Structure, dynamics and interactions of the N-terminal domain of the androgen receptor

Eva De Mol

ADVERTIMENT. La consulta d'aquesta tesi queda condicionada a l'acceptació de les següents condicions d'ús: La difusió d'aquesta tesi per mitjà del servei TDX (**www.tdx.cat**) i a través del Dipòsit Digital de la UB (**diposit.ub.edu**) ha estat autoritzada pels titulars dels drets de propietat intel·lectual únicament per a usos privats emmarcats en activitats d'investigació i docència. No s'autoritza la seva reproducció amb finalitats de lucre ni la seva difusió i posada a disposició des d'un lloc aliè al servei TDX ni al Dipòsit Digital de la UB. No s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX o al Dipòsit Digital de la UB (framing). Aquesta reserva de drets afecta tant al resum de presentació de la tesi com als seus continguts. En la utilització o cita de parts de la tesi és obligat indicar el nom de la persona autora.

ADVERTENCIA. La consulta de esta tesis queda condicionada a la aceptación de las siguientes condiciones de uso: La difusión de esta tesis por medio del servicio TDR (**www.tdx.cat**) y a través del Repositorio Digital de la UB (**diposit.ub.edu**) ha sido autorizada por los titulares de los derechos de propiedad intelectual únicamente para usos privados enmarcados en actividades de investigación y docencia. No se autoriza su reproducción con finalidades de lucro ni su difusión y puesta a disposición desde un sitio ajeno al servicio TDR o al Repositorio Digital de la UB. No se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR o al Repositorio Digital de la UB (framing). Esta reserva de derechos afecta tanto al resumen de presentación de la tesis como a sus contenidos. En la utilización o cita de partes de la tesis es obligado indicar el nombre de la persona autora.

WARNING. On having consulted this thesis you're accepting the following use conditions: Spreading this thesis by the TDX (**www.tdx.cat**) service and by the UB Digital Repository (**diposit.ub.edu**) has been authorized by the titular of the intellectual property rights only for private uses placed in investigation and teaching activities. Reproduction with lucrative aims is not authorized nor its spreading and availability from a site foreign to the TDX service or to the UB Digital Repository. Introducing its content in a window or frame foreign to the TDX service or to the UB Digital Repository is not authorized (framing). Those rights affect to the presentation summary of the thesis as well as to its contents. In the using or citation of parts of the thesis it's obliged to indicate the name of the author.

Universitat de Barcelona

Facultat de Farmàcia

Institut de Recerca Biomèdica, Barcelona

Structure, dynamics and interactions of the N-terminal domain of the androgen receptor

Eva De Mol

2014

UNIVERSITAT DE BARCELONA

FACULTAT DE FARMÀCIA

INSTITUT DE RECERCA BIOMÈDICA, BARCELONA

PROGRAMA DE DOCTORAT EN BIOMEDICINA

STRUCTURE, DYNAMICS AND INTERACTIONS OF THE N-TERMINAL DOMAIN OF THE ANDROGEN RECEPTOR

Memòria presentada per Eva De Mol per optar al títol de doctor per la Universitat de Barcelona

TESI DUTA A TERME A L' INSTITUT DE RECERCA BIOMÈDICA, BARCELONA

DIRECTOR I TUTOR

Xavier Salvatella Giralt

DOCTORANDA

Eva De Mol

Eva De Mol, 2014

Contents

Acknowledgements

Many people have contributed to the work I present in this PhD thesis and I would like to express my gratitude to all of them for their help, guidance and support throughout these last couple of years.

In the first place I want to thank my thesis supervisor, Dr. Xavier Salvatella. Xavier, thank you for giving me the opportunity to come to your lab and to work on a project that has fascinated me from the beginning. I think we are now one step closer to unravelling the mysteries of this protein and I have enjoyed the process of getting to this point. Thank you for introducing me into the world of protein NMR and for the countless hours you have spent with me to teach me and to discuss data with me.

I am grateful to la Caixa, for awarding me with a fellowship to carry out this PhD research and for providing me with the means to attend numerous courses and conferences throughout the years, from which I have learned a lot.

I would like to thank the former and current members from our group, both for collaborating on the projects described in this thesis and for sharing day to day life in the lab. I would like to acknowledge Dr. Robert Bryn Fenwick for assigning AF1*c and AF1*, Dr. Carlos W. Bertoncini, Dr. Marianela Masin and Dr. Santiago Esteban Martín for their collaboration on the RAP74 project and Dr. Christopher Phang for his collaboration on the EPI-001 project. Dr. Carlos W. Bertoncini cloned the RAP74NMR construct and assigned its protein backbone, as well as produced RAP74NMR protein for some of the NMR experiments conducted. He further analyzed the T_2 relaxation data of AF1*. Dr. Marianela Masin cloned the mutants of both full-length AR (for the luciferase-based reporter gene assays) and of AF1*c for the interaction study with RAP74. Dr. Santiago Esteban Martín performed molecular

dynamics simulations to complement the experimental work on the interaction of AR and RAP74 and performed the global fitting for all the titrations. Dr. Christopher Phang prepared the samples for the study of the covalent reaction between AF1* and EPI-001 and carried out the analysis of the subsequent MS experiments in collaboration with the MS core facility at IRB Barcelona. He further analyzed the STD data.

Bryn, thank you for teaching me how to produce protein and how to analyze NMR data. A special thank you for assisting me with the assignment of AF1*b, for your patience during the process and for explaining me how to solve the "puzzle" step by step. Thanks to Dr. Maria Mossuto for teaching me all the practical things in the lab and showing me how to measure CD. Carlos, thank you for assisting me with the acquisition of numerous NMR experiments, for helping to interpret my data, for your suggestions and ideas that have helped a lot to advance on the various projects I was working on. Marianela, thank you for shedding light on the more biological aspects of the AR and RAP74 interaction. Chris, thank you for contributing so much to the EPI-001 project. Santi, thank you for everything, for teaching me to critically analyze my data and for helping me to grow as a person. Of course, I also want to thank the other students in the lab, with whom I have spent so many hours working side by side. Giulio, Anna, Bahar, thank you for all the nice moments we have shared together and simply for making these last couple of years an unforgettable part of my life. To Jomi, our lab technician, a very special thank you. Without your help to prepare proteins it would not have been possible to do many of the experiments in this PhD thesis. Thank you for caring about my proteins the same way I do and for sometimes doing the impossible to get the protein on time for the NMR. Also thank you to Claudio, Victor and Busra.

A special thank you also for Dr. Margarida Gairí. Marga, thank you for helping me with all my NMR doubts, no matter how tiny they were, for making me confident to record 3D experiments on my own, and to believe in me. I would also like to thank Dr. Jesús García for assistance with NMR acquisition and processing, in particular for the STD experiments, and for useful discussions of my results in general. Thank you also to Dr. Miguel Feliz for introducing me to STD NMR.

I would like to thank our collaborators. Dr. Eva Estébanez Perpiñá for inviting me to her lab to perform the cloning of the AR constructs together with Dr. Victor Buzón. Prof. McEwan and Dr. Kate Watson and Dr. Irene Hunter from his lab who carried out the luciferase-based reporter gene assays. Prof. Antoni Riera and Àlex de la Fuente in his lab who synthesized the four stereoisomers of EPI-001. Dr. Philipp Selenko for useful discussions on the phosphorylation pattern of AR and for the preliminary experiments already performed in his lab.

I would also like to thank Dr. Nick Berrow and his team from the protein expression core facility, as well as Dr. Marta Vilaseca and her team from the mass spectrometry core facility. In particular Dr. Gonçalo Costa, who performed the trypsin digestion experiments and the following MS analysis for the EPI-001 project.

None of this would have been possible without the support of my family and the unconditional love from Nacho. To my family, a very special thank you for supporting my decision to come to Barcelona for my PhD and for always being there for me even when we were several hundreds of kilometers apart. Nacho, thank you for everything, for believing in me, for supporting me, for feeding me and above all for loving me. A special thank you also to Julio, Pilar, Elena and David, Cristina and Alberto, Chiqui and Paqui, Pilar and Rolphi, for your support these last couple of months.

I could not have reached the point I am at today without the help of all of you. Thank you. I wish you all the best and hope we stay in touch.

List of Abbreviations

AABUF average area buried upon folding

ACBP acyl coenzyme A binding protein

ACTR activator for thyroid hormone and retinoid receptors

ADT androgen deprivation therapy

AF1 activation function 1

AF2 activation function 2

AIS androgen insensitivity syndrome

AR androgen receptor

ARE androgen response element

BF-3 binding function 3

BMRB BioMagResBank

CBP CREB-binding protein

CD circular dichroism

centFCP1 central part of FCP1, residues 579–600

centFCP1-PO4 central part of FCP1, residues 579–600, with phosphorylated T584

clARE classical androgen response element

CREB cyclic-AMP response element binding protein

CRPC castration-resistant prostate cancer

cryo-EM cryo-electron microscopy

CTD C-terminal domain

CTE C-terminal extension

cterFCP1 C-terminal part of FCP1, residues 937–961

cterRAP74 C-terminal part of RAP74

CV column volume

DBD DNA-binding domain

DHT dihydrotestosterone

DMSO dimethyl sulfoxide

DSS-d_6 3-(trimethylsilyl)-1-propanesulfonic acid-d_6 sodium salt

EMA European Medicines Agency

ER estrogen receptor

FCP1 transcription factor IIF (TFIIF)-associated C-terminal domain (CTD) phosphatase

FDA US Food and Drug Administration

FPLC fast protein liquid chromatography

FTIR Fourier transform infrared spectroscopy

GR glucocorticoid receptor

GTF general transcription factor

GTM general transcription machinery

HDAC histone deacetylase

HPLC high-performance liquid chromatography

HSQC heteronuclear single quantum correlation

HTH helix-turn-helix

ID intrinsically disordered, intrinsic disorder

IDP intrinsically disordered protein

IDR intrinsically disordered region

IEC ion exchange chromatography

IL-6 interleukin-6

IPTG isopropyl β-D-1-thiogalactopyranoside

JAK Janus kinase

KID kinase-inducible transcriptional activation domain

KIX the KID-binding domain of CBP

LBD ligand-binding domain

MAGE-11 melanoma antigen gene protein 11

MALDI-TOF MS matrix-assisted laser desorption/ionization-time of flight mass spectrometry

MAPK mitogen-activated protein kinase

MoRFs molecular recognition features

MR mineralocorticoid receptor

MS mass spectrometry

MST microscale thermophoresis

MTSL (1-oxyl-2,2,5,5-tetramethyl-Δ3-pyrroline-3-methyl) methanethiosulfonate

NCBD nuclear receptor co-activator-binding domain

NCoR nuclear receptor co-repressor

NES nuclear export signal

NLS nuclear localization signal

NMR nuclear magnetic resonance

NR nuclear receptor

NTD N-terminal domain

OD optical density

p-TEFb positive transcription elongation factor b

p/CAF p300/CBP-associated factor

PCa prostate cancer

PDB Protein Data Bank

PECF protein expression core facility

PIC protease inhibitor cocktail

PIC pre-initiation complex

PMSF phenylmethylsulfonyl fluoride

pol II RNA polymerase II

PPAR peroxisome proliferator-activated receptor

PR progesterone receptor

PRE paramagnetic relaxation enhancement

PSA prostate-specific antigen

PTM post-translational modification

RAP30 RNA polymerase II-associated protein 30

RAP74 RNA polymerase II-associated protein 74

RAR retinoic acid receptor

RPB1 largest subunit of RNA polymerase II

RPB2 second largest subunit of RNA polymerase II

RTK receptor tyrosine kinase

RXR retinoic X receptor

SBMA spinal and bulbar muscular atrophy

SDS-PAGE sodium dodecyl sulfate polyacrylamide gel electrophoresis

SEC size exclusion chromatography

selARE selective androgen response element

SMRT silencing mediator for retinoid and thyroid hormone receptors

SR steroid receptor

SRC steroid receptor co-activator

STAT3 signal transducer and activator of transcription 3

STD saturation transfer difference

SW sweep width

Tau-1 transcription activation unit 1

Tau-5 transcription activation unit 5

TBP TATA-binding protein

TCEP tris(2-carboxyethyl)phosphine hydrochloride

TFE trifluoroethanol

TMAO trimethylamine N-oxide

TR thyroid hormone receptor

VDR vitamin D receptor

WT wild type

1 Introduction

1.1 Prostate cancer and the androgen receptor

Prostate cancer (PCa) is the second most common cancer in men, after lung cancer (see Fig. 1.1).[1] An estimated 1.1 million men worldwide were diagnosed with PCa in 2012.[1] With an estimated 307,000 deaths in 2012, PCa is the fifth leading cause of death from cancer in men (6.6% of the total men deaths, after lung, liver, stomach and colorectal cancer).[1] In European men, PCa is the most common cancer (22.8%) and the third most common cancer-related cause of death (9.5%), after lung and colorectal cancer.[1] PCa killed over 90,000 European men in 2012.[1] About one man in seven will be diagnosed with PCa during his lifetime.[2] Even though it can be a devastating disease, most men diagnosed with PCa do not die from it.[1]

PCa growth depends on the ratio of cells proliferating to those dying. Androgens, the male sex hormones, are the main regulator of this ratio both by stimulating proliferation and inhibiting apoptosis. Therefore, PCa depends on androgens for growth and survival [1]. The androgen receptor (AR) is a nuclear hormone receptor that is activated by androgenic hormones and the protein through which the physiological effects of androgens are mediated [2]. The AR is necessary for normal prostate development, growth and physiology, but also plays a major role in PCa disease progression. Like other members of the nuclear receptor (NR) superfamily, the AR is composed of four functional domains: an N-terminal transactivation domain (NTD), a DNA-binding domain (DBD) containing two Zn fingers, a hinge region (H), and a ligand-binding domain (LBD) that binds testosterone or dihydrotestosterone (DHT) (see Fig. 1.2) [3].

[1] Source: http://globocan.iarc.fr

[2] Source: http://www.cancer.org/cancer/prostatecancer/detailedguide/prostate-cancer-key-statistics

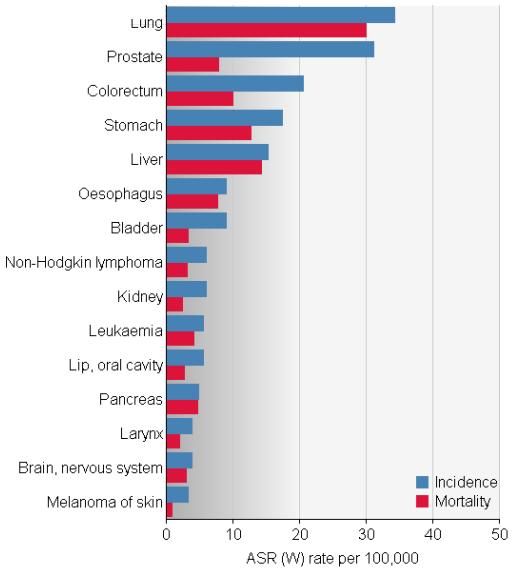


Figure 1.1: Estimated age-standardised incidence and mortality rates for cancer in men, worldwide (source: http://globocan.iarc.fr).

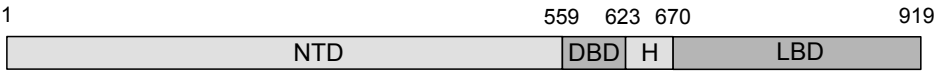


Figure 1.2: The domain structure of the AR. The AR contains an N-terminal transactivation domain (NTD), a DNA-binding domain (DBD), a flexible hinge region (H), and a ligand-binding domain (LBD). Numbers indicate the residues that correspond to the different domains.

The AR presents two well-defined polymorphic tracts, a polyQ tract near the N-terminus and a polyG tract near the center of the sequence, that can due to their variability generate uncertainties in the numbering of residues. Throughout this thesis the numbering used by the Uniprot database is used, which attributes lengths of 21 and 24 to these two tracts, respectively.

1.1.1 Mechanism of androgen action via the AR

The androgen signaling axis, mediated via the AR, is shown in figure 1.3 [1]. Testosterone, the main circulating androgen, is secreted primarily by the testes and circulates in the blood bound to albumin or to sex-hormone-binding globulin (SHBG), with a small fraction freely dissolved in serum. Testosterone in its unbound form can enter prostate cells by passive diffusion. In the cytoplasm of these cells, testosterone is converted to DHT by the enzyme 5α-reductase. DHT has a five- to tenfold higher affinity for AR compared to testosterone, and is consequently the primary androgen binding to AR [1, 4]. In the absence of hormone, the AR is present in the cytoplasm in a complex with molecular chaperones, such as heat shock proteins Hsp90, Hsp70 and Hsp40, and other proteins. This protects the unliganded AR from degradation. Binding of testosterone or DHT to the LBD causes a conformational change that leads to the dissociation of the complex, induces the formation of a binding site for co-regulators in the LBD and allows the interaction with kinases that phosphorylate several serine residues [5, 6]. Furthermore, ligand-binding exposes the nuclear localization signal, thus allowing translocation of the AR to the nucleus, where it homodimerizes [7]. In the nucleus, the AR recognizes and binds to specific elements on the DNA, called androgen response elements (AREs), located in the promoter and/or enhancer regions of target genes. The activated DNA-bound AR recruits co-regulatory proteins, co-activators or co-repressors, to the AR complex. The co-activators allow relaxation of the chromatin and facilitate the interaction of the AR complex with the general transcription machinery to stimulate target gene transcription [8]. Co-repressors prevent the interaction with the general transcription machinery, inhibiting the transcription of target genes. Many AR target genes have been identified [9] and many are involved in growth and survival. Prostate-specific antigen (PSA) is one such AR-regulated gene and is a clinically important marker in the detection of PCa [1].

1.1.2 Current treatments for PCa

The treatment for PCa depends on the stage of the disease. Localized PCa and early stage disease may be successfully treated with surgery (prostatectomy, i.e. surgical

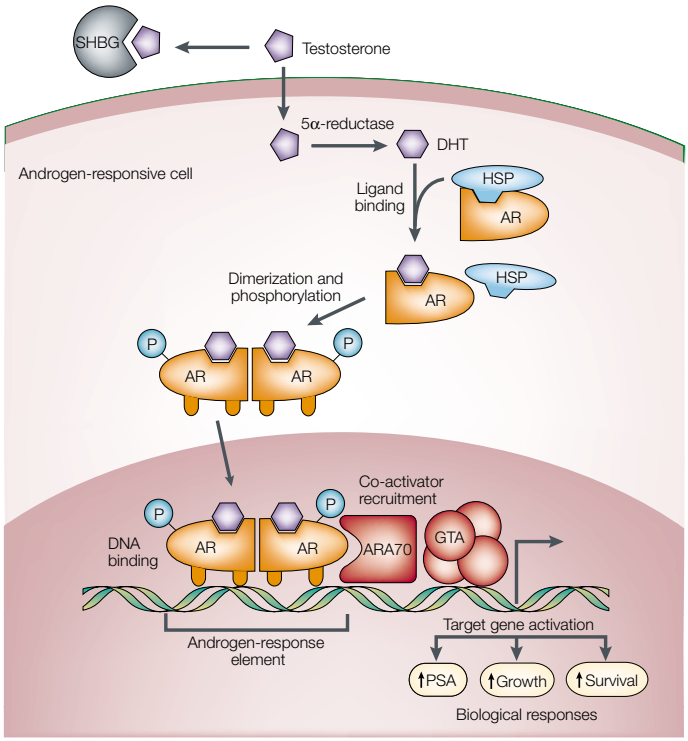


Figure 1.3: The mechanism by which androgens enter healthy or cancerous prostate cells and induce transcription of genes that are involved in normal development and reproduction in males or in the progression of PCa. From [1]. It should be noted that, according to current models, homodimerization of the AR takes place in the nucleus [7]. SHBG: sex-hormone-binding globulin, DHT: dihydrotestosterone, HSP: heat shock protein, AR: androgen receptor, P: phosphorylation, ARA-70: AR-associated protein 70, a co-activator of AR, GTA: general transcription apparatus and PSA: prostate-specific antigen.

removal of (part of) the prostate gland) and/or radiotherapy, whereas the standard form of treatment for advanced or metastatic disease has been androgen deprivation therapy (ADT), also called androgen ablation therapy, via surgical or chemical castration for the last 70 years [10]. ADT was developed based on the early recognition that PCa progression is dependent on androgens and the observation of Huggins and Hodges in the early 1940s that removal of androgens caused advanced PCa to regress [11].

ADT is aimed at reducing the circulating levels of testosterone in the blood to block the androgen signaling axis described in figure 1.3. Testosterone is primarily produced in the testes and to a lesser extent in the adrenal glands. Although surgical castration by orchiectomy (surgical removal of the testes) effectively depletes androgens, chemical castration by the administration of luteinizing hormone-releasing hormone (LHRH) antagonists (cetrorelix and degarelix) or agonists (goserelin and leuprolide), is a widely used alternative. These agents suppress the secretion of luteinizing hormone (LH) and follicle-stimulating hormone (FSH), which are the hormones that stimulate the testes to produce and secrete testosterone [1, 12]. Importantly, this therapy inhibits the production of testosterone by the testes, but does not affect the testosterone synthesis by the adrenal glands.

Often ADT is combined with the administration of an AR antagonist, an anti-androgen that competes with the endogenous ligand of AR and preferentially binds to the protein but does not activate it. This combined therapy is referred to as maximal androgen blockade [13]. AR antagonists also prevent androgens produced in the adrenal glands from binding to AR and prevent the initial flare-up phenomenon of LHRH agonists that have an initial stimulatory action on steroid production [14–16]. Anti-androgens that are used in combination with ADT include flutamide, hydroxyflutamide, nilutamide and bicalutamide (see Fig. 1.4) [17, 18]. In about 15–30% of patients treated with these anti-androgens a so-called "anti-androgen withdrawal syndrome" is observed over time [17]. This is characterized by a clinical worsening of patients when treated with the anti-androgen, which is improved when the administration of the anti-androgen is stopped. This has been linked to the acquirement of mutations by the AR as a response to selective pressure of the treatment, that allows the antagonist to act as an agonist and stimulate the AR [1]. Anti-androgen withdrawal has been observed for flutamide [19, 20], hydroxyflutamide [21], nilutamide [22, 23] and bicalutamide [24, 25]. When observed in a patient, treatment with the corresponding anti-androgen must be stopped.

Although the majority of the patients shows a positive initial response to ADT, the disease progresses inevitably to a state that is resistant to androgen ablation, generally referred to as castration-resistant prostate cancer (CRPC). In most cases, patients

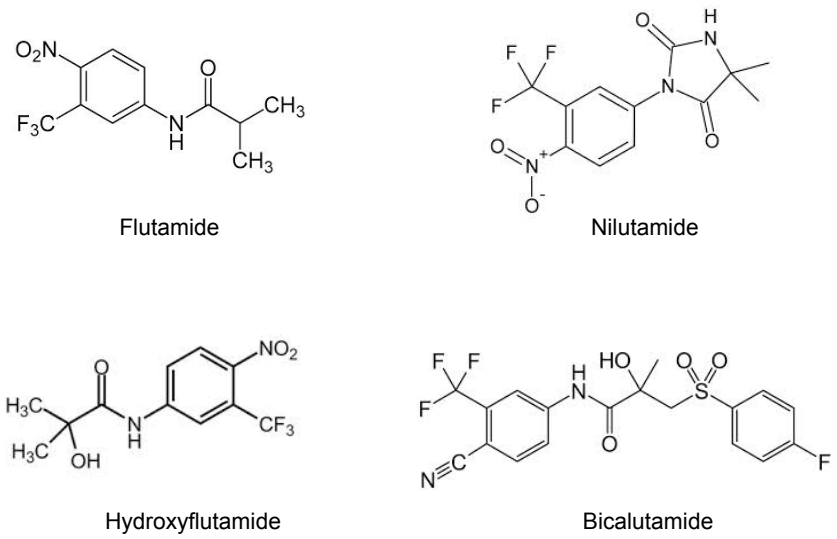


Figure 1.4: Chemical structure of anti-androgens flutamide, hydroxyflutamide, nilutamide and bicalutamide.

develop therapy resistance within about 12 to 24 months [26]. During the progression to CRPC, the tumors develop several mechanisms to maintain the androgen signaling axis in spite of the conditions of ADT in combination with anti-androgens, which allows them to survive. These mechanisms include AR mutations that allow hypersensitization to low levels of androgens or promiscuous activation by other ligands such as other steroid hormones or even antagonists, AR amplification, alterations in the expression of co-regulators, expression of constitutively active splice variants and cell signaling cross-talk, and will be discussed in more detail in the next section. Consequently, the two main therapeutic strategies, androgen synthesis inhibition and inhibition of the AR by competitive binding of AR antagonists, eventually fail. Rising PSA levels are indicative of renewed tumor growth and herald the transition of the disease to a state that is insensitive to ADT. Once PCa reaches this castration-resistant stage of the disease, treatment options are limited and although they prolong the mean life span of patients with some months, they mainly serve to improve their quality of life. At present no effective treatment is available for CRPC, rendering this stage of the disease lethal [27]. The mean survival of patients after developing CRPC is 12 months [16]. The standard first-line treatment after castration-resistance emerges is a combination of the chemotherapeutic agent docetaxel and prednisone, a corticosteroid [28]. This treatment increases the overall survival of patients for about two to three months, reduces pain and improves their quality of life [17]. However, patients will usually experience disease progression due to inherent or acquired resistance to docetaxel [28–30] and succumb to the disease.

Before 2010, second-line treatment options were limited for patients whose disease had progressed during or after receiving docetaxel treatment [28]. At this stage of the disease, no treatments were available that showed an increase in overall survival. In recent years, however, additional treatment options have become available for patients that have developed resistance to docetaxel. Significant advances in the field have led to the approval of five new agents in the last four years for the treatment of metastatic CRPC, all with different mechanisms of action: cabazitaxel, abiraterone acetate, enzalutamide, radium-223 chloride and sipuleucel-T [31, 32]. Even though all of these agents have shown to improve overall survival and the quality of life of the patients, unfortunately none of them has been found to cure the disease. Only two of these agents directly target the androgen signaling axis, either by inhibiting androgen synthesis (abiraterone acetate) or by blocking the AR directly (enzalutamide). Cabazitaxel indirectly affects this pathway by inhibiting AR nuclear translocation through affecting microtubule stabilization [33]. The remaining therapeutics either directly target bone metastases (radium-223 chloride) or trigger an immune response (sipuleucel-T).

Cabazitaxel is a chemotherapeutic agent that has shown activity against tumor cells that are resistant to docetaxel both *in vitro* and *in vivo* in preclinical studies [28, 34]. It was demonstrated to have an overall survival benefit in patients with metastatic CRPC who have previously been treated with docetaxel and is indicated for the treatment of these patients. It was approved by the US Food and Drug Administration (FDA) in 2010 and by the European Medicines Agency (EMA) in 2011 [13, 28].

Abiraterone acetate (Zytiga®) is an oral drug that after uptake is converted to the active compound abiraterone (see Fig. 1.5) [13]. It is a selective irreversible inhibitor of Cytochrome P450 17 (CYP17), a key enzyme in the production of androgens, estrogens and glucocorticoids within the adrenal steroid hormone synthetic pathway. [13, 35] Consequently, it is a potent inhibitor of the production of testosterone by the adrenal glands. As mentioned before, part of the testosterone circulating in the blood is produced in the adrenal glands and standard ADT only inhibits testosterone synthesis by the testes. Blocking the adrenal synthetic pathway is therefore complementary to ADT and CYP17 inhibition has been shown to result in a further decrease of androgen levels both circulating in the blood and in the tumors of CRPC patients [36]. Furthermore, a number of studies have demonstrated expression of CYP17 in castration-resistant prostate tumors [37, 38], indicating that this enzyme is also involved in intratumoral androgen synthesis. Abiraterone acetate therefore inhibits both the production of testosterone by the adrenal glands and intratumoral testosterone production [39]. Treatment with abiraterone acetate requires concomitant steroids ad-

ministration [40] as a consequence of the interrupted synthesis pathway of other steroid hormones like estrogens and glucocorticoids by CYP17 inhibition. In 2011, abiraterone acetate was approved by the FDA in combination with prednisone for the treatment of CRPC patients who have received prior docetaxel chemotherapy [13]. In addition, it was approved for treatment in the pre-chemotherapy phase of CRPC in 2013 [13].

Although clinical responses to abiraterone have been impressive, not all men respond, the duration of the response is variable and, importantly, a majority of patients treated with abiraterone eventually becomes refractory as indicated by a rising PSA titer [39]. Even though the mechanisms determining resistance to abiraterone have not been fully elucidated, emerging data from preclinical and clinical studies suggest several possibilities [39]. One of these is the development of AR mutations which allow AR activation by exogenous corticosteroids, that are co-administered to reduce the side effects of abiraterone, or steroid precursors upstream of CYP17 [39, 41]. Inhibition of CYP17 is associated with a rise in circulating levels of upstream progesterone precursors [42, 43], which have been shown to activate AR bearing certain mutations [44, 45]. In addition, exogenous glucocorticoids can activate the glucocorticoid receptor (GR), and it has been shown that signaling via the GR can activate AR-regulated genes in the absence of androgens [46]. In addition, the expression of constitutively active splice variants of the AR is increased following treatment with abiraterone acetate, which has been proposed to contribute to therapy failure [47].

Enzalutamide (also known as MDV3100, Xtandi®) is a selective next-generation AR antagonist that is also administered as an oral drug (see Fig. 1.5). It was rationally designed to overcome the antagonist-to-agonist conversion of first-generation AR antagonists [48]. It potently binds the AR due to its high affinity and selectivity, thereby preventing the binding of testosterone or DHT to the protein. Consequently, AR signaling is efficiently suppressed as this blocks androgen-induced nuclear translocation of the AR, binding to DNA and co-activator recruitment to form a transactivation competent complex [48]. Moreover, enzalutamide is able to bind to and inhibit not only wild type AR but also mutant ARs that can be activated by other anti-androgens, such as a bicalutamide-resistant AR mutant [32, 45]. Two large phase III clinical studies, AFFIRM and PREVAIL, the latter still ongoing, were initiated to test enzalutamide efficacy in patients who received prior docetaxel therapy and in chemotherapy-naive patients, respectively [13]. Based on the AFFIRM study, enzalutamide in the metastatic CRPC post-docetaxel setting received FDA approval in August 2012 and EMA approval in April 2013 [32].

In spite of its superiority over first-generation AR antagonists, resistance to enzalutamide treatment has already been reported. Notably, a recent study revealed a new

mutation in the AR LBD (F876L) that conferred resistance to enzalutamide by converting it into an AR agonist [49–51]. This illustrates that in spite of the efforts to optimize next-generation AR antagonists, the AR is expected to acquire new mutations that will allow its stimulation by novel anti-androgens. Furthermore, other possible mechanisms conferring enzalutamide resistance have already been reported, including an increase in constitutively active AR splice variants [52, 53] and cross-talk with other cellular pathways [54]. These mechanisms of resistance will be discussed in more detail in the next section.

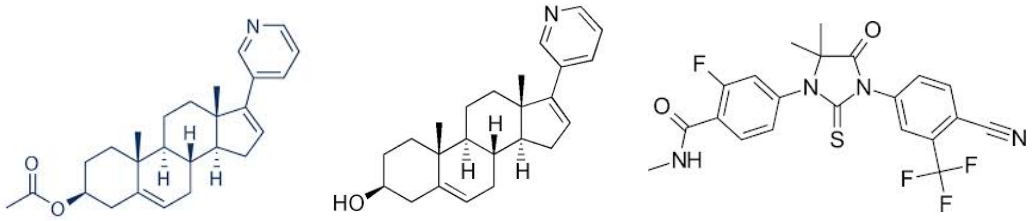


Figure 1.5: Chemical structure of abiraterone acetate, abiraterone and enzalutamide.

Radium-223 chloride (Alpharadin®) is a radiopharmaceutical that is used in CRPC patients to reduce bone pain [12]. Bone metastases represent the most common type of metastases in PCa, and consequently the bone is an important therapeutic target for advanced PCa [13]. Alpharadin is an intravenous formulation of radium-223 that is an α emitter and a calcium mimetic [32, 55]. Because of its properties as a calcium mimetic it is taken up into the bone, especially in osteoblastic metastases, where it delivers high-energy, short-range α irradiation inducing DNA double strand breaks [32]. The recently reported ALSYMPCA phase III study has shown significant overall survival benefit in favor of Alpharadin treatment in CRPC patients [55]. Based on the results from this trial, Alpharadin has recently been approved for patients with metastatic CRPC with metastases to bone but not other organs by the FDA and is under review by the EMA [32].

Sipuleucel-T (Provenge®) is an immunotherapeutic compound designed to stimulate an immune response against CRPC [17]. It has minimal side effects and can be used to treat PCa in combination with anti-androgens or with cytotoxic therapies allowing for earlier treatment of CRPC [56–58]. Sipuleucel-T was approved by the FDA for the treatment of asymptomatic or minimally symptomatic CRPC [13]. The limited availability, especially outside the US, and the high therapy costs currently restrict Sipuleucel-T as a standard treatment option for CRPC [13].

In addition to these drugs, other agents are being used in the treatment of CRPC.

These therapeutics target more general aspects of cancer biology, such as inhibitors of angiogenesis, or target other cellular pathways that are involved in the transition of PCa to CRPC, such as growth factor receptor inhibitors and kinases inhibitors [12, 13].

Most of the recently approved agents were evaluated for their use in metastatic CRPC for patients who had already been treated with docetaxel and developed resistance against it. It is however possible that some of them also have a beneficial effect for patients at earlier stages of the disease. Such an effect has already been shown for abiraterone acetate which has led to its approval as a drug prior to chemotherapy treatment. Currently, most of the other agents are under preclinical or clinical evaluation to assess their potential use in earlier stages of the disease [32].

In addition, new compounds are being developed to circumvent the resistance mechanisms already reported for some of these new agents, including abiraterone acetate and enzalutamide. Many of them show promising results in preclinical trials and are currently being further evaluated [32].

Taken together, only minimal improvement has been achieved in overall survival of metastatic CRPC patients and no strategy to cure the disease has been found despite recent advances in the field. In addition, resistance to the two novel agents that target the androgen signaling axis directly, abiraterone acetate and enzalutamide, has already emerged. These new agents are based on the same two therapeutic strategies as the first-generation therapeutics, i.e. androgen synthesis inhibition and inhibition of the AR by competitive binding of AR antagonists. Although they are superior to the first-generation agents, the emerging resistance mechanisms against them appear to be in part similar to those acquired against first-generation drugs. This suggests that these therapeutic strategies will eventually fail. This notion underlines the need to gain a better understanding of the mechanisms leading to aberrant AR activation in CRPC patients to develop novel therapeutic strategies.

1.1.3 Progression of PCa to CRPC

Although ADT efficiently reduces the serum levels of testosterone and AR antagonists efficiently compete for AR binding with its endogenous ligand, PCa cells inevitably progress to a castration-resistant stage of the disease that is currently incurable. Tumors adapt to the conditions of androgen deprivation therapy and the presence of AR antagonists through a multitude of mechanisms that lead to hypersensitization of the AR to very low levels of its endogenous hormone, allow promiscuous AR activation by binding of alternative hormones or even antagonists, enable ligand-independent AR activation or completely bypass the androgen signaling axis by activating parallel

signaling pathways important for cell survival. Combined, these strategies maintain AR activity and transcription of its target genes despite maximal androgen ablation therapy.

In this section, a summary of the current understanding of the molecular mechanisms underlying the progression of PCa to a castration-resistant stage is provided. PCa biology is highly complex and the transition of PCa to castration-resistant growth involves many different mechanisms that are likely not mutually exclusive. Various mechanisms may be active within one cell population, and possibly act synergistically to maintain AR transcriptional activity under conditions of androgen ablation.

AR gene amplification and protein overexpression *AR* gene amplification and/or overexpression of the AR have been shown to occur more frequently in CRPC than in benign tissue and hormone-naive cancers [59–62]. In approximately 20–30% of tumors that become castration-resistant after androgen ablation the *AR* gene was found to be amplified and most tumors showed a corresponding increase in AR protein expression [63, 64]. Higher AR protein expression as a result of *AR* gene amplification sensitizes PCa cells to low levels of androgens and may enable the AR to be activated [18, 65]. Notably, *AR* gene amplification does not always result in increased AR protein expression [62]. Furthermore, an increase in AR mRNA levels without gene amplification has also been described [66].

Intratumoral synthesis of androgens and enhanced steroid uptake Even though ADT efficiently reduces the serum levels of testosterone, the intratumoral androgen levels are generally not equally reduced and can remain sufficient to activate the AR [39]. This is related to an increased local production of androgens in PCa cells following ADT to compensate for the overall decline in circulating testosterone. On the one hand, PCa cells can locally convert adrenal precursors to more active androgens like testosterone and DHT via their own enzymes, many of which are overexpressed in CRPC [67, 68]. On the other hand, an increase in the 5α-reductase activity within the PCa cells enhances the rate of conversion of testosterone to the more potent hormone DHT [1]. Downregulation of androgen-metabolizing enzymes has also been proposed as an additional mechanism by which prostate tumors can maintain high levels of potent androgens [13, 69]. In addition, several organic acid transporters are significantly overexpressed in metastatic CRPC tissues compared to primary cancers [70]. These proteins are involved in the transport of various steroid hormones across cell membranes, thus their overexpression facilitates the uptake of androgen precursors into prostate tumors [68]. These combined approaches facilitate continued AR signaling

even with significantly lower levels of serum testosterone.

AR mutations AR mutations are more frequently observed in CRPC than in untreated, hormone-sensitive PCa, and are often associated with gain-of-function [71, 72]. The higher incidence of AR mutations after ADT suggests an adaptive response to the treatment leading to clonal selection of some AR mutants [21, 73]. Mutations of the AR gene most frequently localize to exons that encode the LBD (ca. 49%), followed by the NTD (ca. 40%), the DBD (ca. 7%) and the hinge (ca. 2%). Mutations are rarely found in untranslated regions [71].

Several mutations in the LBD have been described to broaden the ligand specificity of the AR, allowing inappropriate activation by various non-androgenic steroids and AR antagonists [1]. The T877A mutation was identified in LNCaP prostate cancer cells [74], and allows other steroid hormones, like progestins and estrogens, as well as the AR antagonist flutamide to bind to the LBD and activate the AR [1, 75]. Furthermore, it was shown that the T877A mutation is more frequently present in patients that have been treated with flutamide, indicating that this mutation occurs in response to strong selective pressure from flutamide treatment [21]. Since this mutation converts flutamide from an AR antagonist into an AR agonist, it is directly related to the development of flutamide withdrawal syndrome in some patients [19]. In these patients, treatment with flutamide causes clinical worsening whereas its withdrawal results in improvement. As mentioned before, withdrawal syndrome has been observed for other anti-androgens, including hydroxyflutamide, nilutamide, bicalutamide and even enzalutamide, and has been linked to specific mutations in the AR LBD that allow aberrant AR activation by these compounds [49–51].

Another mutation in the LBD, L701H, enhances the binding of other adrenal corticosteroids, particularly the glucocorticoids cortisol and cortisone. In combination with the T877A mutation, the affinity of the AR for glucocorticoids is increased by 300% compared to for the L701H mutation alone [41]. It is therefore likely that physiological levels of circulating cortisol and cortisone would be sufficient to promote tumor growth in patients with this double mutation. The H874Y mutation was also identified in CRPC patients treated with flutamide, and increases ligand promiscuity allowing estradiol, progesterone and hydroxyflutamide to activate transcription in various model systems [76, 77].

Mutations outside of the LBD have been reported to influence nuclear localization, co-regulator binding to AR, AR post-translational modifications (PTMs), protein stability and promoter selectivity [68, 78]. In particular, mutations in the NTD may allow AR interaction with co-activators in the absence of a ligand [79–81]. Mutations

G142V, M523V, G524D and M537V, all in the NTD, were shown to render the AR constitutively active [82].

Changes in the post-translational modifications of the AR Many PTMs are involved in the physiological function of AR in healthy prostate cells [83]. Changes to the post-translational status of AR may influence its transcriptional activity. During the progression of PCa to castration-resistant disease, several factors can influence the PTMs of the AR and thereby affect the androgen signaling axis.

Residues that are usually post-translationally modified can be mutated in CRPC, preventing PTM at that position. Mutations at SUMOylation sites within the NTD, for instance, can lead to decreased SUMO-1 binding to AR and enhanced transcriptional activity of AR [83–86].

In addition, interaction of co-regulators with the AR can lead to AR activation in the absence of ligand by affecting its PTMs. AR acetylation by its co-activator p300 enhances binding of co-activators to AR and inhibits co-repressor binding [87]. Changes in the acetylation status of AR induced by Tat interactive protein 60 kDa (Tip60) have also been proposed to influence the transcription of AR target genes [88].

Finally, cross-talk with other cellular pathways can also modulate PTMs of the AR. AR phosphorylation is frequently observed in CRPC and is often related to cross-talk between AR and several kinase signaling pathways that are upregulated in CRPC [89]. This will be discussed in more detail below.

Alterations in AR co-regulators A large number of co-regulators has been implicated in stimulating AR-dependent transcription as a mechanism of resistance to castration in various studies, essentially either by loss of repressive function or by gain of activation function [90–92]. Co-regulators can alter the ligand specificity of the AR and/or allow AR transactivation at low levels of androgen [93]. Additionally, co-regulators are involved in androgen-independent AR transactivation by affecting AR PTMs or by promoting its nuclear translocation in the absence of androgens. The expression of many AR co-regulators is deregulated in CRPC which can lead to either increased or decreased AR transcriptional activity and contributes to disease progression [94, 95].

The steroid receptor co-activators (SRC) SRC-1, SRC-2 and SRC-3 constitute one important family of AR co-activators. They are overexpressed in PCa and even more so in CRPC [93–95], which may lead to increased AR sensitivity to weak agonists. [95] SRC-3, for instance, facilitates RNA polymerase II recruitment to a distant enhancer

element of the *PSA* gene, resulting in increased PSA levels in response to very low levels of adrenal androgens [96, 97]. Other co-activators known to modulate the ligand specificity of the AR or to allow activation of the AR at low levels of endogenous hormone include CREB-binding protein (CBP), AR-associated protein 70 (ARA70), ARA55, β-catenin and Tip60 [87, 98–101]. CBP and ARA70 are overexpressed in CRPC and enable the anti-androgen hydroxyflutamide to function as an AR agonist [99, 102, 103]. In addition, ARA70 overexpression also allows bicalutamide to function as an AR agonist [103] and enables AR activation at low concentrations of adrenal androgens or estradiol [101, 104]. The co-activators ARA55 and β-catenin also alter AR ligand specificity and enhance AR transactivation in response to estradiol [102]. The expression and nuclear translocation of co-activator Tip60 are increased in response to androgen withdrawal in CRPC cells [105]. Consequently, in CRPC tumors, the predominant nuclear location of Tip60 may mediate increased AR sensitivity to low concentrations of androgens [88]. Tip60 may also affect the acetylation status of the AR and thereby affect its transactivation in the absence of ligand, as mentioned above.

Furthermore, alterations in the relative ratios of the recruitment of co-repressors and co-activators may play a role in the development of CRPC, especially in the presence of AR antagonists [16, 68]. Bicalutamide, for instance, largely increases the interaction between the AR and its co-repressor silencing mediator for retinoid and thyroid hormone receptors (SMRT) [16]. However, upon overexpression of co-activator SRC-1, the AR has been shown to interact preferentially with SRC-1, annulling the inhibiting effect of bicalutamide on AR transactivation [16].

Cell signaling cross-talk In the absence of ligand, the AR can be activated through cross-talk with several cellular signaling pathways. Some of these intracellular cascades are activated by binding of extracellular peptides, like growth factors and cytokines, to their receptors on the cell membrane of prostate cells. These pathways generally contribute to the activation of AR target genes in the healthy prostate by amplifying the AR activity at low levels of hormone. However, most of these pathways are upregulated in CRPC and can therefore substantially contribute to the AR transcriptional activity, even in the absence of ligand [89]. The upregulation of these pathways can often also contribute to the progression of PCa to CRPC, independent of the AR or the AR signaling axis, by increasing cell proliferation and decreasing apoptosis. These bypass mechanisms will be discussed in more detail later on.

In the following a brief overview of selected pathways that are upregulated in CRPC will be given, but it needs to be noted that many other cellular signaling pathways are involved in AR activation in an androgen-deprivation environment. These include

mTOR signaling, the TNF-α pathway, NFκB and endothelin A receptor pathways and several other kinase signaling pathways that lead to AR phosphorylation (protein kinase A, protein kinase C, sarcoma-related kinase (Src)-family of kinases, Ack1 kinases) [12, 68].

Pathways activated by binding of growth factors to receptor tyrosine kinases

Several growth factors, including insulin-like growth factor-1 (IGF-1), keratinocyte growth factor (KGF) and epidermal growth factor (EGF), and their corresponding receptors have been reported to be overexpressed in advanced PCa [106, 107]. They have been implicated in activating the AR in the absence of ligand or in sensitizing AR to subphysiological androgen concentrations [1, 89, 108]. Binding of the growth factors to their receptors, receptor tyrosine kinases (RTKs), initiates complex intracellular signaling cascades, including the mitogen-activated protein kinase (MAPK) and AKT kinase signaling pathways. Activation of these kinase pathways can lead to enhanced AR transcriptional activity by direct phosphorylation of the AR and/or by modulating the binding of AR co-activators [89].

Various members of the MAPK cascade are amplified in CRPC [109], which results in an amplification of the AR activity. Direct phosphorylation of the AR at S515 by MAPK renders AR sensitive to low levels of the synthetic androgen R1881 [110]. In addition, AR phosphorylation by MAPK enhances the recruitment of AR co-regulators and positively modulates the expression of AR target genes [111]. Moreover, it was reported that MAPK may increase AR activity by phosphorylating AR co-activator SRC-1, independent of AR phosphorylation [4, 112]. The interaction of phosphorylated SCR-1 with the AR, in the absence of hormone, results in activation of the AR to the same magnitude as that obtained by DHT [113]. The MAPK signaling pathway has also been associated with the phosphorylation of c-Jun [16]. When phosphorylated, c-Jun can homodimerize (c-Jun/c-Jun) or heterodimerize with c-Fos (c-Jun/c-Fos) to form AP-1. AP-1 can further form a complex with the AR, which prevents the AR from binding to DNA and therefore results in a decreased gene expression [114]. Alternatively, AP-1, when not bound to the AR, can act as a transcription factor and bind to TPA responsive elements (TREs), present in for example the PSA promoter, to induce transcriptional activation [114]. AP-1 has been proposed to influence the progression to CRPC by competing with the AR to alter the expression of androgen-related genes [114].

The AKT signaling cascade is also frequently activated in PCa cells and can modulate AR activity via phosphorylation [115]. AKT, or protein kinase B, can be activated by binding of growth factors to RTKs as well as by inhibition of PTEN. PTEN is a

lipid phosphatase that removes phosphate groups from inositol lipids. This results in inhibition of AKT activity as AKT is activated by these inositol lipids when they are phosphorylated. Mutations inactivating PTEN are common in PCa and PTEN is frequently functionally inactivated in metastatic disease [116, 117], which contributes to the activation of the signaling pathway downstream of AKT. Phosphorylation of the AR at S213 and S791 by AKT in PCa cells positively modulates its transcriptional activity in the absence of ligand [1, 118]. However, the AKT pathway mainly contributes to progression of PCa by affecting the cell cycle progression and its anti-apoptotic activity, independent of the AR, which will be discussed under bypass mechanisms.

AR activation via cross-talk with cytokine signaling pathways
Binding of the cytokine interleukin-6 (IL-6) to its receptor in the cell membrane activates the intracellular JAK/STAT3 pathway (JAK = Janus kinase; STAT3 = signal transducer and activator of transcription 3) and can promote AR transcriptional activity in the absence of androgens (see Fig. 1.6) [89, 113, 119]. Levels of activated STAT3 are significantly higher in CRPC than in androgen-sensitive PCa [120]. STAT3 is phosphorylated by JAK upon IL-6 binding to the receptor, allowing it to dimerize. The activated STAT3 dimer can bind AR in the absence of ligand and promote its translocation to the nucleus [121]. The complex formed by AR and STAT3 can subsequently bind to DNA and activate androgen-regulated transcription [121, 122].

Although the predominant effect of IL-6 is the stimulation of JAK and activation of STAT3, IL-6 can also activate the MAPK and AKT pathways depending on the context [123, 124].

Interaction of AR with other proteins/pathways
The AR might also be activated through interaction with other proteins, including other transcription factors, like ETS oncogenic transcription factors, and pioneer factors, like FOXA1. Such interactions may promote nuclear translocation of the AR in the absence of ligand and facilitate its binding to the DNA, as well as allow for indirect DNA binding. This may also affect the expression and sets of genes associated with biological pathways [125]. AR ubiquitination via the ubiquitin E3 ligase RNF6 enhances AR transcriptional activity and can alter co-factor recruitment. In CRPC, RNF6 is overexpressed and it has been linked to PCa cell growth following ADT [83, 85, 126].

AR splice variants In recent years several splice variants of the AR have been identified, both in healthy tissue and in PCa. These are truncated AR isoforms that have in common that they lack the C-terminal LBD and, in addition, some lack the hinge region or even portions of the DBD (see Fig. 1.7) [127–132]. The majority of these splice variants displays constitutive activity in the absence of ligand, and they

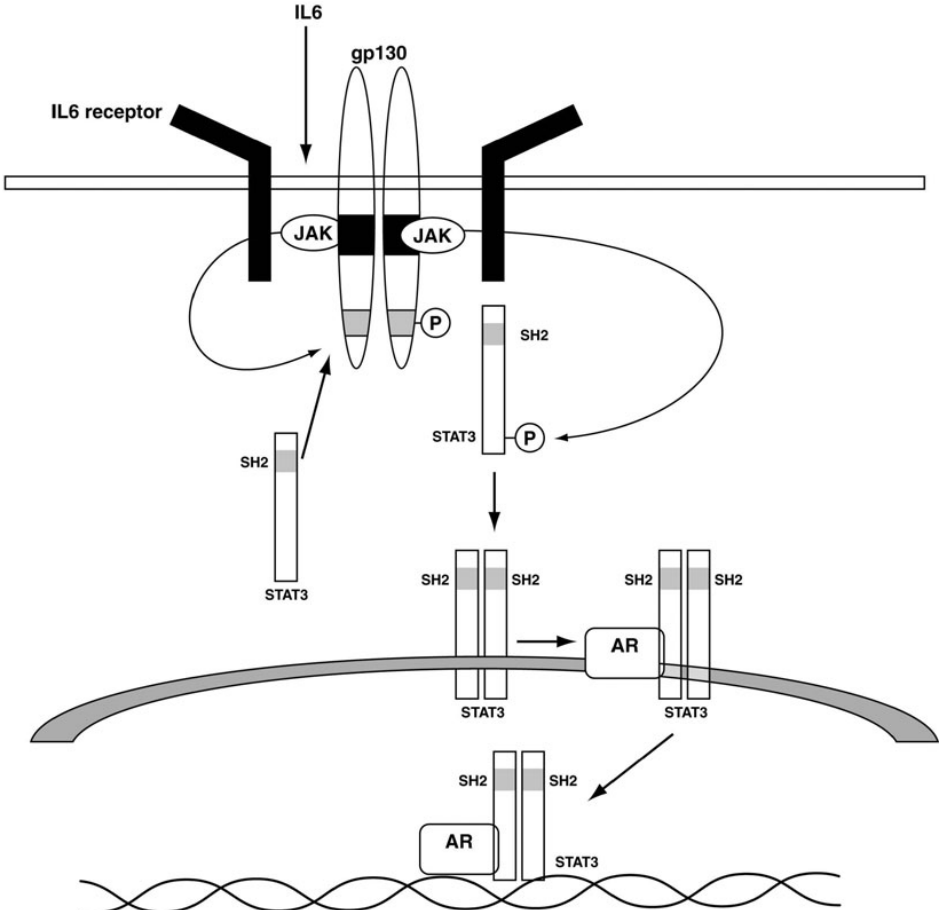


Figure 1.6: Activation of the AR in the absence of ligand by the JAK/STAT3 pathway, activated upon binding of IL-6 to its receptor in the cell membrane. From [16].

have been proposed to play a role in the progression of PCa and resistance to ADT. AR splice variants are upregulated in CRPC compared to hormone-naive cancers and healthy tissue, and can emerge as an adaptive response to therapies targeting the androgen signaling axis, in particular to the recently approved abiraterone acetate and enzalutamide [47, 53, 68, 133]. Since these splice variants lack the LBD, therapeutic strategies that target the binding of hormone to the LBD (either by inhibiting hormone synthesis, such as abiraterone acetate, or by using potent competitors of androgens to bind to the LBD, such as enzalutamide) appear not to have an effect on the constitutive activity of these splice variants.

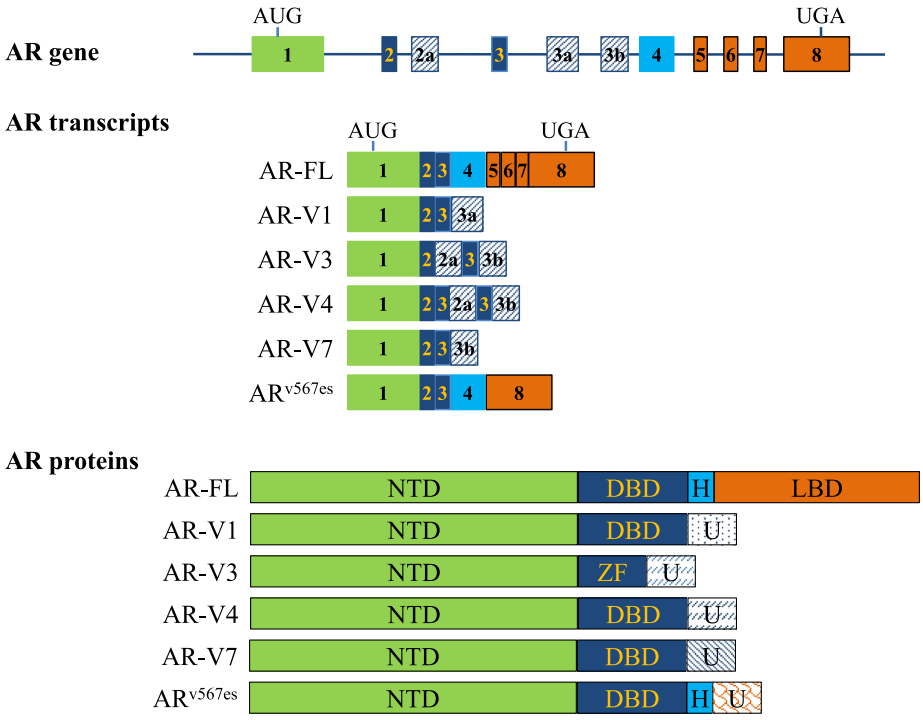


Figure 1.7: Schematic representation of the transcripts and proteins of the full-length AR (AR-FL) and various splice variants (AR-V). NTD: N-terminal domain, DBD: DNA-binding domain, H: hinge region, LBD: ligand-binding domain, ZF: Zn finger, U: untranslated region. From [68].

Several natural or synthetic compounds have been identified to inhibit the function of AR splice variants [134–140]. Further studies are needed to understand their mode of action and to explore the potential of such agents to inhibit PCa progression. Several of these compounds are currently in preclinical or clinical trials.

Bypass mechanisms PCa cells also adopt survival mechanisms that bypass the AR and the AR signaling axis completely, and promote growth and survival of the cancerous cells by increasing cell proliferation and inhibiting apoptosis. Many of the upregulated

signaling pathways in CRPC, including the MAPK and AKT pathway, contribute to cell survival in an AR-independent fashion [16]. The MAPK cascade may influence cell cycle regulation and/or increase cell proliferation via the transcription factors AP-1, c-Myc and NFκB [16, 110, 141, 142]. There are many ways by which the stimulation of the AKT pathway might influence the development of CRPC, including the inactivation of pro-apoptotic proteins, such as BAD and procaspase-9, by phosphorylating them, signaling for G1 cell-cycle progression and inactivation of the Forkhead family of transcription factors to decrease expression of cell-cycle regulator protein $p27^{KIP1}$ [16]. Other transcription factors, including other SRs, have been implicated in the activation of AR target genes, independently of the AR. These include c-Myc, NFκB, AP-1, nuclear factor erythroid 2-related factor 1 and 2 (Nrf1 and Nrf2) and the GR [16, 46, 143]. Hedgehog signaling, which does not involve the AR, has also been implicated in PCa regeneration, invasion and metastasis [12, 144].

Taken together, tumors use a multitude of mechanisms to escape normal growth control under androgen ablation conditions and in the presence of AR antagonists. These mechanisms confer resistance to the current therapies and although substantial advances have been made in recent years it is still not possible to cure the disease once it reaches the CRPC stage. Importantly, however, the new insights gained into the molecular mechanisms by which PCa evolves to CRPC and the factors that lead to therapy resistance, demonstrate that the AR remains the key regulator and driver of tumor growth, spread and survival in castration-resistant tumors, and as such represents the most promising therapeutic target [13, 145].

Furthermore, the AR NTD appears to play a crucial role in the constitutive activity of the AR splice variants in CRPC tumors, which has been proposed to be involved in the development of tumor resistance to some of the recently approved therapeutics, including abiraterone acetate and enzalutamide [47, 53]. The transactivation by the NTD is independent of androgens and therefore circumvents the two current therapeutic strategies that directly target the AR signaling axis, i.e. by inhibition of androgen synthesis or by competitive binding of a potent AR antagonist preventing the binding of the endogenous ligand.

1.1.4 The AR NTD as a drug target for CRPC patients

The transcriptional activity of the AR is mainly mediated via its NTD, both in the presence and in the absence of hormone [146–148]. The AR has two activation functions, namely activation function 1 (AF1) in its N-terminal transactivation domain and AF2 in its LBD. In contrast to most NRs, AF1 is the most potent transactivation function,

and it is key for the recruitment of co-regulators and the further assembly of the pre-initiation complex (PIC) for transcription [149, 150]. Furthermore, the importance of the NTD for transcriptional activity in the absence of ligand is illustrated by the constitutive activity of splice variants lacking the LBD.

Because it was recognized that the NTD plays a crucial role in AR transactivation, both in the presence and in the absence of hormone, the NTD was suggested as a therapeutic target for CRPC [147, 148]. Quayle *et al.* showed that overexpression of an AR NTD protein construct (AR 1–558) inhibited the growth of PCa tumors. The proposed mechanism was that this NTD construct functioned as a decoy molecule that competitively bound the interacting proteins required for activation of endogenous full-length AR. Nevertheless, the specific factors retained by these decoys were not identified [147].

However, the conformational properties of the NTD have hampered the development of treatments targeting this domain. In contrast to the DBD and the LBD, the structures of which are well-characterized, the NTD is predicted to be intrinsically disordered (ID) and experimental data support a high degree of ID in this domain. ID in proteins is characterized by the lack of a well-defined secondary and/or tertiary structure in spite of functional roles of such regions, and has recently been recognized to be crucial in various cellular processes, including transcription [151, 152]. A more detailed description of ID as well as an overview of the studies of the structural properties of the AR NTD is provided under section 1.2.4.

ID is present in essentially all transactivation domains of transcription factors and has been linked to the need of these to establish transient interactions with a wide range of different binding partners, that are likely to change depending on cell type and during the various stages of transcription (initiation, elongation and termination) [153]. A recent review by Hilser and Thompson [154] also highlights the importance of ID in the NTDs of NRs and transcription factors in general. Consequently, the ID nature of the AR NTD is related to its functional role in transcription. However, this property also renders the NTD unamenable to the conventional tools of structure determination such as X-ray crystallography, and has limited its structural characterization thus far [155, 156]. In addition, in spite of the fact that over 150 co-regulators have been reported to interact with the AR AF1 [93, 150, 157, 158], little structural information is available for these complexes. This lack of structural detail hinders rational targeting of the NTD and/or of the protein-protein interactions between the NTD and its binding partners. Nevertheless, several compounds that specifically interact with the AR NTD and inhibit the AR transcriptional activity have been identified recently. This is particularly important because these are the first compounds that do not target

the AR LBD to prevent androgen-dependent AR activation, but instead aim to inhibit transactivation by targeting the NTD.

1.1.4.1 Small molecules that target the NTD

To date, three compounds have been identified to target the AR NTD and to inhibit AR transcriptional activity in PCa cells. These are sintokamide A, a chlorinated peptide extracted from the sponge *Dysidea* sp. [139], EPI-001, a bisphenol A analog isolated from the sponge *Geodia lindgreni* [134], and niphatenone B, a glycol ether from the sponge *Niphates digitalis* [159] (see Fig. 1.8).

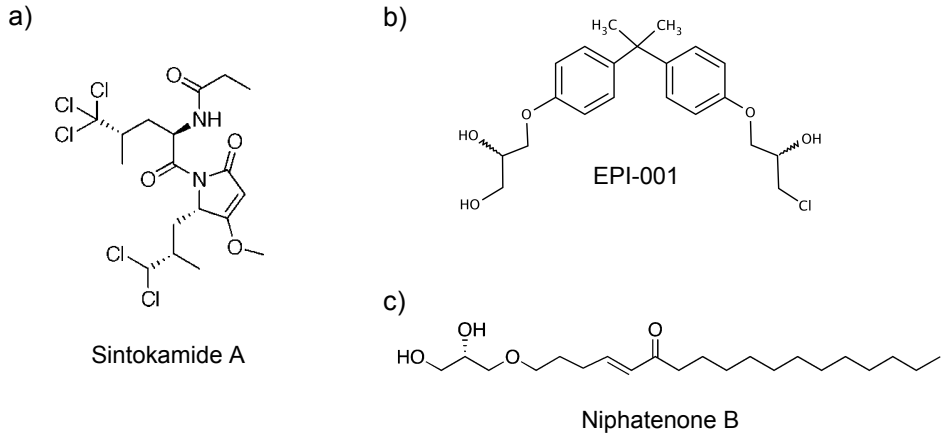


Figure 1.8: Chemical structures of a) sintokamide A [139], b) EPI-001 [134] and c) niphatenone B [159].

Sintokamide A The first small molecule that was identified to selectively block transactivation mediated by the NTD of the AR in PCa cells is sintokamide A. This is a chlorinated peptide that was present in an extract from the marine sponge *Dysidea* sp. Sintokamide A was shown to inhibit both androgen-induced and androgen-independent AR transactivation in LNCaP cells, which are androgen-dependent PCa cells. Since it inhibited transactivation of an AR NTD-Gal4DBD chimera protein, lacking the LBD, it was shown to target the AR NTD [139].

EPI-001 Like sintokamide A, EPI-001 was identified in a screening assay of marine compounds extracted from sponges. Importantly, it was shown to cause regression of CRPC in cell lines and an animal model of this disease, in addition to inhibiting androgen-induced proliferation in androgen-sensitive PCa. It was found to bind covalently with the AF1 region in the NTD and was proposed to inhibit essential

protein-protein interactions between that region of AR and co-regulators or members of the transcription machinery, including AR co-activator CBP and RAP74, a subunit of general transcription factor TFIIF. In addition, EPI-001 reduced AR interaction with AREs on DNA and inhibited transactivation of constitutively active AR splice variants. EPI-001 and its analogs are, at present, the only known inhibitors of constitutively active AR splice variants that have been shown to target the AR NTD [134, 160]. The interaction of EPI-001 with AF1 was proposed to take place in a two-step mechanism, involving a reversible interaction between the two partners that allows for the formation of an intermediate epoxide at the chlorohydrin moiety, followed by a nucleophilic attack of a side chain of AF1 to form a covalent bond [160].

Niphatenone B Niphatenone B, a glycerol ether extracted from the sponge *Niphates digitalis*, was found to inhibit androgen-induced proliferation of LNCaP cells. Like EPI-001, niphatenone B was shown to bind covalently to the AF1 region of the AR NTD. Examination of a series of synthetic analogs revealed that the Michael acceptor enone functionality is not required for activity of the compound and that the glycerol ether oxygen, glycerol alcohols and a saturated alkyl chain substituent on the ketone of reasonable length are important [159].

Importantly, several other compounds have been identified to reduce the transcriptional activity of both full-length AR and AR splice variants, mainly through enhanced AR protein degradation and/or destabilization of the AR mRNA which leads to reduced AR expression. These include nigericin [138], berberine [136], kava components kavalactones and flavokawain B [137], ASC-J9 [140] and 20(S)-protopanaxadiol-aglycone [135]. However, the mode of action of these compounds remains largely elusive and it is not clear whether they exercise their effect by targeting the AR NTD.

Many of these compounds were identified from high-throughput screening assays of existing chemical libraries or libraries of new marine compounds extracted from sponges. Until a better understanding is obtained from the structural properties of the AR NTD and/or of the mode of action of these compounds, the rational design of novel compounds that target the NTD will be challenging and it will be necessary to rely on high-throughput screening efforts to identify new hits.

In the remaining part of the introduction a general overview will be given of the AR and its role in the transcription of its target genes, with particular attention to the structural and functional properties of the AR NTD and the interaction of the AR and the general transcription factor TFIIF.

1.2 The androgen receptor

1.2.1 General biology

The human AR is a 919-residue NR that is expressed in most tissues including the brain, liver, kidneys, muscle, skin, bone and prostate and is important for the development of the male sexual phenotype [161, 162]. Its gene is located on the long arm of the X chromosome, Xq11–12, and consists of 8 exons (see Fig. 1.9) [163, 164]. Exon 1 codes for the NTD, exons 2 and 3 for the DBD and exons 4 to 8 for the hinge region and the LBD. In the late 1980s several groups cloned the human *AR* complementary DNA (cDNA), which paved the way for the structural and functional characterization of the protein [165–168].

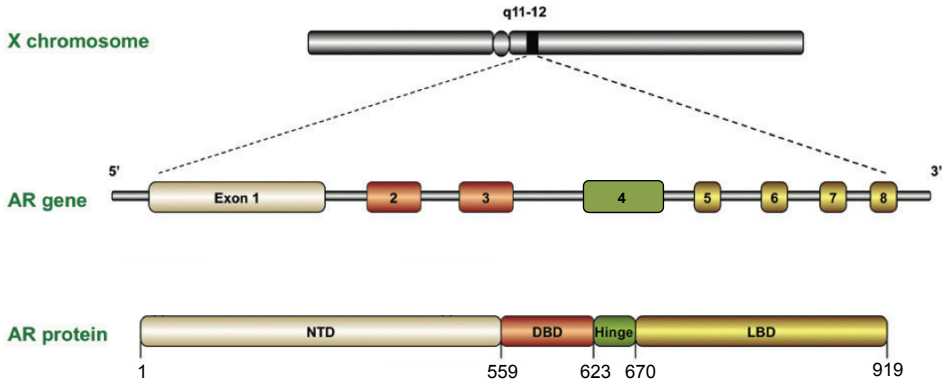


Figure 1.9: Schematic representation of the human *AR* gene and the AR protein. Adapted from [169].

As previously mentioned, various splice variants of the AR were identified recently in addition to full-length AR. Those splice variants all lack the C-terminal LBD and some lack in addition the hinge region and/or parts of the DBD. They are more frequently present in advanced PCa than in healthy tissue and have been linked to the survival of CRPC cells, due to their constitutive activity [68, 170]. Whether they have a functional role in normal cellular processes remains to be determined [170, 171].

In addition to the alternatively spliced AR variants, a shorter isoform of the AR has also been observed in human tissues, and is mainly expressed in heart and sketelal muscle [171, 172]. This isoform, AR-A (also called AR45), is not considered a splice variant since it is translated from the same mRNA as full-length AR, but translation starts at an alternative initiation codon [173]. AR-A is an N-terminally truncated

AR isoform, in which residues 1–531 are missing and the first seven residues differ from the corresponding residues in full-length AR. It represents 10–26% of total AR in some tissues, but its physiological role remains largely elusive [172, 174]. The AR-A isoform was shown to bind to androgens and subsequently translocate into the nucleus where it binds AREs. Furthermore, it was found to inhibit AR transcriptional activity and in mammalian two-hybrid assays it was shown to interact with the AR NTD in an androgen-dependent manner. The absence of stimulatory activity of the AR-A isoform was proposed to be linked to poor co-activator recruitment due to the absence of the NTD and/or the formation of a heterodimer between AR-A and the AR NTD. Interestingly, upon overexpression of co-activators that are known to bind to the LBD, the isoform could stimulate transcription of AR target genes [172].

Androgens mainly exert their physiological effects via binding to the AR, which leads to nuclear translocation, binding to DNA and regulation of the transcription of AR target genes. This mechanism was described in the first part of the introduction. However, androgens can also trigger rapid cellular responses, dependent or independent of the AR. These occur on a timescale of seconds to minutes, and consequently do not involve transcription and subsequent translation of androgen-target genes. Non-genomic actions of androgens can originate at the cell membrane or in the cytoplasm and involve the interaction of androgens with proteins in these locations. This can then lead to, e.g., the activation of kinase signaling pathways, such as MAPK, AKT, protein kinase A, and protein kinase C cell-signaling pathways, or trigger an increase in intracellular calcium concentrations [175–177].

Mutations in the AR or malfunctioning of the protein are associated with several diseases. These include PCa, Kennedy's disease or spinal and bulbar muscular atrophy (SBMA) which is linked to an expansion of the polyQ tract in the AR NTD and will be discussed in more detail in section 1.2.2.4, and androgen insensitivity syndrome (AIS) [175, 178]. Germ line mutations are frequently present in AIS patients, and are usually associated to loss of function, leading to lowered virilization in men (partial AIS = PAIS) or even a fully female external phenotype (complete AIS = CAIS) and infertility [175]. Furthermore, changes in AR activity with ageing have been linked to the development of cognitive impairment, osteoporosis and atherosclerosis [175, 179–182].

1.2.2 Domains of the AR

The AR is a member of the NR superfamily, and belongs to the steroid receptors (SRs), together with the estrogen receptor (ER), the glucocorticoid receptor (GR),

the mineralocorticoid receptor (MR) and the progesterone receptor (PR) [155]. Non-steroid nuclear receptors include the thyroid hormone receptor (TR), the retinoic acid receptor (RAR), the vitamin D receptor (VDR), the peroxisome proliferator-activated receptor (PPAR) and the retinoic X receptor (RXR) [155, 183]. Nuclear receptors are ligand-activated transcription factors and the members of this superfamily share a specific domain organization. As previously mentioned, this is also the case for the AR [3, 183]. Its NTD is located between residues 1 and 559 and contains AF1, important for AR transactivation. The central DBD between residues 560 and 622 contains two Zn fingers and mediates DNA-binding of the AR. A flexible hinge region (residues 623–670) separates the DBD from the LBD. Finally, the LBD is located between resides 671 and 919 and is the domain of the protein to which androgens bind and which contains the less potent AF2 activation function [184].

It should be noted that there is no consensus in the literature as to where precisely the boundaries between these different functional domains are located. This is in part due to the presence of two polymorphic repeats in the AR NTD, a polyQ tract (between residues 58 and 78) and a polyG stretch (between residues 449 and 472), which has led to discrepancies in the AR numbering in early publications [2]. These polymorphic repeat regions will be discussed in more detail under section 1.2.2.4. As previously mentioned, the numbering used by the Uniprot database, which attributes lengths of 21 and 24 to these two polymeric tracts, respectively, and corresponds to a full-length AR of 919 residues, will be used throughout this thesis. This sequence can be found in appendix A.1.

Each of these domains will be discussed in more detail with particular focus on its structural and functional properties. In figure 1.10 a schematic representation is provided indicating the AR domain organization and the various stretches of sequence that have been identified to be important for AR function.

1.2.2.1 Ligand-binding domain

Despite a relative low sequence similarity between the LBDs of various NRs, these domains adopt highly similar three-dimensional structures [185, 186]. Since the LBD of NRs generally contains 12 α-helices, the helices of the AR LBD are numbered 1–12, even though the AR LBD lacks what would be helix 2 in the other LBDs.

In the presence of ligand, the LBD of AR adopts a twelve α-helical sandwich fold with a central ligand-binding cavity, similar to the LBD of other NRs (see Fig. 1.11a) [187, 188]. Ligand binding is stabilized by several contacts between the hormone and the LBD. The position of helix 12 is especially important since it aids to prevent ligand

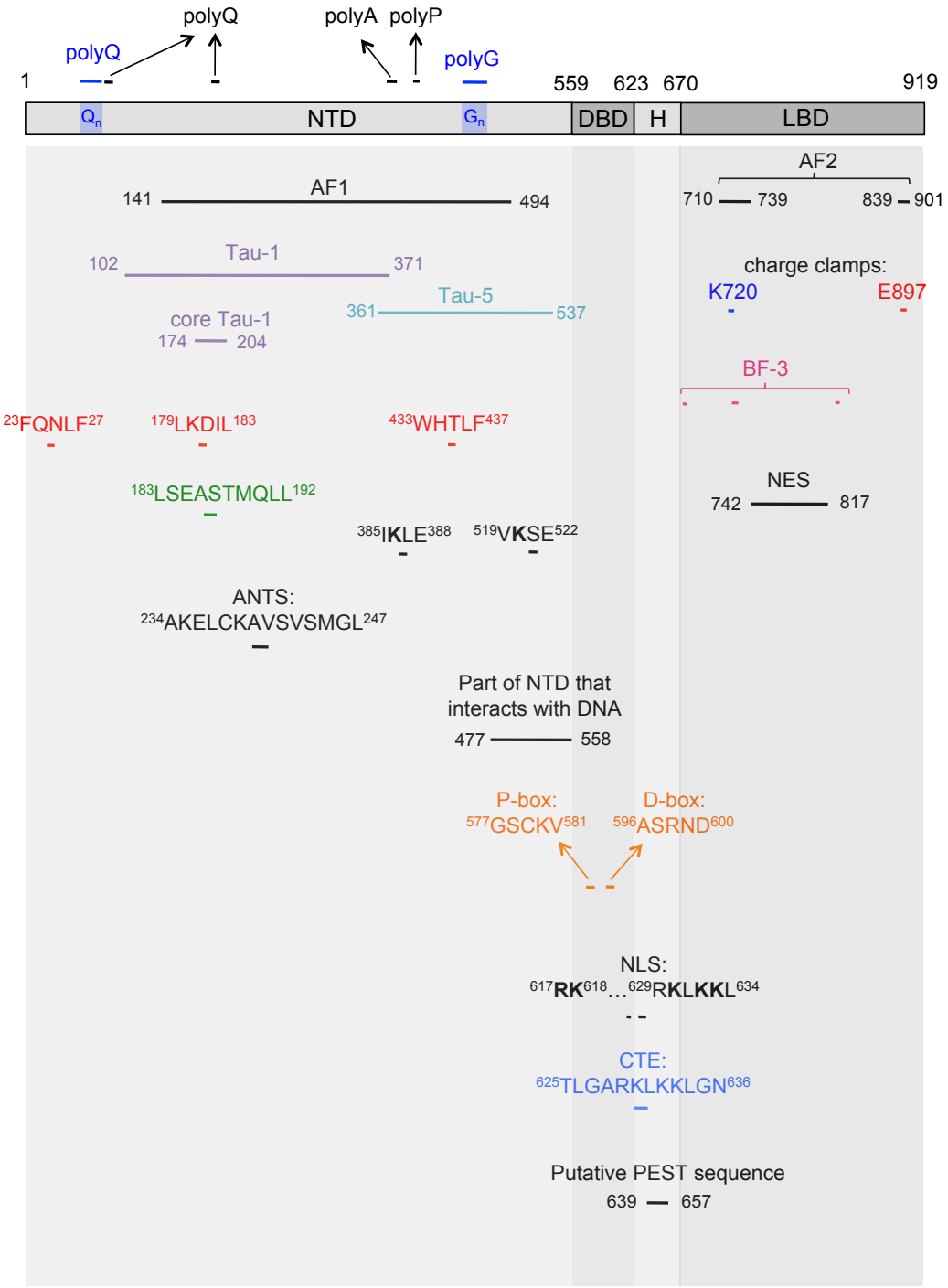


Figure 1.10: Schematic representation of the domain structure of the AR, with indications of the various stretches of sequence that have been identified to be important for AR function. See main text for more details. NTD: N-terminal domain, DBD: DNA-binding domain, H: hinge region, LBD: ligand-binding domain, Q_n: polymorphic polyglutamine stretch, G_n: polymorphic polyglycine stretch, polyQ: polyglutamine, polyA: polyalanine, polyP: polyproline, polyG: polyglycine, AF1: activation function 1, AF2: activation function 2, Tau-1: transcription activation function 1, Tau-5: transcription activation function 5, ANTS: AR NTD signature sequence, BF-3: binding function 3, NLS: nuclear localization signal, CTE: C-terminal extension, NES: nuclear export signal.

escape and is crucial for the interaction of the LBD with both co-regulators and the NTD. Attempts to crystallize the AR LBD in the absence of ligand have not been successful. It is, however, assumed that the structure of the AR apo-LBD (without ligand) is similar to that of apo-LBD domains of other NRs. In the apo-state, helix 12 of these LBDs is oriented away from the rest of the LBD. Upon ligand binding, helix 12 folds back onto the LBD to "trap" the ligand, in addition to other smaller conformational changes (see Fig. 1.11b) [189]. Importantly, this repositioning of helix 12 results in the formation of a hydrophobic cleft on the surface of the LBD, which serves as a binding site for numerous co-regulatory proteins, and is referred to as activation function 2 (AF2) [190]. AF2 is also present in the ligand-bound AR LBD and is formed by helices 3, 4, 5 and 12 (see Fig. 1.12a) [191].

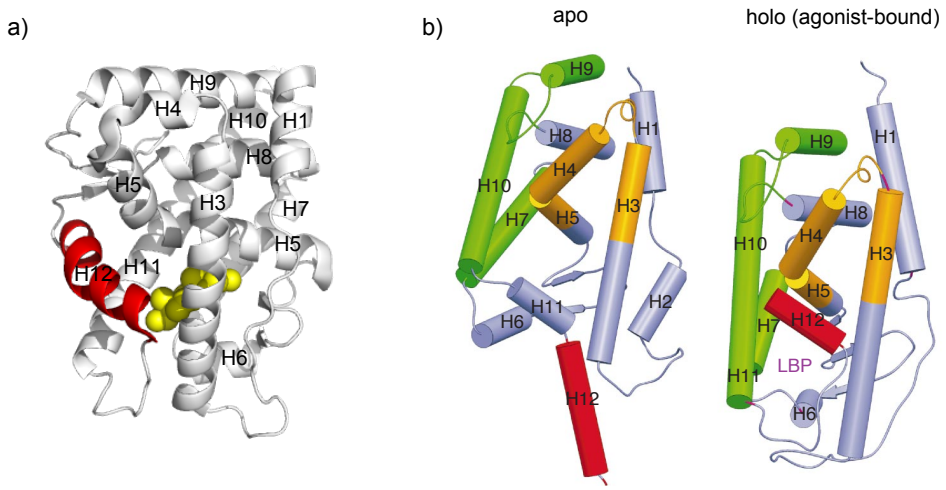


Figure 1.11: a) The structure of the AR LBD bound to its endogenous ligand, DHT (PDB 2AMA, [192]). The helices on the LBD are indicated as H1 to H12 (H2 is missing in the AR LBD). Helix 12 is indicated in red and DHT is represented as yellow spheres. b) Schematic drawing of the LBD of nuclear receptors in the apo and holo state. From [187]. On the left the unliganded RXR LBD is shown and on the right agonist-bound RAR LBD. Also here, the position of the helices is indicated, and helix 12 is represented in red. LBP: ligand binding pocket.

Typically, the AF2 region of NRs interacts strongly with LxxLL signature motifs –in which L is leucine and x is any amino acid– present in several NR co-activators, like the family of p160 co-activators (SRC-1, SRC-2 and SRC-3) [193]. These LxxLL motifs adopt a helical conformation and dock into the hydrophobic cleft formed by AF2 on the surface of the LBD. Two charge clamp residues, K720 (in helix 3) and E897 (in helix 12), form part of the AF2 surface and further stabilize the interaction between AF2 and these co-activator motifs. The AF2 of AR is unique among NRs in the sense that it does not have a high affinity for such LxxLL signature motifs, even though it can accommodate LxxLL containing peptides. Instead, the AR AF2 preferentially binds

FxxLF motifs –in which F is phenylalanine, L is leucine and x is any amino acid–, such as the ^{23}FQNLF^{27} motif located in the N-terminus of the AR (see Fig. 1.12b and c) [194, 195]. This leads to an N/C interaction that is essential for AR transactivation and is regulated in a spatiotemporal fashion [7]. This interaction will be extensively discussed in section 1.2.3. FxxLF motifs are also present in AR co-regulators, such as ARA70, and can bind to the AF2 cleft.

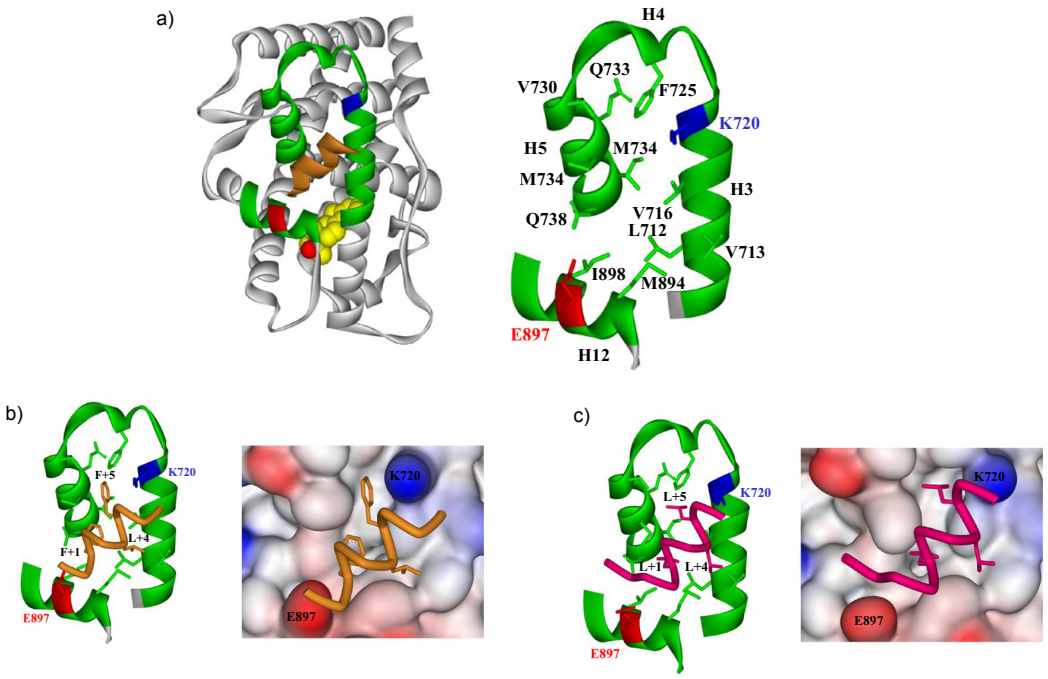


Figure 1.12: a) The AR LBD when bound to the ^{23}FQNLF^{27} motif in the AR NTD. The LBD is represented in white ribbon, whereas the position of the AF2 hydrophobic cleft is indicated in green ribbon and a more detailed view is provided on the right. Charge clamp residues K720 and E897 (see main text) are indicated in blue and red, respectively. The ^{23}FQNLF^{27} peptide is represented in orange and the hormone R1881 is represented as space-filled atoms (yellow: carbon, red: oxygen). b) Different representation of the AR ^{23}FQNLF^{27} bound to the AF2 surface of the AR LBD. c) Representation of the LxxLL motif in co-activator SRC-2 bound to the AF2 cleft on the AR LBD. Peptide backbones in b) and c) are shown as orange and magenta Cα coils, respectively. Side chains of residues at positions +1, +4 and +5 are shown as sticks. The charged residues K720 and E897 are indicated for orientation. Compared to the FxxLF peptides, LxxLL peptides are shifted in the AF2 groove towards K720, resulting in different interactions, and thus stabilizations, between the two types of peptides with then AR LBD surface [158]. Figures a), b) and c) from [158].

The low binding affinity of AF2 for LxxLL motifs commonly present in AR co-activators is directly related to its low intrinsic transactivation potential. Co-activators that typically bind to AF2 in other NRs are preferentially recruited to AF1 in the AR NTD (via a distinct binding surface). Consequently, the main transactivation of the AR is mediated by AF1 instead of AF2.

In addition to the ^{23}FQNLF^{27} motif in the NTD and the LxxLL or FxxLF motifs present in the AR co-activators, the AF2 hydrophobic cleft can also interact with LxxxIxxx(I/L) motifs of co-repressors –in which L is leucine, I is isoleucine and x is any amino acid– such as nuclear receptor co-repressor (NCoR) and silencing mediator for retinoid and thyroid hormone receptors (SMRT) [188, 196].

Binding of androgens to the AR LBD also exposes the binding function 3 (BF-3) surface that was proposed to allosterically regulate co-activator binding. BF-3 forms another hydrophobic cleft at the junction of helix 1, the helix 3–helix 5 loop and helix 9 (see Fig. 1.13). Several compounds were identified to bind to the BF-3 surface and in doing so remodeled the adjacent AF2 interaction site to weaken co-activator binding [197].

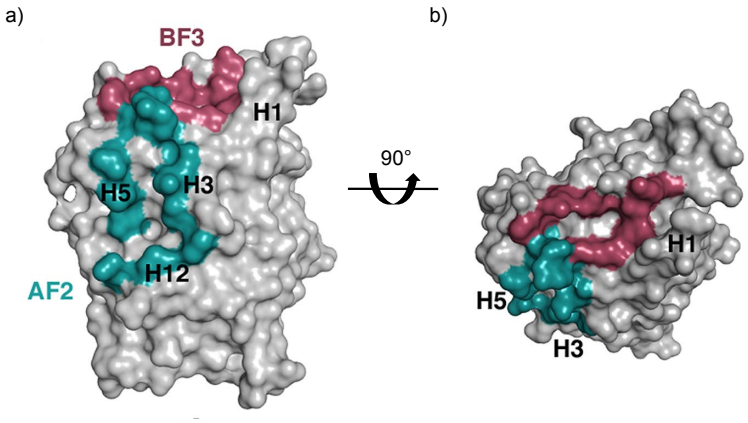


Figure 1.13: a) Space-filling model of the AR LBD showing residues in AF2 (cyan) and BF-3 (red). b) As in a, but rotated 90 ° about the x-axis to reveal BF-3. From [197].

As explained in the first part of the introduction, the AR LBD cannot only bind to its endogenous ligands, testosterone and DHT, but also to several other ligands that either have an agonistic effect (activating) or an antagonistic effect (repressing) on AR transcriptional activity. Both agonists and antagonists bind to the same ligand binding pocket as DHT, but binding of antagonists results in a different reorientation of helix 12, which impedes the correct formation of the AF2 binding pocket and hence prevents the binding of co-activators. Instead, favorable surfaces for the interaction with co-repressors can be formed [187].

The LBD also contains a nuclear export signal (NES) between residues 742 and 817 that is inactive in the presence of ligand, allowing androgen-induced nuclear translocation of the AR mediated by the nuclear localization signal (NLS) in the hinge region [175, 198].

Interestingly, the LBD was already shown in the early 90s to have an inhibitory

effect on AR transactivation in the absence of hormone. Deletion of the LBD resulted in constitutive activity to, depending on the promoter context, approx. 10% of or approx. the same level as the activity from full-length AR in the presence of hormone [146, 199]. This observation indicated that the LBD acts as a regulatory domain that represses transcriptional activation in the absence of androgens [146, 199]. This was further confirmed by the discovery of constitutively active splice variants lacking the NTD more than a decade later.

1.2.2.2 DNA-binding domain

The DBD is the domain of the AR through which this SR recognizes and binds specific DNA elements in regulatory promoter or enhancer regions of its target genes, called AREs. The AR DBD, similar to the DBD of all SRs, binds as a homodimer to its response elements on DNA that are typically present in pairs [155, 200]. DNA binding by NRs is highly evolutionary conserved, as the DBD is the most conserved domain among NRs [161, 184]. DBDs are organized in two Zn fingers, which are modules in which a Zn atom coordinates four cysteine residues (see Fig. 1.14). The first Zn finger contains the P-box and forms an α-helix that enters the major groove of DNA, conferring responsibility for sequence recognition and DNA binding [201]. The P-box residues that make these base-specific contacts are identical for the AR, GR, MR and PR, and correspond to residues ^{577}GSCKV^{581} in the AR [184]. The second Zn finger is involved in the DNA-dependent dimerization of the AR DBD via the so-called D-box residues ^{596}ASRND^{600}, which are also conserved between the AR, GR, MR and PR [202, 203]. These D-box interactions are sustained by a network of hydrogen bonds between individual amino acid residues in the D-box and by an extensive complementary surface [7].

Various types of AREs have been identified in regulatory regions on AR target genes. Classical AREs (clAREs) consist of an inverted repeat of two hexameric half-sites that are separated by three nucleotides (IR3). These half-sites have consensus sequence 5'–AGAACA–3' and IR3 AREs are recognized by all members of the SR family, except for ER that preferentially binds a slightly different nucleotide sequence [155]. AREs that are selective for the AR and are not bound by other SRs (selAREs) have also been identified and are organized as direct repeats of the same hexamer sequence spaced by three nucleotides (DR3). In addition, non-canonical AREs have been described, including single ARE half-sites or half-sites separated by more than three nucleotides [204–207].

The AR DBD has not been crystallized on an IR3 clARE, but the DBD dimer of

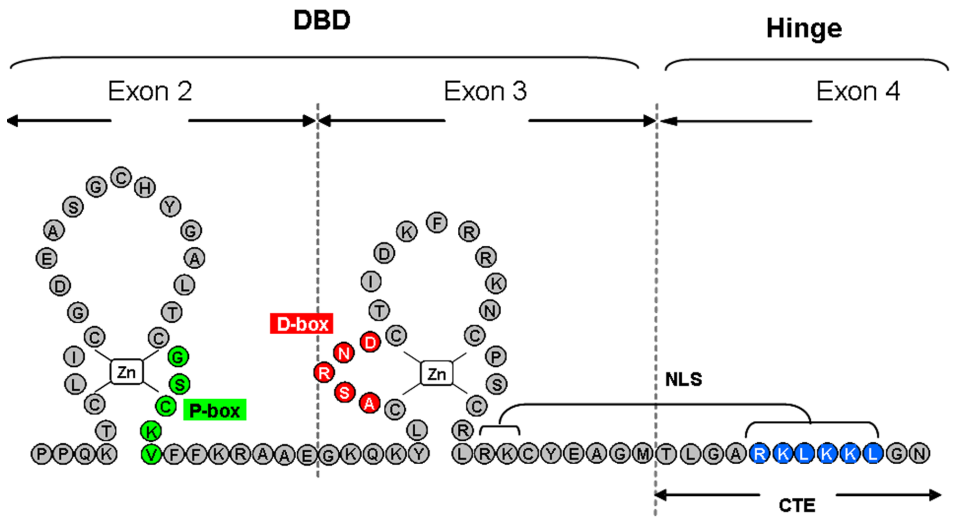


Figure 1.14: Schematic representation of the two Zn fingers of the AR DBD and the C-terminal nuclear localization signal (NLS) and C-terminal extension (CTE) that are (in part) located in the hinge region. The P-box and D-box residues are highlighted, as well as the position of the NLS and CTE. The single letter code for amino acids is used. From [184].

other SRs like ER and GR adopts a head-to-head orientation when bound to such a DNA sequence [208, 209]. The P-box residues interacting with the DNA fix the position of the DBD dimer and this is optimally combined with an inverted repeat sequence separated by three nucleotides. This explains why SRs bind with high affinity to such IR3 sequences [184].

An X-ray crystal structure is, however, available for the AR DBD dimer on DR3 selAREs (PDB 1R4I, [210], see Fig. 1.15). Contrary to what was expected, the AR DBD dimer adopts a head-to-head orientation on these direct repeats. Other NRs that have been crystallized on direct repeats bind in a head-to-tail orientation [210]. Furthermore, this head-to-head dimer of the AR DBD on DR3 is nearly identical to the head-to-head dimers observed for ER and GR DBD dimers bound to IR3 repeats [208–210]. The head-to-head orientation of the AR DBD on DR3 results in one AR DBD binding to its cognate half-site with high affinity and the partner DBD with low affinity, thereby reducing specific interactions with the target DNA [207, 210]. This is compensated by a relatively stronger AR D-box dimerization interface compared to those of other SRs bound to IR3s. Hydrogen bonds and van der Waals contacts are present between S597 in both DBD monomers, as well as an additional pair of hydrogen bonds between A596 and T602 in the opposing DBD, and vice versa [7, 210]. In addition, residues in the carboxyl-terminal extension (CTE) (amino acids 625–636, already located in the hinge region) were proposed to provide an additional dimer interface for the AR DBD that acts in concert with amino acids in the second Zn finger to mediate specific and

high-affinity DNA binding [207, 211–214].

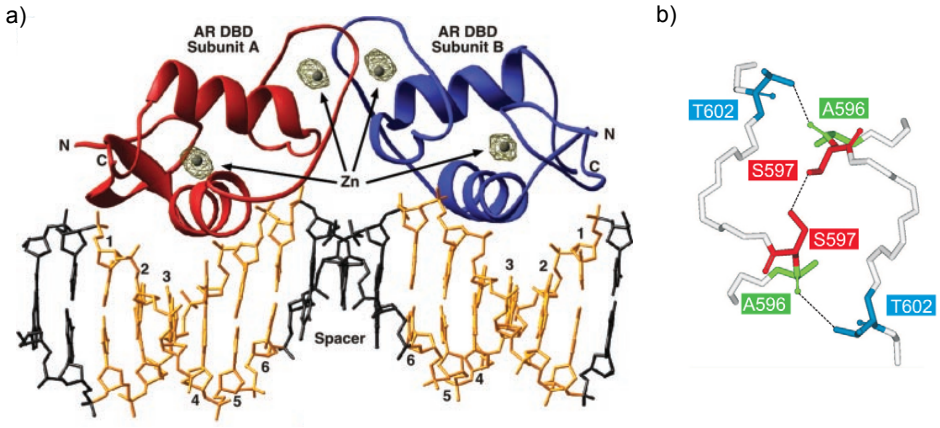


Figure 1.15: a) X-ray crystal structure of the AR DBD dimer on DR3 selAREs. PDB 1R4I, [210]. From [210]. b) zoom of the D-box interactions, showing the specific contacts between S597 in both DBD monomers as well as an additional pair of hydrogen bonds between A596 and T602 in the opposing DBD, and vice versa. From [7].

This illustrates that the role of DNA is not merely to provide a sequence that can be recognized by the AR, but the sequence and orientation of the hexameric repeats as well as the spacer length contribute to the nature of the DBD dimer formed, which further determines the conformation of the full-length AR when bound to DNA. It is believed that the second Zn finger and CTE have important roles in regulation [155, 175, 207].

Furthermore, the DBD contains part of the bipartite ligand-dependent NLS for AR nuclear import. The NLS consists of two basic amino acid clusters (underlined) found in the DBD and hinge, corresponding to residues 617–618 (in the DBD) and residues 629–633 (in the hinge) (^{617}RKCYEAGMTLGARKLKK^{633}) (see Fig. 1.14) [215]. The conformational changes in the AR upon binding of DHT expose the NLS and allow nuclear translocation of the AR.

1.2.2.3 Hinge region

The hinge region connects the DBD and LBD of the AR and has for a long time been considered as merely a flexible polypeptide linker. Even though its sequence is poorly conserved among NRs, the hinge region of all SRs contains part of a bipartite NLS, as the one present in AR [215, 216]. Mutations in the NLS region were shown to result in the loss of nuclear import, demonstrating the importance of this region for nuclear translocation [2].

In addition to its role in nuclear import, several other functions for the hinge re-

gion in the regulation of AR activity are emerging. It is involved in DNA selectivity and affinity, transactivation potential of the AR by, for instance, modulating the N/C interaction between the NTD and LBD and co-activator or co-repressor recruitment [84, 184, 217].

As discussed in the previous section, the hinge region contains a CTE sequence, which plays an important role in differentiating between clAREs and selAREs. Binding of the AR DBD to selAREs involves the CTE, while interaction with clAREs only involves the two Zn fingers. Furthermore, the hinge region of NRs has been associated with the relative orientation of the LBDs and DBDs of recently solved structures of (nearly) full-length NRs as dimers on DNA. The flexibility and the length of the hinge region in these NRs were found to contribute to the correct positioning of the receptors on a specific DNA element [155, 218–221].

The AR hinge is also an important site for PTMs, like acetylation, methylation, ubiquitination and phosphorylation, all influencing receptor activity [83, 84]. In addition, various mutations in the hinge region that affect AR function have been found in AIS and PCa patients, further illustrating its role in AR function [84].

The ^{629}RKLKKLGN^{636} motif, which is part of the CTE and contains part of the NLS, was identified to have an inhibitory role as deletion of this sequence resulted in an increased AR transactivation potential. The N/C interaction between the NTD and LBD was shown to be strongly enhanced by deletion of this motif. Paradoxically, deletion of this region or mutations in it also resulted in an impairment of nuclear translocation as well as an apparent loss of *in vitro* DNA binding [217]. Somatic mutations in R629 and K630 have been reported in PCa biopsies [222, 223]. Furthermore, acetylation of lysine residues K630, K632 and K633 has been reported in the AR [87, 224–226] and is important for nuclear translocation as well as for the regulation of AR activity [84, 184, 217]. Methylation of K630 and K632 has also been reported and was found to positively modulate AR transactivation by enhancing the N/C interaction as well as facilitating the recruitment of AR to its target genes [227, 228].

Moreover, the hinge region also contains a putative PEST sequence between residues 639 and 657. Such sequences are rich in proline, glutamate, serine and threonine residues and can function as signals for enhanced degradation via regulation of ubiquitination. However, deletion of this sequence did not affect AR activity or its expression level [84, 175, 184, 217].

1.2.2.4 N-terminal transactivation domain

The NTD constitutes about 60% of the AR and predominantly mediates the AR transcriptional activity. It is the least conserved domain among NRs, both in size and sequence, with only about 15% homology between different SRs [229].

Whereas the relationship between the structure and functional roles of the DBD and LBD is well-established, no structure of the NTD, or of fragments of it, is available. This is largely due to its substantial flexibility that has prevented its characterization by X-ray crystallography and its size (559 amino acids), which has delayed its study using solution nuclear magnetic resonance (NMR) spectroscopy [156]. In spite of this, several regions have been identified within the NTD that are important for its function. Its transcription activity has been mapped to specific regions containing linear motifs. In addition, many co-regulators of the AR that interact with the NTD have been identified [93, 175] and although there is little structural information available for these complexes, most of these co-regulators were found to bind within the functional regions of the NTD.

Functional regions within the NTD The NTD harbors the most potent activation function of the AR, AF1, and as such is indispensable for AR transactivation, both in the presence and in the absence of androgens [146–148]. In contrast to AF2, which is a well-defined patch on the globular LBD, AF1 is an ill-defined region in the ID NTD and has been loosely defined as residues 141–494 based on early chloramphenicol acetyltransferase (CAT) reporter assays with several deletion mutants of the NTD [149, 199].

Further transcription activation assays evaluating the effect of deletions in the NTD identified two transcription activation units (Tau-1 and Tau-5) that are indispensable for AR transactivation [149]. Importantly, depending on the context, distinct regions of the AR NTD were found to be essential for AR transcriptional activation. Both the size and the location of the transcription activation unit was shown to differ depending on the promoter context, the presence or the absence of the LBD, the nature of the DBD (AR DBD or Gal4DBD), the hormone level, the sequence and organization of the AREs and the cell type. This illustrates that the AR is capable of using different regions in the NTD as transcription activation units depending on the cellular context [146, 149, 230, 231]. Tau-1, located between residues 102 and 371, is essential for the transactivation of full-length AR in the presence of hormone, whereas Tau-5, located between residues 361 and 537, appears to be particularly important for transactivation of full-length AR at low concentrations of androgens (comparable to castrate levels) as

well as for the constitutive activity (i.e. in the absence of androgens) of splice variants of the AR lacking the LBD that are frequently present in CRPC patients [149, 232].

Callewaert *el al.* further defined the core region of Tau-1 as residues 174–204 (SCSAD**LKDIL**SEASTMQLLQQQQQQEAVSEGS), based on a comparison with the core Tau-1 of the GR and the presence of predicted helices in the AR between residues 178 and 186, and residues 188 and 200. They showed that simultaneous mutation of residues I182, L183, L192 and L193 (underlined residues) in this core Tau-1 resulted in a significant loss of AR transcriptional activity, and almost completely inactivated full-length AR when combined with deletion of the Tau-5 region [157]. In addition, the core Tau-1 region contains an ^{179}LKDIL^{183} linear motif (shown in bold), which is important for its role in AR transactivation and will be discussed in more detail in section 1.2.2.4.

Similarly, the linear motif ^{433}WHTLF^{437} was identified as the core sequence that mediates Tau-5 activity [232], despite the fact that the integrity of the complete Tau-5 appears to be required for its optimal transcriptional activity as illustrated by the loss of its transactivation properties by deletion of any part of Tau-5 [149, 157]. The ^{433}WHTLF^{437} motif was found to be responsible for approximately 50% of androgen-independent AR transcriptional activity, while its function in androgen-dependent AR transactivation appears to be limited, in agreement with the predominant role of Tau-5 in the absence of or at low levels of androgens [232]. This linear motif will also be further discussed in section 1.2.2.4.

AF1 is the main transactivation function of the AR and mediates protein-protein interactions with numerous co-regulatory proteins and members of the transcription machinery [150]. Members of the p160 co-activator family (SRC-1, SRC-2 and SRC-3) that bind the AF2 cleft in the LBD of other NRs via LxxLL motifs, were found to interact weakly with the AR AF2 and instead preferentially bind to the Tau-5 region of AR AF1 via glutamine rich regions, independent of their LxxLL motifs [157, 231, 233, 234]. Although p160 co-activators do not bind to the Tau-1 region directly, this region attenuates the interaction of p160 co-activators with the Tau-5 region of AR [157]. Since there is no evidence for direct intradomain communication between Tau-1 and Tau-5, this is suggested to be an indirect effect, e.g. via induction of a conformational change or the recruitment of a secondary binding partner [157, 184].

Despite the fact that AF1 is the main site for recruitment of co-regulatory proteins, various co-regulators have been identified that bind to the DBD or the hinge region or that can bind weakly to the AF2 in the LBD via LxxLL or FxxLF motifs [175, 184].

Linear motifs There are three linear motifs present in the AR NTD that resemble LxxLL motifs: ^{23}FQNLF^{27}, ^{179}LKDIL^{183} and ^{433}WHTLF^{437}. Each of those has been found to play a key role in AR transactivation and/or interaction with co-regulators.

The ^{23}**FQNLF**^{27} motif is located close to the N-terminus of the AR, and even though it resides outside Tau-1 and Tau-5 it plays an important role in AR transactivation. This motif is highly conserved among the AR of different species and is the main motif through which the N/C interaction is mediated [235, 236]. Binding of the ^{23}FQNLF^{27} motif to the AF2 pocket of the LBD is androgen-dependent and involves the folding-upon-binding of this region of sequence into a helix (PDB 1XOW, [191]). It was shown that amino acids 3–13 further modulated the N/C interaction without directly interacting with the LBD [237]. The binding affinity between the ^{23}FQNLF^{27} motif and the AF2 pocket was determined to be $1.2 \pm 0.2\,\mu$M by isothermal titration calorimetry (ITC) [238] and $9.2 \pm 0.4\,\mu$M by fluorescence polarization [191]. In addition to its role in the formation of the N/C interaction, ^{23}FQNLF^{27} also binds cyclin D1 [239] and co-activator melanoma antigen gene protein 11 (MAGE-11) [240].

The ^{179}**LKDIL**^{183} motif is located in the core Tau-1 region and its importance in mediating Tau-1 activity has been confirmed in various cell lines [149, 157, 230]. Despite its prominent role in transactivation the co-regulatory proteins that are thought to interact with this region remain largely unidentified [170, 241]. Recently, Jin *et al.* reported that MED1, a subunit of the Mediator complex, interacts with the Tau-1 region of the AR in the presence of hormone. Deletion of the core Tau-1 region resulted in a loss of binding, which suggested that the core Tau-1 region is crucial for AR-MED1 interaction [241]. The region in which ^{179}LKDIL^{183} is located is predicted to form an amphipathic helix. Mutational analysis showed that both the hydrophobic side chains and the negative charges of this putative amphipathic helix are important for the transactivating capacity of the AR [157]. In addition, the ^{179}LKDIL^{183} motif has been described to contribute to the N/C interaction by association with AF2 [157, 233]. Mutations enhancing the hydrophobic nature of the flanking regions of the ^{179}LKDIL^{183} motif were found to enhance the N/C interaction [157]. However, the affinity of the ^{179}LKDIL^{183} motif for the AR LBD is very weak and most likely not significant [184, 233]. Overlapping with the ^{179}LKDIL^{183} motif, there is a ^{183}LSEASTMQLL^{192} (Lx$_7$LL) motif, which is evolutionary conserved among SRs and also forms part of core Tau-1. This Lx$_7$LL motif serves as a binding site for TAB2, a component of the NCoR complex [242].

The ^{433}**WHTLF**^{437} motif was identified as the key motif within Tau-5 for AR transactivation in the absence of hormone. [232] Androgen-dependent AR transactivation is mediated via Tau-1 and the Tau-5 region plays an inactive or inhibitory role under

these conditions. However, for full-length protein in the absence of hormone, or for splice variants lacking the LBD, Tau-5 becomes the predominant transactivation unit [146, 149, 199, 232]. The ^{433}WHTLF^{437} motif in the Tau-5 region selectively enhances transcriptional activity in the absence of or at low levels of hormone in CRPC cells [232]. It was therefore proposed that targeting the ^{433}WHTLF^{437} motif may represent a novel strategy to selectively inhibit aberrant AR activity in CPRC cells [232]. Nevertheless, the molecular mechanisms underlying this functional role of the ^{433}WHTLF^{437} motif as well as the nature of a potential binding partner through which it could exert this function remain unclear.

Furthermore, the ^{433}WHTLF^{437} motif was reported to bind to the LBD, in an androgen-independent manner, to form a secondary binding site that stabilizes the N/C interaction [243]. However, it is not clear whether the ^{433}WHTLF^{437} motif interacts with the LBD via the AF2 groove or whether instead it interacts with a part of the LBD located outside of AF2 [243–245]. Although there is no structure available for the wild-type ^{433}WHTLF^{437} sequence in complex with the LBD, a WxxLF containing peptide, obtained via phage display, has been crystallized in complex with the LBD and is located in the AF2 cleft in a similar way as the ^{23}FQNLF^{27} motif (PDB 1T74, [195]). The WxxLF peptide adopted a more helical conformation upon binding to the AF2 cleft, but remained relatively disordered and was suggested to bind with lower affinity than the ^{23}FQNLF^{27} peptide [195]. In general, the contribution of the ^{433}WHTLF^{437} motif to the N/C interaction remains unclear and it is even uncertain it is involved in this interaction at all [184, 207].

The ^{433}WHTLF^{437} motif has also been reported to interact with histone acetyltransferase p300 as part of an interaction network that leads to upregulated AR transcription [246]. This interaction is described to only take place when the ^{23}FQNLF^{27} motif interacts with MAGE-11 (see above) which exposes the ^{433}WHTLF^{437} motif so it is available for interaction with p300 [246].

AR NTD signature sequence The AR NTD signature sequence (ANTS) is located between residues 234 and 247 (AKELCKAVSVSMGL) and is fully conserved across all species. This is remarkable given the low degree of sequence conservation of the AR NTD across species and, even more so, because this sequence is not present in any of the other SRs [2]. This stretch of fourteen amino acids was found to interact with the COOH terminus of the Hsp70-interacting protein E3 ligase (CHIP), which promotes AR degradation [247].

Amino acid repeats in the NTD Several amino acid repeats are present in the NTD, including two large polymorphic glutamine and glycine stretches (residues 58–78 and residues 449–472, respectively), two shorter glutamine repeats of six and five residues (residues 84–89 and residues 193–197) and short repeats of five alanine (residues 398–402) and eight proline residues (residues 372–379). The polymorphic glutamine repeat (polyQ) is encoded by trinucleotide CAG repeats in exon 1 and expansion of this stretch to more than 38 glutamine residues is associated with Kennedy's disease or SBMA, a neurodegenerative disease affecting the motor neurons in the brain stem and spinal cord [248]. The polyglycine stretch (polyG) is encoded by GGN repeats –in which N is any nucleotide– in exon 1 and is also variable in length [2].

The roles of the polymorphic polyQ and polyG repeats have not been fully elucidated, but they have been proposed to provide inhibitory control over the NTD [2]. Shorter polyQ and polyG repeat lengths are associated with increased AR activity [249–251], and deletion of the polyQ stretch results in a much more active AR [252]. The N/C interaction was also found to be stronger for shorter repeat lengths [175, 252, 253]. Furthermore, the length of both the polyQ and polyG stretch is inversely correlated with the risk of developing PCa [254–256], and somatic mutations that shorten the CAG repeat are commonly observed in PCa patients [257]. In contrast, longer polyQ stretches are correlated with an increased risk of SBMA and an earlier age of onset of this disease [248].

1.2.3 AR interdomain communication and allosteric regulation of the AR activity

Even though each of the AR domains is responsible for a prime function (transactivation, DNA or ligand binding), there is ample evidence for functional communication between the different domains. Also the structural and dynamical properties of the AR domains are coupled and this coupling plays and important role in the activation of the AR by androgens. Furthermore, allosteric communication can be induced by ligand or DNA binding [154, 155]. In this section, the interdomain communication and the allosteric mechanisms reported for the AR will be described.

Androgen binding to the AR LBD causes intradomain structural changes that expose the AF2 and BF-3 binding surfaces [190, 197]. However, in addition to these local effects, ligand binding also causes conformational changes in other domains of the AR. It exposes for instance the NLS, located in part in the DBD and in part in the hinge region, which allows interaction with proteins that translocate the AR to the nucleus [258]. Ligand binding further induces the N/C interaction between the NTD

and the LBD and the concomitant release of molecular chaperones, which changes the conformational properties of the NTD and possibly other parts of the AR [7].

Binding of the AR DBD to AREs does not only cause conformational changes in the DBD, but was also shown to induce a certain degree of helicity in the NTD [259, 260]. This was proposed to act as a mechanism to regulate gene expression by enhancing the interaction of the NTD with certain co-activators [259]. Similarly, DNA binding also caused structural changes within the NTD of other SRs, including GR [261], PR [262] and ERα [263–266]. Furthermore, the AR NTD also modulates the DNA binding properties of the DBD by reducing its binding affinity for AREs [259]. Liu *et al.* reported that a region of the AR NTD immediately adjacent to the DBD and spanning 81 residues (residues 477–558) binds directly to the DBD and thereby inhibits DBD-ARE interaction *in vitro* and AR transactivation *in vivo* [267].

The ligand-induced N/C interaction between the NTD and LBD of the AR is one of the best characterized interdomain interactions of the protein. Similar ligand-dependent N/C interactions have also been reported for other SRs, including MR [268], PR [269, 270] and ERα [271, 272], but none of them was found to be as strong as the one in AR. The N/C interaction of the AR has been described to be required for its full transcriptional activity depending on the context, including the promoter sequence [244, 273]. It can be formed by direct interaction between the NTD and the LBD or instead N/C interaction can be indirect via the binding of SRC co-activators to both the NTD and the LBD of the AR [2, 274, 275]. Direct N/C interaction primarily involves ligand-induced binding of the ^{23}FQNLF^{27} motif to the AF2 hydrophobic cleft in the LBD. As explained before, the contribution of both the ^{179}LKDIL^{183} and ^{433}WHTLF^{437} linear motifs in the NTD to this interaction remains largely unclear.

It has been proposed that the efficient recruitment of co-activators requires a functional association of the NTD with the LBD via the N/C interaction [233]. This interaction is thought to facilitate co-regulator binding by generating alternative binding sites at the AR NTD and hinge region [184, 276]. It was, for instance, shown to be required for AR binding with the SWI/SNF chromatin remodeling complex that primarily interacts with the AR hinge region [277] and was suggested to induce a conformational change in the NTD that renders the Tau-1 domain accessible for MED1 binding [241]. Additionally, the N/C interaction may regulate the interaction of co-regulators with the AF2 hydrophobic groove by occupying it to prevent untimely and/or unfavorable protein-protein interactions. Furthermore, the N/C interaction has been suggested to play a role in the stability of the ligand-bound state [194, 245, 278].

The N/C interaction can occur both intramolecularly and intermolecularly and is

regulated in a spatiotemporal fashion (see Fig. 1.16) [7]. Upon ligand binding of the AR in the cytoplasm, intramolecular N/C interaction takes place [278, 279], which is followed by nuclear translocation. In the nucleus, the majority of AR molecules forms head-to-tail dimers stabilized by intermolecular N/C interactions [7]. These can bind to AREs on DNA. It is not clear, however, whether the N/C interaction is lost in DNA-bound AR or not. Data from van Royen *et al.* strongly suggest that the N/C interaction is abolished upon DNA binding [278]. This was proposed to facilitate interaction of AR with co-activators to initiate transcription [7, 278]. When the N/C interaction is lost, alternative surfaces on the LBD as well as the NTD are expected to become available for the interaction with co-regulatory proteins [184]. In contrast, Klokk *et al.* found that there are significant intramolecular N/C interactions in AR in its DNA-bound and transcriptionally active state. Intermolecular N/C interaction was also observed in DNA-bound AR, but to a lesser extent [280]. Interestingly, chromatin immunoprecipitation (ChIP) data from Li *et al.* indicated that, for the promoters tested, N/C interaction was required for binding to chromatin but not necessarily for binding to plasmid-based AREs [277]. This further supports the need for N/C interaction to occur in the AR when it is bound to DNA in cells. However, it remains unclear whether the N/C interaction is generally lost upon DNA binding or not. It is likely that the cellular context determines the nature of the N/C interaction formed and the role it plays in AR signaling [207].

In spite of what it already known about the communication between the AR domains, many aspects of its allosteric regulation remain largely undiscovered. It is not clear whether some of these interactions occur intra- or intermolecular (or both), how they are regulated in time and how other proteins or PTMs affect them. It appears that the N/C interaction of the AR, as well as its interdomain communication and its interaction with binding partners are dynamic processes that depend on the context and can change during one or more of the steps in transcription activation [155, 184].

1.2.4 Structural properties of the NTD

Unlike the DBD and LBD, the NTD has no stable secondary or tertiary structure and according to biophysical experiments it can be considered as ID [156]. The NTD of other NRs is similarly disordered and this flexibility is related to the function of this domain [150].

The AR AF1, in isolation, was shown to contain 13–16% helical secondary structure by circular dichroism (CD) and Fourier transform infrared spectroscopy (FTIR) [281, 282]. However, as described in the previous section, interdomain communication and

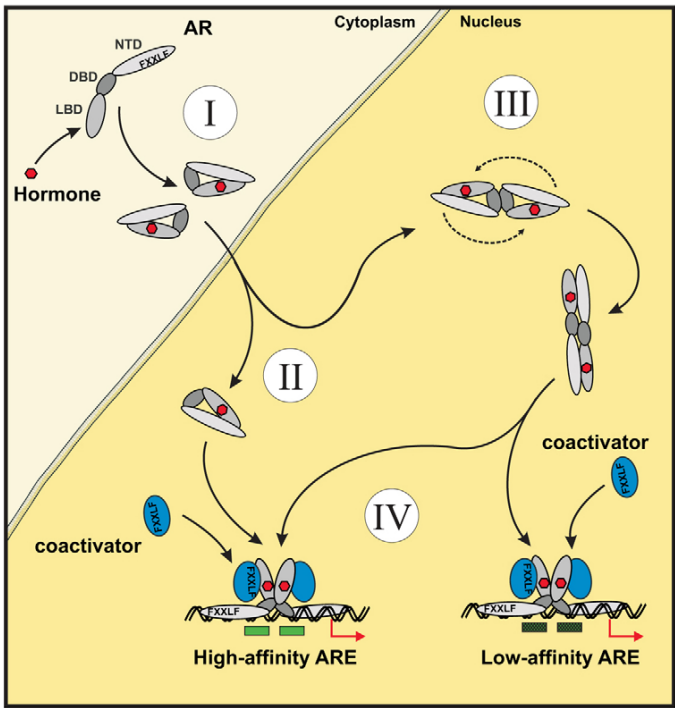


Figure 1.16: Schematic representation of the spatiotemporal organization of N/C interactions in AR function (see main text). From [7]. Note that in this schematic representation, the N/C interaction of AR is lost upon DNA binding. It is, however, not clear at present whether N/C interaction occurs when AR is bound to DNA or not (see main text). Therefore, the mechanism described here represents one possible scenario for DNA-bound AR.

allosteric mechanisms can increase the helical content of AF1. The N/C interaction induces an α-helical conformation in the ^{23}FQNLF^{27} binding motif and is has been proposed to affect the conformational properties of other regions in the NTD and possibly other domains of the AR [191, 241]. When the NTD is linked to the DBD, binding to DNA will also induce helicity in the NTD [259]. This illustrates that in the context of full-length AR and/or when the AR is bound to DNA, the NTD adopts a more folded conformation than when it is studied in isolation.

Interestingly, other factors have also been found to increase the helical content of the AR NTD. The domain becomes more helical in the presence of a natural osmolyte trimethylamine N-oxide (TMAO), or a hydrophobic organic solvent, trifluoroethanol (TFE). [281] TMAO is thought to stabilize proteins in a native folded conformation [183], and TFE is a helix-stabilizing solvent [281]. Binding of the AR AF1 with the large subunit of TFIIF, RAP74, has also been shown to increase the helical content of AF1, similar to the induced helicity by addition of TMAO [281, 282]. In the presence of TMAO or RAP74 the α-helical content was increased three-fold [282]. Furthermore, addition of TMAO reduces limited proteolysis of AF1 and steady state fluorescence emission spectra indicated that the two tryptophan residues within AF1 (W397 and W433) become less solvent exposed when TMAO is added, which is consistent with an increased conformational stability of the AF1 region [281]. Moreover, the increased α-helical content in AF1 in the presence of TMAO or upon RAP74 binding resulted in enhanced binding of co-activator SRC-1. [282] This suggests that induced folding of AF1 creates surfaces for further protein-protein interactions and may illustrate a mechanism for regulating assembly of distinct transcriptionally competent complexes [150].

AF1 of the AR has more recently been described to exist in a "collapsed disordered" conformation, with structural properties that resemble those of a molten globule [283]. Molten globules are characterized by native-like secondary structure elements but without adopting a stable tertiary structure. They are collapsed states of tightly packed proteins, but lack the close packing of side chains [284, 285].

Current models for the NTD, and in particular AF1, of the AR suggest that it exists as an ensemble of conformations with limited secondary and tertiary structure, which are primed to bind co-regulatory proteins (see Fig. 1.17). Regions with residual structure in the NTD might be stabilized by interdomain communication. In addition, DNA binding of the AR and binding of co-regulatory proteins to the NTD can induce a more helical conformation and/or stabilize a particular conformation from this ensemble. This further facilitates interaction of other co-regulatory proteins and the assembly of a transcriptionally competent AR complex. Such a model, in which

structure can be induced in the NTD and new binding sites can be created, allows for specificity and multiple target protein binding to the NTD. Possibly some interactions induce folding, whereas others require a more ordered conformation to interact with the NTD [150, 156].

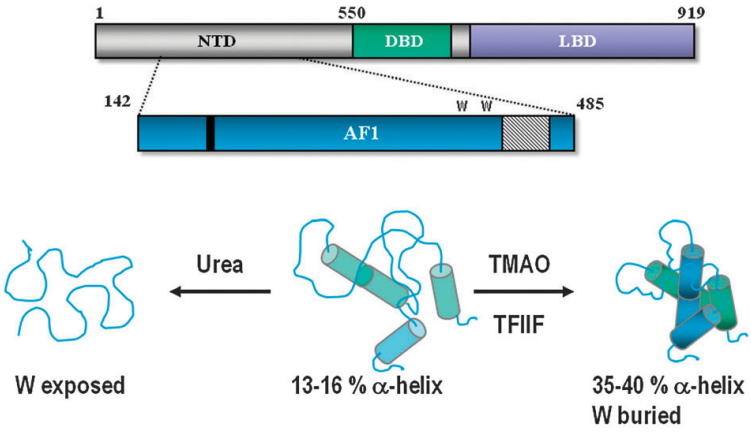


Figure 1.17: Current model of the structural properties of the AR NTD, and particularly the AF1 region. The AR NTD/AF1 is proposed to exist as an ensemble of conformations having limited stable structure, described as a "collapsed disordered" conformation. However, this domain will adopt a more helical conformation in the presence of the natural osmolyte TMAO or a co-regulatory binding protein (i.e. TFIIF). Folding or stabilization of the AR NTD/AF1 conformation is also regulated by DNA binding and possibly by PTMs. From [156].

The ID nature of the NTD is exquisitely suited for such a regulatory mechanism involving multiple protein-protein interactions that can vary depending on the context. ID allows interaction with high specificity without the need for high affinity, allowing transient regulatory interactions [151]. In addition, it allows interaction with multiple binding partners, possibly via the same region of sequence that can adopt a distinct conformation depending on the binding partner [286].

Even though the NTD adopts a more ordered conformation in the context of full-length AR and/or when it is bound to DNA and in the presence of co-regulators, experimental data indicate it is highly unlikely to adopt a stable globular fold like the DBD and the LBD. No electron density for the NTD was observed in crystals obtained in an attempt to crystallize full-length AR in the presence of DNA elements and co-activator motifs [287]. In addition, to date, there is only one structure available in which the NTD of a NR was observed when bound to DNA as a heterodimer with its binding partner. This is the cryo-electron microscopy (cryo-EM) structure of DNA-bound RXR-VDR on DR3, hexametric half-sites organized as direct repeats spaced by 3 nucleotides [220]. The NTD of VDR is short and was found to interact with the major groove of the DNA response element next to the recognition helix of VDR. In general, the NTD of NRs remains invisible in the structures obtained of (nearly) full-

length NRs bound to DNA, suggesting this domain remains structurally flexible when the full-length protein is bound to DNA [155].

Intrinsically disordered proteins Intrinsically disordered proteins (IDPs) or intrinsically disordered regions (IDRs) in proteins lack a well-structured three-dimensional fold but, in spite of this, they are often functional [151]. This has challenged the structure-function paradigm according to which a well-defined structure is required for protein function and has led us to adjust our concept of a functional protein to one that includes several conformations that are consistent with various aspects of its function [151, 152]. Although these proteins are abundant in all species, they are significantly more represented in eukaryotes than in prokaryotes [288]. About 10–15% of prokaryotic and about 15–45% of eukaryotic proteins are conservatively estimated to contain significant disorder, i.e. long disordered regions of at least 30 residues in length [289].

IDPs have been found to play crucial roles in several cellular processes involving signaling and regulation. Functions of IDPs or IDRs include the regulation of transcription and translation, cellular signal transduction, facilitating protein PTM, interaction with many different partners as hubs in protein-protein interaction networks and the regulation of the self-assembly of large multi-protein complexes such as the ribosome [151, 286].

Binding of an IDP/IDR with its target is often coupled to folding. Coupled folding and binding is the process in which an IDP or IDR folds into an ordered structure concomitant with binding to its target [151]. One example is the interaction between the ID kinase-inducible transcriptional activation domain (KID) of cyclic-AMP response element binding protein (CREB) with the KID-binding (KIX) domain of CREB-binding protein (CBP). The KID polypeptide is ID in solution, both as an isolated peptide and in full-length CREB [290, 291], but it folds into a pair of orthogonal helices upon binding to KIX [292] (see Fig 1.18a). Mutual synergistic folding of two IDPs upon binding has also been observed, for instance for the interaction of the activator for thyroid hormone and retinoid receptors (ACTR) domain of p160 co-activators and the nuclear receptor co-activator-binding domain (NCBD) of CBP [293] (see Fig. 1.18) and for the interaction of transcription factor c-Myc and its binding partner Max that fold into a basic helix-loop-helix leucine zipper [294]. In addition, some IDPs remain (largely) disordered even after binding to their targets [295]. An example is the interaction between the Smad2 Mad homology domains and the Smad binding domain (SBD) of Smad anchor for receptor activation (SARA). [296] Protein complexes with a significant amount of structural disorder or polymorphism have been termed “fuzzy complexes” [297]. Another example of such a fuzzy complex is shown in figure 1.18c, which repre-

sents a complex with more than one bound conformation for the interaction of T cell factor 4 (Tcf4) with β-catenin [297, 298].

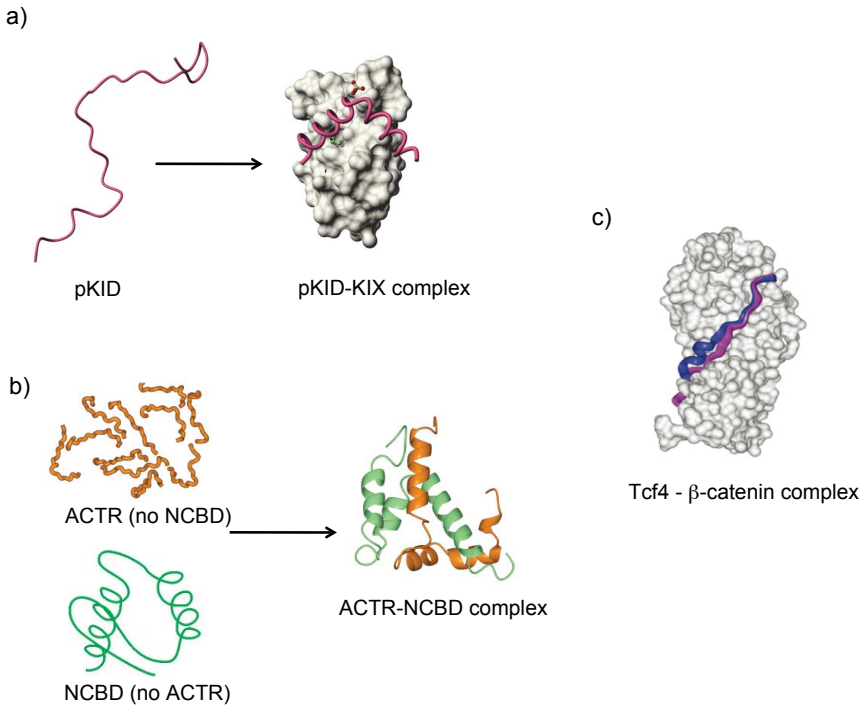


Figure 1.18: Examples of coupled folding and binding of IDPs upon interacting with their target proteins. a) pKID + KIX (from [151]), b) ACTR + NCBD (adapted from [151]) Uncomplexed ACTR is intrinsically disordered and is represented as a conformational ensemble (orange), whereas uncomplexed NCBD is highly helical but does not have a stable tertiary structure and is molten globule-like (green). c) Interaction of Tcf4 with β-catenin (PDB 1JDH, blue), where alternative contacts with the charge buttons could be realized. The alternative model is constructed based in the structure of Tcf3 in complex with β-catenin (PDB 1G3J, magenta) and biochemical data [297, 298]. (from [297])

While coupled folding and binding can involve a complete domain to fold into a globular structure upon interaction with its binding partner, most coupled folding and binding events involve relatively short amphipathic motifs contained within longer disordered sequences [151, 295, 299]. These folding motifs are called molecular recognition features (MoRFs) [300], also referred to as molecular recognition elements (MoREs), and can often fold into distinct structures on binding different target proteins. They are suggested to be short disordered regions that sample structured states within the conformational ensemble, and become fully ordered upon binding to the partner [300]. They have been found to fold into α-helix, β-strand, or form irregular structure on binding to a target protein [295].

ID offers specific advantages in molecular recognition. It allows for transient interactions to occur with high specificity, which is crucial for signal transduction. Proteins

involved in signal transduction must not only associate specifically to initiate the signaling process, but must also dissociate again when signaling is complete. Coupled folding and binding of an IDR allows for such high specificity in combination with a relatively low affinity resulting in highly specific and transient interactions [151]. Coupled folding and binding also allows the burial of extremely large surface areas even when the interacting domains are quite small. An example of this is the interaction between hypoxia-inducible factor-1α (HIF1α) and the TAZ1 domain of CBP [151, 301]. The HIF1α polypeptide is almost entirely wrapped around the TAZ1 domain in the complex, forming an extensive intermolecular interface (see Fig. 1.19, PDB 1L8C, [301]). This results in a binding affinity that is much higher than what would be achieved by interaction between two stable, folded proteins of comparable sizes. ID also maximizes allosteric coupling between binding sites when binding is accompanied by folding [302]. Consequently, ID has been proposed to play a crucial role in the allosteric communication of transcription factors [154].

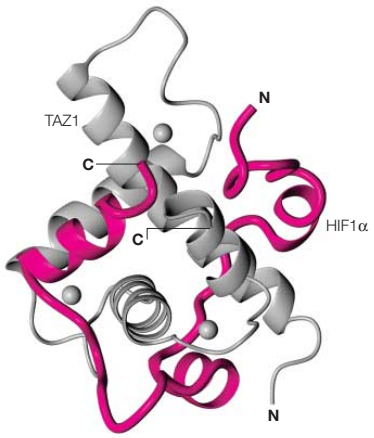


Figure 1.19: Structure of the TAZ1 domain of CBP (in grey) in complex with HIF1α (in pink). Spheres show the location of Zn atoms. PDB 1L8C, [301]. Figure from [151].

Binding of IDPs/IDRs to their targets is often regulated by PTMs, which can serve as biological switches [151]. Conformational flexibility in IDRs facilitates interaction with multiple targets. It allows for interaction of the protein with enzymes that can post-translationally modify it, in addition to the interaction with the physiological binding partner [151]. Signal transduction is often mediated through phosphorylation in signaling proteins and phosphorylation sites are frequently located in IDRs. Phosphorylation can act at different levels, affecting either the conformational propensities in the uncomplexed state or the interactions between the IDP and its binding partners. Further, phosphorylation can function as a switch to fine-tune the biological response [295, 303]. Interaction of the intrinsically disordered KID of CREB with the KIX domain of CBP, for instance, requires phosphorylation of KID at S133. Phosphorylation

does not affect the structural properties of the unbound KID [290], but pKID binds with a high affinity to KIX and folds into a pair of helices [292], whereas the affinity of unphosphorylated KID for KIX is much lower. IDPs also often play the role of hubs in protein-protein interaction networks. The ability of IDPs to interact with multiple binding partners and to adopt various conformations depending on the interacting protein makes them exquisitely suited as hubs [286].

ID also plays an important role in the assembly of large multi-protein complexes. On the one hand, IDPs/IDRs have a large capture radius that facilitates the diffusive search for a binding target through the so-called "fly-casting" mechanism [295, 304]. Thus, the IDP/IDR can recruit binding partners to incorporate them in a protein assembly. In addition, ID in regions connecting the various domains of multi-domain proteins also allows for sufficient structural flexibility for the various domains to interact with their binding partners while being part of a large assembly. On the other hand, the complexes IDPs form with their binding partners are usually composed of multiple interactions [305]. Multiple weak interactions that are transient in nature allow for rapid assembly or disassembly of protein complexes as a response to changes in the environment. Furthermore, some IDPs can interact with various partners simultaneously, which also results in inherent stability of the protein assembly [153]. ID plays a crucial role in transcription, since the assembly of the transcription machinery involves a complex set of highly specific and transient interactions with multiple partners that often have variable architecture. IDPs and IDRs are exquisitely suited to precisely coordinate such interactions, because of their specific molecular recognition features as described above [153].

Furthermore, ID is frequently present in proteins that are associated with disease, such as cancer, neurodegenerative diseases, cardiovascular diseases and diabetes [306, 307]. Consequently, IDPs are important targets for drug development although they are considered very difficult to target because they rarely form binding pockets such as those available on the surface of globular proteins [308]. Since they are often engaged in protein-protein interactions, therapeutic approaches that aim at interfering with these interactions via small molecules have been suggested for drug development [309]. This approach is viable, as exemplified by the successful development of nutlins, small molecules that inhibit the interaction between p53 and murine double minute 2 (MDM2), reactivating the p53 pathway in cancer cells [310]. However, in this case, the target to which nutlins bind is the globular MDM2 and not the disordered domain of p53. IDPs have also been successfully targeted themselves by small molecules, as demonstrated for the onco-protein c-Myc [311, 312], the oncogenic fusion protein EWS-Fli1 [313] and amyloid precursor protein (APP) and Aβ in Alzheimer's disease [314].

However, the rational design of small molecules to interact with IDPs/IDRs and interfere with specific protein-protein interactions remains challenging due to their highly dynamic nature [308].

1.2.5 Known mutations in AR

To date, over 1000 mutations of the AR have been reported and they are available in the Androgen Receptor Gene Mutations Database[3] [71]. Since the *AR* gene is located on the X-chromosome it is not necessary for survival. Therefore, germ line mutations are not lethal. Instead, germ line loss-of-function mutations in the AR frequently result in AIS [1, 162]. Somatic mutations, often associated with gain-of-function, are frequently observed in PCa. Various studies have reported few AR mutations in primary PCa tumors samples, whereas an increased number of AR mutations is present in advanced and/or metastatic PCa [4, 73, 315, 316]. Furthermore, it was shown that at least some of these mutations occur as an adaptive response to PCa treatment, including specific mutations in the LBD that allow promiscuous AR activation and mutations in the NTD that render the AR constitutively active [49, 78, 317]. In 2012, 159 AR mutations had already been identified in PCa tissue, and while the majority occurs in the LBD (ca. 45%), a large fraction of these mutations is present in the NTD (ca. 30%) [71]. This further highlights the importance of the NTD for the functional role of AR in CRPC survival mechanisms.

In a study to compare the AR mutations in patients that were treated with an anti-androgen (flutamide of bicalutamide) and patients that were not treated with an anti-androgen, Steinkamp *et al.* found that recurring missense mutations in metastatic PCa samples mostly occurred in the NTD. About half of these mutations occurred only following anti-androgen treatment while the other half was present both in patients that were treated with anti-androgens and patients that were not treated with anti-androgens (see Fig 1.20). Notably, some of these mutations occur in motifs in the NTD identified to be important for function, including mutations of residue "L194" (corresponds to L192 in the numbering system used in this thesis[4]) located in the ^{183}Lx$_7$LL^{192} motif, and "W435"[5] (W433), which is part of the ^{433}WHTLF^{435} motif.

Mutations in several regions of the AR described above to be important for function result in impaired AR regulation, including somatic mutations in the hinge region

[3] http://androgendb.mcgill.ca

[4] numbering Steinkamp *et al.*/numbering used in this thesis: E43/E43, Q58/Q58, Q86/Q84, S121/S119, L194/L192, E213/E211, T229/T227, A253/A251, A358/A356, R362/R360, G416/G414, W435/W433, T440/T438, G456/G454, G457/G455, R485/R484, T498/T497, A499/A498, P500/P499, V509/V508, E666/E665, R761/R760, R787/R786, Q868/Q867.

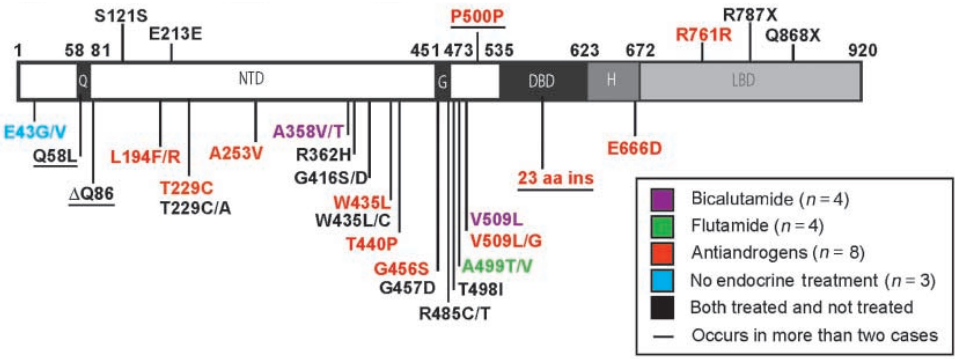


Figure 1.20: Recurring *AR* mutations from PCa metastases. Mutations found in multiple cases are shown. For codons carrying mutations to different amino acids, both changes are indicated. Color corresponds to the treatment group of the patients (see legend). From [78]. The numbering used to indicate the residues differs from the one used in this document.[5]

observed in AIS and PCa patients [84], mutations in the D-box constituting the DBD dimer interface [207] as well as mutations in the core Tau-1 region and the CHIP-interacting motif that have been described in PCa biopsies [184].

1.2.6 Known PTMs in AR

Many PTMs have been described in the AR and an overview of these is provided in figure 1.21. Several of the enzymes that modify the AR are differently regulated in PCa, which can lead to differences in the PTM pattern of the AR and corresponding changes in its regulation [83].

Acetylation or methylation of key residues in the NLS (residues K630, K632 and K633) plays an important role in the nuclear translocation of the AR. Two SUMOylation sites in the NTD (K386 and K520) are located in regions that were originally identified as negative regulatory motifs, NRM1 and NRM2, spanning residues ^{385}IKLE^{388} and ^{519}VKSE^{522}, respectively. [2] SUMOylation results in repression of the AR transcriptional activity. In contrast, ubiquitination at residues K845 and K847 promotes AR transcriptional activity. Many phosphorylation sites have been identified, most of which are present in the NTD. Phosphorylation is presumed to affect AR activity by increasing or decreasing protein interactions that occur proximal to the phosphosite. Depending on the phosphosite the AR can be phosphorylated in the absence or presence of androgen, which in some cases can lead to stimulation of AR activity in the absence of hormone [83].

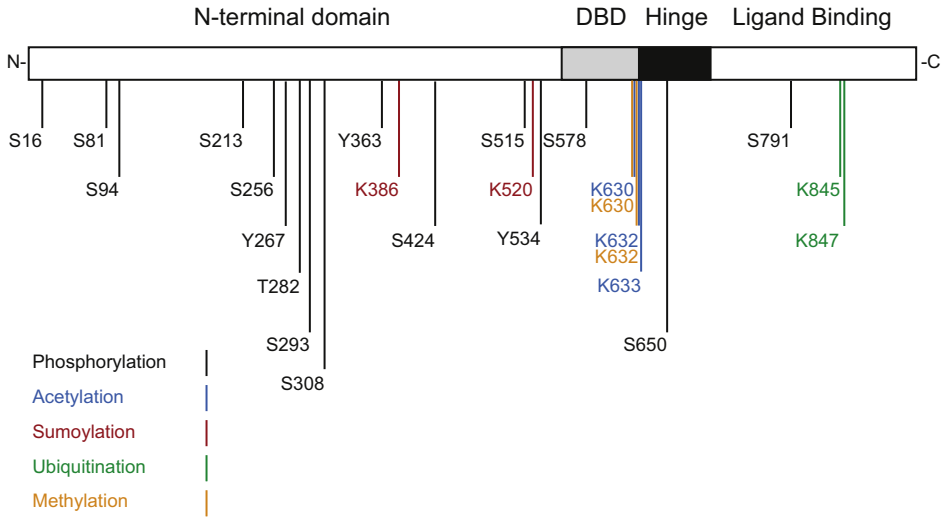


Figure 1.21: Overview of the PTMs of the AR. Phosphorylation is indicated in black, acetylation in blue, SUMOylation in red, ubiquitination in green and methylation in orange. From [83]. The numbering used to indicate the residues corresponds to the one used in this document, with the exception of T282 and S293, which in the numbering system used in this thesis correspond to T280 and S291, respectively.

1.3 The AR and transcription

1.3.1 Formation of the AR transcriptional complex

As is the case for other transcription factors, enhanced transcription by the AR depends on the recruitment of RNA polymerase II (pol II) and the general transcription factors (GTFs: TFIIFA, TFIIB, TFIID, TFIIE, TFIIF and TFIIH) to the promoter and/or enhancer sites of its target genes by ARE-bound AR. This results in the assembly of the PIC at the promoter site, which in eukaryotes is usually initiated by binding of TFIID to the TATA box found in most core promoters and is followed by either sequential recruitment of pol II and the other GTFs or recruitment of a pre-assembled pol II holoenzyme complex [318].

In the sequential assembly pathway, TFIID, which is composed of TATA-binding protein (TBP) and TBP-associated factors (TAFs), first binds to the promoter region, followed by the entry of TFIIA and TFIIB that help stabilize the promoter-bound TFIID. TFIIB further recruits pol II and TFIIF, which is tightly associated to the polymerase, which ensures specific interaction of pol II at the promoter. In turn, TFIIF recruits TFIIE and TFIIH that play a role in promoter clearance and the formation of an initiation competent pol II. TFIIH has ATPase activity for transcription initiation and promoter clearance, helicase activity for promoter opening, and kinase activity for

phosphorylating the C-terminal domain of pol II, which is required for transcription elongation. In the pre-assembled pathway, first a complex formed by TFIID and TFIIA binds to the promoter site, which is followed by the recruitment of a pre-assembled pol II holoenzyme complex containing pol II and the remaining GTFs [93, 318].

The AR can interact directly with components of the general transcription machinery (GTM) assembled in the PIC, or indirectly via a wide range of co-regulators that facilitate and mediate communication between the AR and the GTM.

Direct interactions between the AR and several proteins of the GTM have been reported. Biochemical binding assays demonstrated that the AR interacts via its NTD with the large subunit of TFIIF, RNA polymerase II-associated protein 74 (RAP74) [319, 320]. As previously mentioned when discussing the structural properties of the AR NTD, binding of RAP74 to the AR NTD was found to induce α-helical structure in the NTD and facilitated its interaction with peptides of the SRC-1 co-activator. [281, 282] The AR was further shown to interact directly with general transcription factor TFIIH under physiological conditions [321]. Moreover, overexpression of the cdk-activating kinase subunit of TFIIH was found to markedly stimulate AR-mediated transcription [321]. One important function of TFIIH is to phosphorylate the C-terminal domain (CTD) of the largest subunit of pol II (RPB1), which is required for efficient transcriptional elongation. Therefore, the interaction between AR and TFIIH may enhance phosphorylation of the CTD of RPB1, implicating a role for the AR in increasing the efficiency of transcriptional elongation. Alternatively, TFIIH might directly phosphorylate the AR and enhance AR transactivation, similar to the TFIIH-mediated phosphorylation of the ER which has been shown to enhance ER transactivation [322, 323]. Consistent with a potential role of the AR in transcriptional elongation, the AR was also found to interact with positive transcription elongation factor b (p-TEFb) [324]. The small subunit of p-TEFb, PITALRE (also known as cdk9), possesses protein kinase activity and is also able to phosphorylate the CTD of RPB1, which is necessary to progress from PIC formation on the promoter to transcriptional elongation [325, 326]. Remarkably, both TFIIH and p-TEBb display RPB1 CTD kinase activity, but these act at different stages of transcription. Phosphorylation by TFIIH is required for promoter clearance [327], while p-TEFb is required to prevent arrest of the pol II within a few hundred nucleotides of the promoter [328]. In addition, the AR interacts directly with pol II through association with its second largest subunit, RPB2 [329]. Furthermore, co-expression of RPB2, which is also involved in transcriptional elongation, was found to stimulate AR-mediated transcription of its target genes [329]. Taken together, the direct interaction of the AR with TFIIF, TFIIH, p-TEFb and RPB2 may increase the efficiency of RPB1 CTD phosphorylation and lead to enhanced

transcription of AR target genes. It was therefore proposed that the AR may modulate transcription of its target genes by regulating both transcriptional initiation and elongation events [93, 322]. Interestingy, both TFIIF and TFIIH interact with the AR NTD [175, 320, 321]. The AR domain to which the PITALRE subunit of p-TEFb binds was not determined. Nevertheless, PITALRE was found to interact with the AR in a ligand-independent manner under physiological conditions and was retained on a resin containing a histidine-tagged AR construct consisting of the AR NTD and DBD [324]. This suggests PITALRE interacts with the AR NTD or DBD. The RPB2 subunit of pol II was shown to mainly bind to the AR LBD [329].

In addition to the direct interactions described above, the AR can also indirectly communicate with the GTM via co-regulators. These are proteins that are recruited to gene promoter regions by the AR and either enhance (i.e. co-activators) or reduce (i.e. co-repressors) its transactivation. AR co-regulators do not generally bind to DNA themselves, but can facilitate DNA binding of the AR through chromatin remodeling or histone acetyltransferase activity. Furthermore, they can recruit additional co-regulators, influence the recruitment of GTFs to the promoter site and facilitate their assembly into a stable PIC, or form a bridge between the AR and the GTFs. In general, co-regulators display a wide range of functions and can affect the AR transcriptional activity through diverse mechanisms. More than 160 proteins were identified as potential AR co-regulators, and can exert their function either by binding directly to the AR, usually in a ligand-dependent way, or via indirect interaction through other co-regulators [93].

The members of the p160 or steroid receptor co-activator family (SRC-1, SRC-2 and SRC-3) have been found to bind directly to the AR and activate AR transactivation via histone acetyltransferase activity as well as by recruiting additional co-activators, including p300, the p300 homolog CBP, as well as p300/CBP-associated factor (p/CAF). These recruited co-activators all display histone acetylase activity that is intrinsically stronger than that found in the SRC co-activators. Acetylation of histones results in a less compact packing of nucleosomes and is associated with transcriptional activation. In addition, both p300 and p/CAF were found to acetylate the AR directly at lysine residues in the NLS (K630, K632, K633) [83]. Furthermore, p300, CBP and p/CAF can also directly interact with the AR [93].

The two best characterized AR co-repressors are NCoR and SMRT. As mentioned before, they bind via conserved LxxxIxxx(I/L) motifs to the AR AF2. In the absence of ligand, the AR does not strongly interact with NCoR or SMRT, but these co-repressors are recruited when the AR is bound to antagonists, such as hydroxyflutamide and bicalutamide [330–332]. Interestingly, also in its agonist-bound state, the AR can be

inhibited by NCoR and SMRT through direct AR binding [158, 333–335]. NCoR and SMRT recruit histone deacetylases (HDACs) that by deacetylating histones compact nucleosomes into tight and inaccessible structures, which results in transcriptional repression. Similar to NCoR and SMRT, also other co-repressors directly or indirectly recruit HDACs to promoter and enhancer sequences [158].

Most studies to reveal the assembly of the AR transcriptional complex were performed with the *PSA* gene as a model system. As is the case for most other AR target genes identified to date, androgen regulation of the *PSA* gene involves promoter and enhancer elements [93]. The *PSA* gene contains two AREs in the proximal promoter region (ARE I and ARE II) and one ARE in the enhancer region (ARE III), which is centered at approximately 4.2 kb [336, 337]. Even though the *PSA* promoter and enhancer region each display androgen responsiveness, maximal androgen regulation requires the involvement of both regions [336, 337]. It has been proposed that the formation of an activation complex involves recruitment of regulatory proteins to both the promoter and enhancer regions, whereas the formation of a repression complex only involves factors bound at the promoter [330]. Thus far, a rather limited set of co-regulators has been identified in the formation of the AR transcriptional complex at the *PSA* gene [97, 330, 338, 339]. These are mainly regulatory proteins that are common core components of the transcriptional complexes formed by many nuclear receptors and specific transcription factors, such as SRC-1, SRC-2, SRC-3, p300, p/CAF, Brahma-related gene 1 (BRG1), the MED1 subunit of the Mediator complex and coactivator-associated arginine methyltransferase 1 (CARM-1). BRG1 is an ATPase required for nucleosome repositioning by the SWI/SNF (mating type switching/sucrose nonfermenting) chromatin remodeling complex. The Mediator complex is a multi-subunit complex that plays an important role in AR transcriptional activation and contacts the AR directly via its MED1 subunit. Finally, CARM-1 is a histone methyltransferase, which does not interact directly with the AR, but instead is recruited as a secondary co-activator via interaction with mainly SRC co-activators. Interestingly, many of these proteins have been found to interact directly with the AR, and in particular with its NTD, including SRC1, SRC-2, SRC-3, p300 and MED1 [93, 175, 241, 246]. p/CAF also interacts directly with the AR, but the domain to which it binds has not been determined [93]. Upon androgen stimulation, the AR is recruited to both the promoter and enhancer of the *PSA* gene, followed by the recruitment of these co-activators and pol II. Contrary to pol II, which displays an overall higher occupancy at the promoter region, the relative abundance of the AR and its associated co-activators is higher at the enhancer region [93]. A model has been proposed in which the recruitment of the AR and its co-activators at both the promoter and enhancer regions creates a chromosomal loop that allows promoter-bound AR and enhancer-bound AR to share a common co-activator

complex. Furthermore, it permits pol II to track from the enhancer along the looped chromatin to the promoter region (see Fig. 1.22) [93, 339]. At present, it remains unclear whether the assembled co-activator complex at the *PSA* gene can be generalized to other AR target genes.

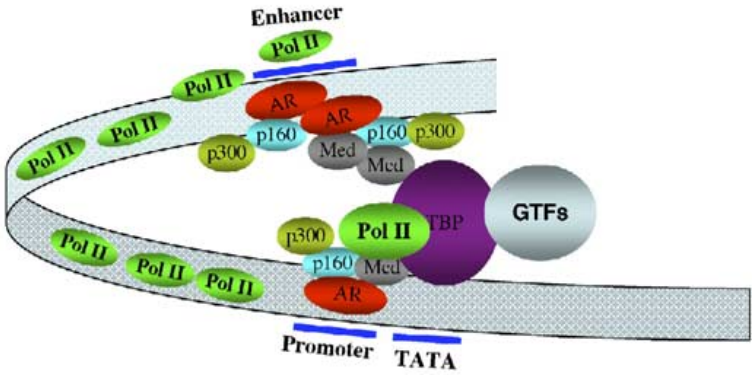


Figure 1.22: Model of the AR transcriptional complex assembled at *PSA* regulatory regions. The AR co-activator complex is predominantly recruited to the *PSA* enhancer, which communicates with the AR transcriptional complex weakly associated with the *PSA* promoter through the 4 kb chromosomal loop on which pol II tracks. Med: Mediator. From [339].

1.3.2 Interaction of the AR and TFIIF

TFIIF is a general transcription factor that tightly associates with pol II and facilitates its recruitment to the promoter where the PIC is formed. It contains two subunits, RNA polymerase II-associated proteins 30 (RAP30) and 74 (RAP74), that each can interact with several binding partners (see Fig. 1.23) [318].

The N-terminal domains of the RAP30 (residues 2–119) and RAP74 (residues 2–172) form a heterodimer [340]. The central region of RAP30 (residues 107–170) can interact with pol II, the region between residues 27 and 152 can interact with TFIIB, and its C-terminal domain (residues 162–249) can adopt a winged helix-turn-helix (winged HTH) fold between residues 175 and 243. The same is true for the C-terminal residues of RAP74 (residues 449–517). Winged HTH folds are commonly observed in eukaryotic DNA binding domains [318, 341]. The winged HTH motifs in both RAP30 and RAP74 indeed display cryptic DNA binding activity, which has been proposed to contribute to the stability of the pol II as part of the promoter complex by providing additional protein-DNA contact surfaces, and to reduce unspecific interactions with the DNA [318, 342]. However, as is indicated in figure 1.23, the various domains of both subunits can interact with a wide array of binding partners. The globular N-terminal domain of RAP74 can further interact with the TAF1 (TATA-binding protein associated factor

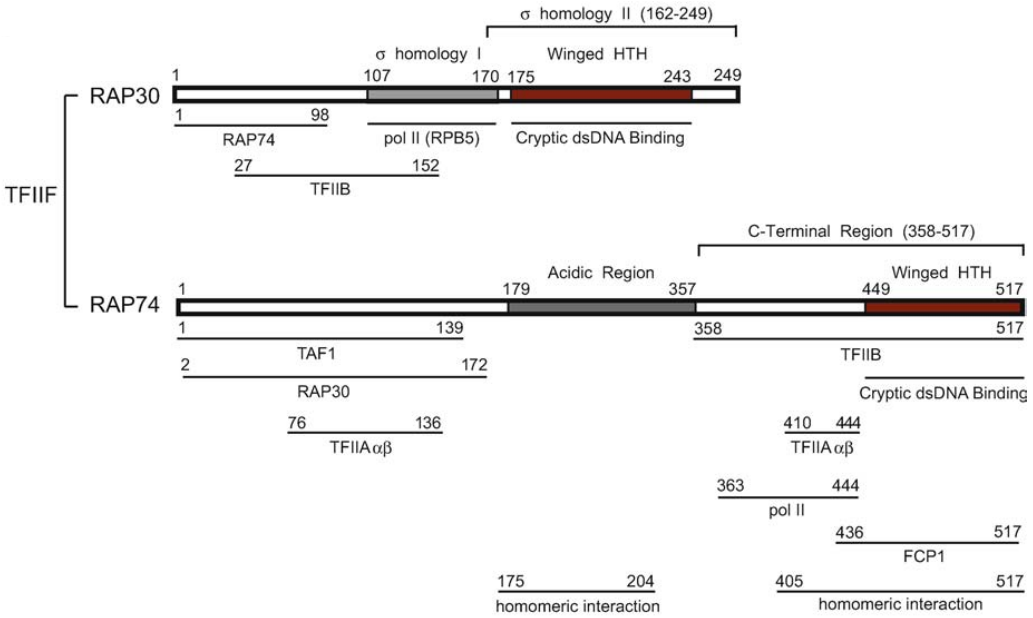


Figure 1.23: Schematic representation of the structural domains and interacting regions of general transcription factor TFIIF. The numbers above the schematic representation of both subunits corresponds to the boundaries of the different protein domains. Solid lines below the schematic representation of both subunits correspond to amino acid residues within TFIIF that are shown to interact with binding partners. winged HTH is winged helix-turn-helix motif; pol II is RNA polymerase II; TAF1 is the TATA-binding protein associated factor 1, a subunit from TFIID; FCP1 is TFIIF-associated CTD phosphatase. From [318].

1) subunit of TFIID and TFIIA$\alpha\beta$. Its central domain (residues 179–357) is the site of various phosphorylations and contains many negatively charged glutamate and aspartate residues and few hydrophobic amino acids and it is hypothesized to be exposed and unstructured within the PIC [343]. The C-terminal domain of RAP74 (residues 358–517) contains a disordered region of sequence (residues 358–449) as well as the winged HTH globular region (residues 449–517). It has been described to interact with TFIIB [344], TFIIA$\alpha\beta$ [318], pol II [345], FCP1 (TFIIF-associated C-terminal domain (CTD) phosphatase) [346] and the NTD of the AR [319, 320], in addition to DNA. FCP1 specifically dephosphorylates the carboxy-terminal domain of the RPB1 subunit of pol II to regenerate a hypophosphorylated form of pol II competent for (re)initiation of transcription [347]. Furthermore, RAP74 can also interact with itself through homodimeric interaction of regions of sequence either in the C-terminal domain (residues 405–517) or in the linker region (175–204) [342].

Recently, the architecture of the pol II transcription PIC was described, both in human [348] and in yeast [349]. Human transcription initiation was studied based on an *in vitro* reconstituted system by cryo-EM, whereas the study of the PIC complex in yeast, as revealed by cryo-EM and chemical cross-linking, reported on a complete set of GTFs, assembled with pol II and promoter DNA. According to both structures, the heterodimer formed by the N-terminal domains of RAP30 (or its yeast analog Tfg1) and RAP74 (or its yeast analog Tfg2) is bound to pol II, and the winged HTH domain of either RAP30 [348] or RAP74 (Tfg2) [349] was found to be located close to the DNA, and in contact with TBP, TFIIA and TFIIE [348] or TFIIA, TFIIB and TFIIE [349].

The structure of the winged HTH domain in both RAP30 and RAP74 has been solved (PDB 1BBY, [341] and 1I27, [350], respectively) (see Fig. 1.24). For RAP30, the winged HTH motif is comprised of residues 175–243 and contains an α-helix (H1, residues 179–193) followed by a short β-strand (S1, residues 196–197), leading to the HTH motif, which is composed of an α-helix (H2, residues 199–205), turn (residues 206–209), and another α-helix (H3, residues 210–219). Following the HTH motif, there is an anti-parallel β-sheet formed by β-strands between residues 221–225 and 231–234, that are connected by a short loop of 5 amino acids (residues 226–230), forming the "wing" of RAP30 [341].

Similarly, for RAP74, the winged HTH motif encompasses residues 449–517 and is formed by an α-helix (H1, residues 456–465) followed by the HTH motif, which is composed of a second α-helix (H2, residues 470–475), a long loop (residues 476–485) (containing one helix turn), and a third α-helix (H3, residues 486–500). This HTH motif is followed by a short anti-parallel β-sheet (β-strands correspond to residues 503–507 and residues 510–514) connected by a short turn (residues 508–509) that forms the

"wing".

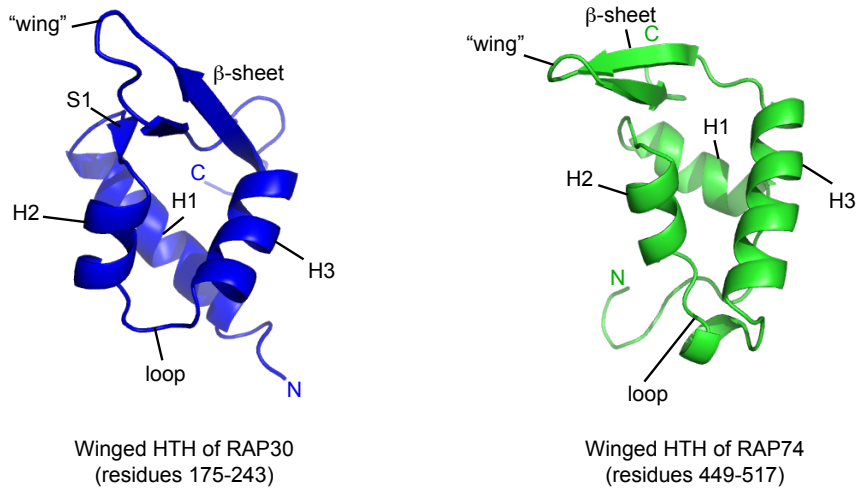


Figure 1.24: Winged helix-turn-helix of RAP30 (left) and RAP74 (right). The helices, turns (loops) and β-sheets characteristic of this fold are indicated. H is helix, S is strand. The N-terminal and C-terminal are indicated with N and C, respectively. The structure for the RAP30 winged HTH corresponds to PDB 1BBY, [341]; that for the RAP74 winged HTH to PDB 1I27, [350].

The interaction between the AR NTD and RAP74 has been studied by various biochemical binding assays and biophysical techniques. Furthermore, TFIIF was co-immunoprecipitated with the AR in cellular assays [351]. TFIIF was identified in a protein-protein interaction screening assay to selectively bind a construct of the AR spanning residues 143–494[5]. The AR construct interacted selectively with the RAP74 subunit of TFIIF and showed only modest binding to the RAP30 subunit. Furthermore, this interaction was shown to be sufficient to reconstitute AR transcription under cell-free conditions [319].

Interaction studies with a series of deletion mutants further revealed that AR 143–494 can interact both with the N-terminal 136 residues of RAP74 (RAP74-NTD) and with the C-terminal domain of RAP74 corresponding to residues 363–517 (RAP74-CTD). However, the main site of interaction was mapped to RAP74-CTD [320]. More recently, mutation of hydrophobic residues V490, L493 and L497 in helix 3 of the winged HTH motif of the RAP74-CTD to prolines was shown to selectively disrupt binding to AR 143–494 [352].

The binding site of RAP74-CTD in the AR NTD has not been determined. Nevertheless, based on mutational studies, residues M245, L247 and V249 of the AR and the surrounding residues were proposed to form part of the binding site, while residues S160

[5]In the original publications the construct is defined as AR residues 142–485, but this corresponds to residues 143–494 in the numbering system used in this thesis.

and S163, located in the ^{159}PSTLSL^{164} sequence, were suggested to have an indirect effect on the binding by altering the structural flexibility of the AR NTD [320, 353]. Mutation of residues S341 and S344, located in a similar ^{340}PSTLSL^{345} sequence, was found to only modestly affect RAP74-CTD binding [320].

Moreover, binding of RAP74-CTD was found to result in enhanced protease resistance of AR 143–494, consistent with an induced folding of the AR construct. FTIR experiments confirmed that AR 143–494 adopts a more helical conformation in the presence of RAP74-CTD, similar to the induced helicity observed in AR 143–494 by addition of TMAO [281, 282]. In addition, it was found that the interaction of AR 143–494 with SRC-1 peptides was facilitated in the presence of either TMAO or RAP74-CTD. These data strongly suggest that induced folding of the AR 143–494 construct by interaction with RAP74-CTD or the presence of TMAO leads to a conformation that enhances subsequent binding of SRC-1 [282].

The binding kinetics for the interaction of full-length RAP74, RAP74-CTD and RAP74-NTD with AR 143–494 were determined by surface plasmon resonance, for which the AR 143–494 construct was immobilized on the sensor chip surface using amine coupling chemistry and solutions of the RAP74 protein or constructs were passed over the immobilized AR construct. The binding affinity of RAP74 for AR 143–494 was determined to be $0.17 \pm 0.04\,\mu$M, that of RAP74-CTD was found to be $0.63 \pm 0.08\,\mu$M and that of RAP74-NTD was $0.29 \pm 0.09\,\mu$M [352].

Furthermore, EPI-001, the recently discovered small molecule inhibitor that selectively targets the AR NTD, impaired the interaction between AR 143–494 and RAP74-CTD by 21% [134].

1.4 Conclusion

Many of the described mechanisms that lead to CRPC sensitize the AR to very low levels of androgens, allow the AR to function promiscuously (activation by non-endogenous ligands, including other steroid hormones and antagonists that become agonists) or render it constitutively active (activation in the absence of ligand). They maintain the androgen signaling axis despite androgen depletion and the administration of AR antagonists. These mechanisms allow nuclear translocation of the AR and binding to DNA in the absence of endogenous ligand, but still depend on the AR to recruit coregulatory proteins and interact with the general transcription machinery to activate transcription of its target genes. Preventing these protein-protein interactions that are critical for transcription represents another opportunity to block the AR signaling axis.

This notion is particularly important given the fact that the two currently used therapeutic strategies that target the AR signaling axis (androgen synthesis inhibition and the use of AR antagonists) eventually fail due to acquired resistance. This is true in spite of renewed efforts that have led to secondary hormonal therapies and the development of highly selective and potent second-generation AR antagonists. Resistance to these novel agents also emerges and this resistance appears to be in part due to similar resistance mechanisms as those that arose to the first-generation drugs, i.e. antagonist conversion to agonist due to acquired mutations (enzalutamide) and promiscuous AR activation by other hormones (abiraterone acetate). In addition, overexpression of constitutively active splice variants has also been linked to the acquired resistance to these novel compounds, underlining the importance of the NTD for AR transactivation.

In general, the AR NTD, and particularly the AF1, appears to be the key to sustained survival and proliferation of tumors, as it is indispensable for AR transactivation, both in the presence and in the absence of androgens. In spite of that, it is not directly targeted by the current therapeutic strategies that are instead focused on the LBD. Full-length AR might be translocated to the nucleus and bind to DNA by mechanisms independent of ligand, but protein-protein interactions between the NTD and co-regulators as well as members of the transcription machinery stay crucial for the activation of transcription. The constitutive activity of AR splice variants is mediated via its NTD and independent on both androgens and the binding of any ligand to the AR. Consequently, the inhibition of androgen synthesis and the administration of potent AR antagonists do not influence the activity of this domain.

EPI-001, a recently discovered small molecule that targets the NTD, has been shown to effectively inhibit AR transactivation mediated by the NTD both for full-length AR and for splice variants. Furthermore, it was proposed to interfere with protein-protein interactions between the AR AF1 and binding partners such as co-regulators and members of the transcription machinery. This suggests that it is possible to target the androgen signaling axis via a third therapeutic strategy, namely preventing the protein-protein interactions that are crucial for transcription between the NTD and its binding partners.

The AR NTD is consequently a promising, if not the most promising, therapeutic target for CRPC at present. Although large screening assays have led to the identification of several compounds that interact with the NTD and inhibit its transactivation function, the ID nature of this domain severely hampers the rational design of new agents targeting it. Therefore, there is a pressing need to gain better insight into the molecular mechanisms by which the NTD activates transcription of AR target genes, both in the presence and absence of androgens. It will consequently be necessary to

characterize the structural properties of the NTD, and especially the AF1 region, both in isolation and when in complex with its binding partners. A more detailed molecular view on the role of the NTD in AR transactivation can potentially lead to unraveling the mode of action of the small compounds that target the NTD identified to date. This might allow the rational design of more potent compounds in the near future.

2 Objectives

There is a pressing need to understand the mechanisms of aberrant AR transactivation in CRPC patients as they allow cancer cells to proliferate. The NTD of AR harbors AF1, the most potent activation function of the protein, and is indispensable for both ligand-dependent and ligand-independent AR transactivation. Hence, it is one of the main therapeutic targets for PCa.

Targeting AF1, as opposed to AF2 in the well-defined LBD of AR, has, however, been little explored due to the lack of information about its structural and dynamical properties. This stresses the importance of gaining a better insight in the structure, dynamics and interactions of the AR NTD, as it is highly relevant to the study of androgen-independent activation of AR in CRPC.

The main objectives of my thesis project were therefore, firstly, to characterize the conformational properties of the intrinsically disordered NTD of the AR, and in particular of the AF1 region, and secondly, to study the interactions of this domain, at atomic resolution, with some of its biological binding partners and with small molecules that are drug candidates.

2.1 Characterization of the conformational properties of the intrinsically disordered NTD of the AR

To gain better insight into the molecular mechanisms by which this intrinsically disordered domain mediates the function of the AR and causes disease, we defined as an objective to perform a thorough NMR characterization of the conformational properties of this domain measuring NMR observables that allow the backbone assignment of AF1

and that report on its secondary structure and dynamics. This will be complemented with information from other biophysical techniques, such as CD.

2.2 Characterization of the interaction of the AR with biological binding partners involved in transcription

To investigate the role of the AR NTD in transcriptional activation at a molecular level, we wished to study its interaction with RAP74, a subunit of general transcription factor IIF (TFIIF), which has been reported to interact with the AR NTD [282, 319], by performing chemical shift perturbation experiments. These experiments do not only yield the exact binding epitope, but can also provide additional information, such as the binding affinity and induction of structural changes upon interaction.

2.3 Study of the interaction of the AR with small molecules that are drug candidates

To understand how the NTD can be targeted by small molecules, we wish to elucidate the mode of action of EPI-001, a small molecule that has been recently identified to be a potent AR inhibitor that interacts with the AR NTD and causes regression of CRPC [134].

3 Methodology

3.1 Molecular biology

3.1.1 Recombinant Gateway technology

The recombinant Gateway® Technology with Clonase™ II of LifeTechnologies (Cat. No. 12535-029 and 12535-037) was used for the cloning of four designed AR constructs (AF1*a, AF1*b, AF1*c and AF1*) in several expression vectors and for the generation of a RAP74 construct (RAP74 residues 450–517) with an N-terminal HisMBP-tag. This technology allows to clone DNA sequences of interest via recombination into a donor vector and to subsequently transfer it from the donor vector to a wide range of available destination vectors (bacterial expression vectors, mammalian vectors and baculoviros vectors).

The Gateway technology is based on the bacteriophage lambda site-specific recombination system which facilitates the integration of lambda into the *E. coli* chromosome. Recombination involves two major components: the DNA recombination sequences (attachment sites (*att* sites)) and the enzymes that mediate the recombination reaction, i.e. Clonase II enzyme mix.

Lambda recombination occurs between site-specific *att* sites: *att*B sites on the *E. coli* chromosome and *att*P sites on the lambda chromosome. The *att* sites serve as binding site for the recombination enzymes. These enzymes bring together the target sites to which they bind, cleave them and covalently attach the DNA again after recombination. The DNA segments flanking the recombination sites are switched, such that after recombination, the *att* sites are hybrid sequences comprised of sequences donated by each parental vector. For example, recombination between *att*B and *att*P sites gives

rise to *att*L and *att*R sites.

The two recombination reactions that constitute the basis of the Gateway technology, the BP reaction and the LR reaction, are based on this recombination principle (see Fig. 3.1). The BP reaction facilitates recombination of an *att*B substrate (*att*B-PCR product or a lineralized *att*B expression clone) with an *att*P substrate (donor vector) to create an *att*L-containing entry clone (see Fig. 3.1a). This reaction is catalyzed by BP Clonase II enzyme mix. The LR reaction facilitates recombination of an *att*L substrate (entry clone) with an *att*R substrate (destination vector) to create an *att*B-containing expression clone (see Fig. 3.1b). This reaction is catalyzed by LR Clonase II enzyme mix.

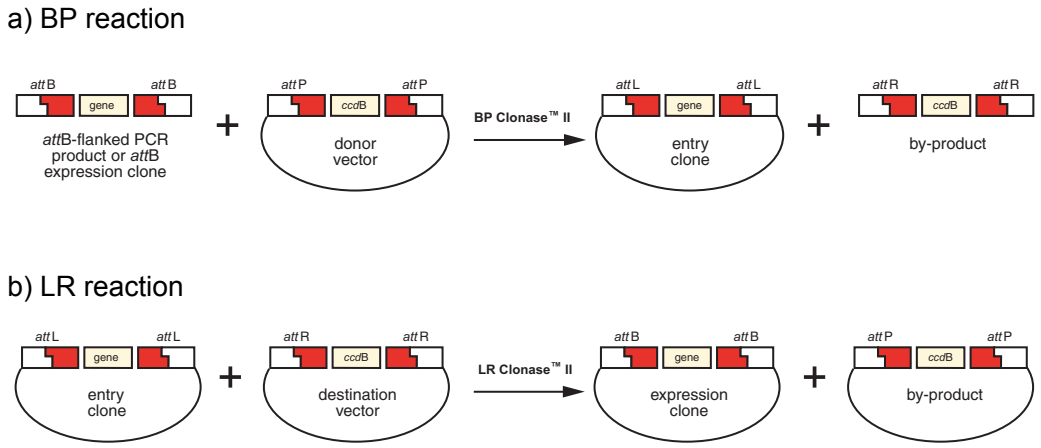


Figure 3.1: The Gateway recombination reactions. a) BP reaction to insert the gene of interest, flanked by *att*B recombination sites, in a donor vector that contains *att*P recombination sites. This creates an entry clone containing the gene of interest. b) LR reaction to further transfer the gene of interest from the entry clone (*att*L recombination sites) to a destination vector (*att*R recombination sites) to create an expression clone containing the gene of interest. Figure from the Gateway manual available on the website of LifeTechnologies.

pDONR/Zeo (LifeTechnologies, 12535035) was used as donor vector in this thesis. The designed AR constructs were further cloned in several destination vectors: pDEST17 (LifeTechnologies, 11803012), pDEST15 (LifeTechnologies, 11802014) and pDEST-HisMBP (Addgene, 11085). These are all bacterial expression vectors that produce an N-terminal fusion protein corresponding to the inserted gene and a specific N-terminal tag. The pDEST17 vector adds an N-terminal histidine tag (His-tag) to the expressed protein, the pDEST15 vector adds an N-terminal glutathione S-transferase tag (GST-tag) and the pDEST-HisMBP vector adds an N-terminal tag containing both consecutive histidine residues and maltose binding protein (HisMBP-tag). Both glutathione S-transferase and maltose binding protein are soluble proteins and for this reason GST-tags and MBP-tags are frequently used to promote solubility of the ex-

pressed protein of interest fused to the tag. The sequences of the N-terminal tags encoded by each of the destination vectors used are shown in appendix A.2.

To enable recombinational cloning and efficient selection of entry or expression clones most Gateway vectors, including the ones used in this thesis, contain two *att* sites flanking a cassette containing the *ccd*B gene for negative selection and a chloramphenicol resistance gene for counterselection. After a BP or LR recombination reaction, this cassette is replaced by the gene of interest to generate the entry clone and expression clone, respectively (see Fig. 3.1). The ccdB protein is lethal for most *E. coli* strains. Cells that take up unreacted vectors carrying the *ccd*B gene or by-product molecules retaining the *ccd*B gene will consequently fail to grow. This allows high-efficiency recovery of the desired clones.

To be suitable as substrate in a Gateway BP recombination reaction with a donor vector, the PCR product of the gene of interest needs to contain *att*B sites. This can be achieved by designing primers for PCR that create these *att*B sites. Recommendations for the design of both forward and reverse primers are provided by LifeTechnologies.

3.1.2 Cloning of the AR constructs

Four AR constructs were designed for biophysical studies of the AR AF1 (see chapter 4), corresponding to AR residues 142–275 (AF1*a), AR 265–340 (AF1*b), AR 330–448 (AF1*c) and AR 142–448 (AF1*). These were cloned into several bacterial expression vectors using the Gateway technology.

The protein constructs were designed to include an N-terminal fusion tag (His-tag, GST-tag or HisMBP-tag), followed by a TEV protease cleavage site (amino acids ENLYFQG) to remove the tag during purification, and the AR residues corresponding to the designed constructs. The TEV cleavage site was introduced in the PCR product before inserting the DNA in the donor vector. This required the insertion of 21 additional basepairs in the PCR product, apart from the *att*B sites that need to be introduced to allow recombination with the *att*P sites of the donor vector. Therefore, the PCR reaction was performed in two steps: PCR 1 and PCR 2 (see Fig. 3.2). In the PCR 1 reaction a large part of the TEV protease cleavage site was introduced N-terminal to the AR DNA sequence, whereas C-terminal to the AR sequence *att*B2 sites were introduced. In the subsequent PCR 2 reaction *att*B1 sites and the remaining part of the TEV protease cleavage site were introduced at the N-terminus and the *att*B2 site at the C-terminus was completed.

The forward and reverse primers used for PCR 1 and PCR 2 reactions of each of the

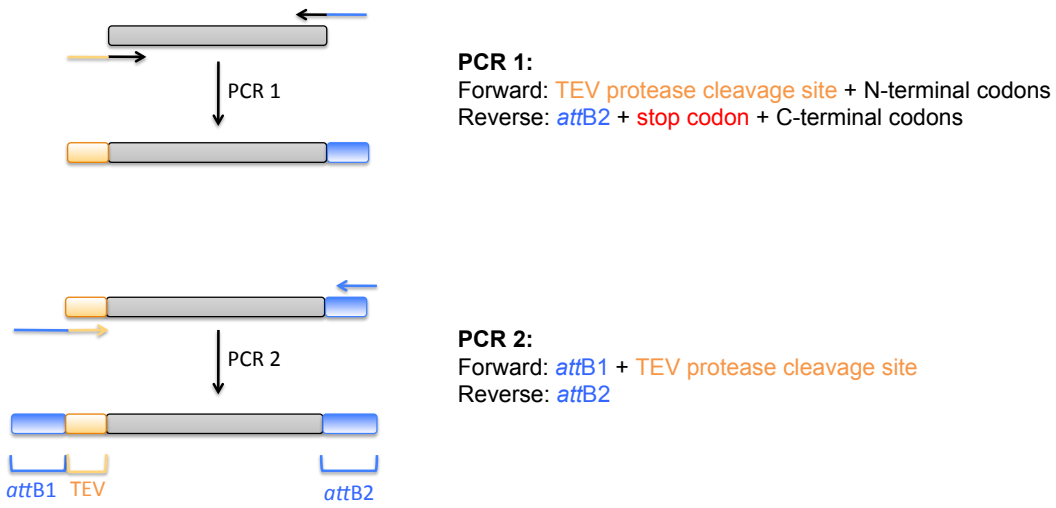


Figure 3.2: Schematic representation of the two consecutive PCR reactions used to obtain the PCR products that were inserted in the pDONR/Zeo vector. The *att*B sites are indicated in blue, the TEV protease cleavage site is indicated in orange and AR DNA sequence is indicated in grey. Arrows indicate the forward and reverse primers used in both PCR reactions. The colors of the arrows correspond to the sequence the codons code for (blue: *att*B1 or *att*B2, orange: TEV cleavage site, black: AR sequence).

constructs are shown in figure 3.3. For the PCR 1 reaction of the AF1* construct, the same forward primer as for the PCR 1 reaction of the AF1*a construct could be used as well as the same reverse primer as for the PCR 1 reaction of the AF1*c construct. For the PCR 2 reaction the same forward and reverse primers could be used for the four designed constructs because no AR specific sequence was introduced in this step. The primers were ordered from Sigma (synthesis scale: 0.025 μmol, purification: desalt; format: dry) and resuspended in autoclaved milliQ water to a final concentration of 100 μM.

PCR reactions and cloning were performed in collaboration with Dr. Eva Estébanez Perpiñá (IBUB, Spain) and according to the Gateway manual recommendations. For the AF1*c construct protocol 1 was used, whereas for AF1*a, AF1*b and AF1* protocol 2 was used (see below). The template DNA used for PCR 1 reactions was the pET-AR-AF1 vector containing human AR residues 143–494, kindly provided by Prof. Iain McEwan (University of Aberdeen, UK). The vector maps of pET-AR-AF1 and of the vector from which is was derived, pET-19bm, are shown in appendix A.3.

Protocol 1: used for PCR 1 and PCR 2 to generate the AF1*c construct
PCR 1 reaction
(final volume = 50 μL)
41 μL autoclaved milliQ

PCR 1

FORWARD

AR-Fwd-142:

```
5′-AAC CTG TAC TTC CAG GGC ggg ctg ccg cag cag ctg cca gca-3′
    N   L   Y   F   Q   G   G   L   P   Q   Q   L   P   A
```

AR-Fwd-265:

```
5′-AAC CTG TAC TTC CAG GGC tgc atg tac gcc cca ctt ttg gga-3′
    N   L   Y   F   Q   G   C   M   Y   A   P   L   L   G
```

AR-Fwd-330:

```
5′-AAC CTG TAC TTC CAG GGC gca gca ggg agc tcc ggg aca ctt-3′
    N   L   Y   F   Q   G   A   A   G   S   S   G   T   L
```

REVERSE

AR-Rev-275:

```
5′-GTA CAA GAA AGC TGG GTC cta ggg tgg aac tcc caa aag tgg-3′
                          stop P   P   V   G   L   L   P
```

AR-Rev-340:

```
5′-GTA CAA GAA AGC TGG GTC cta cgg cag ttc aag tgt ccc gga-3′
                         stop P   L   E   L   T   G   S
```

AR-Rev-448:

```
5′-GTA CAA GAA AGC TGG GTC cta aca cgg tcc ata caa ctg gcc-3′
                         stop C   P   G   Y   L   Q   G
```

PCR 2

FORWARD

```
5′-GGGG ACA AGT TTG TAC AAA AAA GCA GGC TCC GAA AAC CTG TAC TTC CAG-3′
                                            E   N   L   Y   F   Q
```

REVERSE

```
5′-GGGG AC CAC TTT GTA CAA GAA AGC TGG GTC cta-3′
                                          stop
```

Figure 3.3: Forward and reverse primers designed for PCR 1 and PCR 2 reaction to obtain the PCR product corresponding to AF1*a, AF1*b, AF1*c and AF1* to subsequently introduce in the pDONR/Zeo vector. For the PCR 1 reaction, three forward primers were designed (AR-Fwd-142 for AF1*a and AF1*, AR-Fwd-265 for AF1*b and AR-Fwd-330 for AF1*c) as well as three reverse primers (AR-Rev-275 for AF1*a, AR-Rev-340 for AF1*b and AR-Rev-448 for AF1*c and AF1*). For the PCR 2 reaction the same forward and reverse primer could be used for each of the constructs. The DNA sequence of the primers is indicated with the corresponding amino acids shown below. Blue corresponds to *att*B sites, orange to the TEV protease cleavage site, red to the stop codon and black to AR sequence.

1 μL AccuPrime *Pfx* DNA Polymerase (LifeTechnologies, 12344024)
5 μL 10X AccuPrime *Pfx* Reaction mix (LifeTechnologies, buffer supplied with the polymerase, already contains dNTPs)
1 μL of forward primer for PCR 1 (AR-Fwd-330)
1 μL of reverse primer for PCR 1 (AR-Rev-448)
1 μL of template DNA: pET-AR-AF1

PCR 2 reaction
(final volume = 50 μL)
41 μL autoclaved milliQ
1 μL AccuPrime *Pfx* DNA Polymerase (LifeTechnologies, 12344024)
5 μL 10X AccuPrime *Pfx* Reaction mix (LifeTechnologies, buffer supplied with the polymerase, already contains dNTPs)
1 μL of forward primer for PCR 2
1 μL of reverse primer for PCR 2
1 μL of template DNA: DNA obtained from PCR 1

The same PCR program was run for the PCR 1 and the PCR 2 reaction. The PCR equipment used was from Applied Biosystems (GeneAmp PCR System 2400).
one hold: 95 ℃, 2 min
30 cycles of:
    95 ℃, 30 s
    55 ℃, 45 s
    68 ℃, 5 min
two holds: 68 ℃, 15 min
lower temperature to 4 ℃

Protocol 2: used for PCR 1 and PCR 2 to generate the AF1*a, AF1*b and AF1* constructs
PCR 1
(final volume = 50 μL)
35.5 μL autoclaved milliQ
0.5 μL Taq polymerase Expand High Fidelity^{PLUS} PCR System (Roche, 03300226001)
1 μL dNTPs: dNTP Mix (KAPA Biosystems, KN1007)
10 μL 5X Expand High Fidelity^{PLUS} PCR System Reaction Buffer with $MgCl_2$ (Roche, supplied with the polymerase)
1 μL of forward primer for PCR 1 (AR-Fwd-142 for AF1*a and AF1* or AR-Fwd-265 for AF1*b)
1 μL of reverse primer for PCR 1 (AR-Rev-275 for AF1*a, AR-Rev-340 for AF1*b or

AR-Rev-448 for AF1*)
1 μL of template DNA: pET-AR-AF1

PCR 2
(final volume = 50 μL)
35.5 μL autoclaved milliQ
0.5 μL Taq polymerase Expand High Fidelity^{PLUS} PCR System (Roche, 03300226001)
1 μL dNTPs: dNTP Mix (KAPA Biosystems, KN1007)
10 μL 5X Expand High Fidelity^{PLUS} PCR System Reaction Buffer with $MgCl_2$ (Roche, supplied with the polymerase)
1 μL of forward primer for PCR 2
1 μL of reverse primer for PCR 2
1 μL of template DNA: DNA obtained from PCR 1

The PCR program was the same for the PCR 1 and the PCR 2 reaction. The PCR equipment used was from Applied Biosystems (GeneAmp PCR System 2400).
one hold: 94 ℃ , 2 min
30 cycles of:
    94 ℃, 30 s
    55 ℃, 45 s
    72 ℃, 5 min
two holds: 72 ℃, 10 min
lower temperature to 4 ℃

Both PCR 1 and PCR 2 products were run on an agarose gel to confirm they ran at the expected size. For the smaller DNA constructs (PCR 1 and PCR 2 products of AF1*a, AF1*b and AF1*c) agarose gels of 1% were prepared, whereas agarose gels of 1.5% were prepared for PCR 1 and PCR 2 products of AF1*. Agarose (Pronadisa/CONDA, 8014) was dissolved in TAE buffer (1X) (1% gel: 0.5 g agarose in 50 mL TAE buffer; 1.5% gel: 0.75 g agarose in 50 mL TAE buffer). 1.5–2 μL SYBR Safe DNA gel stain (10,000X) (LifeTechnologies, S33102) was added to 50 mL agarose before the gel was cast and solidified. Gel samples (18 μL) were prepared: 15 μL DNA + 3 μL DNA loading buffer (6X). The benchtop 1 kb dna ladder (Promega, g7541) was used as molecular marker. Agarose gels were run on a BIO-RAD device at 100 V for 30–45 min.

The obtained PCR 2 products were inserted into pDONR/Zeo vectors via the BP recombination reaction, following the Gateway protocol. The components (see below) were mixed, vortexed and spun down. The reaction was placed at 25 ℃ for 1 hour, after which 1 μL of Proteinase K solution (LifeTechnologies, supplied with the BP Clonase II enzyme mix) was added to terminate the reaction. After adding Proteinase K, the

reaction mixture was incubated for 10 min at 37 ℃. Following incubation, the BP products were transformed into OmniMAX 2 T1 phage-resistant cells[1] (purchased from the Protein Expression Core Facility (PECF), IRB Barcelona) (1 μL BP reaction to 20 μL cells) and plated on LB agar plates containing 50 μg/mL zeocin (LifeTechnologies, R250-01). OmniMAX 2 T1 phage-resistant cells are also commercially available from LifeTechnologies (C8540-03) and will be referred to as "OmniMAX cells" from now on. The plates were incubated for 16 hours at 37 ℃, after which they were stored at 4 ℃ until individual colonies were picked and grown in 5 mL Lennox L Broth (LB) medium (Melford, GL1703) with zeocin for selection (16 hours at 37 ℃ and 220 rpm). After 16 hours of growth, DNA was extracted from these saturated cultures using the NucleoSpin® Plasmid purification kit (Macherey-Nagel, 740588.250; miniprep). The concentration of the extracted DNA was measured with a NanoDrop Spectrophotometer and the purity was estimated based on the 260/280 ratio. Afterwards the DNA was sequenced (GATC Biotech).

BP reaction
(final volume = 10 μL)
6 μL TE buffer (1X)
1 μL pDONR/Zeo vector (LifeTechnologies, 12535035)
1 μL PCR 2 product
2 μL BP Clonase II enzyme mix (LifeTechnologies, 11789-020)

The BP products were further cloned into pDEST17, pDEST15 and pDEST-HisMBP vectors via the LR recombinant reaction, following the Gateway protocol. Similarly, the reaction components (see below) were mixed, vortexed and spun down, after which the reaction was placed at 25 ℃ for 4–5 hours. The reaction was terminated by adding 1 μL of Proteinase K solution, and further incubated for 10 min at 37 ℃. Following incubation, the LR products were transformed into OmniMAX cells (1 μL LR reaction to 20 μL cells) and plated on LB agar plates containing 100 μg/mL ampicillin (Melford, A0104-25G). The plates were incubated for 16 hours at 37 ℃, after which they were stored at 4 ℃ until individual colonies were picked and grown in 5 mL LB medium with ampicillin for selection (16 hours at 37 ℃ and 220 rpm). After 16 hours of growth, DNA was extracted from these saturated cultures using the NucleoSpin® Plasmid purification kit. The concentration of the extracted DNA was measured with a NanoDrop Spectrophotometer and the purity was estimated based on the 260/280 ratio. Afterwards the DNA was sequenced (GATC Biotech).

[1] For more information: see http://tools.lifetechnologies.com/content/sfs/manuals/oneshot_omnimax2_man.pdf

LR reaction
(final volume = 10 μL)
6 μL TE buffer (1X)
1 μL pDEST17 vector (LifeTechnologies, 11803012), pDEST15 vector LifeTechnologies, 11802014) or pDEST-HisMBP vector (Addgene, 11085)
1 μL BP reaction product
2 μL LR Clonase II enzyme mix (LifeTechnologies, 11791-020)

The sequenced expression vectors obtained from minipreps of the LR reaction products were stored at -20 °C. This DNA was further transformed into the Rosetta(DE3)pLysS *E. coli* expression strain (purchased from the PECF, IRB Barcelona) for expression of the various protein constructs. Rosetta(DE3)pLysS cells are also commercially available from Novagen/Merck (see Novagen user protocol TB009 Rev. F0104 for more information[2]), and will be referred to as "Rosetta cells" from now on.

For transformation into OmniMAX (for preserving DNA) or Rosetta cells (for expression) cells were purchased from the PECF at IRB Barcelona as 50 μL aliquots. Cells were thawed on ice and transferred to autoclaved eppendorfs. Depending on the concentration, 1 to 2 μL DNA (pDONR/Zeo, pDEST17, pDEST15 or pDEST-HisMBP clone) was added to one aliquot of cells. After incubation of 30 min on ice the DNA was internalized by a heat shock of 30 seconds at 42 °C, after which the cells were placed back on ice. 1 mL of autoclaved LB medium was added and cells were further incubated for 1 h at 37 °C and 500 rpm. After incubation, the cells were spun down gently (3 min at 3 krpm on the bench rotor), the supernatant was tipped off and the cells were gently resuspended in the remaining supernatant ($\pm$150 μL). The resuspended cells were plated on selective LB agar plates (for OmniMAX cells: 50 μg/mL zeocin for pDONR/Zeo clones and 100 μg/mL ampicillin for pDEST17, pDEST15 and pDEST-HisMBP clones; for Rosetta cells: 100 μg/mL ampicillin + 50 μg/mL chloramphenicol (Melford, C0113-25G) + 1% glucose for pDEST17, pDEST15 and pDEST-HisMBP clones). Colonies on the plates were left to grow for 16 hours at 37 °C.

In collaboration with Dr. Nick Berrow (PECF of IRB Barcelona) the AF1*a construct was further cloned in three additional vectors: pOPINS (N-terminal His-SUMO tag) [354, 355], pPEU10 (N-terminal His-thioredoxin tag with 3C protease cleavage site) and pPEU11 (N-terminal His-Z tag with 3C protease cleavage site).

[2] http://www.med.unc.edu/pharm/sondeklab/Lab%20Resources/manuals/novagen_competent_cells2.pdf

3.1.3 Cloning of the RAP74NMR construct

A protein construct corresponding to RAP74 residues 450–517 was also cloned into a pDEST-HisMBP expression vector via the Gateway technology. In this case, the pDONR clone containing the RAP74 sequence of interest and an N-terminal TEV protease cleavage site was directly purchased from GeneArt (now LifeTechnologies). The pDONR vector was pDONR221, which is very similar to pDONR/Zeo but carries kanamycin resistance instead of zeocin resistance. The RAP74 construct was further transferred to the pDEST-HisMBP vector via an LR recombination reaction, following the manufacturer's protocol. The LR reaction product was transformed into OmniMAX cells and plated on an ampicillin selective LB agar plate. After 16 hours at 37°C individual colonies were picked from the plate and grown in LB medium containing ampicillin. The DNA was extracted from these saturated cultures by minipreps and was sent for sequencing (GATC Biotech). The sequencing results confirmed the gene had been inserted correctly in the pDEST-HisMBP expression vector.

3.2 Protein expression and purification

Proteins were purified by fast protein liquid chromatography (FPLC) using an ÄKTA Purifier or ÄKTA Explorer system (GE Healthcare).

3.2.1 AF1*a

AF1*a could not be successfully produced via conventional expression methods in the *E. coli* Rosetta expression strain, in spite of a thorough screening for expression conditions of AF1*a cloned in the three Gateway vectors (pDEST17, pDEST15 and pDEST-HisMBP) and in three additional expression vectors (pOPINS, pPEU10 and pPEU11). Although AF1*a fused to various tags was expressed and was mainly present in the soluble fraction of the cell lysates, it appeared to already degrade during expression under all conditions tested. Therefore, it could not be purified from the cell lysates for further experiments (see chapter 4).

Using the Cell-free Expression EasyXpress Protein Synthesis Kit (Qiagen), His-tagged AF1*a was successfully expressed on a small test scale (not isotopically labeled) (see chapter 4). Double isotopically (^{13}C, ^{15}N) labeled His-tagged AF1*a was further expressed as part of a workshop organized by the University of Gothenburg (Prof. Göran Karlsson). Although the protein was expressed (see chapter 4), the yield was

not high enough to further purify it for subsequent NMR experiments. In addition, also in this case, degradation of the His-tagged AF1*a protein construct was observed.

In conclusion, AF1*a was never successfully expressed and purified in high enough yields to perform NMR experiments, despite a thorough screening for expression conditions in collaboration with the PECF at IRB Barcelona, and the use of cell-free expression techniques.

3.2.2 AF1*b, AF1*c and AF1*

Expression and purification protocols of His-tagged AF1*b, AF1*c and AF1* starting from pDEST17-AF1*b, pDEST17-AF1*c and pDEST17-AF1* were developed and optimized. Since the obtained protocols for these three fusion proteins are similar, a general protocol is described in detail below in which the differences for the various protein constructs are highlighted.

For the production of non-isotopically labeled proteins, LB medium was used, whereas the production of uniformly single (^{15}N) or double (^{13}C, ^{15}N) isotopically labeled proteins was performed in minimal MOPS medium to which $^{15}NH_4Cl$ was added as nitrogen source and ^{13}C-glucose (double isotopically labeled sample) or unlabeled (single isotopically labeled sample) glucose as carbon source (see section 3.2.2.4 for the preparation of MOPS medium).

3.2.2.1 Expression of isotopically labeled protein

The pDEST17-clone containing the AR construct of interest was transformed into Rosetta cells (see section 3.1.2 for the transformation protocol) and plated on an LB agar plate containing 100 μg/mL ampicillin for selection, 35 μg/mL chloramphenicol and 1% glucose (w/v). The colonies were grown for 16 hours at 37 °C, after which the plate was stored at 4 °C until a single colony was picked to inoculate 5 mL autoclaved LB medium containing 100 μg/mL ampicillin and 35 μg/mL chloramphenicol. Alternatively, a glycerol stock of the pDEST17-clone in Rosetta cells from a previous expression was used to inoculate 5 mL autoclaved LB medium containing the two antibiotics. The LB preculture was incubated for 16 hours at 37 °C and 250 rpm. Minimal MOPS medium (see section 3.2.2.4) to which $^{15}NH_4Cl$ was added as nitrogen source and ^{13}C-glucose (double isotopically labeled sample) or unlabeled (single isotopically labeled sample) glucose as carbon source was inoculated 1/50 with the saturated LB preculture and grown for 16 hours at 37 °C and 220 rpm to obtain a saturated MOPS preculture. A glycerol stock was prepared from the saturated LB preculture by adding

625 μL autoclaved 80% glycerol to 375 μL cells (final concentration glycerol is 50%), and was stored at -80 ℃.

After 16 hours of incubation of the MOPS preculture, minimal MOPS medium to which $^{15}NH_4Cl$ was added as nitrogen source and ^{13}C-glucose (double isotopically labeled sample) or unlabeled (single isotopically labeled sample) glucose as carbon source, and containing 100 μg/mL ampicillin and 35 μg/mL chloramphenicol, was inoculated 1/45 with the saturated MOPS preculture. Cells were grown at 37 ℃ and 220 rpm to optical density (OD) measured at 600 nm of 0.8 and induced with isopropyl β-D-1-thiogalactopyranoside (IPTG) (final concentration 0.5 mM for AF1*b, 1 mM for AF1*c and 1 mM for AF1*). Induced cultures were further incubated for 16 hours at 25 ℃ (AF1*b and AF1*c) or 20 ℃ (AF1*) and 220 rpm.

After 16 hours of incubation, the cell cultures were centrifuged (Beckman Avanti J-25 centrifuge, rotor JA-10, 6000 g, 4 ℃, 20 min), pellets were resuspended in lysis buffer without urea (see section 3.2.2.4 for buffer protocol) and protease inhibitor cocktail (PIC: P8849-5mL, Sigma) was added (500 μL PIC to resuspended cells from 2 L expression). The resuspended pellets were stored at -80 ℃ until purification.

3.2.2.2 Expression of non-isotopically labeled protein

The expression protocol for non-isotopically labeled His-tagged AF1*b, AF1*c and AF1* is essentially the same as for the isotopically labeled fusion proteins, with the exception that only one preculture was required (grown in LB medium). For the expression of non-isotopically labeled proteins, autoclaved LB medium containing ampicillin (100 μg/mL) and chloramphenicol (35 μg/mL) was inoculated 1/100 with the saturated LB preculture, obtained either from a glycerol stock of the pDEST17-clone of interest or from a new transformation of this pDEST17-clone in Rosetta cells. Cells were subsequently grown at 37 ℃ and 220 rpm to OD 0.8 and the expression protocol described above was further followed.

3.2.2.3 Purification

The frozen resuspended cell pellets were thawed and incubated with DNases (Sigma, D4527-10KU; 25 μL to cell lysate from expression of 2 L culture) for 2 hours at 4 ℃, while turning on a wheel. After incubation with DNases, the cells were lysed by one pass through a cell disruptor (TS1.1 kW model from Constant Systems ltd. (UK)) at 20 KPSI. Protease inhibitor phenylmethylsulfonyl fluoride (PMSF: Melford, MB2001) was added immediately to the lysed cells (500 μL at 100 mM PMSF to cell lysate from

2 L expression). Lysed cells were centrifuged (Beckman Avanti J-25 centrifuge, rotor JA-25.50, 25000 rpm, 4 ℃, 30 min) to separate the soluble from the insoluble fraction. The fusion proteins express as insoluble proteins and accumulate in inclusion bodies. The inclusion bodies in the pellet were resuspended in lysis buffer containing 8 M urea (see section 3.2.2.4 for buffer protocol) by leaving the pellets to which the buffer was added turning on a wheel for 16 hours at 4 ℃.

The resuspended pellets were passed again through the cell disruptor, at 25 KPSI to break the inclusion bodies, and centrifuged afterwards (Beckman Avanti J-25 centrifuge, rotor JA-25.50, 25000 rpm, 4 ℃, 30 min). The protein was recovered in the soluble fraction, which was filtered (pore diameter filter: 0.22 μm) before loading on a HisTrap HP 5mL column (GE Healthcare) for Ni^{2+} affinity chromatography. The column was equilibrated in lysis buffer with 8 M urea and proteins samples were loaded at a flow rate of 3 mL/min. The His-tagged fusion proteins were retained on the column. After loading the sample, lysis buffer with 8 M urea was passed at 5 mL/min (the double volume of the loaded sample volume) to remove unspecifically bound species. Next, the protein was eluted with an imidazole gradient (1 mM imidazole to 500 mM imidazole over 60 column volumes (CV), i.e. from 0% elution buffer with 8 M urea to 100% elution buffer with 8 M urea, see section 3.2.2.4 for buffer protocol).

The eluted His-tagged protein was concentrated (AF1*: Amicon Ultra Centrifugal Filters, 10,000 MWCO (Millipore, UFC901024); AF1*b and AF1*c: Vivaspin 20, 5,000 MWCO (Sartorius, VS2012)) and subjected to two steps of dialysis against cleaving buffer (see section 3.2.2.4 for buffer protocol) to remove urea and imidazole prior to TEV protease cleavage. After dialysis and before TEV protease cleavage, EDTA was added to final concentration 0.5 mM. Subsequently, TEV protease (purchased from the PECF at IRB Barcelona) was added (1/50 (mol/mol) for AF1*b and AF1*c and 1/10 (mol/mol) for AF1*) to cleave the His-tag (incubation at 4 ℃ for 16 hours or at 25 ℃ for 3 hours). Efficient cleavage of the His-tag was confirmed by SDS-PAGE (sodium dodecyl sulfate polyacrylamide gel electrophoresis).

Following TEV cleavage, the protein was passed again over the HisTrap HP 5mL column for reverse Ni^{2+} affinity chromatography. For AF1*b and AF1*c, the column was equilibrated in lysis buffer with 8 M urea and 8 M urea was added to the protein sample before it was loaded on the HisTrap HP 5mL column. For AF1*, the reverse Ni^{2+} affinity chromatography was run in lysis and elution buffers without urea. Uncleaved His-tagged protein constructs, the His-tagged TEV protease and the cleaved His-tag were retained on the column, whereas the cleaved protein construct was present in the flow-through of the load.

The obtained cleaved protein constructs were further purified by size exclusion chromatography (SEC) (HiLoad 26/60 Superdex 75 prep grade 17-1070-01 column (GE Healthcare)). The column was equilibrated in SEC buffer (see section 3.2.2.4 for buffer protocol) and the sample was passed at 2 mL/min. Finally, pure eluted protein was concentrated (AF1*: Vivaspin 20 5,000 MWCO (Sartorius, VS2012); AF1*b and AF1*c: Amicon Ultra Centrifugal Filters 3,000 MWCO (Millipore, UFC900324)), protease inhibitors (PIC and PMSF) were added 1/500 (v/v) each and the protein was flash frozen and stored at -80 °C. The sequences of the obtained AR protein constructs are shown in appendix A.2.

At several steps of the purification, the quality of the protein samples was assessed by SDS-PAGE. NuPAGE Novex 4–12% Bis-Tris protein gels were purchased (LifeTechnologies, NP0323BOX) as well as the corresponding running buffer (NuPAGE MES SDS running buffer, LifeTechnologies, NP0002), the loading buffer (LDS sample buffer, LifeTechnologies, NP0007) and NuPAGE sample reducing agent (LifeTechnologies, NP0009). Gel samples were prepared by adding 3.5 μL loading buffer and 1.5 μL sample reducing agent to 10 μL protein sample and heating the samples for 5 min at 98 °C. Gels were run with the default running parameters according to the NuPAGE protocol. SeeBlue®Plus2 Pre-Stained Standard (LifeTechnologies, LC5925) was used as a ladder.

3.2.2.4 MOPS medium and buffers

MOPS minimal medium
MOPS medium was prepared based on the protocol by Neidhardt *et al.* [356].
(MOPS = 3-(N-morpholino)propanesulfonic acid, 4-morpholinepropanesulfonic acid)

For the preparation of 500 mL MOPS minimal media, the following components were mixed:

- 440 mL autoclaved milliQ water
- 50 mL of 10X MOPS
- 5 mL of 0.132 M K_2HPO_4
- 5 mL of filter-sterilized glucose solution (7.56 g glucose in 21 mL milliQ water)
- 2.5 mL of $^{15}NH_4Cl$ filter-sterilized solution (1.05 g $^{15}NH_4Cl$ in 10.5 mL milliQ water)
- 500 μL chloramphenicol (stock: 35 mg/mL)

- 500 μL ampicillin (stock: 100 mg/mL)

For expression of single isotopically labeled protein (^{15}N) unlabeled glucose was used, while for the expression of double isotopically labeled protein (^{13}C, ^{15}N) ^{13}C-labeled glucose was used.

The 10X stock of MOPS minimal medium was prepared as follows.

1. Add 83,72 g of MOPS (Sigma, M1254-250G) and 7.17 g tricine (Sigma, T0377-250G) to ca. 300 mL autoclaved milliQ water in a 1 L beaker with a stir bar.
2. Add 10 M KOH to a final pH of 7.4 and adjust the total volume to 440 mL by adding autoclaved milliQ water.
3. Prepare a solution of 10 mM $FeSO_4.7H_2O$ (28 mg $FeSO_4.7H_2O$ in 10 mL autoclaved milliQ) and add it to the MOPS/tricine solution.
4. Add the following solutions to the MOPS/tricine/$FeSO_4.7H_2O$ solution in the indicated order.
   10 mL of 0.276 M K_2SO_4
   0.25 mL of 0.02 M $CaCl_2.2H_2O$
   2.1 mL of 2.5 M $MgCl_2$
   100 mL of 5 M NaCl
   0.2 mL of micronutrient stock (see below)
   447 mL of autoclaved milliQ water
5. Filter-sterilize under sterile conditions, using a filter with pore size 0.22 μm.
6. Store in 500 mL volumes at room temperature.

The micronutrient stock (100 mL) was prepared by mixing the components listed below in 40 mL autoclaved milliQ water and bringing the total volume to 50 mL. It was stored at room temperature.

ammonium molybdate ($(NH_4)_6Mo_7O_{24}.H_2O$): 0.009 g
boric acid (H_3BO_3): 0.062 g
cobalt chloride ($CoCl_2$): 0.018 g
cupric acid ($CuSO_4$): 0.006 g
manganese chloride ($MnCl_2$): 0.040 g
zinc sulfate ($ZnSO_4$): 0.007 g

The 0.132 M K_2HPO_4 stock solution was prepared by solubilizing 23 g K_2HPO_4 in 1 L milliQ water. The stock solution was autoclaved and stored at room temperature. The stock solution of glucose was prepared just before use by solubilizing 7.56 g glucose

(^{13}C-labeled or unlabeled) in 21 mL autoclaved milliQ water (for 2 L expression) and sterilizing it by passing it through a filter with pore diameter 0.22 μm. Likewise the stock solution of $^{15}NH_4Cl$ was prepared just before use by solubilizing 1.05 g $^{15}NH_4Cl$ in 10.5 mL autoclaved milliQ water (for 2 L expression) and sterilizing it by passing it through a filter with pore diameter 0.22 μm.

BUFFERS

Lysis buffer without urea
20 mM Tris-HCl pH 8.0
500 mM NaCl
1 mM imidazole
0.05% NaN_3
β-mercaptoethanol (added 1/1000 (v/v))
adjust pH to 7.8 and filter-sterilize (0.22 μm)

Lysis buffer with 8 M urea
20 mM Tris-HCl pH 8.0
500 mM NaCl
1 mM imidazole
8 M urea
0.05% NaN_3
β-mercaptoethanol (added 1/1000 (v/v))
adjust pH to 7.8 and filter-sterilize (0.22 μm)

Elution buffer without urea
20 mM Tris-HCl pH 8.0
500 mM NaCl
500 mM imidazole
0.05% NaN_3
β-mercaptoethanol (added 1/1000 (v/v))
adjust pH to 7.8 and filter-sterilize (0.22 μm)

Elution buffer with 8 M urea
20 mM Tris-HCl pH 8.0
500 mM NaCl
500 mM imidazole
8 M urea
0.05% NaN_3
β-mercaptoethanol (added 1/1000 (v/v))

adjust pH to 7.8 and filter-sterilize (0.22 μm)

Cleaving buffer
50 mM Tris-HCl pH 8.0
1 mM dithiothreitol (DTT)
filter-sterilize (0.22 μm)

SEC buffer
20 mM sodium phosphate pH 7.4
1 mM tris(2-carboxyethyl)phosphine hydrochloride (TCEP)
adjust pH to 7.4 and filter-sterilize (0.22 μm)

3.2.3 RAP74-CTD

The pET-23d plasmid encoding the RAP74-CTD construct [357] was kindly provided by Prof. Iain McEwan (University of Aberdeen, UK). The RAP74-CTD protein construct corresponds to the C-terminal residues of RAP74 (residues 363 to 517) and contains a C-terminal His-tag (LEHHHHHH). The expression and purification protocol of RAP74-CTD was previously published by Lavery *et al.* [352]. RAP74-CTD was produced according to this protocol to which small changes were made. Since RAP74-CTD does not contain cysteine residues, no reducing agents were added to the buffers used for protein purification.

The pET-23d-RAP74-CTD plasmid was transformed into the *E. coli* Rosetta expression strain (see transformation protocol above). Cultures were grown in LB medium with ampicillin (100 μg/mL) for selection at 37 °C and 220 rpm to OD 0.4–0.6. After induction by 1 mM IPTG the cells were further incubated at 37 °C and 220 rpm for 1.5 hours. Bacterial cells were pelleted by centrifugation and resuspended in lysis buffer (20 mM Tris-HCl pH 7.9, 500 mM NaCl, 5% glycerol (v/v), 5 mM imidazole). PIC was added (250 μL to resuspended pellet from 2 L expression) and the resuspended pellets were stored at -80 °C until purification.

For purification, the resuspended cell pellets were thawed and the cells were lysed by one pass through the cell disruptor at 20 KPSI. 250 μL at 100 mM PMSF was added immediately to the lysed cells (cell lysate from 2 L expression), as well as $MgCl_2$ to a final concentration of 10 mM and $MnCl_2$ to a final concentration of 1 mM. Finally, DNases were added (25 μL to cell lysate from expression of 2 L culture) and the cell lysate was incubated with the DNases for 2 hours at 4 °C turning on a wheel. After 2 hours of incubation, the cell lysate was centrifuged and the filtered soluble fraction was purified by Ni^{2+} affinity chromatography (HisTrap HP 5mL column). His-tagged

protein was eluted with an imidazole gradient (5 mM imidazole to 200 mM imidazole over 60 CV. Elution buffer: 20 mM Tris-HCl, pH 7.9, 500 mM NaCl, 5% glycerol (v/v), 200 mM imidazole), following a wash step of 6 CV with lysis buffer containing 40 mM imidazole. The obtained protein was further purified by SEC (HiLoad 26/60 Superdex 75 prep grade 17-1070-01 (GE Healthcare)). The run was performed at 2 mL/min in 20 mM sodium phosphate buffer with 500 mM NaCl and 0.5 mM EDTA at pH 7.4. Pure protein was concentrated, flash frozen and stored at -80 °C. Prior to NMR experiments with AR constructs, the protein was dialyzed against 20 mM sodium phosphate buffer with 1 mM TCEP at pH 7.4. The sequence of the obtained RAP74-CTD protein construct is shown in appendix A.2.

3.2.4 RAP74NMR

The RAP74NMR construct (RAP74 450–517) was expressed as a HisMBP-fusion protein from the pDEST-HisMBP-RAP74NMR clone described above. For the production of non-isotopically labeled RAP74NMR the protein was expressed in LB medium, whereas isotopically single (^{15}N) or double (^{13}C, ^{15}N) labeled RAP74NMR protein was expressed in minimal MOPS medium to which $^{15}NH_4Cl$ and ^{13}C-labeled glucose (double labeled sample) or unlabeled glucose (single labeled sample) were added as nitrogen and carbon source.

The protocol for non-isotopically labeled RAP74NMR production is described below. To produce isotopically labeled RAP74NMR, similar changes to the protocol need to be applied as those made in the AR construct protocols when isotopically labeled protein constructs were produced instead of non-isotopically labeled ones. Since RAP74NMR does not contain cysteine residues, no reducing agents were added to the buffers used for protein purification.

Bacterial cultures (Rosetta) were grown in LB medium with ampicillin (100 μg/mL) for selection at 37 °C and 220 rpm to OD 0.7. At OD 0.7, 1 mM IPTG was added for induction, after which the cultures were further incubated at 37 °C and 220 rpm for 3 hours. After 3 hours, the bacterial cultures were pelleted by centrifugation and resuspended in resuspension buffer (50 mM Tris-HCl pH 8.0, 1 M NaCl, 10 mM imidazole). PIC was added (500 μL to pellet from 2 L expression) and the resuspended pellets were stored at -80 °C until purification.

For purification, resuspended cell pellets were thawed and incubated with DNases (50 μL to cell lysate from expression of 2 L culture) for 2–3 hours at 4 °C, turning on a wheel. Following incubation with DNases, cells were lysed by one pass through the cell disruptor at 20 KPSI. 500 μL at 100 mM PMSF was added immediately to the cell

lysate (cell lysate from 2 L expression). The cell lysate was centrifuged and the soluble fraction was purified by Ni^{2+} affinity chromatography (HisTrap HP 5mL column). The HisMBP-tagged fusion protein was eluted by an imidazole gradient (10 mM imidazole to 300 mM imidazole in 60 CV; binding buffer: 50 mM Tris-HCl pH 8.0, 1 M NaCl, and elution buffer: 50 mM Tris-HCl pH 8.0, 1 M NaCl, 300 mM imidazole) following a washing step with binding buffer containing 10 mM imidazole.

After Ni^{2+} affinity chromatography, the eluted protein was pooled, concentrated and dialyzed for 16 hours against 50 mM Tris-HCl pH 8.0, 200 mM NaCl to remove imidazole and to reduce the concentration of NaCl. After dialysis, EDTA was added to final concentration 0.5 mM, as well as 1/50 (mol/mol) TEV protease. HisMBP-RAP74NMR was incubated with TEV protease for 16 hours at 4 ℃. After cleavage, the sample was filtered prior to subjecting it to a reverse Ni^{2+} affinity chromatography step (binding buffer: 50 mM Tris-HCl pH 8.0, 1 M NaCl, and elution buffer: 50 mM Tris-HCl pH 8.0, 1 M NaCl, 300 mM imidazole). The cleaved HisMBP-tag, uncleaved HisMBP-RAP74NMR protein and the His-tagged TEV protease were retained on the column, while the cleaved RAP74NMR protein was present in the flow-through.

After the reverse Ni^{2+} affinity chromatography, the RAP74NMR protein was concentrated and subjected to SEC (HiLoad 26/60 Superdex 75 prep grade 17-1070-01 (GE Healthcare)). The SEC column was equilibrated with 20 mM sodium phosphate buffer with 0.05% NaN_3, at pH 7.4 (unlabeled or isotopically labeled RAP74NMR) or at pH 6.5 (isotopically labeled RAP74NMR). If residual DNA was present in the sample after the reverse Ni^{2+} affinity chromatography, 1 M NaCl was added to the buffer used for SEC. The protein was passed over the column at 1 mL/min. The pure eluted protein was concentrated, flash frozen and stored at -80 ℃. If 1 M NaCl was added before SEC an extra dialysis step against 20 mM sodium phosphate buffer with 0.05% NaN_3, at pH 7.4 or pH 6.5 was performed to remove the NaCl.

If residual DNA was still present after SEC, the protein was further purified by an additional ion exchange chromatography (IEC) step (SOURCE15Q column, GE Healthcare (anionic exchanger)). In this case, the protein eluted from SEC was dialyzed against 20 mM Tris-HCl pH 8.0 prior to IEC. Pure RAP74NMR protein was dialyzed against 20 mM sodium phosphate, 0.05% NaN_3, at pH 7.4 (unlabeled or isotopically labeled RAP74NMR) or at pH 6.5 (isotopically labeled RAP74NMR), flash frozen and stored at -80 ℃. The sequence of the obtained RAP74NMR protein construct is shown in appendix A.2.

3.3 Synthetic peptides

Peptides were purchased from the Peptide Synthesis Unit of the Unitat de Química Combinatòria from the Parc Científic de Barcelona (WT, W433A/L436A/F437A, E440K/E441K, pS430, pS431, pS432, pS430/pS432 peptides (see table 5.1 on p. 184), R2' peptide (see table 6.4 on p. 262) and cterFCP1 peptide (see section **??**)) or from GenScript (pS424, C-Cys (see table 5.1 on p. 184) and R1' peptide (see table 6.4 on p. 262)). They were ordered on a 10–14 mg scale and with purity >95%.

The peptides were delivered as lyophilized powder. They were solubilized by adding autoclaved milliQ water and adjusting the pH with NaOH solution (1 M and 0.1 M) to above the isoelectric point (pI) for each of the peptides. Solubilized peptides were quantified by amino acid analysis by the Unitat de Tècniques Separatives i Síntesi de Pèptids from the Centres Científics i Tecnològics (CCiT) (University of Barcelona). A UV spectrum was measured for peptides containing aromatic amino acids and the extinction coefficient was calculated based on the concentration determined by amino acid analysis. The concentration of further prepared solutions of these peptides could thus be determined based on the absorption at 280 nm with the previously obtained extinction coefficient. Quantified peptides were aliquoted, lyophilized and further stored at 4 °C. For short term storage peptide solutions were flash frozen and stored at -20 °C.

Extra care was taken for peptides containing cysteine residues (C-Cys and R2') and phosphorylated peptides (pS430, pS431, pS432 and pS430/pS432). With these peptides high pH should be avoided, as at high pH the -SH group of cysteine can be deprotonated and disulfide bonds can form readily. Since both the C-Cys and R2' peptide only contain one cysteine residue, such disulfide bonds would form intermolecularly and lead to dimers of the peptides. Also, at high pH, the phosphate group of the phosphoserines can be removed by a β-elimination reaction. Therefore, the identity of these peptides was confirmed by MALDI-TOF MS (see below) before experiments were recorded. MALDI-TOF MS confirmed that the phosphate group(s) was/were present in the phosphopeptides and that the peptides containing cysteine residues were present as monomers.

The R2' peptide (see table 6.4 on p. 262) only contains one charged residue and is 24 amino acids long. As expected, it was poorly soluble without the addition of a co-solvent. To obtain a soluble R2' peptide sample at 25 μM it was necessary to use 1% deuterated dimethyl sulfoxide (DMSO-d_6) as co-solvent. The lyophilized peptide was directly solubilized in 100% DMSO-d_6 and to this solution 20 mM sodium phosphate buffer containing 1 mM TCEP and 10% D_2O at pH 7.4 was added to obtain a final sample containing 1% (v/v) DMSO-d_6. For further NMR experiments to investigate

the interaction of the R2' peptide with EPI-001, the small compound was added to the peptide sample from an EPI-001 stock in 100% dioxane-d_8. The final sample for NMR contained 25 μM R2' peptide, 100 μM EPI-001, 1% DMSO-d_6 and 0.5% dioxane-d_8 in 20 mM sodium phoshate buffer, 1 mM TCEP and 10% D_2O (pH 7.4).

3.4 Spin labeling protocol

Paramagnetic spin labels can be introduced in proteins or peptides at a specific position, by covalent interaction with the -SH group of a cysteine residue, to conduct paramagnetic relaxation enhancement (PRE) studies. Most commonly, an MTSL ((1-oxyl-2,2,5,5-tetramethyl-Δ3-pyrroline-3-methyl) methanethiosulfonate) spin label is used (see Fig. 3.4), which contains a nitroxide radical that causes enhanced relaxation of nearby nuclei. This can be detected by NMR and provides information about which residues are close in space, possibly only a fraction of the time, to the position of the protein or peptide to which the spin label is attached.

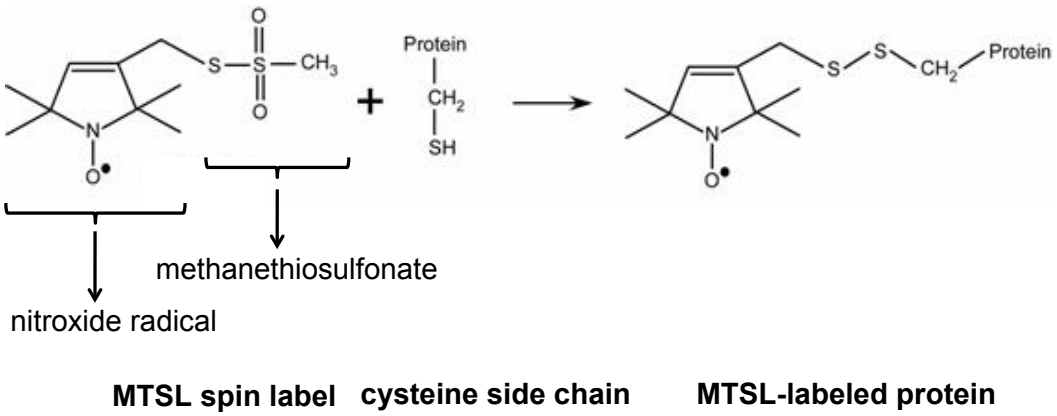


Figure 3.4: Spin labeling reaction with an MTSL spin label that binds covalently to the side chain of a cysteine residue in the protein or peptide, to give a disulfide bond adduct.

To attach an MTSL spin label to the cysteine amino acid in the AR C-Cys peptide (see table 5.1 on p. 184), the following protocol was used. Both the free -SH group of cysteine and the MTSL radical are readily oxidized in the presence of oxygen. Therefore, the MTSL stock was prepared fresh and under argon and all the buffers used were carefully degassed under argon to avoid the presence of oxygen in all steps required for spin labeling. MTSL (Toronto research chemicals, O875000) is also light sensitive and therefore reactions were performed in the dark. First, the C-Cys peptide was solubilized, quantified, aliquoted and lyophilized as described above. To ensure that the cysteine residue would have a reduced -SH group, β-mercaptoethanol was added prior to lyophilization. As β-mercaptoethanol is volatile, it is theoretically removed during

the lyophilization step. Lyophilized C-Cys peptide was solubilized in degassed 10 mM sodium phosphate buffer, pH 8.0, containing 0.01% NaN_3, and the absence of disulfide bonds was confirmed by MALDI-TOF MS (see below). To assure no trace amounts of β-mercaptoethanol, that could compete with the spin label for reaction with the -SH group of the cysteine residue, were present, the solubilized peptide was passed over a PD MidiTrap G10 desalting column (GE healthcare, 28-9180-11, cut-off = 700 Da) equilibrated with degassed 10 mM sodium phosphate buffer, pH 8.0, containing 0.01% NaN_3 prior to the addition of MTSL. The MTSL stock (250 mM) was prepared under argon. The acetonitrile in which the MTSL was dissolved was passed through three "freeze-pump-thaw" cycles to remove all oxygen. Briefly, the acetonitrile was placed in a round bottom flask that was subsequently sealed, it was then flash frozen in liquid nitrogen (freeze) after which the atmosphere above the frozen acetonitrile was removed by vacuum (pump) and replaced by argon. While under argon, the acetonitrile was thawed again (thaw). This procedure was repeated three times. Free -SH groups of the C-Cys peptide were reacted with a five-fold molar excess of MTSL solubilized in acetonitrile, for 70 min at room temperature and 500 rpm in the dark (eppendorf was covered with tin foil). The reaction was followed by HPLC (LC-MS) and yielded 60% MTSL-labeled C-Cys peptide, while 40% of the peptide remained unreacted. Addition of an additional five molar equivalents MTSL did not result in further reaction. To remove unreacted MTSL, the end product of the reaction was passed over a PD MidiTrap G10 desalting column equilibrated in degassed 10 mM sodium phosphate buffer, pH 8.0, containing 0.01% NaN_3 immediately after the reaction and the eluted peptide (60% MTSL-labeled and 40% free peptide) was lyophilized overnight (covered in tin foil to shield it from light). The following day, the lyophilized peptide was solubilized in degassed 10 mM sodium phosphate buffer, pH 8.0, containing 0.01% NaN_3 to a final concentration of 1470 μM. Note that through this procedure the final buffer was 20 mM sodium phosphate, pH 8.0, and 0.02% NaN_3. This peptide solution was mixed 1/1 with ^{15}N-RAP74NMR at concentration 100 μM in buffer 20 mM sodium phosphate, pH 8.0, 0.02% NaN_3 and 20% D_2O, yielding a final NMR sample of 50 μM ^{15}N-RAP74NMR + 785 μM C-Cys peptide (60% MTSL-labeled and 40% unlabeled) in 20 mM sodium phosphate, pH 8.0, 0.02% NaN_3 and 10% D_2O. Subsequently, the pH was lowered to 6.5 and the sample was measured immediately by NMR.

After measuring the [^{1}H,^{15}N]-HSQC spectrum of the oxidized sample, the spin label was reduced by the addition of three molar excess ascorbic acid (2 μL of 438 mM). Reaction was allowed to take place for 1 hour at room temperature in the dark. The pH was readjusted to 6.5 before recording the [^{1}H,^{15}N]-HSQC spectrum of the reduced sample.

3.5 Native gels

The native gels were prepared based on the "basic-native gel protocol" for acidic and neutral proteins (pI<7) from Dr. Mario Lebendiker[3], with small modifications (the pI of AF1* is 4.55). The loading buffer (5X), running buffer (1X) and resolving gel (10% acrylamide) were prepared based on the Lebendiker protocol. For the resolving gel 30% acrylamide:bis acrylamide at 1:29 was used. No stacking gel was used. Staining solution and destaining solution were prepared as follows and stored at room temperature, covered by tin foil.

Staining solution
500 mL methanol
100 mL acetic acid
400 mL milliQ water
2.5 g Coomassie Brilliant Blue

Destaining solution
150 mL methanol
100 mL acetic acid
750 mL milliQ water

Gel samples were prepared by adding 4 μL loading buffer (at room temperature) to 16 μL protein sample (kept on ice before addition of loading buffer). After adding loading buffer, the samples were loaded immediately and gels were run for 20.5 hours at 4 °C and at 40 V (BIO-RAD device), after which they were stained and destained using the staining and destaining solution, respectively.

3.6 NMR spectroscopy

3.6.1 NMR as a tool to study IDPs and biomolecular interactions

Nuclear magnetic resonance (NMR) spectroscopy has emerged as the main tool for the analysis of the conformational properties of disordered proteins [358–362]. The structural flexibility of IDPs or IDRs prohibits their characterization by X-ray crystallography. NMR is unique in its ability to yield residue specific information for IDPs and IDRs that can be related to their structural properties such as secondary structure propensity, interatomic distances and torsion angles. It is therefore the technique of

[3] http://wolfson.huji.ac.il/purification/Protocols/PAGE_Basic.html

choice to study the structural properties of the AR NTD.

Furthermore, NMR is an exquisite tool to study biomolecular interactions at atomic resolution. NMR chemical shift perturbation experiments can be used to reveal the binding epitopes of binding partners at single residue level and to determine the binding affinity. Furthermore, this technique allows to detect induced conformational changes upon binding, which are commonly observed for IDPs upon binding to their target.

3.6.2 General

NMR experiments were acquired on a Bruker Avance III UltraShield Plus 600 MHz spectrometer equipped with a triple resonance TCI cryoprobe (referred to as 'Bruker 600US' from now on) or on a Bruker DRX 800 MHz spectrometer equipped with a triple resonance TCI cryoprobe (referred to as 'Bruker 800' from now on).

Deuterated DSS (DSS-d_6: 3-(trimethylsilyl)-1-propanesulfonic acid-d_6 sodium salt (Isotec, 613150)) was added to the samples (30 μM) as a reference compound. The DSS-d_6 ^{1}H chemical shift corresponds to 0.000 ppm and is not dependent on temperature or pH (pH between 2 and 11) [363]. Indirect referencing in the ^{15}N and ^{13}C dimensions was based on the conversion factors from Wishart *et al.* [363]. All samples contained 10% deuterated water (D_2O), necessary to lock the magnetic field [362]. Samples for saturation transfer difference (STD) experiments (see below) were prepared in fully deuterated buffer to minimize the residual water signal.

To allow complete equilibration of the sample temperature, the measurements were only started 30 minutes after the samples were introduced into the NMR spectrometer that was already at the temperature at which the experiment was conducted. Standard Bruker pulse sequences, in some cases slightly modified, were used for the acquisition of most experiments (see below). The acquired spectra were processed using NMRDraw and NMRPipe [364]. The processed spectra were further analyzed in CcpNmr Analysis [365].

Experiments for AR constructs (AF1*b, AF1*c and AF1*) were acquired at 278 K and in 20 mM sodium phosphate buffer containing 1 mM TCEP and at pH 7.4, whereas NMR experiments of RAP74NMR were acquired at 298 K and in 20 mM sodium phosphate buffer at pH 7.4 or 6.5. For interaction experiments between AR constructs and RAP74NMR, 1 mM TCEP was added to the buffer to ensure reduction of the cysteine side chains present in AR constructs. When AR peptides were used for the interaction study with RAP74NMR, addition of 1 mM TCEP was not required as neither the peptides nor RAP74NMR contain cysteine residues.

The hard 90° pulse for ^{1}H was calibrated immediately before the experiment of choice using a standard Bruker zg pulse sequence. The carrier position (O1) for optimal water suppression was carefully determined based on a zgpr experiment recorded directly before the experiment of choice.

3.6.3 Experiments used in this thesis

3.6.3.1 1D NMR experiment for interaction study

The effect of AR constructs or peptides on the resonances of EPI-001 was monitored using one-dimensional (1D) ^{1}H experiments that were acquired on the Bruker 600US spectrometer. The Bruker pulse sequence zggpw5 was used for optimal shimming and water suppression. The samples contained 0.5% (v/v) dioxane-d_8, and in the case of the R2' peptide an additional 1% DMSO-d_6, and were acquired at 278 K, which complicated shimming. A delay (D1) of 7 s was used to ensure complete relaxation of the nuclei before the next acquisition cycle was started. Spectra were acquired with a sweep width (SW) of 11 ppm, with 128 scans and receiver gain (RG) 512.

3.6.3.2 [^{1}H,^{15}N]-HSQC

The [^{1}H,^{15}N]-HSQC (Heteronuclear Single Quantum Correlation) experiment provides a fingerprint spectrum of the protein and is the NMR experiment that was most frequently used in this thesis. This heteronuclear NMR experiment correlates the amide proton and amide nitrogen of each amino acid in the protein backbone [366]. It involves excitation of the proton (^{1}H), followed by transfer of magnetization to the ^{15}N heteronucleus and subsequent reverse polarization transfer to the proton for detection, as it is highly advantageous from a sensitivity point of view to detect magnetization on the proton [366]. In the resulting [^{1}H,^{15}N]-HSQC spectrum, a cross peak is visible for every amino acid of the protein (construct) under study, except for proline residues and the amino-terminal residue. Proline side chains do not contain an amide proton and therefore do not have a corresponding peak in an [^{1}H,^{15}N]-HSQC spectrum. The same is true for the amino-terminal residue of the protein (construct) that contains an NH_3^+ group at physiological pH and is not coupled to the carbonyl of a previous residue. Each cross peak corresponds to the chemical shifts of the amide proton and amide nitrogen from a particular residue of the protein backbone. Backbone assignment of the AR and RAP7 constructs used in this thesis (see below) revealed the identity of the corresponding amino acid for every observed peak in the [^{1}H,^{15}N]-HSQC spectra.

These assigned spectra formed the basis of many NMR-based experiments in this thesis. Residue specific changes observed in these spectra as a response to, for example, the addition of a binding partner (protein or small molecule) revealed the binding epitope of AR and RAP74, and of AR and EPI-001, and could be used to determine the binding affinity of AR and RAP74. Furthermore, the effect of specific mutations could be monitored using [^{1}H,^{15}N]-HSQC experiments.

Uniformly single ^{15}N-labeled protein samples were used to measure [^{1}H,^{15}N]-HSQC experiments. For interaction studies, the ^{15}N-labeled partner was observed by NMR, whereas the binding partner was not labeled with ^{15}N and therefore NMR-silent. Bruker pulse sequences fhsqcf3gpph or hsqcetfpf3gpsi2, the latter with sensitivity enhancement, were most commonly used to acquire [^{1}H,^{15}N]-HSQC spectra. The water suppression for both pulse sequences is based on a watergate scheme.

[^{1}H,^{15}N]-HSQC spectra of AF1* were acquired at 278 K on the Bruker 800 spectrometer, using 2048 and 512 complex incremental data points in F2 (^{1}H) and F1 (^{15}N) dimensions, with 8 to 64 scans per increment depending on the concentration. A SW of 10 ppm for ^{1}H and 22 ppm for ^{15}N was used. The concentration of AF1* was 50 μM for the interaction study with RAP74NMR and 25 μM for the interaction study with EPI-001. [^{1}H,^{15}N]-HSQC spectra of AF1* were measured at concentrations between 10 μM and 680 μM. [^{1}H,^{15}N]-HSQC spectra of AF1*b, AF1*c and RAP74NMR were acquired on the Bruker 600US spectrometer, at 278 K for the AR constructs and at 298 K for RAP74NMR. [^{1}H,^{15}N]-HSQC spectra of AF1*b were measured using 2048 and 512 complex incremental points in F2 (^{1}H) and F1 (^{15}N) dimensions, with 16 scans per increment and SW of 10 ppm for ^{1}H and 22 ppm for ^{15}N. Concentrations were 150 μM and 200 μM. For AF1*c, 2048 and 512 complex incremental data points in F2 (^{1}H) and F1 (^{15}N) dimensions were used, with 8 to 64 scans per increment depending on the concentration. The SW was set to 10 ppm for ^{1}H and 22 ppm for ^{15}N. [^{1}H,^{15}N]-HSQC spectra of AF1*c were measured at concentrations between 10 μM and 465 μM, and for interaction studies with RAP74NMR, 50 μM of AF1*c was used. [^{1}H,^{15}N]-HSQC spectra of RAP74NMR were measured using 2048 and 256 complex incremental points in F2 (^{1}H) and F1 (^{15}N) dimensions, with 16 scans per increment and SW of 10 ppm for ^{1}H and 24 ppm for ^{15}N. The concentration of RAP74NMR used for interaction studies with AR protein constructs and peptides was 50 μM.

3.6.3.3 Triple resonance experiments for backbone assignment

Three-dimensional (3D) triple resonance experiments were used to assign the backbone of AF1*b, AF1*c, AF1* and RAP74NMR. These experiments serve to establish the

sequential connectivity between neighboring residues and to subsequently assign the corresponding amino acids to the obtained strings of neighboring residues [367]. The experiments that reveal these connectivities correlate the amide proton H^N and nitrogen N^H of a given residue to the carbon atoms (C_α, C_β and carbonyl CO) of the preceding residue (i-1) and/or of the same residue (i) (see Fig. 3.5). They are organized in pairs with one experiment giving information about both intra-residue (i) and inter-residue (i-1) correlations and a second experiment that only provides information about the inter-residue (i-1) correlation. The sequential connectivity can be followed through the C_α [HNCA and HN(CO)CA], the C_β [CBCANH and CBCA(CO)NH] and the carbonyl CO [HNCO and HN(CA)CO] nuclei. For example, the HNCA experiment correlates each amide proton and nitrogen with both the intra-residue and inter-residue C_α, while the HN(CO)CA experiment only correlates this amide proton and nitrogen with the inter-residue C_α. Because the three sets of triple resonance experiments (based on C_α, on C_β or on CO) all provide information on the sequential connectivity, ambiguities, due to for instance overlapped signals, can generally be resolved. Segments of connected spin systems are then mapped onto the sequence of the protein on the basis of characteristic chemical shift values for specific amino acid types.

Standard Bruker pulse sequences were used to acquire 3D triple resonance experiments for AF1*b, AF1*c, AF1* and RAP74NMR (see table 3.1). [^{1}H,^{15}N]-HSQC experiments were acquired in between the 3D experiments for control of sample quality. When the complete set of assignment experiments was run, no degradation of the sample was observed in any of the cases.

The backbone H^N, N^H, CO, C_α and C_β nuclei of AF1*b (AR 265–340), AF1*c (AR 330–448) and AF1* (AR 142–448) were assigned at 200 μM (AF1*b), at 290 μM (AF1*c) or at 340 μM (AF1*), in 20 mM sodium phosphate buffer containing 1 mM TCEP, at pH 7.4 and 278 K. The backbone H^N, N^H, CO and C_α nuclei of RAP74NMR (RAP74 450–517) were assigned in 20 mM sodium phosphate, at pH 6.5 and 298 K. The acquisition parameters used for the assignment experiments are shown in tables 3.2, 3.3, 3.4 and 3.5.

Table 3.1: Pulse sequences used for the 3D triple resonance experiments to assign the backbone of AF1*b, AF1*c, AF1* and RAP74NMR.

experiment	pulse sequence	references
HNCA	hncagpwg3d	[368, 369]
HN(CO)CA	hncocagpwg3d	[368, 369]
CBCANH	cbcanhgpwg3d	[370]
CBCA(CO)NH	cbcaconhgpwg3d	[371, 372]
HNCO	hncogpwg3d	[368, 369]
HN(CA)CO	hncacogpwg3d	[369, 373]

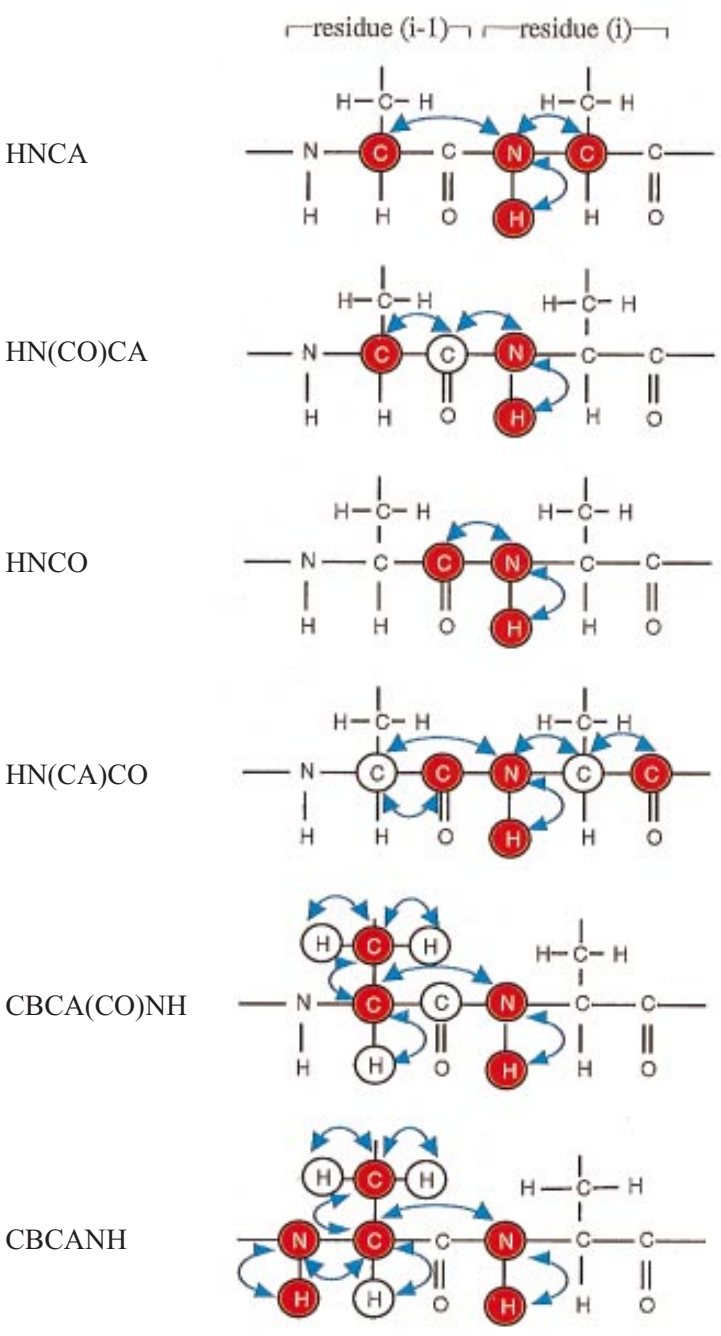


Figure 3.5: The magnetization transfer pathways for the commonly used triple resonance experiments for the assignment of protein backbone residues. The atoms involved in the transfer pathways are circled and the resonances that are observed in the spectra are colored in red. From [367].

Table 3.2: Parameters used for the acquisition of 3D triple resonance experiments to assign the backbone of AF1*b. TD: time domain, SW: sweep width (in ppm), O1, O3 and O2 are the carrier positions for the ^{1}H, ^{15}N and ^{13}C dimension, respectively (in ppm), NS: number of scans, RG: receiver gain.

experiment	TD			SW			O1	O3	O2	NS	RG
	^{1}H	^{15}N	^{13}C	^{1}H	^{15}N	^{13}C	^{1}H	^{15}N	^{13}C		
HNCA	1600	64	64	10	24	24	4.7	118	52	16	512
HN(CO)CA	1600	64	64	10	24	24	4.7	118	52	16	512
CBCANH	1600	64	128	10	24	60	4.7	118	40	16	256
CBCA(CO)NH	1600	64	128	10	24	60	4.7	118	40	8	256
HNCO	1600	64	64	10	24	6	4.7	118	173	8	512
HN(CA)CO	1600	64	64	10	24	6	4.7	118	173	16	512

Table 3.3: Parameters used for the acquisition of 3D triple resonance experiments to assign the backbone of AF1*c. TD: time domain, SW: sweep width (in ppm), O1, O3 and O2 are the carrier positions for the ^{1}H, ^{15}N and ^{13}C dimension, respectively (in ppm), NS: number of scans, RG: receiver gain.

experiment	TD			SW			O1	O3	O2	NS	RG
	^{1}H	^{15}N	^{13}C	^{1}H	^{15}N	^{13}C	^{1}H	^{15}N	^{13}C		
HNCA	1600	64	64	10	24	24	4.7	118	52	16	512
HN(CO)CA	1600	64	64	10	24	24	4.7	118	52	16	512
CBCANH	1600	64	128	10	24	70	4.7	118	40	16	512
CBCA(CO)NH	1600	64	128	10	24	70	4.7	118	40	8	512
HNCO	1600	64	64	10	24	6	4.7	118	173	8	512
HN(CA)CO	1600	64	64	10	24	6	4.7	118	173	32	512

Table 3.4: Parameters used for the acquisition of 3D triple resonance experiments to assign the backbone of AF1*. TD: time domain, SW: sweep width (in ppm), O1, O3 and O2 are the carrier positions for the ^{1}H, ^{15}N and ^{13}C dimension, respectively (in ppm), NS: number of scans, RG: receiver gain.

experiment	TD			SW			O1	O3	O2	NS	RG
	^{1}H	^{15}N	^{13}C	^{1}H	^{15}N	^{13}C	^{1}H	^{15}N	^{13}C		
HNCA	1600	64	64	10	24	24	4.7	118	52	8	512
HN(CO)CA	1600	64	64	10	24	24	4.7	118	52	8	512
CBCANH	1600	64	128	10	24	70	4.7	118	40	8	512
CBCA(CO)NH	1600	64	128	10	24	70	4.7	118	40	8	512
HNCO	1600	64	64	10	24	6	4.7	118	173.5	8	512
HN(CA)CO	1600	64	64	10	24	6	4.7	118	173.5	8	512

Table 3.5: Parameters used for the acquisition of 3D triple resonance experiments to assign the backbone of RAP74NMR. TD: time domain, SW: sweep width (in ppm), O1, O3 and O2 are the carrier positions for the ^{1}H, ^{15}N and ^{13}C dimension, respectively (in ppm), NS: number of scans, RG: receiver gain.

experiment	TD			SW			O1	O3	O2	NS	RG
	^{1}H	^{15}N	^{13}C	^{1}H	^{15}N	^{13}C	^{1}H	^{15}N	^{13}C		
HNCA	2048	64	64	10	24	28	4.7	118	56	8	512
HN(CO)CA	2048	64	64	10	24	28	4.7	118	56	8	512
CBCANH	2048	64	64	10	24	56	4.7	118	42	8	912
CBCA(CO)NH	2048	64	64	10	24	56	4.7	118	42	8	512
HNCO	2048	64	64	10	24	11	4.7	118	173.5	8	512
HN(CA)CO	1600	64	64	10	24	11	4.7	118	173.5	8	512

3.6.3.4 Transverse ^{15}N relaxation

Transverse ^{15}N relaxation rates were measured for AF1* at 140 μM on the Bruker 600US spectrometer using the Bruker pulse sequence hsqct2etf3gpsi. The relaxation delays (d20) were set to 20 ms, 35 ms, 50 ms, 90 ms, 120 ms, 160 ms, 200 ms and 250 ms in the various experiments. The experiments at relaxation delay 35 ms and 200 ms were acquired twice to determine the experimental error in the analysis. 2048 and 256 complex incremental data points were acquired in F2 (^{1}H) and F1 (^{15}N) dimensions, the number of scans was 16 and the SW was set to 10 ppm for ^{1}H and to 24 ppm for ^{15}N.

Transverse relaxation rates for each residue were determined by fitting an exponential function to the intensity decay of the signals at increasing relaxation delays (see equation 3.1). In this equation y is the peak intensity of the resonance of a given residue at corresponding relaxation delay x (expressed in s), and A and B are adjusted by the fitting procedure. The obtained fitted value B, for each amino acid, corresponds to its transverse relaxation rate (R_2) (and is given in s^{-1}).

$$y = A \cdot e^{(-B \cdot x)} \tag{3.1}$$

3.6.3.5 CON

The CON experiment is a heteronuclear experiment that results in a spectrum with cross peaks for the backbone amide nitrogen ^{15}N and the carbonyl ^{13}CO of the preceding residue [374]. Because the detection of magnetization in this experiment happens on carbon (^{13}C) the sensitivity is low, and thus longer acquisition times are required. This experiment is especially useful as cross peaks are obtained for proline residues, which are not detectable in [^{1}H,^{15}N]-HSQC spectra as explained before. It therefore allows

the determination of the amide nitrogen chemical shift of proline residues.

A CON experiment was acquired for AF1* in the Bruker 800 spectrometer at 278 K using pulse sequence c_con_iasq. The sample contained 390 μM uniformly double labeled (^{13}C and ^{15}N) AF1* in 20 mM sodium phosphate buffer with 1 mM TCEP and at pH 7.4. 256 complex incremental data points were acquired for the ^{15}N dimension and 1024 for the ^{13}C dimension, with 128 scans per increment. The SW was set to 40 ppm for ^{15}N and 14 ppm for ^{13}C, with carrier positions at 125 ppm and 173 ppm, respectively.

3.6.3.6 Saturation transfer difference (STD) NMR

Saturation transfer difference (STD) NMR is one of the most powerful NMR techniques for detection and characterization of transient receptor-ligand interactions in solution [375]. The technique is based on the transfer of saturation from the selectively irradiated receptor (protein) to the bound ligand (small molecule) that is subsequently detected when the ligand is free in solution again [376, 377]. STD takes the difference of two ^{1}H NMR experiments, the off-resonance experiment and the on-resonance experiment (see Fig. 3.6). In the off-resonance experiment, a train of frequency selective radio frequency (rf) pulses is applied off-resonance, i.e. at a frequency far away from both the protein and small molecule resonances (e.g. at +60 ppm) so neither of them are perturbed. In the on-resonance experiment, the protein is selectively saturated for a specific period (saturation time) via an identical train of frequency selective rf pulses, generally applied to the higher field region of the spectrum (0.0 to -1.0 ppm for proteins). Importantly, irradiation must saturate the protein but cannot saturate the small molecule directly. Therefore, the frequency at which is irradiated for a particular system is chosen corresponding to a chemical shift at which a protein signal is observed, and no ligand resonance is present. For globular proteins, often on-resonance irradiation at frequencies corresponding to chemical shift values below 0 ppm is suitable, as typically upfield shifted methyl groups are present in this region of the spectrum whereas resonances from small molecules are rarely detected here. The saturation of these protein resonances propagates from these protons to the entire protein via a spin diffusion process, which is quite efficient in globular proteins due to the vast network of intramolecular ^{1}H-^{1}H cross-relaxation pathways available in a large molecular weight protein. In IDPs such a network is absent. When a small molecule binds to the protein, saturation is transferred from the protein to the ligand via intermolecular ^{1}H-^{1}H cross-relaxation at the protein-ligand interface. Finally, the ligand dissociates from the protein and is free to diffuse in solution again, where the saturated state persists due to the small longitudinal relaxation rates of small molecules. Consequently, the resonances of small molecules that bind to the protein will be broadened in the ac-

quired on-resonance spectrum compared to in the off-resonance spectrum where the corresponding protons were not saturated. The STD spectrum is a difference spectrum obtained by subtracting the on-resonance spectrum from the off-resonance spectrum. If a small molecule binds to the protein, this difference spectrum will contain resonances of this small molecule because its resonances have been attenuated by saturation transfer. For compounds that do not bind to the protein, the resonances in the on- and off-resonance spectra are identical and subtraction leads to canceling out of the signals. As a consequence, the STD spectrum contains only resonances of small molecules that bind the protein and resonances of the protein itself, because the protein resonances have also experienced saturation. However, protein resonances are generally not observed, either due to the low protein concentration or because they are eliminated by applying a transverse relaxation filtering before detection (spin-lock filter). Therefore, the STD spectrum contains only signals of protons from small molecules that bind to the protein under study.

STD experiments were acquired at 278 K on the Bruker 800 spectrometer for a sample containing 25 μM AF1* and 250 μM EPI-001 in 20 mM sodium phosphate buffer containing 1 mM TCEP and 0.5% dioxane-d_8 and at pH 7.4. The buffer was fully deuterated (100% D_2O) to reduce the water signal as much as possible. The pulse sequence stddiffesgp.3 with water suppression using excitation sculpting with gradients and with a spin-lock filter to suppress protein signals was selected [376, 379]. For selective saturation of the protein, a train of Gaussian pulses with a length of 50 ms each was used for a total saturation time of 2 s. To remove residual protein signals, the length of the spin-lock filter mixing time was set to 20 ms. The number of scans was set to 256. Selective irradiation of the protein signals was applied at -0.4 ppm (on-resonance) and irradiation was applied at +60 ppm for acquisition of the off-resonance spectrum. Interestingly, at -0.4 ppm, no protein signals were present in the ^{1}H spectrum of the protein, as expected for IDPs. Due to the absence of a packed hydrophobic core, IDPs do not have upfield shifted methyl groups [362, 380]. The first AF1* proton resonance at high field was present around +0.7 ppm and the first EPI-001 proton signal, corresponding to the two methyl groups, was present around 1.6 ppm. In spite of the fact that no proton signals were observed in the ^{1}H NMR spectrum of AF1* at -0.4 ppm, the STD spectrum acquired for on-resonance -0.4 ppm showed clear EPI-001 signals, indicative of interaction. It is possible that by irradiating at -0.4 ppm a small population of a more structured conformational state of AF1* is saturated and that saturation is transferred to EPI-001 because it interacts with this state of the protein. Saturation of a conformational state of AF1* with a relatively high structural content would also be compatible with efficient propagation of saturation in AF1*, leading to a clear STD effect.

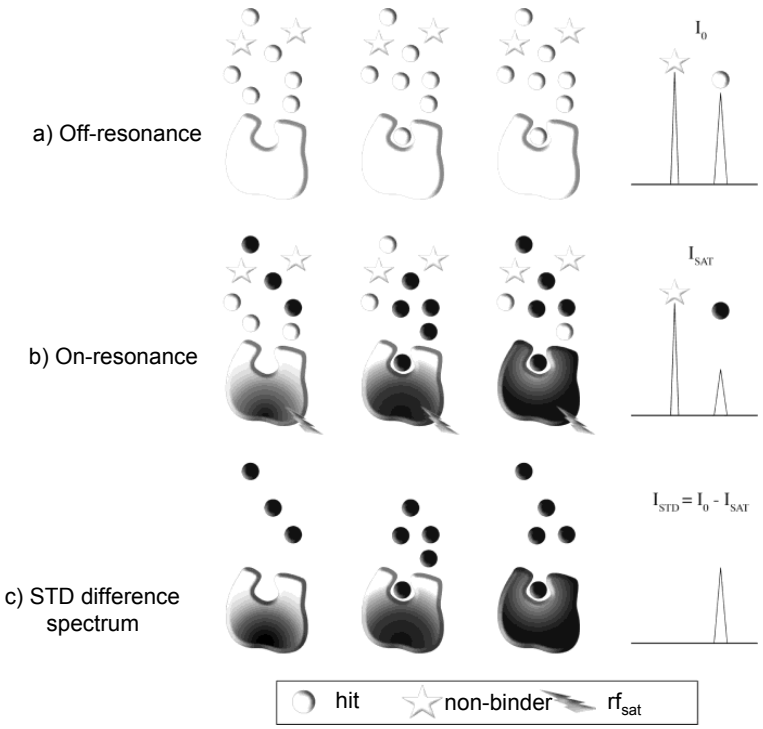


Figure 3.6: Schematic representation of the STD experiment. Circles and stars indicate compounds that bind and do not bind to the protein, respectively. STD involves two experiments: an off-resonance experiment and an on-resonance experiment. a) In the off-resonance experiment, radio frequency (rf) is applied off-resonance from both protein and compound protons. Detection yields spectra with intensity I_0. b) In the on-resonance experiment, the rf irradiation selectively saturates the receptor and is transferred to any binding compounds (indicated by dark shading). This results in a decreased signal intensity I_{SAT} for compounds that bind. c) The STD difference spectrum results from the subtraction of the on-resonance spectrum from the off-resonance spectrum. I_{STD}=I_0-I_{SAT}. This yields only resonances of binding compounds. Protein resonances are usually invisible due to either low concentration or relaxation filtering (spin-lock). From [378].

3.6.4 NMR experiments to study the interaction of AR and EPI-001

As will be explained in chapter 6, both the sample preparation and the NMR experiments to study the interaction between AR and the small molecule EPI-001 had to be carried out extremely carefully.

3.6.4.1 Sample preparation

In order to obtain identical conditions for all the samples, they were prepared as follows. A mother solution of 25 μM of ^{15}N-labeled AF1* protein was prepared in 20 mM sodium phosphate buffer at pH 7.4 containing 10% D_2O and 30 μM DSS-d_6, which was used as for reference purposes. Aliquots of 350 μL (volume used for measurements in D_2O-matched Shigemi tubes) were prepared, flash frozen and stored at -20 °C until they were used for NMR experiments. To measure an [^{1}H,^{15}N]-HSQC spectrum of 25 μM ^{15}N-labeled AF1* an aliquot was simply thawed and transferred to a Shigemi tube. For the blank measurements, 0.5% (v/v) dioxane-d_8 was added to one of these aliquots after thawing (1.75 μL dioxane-d_8 for 350 μL protein sample). To add EPI-001 in a specific concentration to 25 μM ^{15}N-AF1*, 1.75 μL of a stock solution of EPI-001 in 100% dioxane-d_8 was added to one of these aliquots. To reach a final concentration of 250 μM EPI-001, a stock solution of 50 mM EPI-001 in 100% dioxane-d_8 was prepared. Different stock solutions were used to reach varying final concentrations of EPI-001 (see table 3.6).

Table 3.6: Concentration stock solutions of EPI-001 in 100% dioxane-d_8 to prepare samples containing various concentrations of EPI-001 in 0.5% dioxane-d_8.

stock concentration	final concentration sample
10 mM	50 μM
30 mM	150 μM
50 mM	250 μM
70 mM	350 μM

To prepare the stock solution of EPI-001 in 100% dioxane-d_8, about 10 mg of EPI-001 (Sigma, 92427, bisphenol A (3-chloro-2-hydroxypropyl)(2,3-dihydroxypropyl)ether) was weighed using a balance with micromolar precision. The exact volume of dioxane-d_8 to add to the compound in order to obtain the desired stock concentration was calculated based on the exact weight of the EPI-001 obtained like this.

3.6.4.2 NMR acquisition

The [^{1}H,^{15}N]-HSQC spectra were acquired in the Bruker 800 spectrometer with the following parameters: 2048 complex points in the F_2 dimension (^{1}H) and 512 increments in the F_1 dimension (^{15}N). The SW was set to 8013 Hz (10 ppm) (^{1}H) and 1784 Hz (22 ppm) (^{15}N), and the number of scans to 16. The x-carrier frequency was determined by referencing to internal DSS-d_6. The DSS-d_6 frequency was obtained from a 1D experiment, recorded immediately before the 2D [^{1}H,^{15}N]-HSQC. Indirect referencing was used in the ^{15}N dimension by use of the conversion factors from Wishart *et al.* [363].

3.6.4.3 NMR processing

The spectra were processed using NMRDraw and NMRPipe [364]. Spectrometer frequencies and carrier frequencies in ppm were inserted with six decimals. (0.000001 ppm = 1 ppb, so ppb precision). Forward-backward linear prediction and zero filling to the nearest power of two were used in the indirect dimension of the HSQC (^{15}N). Zero filling to the nearest power of two was also applied to the direct dimension (^{1}H).

Table 3.7: Experimentally determined reproducibility of chemical shift measurements and the digital resolution of the NMR experiments after processing of the data with relevant acquisition and processing parameters.

	^{1}H	^{15}N
Number of points/increments	2048	512
Sweep width in Hz	8013	1784
Linear prediction (LP)	none	1024
Zero filling (ZF)	4096	2048
Digital resolution in ppb	4.88	43
Digital resolution after LP and ZF in ppb	2.44	10.7
Experimental standard deviation in ppb	±0.5	±2.4

3.6.4.4 NMR data analysis

The processed spectra were analyzed in CcpNmr Analysis [365]. Peaks that are not overlapped were picked using the automated peak picking routine in CcpNmr Analysis. Overlapped peaks were carefully picked manually.

3.7 Circular dichroism

3.7.1 Data acquisition

Far-UV CD spectra were acquired on a Jasco J-810 spectropolarimeter (CCiT, University of Barcelona) with a temperature control unit (Peltier) from 190 nm to 250 or 260 nm. Together with the ellipiticity at each wavelength (Θ_λ), the trace of the High Tension (HT) voltage, the voltage applied to the photomultiplier, was monitored. This trace should generally be less than 700 V to obtain reliable data, but this value depends on the particular instrument used [381]. For all the acquired CD data on the spectropolarimeter used, the data was considered not to be reliable and thus not included in the analysis when the HT signal was higher than 600 V, according to recommendations from CCiT personnel.

The total absorbance of the sample should be below 1.0 in order to avoid excessive noise [381]. Consequently, in order to measure at various concentrations of protein or peptide, the path length of the cuvette should be adjusted accordingly to meet the absorbance criterion [381]. Therefore, cuvettes with path length 1.0 mm (Hellma, Precision cells made of Quartz SUPRASIL, 100-QS 1 mm) and 0.1 mm (Hellma, 121-0.10-40, type: 121.000-QS) were used. The cuvette with path length 0.1 mm was cylindrical and thus not compatible with the Peltier temperature control unit. Most of the CD experiments were therefore acquired at room temperature (without temperature control), and for AF1*c also a control experiment was run at 5 °C.

The concentration, path length, temperature and range of wavelengths for which the various experiments were measured are indicated in table 3.8. All data were acquired with the following parameters: data pitch = 0.2 nm, scan rate = 50 nm/min, response = 4 s, band width = 2 nm, number of accumulations = 20. For each protein or peptide sample, a blank sample with identical composition but without the protein or peptide was acquired with the same experimental parameters.

A far-UV CD spectrum was measured for AF1*b in 20 mM sodium phosphate buffer with 100 mM NaCl, 1 mM TCEP, pH 7.2. For AF1*c and AF1*, far-UV CD spectra were acquired at various concentrations (see table 3.8) and all in 20 mM sodium phosphate buffer with 1 mM TCEP, pH 7.4. The far-UV CD spectra for peptides (WT, pS430, pS431 and pS432) were acquired with milliQ water as buffer system.

Table 3.8: Experimental conditions of the acquired CD spectra. rt stands for room temperature (temperature control unit not used).

protein/peptide	concentration (μM)	path length (mm)	temperature (℃)	wavelength (nm)
AF1*b	7	1	20	190–250
AF1*c	10	1	5	190–250
AF1*c	10	1	rt	190–260
AF1*c	25	1	rt	190–260
AF1*c	100	0.1	rt	190–260
AF1*c	250	0.1	rt	190–260
AF1*c	400	0.1	rt	190–260
AF1*	5	1	rt	190–260
AF1*	10	1	rt	190–260
AF1*	25	0.1	rt	190–260
AF1*	50	0.1	rt	190–260
AF1*	100	0.1	rt	190–260
AF1*	250	0.1	rt	190–260
WT peptide	45	1	rt	190–250
WT peptide	445	0.1	rt	190–250
pS430 peptide	428	0.1	rt	190–250
pS431 peptide	413	0.1	rt	190–250
pS432 peptide	428	0.1	rt	190–250

3.7.2 Data analysis

The measured data is the ellipticity at each wavelength (Θ_λ) and is expressed in millidegrees (mdeg). This data was analyzed by subtracting the acquired ellipticity from the corresponding blank sample for each wavelength, and further normalization according to equation 3.2 for the number of peptide bonds (i.e. the number of residues minus one), the path length of the cuvette used and the concentration of the sample. The mean residue ellipticity for each wavelength (MRE_λ) obtained like this is expressed in deg.cm^2.dmol^{-1}.

$$MRE_\lambda(deg.cm^2.dmol^{-1}) = \frac{\Theta_{\lambda,sample}(mdeg) - \Theta_{\lambda,blank}(mdeg)}{(N-1)\cdot d(m)\cdot c(mM)} \tag{3.2}$$

In equation 3.2, MRE_λ indicates the mean residue ellipticity for each wavelength, $\Theta_\lambda(mdeg)$ is the observed ellipticity for each wavelength λ, for the sample or for the blank, and expressed in millidegrees, N is the number of residues in the protein or peptide, $d(m)$ is the path length expressed in meter and $c(mM)$ is the concentration of protein or peptide in millimolar.

3.8 Mass spectrometry

3.8.1 Liquid chromatography coupled to mass spectrometry

The mass of expressed fusion-tagged protein constructs and of the protein constructs after removal of the fusion tag was verified by liquid chromatography coupled to mass spectrometry (LC-MS), by the mass spectrometry core facility of IRB Barcelona.

3.8.1.1 LC conditions

Chromatographer: quaternary pump, Finnigan, Mod. Surveyor MS (Thermo Electron Corporation)
Autosampler: Finnigan, Mod. Micro AS (Thermo Electron Corporation)
Column: BioSuite pPhenyl 1000 RPC, 10 μm, 2.0x75 mm (Waters)
Eluents:
A. H_2O 0.1% formic acid
B. acetonitrile 0.1% formic acid
Gradient: from 5% to 80% of B in 60 min.
Flow rate: 100 μL/min.
The column outlet was directly connected to an Advion TriVersa NanoMate (Advion) fitted on an LTQ-FT Ultra mass spectrometer (Thermo).

3.8.1.2 MS conditions

LTQ-FT Ultra (Thermo Scientific)
NanoESI
Positive ionization
Capillary Temperature: 200 °C
Tube Lens: 100 V
Ion Spray Voltage: 1.7 kV
m/z 400-2000 a.m.u

3.8.1.3 Data processing

LTQ-FT Ultra is powered by Xcalibur software vs 2.05R2. Ion deconvolution to zero charged monoisotopic masses was performed with Xtract algorithm in Xcalibur software.

3.8.2 Trypsin digestion and following mass spectrometry analysis

To detect covalent binding between AF1* and EPI-001 mixtures of the protein and small compound were incubated for different time points and further analyzed by trypsin digestion followed by mass spectrometry analysis. The sample preparation was performed by Dr. Christopher Phang (IRB Barcelona) and the trypsin digestion and mass spectrometry analysis were performed by Dr. Gonçalo Costa (MS core facility, IRB Barcelona).

3.8.2.1 Sample preparation

Samples of AF1* and EPI-001 (protein/small compound ratio 1/1.25 or 1/10) were prepared in 20 mM sodium phosphate buffer, pH 7.4 containing 1 mM TCEP and 0.5% (v/v) dioxane. Samples were oxygen degassed by three freeze-pump-thaw cycles and the atmosphere was replaced with argon (see section 3.4). Samples were then incubated at 310 K for specified time points (1 h, 2.5 h, 4 h, 7 h, 10 h and 24 h). After the specified time, the reaction was stopped by flash freezing with liquid nitrogen and samples were stored at -20 ℃ until further manipulation for trypsin digestion.

3.8.2.2 Trypsin digestion

Samples were diluted 1/2 with 50 mM ammonium bicarbonate. They were first reduced with DTT (2 mM) for 1 hour and then reacted with 2-iodoacetamide (5 mM) for 30 minutes in the dark. The reaction with 2-iodoacetamide was quenched with DTT (2 mM) and proteins were then digested with trypsin (2% w/v) overnight at 310 K. Trypsin activity was quenched by addition of formic acid to a final concentration of 1%.

3.8.2.3 Mass spectrometry analysis

Samples were loaded to a 180 μm × 2 cm C18 Symmetry trap column (Waters) at a flow rate of 15 μL/min using a nanoAcquity Ultra Performance LCTM chromatographic system (Waters Corp., Milford, MA). Peptides were separated using a C18 analytical column (BEH130™ C18 75 μm × 25 cm, 1.7 μm, Waters Corp.) with a 90 min run, comprising three consecutive steps with linear gradients from 1 to 35% B in 60 min, from 35 to 50% B in 5 min, and from 50% to 85% B in 3 min, followed by isocratic elution at 85% B in 10 min and stabilization to initial conditions (A = 0.1% formic acid (FA) in water, B = 0.1% FA in acetonitrile). The column outlet was directly

connected to an Advion TriVersa NanoMate (Advion) fitted on an LTQ-FT Ultra mass spectrometer (Thermo). The mass spectrometer was operated in a data-dependent acquisition (DDA) mode. Survey MS scans were acquired in the FT with the resolution (defined at 400 m/z) set to 100,000. Up to six of the most intense ions per scan were fragmented and detected in the linear ion trap. The spectrometer was set-up in positive polarity mode.

A database search was performed with Bioworks v3.1.1 SP1 (Thermo Scientific) using Sequest search engine and SwissProt database, which included SwissProt HUMAN (release 2013˙06), AF1* protein and the common repository of adventitious proteins (http://www.thegpm.org/crap/index.html). Peptide mass tolerance was 10 ppm and the MS/MS tolerance was 0.8 Da. Peptides with a q-value lower than 0.1 and a FDR<1% were considered as positive identifications with a high confidence level.

3.8.3 MALDI-TOF MS

When necessary, the molecular weight of peptides was determined by matrix-assisted laser desorption/ionization-time of flight mass spectrometry (MALDI-TOF MS) using an Applied Biosystem 4700 instrument. 1 μL of peptide solution was mixed with 1 μL of α-cyano-4-hydroxycinnamic acid (ACH) matrix, and seeded on the MALDI plate where it was air-dried before the measurement. The ACH matrix was prepared as a 10 mg/mL ACH solution in acetonitrile/water 1/1 containing 0.1% trifluoroacetic acid (TFA).

MALDI-TOF MS confirmed that no dimers were present for the cysteine-containing peptides prior to further experiments. Likewise, MALDI-TOF MS was used to confirm that the phosphorylated peptides had not undergone β-elimination. Finally, labeling of the C-Cys peptide with the nitroxide MTSL label was verified by MALDI-TOF MS.

3.9 High-performance liquid chromatography

High-performance liquid chromatography (HPLC) experiments were performed on an HPLC 1200 Series (Agilent Technologies). For size exclusion chromatography by HPLC a Yarra 3u SEC-3000 C4 column (Phenomenex, product number 00H-4513-K0) was used. The column was equilibrated in 20 mM sodium phosphate buffer with 1 mM TCEP, pH 7.4. Samples were injected manually (10 μL for samples at 5 μM, 10 μM and 25 μM and 5 μL for samples at 50 μM, 100 μM, 250 μM and 400 μM) and buffer was passed for 15 minutes at 0.2 mL/min. Blank runs for which 10 μL of buffer was injected

were run in between protein samples to assure no protein was retained on the column. Detection was done at wavelength 215 nm.

3.10 Analytical size exclusion chromatography by FPLC

Analytical size exclusion chromatography was performed on a calibrated Superdex 200 10/300 GL column (GE Healthcare). The column was equilibrated in 20 mM sodium phosphate buffer with 1 mM TCEP, pH 7.4. For AF1*c, 40 μL at 593 μM was injected, and for AF1*, 40 μL at 1 mM was injected. Buffer was passed at 0.5 mL/min and fractions of 500 μL were collected. The absorption was detected at 280 nm and at 260 nm.

3.11 Microscale thermophoresis

Microscale thermophoresis (MST) is a recently developed technology for the measurement of biomolecular interactions under close-to-native conditions [382, 383]. It can be used to study the interaction between many types of partners (protein-protein, protein-nucleic acid, protein-peptide, protein-small molecule, etc.) and it allows the determination of binding affinities. Compared to other techniques for characterizing binding, MST offers additional advantages such as the use of very small sample volumes, no need to immobilize one of the binding partners on a surface, and it is compatible with almost any buffer system.

MST is based on a physical principle, thermophoresis, which is the motion of molecules along temperature gradients [383]. Thermophoresis is very sensitive to changes in size, charge and solvation shell of a molecule [384, 385]. Since at least one of these parameters is changed by a binding event, thermophoresis can be used to study biomolecular interactions [383].

In MST, an infrared-laser (IR-laser) is used to generate precise microscopic temperature gradients within thin glass capillaries that are filled with a sample in a buffer or bioliquid of choice. The readout is based on fluorescence. Therefore, either fluorescently labeled proteins/peptides/nucleic acids can be used or the intrinsic tryptophan fluorescence of the biomolecule under study can be exploited. The fluorescence of these molecules is used to monitor their motion along the created temperature gradients. The experimental set-up of MST and a typically obtained MST signal are shown in figure 3.7.

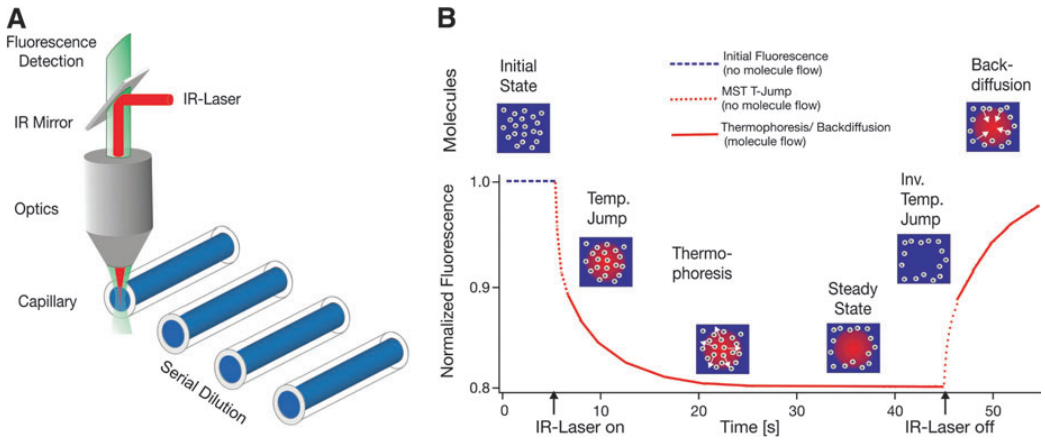


Figure 3.7: Experimental set-up of MST and the MST signal. A) MST is measured in a capillary. The fluorescence in the capillary is excited and measured through the same optical element. An IR-laser is used to locally heat the sample volume that is observed by fluorescence. Temperature jump (T-jump) and thermophoresis are directly observed as a change in fluorescence at different time scales. B) A typical MST signal for a given capillary. Initially, the molecules are homogeneously distributed and a constant "initial fluorescence" is measured. As soon as the IR-laser is turned on, a fast T-jump is observed, followed by thermophoretic molecule motion. The fluorescence decrease is measured for about 30 seconds. When the IR-laser is turned off, an inverse T-jump is observed, followed by the "back diffusion" of molecules. From [383].

The binding affinity of the double phosphorylated pS430/pS432 AR peptide (see table 5.1 on p. 184) and RAP74NMR was measured by label-free MST, on a Monolith NT.LabelFree device (NanoTemper Technologies GmbH). The AR peptide contains three fluorescent amino acids (W433, F437 and Y445), whereas RAP74NMR does not contain any tryptophan residues, only contains one tyrosine residue (Y462) and two phenylalanine residues (F477 and F513). Since the contribution of phenylalanine to fluorescence is negligible compared to that of tryptophan and tyrosine, the intrinsic fluorescence of the AR peptide could be used to perform label-free MST experiments, avoiding the need to attach a fluorescent label to one of the binding partners. The experiments were performed at room temperature and in a 20 mM sodium phosphate buffer at pH 7.4.

Sixteen samples were prepared as mixtures of RAP74NMR (non-fluorescent) at varying concentration and AR pS430/pS432 peptide (fluorescent) at constant concentration, as follows. A stock solution of 3.13 mM RAP74NMR was prepared in 20 mM sodium phosphate buffer at pH 7.4, as well as a stock solution of 10 μM pS430/pS432 peptide in 20 mM sodium phosphate buffer at pH 7.4 and containing 0.1% Tween. First, a dilution series of RAP74NMR was prepared. For this purpose 20 μL of the RAP74NMR stock solution was pipetted into the first of sixteen PCR eppendorfs and 10 μL of 20 mM sodium phosphate buffer at pH 7.4 was pipetted into the remaining fifteen eppendorfs

that were labeled 2 to 16. Subsequently, 10 μL of concentrated RAP74NMR solution was transferred from eppendorf 1 to eppendorf 2, and the solution was mixed by pipetting up and down. Next, 10 μL was transferred from eppendorf 2 to eppendorf 3, and the solution was mixed by pipetting up and down. This procedure was repeated until 10 μL of eppendorf 15 was transferred to eppendorf 16. Finally, 10 μL of the 20 μL in eppendorf 16 was discarded. To add the fluorescent binding partner to each of the samples at constant concentration, 10 μL of the stock solution containing 10 μM pS430/pS432 peptide and 0.1% Tween in 20 mM sodium phosphate buffer at pH 7.4 was added to each of the sixteen eppendorfs containing distinct concentrations of RAP74NMR.

As a result, sixteen samples were obtained in which the concentration of pS430/pS432 peptide was 5 μM and the concentration of RAP74NMR varied between 1565 μM and 48 nM, in 20 mM sodium phosphate buffer at pH 7.4 containing 0.05% Tween. The addition of 0.05% Tween to the samples was necessary because one of the binding partners was sticking to the capillaries and this was found not to be the case in the presence of 0.05% Tween. We note that the addition of 0.05% Tween is a common procedure and is recommended by the manufacturer in case one of the binding partners sticks to the capillaries. These sixteen samples were loaded in standard glass capillaries (NanoTemper Technologies GmbH) and placed into the tray of the Monolith NT.LabelFree device from high concentration to low concentration. The data were acquired at laser power 40%.

4

Conformational properties of the AR NTD

The AR N-terminal transactivation domain is indispensable for AR function as it regulates transcription of its target genes by interacting with the transcription machinery and transcriptional co-regulators. It is also one of the main therapeutic targets for PCa, since the proliferation of PCa cells stays dependent on active AR. In late stage PCa the AR can be activated even in the absence of hormone. This aberrant transactivation is believed to take place through activation function AF1 within the NTD. It is therefore crucial to understand the role of the NTD in AR activation, both in healthy cells and in cancer cells.

In spite of its biological importance, the structural properties of the NTD have remained little understood. This is due to the ID nature of the domain, which has prevented its characterization using X-ray crystallography, and its size (559 amino acids), which has delayed its study using solution nuclear magnetic resonance (NMR) spectroscopy.

To gain better insight into the molecular mechanisms by which this ID domain regulates the function of the AR and causes disease, we performed a thorough NMR characterization of the conformational properties of this domain by measuring NMR observables that report on its secondary structure and dynamics. This was complemented with information from other biophysical techniques.

4.1 Predicted structural properties of the NTD as part of the AR

A wide range of bioinformatics tools is available to assess the transient structural properties of IDPs based on their sequence. To complement the limited low-resolution data available in literature of the structural properties of the NTD, we used some of these predictors.

We used several protein disorder predictors, PONDR-Fit[1], DISpro[2] and DISOPRED2[3], to assess the disorder probability. PONDR-Fit (PONDR: Predictor Of Naturally Disordered Regions) is a meta-predictor of ID amino acids and gives reliable results for fully disordered and fully structured proteins, as well as for proteins in which both structure and disorder are present [386]. DISpro gives predictions of protein disordered regions by mining protein structure data [387], whereas DISOPRED2 employs support vector machines to predict native disorder in proteins [388]. The results from PONDR-Fit, which are representative for the tested disorder predictors, are given in figure 4.1a.

These predictors reveal that, in contrast to the DBD and the LBD, which are folded domains, the NTD is mainly predicted to be disordered. Nevertheless, it contains some stretches of sequence with higher order probability. One of these is the FQNLF motif between residues 23 and 27 of the AR, which is known to fold into a helix when it interacts with the AF2 region of the LBD [389].

To assess the secondary structure propensity, we used PSIPRED v3.3[4] [390, 391], and SSpro v4.5[5] [392], both secondary structure prediction methods. The NTD is predicted by both to have a low content of secondary structure, with almost no β-strand but with some α-helical propensity in specific regions. The results from SSpro, which are representative for both predictors, are given in figure 4.1b. Interestingly, some regions of sequence in the NTD that are predicted by PONDR to be more ordered coincide with stretches of sequence predicted to be helical by SSpro, e.g. the ^{23}FQNLF^{27} motif, residues 240–260 and the residues around 350. Agadir, an algorithm to predict the helical content of peptides [393], was used to further explore this predicted residual helicity. Also according to Agadir, the NTD is predicted to have a low helical content with the highest helical propensity (ca. 30%) around A400 (see Fig. 4.1c). Stretches of sequence in the NTD predicted by Agadir to have helical propensity also coincide with

[1] http://pondr.com/index.html
[2] http://scratch.proteomics.ics.uci.edu
[3] http://bioinf.cs.ucl.ac.uk/psipred/?disopred=1
[4] http://bioinf.cs.ucl.ac.uk/psipred/
[5] http://scratch.proteomics.ics.uci.edu

minima in the PONDR prediction, i.e. regions with less disorder probability.

Regions with a relatively high degree of residual structure in IDPs can indicate molecular recognition features (MoRFs), also known as molecular recognition elements (MoREs). These MoRFs are short loosely structured regions located within longer IDRs that are predicted to bind to protein partners via disorder-to-order transitions. They are therefore considered as disordered regions with molecular recognition and binding functions [300]. We used MoRFpred[6] to locate putative MoRFs in AR [394]. Interestingly, some MoRFs are predicted at minima in the disorder plots, such as the ^{23}FQNLF^{27} region and the ^{433}WHTLF^{437} region (see Fig. 4.1d). The ^{433}WHTLF^{437} motif of AR is located in the Tau-5 region and known to be important for androgen-independent transactivation [232]. This motif is predicted by PONDR to have some order propensity (see Fig. 4.1a), but neither SSpro nor Agadir predict helical propensity (see Figs. 4.1b and c). Nevertheless, it is predicted to be a MoRF (see Fig. 4.1d).

The hydrophobicity of the AR sequence was probed by the "average area buried upon folding" (AABUF)[7], defined for each amino acid as the difference in accessible surface area between a solvent-exposed standard state and a mean solvent accessibility of the amino acid in a database of folded proteins [395]. This parameter, which incorporates both residue size and hydrophobicity, is characteristic of individual amino acids and is usually averaged over a window of several residues to account for groupings of amino acids of particular types [396]. AABUF recognizes the potential for hydrophobic interactions inherent in long amino acid side chains, even if they also contain hydrophilic or charged groups [397]. For example, a lysine residue, which is identified in classic hydrophobicity scales as "charged", in fact contains an extensive aliphatic chain, which is certainly capable, in the context of an unfolded protein, of making hydrophobic interactions [397, 398]. Regions of high AABUF have been shown to initiate folding by forming foci for local hydrophobic collapse and by attracting and consolidating transient long-range interactions that are required for progression toward the folded state [397, 399]. Furthermore, for apomyoglobin, it was shown that regions of sequence with high AABUF values correspond with those parts of the polypeptide chain that fold first and where secondary structure becomes stabilized in kinetic folding intermediates [396].

The AABUF of AR, calculated with a nine-residue moving-average window, is shown in figure 4.2.

The AR is predicted by Paircoil2[8] [400] to have coiled-coil propensity between

[6] http://biomine-ws.ece.ualberta.ca/MoRFpred/index.html
[7] http://web.expasy.org/protscale/
[8] http://groups.csail.mit.edu/cb/paircoil2/

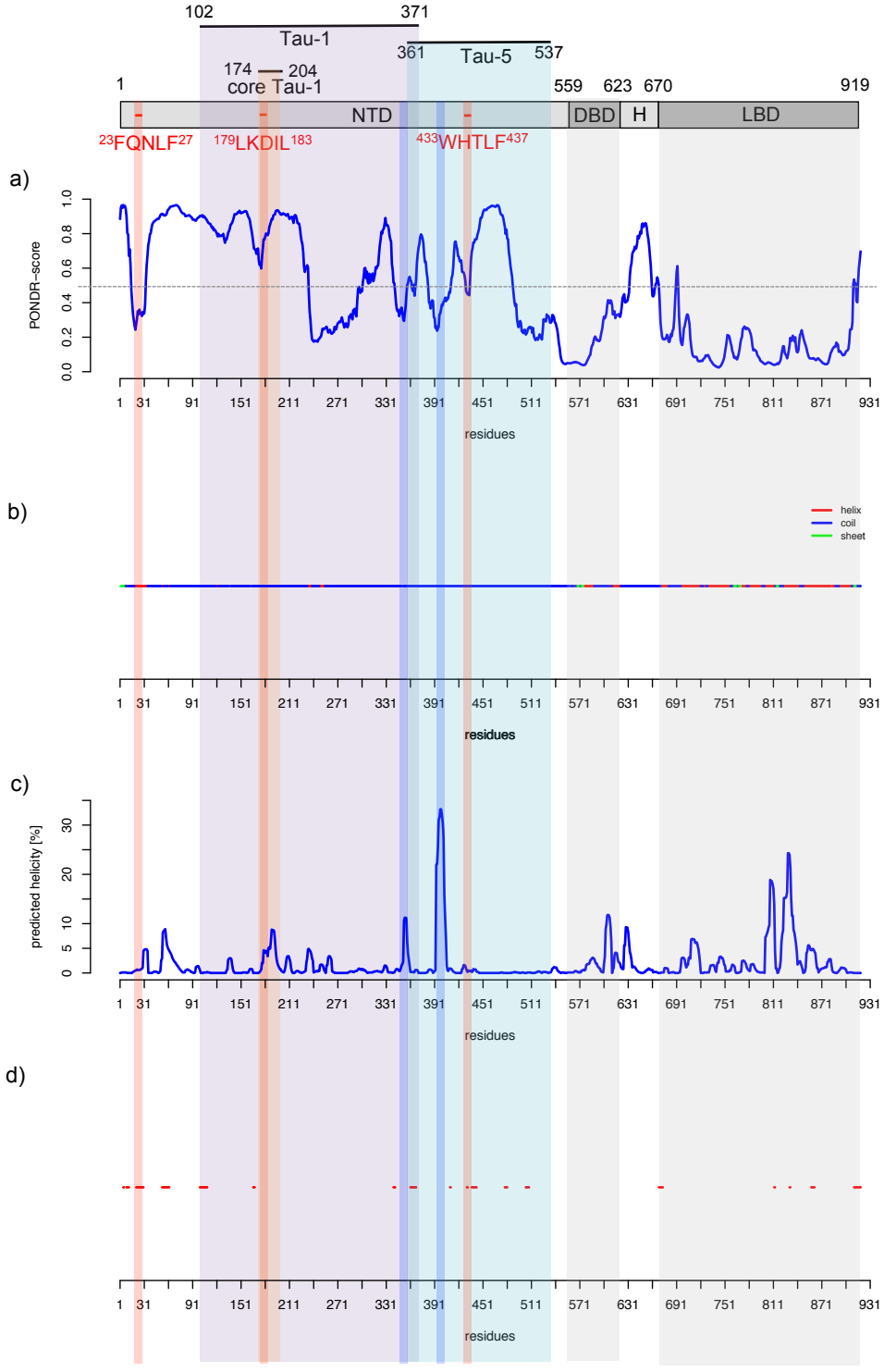


Figure 4.1: a) PONDR-Fit: values higher than 0.5 indicate disorder whereas values lower than 0.5 indicate order, b) SSpro (helix in red, coil in blue, sheet in green), c) Agadir: predicted helicity in percentage, d) prediction of MoRFs, molecular recognition features, within the AR (represented in red). The domain structure of the AR and functional regions of the protein are indicated.

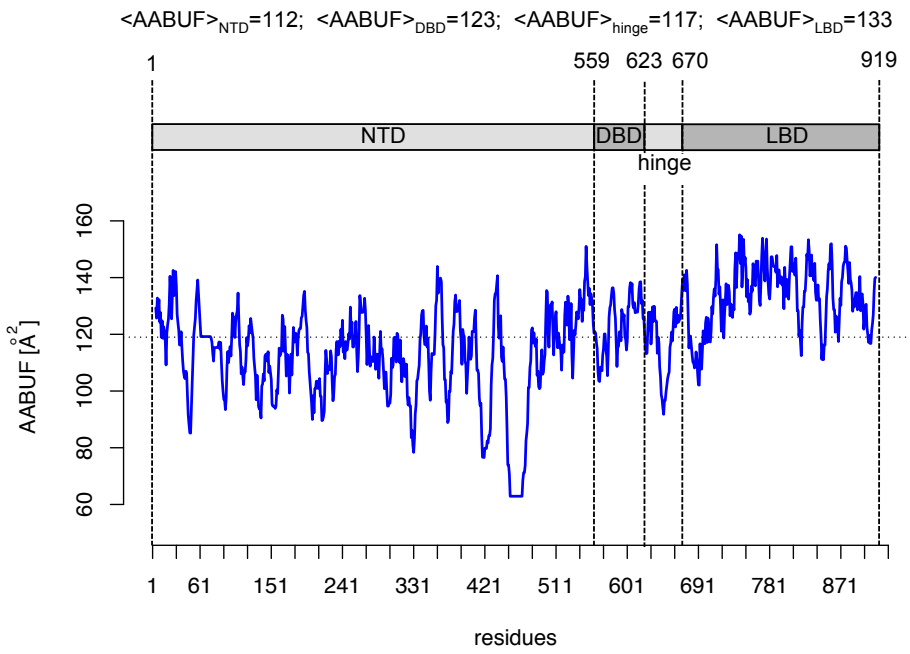


Figure 4.2: The hydrophobicity pattern of the AR as given by AABUF, using a nine-residue moving-average window (default window size). The dashed line indicates the mean value of AABUF for the entire chain (119 Å^2). The average value of AABUF is given for each of the AR domains (NTD, DBD, hinge and LBD).

residues 51 and 81 and between residues 171 and 215 (see Fig. 4.3). The former region corresponds to the polyQ stretch (58-78) with flanking residues, whereas the latter contains the core Tau-1, known to be important for AR function. Although the regions are identified by the predictor, the corresponding p-scores are rather high for all residues (>0.025), especially for the ones in the polyQ region, indicating these residues are not predicted to be in a coiled-coil using a default p-score cut-off value of 0.025. Only when a cut-off value of 0.080 is used, the residues in core Tau-1 would be recognized as having coiled-coil propensity.

4.2 Design constructs for NMR studies

The size of the NTD (559 residues) renders the assignment of the backbone chemical shifts of this domain very challenging using classical triple resonance 3D experiments even by using the highest field NMR spectrometers that are currently available (a 1 GHz spectrometer is installed in Les Houches, France). However, it has previously been observed that IDPs display certain spectral peculiarities that can play to our advantage in obtaining their full-length assignment. It is possible to assign the resonances of

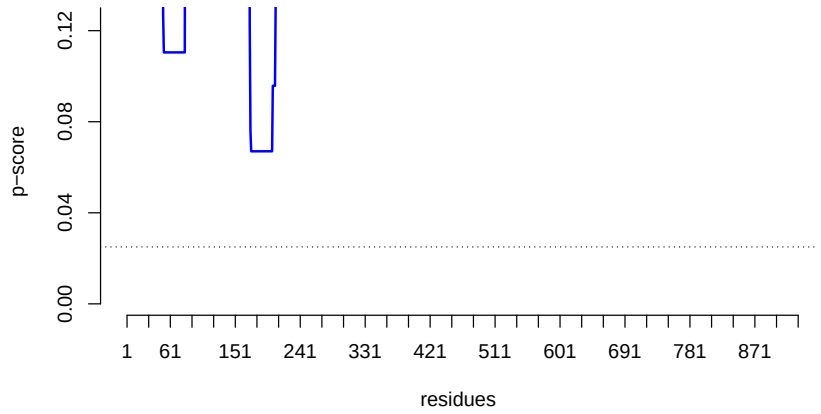


Figure 4.3: Coiled-coil prediction of the AR as given by Paircoil2. A default minimum search window of 28 residues was used. The dotted line indicates the default p-score cut-off value used by the software, 0.025. Residues below this score are predicted to be in a coiled-coil.

very large disordered proteins by assigning the resonances of several constructs that contain fractions of this protein –which are short enough to assign using conventional experiments– individually, and comparing these spectra with the spectrum of the full-length protein. This is possible because the chemical shifts of the vast majority of residues depend to a substantial extent on their immediate chemical environment. This *divide-and-conquer* strategy has been used by Zweckstetter and co-workers to characterize the structure and dynamics of Tau, a 441-residue IDP that forms intracellular neurofibrillary tangles in Alzheimer's disease [401].

Even using a *divide-and-conquer* approach, it would be challenging to assign the full-length NTD of the AR (559 residues). The regions within the NTD that have been described to be indispensable for function are located within activation function AF1, which spans residues 141 to 494 (see section 1.2.2.4). We therefore focussed our efforts on this part of the protein. We did not include the polyglycine stretch (24 glycines: residues 449–472) at the C-terminal end of AF1 in any of the NMR constructs, because such a repeat region would cause severe signal overlap where it appears in the spectra and is unlikely to contribute to any functional role of the AF1. We neither included the region of the NTD containing the polyglutamine stretch (21 glutamines: residues 58–78) as this region is known to be aggregation prone and involved in SBMA, a neurodegenerative disease caused by an expansion of the polyQ tract over a threshold of 38 residues [248]. Therefore, the construct we used to study the conformational properties of the N-terminal transactivation domain of the AR, AF1*, spans residues 142–448 (308 residues), largely including transactivation units Tau-1 and Tau-5, and excluding the polyglycine stretch (G449–G472) and the polyglutamine stretch (Q58–Q78). A

schematic representation of the construct design is given in figure 4.4. To assign AF1* we used three shorter protein constructs that have a 10-residue overlap. The design of these constructs was based on the low resolution structural information described for the AR NTD (see section 1.2.4), complemented with the predicted structural properties of the NTD (see section 4.1) and the limited functional annotation of this domain obtained mainly through mutational studies (see section 1.2.2.4).

The regions of the AR NTD with low predicted helicity by Agadir coincide with stretches of the NTD predicted by PONDR VL-XT[9] to be ordered (see Fig. 4.4). Even though the predicted helical content is low, AF1 has been proposed to exist as a molten globule with residual helical structure that can be increased in the presence of helix-promoting osmolytes, like TMAO [156]. To avoid possible disruption of helical elements, we chose the boundaries of the constructs in regions predicted to be disordered by PONDR VL-XT.

Given that the two regions with ordered probability at the N-terminus of AF1 are located within the part of the protein annotated as Tau-1 (residues 102–371) and the three predicted ordered regions towards the C-terminus of AF1 are located within the Tau-5 region (residues 361–537), we designed the protein constructs in such a way that the putative structural elements that form these functional units would be contained in the same construct (see Fig. 4.4).

The assignment of protein constructs is facilitated when they are not too long (i.e. max. 150 residues), not too repetitive in sequence and when they do not contain many proline residues as those interrupt the sequential connectivity. Taken this into account, we designed a first short protein construct (AR 142–275 = AF1*a, 135 residues) that includes two regions predicted to be slightly ordered in an α-helical conformation in Tau-1, and a second one (AR 330–448 = AF1*c, 120 residues) that contains the three C-terminal regions with high order probability in Tau-5. A third protein construct that spans the region of sequence in between those two (AR 265–340 = AF1*b, 77 residues) includes a region predicted to be ordered in a non-helical conformation (see Fig. 4.4).

The local environment of the terminal residues is expected to be different in the various protein constructs. For instance, for AF1*a the chemical shifts of the C-terminal residues will only be influenced by the residues N-terminal to them, whereas in AF1* the chemical shifts of the corresponding residues will be influenced by the neighboring residues both towards the N-terminus and the C-terminus. This will cause small de-

[9]When we designed the constructs for NMR studies the meta-predictor PONDR-Fit was not yet available. PONDR VL-XT is a merger of three predictors of the PONDR family, one trained on Variously characterized Long disordered regions and two trained on X-ray characterized Terminal disordered regions. It is useful for predicting MoRFs.

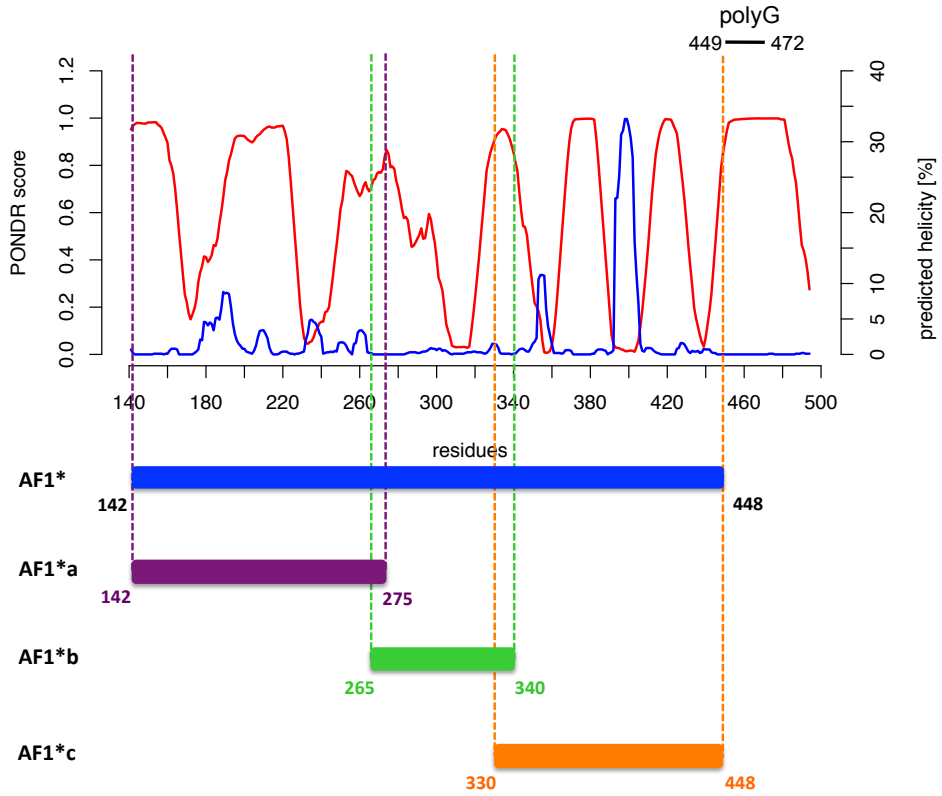
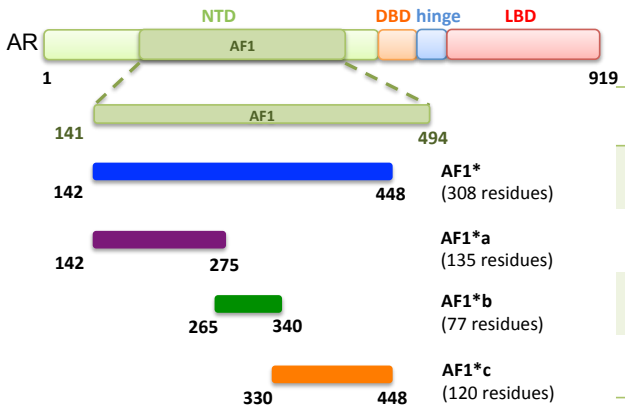


Figure 4.4: Schematic overview of the construct design for NMR studies. The PONDR VL-XT plot and Agadir plot for AF1 (141–494) are given in red and blue, respectively. PONDR scores higher than 0.5 indicate disorder and scores lower than 0.5 indicate order propensity. Predicted helicity is indicated in percentage. Residues 449–472 correspond to the polyglycine stretch. Bars under the predictors indicate the designed constructs: AF1*a (residues 142–275), AF1*b (residues 265–340), AF1*c (residues 330–448) and AF1* (residues 142–448).

viations between the peak positions of terminal residues in the NMR spectra of both protein constructs. The N-terminal residues of AF1*a and AF1*, however, do have the same local environment as both protein constructs start at the same amino acid. To be able to correct for these deviations between different protein constructs, we designed the constructs with a 10-residue overlap at the termini. Since the proposed persistence length of unfolded polypeptide chains is seven amino acids [402], we anticipated that ten residues overlap would be sufficient to transfer the assignments confidently between constructs.

4.3 Cloning, expression and purification of the AR constructs for NMR studies

The four designed constructs were successfully cloned into several Gateway vectors and, all but AF1*a, successfully expressed and purified for biophysical studies.



Construct	Cloning	Expression	Yield (per L culture)
AF1*	✔	✔	3.5 mg
AF1*a	✔	Cell-free expression	-
AF1*b	✔	✔	2 mg
AF1*c	✔	✔	5 mg

Figure 4.5: The AR with an indication of activation function AF1 and the constructs designed for NMR studies of this part of the protein.

The protein constructs were cloned using the recombinant Gateway® technology of LifeTechnologies, which is explained in more detail under methodology (chapter 3).

Two PCR reactions, PCR 1 and PCR 2, were used to introduce a TEV protease cleavage site (amino acids ENLYFQG) N-terminal to the gene construct corresponding to the designed AR constructs as well as *att*B recombination sites N-terminal to this cleavage site and C-terminal to the gene construct. The forward and reverse primers for PCR 1 and PCR 2 reactions were designed for each of the constructs (see chapter 3) and both PCR reactions were run with the conditions indicated in chapter 3. The template DNA used for PCR 1 reactions was the pET-AR-AF1 vector containing human AR residues 143–494, kindly provided by Prof. Iain McEwan (University of Aberdeen,

UK). The products obtained by the PCR 1 reactions were used as templates for the subsequent PCR 2 reactions.

The expected size of the PCR 1 and PCR 2 product is about 405 bp for the AF1*a construct (AR 142–275: 135 amino acids), 231 bp for the AF1*b construct (AR 265–340: 77 amino acids), 360 bp for the AF1*c construct (AR 330–448: 120 amino acids) and 924 bp for AF1* (AR 142–448: 308 amino acids). The obtained PCR 1 and PCR 2 products were run on an agarose gel (1% or 1.5%) and as can be seen in figure 4.6 the bands correspond to the expected size for each of the constructs.

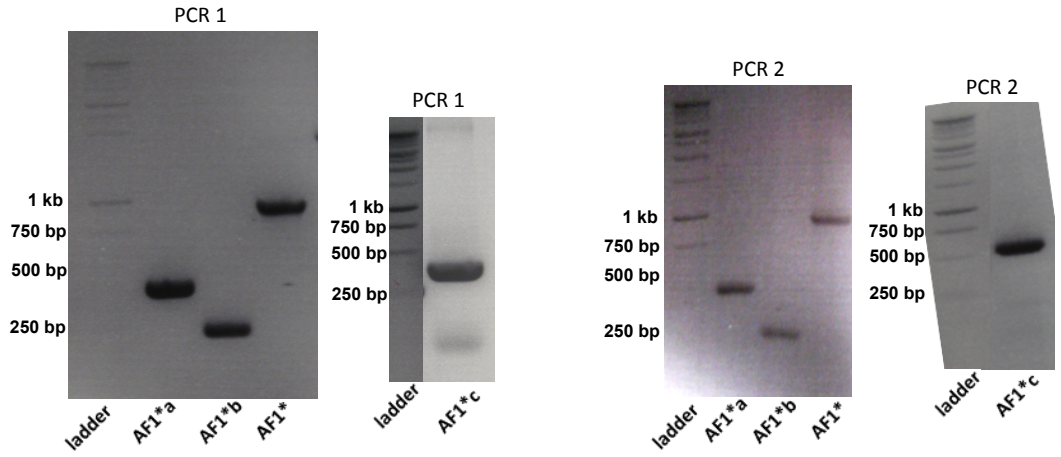


Figure 4.6: Agarose gel for the PCR 1 (left) and PCR 2 (right) products of the four designed AR constructs. Bands are expected around 405 bp for PCR products corresponding to AF1*a, around 231 bp for PCR products corresponding to AF1*b, around 360 bp for PCR products corresponding to AF1*c and around 924 bp for PCR products corresponding to AF1*.

The DNA constructs obtained from the PCR 2 reactions were further inserted into a pDONR/Zeo entry vector (purchased from LifeTechnologies) by BP recombination reaction. The conditions used for the BP reaction are detailed in chapter 3. The obtained BP reaction products were transformed into OmniMAX cells and plated on LB agar plates containing the antibiotic zeocin. Colonies were incubated for 16 hours at 37 ℃, after which the plate was stored at 4 ℃ until individual colonies were picked and grown in LB medium to which zeocin was added. After 16 hours of incubation at 37 ℃ and 250 rpm, the DNA was extracted from this saturated culture with a miniprep and sent for sequencing. The sequencing results confirmed that the inserted DNA sequences in the pDONR/Zeo vector correspond to the designed AF1*a, AF1*b, AF1*c and AF1* constructs with an N-terminal TEV protease cleavage site.

The constructs were then transferred from the pDONR/Zeo entry vector to various destination vectors (purchased from LifeTechnologies or Addgene): pDEST17 (N-terminal His-tag), pDEST15 (N-terminal GST-tag) and pDEST-HisMBP (N-terminal

HisMBP-tag) by LR recombination reaction. The conditions used for the LR reaction are detailed in chapter 3. Similarly, the obtained LR reaction products were transformed into OmniMAX cells and plated on ampicillin containing LB agar plates, as each of these destination vectors has ampicillin resistance. Colonies were incubated for 16 hours at 37 ℃, after which the plate was stored at 4 ℃ until individual colonies were picked and grown in LB medium to which ampicillin was added. After 16 hours of incubation at 37 ℃ and 250 rpm, the DNA was extracted from this saturated culture with a miniprep and sent for sequencing. The four AR constructs were successfully cloned in these three destination vectors.

Following the cloning, expression trials were performed with each of the clones. For this purpose, the sequenced plasmids from the various AR constructs in the three destination vectors were transformed into Rosetta cells. They were grown in LB medium to which ampicillin and chloramphenicol was added to OD 0.6, 0.7, 0.8 or 0.9, induced with various IPTG concentrations (0.1 mM, 0.5 mM or 1 mM), incubated for different combinations of temperature (20 ℃, 25 ℃ or 37 ℃) and time (1.5 h, 3 h or 16 h) and spun down afterwards. The cell pellets were resuspended in buffer, lysed by sonication and separated in a soluble and insoluble fraction by centrifugation. SDS-PAGE gels were run to determine whether the protein had expressed and whether it was present in the soluble or insoluble fraction of the cell lysate.

We found that all constructs, except for AF1*a, cloned in pDEST17, pDEST15 and pDEST-HisMBP vectors expressed well and were predominantly present in inclusion bodies in the insoluble fraction of the cell lysates in all the conditions tested. It is commonly observed that IDPs are directed to inclusion bodies by the cell when they are overexpressed [403]. This is true, even when the IDPs are fused to a solubility-promoting tag (like a GST-tag or an MBP-tag) [403]. Moreover, fusion of IDPs to tags that promote the formation of inclusion bodies has been suggested as a strategy to promote their expression [403]. We developed and optimized an expression and purification protocol for His-tagged AF1*b, AF1*c and AF1* starting from the pDEST17-AF1*b, pDEST17-AF1*c and pDEST17-AF1* clones. The detailed protocols are given in chapter 3.

In contrast to the other constructs, the expressed AF1*a construct was mainly soluble with all three tags and thus present in the supernatant of the cell lysate where it appeared to already degrade during the expression. We attempted to drive the protein into inclusion bodies during expression by inducing with high concentrations of IPTG (2 mM) and incubating for a short time at high temperature after induction (37 ℃ for 3 hours). However, this approach was not successful. In collaboration with Dr. Nick Berrow (PECF of IRB Barcelona) the AF1*a construct was further cloned in three ad-

ditional vectors: pOPINS (N-terminal His-SUMO tag) [354, 355], pPEU10 (N-terminal His-thioredoxin tag with 3C protease cleavage site) and pPEU11 (N-terminal His-Z tag with 3C protease cleavage site). A thorough screening for expression conditions of AF1*a cloned in these three vectors and in the three Gateway vectors (pDEST17, pDEST15 and pDEST-HisMBP) did, however, not yield the AF1*a protein construct. Other possible explanations for the degradation pattern we observed on gels of AF1*a in the soluble fraction of the cell lysate are toxicity of the protein fragment and/or rare codon usage due to the fact that a human protein construct is expressed in a bacterial cell.

Since we were not able to express AF1*a via conventional expression in *E. coli*, cell-free expression techniques were explored. Using the Cell-free Expression EasyXpress Protein Synthesis Kit (Qiagen) we successfully expressed His-tagged AF1*a on a small test scale (not isotopically labeled) (see Fig. 4.7a). To obtain a double isotopically labeled (^{15}N and ^{13}C) sample suitable for NMR experiments we further attempted to express His-tagged AF1*a via cell-free expression as part of a workshop organized by the University of Gothenburg (Prof. Göran Karlsson). Although isotopically labeled His-tagged AF1*a was expressed (see Fig. 4.7b), the yield was not high enough to further purify the protein for subsequent NMR experiments. In addition, also here, degradation of the His-tagged AF1*a protein construct was observed. Cell-free expression to obtain isotopically labeled proteins requires the addition of ^{15}N- and ^{13}C-labeled amino acids, which substantially increases the cost of the protein production. Since the backbone assignment of AF1*a was not strictly necessary for the subsequent assignment of the larger AF1* construct, we did not further pursue the expression of AF1*a. The resonances of AF1* corresponding to residues of AF1*a (AR 142–275) were assigned by first transferring the assignment of AF1*b (AR 265–340) and AF1*c (AR 330–448) to AF1* (AR 142–448) and subsequently assigning the remaining 124 residues (AR 142–265) conventionally (see section 4.5).

His-tagged AF1*b, AF1*c and AF1* were expressed and purified according to the protocols in chapter 3. Non-isotopically labeled proteins were produced in LB medium, whereas uniformly isotopically labeled proteins were expressed in minimal MOPS medium to which $^{15}NH_4Cl$ was added as a nitrogen source and ^{13}C-labeled glucose (double labeled sample) or unlabeled glucose (single labeled sample) as a carbon source. Briefly, an LB preculture was prepared either from a glycerol stock or a new transformation in Rosetta cells of the pDEST17 plasmid containing the AR insert of choice. For non-isotopically labeled expression, LB medium was inoculated 1/100 with saturated preculture, while for isotopically labeled expression, a second preculture in minimal MOPS medium was prepared by inoculating 1/50 with saturated LB precul-

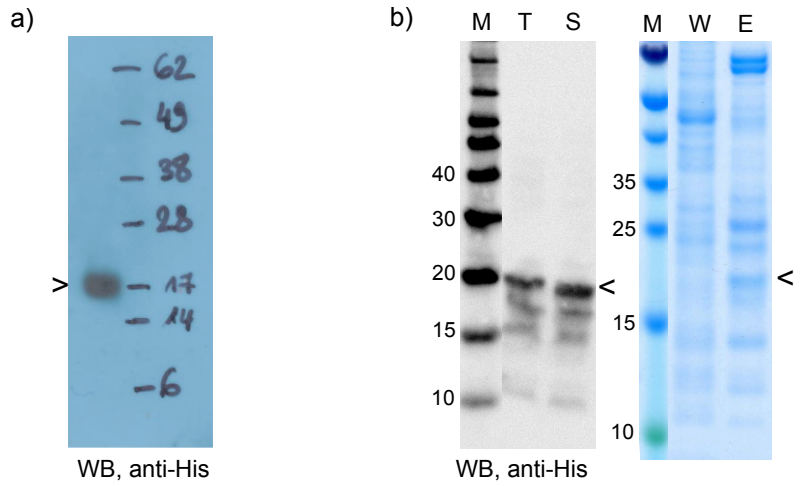


Figure 4.7: Cell-free expression of His-tagged AF1*a (17.2 kDa). a) Small scale cell-free expression with Expression EasyXpress Protein Synthesis Kit (Qiagen). A western blot is shown in which the His-tagged protein is detected by a His-antibody (indicated by an arrow (>)). b) Cell-free expression of isotopically labeled His-tagged AF1*a in workshop University of Gothenburg. On the left, a western blot is shown in which the protein is detected by a His-antibody (M: marker, T: total cell lysate, S: soluble fraction cell lysate). On the right, an SDS-PAGE gel is shown of the Ni^{2+} column run of the expressed protein (M: marker, W: wash (no His-tagged AF1*a present), E: eluted fraction Ni^{2+} column). The protein is indicated by an arrow (<).

ture. The following day, minimal MOPS medium was inoculated 1/40 with saturated MOPS preculture. Bacteria were grown to OD 0.8, induced with IPTG (0.5 mM for AF1*b and AF1*, 1 mM for AF1*c) and further incubated for 16 hours at 25 ℃ (AF1*b and AF1*c) or 20 ℃ (AF1*) and 220 rpm. After 16 hours of incubation, the cell cultures were centrifuged and the pelleted cells were resuspended in lysis buffer (see chapter 3), treated with DNases and lysed using a cell disruptor (one pass at 20 KPSI). Lysed cells were centrifuged to separate the soluble from the insoluble fraction. Since the fusion proteins were present in inclusion bodies, pellets were resuspended in lysis buffer containing 8 M urea (see chapter 3) and filtered (diameter filter pores is 0.22 μM) before loading on a Ni^{2+} column for affinity chromatography. The produced His-tagged protein was retained on the Ni^{2+} column and eluted with an imidazole gradient. Urea was removed via multiple dialysis rounds against buffer without urea. Subsequently, TEV protease was added (1/50 (mol/mol) for AF1*b and AF1*c and 1/10 (mol/mol) for AF1*) to cleave the His-tag (incubation at 4 ℃ for 16 hours or at 25 ℃ for 3 hours). TEV protease cleaves between the glutamine (Q) and glycine (G) residues of the TEV protease recognition site (ENLYFQG), and consequently TEV cleavage results in a non-native glycine residue at the N-terminus of the designed AR constructs. After TEV cleavage, the reaction mixture was again passed over a Ni^{2+} column (reverse Ni^{2+} affinity chromatography) equilibrated in buffer containing 8 M urea (AF1*b and AF1*c) or without urea (AF1*)). Uncleaved His-tagged protein constructs, the His-tagged TEV protease and the cleaved His-tag were retained on the column, whereas the cleaved protein construct was present in the flow-through of the load. Finally, the cleaved protein construct was passed over a size exclusion chromatography column to separate it from remaining impurities and to simultaneously exchange the buffer to the one used for NMR (20 mM sodium phosphate, 1 mM TCEP, pH 7.4).

In figure 4.8 representative gels are shown of the first Ni^{2+} column for each of the AR protein constructs, showing the eluted His-tagged protein constructs, and the final obtained AF1*b, AF1*c and AF1* eluting from the SEC column and used for further experiments. The mass of both the His-tagged and cleaved protein constructs was confirmed by mass spectrometry analysis.

4.4 The designed protein constructs are IDPs

We performed exploratory biophysical experiments for each of the constructs to determine whether they behaved as IDPs, as predicted. All experiments were carried out at physiological pH (pH 7.4), in 20 mM sodium phosphate buffer and in the presence of reducing agent (1 mM TCEP) as the AF1* construct contains eight cysteine residues

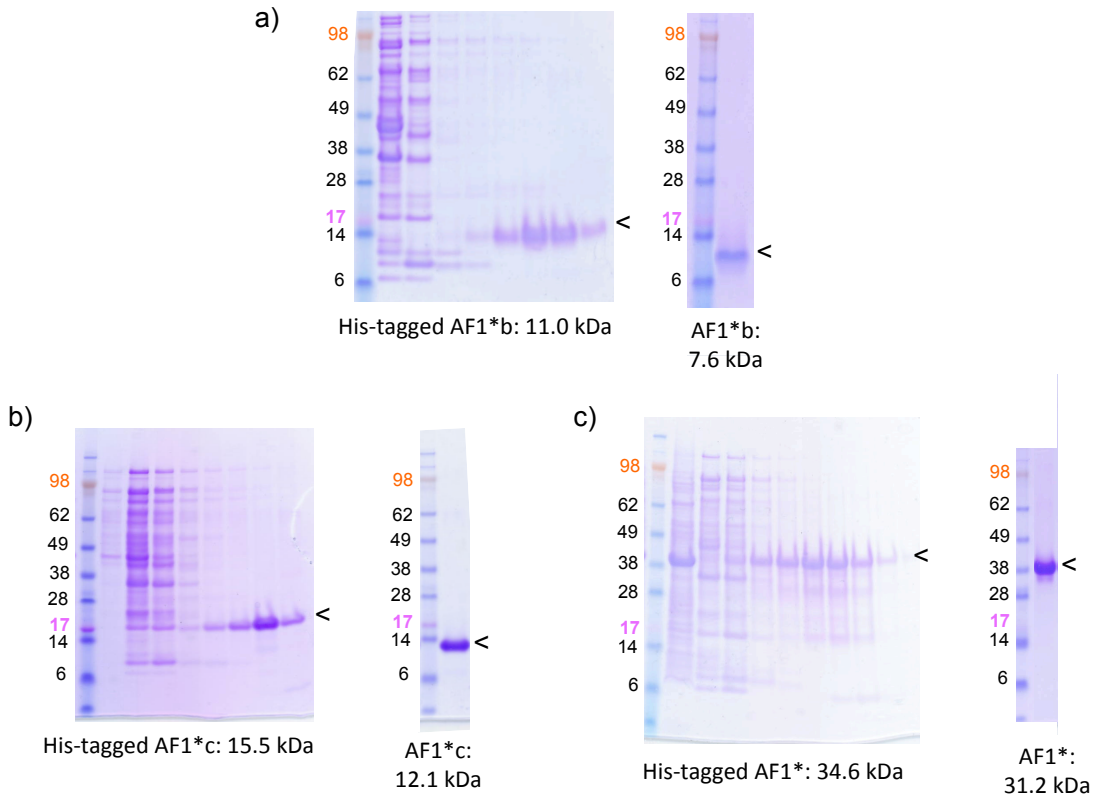


Figure 4.8: Representative SDS-PAGE gels of His-tagged protein constructs eluting from the first Ni^{2+} column and the final obtained protein without the His-tag for AF1*b (a) AF1*c (b) and AF1* (c). The position of the proteins is indicated with an arrow (<).

(C175,C238, C265, C282, C288, C325, C404 and C448). NMR experiments and size exclusion chromatography were performed at 278 K and CD either also at 278 K or at room temperature.

4.4.1 NMR

The [^{1}H,^{15}N]-HSQC spectrum of each of the constructs is that expected for an IDP, with low H^N chemical shift dispersion and relatively wide dynamic range of peak intensities [359] (see Fig. 4.9).

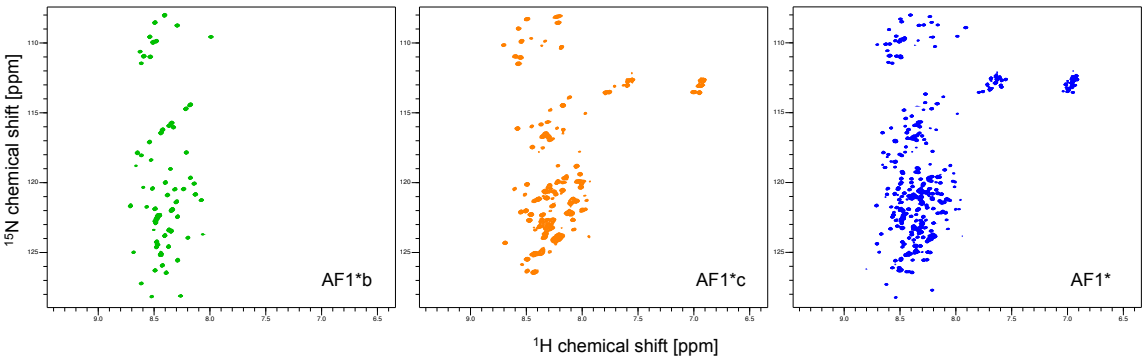


Figure 4.9: [^{1}H,^{15}N]-HSQC spectra of AF1*b (green), AF1*c (orange) and AF1* (blue). All spectra were recorded at 278 K and pH 7.4.

To determine whether the observed wide dynamic range (at 278 K and pH 7.4) was related to solvent exchange of the amide protons, or instead was due to conformational exchange, experiments were carried out at various pH values and temperatures (see Figs. 4.10, 4.11 and 4.12). Amide protons in IDPs exchange rapidly with the solvent due to the lack of protection from structure formation [362]. Because of this, it is necessary to minimize solvent exchange, especially for these proteins, by optimizing the conditions of pH and temperature, as the exchange rate depends on these two parameters [404]. The rate of solvent exchange is minimized at pH values between 3 and 5 and at low temperatures [404, 405].

Therefore, we monitored the peak intensities of AF1*c at decreasing values of pH, while maintaining the temperature at 278 K (see Fig. 4.10). As can been seen in figure 4.10 the peak intensity of the majority of the residues is not affected by the pH, indicating that these resonances are not broad due to solvent exchange. The peak intensity of residues 330–352 and 414–433 does increase for lower pH values, indicating solvent exchange contributes to the wide dynamic range of these resonances. In addition, HSQC spectra of both AF1*b and AF1*c measured at increasing temperature while maintaining the pH constant showed increased line broadening with temperature (see Figs. 4.11

and 4.12). These results illustrate the line broadening at higher temperature due to solvent exchange when measuring at physiological pH, and confirm the need to record HSQC spectra of AR constructs at low temperature for this pH.

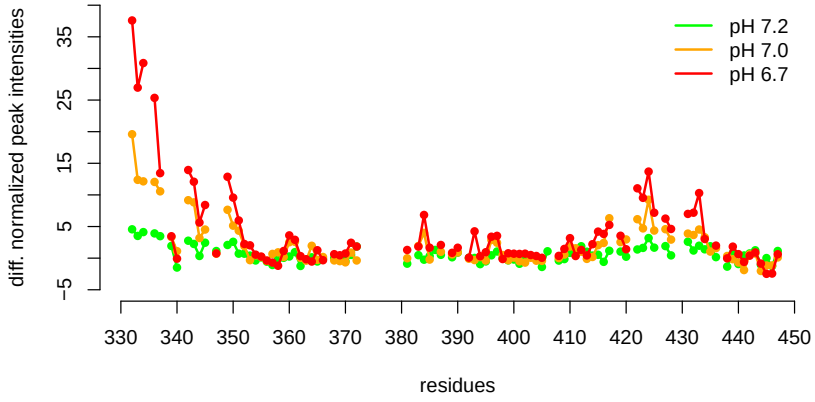


Figure 4.10: Difference in normalized peak intensities of AF1*c for [^{1}H,^{15}N]-HSQC spectra recorded at 278 K and pH 7.2, pH 7.0 and pH 6.7, relative to the normalized peak intensities at pH 7.3.

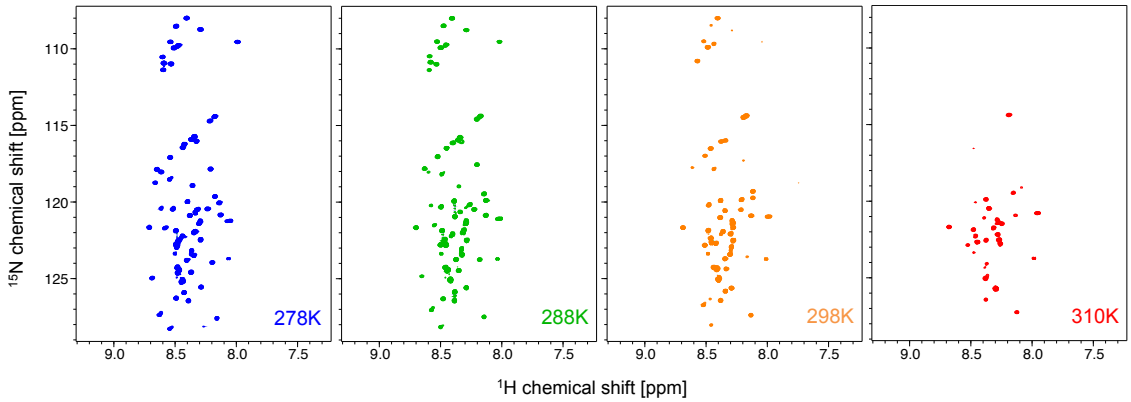


Figure 4.11: [^{1}H,^{15}N]-HSQC spectrum of AF1*b at different temperatures (278 K, 288 K, 298 K and 310 K) and pH 7.15.

The conditions under which we performed our NMR studies (pH 7.4 and 278 K) are a compromise between physiological pH and low temperature, which is common for the study of IDPs ([362, 406] and appendix A.4 in which table S1 of the supplementary information of [406] is shown, listing more than 30 reported experimental conditions that have been used for NMR studies on IDPs.). Under these conditions, the rate of exchange of amide protons with solvent is slowed down, as demonstrated in figures 4.10, 4.11 and 4.12. Consequently, the wide dynamic range observed for the amide cross peaks in the HSQC spectra recorded at this pH (pH 7.4) and temperature (278 K) is

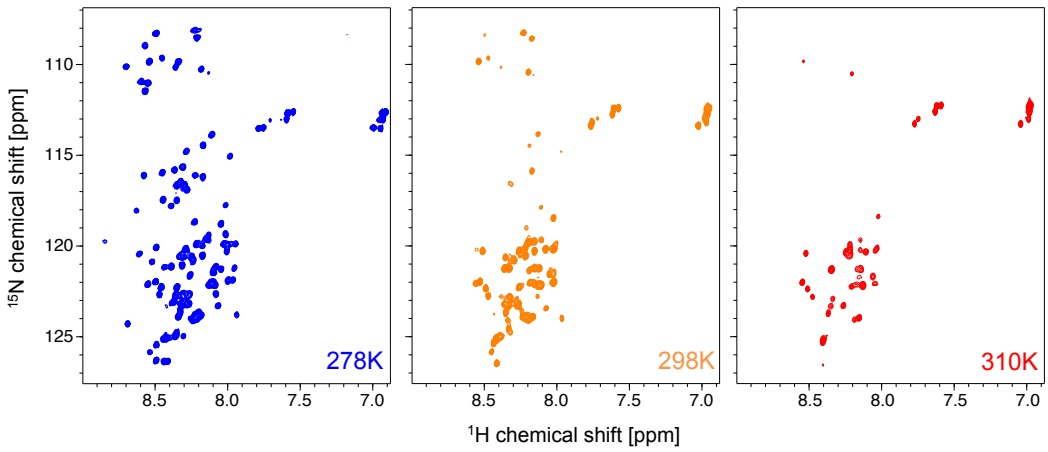


Figure 4.12: [^{1}H,^{15}N]-HSQC spectrum of AF1*c at different temperatures (278 K, 298 K and 310 K) and pH 7.4.

mainly related to conformational exchange and indicative of transient structure of the constructs.

The peak intensities in AF1* are plotted in Fig. 4.13. It is immediately clear that residues over the entire sequence show line broadening due to conformational exchange. The extent of line broadening is substantial, indicating a high degree of fluctuating residual structure. In the C-terminal 120 residues of AF1*, corresponding to the AF1*c construct, this line broadening is particularly high, suggesting a relatively ordered conformation in that part of the protein.

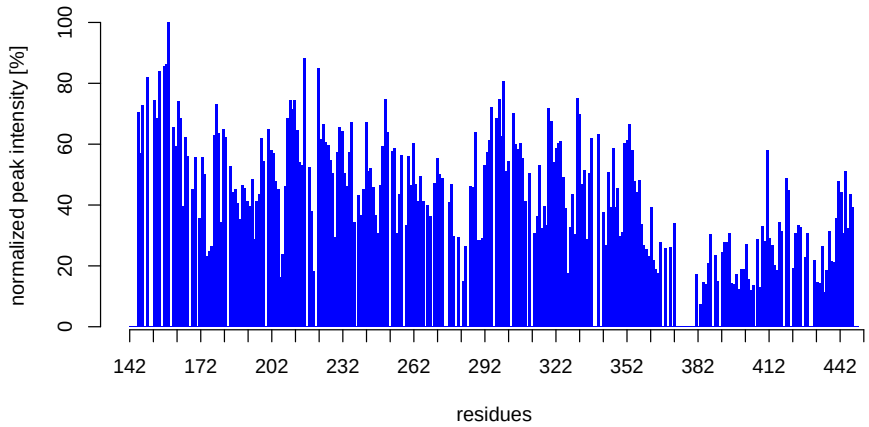


Figure 4.13: Normalized peak intensities for AF1* at 25 μM. Peaks that were assigned to more than one residue were removed for this analysis.

4.4.2 Size exclusion chromatography

Size exclusion chromatography can be used to correlate the retention time of a polypeptide chain of a given molecular weight with its hydrodynamic radius. Since the hydrodynamic radius increases with the degree of disorder in a protein, it is possible to distinguish between more ordered and more disordered conformational states of a protein based on its experimental hydrodynamic radius [362].

To assess the conformational state of a natively unfolded protein one can compare the experimental hydrodynamic radius obtained from size exclusion chromatography with the theoretical hydrodynamic radius for different degrees of collapse of the protein. The theoretical values can be calculated through a set of equations derived by Uversky *et al.* [407]:

$$log\left(R_S^{NU(coil)}\right) = -(0.551 \pm 0.032) + (0.493 \pm 0.008) \cdot log(M) \tag{4.1}$$

$$log\left(R_S^{NU(PMG)}\right) = -(0.239 \pm 0.055) + (0.403 \pm 0.012) \cdot log(M) \tag{4.2}$$

where M is the molecular weight of the protein in Dalton. Equation 4.1 gives the hydrodynamic radius (R_S) for natively unfolded proteins in an extended conformation (denoted as natively unfolded (NU)(coil)) whereas equation 4.2 is for natively unfolded proteins in a more collapsed conformation, i.e. with hydrodynamic characteristics closer to those of a premolten globule (denoted as NU(PMG)) [362].

The experimental hydrodynamic radius can be derived from the elution volume of the protein on a size exclusion column. It is necessary to determine the void volume, V_0, and total volume, V_T, of this column using blue dextran and acetone, respectively. With the elution volume, V_E, of the protein, the partition coefficient, K_{av}, can be calculated:

$$K_{av} = \frac{V_E - V_0}{V_T - V_0} \tag{4.3}$$

The experimental hydrodynamic radius can subsequently be derived using equation 4.4, as obtained from the calibration of the size exclusion column used for these experiments.

$$R_S = 89.141 \cdot \sqrt{-log(K_{av})} - 13.63 \tag{4.4}$$

Analytical size exclusion chromatography was performed with AF1*c and AF1* (see Fig. 4.14). The results show that AF1* exists mainly as an extended coil, whereas AF1*c adopts a more collapsed conformation (Table 4.1). This is in agreement with the more pronounced line broadening observed by NMR in the C-terminal region of AF1*, corresponding to AF1*c (see Fig. 4.13) and correlates with the predicted order

propensity in this part of the protein (see Fig. 4.1a).

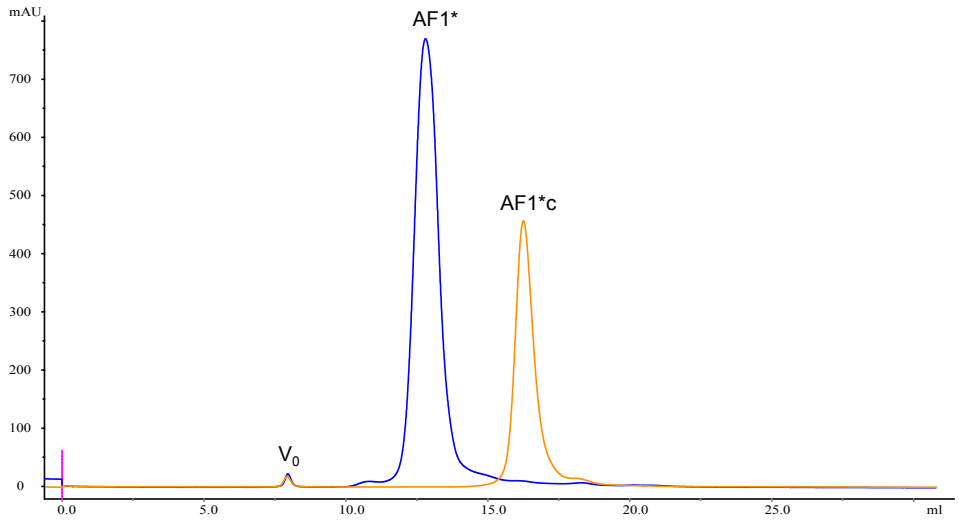


Figure 4.14: Analytical size exclusion chromatography of AF1*c and AF1*. AF1*c eluted at 16.2 mL, and the elution volume of AF1* was 12.7 mL.

Table 4.1: Theoretical hydrodynamic radii for AF1*c and AF1*, both for a fully extended conformation and for a premolten globular conformation, as obtained from equations 4.1 an 4.2. The experimental hydrodynamic radii were obtained using equations 4.3 and 4.4 and the elution volumes in SEC (Fig. 4.14). The experimentally obtained hydrodynamic radius of AF1*c is close to that expected for a premolten globular conformation, whereas that of AF1* is closer to a hydrodynamic radius of an extended coil.

	MW(kDa)	$R_S^{NU(coil)}$(Å)	$R_S^{NU(PMG)}$(Å)	experimental R_S(Å)
AF1*c	12.1	29.0	25.5	26.2
AF1*	31.2	46.2	37.3	46.3

4.4.3 CD spectrum

The far-UV CD spectrum of each of the constructs indicates they are IDPs (see Fig. 4.15). The spectra show a strong negative band near 200 nm, characteristic for disordered polypeptides [408]. Furthermore, no minima at 209 nm and 222 nm that are indicative of α-helical structure [408] are present, which indicates that the protein constructs mainly behave as ID chains. In addition, the far-UV CD spectrum of AF1*c was recorded both at 5 °C and at room temperature, and the secondary structure of this protein construct was found to be comparable at both temperatures.

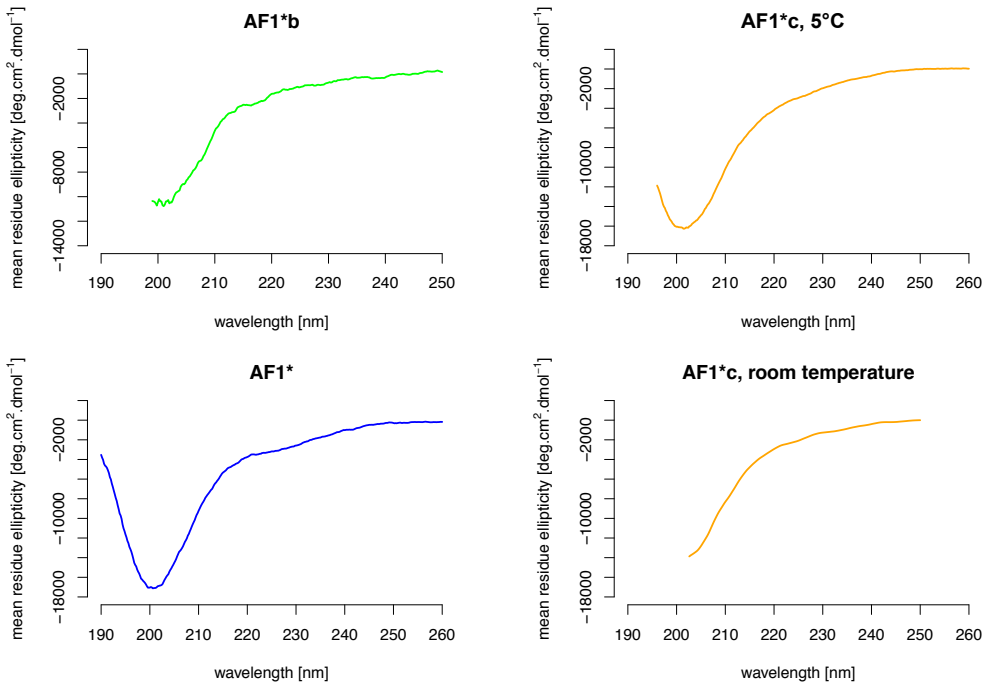


Figure 4.15: Far-UV CD spectra for AF1*b (at 20 °C), AF1*c (at 5 °C and at room temperature) and AF1* (at room temperature).

4.5 Backbone assignment of AF1*

The assignment of the backbone resonances is a prerequisite for structure determination of proteins by NMR. This process involves associating each peak that is observed in a spectrum with one or more nuclei present in the protein. We used 3D triple resonance experiments (see section 3.6.3.3) for the backbone assignment of each of the protein constructs. These require uniformly double labeled samples (^{13}C and ^{15}N).

As explained in section 4.2 we used a *divide-and-conquer* approach for the assignment of AF1* to overcome the severe peak overlap expected for the NMR spectra of AF1* due to its size and the low dispersion of chemical shifts for disordered proteins. This overlap is already reflected in the HSQC spectrum of AF1* (see Fig. 4.9c).

Superposition of the HSQC spectrum of AF1* with those of its individual fragments (Fig. 4.16) revealed that many of the peaks in the shorter constructs overlay with a peak in the longer AF1* construct, which suggested that it would be possible to assign AF1* using the individual assignments of the shorter constructs. We therefore assigned the resonances of the shorter constructs, AF1*b and AF1*c, first individually to subsequently use this information for the assignment of AF1*. All spectra were acquired at 278 K and at pH 7.4.

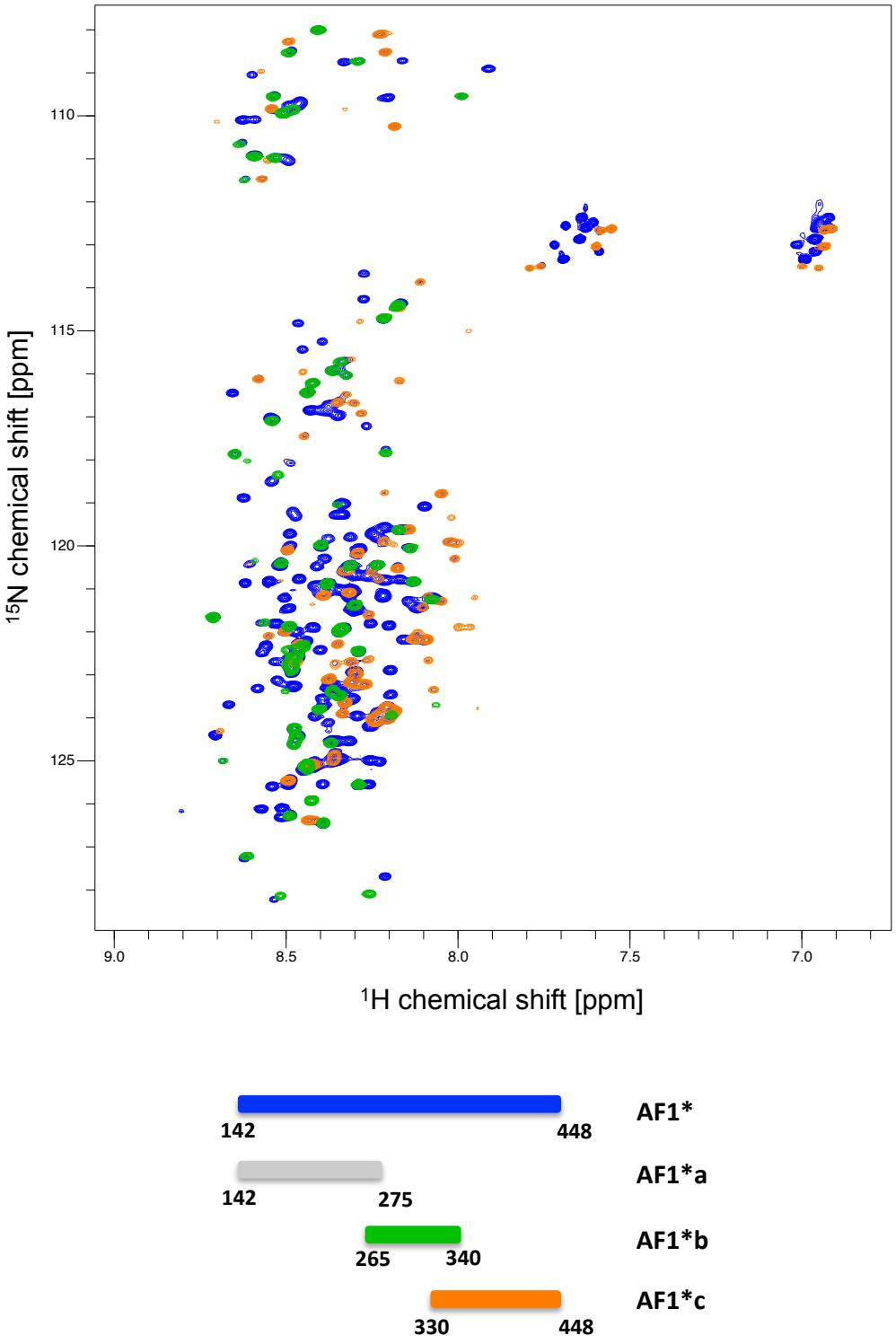


Figure 4.16: Superposition of the HSQC spectra of AF1*b (green), AF1*c (orange) and AF1* (blue) with indication of the protein constructs.

4.5.1 AF1*b

Protein backbone assignment involves two steps. First, to group all the peaks in the various spectra that belong to the same spin systems and to establish the sequential connectivity between those spin systems. Second, to assign each of these spin systems to the corresponding amino acids of the protein.

As explained in the methodology section (3.6.3.3), three sets of triple resonance experiments are used to link protein residues to their direct neighbors. The sequential connectivity can be followed through the C_α [HNCA and HN(CO)CA], the C_β [CBCANH and CBCA(CO)NH] and the carbonyl CO [HNCO and HN(CA)CO] nuclei.

Some residue types are easily recognized, like alanine (low C_β, $C_\beta \approx 19.3$ ppm), threonine and serine ($C_\beta > C_\alpha$, $C_\beta \approx 70$ ppm and $C_\beta \approx 64$ ppm, respectively) and glycine (low C_α, $C_\alpha \approx 45.4$ ppm and no C_β nucleus) [409]. This can be very useful to associate a spin system with a particular amino acid of the protein.

Using the sequential connectivity it is possible to look for the neighboring residues of those that are readily identified to belong to a specific residue type, and identify them as a specific string of amino acids of the protein. The sequential connectivity is broken by proline residues. Since proline side chains do not have an amide proton, there is no peak observed in the HSQC spectrum which contains the cross peaks of the amide proton and amide nitrogen. Nevertheless, proline C_α, C_β and CO chemical shifts can be assigned using the i-1 connectivity of the following residue.

AF1*b is the smallest, and thus least overlapped, of the three protein constructs. We started with the assignment of this construct. The assigned [^{1}H,^{15}N]-HSQC of AF1*b is shown in figure 4.17.

All the non-proline backbone resonances of the AF1*b construct were fully assigned, with the exception of the non-native glycine residue at the very N-terminus of the construct, a result of the cloning strategy, and the amide proton and nitrogen chemical shift of C265 adjacent to that. For these two residues no resonances were observed in the HSQC spectrum. The C_α, C_β and CO resonances of all proline residues were assigned, apart from P274, which precedes another proline residue, and P340, which is the C-terminal residue of the construct.

A complete list of the AF1*b backbone assignment (H^N, N^H, CO, C_α, C_β) can be found in appendix A.5.

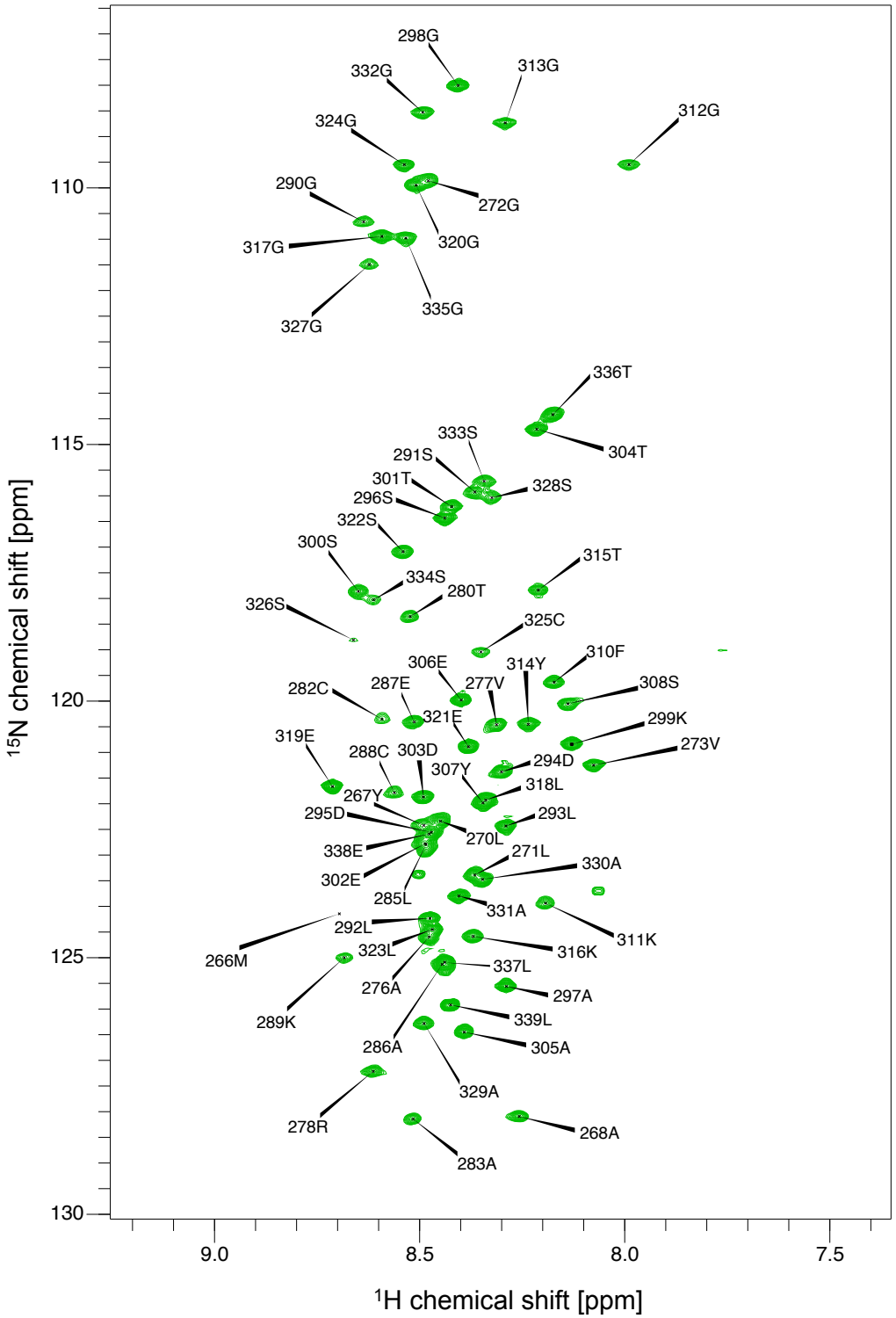


Figure 4.17: The [^{1}H,^{15}N]-HSQC of AF1*b (AR 265–340) with assignment of the backbone amide proton and nitrogen, at 200 μM, in 20 mM sodium phosphate and 1 mM TCEP, at pH 7.4 and 278 K.

4.5.2 AF1*c

The AF1*c construct is 120 residues long and contains several amino acid repeat regions, like polyproline 372–379 (8 consecutive prolines), polyalanine 398–402 (5 consecutive alanines), polyalanine 354–356 (3 consecutive alanines), polyalanine 427–429 (3 consecutive alanines) and polyserine 430–432 (3 consecutive series).

Despite the repetitive nature of the primary sequence, common for IDPs, all residues were fully assigned, with the exception of the H^N and N^H chemical shifts of A330 at the N-terminus of the construct, and the preceding non-native glycine residue as the resonances associated to these residues were not visible in the spectra. The polyproline stretch (P372–P379) could not be assigned, apart from the C_α, C_β and CO chemical shifts of P379. The assigned [^{1}H,^{15}N]-HSQC of AF1*c is shown in figure 4.18.

In the spectra, for some residues preceding a proline residue a minor peak is observed in addition to the main one. This is for instance clearly visible for residues G419 and G446 as well as for Y445, which is in position i-2 of proline residue P447 (see Fig. 4.18). This is related to *cis-trans* isomerization of the peptide bond. All peptide bonds can undergo *cis-trans* isomerization, but usually the *trans* isomer is highly favored over the *cis* isomer as the former causes less steric repulsion. In the case of proline, however, the *cis-trans* equilibrium only slightly favors the *trans* form in peptidyl-proline bonds [410]. Consequently, the chemical environment of the residues preceding prolines varies with the isomerization state of the peptide bond between the residue directly before the proline residue and the proline residue. Therefore, for those residues for which we observed a main and a minor peak, it is expected that the main peak corresponds to that of the residue when the peptide bond between the residue preceding proline and proline adopts the *trans* isomer form, and the minor peak to that of the residue when that peptide bond adopts the *cis* isomer form.

A complete list of the AF1*c backbone assignment (H^N, N^H, CO, C_α, C_β) can be found in appendix A.6.

4.5.3 AF1*: *divide-and-conquer*

As anticipated, the 3D spectra of AF1* still displayed severe peak overlap which complicated the assignment. However, it is expected that superposition of the spectra of AF1* and those of its individual fragments will allow to transfer many of the assignments. Indeed, when overlaying the HSQCs of the different protein constructs, many resonances were observed at identical positions in the shorter constructs and in AF1*

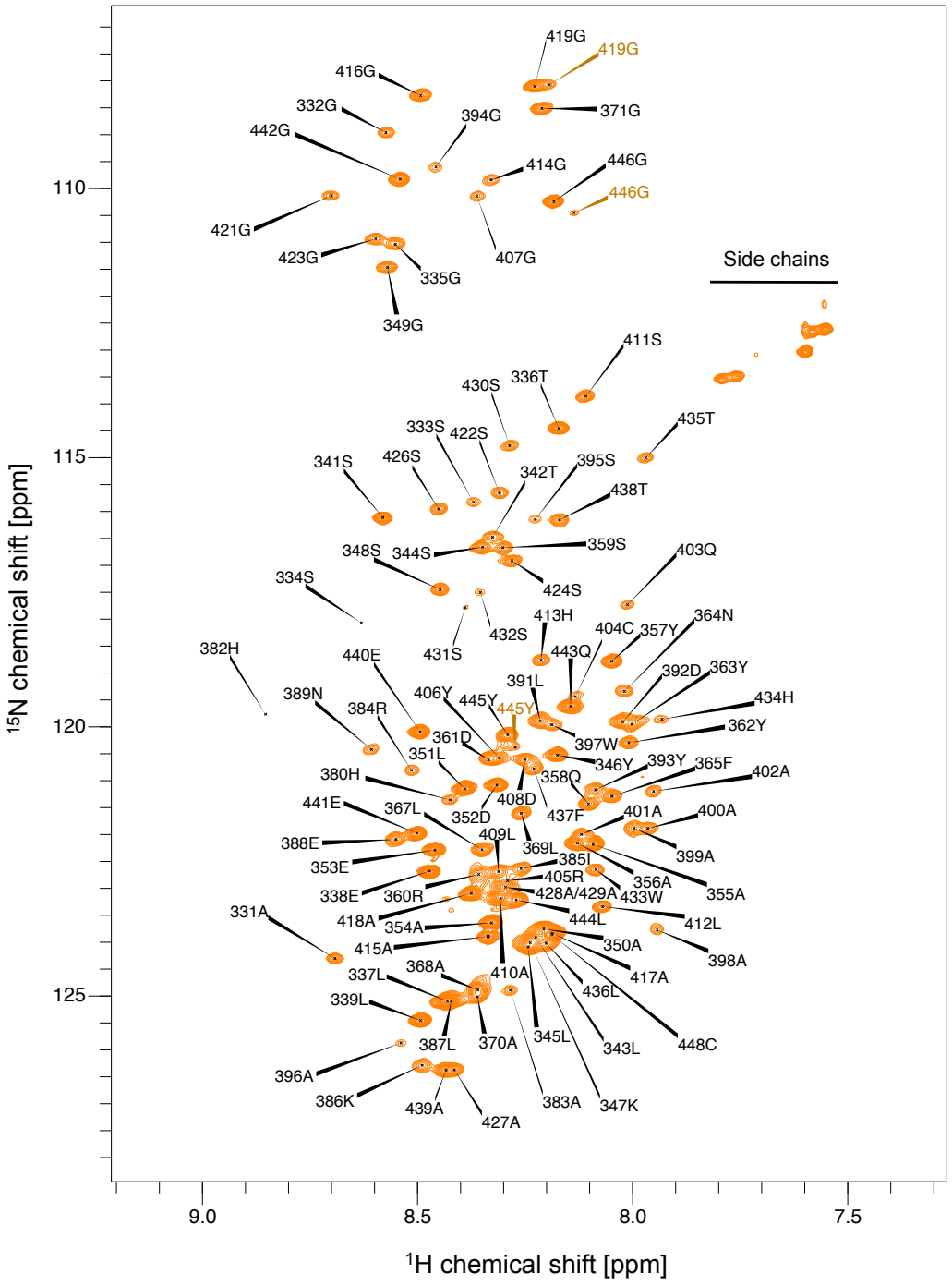


Figure 4.18: The [^{1}H,^{15}N]-HSQC of AF1*c (AR 330–448) with assignment of the backbone amide proton and nitrogen, at 290 μM, in 20 mM sodium phosphate and 1 mM TCEP, at pH 7.4 and 278 K. For residues G419, Y445 and G446, the minor peak present in addition to the main one due to cis-trans isomerization of the proline peptide bond is shown in orange.

(see Fig. 4.19). This was also true when overlaying the 3D spectra. This reflects the ID nature of AF1*.

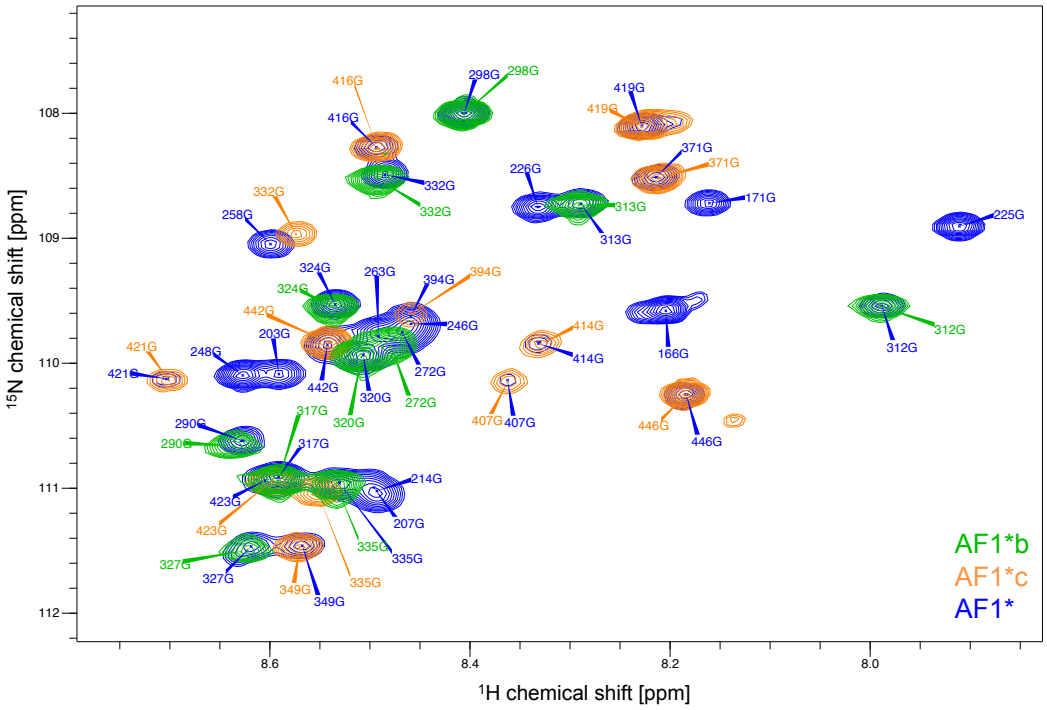


Figure 4.19: Zoom of the glycine region of the HSQC spectra of AF1*b (green), AF1*c (orange) and AF1* (blue) with the assignments for each of the protein constructs in the same colours.

For a large number of peaks we could transfer the assignments from the smaller protein constructs to AF1*. To confidently assign those peaks that did not clearly overlap in the HSQC spectra, we relied on the sequential connectivity between residues that was extracted from the three pairs of triple resonance experiments recorded for AF1*. The assigned [^{1}H,^{15}N]-HSQC of AF1* is shown in figures 4.20 and 4.21.

With this approach we assigned more than 98% of non-proline backbone resonances of AF1*. From the 308 residues, the only ones that were not assigned are the N-terminal non-native glycine residue and the adjacent G142 and L143, and proline residues 150, 274 and 372–378 which are all followed by another proline residue.

We found that the CO chemical shifts of four residues were in a peculiar position compared to the other residues (see Fig 4.22). They were found at uncommonly low values for CO chemical shifts. To assure this was not related to aliasing issues, we recorded the HC plane of the HN(CA)CO with very high spectral width so no folding of the peaks would occur.

The low values for the CO chemical shifts of these four peaks were corroborated

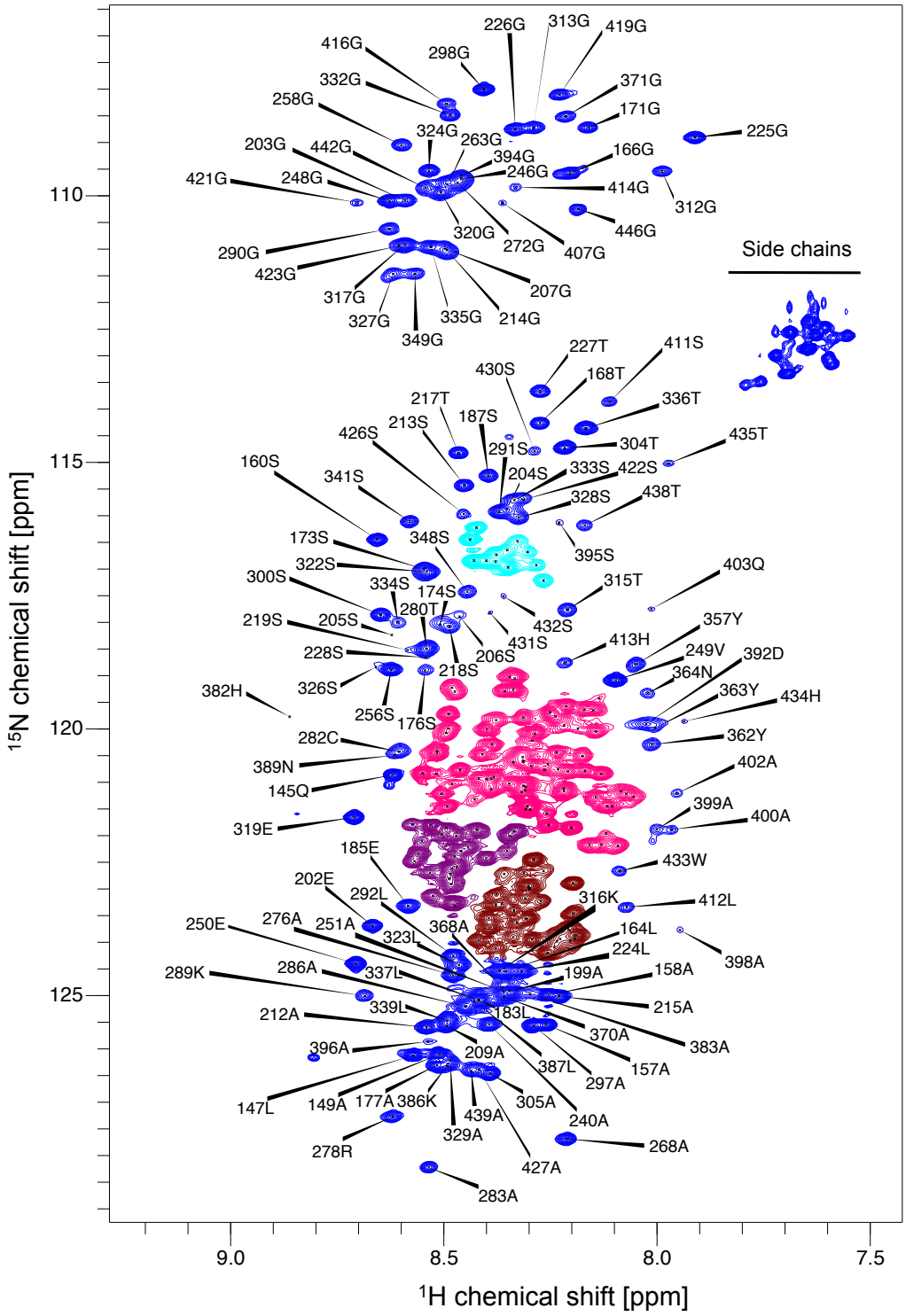


Figure 4.20: The [^{1}H,^{15}N]-HSQC of AF1* (AR 142–448) with assignment of the backbone amide proton and nitrogen, at 340 μM, in 20 mM sodium phosphate and 1 mM TCEP, at pH 7.4 and 278 K. The assignment of the regions colored in cyan, pink, purple and maroon is given in figure 4.21.

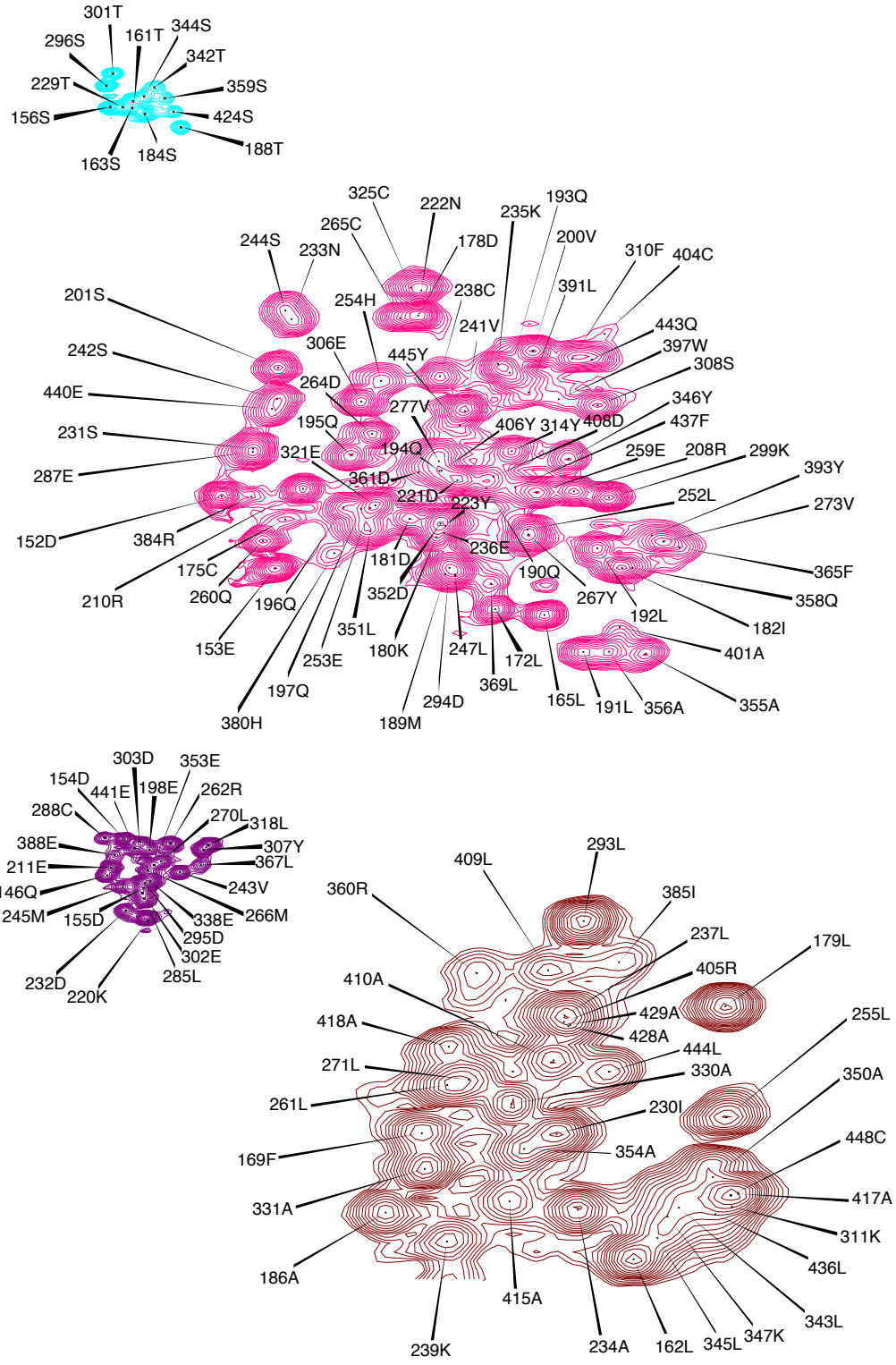


Figure 4.21: The assignment of the backbone amide proton and nitrogen of those resonances of AF1* colored in cyan, pink, purple and maroon in figure 4.20.

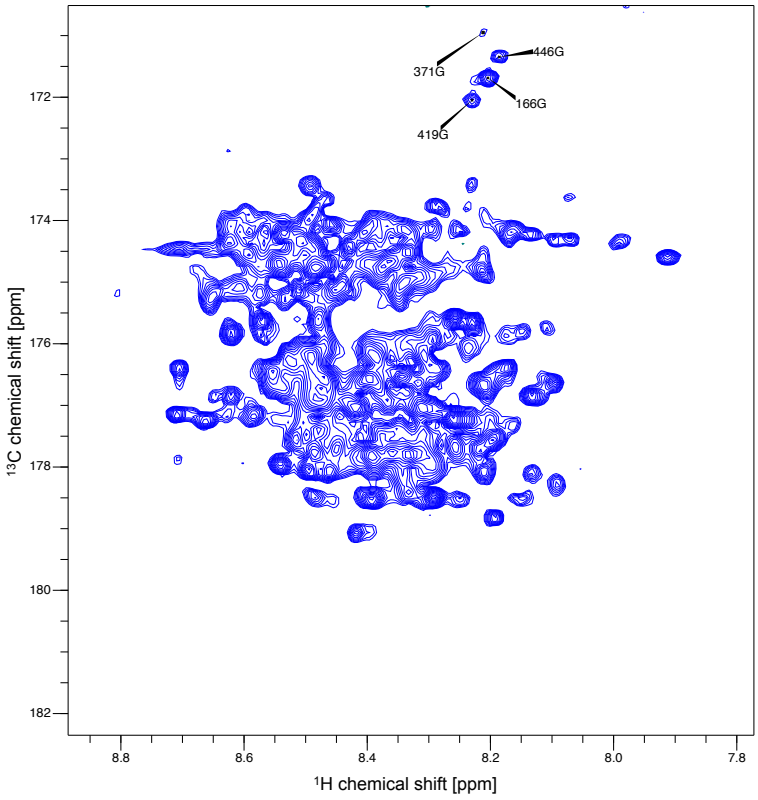


Figure 4.22: The HC plane of the HN(CA)CO of AF1*, recorded with a spectral width of 12 ppm in the ^{13}C dimension and carrier position for ^{13}C at 176.5 ppm. The four peaks with low CO chemical shifts are assigned.

with a CON experiment, which gives cross peaks for the backbone amide nitrogen ^{15}N and the carbonyl ^{13}CO of the preceding residue (see Fig 4.23).

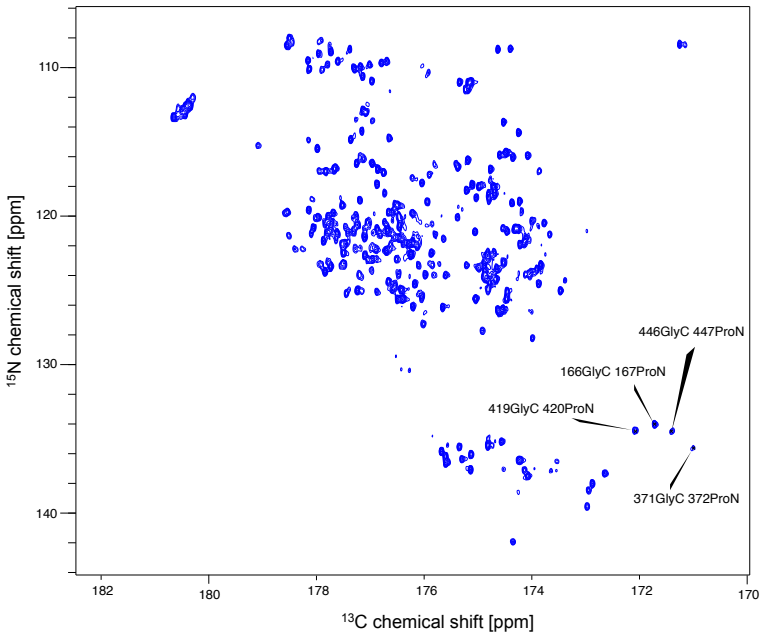


Figure 4.23: CON experiment for AF1*, at 278 K, with spectral width 12 ppm in the ^{13}C dimension and 40 ppm in the ^{15}N dimension.

We identified the four residues with uncommonly low CO chemical shifts as glycine residues followed by a proline. It is well-established that the chemical shifts of residues preceding proline vary substantially from those of the same residues when not followed by a proline [411, 412]. When taking a closer look at the effect of proline on the chemical shifts of all residues of AF1* preceding a proline, we found that the deviation in CO chemical shift, compared to a residue not followed by proline, is in the order of 2 ppm, shifting towards lower chemical shift values. This is not only true for glycine, but also for other amino acid types, as is shown in table 4.2. Given that glycines have the lowest CO chemical shift compared to other residue types, this same offset of ca. -2 ppm gives rise to glycine CO peaks in a distinct region of the spectrum.

A complete list of the AF1* backbone assignment (H^N, N^H, CO, C_α, C_β) can be found in appendix A.7. This dataset forms the basis for probing intramolecular interactions and studying the interactions between AF1 of AR and its binding partners.

As mentioned in section 4.2, we anticipated "boundary conditions" in the chemical shifts of N- and C-terminal residues as a consequence of dividing AF1* in several constructs. The chemical shifts in AF1*c and AF1* were expected to be similar at the C-terminus and to gradually deviate when approaching the N-terminus, whereas the chemical shifts of AF1*b and AF1* were expected to be comparable at the central

Table 4.2: All residues in AF1* preceding prolines with the corresponding average ^{13}CO chemical shift deviation compared to all other residues in AF1* of the same type that are not followed by a proline.

Residues preceding proline	Average ^{13}CO deviation (ppm)
A (149, 158, 215, 268, 283)	-2.91
G (166, 371, 419, 446)	**-2.68**
S (256, 308, 424)	-2.12
L (147, 339)	-2.38
F (169, 365)	-2.01
V273	-2.11
R278	-2.31
T280	-1.63
H380	-1.30
N389	-1.38

region of AF1*b and to deviate going both towards the N- and the C-terminus. We found that these "boundary conditions" are indeed only noticeable up to ten residues from the termini and that they follow the anticipated pattern (see Fig. 4.24).

4.6 Secondary structure propensity

Following the assignment of AF1*, we characterized its secondary structure based mainly on its $^{13}C_\alpha$ chemical shifts. These chemical shifts are sensitive indicators for secondary structure elements in folded proteins and residual secondary structure in IDPs [411].

The main method for determining secondary structure propensity of a protein is analysis of its secondary chemical shifts. These are given by subtracting so-called "random coil" chemical shifts from the experimental chemical shifts. Random coil chemical shifts are empirically determined chemical shifts for a residue in a completely disordered conformation. Typically those have been determined either by studying a series of short disordered peptides [409, 412] or by using available data for coil regions of proteins in the Protein DataBank (PDB) [413] or NMR chemical shift assignments of denatured proteins or IDPs that have been deposited in the BioMagResBank (BMRB) database [414–416]. Since chemical shifts are strongly dependent on the type of amino acid, the random coil value of the corresponding amino acid type is subtracted from the observed chemical shift.

The secondary chemical shifts obtained in this way reliably detect secondary structure elements for folded proteins, but do not perform well for unfolded or partially

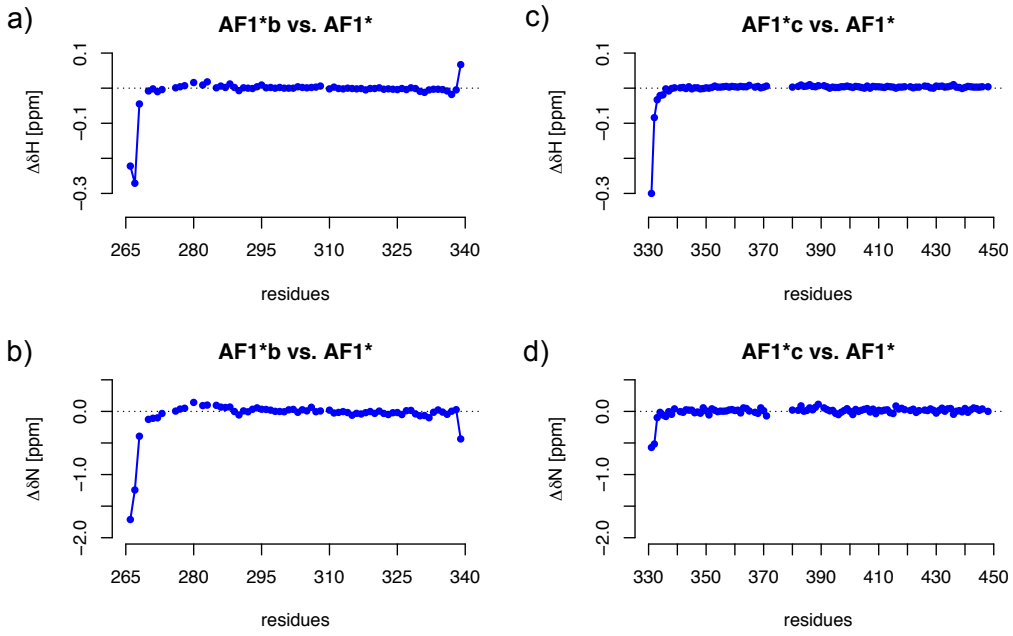


Figure 4.24: Deviation of chemical shifts between the different constructs at the terminal residues of AF1*b, AF1*c and AF1*. a) Difference in ^{1}H chemical shift between AF1*b and AF1*, b) Difference in ^{15}N chemical shift between AF1*b and AF1*, c) Difference in ^{1}H chemical shift between AF1*c and AF1*, d) Difference in ^{15}N chemical shift between AF1*c and AF1*.

structured proteins. This is because chemical shifts are also affected by the immediate local sequence. Local sequence effects are usually too small to affect the determination of secondary structure in folded proteins but they can cause similar deviations from random coil as transiently populated secondary structure elements in IDPs.

Some nuclei, like $^{1}H^{N}$, $^{15}N^{H}$ and ^{13}CO, are highly sensitive to the local amino acid sequence, which makes them less suited for secondary structure analysis but is exploited for sequential assignment, whereas others, like $^{13}C_{\alpha}$ and $^{1}H_{\alpha}$, are less sensitive to local sequence and can therefore be reliably used as indicators of residual secondary structure [411].

For unfolded proteins the difference between observed and random coil chemical shifts is generally much smaller than for folded proteins. When analyzing IDPs, it is therefore imperative to use accurate random coil chemical shift data to be able to distinguish between local deviations from the random coil state and random noise.

There has been a renewed effort in recent years to provide accurate random coil chemical shift data that can be used for the secondary structure analysis of IDPs. These take into account the effect of neighboring residues on the random coil chemical shifts. *ncIDP* (neighbor-corrected Intrinsically Disordered Protein library, [416])

provides neighbor-corrected random coil chemical shifts for IDPs for different sets of random coil data (Schwarzinger *et al.* [411], Wang & Jardetzky [417] and Tamiola *et al.* [416]) by applying neighbor correction factors. Other recent datasets for sequence-corrected random coil shifts, derived by De Simone *et al.* [415] and Kjaergaard & Poulsen [418], are available as separate webservers.[10]

No matter how the random coil chemical shifts are obtained, they can only be approximated using a number of assumptions. The most accurate random coil chemical shift data for a specific protein are given by the chemical shifts of that same protein measured under denaturing conditions, for instance in 8 M urea. However, these data are not always available.

4.6.1 $^{13}C_\alpha$ and $^{13}C_\beta$ chemical shift analysis

The secondary structure propensity for each of the assigned constructs was mainly derived from the $^{13}C_\alpha$ chemical shifts, as this is the most appropriate nucleus to report on secondary structure propensity, and corroborated with the $^{13}C_\beta$ chemical shifts.

In general, large deviations in the chemical shifts of residues preceding proline are observed [411, 412]. Although correction factors exist to account for those deviations, we chose to ignore the residues immediately preceding prolines in the secondary chemical shift analysis because we found that even using the correction factors these residues often gave extreme outliers.

We compared the secondary C_α chemical shifts (ΔC_α) obtained with the random coil values from Wang & Jardetzky [413], Schwarzinger *et al.* [411], De Simone *et al.* [415], Tamiola *et al.* [416] and Kjaergaard & Poulsen [418] as these allow to account for effects of neighboring residues on the random coil values.

As shown in figure 4.25 the different random coil chemical shifts gave similar results. We removed ΔC_α values for glutamate, aspartate and histidine residues when using the random coil shifts from Schwarzinger *et al.*. This was necessary because the charge of the side chains of these residues is different at the pH of our measurements (pH 7.4) and the pH at which the random coil values of Schwarzinger *et al.* were determined (pH 2.0), which gave rise to outliers. Similarly, the ΔC_α chemical shifts of cysteine residues were not taken into account when using the random coil values from De Simone *et al.* The random coil values from De Simone *et al.* were calculated at pH 6.1 (options

[10] Sequence-corrected random coil shifts derived by De Simone *et al.*:
http://www-vendruscolo.ch.cam.ac.uk/camcoil.php.
Sequence-corrected random coil shifts derived by Kjaergaard & Poulsen:
http://www1.bio.ku.dk/english/research/pv/sbin_lab/staff/MAK/randomcoil/script.

pH 2.0 or pH 6.1), and those from Kjaergaard & Poulsen were calculated at pH 7.4 and 278 K.

Continuous stretches of positive secondary C_α chemical shifts report on α-helical propensity, whereas consecutive negative secondary C_α values correspond to propensity of extended conformations like β-strand or polyproline II conformations.

In AF1*, positive ΔC_α values dominate over the entire sequence, indicating its high α-helical propensity. In addition, continuous stretches of negative ΔC_α values are found between residues 142–154 (GLPQQLPAPPDED) and 268–286 (APLLGVPPAVRPTPCAPLA), indicative of β-strand or polyproline II propensities. Since the sequences corresponding to the negative ΔC_α chemical shifts are rich in proline residues, they are unlikely to form β-strand as prolines are known to disrupt both α-helical and β-sheet structural elements when they occur in the middle of the sequence. This suggests these regions have some tendency to adopt an extended conformation other than β-strand, like polyproline II.

Even though the ΔC_α values do not indicate stable secondary structure elements like in globular proteins, the helical propensity is remarkably high compared to other IDPs. Typically, ΔC_α values for IDPs vary mainly between -0.5 ppm and +0.5 ppm (see Fig. 4.26). ΔC_α values between 0 ppm and +0.5 ppm or -0.5 ppm, in stretches of more than five amino acids, are commonly interpreted as indicative of α-helical or β-strand propensities, respectively (see e.g. α-synuclein in Fig. 4.26b), in IDPs. When the ΔC_α shifts increase to values higher than +0.5 ppm or lower than -0.5 ppm the corresponding sequence is almost always recognized as a transient secondary structure element (see e.g. Tau in Fig. 4.26c).

Over the entire sequence of AF1* we observe some degree of α-helical propensity, with values of ΔC_α between 0 ppm and +0.5 ppm, and in addition regions that are clearly more prone to form α-helical structure than expected for an IDP. This is especially true for the region between residues 176 and 202, which overlays with the core region of transactivation unit Tau-1 (^{174}SCSADLKDILSEASTMQLLQQQQQEAVSEGS^{204}). This observation is further discussed in more detail in section 4.6.3. Also between residues 391 and 413 the α-helical propensity is remarkably high. This region corresponds to the residues predicted by Agadir to have a relatively high helical propensity (393–405) (see Fig. 4.1c) and contains a polyalanine stretch.

Secondary C_β shifts can be used to confirm the regions with α-helical propensity and to help identify regions with β-strand propensity, as ΔC_α values are generally poor at distinguishing β-strand from random coil [413]. Opposite to ΔC_α chemical shifts, negative ΔC_β chemical shifts indicate α-helical propensity and positive ΔC_β represent

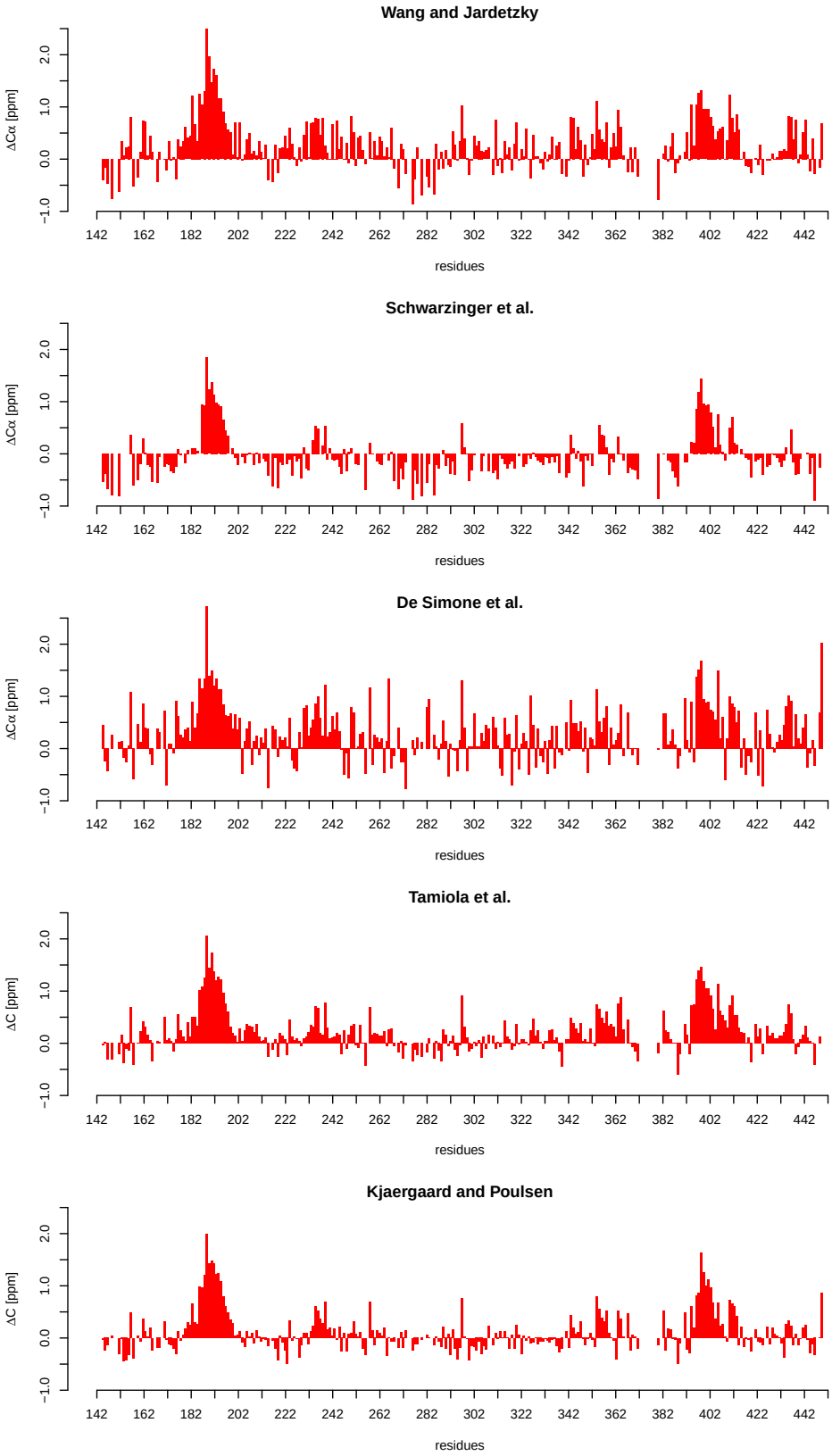


Figure 4.25: Secondary C_α chemical shifts of AF1* obtained with sequence-corrected random coil chemical shifts (from top to bottom) from Wang & Jardetzky, Schwarzinger *et al.*, De Simone *et al.*, Tamiola *et al.*, and Kjaergaard & Poulsen.

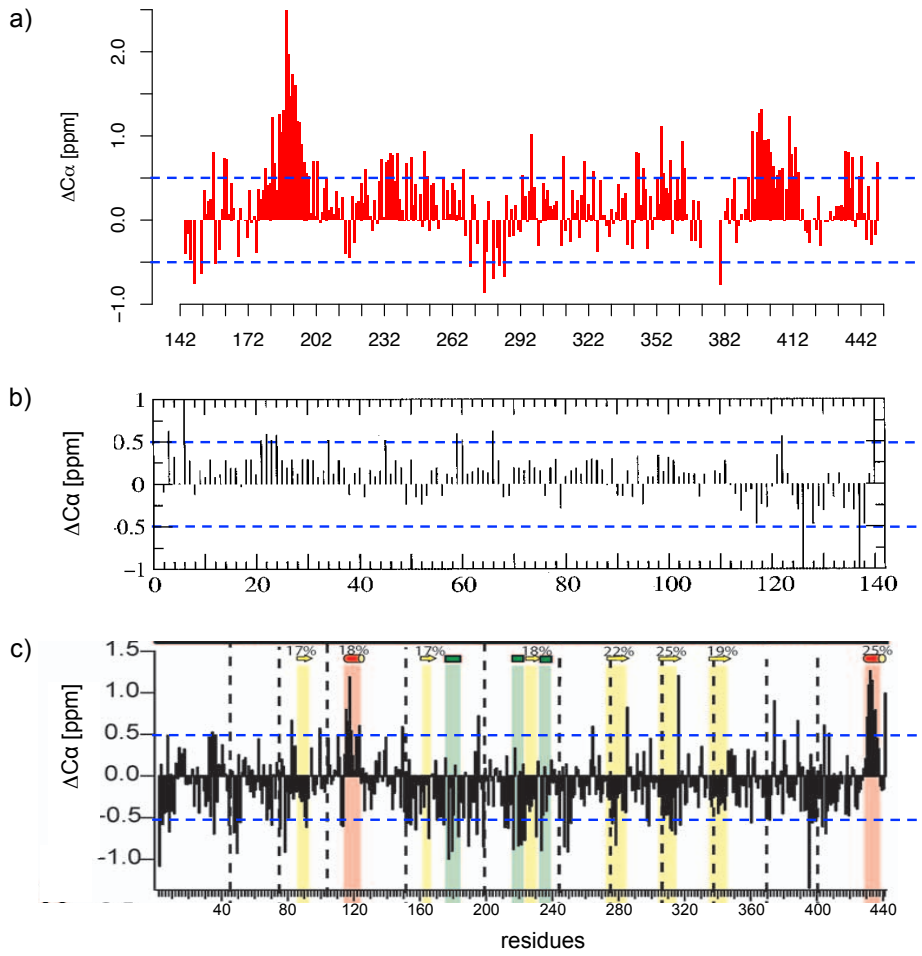


Figure 4.26: a) Secondary C_α chemical shifts of AF1* (measured at pH 7.4 and 278 K) obtained with sequence-corrected random coil chemical shifts from Wang & Jardetzky [413], b) secondary C_α chemical shifts of α-synuclein (measured at pH 7.4 and 283 K, [419]) obtained with random coil values of Wishart *et al.* [420], and c) secondary C_α chemical shifts of Tau (measured at pH 6.8 and 278 K, [401]) obtained with sequence-corrected random coil values from Schwarzinger [411] and random coil values for histidine, glutamate and aspartate from Wishart & Sykes [421].

β-strand propensity. Therefore, stretches of sequence with positive ΔC_α and negative ΔC_β are confidently identified as regions with α-helical propensity and, the other way around, regions with negative ΔC_α and positive ΔC_β are indicating transient β-strand structure. Combining this information in one plot by subtracting the ΔC_β values from the ΔC_α values (ΔC_α-ΔC_β) allows immediate identification of regions with secondary structure propensity.

When overlaying the ΔC_α and ΔC_β chemical shifts determined with the sequence-corrected random coil chemical shifts of Tamiola *et al.* [416], which are representative for the other random coil chemical shift lists tested, several stretches with α-helical propensity can be distinguished (see Fig. 4.27). These are located between residues 176–202, 204–212, 229–239, 353–364, 391–413 and 427–436. In addition, the stretches of sequence we identified before to have negative ΔC_α values also display positive ΔC_β values (residues 142–154 and 268–286). As explained before, it is unlikely these regions will adopt a β-strand conformation due to the presence of several proline residues. Taking into account both ΔC_α and ΔC_β information, the residues flanking the polyproline stretch (residues 367–370 and 385–388) are also identified to have a tendency to adopt a non-helical conformation. The histidine residues gave outliers for the ΔC_β chemical shifts using the random coil values of Tamiola *et al.* and were removed for that reason.

The three helical elements towards the C-terminus are separated from each other by the polyproline stretch and flanking residues on the one hand (^{365}FPLALAGPPPPPPP-PHPHARIKLE^{388}), and by a glycine-rich sequence (^{414}GAGAAGPGSGSPS^{426}) on the other hand. Depending on the random coil shifts used, this glycine-rich region adopts low positive or negative ΔC_α values, which suggests it is a flexible element.

Interestingly, the regions of AF1* in which we observe high helical propensity correspond to functional transactivation units Tau-1 and Tau-5, crucial for AR transactivation (see Fig. 4.27). The main helical element between residues 176 and 202 contains the ^{179}LKDIL^{183} and ^{183}LSEASTMQLL^{192} motifs in the core Tau-1 region, and the three helical elements in the C-terminal part of AF1* (residues 353–364, 391–413 and 427–436) are located in Tau-5.

In figure 4.28 the sequence of the AF1* construct is represented with an overview of the regions identified to have secondary structure propensity.

To estimate the degree of helical propensity in those regions, we normalized the observed C_α chemical shifts using empirically determined C_α chemical shifts expected for the corresponding residue type in a random coil (0% helical) and in a fully formed helix (100% helical). The helical propensity can be calculated for each amino acid and averaged over the residues in an identified helical element. This gives an indication of

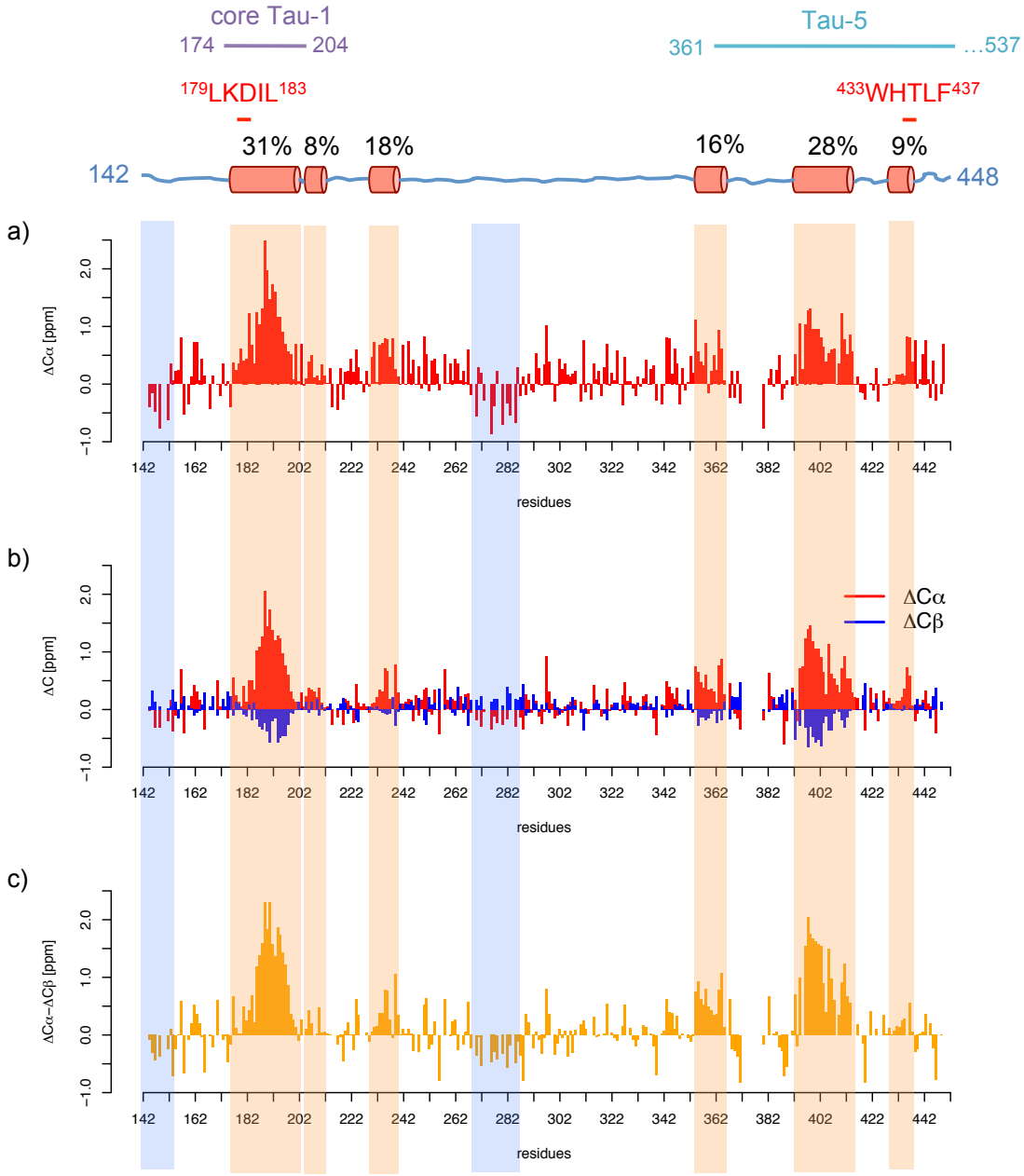


Figure 4.27: a) Secondary C_α chemical shifts of AF1* calculated with the random coil values from Wang & Jardetzky [413], b) Secondary C_α (red) and C_β (blue) chemical shifts of AF1* calculated with the sequence-corrected random coil values from Tamiola *et al.* [416], c) Difference between secondary C_α and C_β chemical shifts of AF1* calculated with the sequence-corrected random coil values from Tamiola *et al.* [416]. Regions with α-helical propensity are shaded in orange and regions with some tendency to adopt an extended conformation other than β-strand are shaded in blue. The estimated degree of helical propensity based on the tabulated shifts of Wang & Jardetzky is indicated for each of the segments identified to possess helical propensity (top of figure). Functional units core Tau-1 and Tau-5, as well as linear motifs ^{179}LKDIL^{183} and ^{433}WHTLF^{437}, are indicated.

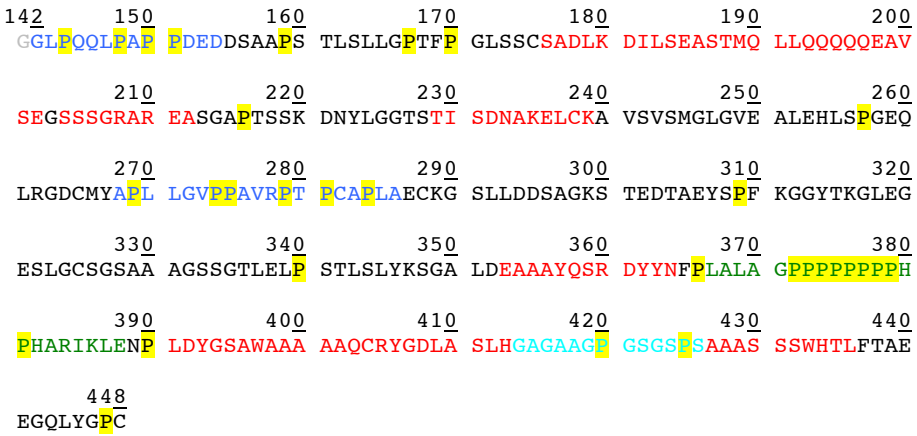


Figure 4.28: Sequence of the AF1* construct of AR with an indication of the secondary structure propensity identified by us in different regions. Regions with helical propensity are indicated in red, regions with tendency to form an extended conformation that is not β-strand are indicated in blue, the polyproline stretch and flanking residues are indicated in green and the glycine-rich stretch between the two C-terminal regions with helical propensity is indicated in cyan. Proline residues are marked in yellow. The non-native glycine residue at the N-terminus of the construct is shown in grey.

the helicity of that region. We used two available datasets containing both random coil shifts and chemical shifts in a fully helical conformation for this purpose: the one from Wang & Jardetzky [413] and the one from the Re-referenced Protein Chemical Shift Database (RefDB) [414]. RefDB is a database of carefully corrected or re-referenced chemical shifts, derived from the BMRB repository [414]. Consequenty, the chemical shifts from RefDB are based on a large number of properly referenced chemical shifts from proteins with known structures, whereas the values from Wang & Jardetzky are based on a smaller number of shifts [413]. Nevertheless, values from both datasets gave comparable results (see Fig. 4.29). The obtained helical content was averaged over a window of five residues for both plots. This was necessary because a representation of the helicity for each of the individual residues was too noisy. This would most likely be less pronounced if random coil chemical shifts from AF1* under denaturing conditions could be used. The helical propensities calculated for each of the helical elements individually are shown in figure 4.27. It should be noted that negative values for helical propensity obtained through this analysis do not indicate helical propensity, but instead correspond to regions with propensity to form extended conformations like β-strand or polyproline II conformations.

An alternative way to represent the secondary structure profile of a protein is by computing the secondary structure propensity (SSP) value for each residue [422]. This method combines chemical shifts from different nuclei into a single SSP score represent-

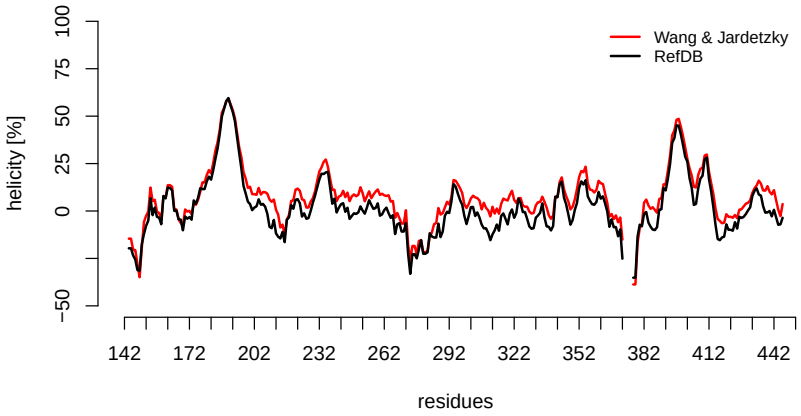


Figure 4.29: Helical content of AF1* according to the experimental $^{13}C_\alpha$ chemical shifts and calculated using random coil and fully helical chemical shifts of Wang & Jardetzky [413] (red) or using random coil and fully helical chemical shifts of RefDB [414] (black). A five-residue window was used in both cases.

ing the expected fraction of α or β secondary structure at a given residue [422]. The contributions of different chemical shifts are weighted by their sensitivity to secondary structure. Positive values represent α-helical structure propensity and negative values represent β-strand structure propensity. An SSP score at a given residue of 1 or -1 reflects fully formed α- or β-structure, respectively, while a score of 0.5 indicates that 50% of the conformers in the disordered state ensemble are helical at that position.

SSP scores for AF1* were calculated using $^{13}C_\alpha$ and $^{13}C_\beta$ chemical shifts, as recommended for IDPs, and with a five-residue weighted averaging. As shown in figure 4.30 the secondary structure profile obtained from SSP scores is very similar to the helical content based on the $^{13}C_\alpha$ chemical shifts.

Taken together, AF1* displays a dynamic equilibrium between a completely unfolded state and different partially helical conformations with contributions of extended conformations other than β-strand in proline-rich regions. Interestingly, its helical propensity is higher than expected for an IDP, especially in those parts of the sequence that were previously identified to be important for function. We observed a high degree of transient helicity in the region corresponding to core Tau-1 (residues 176–202, containing ^{179}LKDIL^{183} and ^{183}LSEASTMQLL^{192}) and towards the C-terminal part of AF1*, between residues 353–364, 391–413 and 427–436, which are all found in transactivation unit Tau-5 (residues 361–537).

One possible reason underlying the observed high helical content in AF1* is the existence of putative long-range contacts, which could be important for function, within

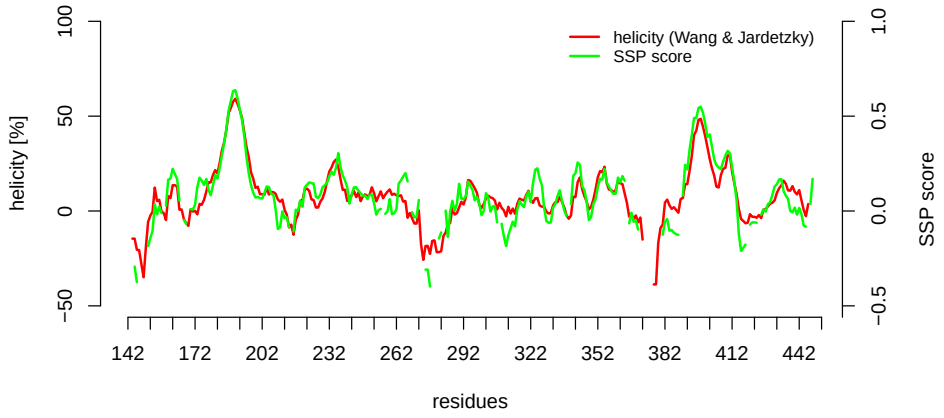


Figure 4.30: Helical content of AF1* according to the experimental C_α chemical shifts and calculated using random coil and fully helical chemical shifts of Wang & Jardetzky [413] (red) and the calculated SSP scores for AF1* (green). Both are averaged over a five-residue window.

the domain. Noticeably, the extent of transient α-helicity observed in the protein is not reflected in the Agadir prediction, for which the temperature was set to 278 K, as shown in figure 4.31. While the regions with helical propensity are identified by the predictor, the experimentally observed helicity is much higher than the predicted one. Considering that Agadir does not take into account possible long-range interactions, because it is an algorithm to predict the helical content of peptides, this suggests that the high helicity observed in AF1* could be related to long-range contacts.

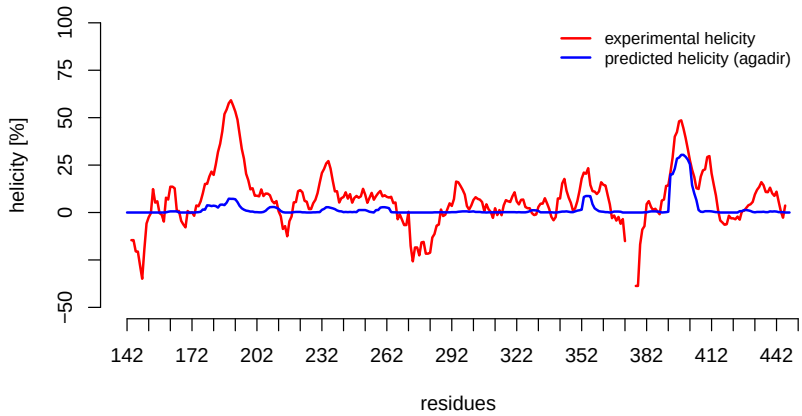


Figure 4.31: Experimentally determined helicity of AF1* based on C_α chemical shifts calculated using random coil and fully helical chemical shifts of Wang & Jardetzky [413] (red) compared to the predicted helicity by Agadir (blue).

4.6.2 Construct dependence of secondary chemical shifts

To investigate whether there are long-range contacts within AF1*, we compared the conformational properties of the smaller constructs individually and in the context of the longer AF1*. We performed the same secondary structure propensity analysis on the shorter protein constructs, AF1*b and AF1*c (see Figs. 4.32 and 4.33, respectively). Again, the ΔC_β chemical shifts for histidine residues where removed when using the random coil values from Tamiola *et al.*

In AF1*b residues 269–287 show negative ΔC_α and positive ΔC_β values, similar to the corresponding region in AF1*, indicating an extended conformation which is also reflected in the SSP scores. AF1*c transiently populates helical structure in the same regions as in AF1* (residues 353–364, 393–412 and 427–435).

To compare the structural propensities of the different constructs, we superimposed the ΔC_α chemical shifts, reliable indicators of the secondary structure propensity, of the two fragments, AF1*b and AF1*c, with the ΔC_α values observed in AF1* (Fig. 4.34). The difference between C_α chemical shifts in AF1* and AF1*b and that in AF1* and AF1*c is given in the upper panel of figure 4.35.

Putative long-range interactions within AF1* are expected to involve transient helical elements of the protein that come together in space and stabilize each other. This would result in an increased helicity when the long-range contact is present. Thus, the helicity of regions that engage in this type of long-range interactions is expected to be higher in the context of AF1* than in the short construct on its own.

As can been seen in figures 4.34 and 4.35, we obtained comparable values for ΔC_α chemical shifts in the short constructs and in the longer AF1*. The ΔC_α values of the different constructs differ mainly between -0.1 ppm and +0.1 ppm, with more pronounced differences at the termini of the constructs as expected due to the "boundary conditions", similar to our observations for ^1H and ^{15}N chemical shifts (see Fig. 4.24). The resolution after application of linear prediction and zero filling for the $^{13}C_\alpha$ chemical shifts was 0.094 ppm. Consequently, the largest chemical shift differences observed are close to the detection level. A difference in ΔC_α values between -0.1 ppm and +0.1 ppm corresponds, respectively, to a decrease or increase in helicity of ca. 4% for the protein sequence under study, as shown in the lower panel of figure 4.35. The difference in helicity between the constructs shown in this figure is calculated with the random coil and fully helical chemical shifts from Wang & Jardetzky [413] and plotted per residue. Residues 330–340 seem to be slightly more helical in the context of AF1* as opposed to in the smaller constructs AF1*b and AF1*c. This is in agreement with the expected higher flexibility of terminal residues (residues 330–340 are located at the

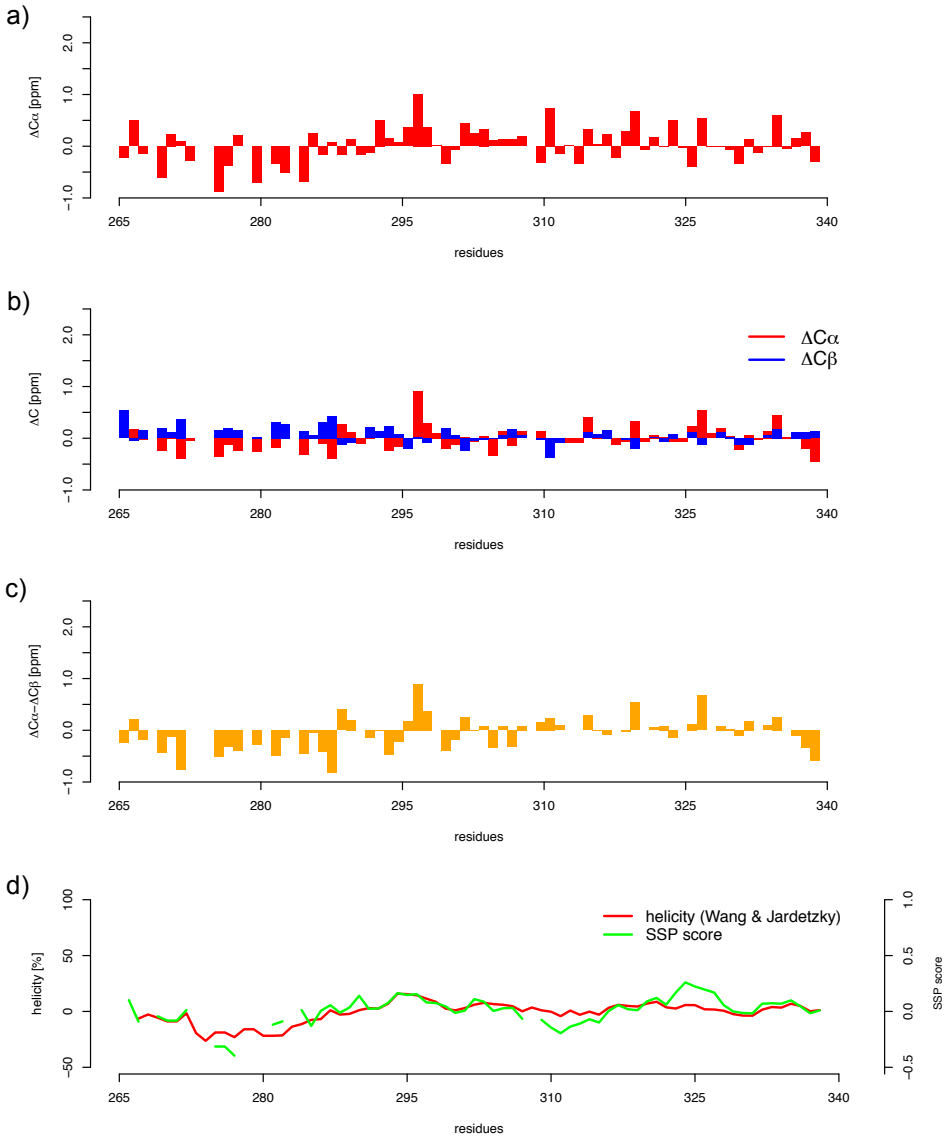


Figure 4.32: a) Secondary C_α chemical shifts of AF1*b calculated with the random coil values from Wang & Jardetzky [413], b) Secondary C_α (red) and C_β (blue) chemical shifts of AF1*b calculated with the sequence-corrected random coil values from Tamiola *et al.* [416], c) Difference between secondary C_α and C_β chemical shifts of AF1*b calculated with the sequence-corrected random coil values from Tamiola *et al.* [416], d) Helical content of AF1*b according to the experimental C_α chemical shifts and calculated using random coil and fully helical chemical shifts of Wang & Jardetzky [413] and with a five-residue window (red) overlaid with SSP scores for AF1*b [422] also averaged over five residues (green).

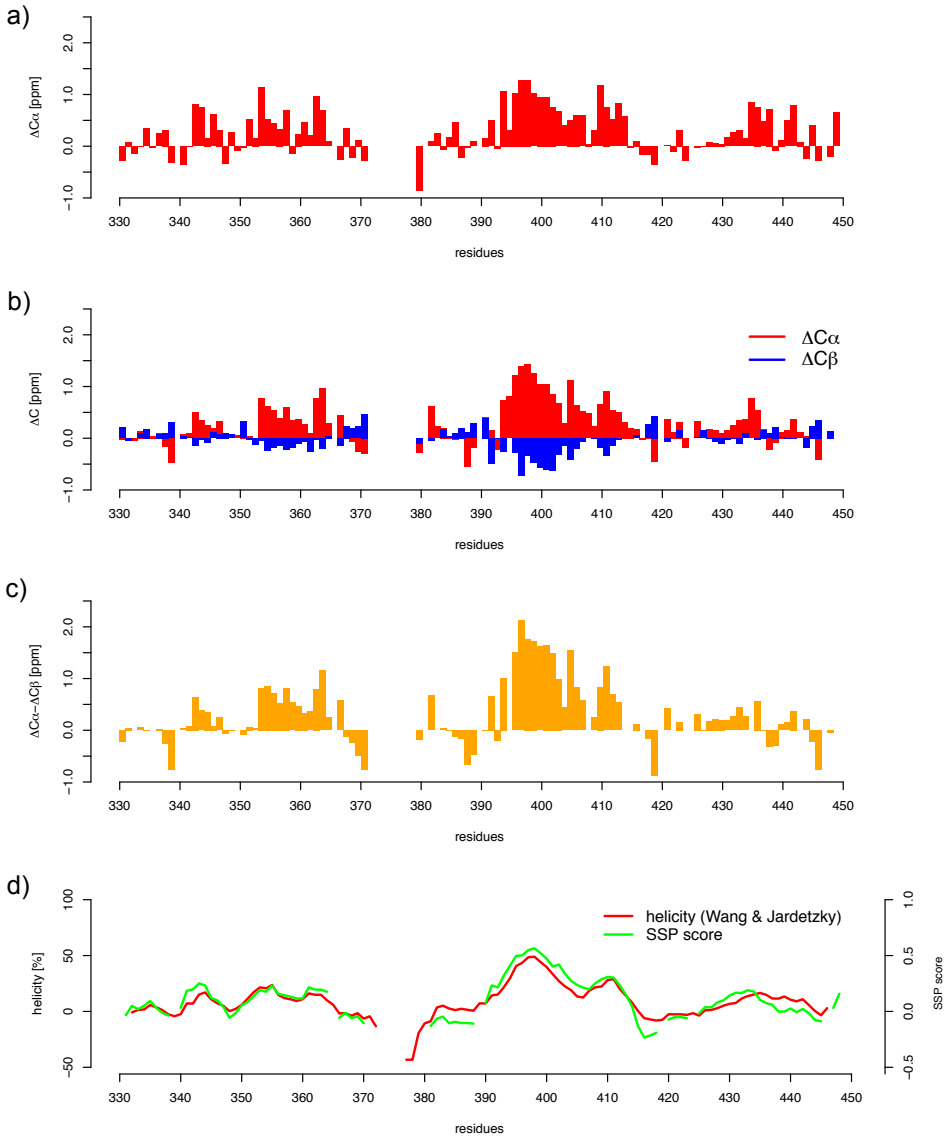


Figure 4.33: a) Secondary C_α chemical shifts of AF1*c calculated with the random coil values from Wang & Jardetzky [413], b) Secondary C_α (red) and C_β (blue) chemical shifts of AF1*c calculated with the sequence-corrected random coil values from Tamiola *et al.* [416], c) Difference between secondary C_α and C_β chemical shifts of AF1*c calculated with the sequence-corrected random coil values from Tamiola *et al.* [416], d) Helical content of AF1*c according to the experimental C_α chemical shifts and calculated using random coil and fully helical chemical shifts of Wang & Jardetzky [413] and with a five-residue window (red) overlaid with SSP scores for AF1*c [422] also averaged over five residues (green).

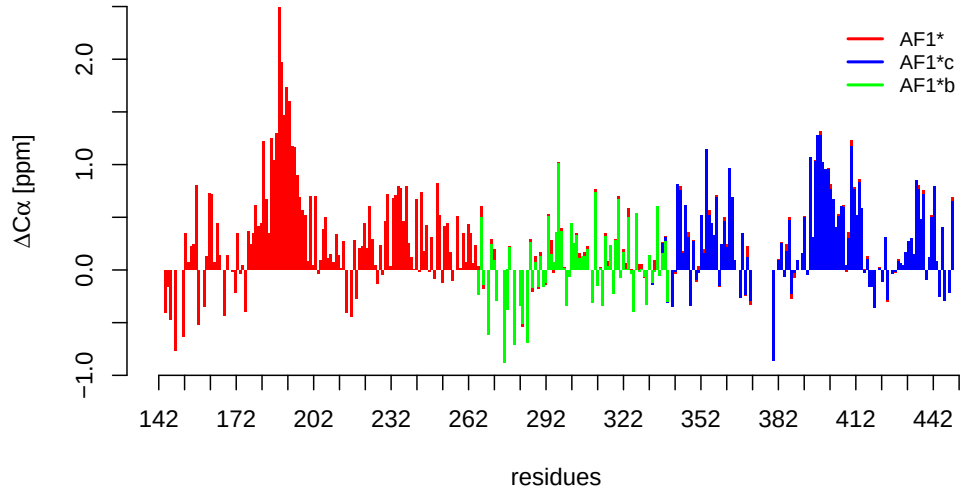


Figure 4.34: Superposition of the ΔC_α chemical shifts observed in AF1* (red) and in shorter constructs AF1*b (green) and AF1*c (blue) (resolution after application of linear prediction and zero filling for $^{13}C_\alpha$=0.094 ppm). The random coil chemical shifts were those of Wang & Jardetzky [413].

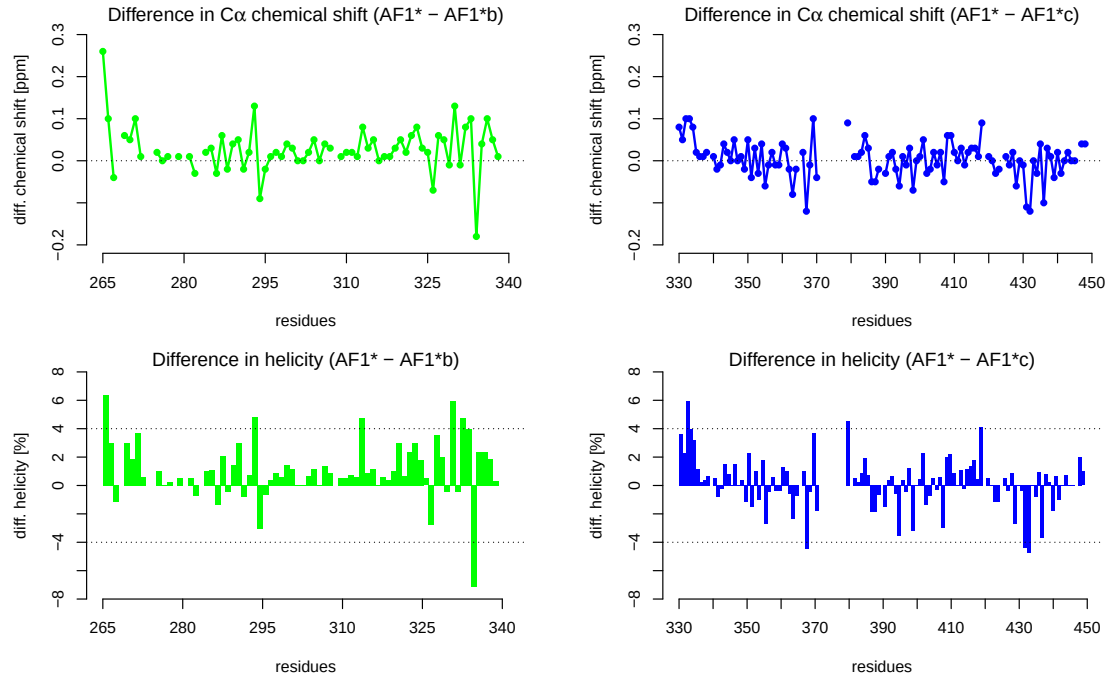


Figure 4.35: The difference in ΔC_α chemical shift between AF1* and AF1*b (top left) and between AF1* and AF1*c (top right) (resolution after application of linear prediction and zero filling for $^{13}C_\alpha$=0.094 ppm). The corresponding difference in helicity is shown in the lower part of the figure. Difference in helicity between AF1* and AF1*b (bottom left) and between AF1* and AF1*c (bottom right), plotted per residue. The helicity in the bottom panel was calculated with the random coil and fully helical chemical shifts from Wang & Jardetzky [413].

C-terminus of AF1*b and at the N-terminus of AF1*c, whereas they are not terminal in the AF1* construct). No further clear contiguous sets of residues for which the difference in helicity is around 4% were visible. The helical propensity for each of the residues of the AF1* construct is shown in figure 4.36. Given that the difference in helicity between the shorter and longer constructs on a residue basis is not larger than ±4% for the vast majority of the residues, these changes are very small compared to the helicity of the AF1* construct that contains residues with helical propensity up to 60%.

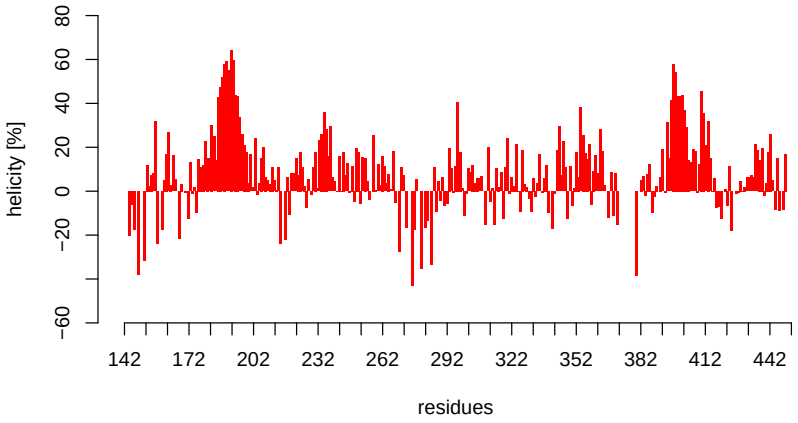


Figure 4.36: Helicity of AF1* at 340 μM, plotted per residue. The helicity was calculated with the random coil and fully helical chemical shifts from Wang & Jardetzky [413].

Based on these observations we concluded that long-range interactions involving helix stabilization do not occur between the three different segments within AF1*, i.e. between AF1*a, AF1*b and AF1*c. However, interactions that do not influence secondary structure cannot be excluded. This type of interaction is, however, unlikely to occur. In addition, it is possible that such long-range helix to helix interactions occur within each of the segments, i.e. within AF1*a, within AF1*b and/or within AF1*c.

Possibly such intrasegment helix to helix interactions occur between the different helical elements in AF1*c (see Fig. 4.37), which could contribute to the observed more collapsed conformation of AF1*c compared to the rest of AF1*. The presence of a polyproline stretch (residues 372–379) between the most N-terminal helical element (residues 353–364, indicated as region 1) and the two more C-terminal helical elements (residues 391–413 and 427–436, indicated as region 2 and region 3, respectively) most likely hinders long-range interactions between regions 1 and 2. This is because polyproline stretches are considered to be relatively stiff units, although kinks can arise from prolines that adopt a *cis*-conformation [423]. Consequently, it is unlikely that the three helical elements in AF1*c would adopt a (loosely packed) three helix-bundle confor-

mation. Nevertheless, relatively collapsed conformations of the AF1*c region of AF1* that involve contacts between regions 2 and 3, or between regions 1 and 3 are likely to exist (see Fig. 4.37). Mutation of R405 (located in region 2) to a serine residue and subsequent phosphorylation has been reported to have an effect on the interaction of the ^{433}WHTLF^{437} motif (located in region 3) with co-activator p300 [246], which further supports the existence of long-range interactions between regions 2 and 3.

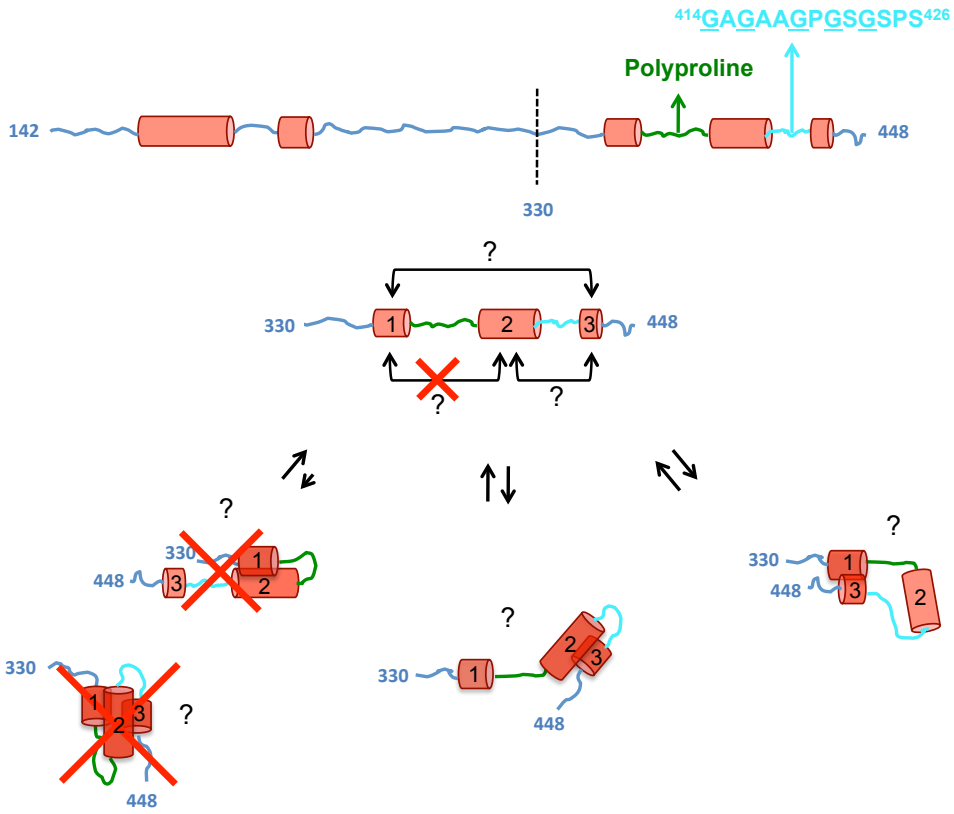


Figure 4.37: Schematic representation of possible intrasegment long-range interactions within the AF1*c region. Stretches with helical propensity in AF1* are indicated by the numbers 1, 2 and 3. Region 1 corresponds to residues 353–364, region 2 corresponds to residues 391–413 and region 3 corresponds to residues 427–436. The polyproline stretch between regions 1 and 2 (indicated in green) is considered to be relatively stiff, whereas the sequence between regions 2 and 3 (indicated in cyan) is rich in glycines and suggested to be a flexible region based on our data.

Based on our findings, we conclude that the high degree of transient helicity observed in AF1* (especially in regions 176–202 and 391–413) is not due to long-range contacts between the different segments of AF1*. It could alternatively be related to helix to helix interactions within the segments of AF1* and/or to other factors that will be discussed later in this chapter.

4.6.3 High helicity in Tau-1

The helicity of the core Tau-1 region is not affected by long-range interactions with regions of the AF1* located outside of the AF1*a region as shown in the previous section, so long-range interactions with residues outside of AF1*a cannot account for the observed high helicity. Since residues 171 to 215 are predicted to have coiled-coil propensity (see section 4.1), we investigated whether core Tau-1 could engage in a coiled-coil interaction.

For this purpose, we recorded an HNCA spectrum of AF1* at 25 μM and compared the C_α shifts obtained from this experiment with those obtained at 390 μM (see Fig. 4.38). If an intermolecular coiled-coil interaction occurs, the helical content of the core Tau-1 region is expected to increase at higher concentration. Interestingly, we found that the ΔC_α shifts at 25 μM were almost identical to those at 390 μM, indicating that the high helicity in core Tau-1 of AF1* is intrinsic to the protein, and not necessarily linked to an intermolecular coiled-coil conformation. Also for these experiments the resolution after application of linear prediction and zero filling for the $^{13}C_\alpha$ chemical shifts was 0.094 ppm. Consequently, the largest chemical shift differences observed are close to the detection level (see Fig. 4.38b). The difference in helical propensity between the AF1* at low and high concentration was calculated with the random coil and fully helical chemical shifts from Wang & Jardetzky [413] (see Fig. 4.38c). From this analysis it is clear that the helicity increases at most 4% when the AF1* concentration is changed from 25 μM to 390 μM. Compared to the helical propensity of AF1* at 25 μM (shown in Fig. 4.38d), this increase in helicity is small but distinguishable. Interestingly, only a small increase in helicity was observed in the core Tau-1 region (between residues 179 and 202), compared to other regions identified to have α-helical propensity, i.e. residues 231–240, residues 393–413 and residues 427–439. This result will be further discussed under section 4.8.1.

Alternatively, the core Tau-1 region could form an intermolecular coiled-coil at concentrations as low as 25 μM, but that would imply that the protein is dimeric at this concentration and that the coiled-coil is very stable. This is not compatible with the remainder of the data. SEC indicates that AF1* exists as a monomer in an extended conformation at concentrations higher than 25 μM. In addition, even if coiled-coil formation would only involve the core Tau-1 region and the rest of the protein would remain flexible and not dimerize, this would be clearly visible in the NMR spectra, as in this case the residues of the core Tau-1 region should have broader peaks with higher dispersion in ^{1}H than what we observe in the AF1* spectra. Furthermore, it is unlikely an intramolecular coiled-coil is formed, because this would imply a kink in the helix centered around the core Tau-1 region and this is not observed. Our data therefore

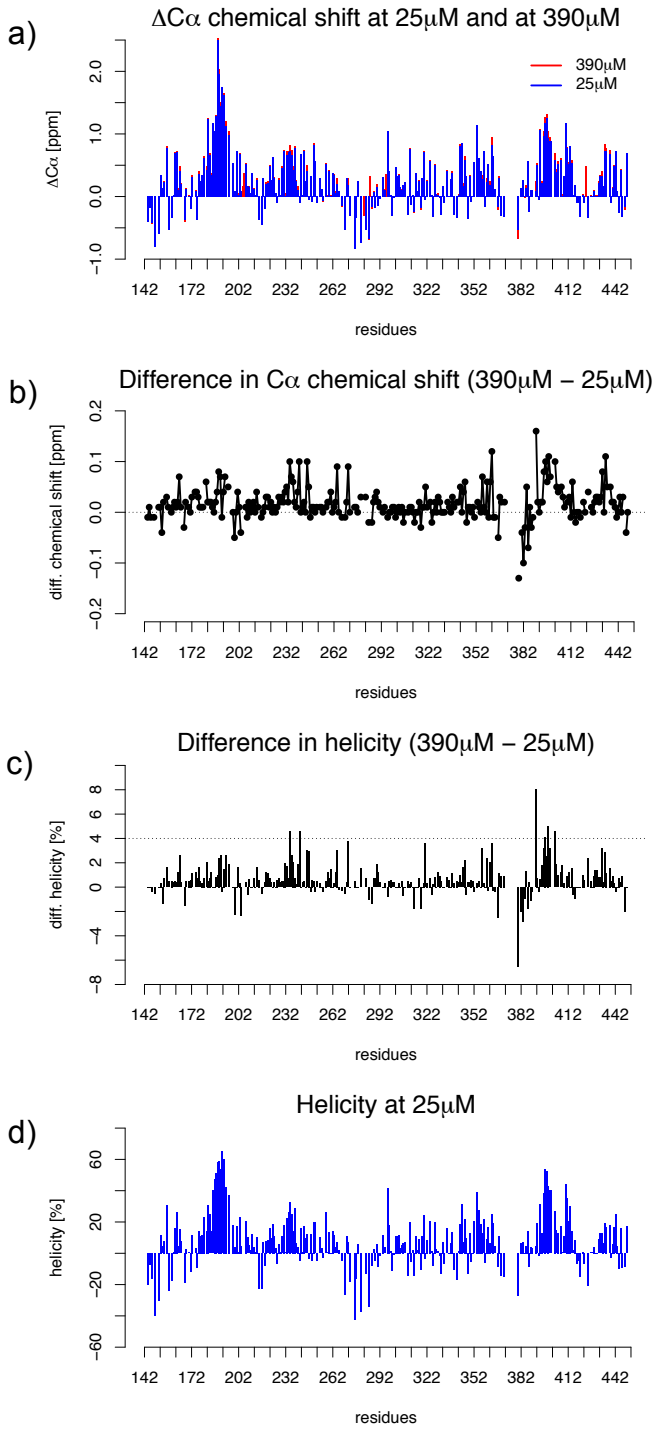


Figure 4.38: a) Comparison of the secondary C_α chemical shifts of AF1* at 25 μM (blue) and 390 μM (red) (resolution after application of linear prediction and zero filling for $^{13}C_\alpha$=0.094 ppm). Random coil values of Wang & Jardetzky were used. b) Difference in C_α chemical shifts between AF1* at 25 μM and 390 μM (resolution after application of linear prediction and zero filling for $^{13}C_\alpha$=0.094 ppm). c) Difference in helicity between 25 μM and 390 μM, plotted per residue. d) Helicity of AF1* at 25 μM, plotted per residue. The helicity in c) and d) was calculated with the random coil and fully helical chemical shifts from Wang & Jardetzky [413].

suggest that the high helicity in the core Tau-1 region does not originate from coiled-coil formation, and neither from long-range contacts between the different segments of the protein, but is intrinsic to AF1*.

The core Tau-1 region may sample a nearly folded helical conformation many times to facilitate the interaction with (some of) its binding partners. This region is key for androgen-dependent transcriptional activity, but despite its prominent role in transactivation not many co-regulatory proteins that interact with this region have been identified [170]. The $^{183}LX_7LL^{192}$ motif, flanking the $^{179}LKDIL^{183}$ motif and part of core Tau-1, was identified as a binding site for TAB2 as a component of an NCoR co-repressor complex [242]. The core Tau-1 region was further found to be crucial for the interaction of AR and MED1, a subunit of the Mediator complex [241]. However, no structural information is available for these complexes.

4.7 Flexibility of the AF1* backbone

To probe the dynamics of AF1* we recorded ^{15}N transverse relaxation (T_2) experiments for this construct. The corresponding relaxation rates (R_2) report on time scales in which conformational transitions occur (μs to ms).

Generally ^{15}N longitudinal and transverse relaxation rates (R_1 and R_2, respectively) and heteronuclear NOE (hetNOE) of the backbone amides are measured to characterize the dynamical properties of a protein backbone. The ^{15}N R_2 relaxation rates were, however, found to be particularly sensitive to any deviation from the random coil [402, 424, 425] as they showed larger variations than the other two relaxation parameters. Furthermore, for several unfolded proteins, the observed variations in R_2 relaxation rates clustered in certain regions of the sequence while they seemed to be more randomly distributed in R_1 relaxation rates and hetNOE [402, 424, 425]. Therefore, the R_2 relaxation rates are particularly informative when characterizing the motional properties of a disordered polypeptide chain.

Flexible parts of a protein relax slower than more rigid parts of the sequence and are therefore characterized by lower R_2 relaxation rates. By contrast, high R_2 relaxation rates are indicative of secondary structural elements, tertiary contacts, local collapse or other phenomena that give the polypeptide chain a certain degree of rigidity. Chemical exchange can also contribute to transverse relaxation. We measured ^{15}N transverse relaxation rates (R_2) of AF1* at 140 μM, 278 K and pH 7.4 (see Fig. 4.39).

The dashed line in figure 4.39 is indicative of the ^{15}N transverse relaxation rates for denatured ubiquitin in 8 M urea, at pH 2.0 and 278 K, representative of a fully

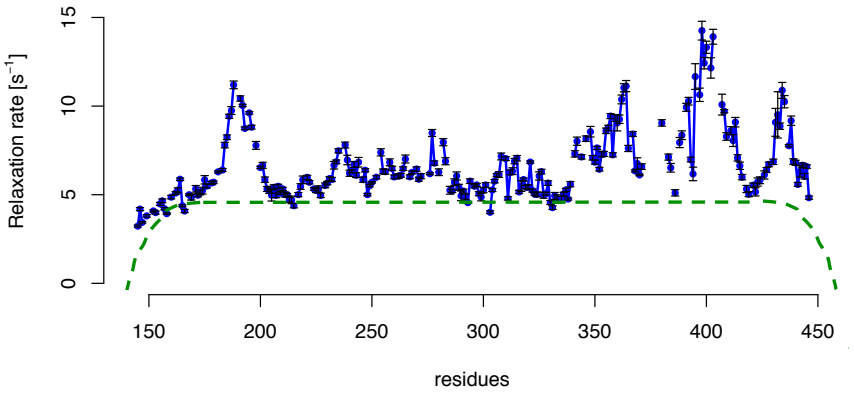


Figure 4.39: ^{15}N transverse relaxation rates of AF1* at 140 μM, 278 K and pH 7.4. The dashed green line is indicative of the ^{15}N transverse relaxation rates for denatured ubiquitin in 8 M urea, at pH 2.0 and 278 K, representative of a fully disordered polypeptide chain, and is used as a baseline [426].

disordered polypeptide chain, and is used as a baseline [426].

The ^{15}N R_2 relaxation rates for each amino acid indicate that AF1* is less flexible than expected for an IDP. Residues 171–211, 346–367, 391–413 and 430–440 show especially high transverse relaxation rates, pinpointing them as parts of the sequence that are less disordered. Indeed, those regions correspond to the parts of the sequence with higher helical propensity (see Fig. 4.40). Residues 276–283 also show higher relaxation rates and correspond to those residues in a proline-rich region adopting an extended conformation (268–286).

The high helical propensity in core Tau-1 and slightly lower relaxation rates could indicate fast forming and breaking of the helix, whereas the high relaxation rates in Tau-5 with slightly lower values of helical propensity could be related to additional sources of "rigidity", other than secondary structure elements, such as local collapse or tertiary contacts within AF1*c. The high relaxation rates at residues 430–440 are of special interest as the termini of a protein are usually more flexible, typically because they are not involved in the tertiary structure of a folded protein. Here, however, we observe remarkably high relaxation rates at the C-terminus of AF1*, which could indicate this region forms a tertiary contact with a different part of the sequence, in agreement with the proposed interactions between the helical elements in AF1*c (see Fig. 4.37). In addition, a relatively high degree of compactness in AF1*c is in agreement with its hydrodynamic radius obtained from size exclusion chromatography (see section 4.4.2) indicating a partially collapsed premolten globular conformation for this part of the protein. An additional source of higher relaxation rates in the C-terminal part

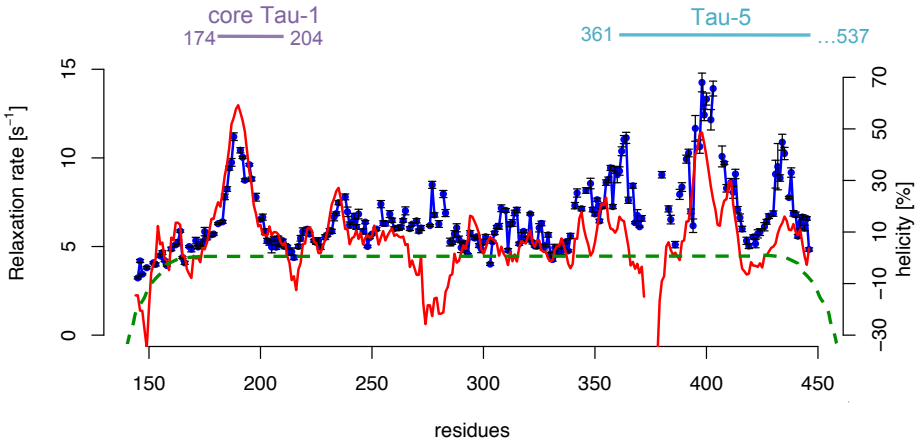


Figure 4.40: The helicity of AF1* as determined using the random coil and fully helical chemical shifts from Wang & Jardetzky in percentage (red) overlaid with the ^{15}N transverse relaxation rates of AF1* in s^{-1} (blue). The functional domains core Tau-1 and Tau-5 are indicated.

of AF1* could be chemical exchange. An equilibrium between a more flexible and a more collapsed conformation, would result in a chemical exchange contribution to the transverse relaxation rate, which will cause line broadening of the peaks and higher transverse relaxation rates for those residues.

Interestingly, we observe the highest degree of line broadening in the AF1*c part of the sequence. The line broadening observed in the residues of AF1* is correlated with the secondary structure propensity and relaxation profiles, confirming the wide dynamic range of the peak intensities is related to the conformational exchange in these different regions of the sequence (see Fig. 4.41).

Importantly, those regions with less backbone flexibility and higher structural propensity coincide with the core Tau-1 and Tau-5 regions, known to be important for AR function (see Fig. 4.40).

Taken together the data from the chemical shifts and the relaxation rates, two regions with higher order than expected for an IDP exist in AF1*, one at the N-terminal and one at the C-terminal part of the protein (residues 176–202 and residues 391–413, respectively). The data indicate, however, that no long-range contacts that involve induction of structure exist between those regions, so they can be regarded as independent units. Our findings suggest that the region corresponding to core Tau-1 populates a helical conformation a large fraction of the time. This helical element appears to be dynamic (fast breaking and forming of helix) and intrinsic to AF1* which suggests the core Tau-1 region pre-samples a helical conformation to facilitate its interaction with

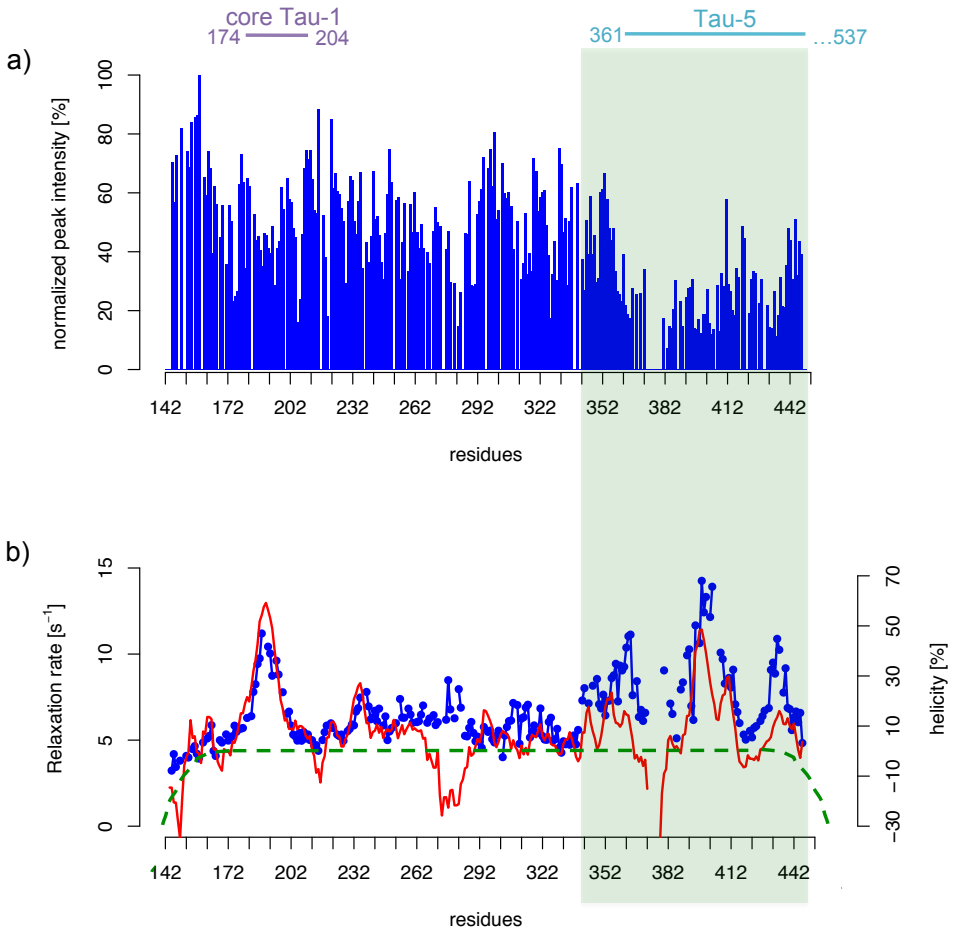


Figure 4.41: a) Normalized peak intensities for AF1* compared with b) the helical propensity of the protein based on its C_α chemical shifts and calculated with random coil and fully helical shifts of Wang & Jardetzky (red) and the ^{15}N transverse relaxation rates in s^{-1} (blue). The region of sequence of AF1* that shows substantial line broadening corresponds to parts of the sequence with high helical propensity and high transverse relaxation rates, and is shaded in green.

biological binding partners. The three helical stretches in the Tau-5 region of AF1* seem to form a relatively collapsed conformation in which the three helical elements could be stabilizing each other through transient contacts (see Fig. 4.37). The conformational exchange arising from this equilibrium between more extended and more collapsed conformations is most likely reflected in the more pronounced line broadening and high relaxation rates in this part of the sequence.

4.8 Oligomerization of AF1*

As explained in the introduction, the AR can exist as a monomer and as a dimer depending on its biological context. AR dimerization is required for its transactivation activity [200] and it is known that the AR binds as a dimer to the DNA [210] where two copies of AREs are each recognized by a DNA-binding domain of AR. It is possible that the N-terminal transactivation domain is also involved in the dimerization, but due to its ID nature a possible role for the NTD in dimerization has remained elusive.

4.8.1 Oligomerization of AF1* and AF1*c studied by NMR

To study whether oligomerization of AF1* occurs, we measured its [^{1}H,^{15}N]-HSQC spectrum at different concentrations. We observed ^{1}H and ^{15}N chemical shifts changes and resonance broadening in the [^{1}H,^{15}N]-HSQC spectra of AF1* from 10 μM to 680 μM, which indicates there is concentration-dependent oligomerization over this range (see Figs. 4.42 and 4.43). Even though the chemical shift changes are clear, they are small for this concentration range. The largest chemical shift changes observed for proton are in the order of 0.03 ppm, whereas for nitrogen they are in the order of 0.25 ppm (resolution after application of linear prediction and zero filling for ^{1}H=0.002 ppm and for ^{15}N=0.011 ppm).

Because AF1* is a protein construct, it is possible that oligomerization of AF1* occurs to higher order oligomers than dimers, whereas in the biological context only dimerization of the NTD can occur as AREs are organized in pairs on DNA. Importantly, it is not possible to distinguish a monomer-dimer equilibrium from an equilibrium between monomer, dimer and higher oligomers of AF1* based on these NMR experiments.

The changes in chemical shift and line broadening with concentration of some residues involved in oligomerization is shown in figures 4.42 and 4.44. The ^{1}H and ^{15}N chemical shift changes are spread over the entire sequence but are more pronounced in

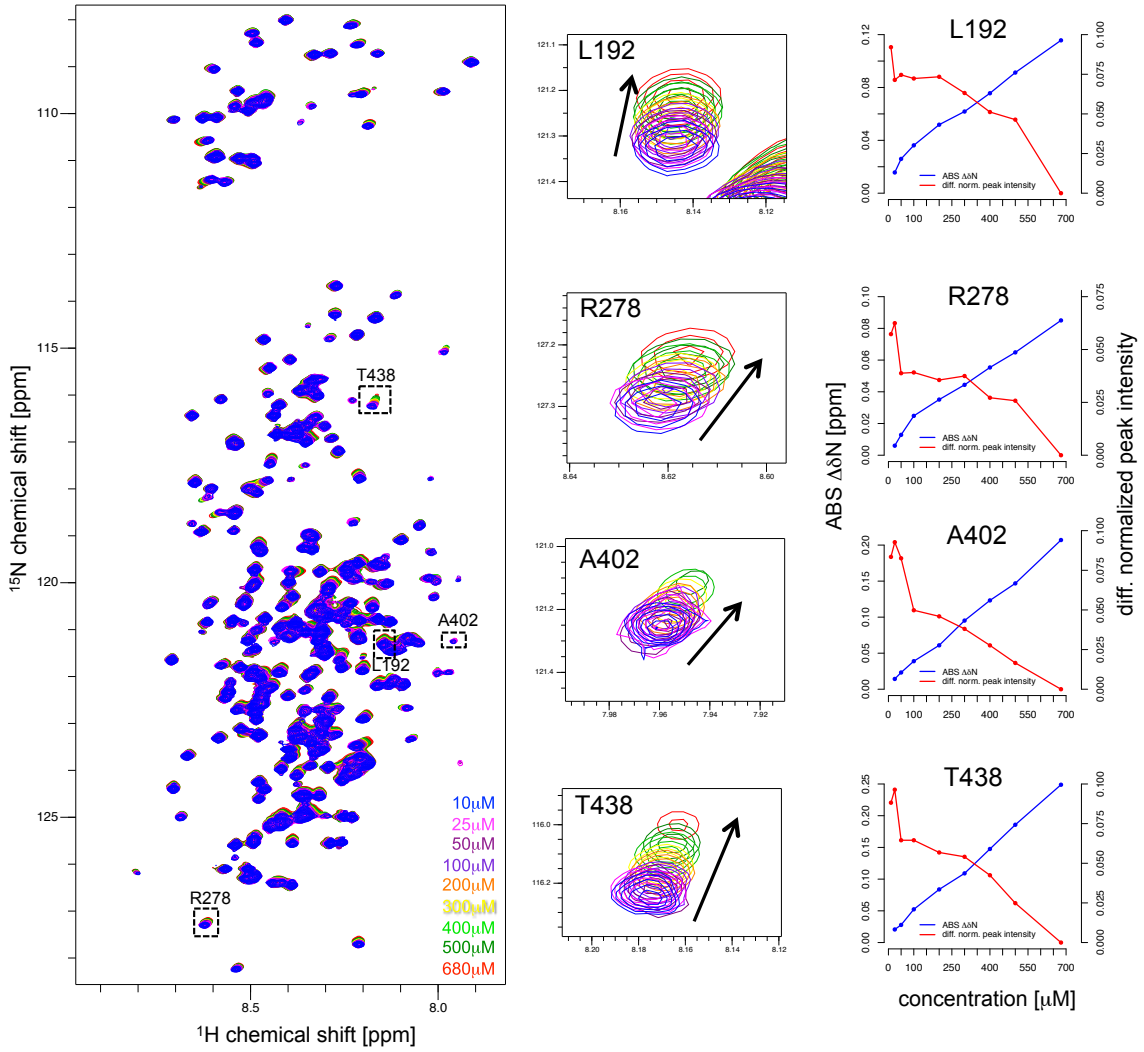


Figure 4.42: HSQC spectra of AF1* at different concentrations with a blow-up of residues L192, R278, A402 and T438 which are involved in dimerization. For these residues the change in ^{15}N chemical shift (blue) and line broadening (red) is also represented as a function of concentration.

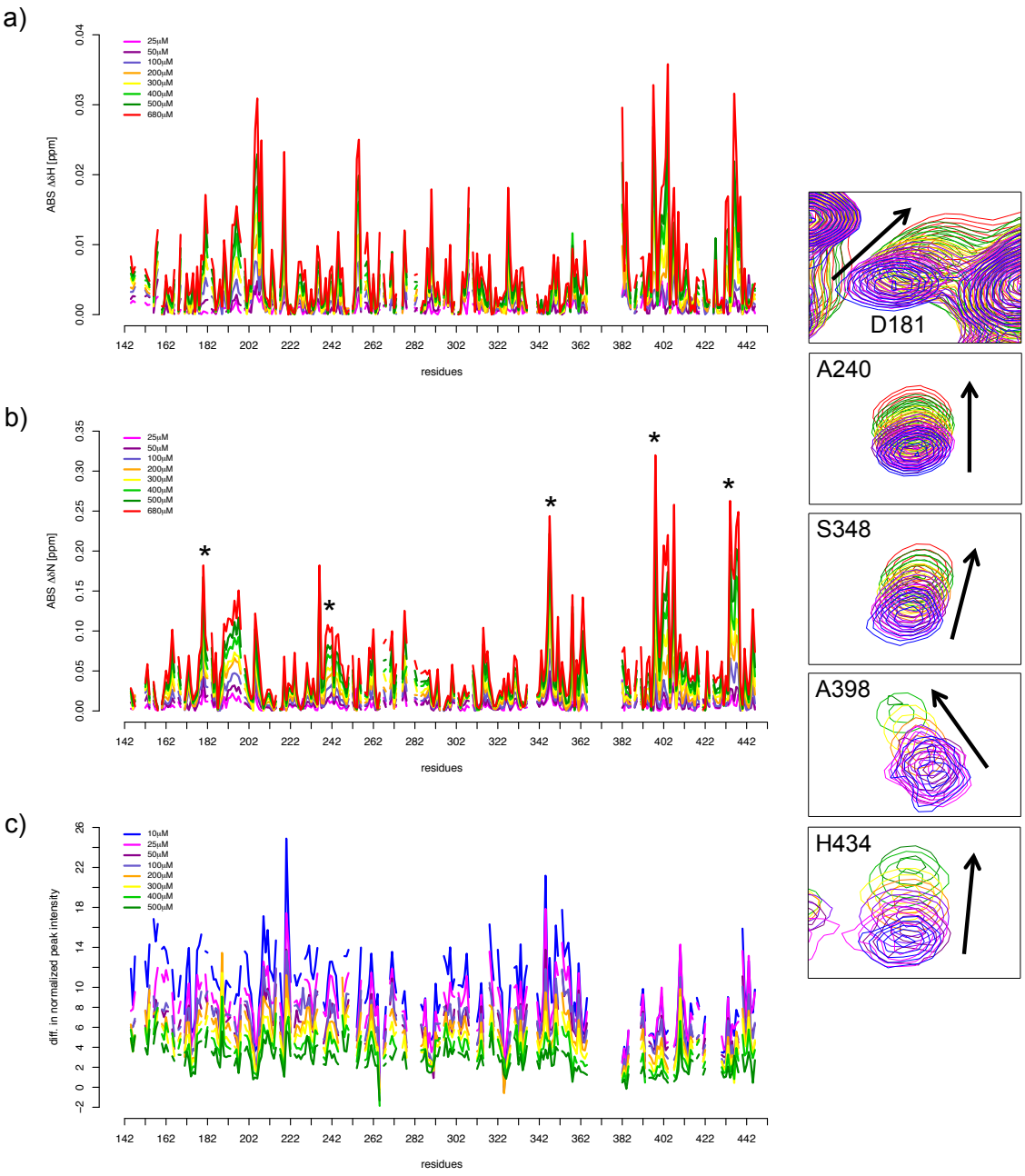


Figure 4.43: a) Absolute value of the difference in ^{1}H chemical shift for AF1* at different concentrations, relative to its chemical shifts at 10 μM. b) Absolute value of the difference in ^{15}N chemical shift for AF1* at different concentrations, relative to its chemical shifts at 10 μM. c) Difference in normalized peak intensities for AF1* at different concentrations, relative to the peak intensities at 680 μM. Residues 372–379 correspond to the polyproline stretch and consequently they do not have proton or nitrogen shifts. The resonances of residues D181, A240, S348, A398 and H434 are shown at the different concentrations. Their position is shown by a * symbol in b).

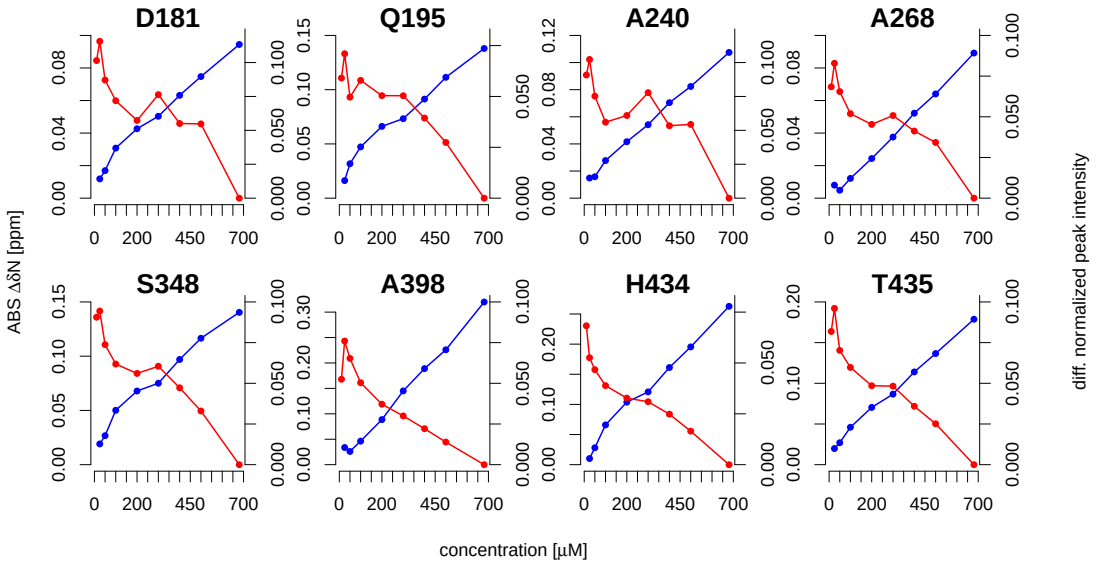


Figure 4.44: The difference in ^{15}N chemical shift (blue) and peak intensity (red) with concentration is shown for selected residues.

specific regions. Most amino acids of AF1* involved in oligomerization coincide with amino acids with a larger helical propensity and less flexibility as reflected in the ^{15}N R_2 relaxation rates (see Fig. 4.45). However, also regions with low helical propensity are involved in oligomerization, like residues 237–254 (LCKAVSVSMGLGVEALEH) and residues 264–291(YAPLLGVPPAVRPTPCAPLAECK). The former sequence is quite hydrophobic and overlaps with the binding site of AR for CHIP, the COOH terminus of the Hsp70-interacting protein, which has been mapped to ^{234}AKELCKAVSVSMGL^{247} [247]. The sequence between residues 264 and 291 on the other hand is rich in prolines. This suggests that these regions might self-interact during oligomerization, respectively, through hydrophobic interactions and adopting polyproline II conformations. It further suggests that oligomerization might affect the interaction with binding partners, such as CHIP. The core Tau-1 region also contributes to oligomerization (see Fig. 4.43), and is predicted to have some coiled-coil propensity (residues 171–215). It is therefore possible that oligomerization goes hand in hand with the formation of a coiled-coil between the core Tau-1 region of different monomers. Even though there are small but distinct changes with concentration of the ^{1}H and ^{15}N chemical shifts in core Tau-1 residues, the observed $^{13}C_\alpha$ chemical shifts between 25 μM and 390 μM are close to the detection limit in this region of sequence (see Fig. 4.38).

Importantly, the small $^{13}C_\alpha$ chemical shift changes that were detected when the AF1* concentration was changed from 25 μM to 390 μM, between residues 231–240, residues 393–413 and residues 427–439 (see Fig. 4.38c), are found in regions that in ad-

dition to having some helical propensity are involved in oligomerization (see Fig. 4.46). This suggests that regions with helical propensity in the monomer are stabilized by helix to helix interactions in the oligomer. These findings further suggest that such interactions occur first in the C-terminal part of AF1* and not in the core Tau-1 region. A coiled-coil interaction between the core Tau-1 regions of AR molecules could further stabilize such an oligomer.

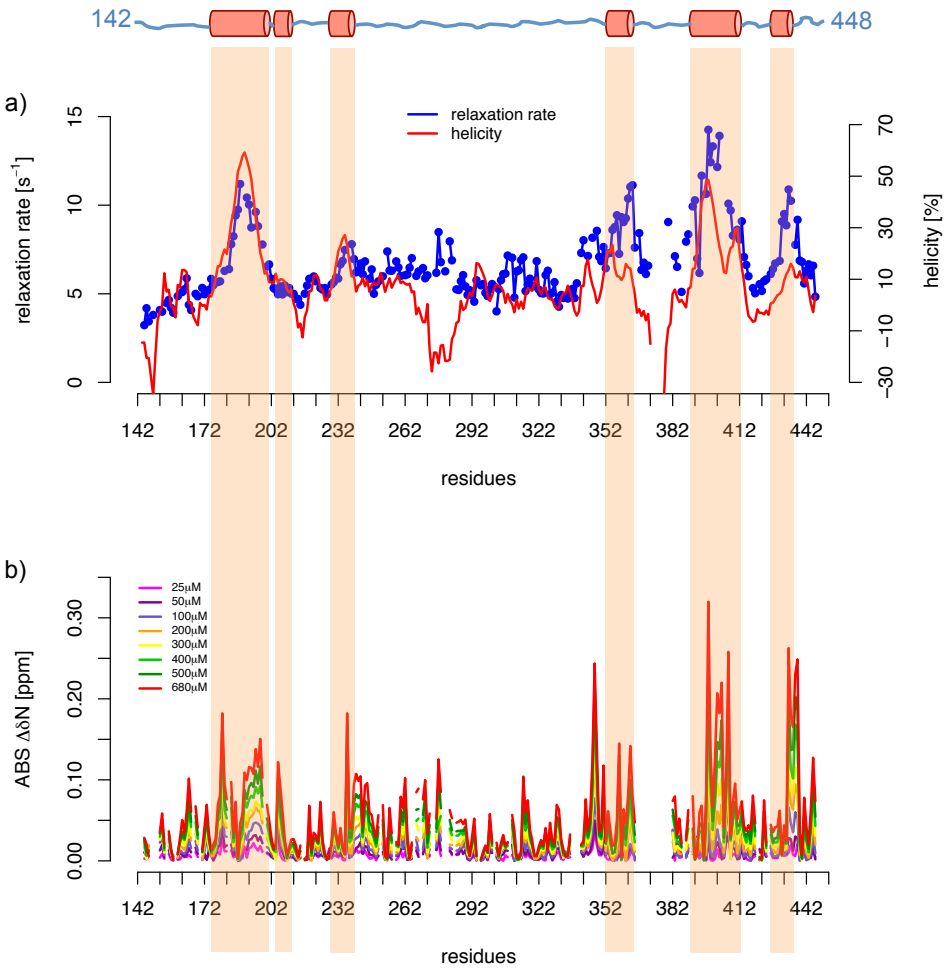


Figure 4.45: a) The helicity of AF1* as determined using the random coil and fully helical chemical shifts from Wang & Jardetzky in percentage (red) overlaid with the ^{15}N transverse relaxation rates of AF1* in s^{-1} (blue). b) Absolute value of the difference in ^{15}N chemical shift for AF1* at different concentrations, relative to its chemical shifts at 10 μM. As indicated, regions with higher helical propensity and higher relaxation rates correspond to regions involved in oligomerization.

Furthermore, the chemical shift changes with concentration in a specific region of residues are not similar over the entire stretch but often show maxima at i and i+4 positions, e.g. A398 (i) and A402 (i+4), and H434 (i) and T438 (i+4). This also

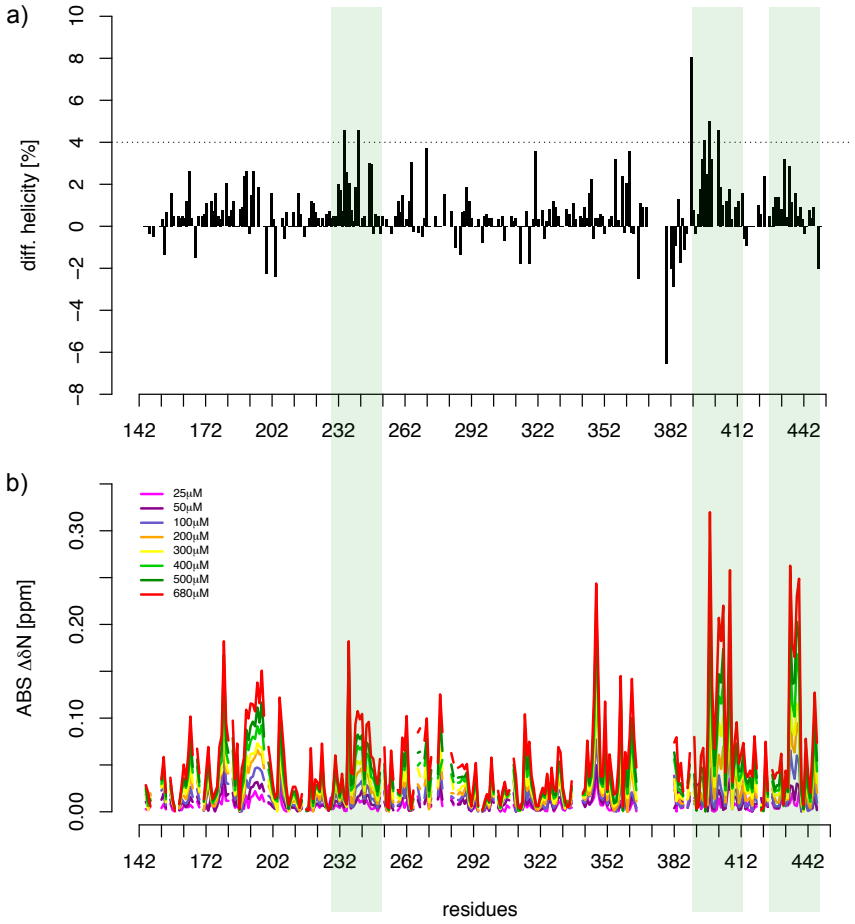


Figure 4.46: a) The increase in helicity of AF1* when the concentration is changed from 25 μM to 390 μM as determined using the random coil and fully helical chemical shifts from Wang & Jardetzky. b) Absolute value of the difference in ^{15}N chemical shift for AF1* at different concentrations, relative to its chemical shifts at 10 μM. As indicated, regions with increased helical propensity with concentration correspond to regions involved in oligomerization.

indicates that interaction between helical elements occurs in the oligomers. Residues corresponding to maxima in chemical shift changes are one helix turn away from each other and are likely to interact with the corresponding residues in an oligomer. Such an interaction would cause relatively larger chemical shift changes in the residues at the interface of the helices (here at positions i and i+4) and smaller, albeit distinct, changes in the remaining residues of the helices.

These observations suggest that oligomerization mainly involves folding of the regions with some helical propensity in the monomer into more stable helices. Additional hydrophobic or other contacts might help to further stabilize the oligomer. The fact that only very small $^{13}C_\alpha$ chemical shift changes are observed between 25 μM and 390 μM, could indicate there are no significant structural changes upon oligomerization, which is unlikely to be the case, or alternatively, that at the highest measured concentration, the population of oligomer is so small the $^{13}C_\alpha$ chemical shift changes are close to the detection limit. By contrast, the small changes in ^{1}H and ^{15}N chemical shifts at this concentration range could be reliably detected because the resolution for these nuclei in the HSQC experiment we conducted was higher than the one for the detection of $^{13}C_\alpha$ chemical shifts using a 3D HNCA experiment (resolution after application of linear prediction and zero filling for ^{1}H=0.002 ppm, resolution for ^{15}N=0.011 ppm and resolution for $^{13}C_\alpha$=0.094 ppm). The data therefore suggest that the AF1* oligomer is low populated and most likely only dimeric with no higher oligomeric species present at concentrations lower than 680 μM.

The residues within AF1* that undergo chemical shift changes with concentration also show line broadening related to the chemical exchange contribution due to the monomer-oligomer equilibrium. However, the extent to which the resonances broaden differs. Towards the C-terminal part of the sequence, the line broadening with concentration is more pronounced compared to the rest of the sequence. This could indicate induction of structure in AF1*c upon oligomerization. We know, however, that in the concentration range we spanned for these experiments the helical propensity of AF1* does not change notably with concentration (see Fig. 4.38). In addition, the signals from the two tryptophan side chains that are present in AF1*, and both in the region corresponding to AF1*c (W397 and W433), do not shift with concentration, indicating they do not undergo a transition between a more flexible conformation (more exposed to the solvent) and a more structured one (less solvent exposed) (see Fig. 4.47). Therefore, the observed additional line broadening in the AF1*c resonances cannot be accounted for by an increase in helicity upon oligomerization in the concentration range that we spanned. Alternatively, the line broadening due to the monomer-oligomer equilibrium could be larger in AF1*c than in the rest of the protein.

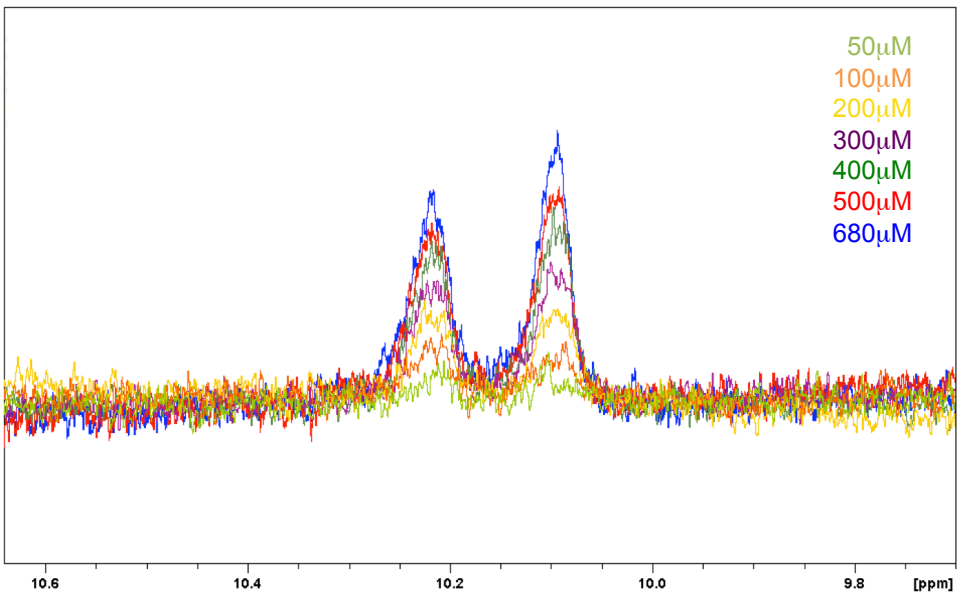


Figure 4.47: Signals of the side chains of W397 and W433 in 1D proton spectra at different concentrations of AF1*.

Equation 4.5 gives the excess transverse relaxation rate caused by the exchange of nuclei between different conformations or states with different chemical shifts, for the fast exchange regime [427]. The larger the excess transverse relaxation rate of a given residue, the broader the lines of the corresponding resonances are.

$$R_{ex,average} = \frac{p_A \cdot p_B \cdot \Delta\omega^2}{k_{ex}} \tag{4.5}$$

The excess exchange contribution depends on the population p_A and p_B of both states (here monomer and dimer), the difference in chemical shift between the two states, $\Delta\omega$, and the exchange rate, k_{ex}. Assuming that oligomerization involves all amino acids, the population of monomer and dimer are the same for each residue, as is the exchange rate k_{ex}. Under this assumption, the only parameter that can be different for different residues, is the chemical shift difference between the monomeric state and the oligomeric state. Note that the dependence of $R_{ex,average}$ on the difference in chemical shift $\Delta\omega$ is quadratic. Therefore, if a specific residue has a larger exchange contribution compared to another residue, this means that the difference in chemical shift between the monomer and oligomer is larger for that specific residue than for the other residues.

Assuming fast exchange for the oligomerization of AF1* (based on the NMR data), this indicates that the chemical shifts between the monomer and oligomer state are more distinct in AF1*c than in AF1*a or AF1*b. This implies that the change in chemical

environment due to oligomerization is larger for AF1*c than for the rest of the protein. This is compatible with a conformational rearrangement of AF1*c from a relatively collapsed conformation with long-range contacts between the different helical elements in monomeric AF1*c to a different arrangement of these helical elements in the context of the oligomer, without considerably changing the helical content of the domain. The part of AF1* that does not belong to AF1*c seems to undergo less conformational changes, in agreement with our hypothesis that helical elements stabilize each other in the context of the oligomer without the need of large conformational changes.

To better understand the molecular mechanism of the AF1* oligomerization, we further investigated whether a smaller protein construct, AF1*c, could oligomerize outside the context of AF1*. This is indeed the case. Our NMR analysis shows chemical shift changes and line broadening in the [^{1}H,^{15}N]-HSQC spectra of AF1*c from 10 μM to 465 μM (see Figs. 4.48 and 4.49).

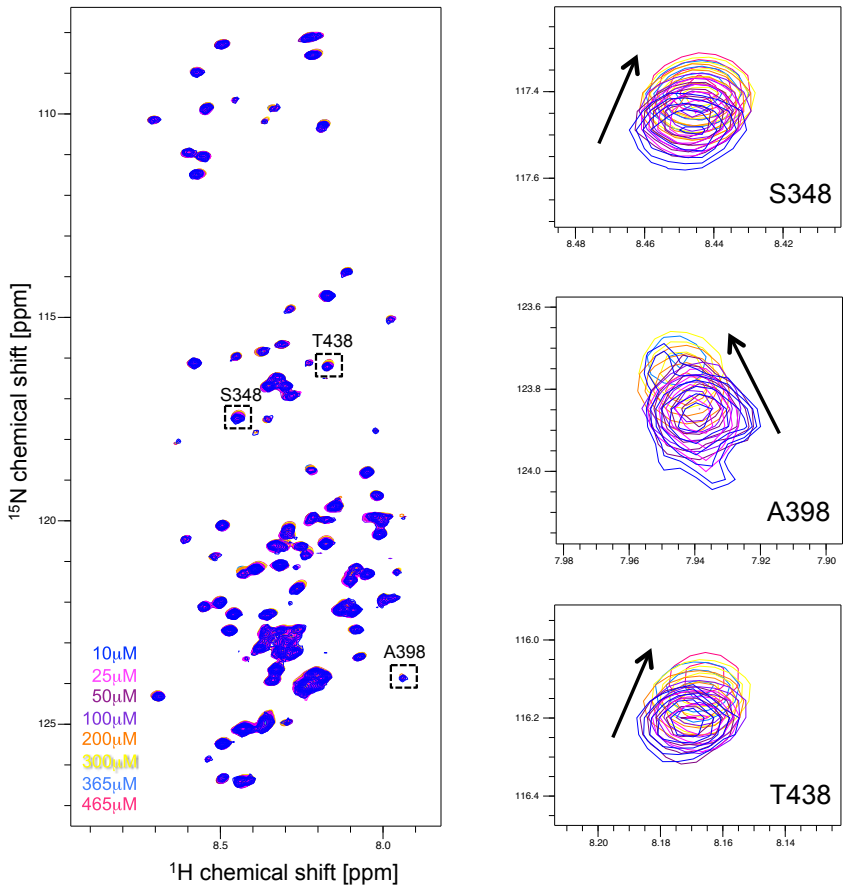


Figure 4.48: HSQC spectra of AF1*c at different concentrations with a blow-up of residues S348, A398 and T438 which are involved in dimerization.

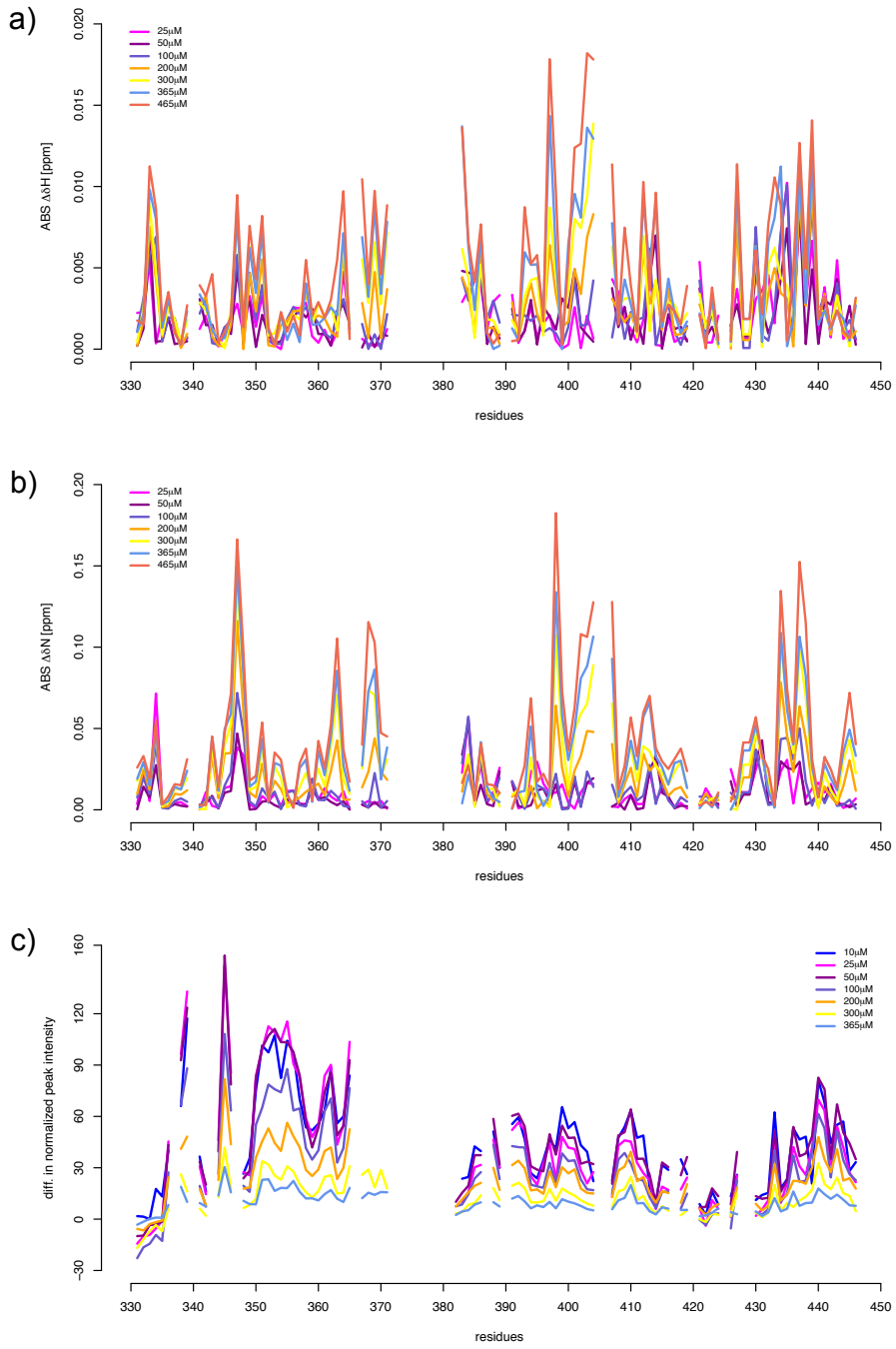


Figure 4.49: a) Absolute value of the difference in ^{1}H chemical shift for AF1*c at different concentrations, relative to its chemical shifts at 10 μM. b) Absolute value of the difference in ^{15}N chemical shift for AF1*c at different concentrations, relative to its chemical shifts at 10 μM. c) Difference in normalized peak intensities for AF1*c at different concentrations, relative to the peak intensities at 465 μM. Residues 372–379 correspond to the polyproline stretch and consequently do not have proton or nitrogen shifts.

When we compare now the chemical shift changes with concentration in AF1* and in AF1*c (see Fig 4.50), we can see that at the same concentration the changes in AF1*c are smaller than in AF1*, i.e. at the same concentration, AF1*c has formed less oligomer than AF1*. These observations suggest a cooperative model for oligomerization in which several stretches of the protein can self-assemble individually but are more prone to do so when other parts of the sequence have already oligomerized. Again, this is in agreement with our hypothesis that oligomerization involves interaction between regions with helical propensity in the monomer that are stabilized into more stable helices in the oligomer.

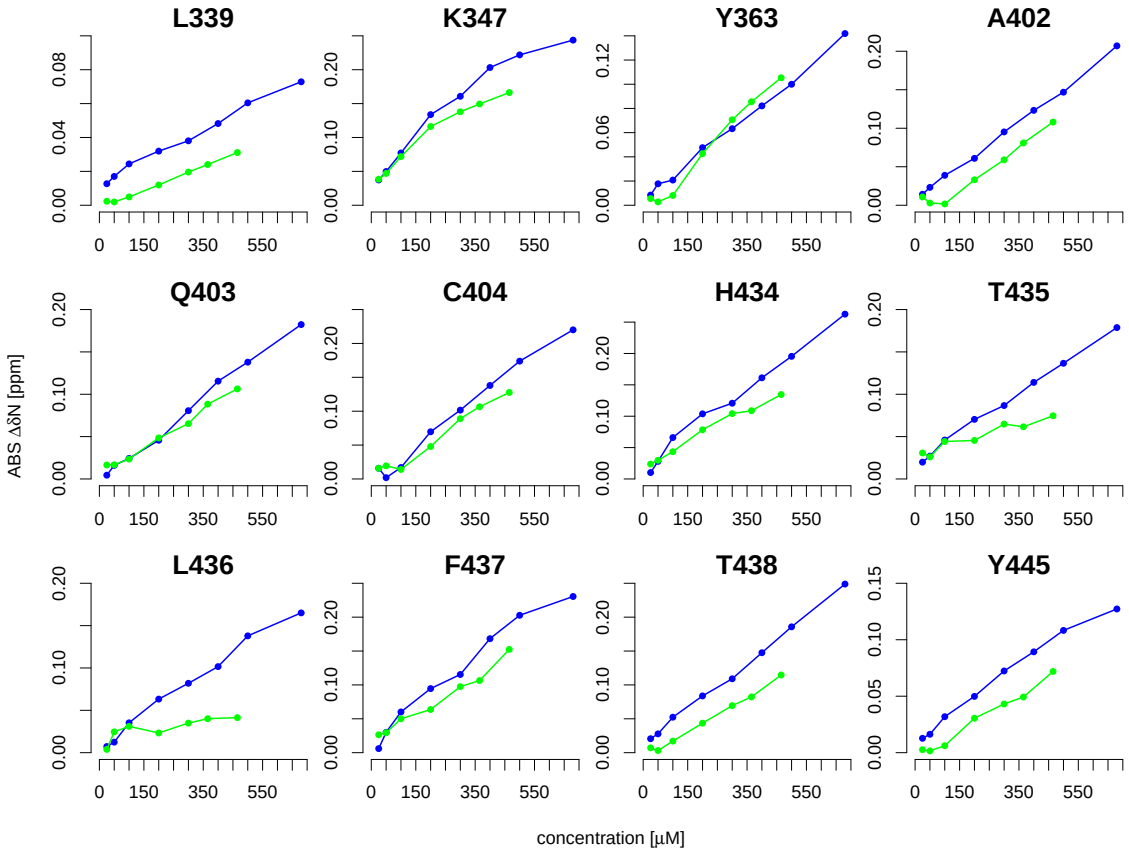


Figure 4.50: Absolute value of the difference in ^{15}N chemical shift with concentration, relative to 10 μM, for AF1* (blue) and AF1*c (green) for selected residues.

The binding constant for self-assembly could not be fitted from the data as neither for AF1* nor for AF1*c saturation was reached at the highest concentrations we could use without risking insolubility of the samples. This is most likely due to a low affinity of AF1* to interact with itself, compatible with a low population of oligomer, but it could also indicate that higher oligomers than dimer are formed of the AF1* construct and no saturation can be reached because AF1* molecules are incorporated in increas-

ingly higher order oligomers. Even though the NMR data seem to indicate the former scenario, we cannot exclude the second possibility based on only this data.

In order to determine whether higher oligomers of AF1* are populated or if only a very small population of dimer is present, and to quantify the populations of the different species, we investigated the oligomerization of AF1* with orthogonal techniques.

4.8.2 Other techniques to study oligomerization

4.8.2.1 Native gels

A simple experiment to detect whether oligomerization takes place and which oligomers are populated is running a native gel. On a native gel, no denaturants are added to the protein sample, in contrast to a typical SDS-PAGE gel for which the proteins are denatured both by the addition of SDS and by boiling the samples at high temperature before loading them into the gel. Therefore, on a native gel it is expected to see bands corresponding to the different oligomeric species if the protein oligomerizes, whereas on a denaturing gel only a band corresponding to the molecular weight of the monomer is expected (possible oligomers have been denatured to monomers).

We ran a native gel of 10% acrylamide for 20 μM AF1* (see Fig. 4.51). The gel showed that, at 20 μM, AF1* is mainly present as monomer and in addition a small population of dimer is present. Since only one band was detected at higher molecular weight, we concluded that no higher oligomeric species were present.

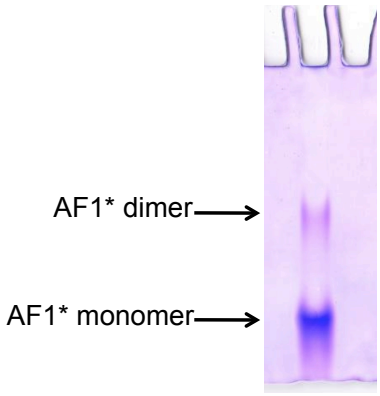


Figure 4.51: Native gel of 20 μM AF1*. The gel contained 10% acrylamide and the sample was run for 20.5 hours at 40 V at 4 °C.

4.8.2.2 Circular dichroism

Even though only very small changes in the structural properties of AF1* could be detected at low and high concentration based on the C_α chemical shifts, we were interested to see whether we could observe differences in the CD spectra of the protein at different concentrations. We studied both AF1* and AF1*c by CD at different concentrations. If oligomerization would induce structural elements that are not present in the monomer, CD spectra at different concentrations would reveal that.

For AF1*, we spanned a concentration range of 5 μM to 250 μM, using different cuvettes to adjust the path length according to the concentration. The CD spectrum at all concentrations is that characteristic of a disordered polypeptide chain (see Fig. 4.52). A slight deviation at 222 nm was observed, which could indicate a very small fraction of α-helical structure, but this is seen for all concentrations. This is compatible with the NMR data that indicate AF1* is a disordered protein that displays a dynamic equilibrium between a completely unfolded state and different partially helical conformations. Even though the helical propensity in core Tau-1 and in regions of Tau-5 is high, it is possible this is not reflected in the CD spectrum of the protein. Core Tau-1 is a stretch of ca. 30 residues (176–202) in a protein of 308 residues, i.e. less than 10%. Likewise, the regions in Tau-5 identified by NMR to display transient helicity are short compared to AF1*. It is likely that the small deviation around 222 nm is due to the residual helicity in core Tau-1 and the regions with helical propensity in Tau-5.

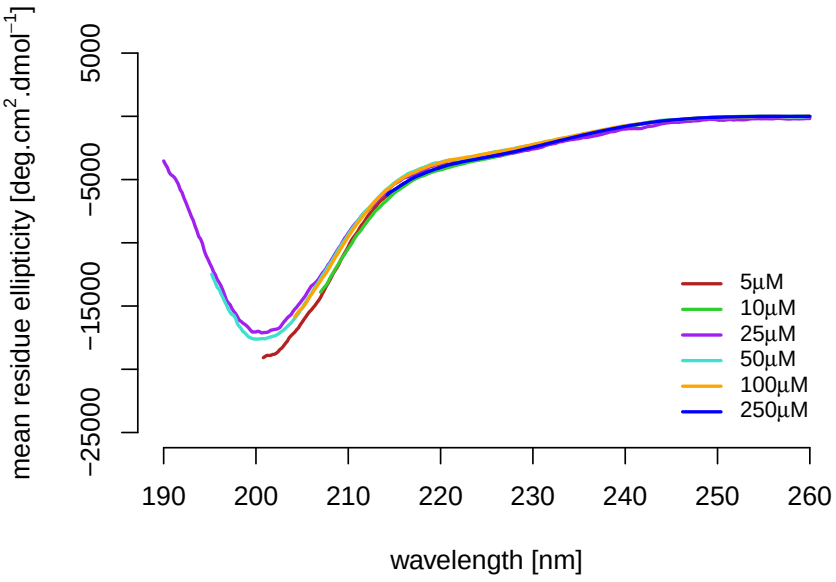


Figure 4.52: Far-UV CD spectra of AF1* at concentrations between 5 μM and 250 μM.

Similarly, the CD spectra of AF1*c measured for concentrations between 10 μM to 400 μM showed no structure induction over this concentration range (see Fig. 4.53).

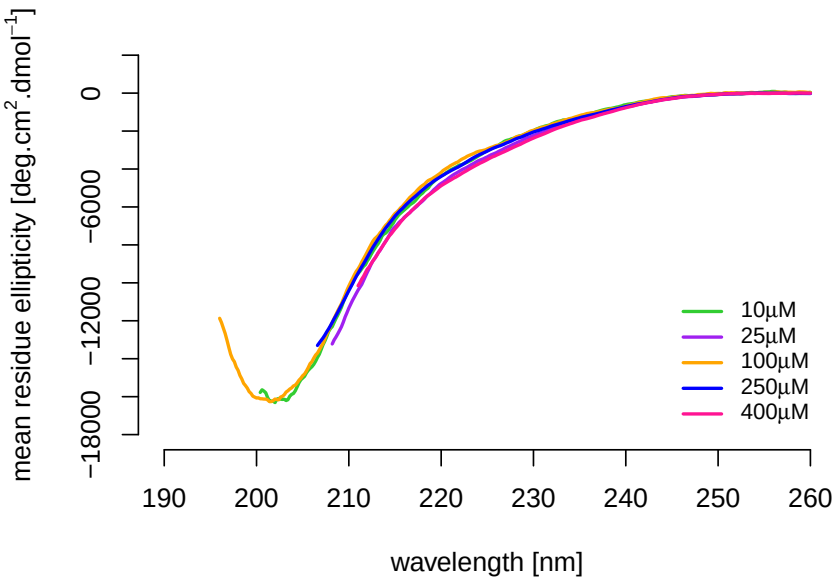


Figure 4.53: Far-UV CD spectra of AF1*c at concentrations between 10 μM and 400 μM.

To exclude the possibility we did not observe any structural changes because we monitored the entire protein, we also measured CD of a 21-residue peptide containing residues 426–446 (Ac-SAAASSSWHTLFTAEEGQLYG-NH_2, acetylated at the N-terminus (Ac) and amidated at the C-terminus (NH_2) to mimic the peptide bonds that would be present in the context of the protein) at low and high concentration. This peptide spans one of the regions most affected by oligomerization (see Fig. 4.43). Also for this peptide, the CD spectra were indicative of disorder both at 45 μM and 445 μM (see Fig. 4.54). No helical induction was observed.

We conclude that for the concentration range we could span by CD, oligomerization does not involve any structural changes that could be observed by this technique. This is in agreement with the observation that the difference in C_α chemical shifts at low and high concentration is close to the detection level.

4.8.2.3 Size exclusion chromatography by HPLC

The elution volumes of monomeric and oligomeric AF1* are sufficiently different to distinguish these species by size exclusion chromatography. We therefore next analyzed

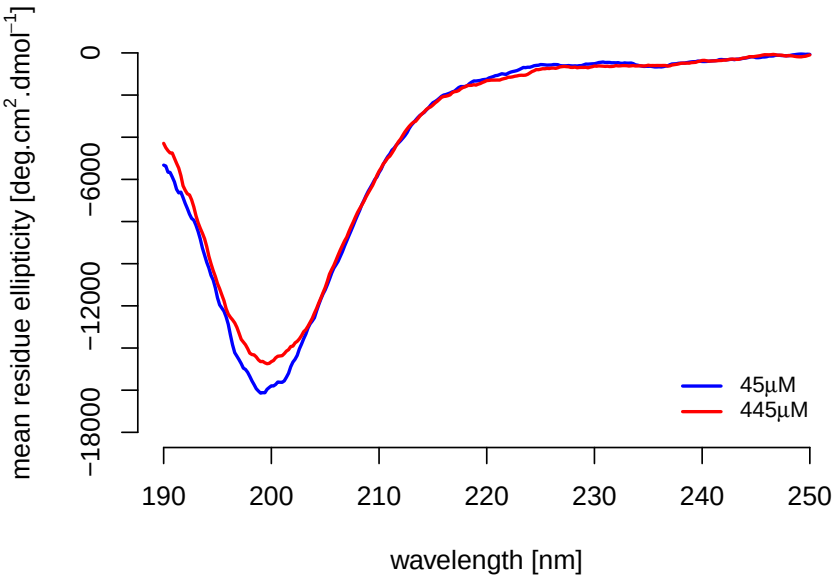


Figure 4.54: Far-UV CD spectra of AR peptide (426–446) at 45 μM and 445 μM.

AF1* using this technique by high-performance liquid chromatography (HPLC). HPLC is sensitive enough to detect oligomers even when they are low populated. Different concentrations of AF1* were injected into the column (5 μM, 10 μM, 25 μM, 50 μM, 100 μM, 250 μM and 400 μM). At low concentrations we could only detect one peak at retention time 7.6–7.7 min, which corresponds to the monomer of AF1* (see Fig. 4.55). Only at concentrations of 50 μM and higher, we detected an additional minor peak at retention time 6.7–6.8 min, which could correspond to a higher oligomer species of AF1* (see Fig. 4.55). Interestingly though, the intensity of this peak relative to the monomer peak did not noticeably increase between 50 μM and 400 μM, neither did its elution volume change considerably. This suggests that the oligomer did not evolve to a higher oligomeric species between these concentrations and, in addition, that the population of oligomer did not increase considerably. This suggests that at concentrations up to 400 μM only a low population of dimeric AF1* is present and no higher oligomeric species are formed.

HPLC does not allow to study higher concentrations of AF1* as this would saturate the system. We therefore moved to fast protein liquid chromatography (FPLC) to continue the analysis at higher concentration. FPLC has the disadvantage it is less sensitive compared to HPLC and it requires more sample, but offers the advantage one can work with higher concentrated samples.

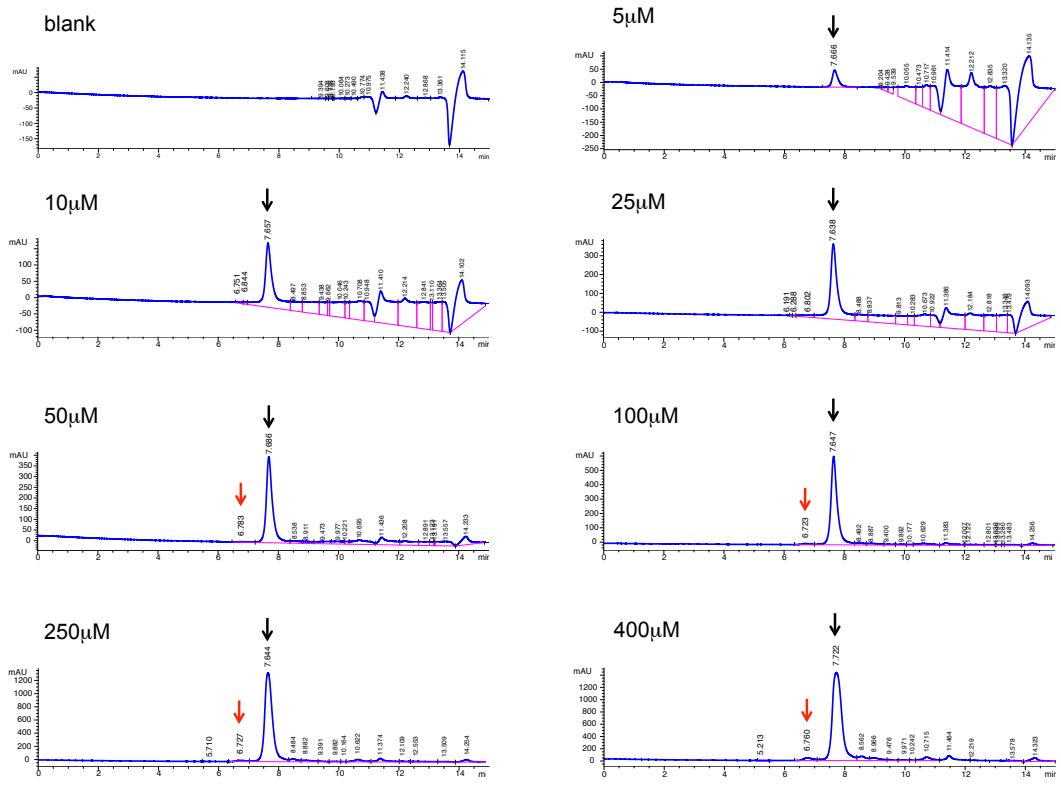


Figure 4.55: Size exclusion chromatography by HPLC of AF1* at concentrations between 5 μM and 400 μM. The peak corresponding to the eluting monomer is indicated with a black arrow, and that corresponding to the eluting oligomer, most likely a dimer, is indicated with a red arrow.

4.8.2.4 Size exclusion chromatography by FPLC

The analytical size exclusion column we used for this experiment was calibrated, which allowed us to differentiate between different oligomeric species. Based on the formulas from Uversky (see p. 125), the theoretical values for the hydrodynamic radius for monomeric, dimeric and trimeric AF1* are those given in table 4.3. The values in column 3 are for natively unfolded proteins in an extended conformation ($NU(coil)$), whereas those in column 4 are for natively unfolded proteins in a more collapsed conformation ($NU(PMG)$).

Table 4.3: Theoretical values for the hydrodynamic radius of mono-, di- and trimeric AF1* based on equations 4.1 and 4.2.

	MW(kDa)	$R_S^{NU(coil)}$(Å)	$R_S^{NU(PMG)}$(Å)
monomeric AF1*	31.2	46.2	37.3
dimeric AF1*	62.4	65.0	49.4
trimeric AF1*	93.6	79.4	58.1

When we injected 1 mM AF1* on the size exclusion column, we could distinguish two peaks (see Fig. 4.56). The major peak elutes at 12.7 mL and therefore corresponds to an experimental hydrodynamic radius of 46.2 Å (equations 4.3 and 4.4). A minor peak is present at 10.8 mL, which corresponds to an experimental hydrodynamic radius of 60.5 Å(equations 4.3 and 4.4). By comparing the experimental hydrodynamic radii with the theoretical ones in table 4.3, it is clear that the major peak corresponds to monomeric AF1* in an extended coil conformation. A hydrodynamic radius of 60.5 Å, corresponding to the species that eluted at 10.8 mL, is incompatible with monomeric AF1*, but could correspond to dimeric AF1* with a conformation that is not fully extended but neither as collapsed as in a premolten globule. Given that there was only a minor population of non-monomeric AF1* according to the chromatogram, it is improbable higher oligomers than dimer were present. Therefore, we did not consider the possibility that the minor peak of AF1* corresponded to a higher oligomer of AF1*.

4.8.3 Orientation of the dimer

Both head-to-head and head-to-tail orientations have been reported for dimers of (domains of) AR. The AR DBD has been crystallized as a head-to-head dimer on a DR3 element [210]. Head-to-tail dimers of full-length AR, stabilized by N/C interactions, were observed in the nucleus when AR was not bound to DNA [278, 279]. However, it remains unclear whether intra- and/or intermolecular N/C interactions occur when AR is bound to DNA [278, 280]. Since we found that AF1* has a tendency to oligomerize,

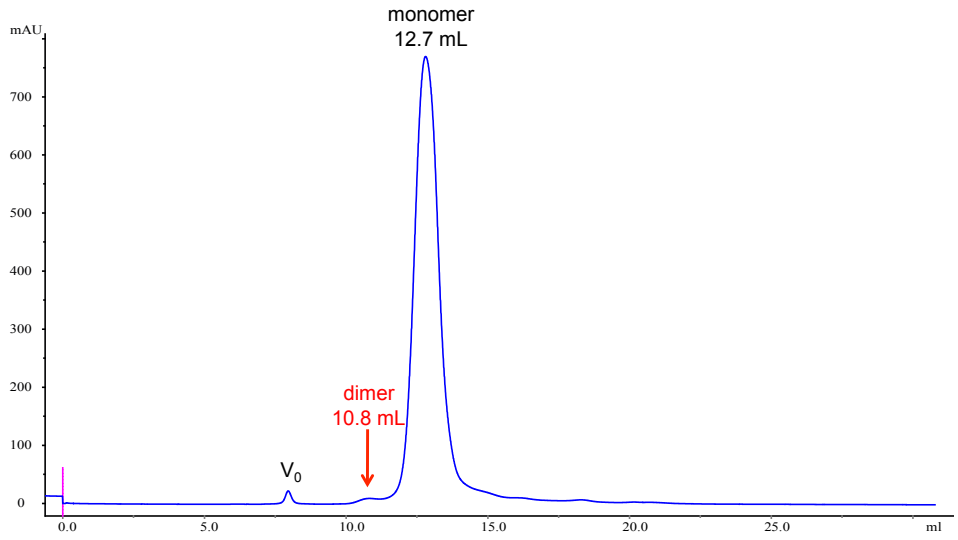


Figure 4.56: Analytical size exclusion chromatography by FPLC of AF1* at 1 mM.

it is possible the dimerization of AR on DNA involves the interaction of the N-terminal domain of two AR monomers. If this is the case, insights into the orientation of the AF1* dimer could help to elucidate the orientation of the AR dimer on DNA.

The AF1*c construct oligomerizes both on its own and in the context of AF1*. In addition, the chemical shift changes observed in AF1*c upon oligomerization are similar both in the individual construct and as part of the AF1* construct, even though AF1*c populates less dimer compared to AF1* at the same concentration. This indicates that AF1*c only requires this part of the sequence to self-interact, which supports a head-to-head orientation of the dimer.

4.8.4 Challenges to study oligomerization

Even though our NMR analysis unequivocally indicates AF1* oligomerizes with concentration, it is challenging to determine the populations of monomer, dimer and potential higher oligomeric species. Our further analyses have shown that AF1* has little tendency to oligomerize and hence the main population is that of the monomer over a wide concentration range. One of the main challenges for the study of the oligomerization of the NTD is the fact that AF1* is a 308-residue protein construct of the full-length AR (919 residues), which is not very prone to form oligomers. If the NTD dimerizes in the biological setting, it is likely to have a much higher tendency to do so than AF1* in solution. Oligomerization of the entire N-terminal domain (559 residues) is expected to be more favorable than that of AF1* (308 residues) as this process likely

involves cooperative interaction of various stretches of the NTD. Importantly, in the biological context, AR dimerization occurs when the NTD is linked to the DBD, which in turn is bound to DNA. DNA binding thus positions the NTD domains of two AR molecules close in space and increases the local concentration substantially, facilitating interaction. In our *in vitro* system, neither the DBD nor the DNA containing AREs are present and AF1* is free to diffuse in solution. Therefore, we cannot reach local concentrations as high as those in the biological context, disfavoring intermolecular interaction. In addition, AF1* could form higher oligomers than dimers in the absence of the remaining AR sequence and the DNA that restrain it to forming a dimer, although this does not seem to happen in the concentration range we tested.

To study the oligomerization properties of the N-terminal domain of the AR, we would strongly recommend to use a protein construct that includes at least the entire NTD (559 residues). In order to avoid artifactual oligomerization to higher oligomers, it would be advisable to also include the DBD in the protein construct and to study oligomerization in the presence of DNA containing AREs. Since it might be a challenging task to express and purify a protein with a large disordered domain (NTD) followed by a globular domain (DBD), an alternative approach would be to artificially link the termini of the NTD to bring them close in space and facilitate the oligomerization. This can be done, for instance, by N-terminal addition of an FGG peptide motif which forms supramolecular interactions with cucurbit[8]uril and induces protein dimerization [428]. In addition, the presence of the LBD is likely to affect the nature of interdomain interactions of DNA-bound AR molecules.

4.9 Summary

Taken together, our data indicate that under the conditions we characterized the conformational properties of AF1* (protein concentration, buffer, temperature, pH), the protein is mainly present as a monomer and only a small population of dimer is present. Oligomerization could not reliably be detected by circular dichroism or size exclusion chromatography, most likely because of the low population of dimer in the concentration range we can span with each technique.

We found, by NMR, that several regions of sequence in AF1* displayed a high helical propensity (see Fig. 4.27). Particularly in the regions spanned by residues 176–202 (in core Tau-1) and residues 391–413 (in Tau-5) this helical propensity was considerably higher than expected for an IDP (see Fig. 4.26). However, by CD we could not detect any helicity in AF1* (see Fig. 4.52). Size exclusion chromatography also indicated that AF1* mainly existed in a monomeric extended conformation (see Fig. 4.56).

Furthermore, the predicted helicity by Agadir was much smaller than the experimentally determined helicity (see Fig. 4.31). This discrepancy is striking, but interestingly a similar discrepancy has been observed for other systems such as acid-denatured acyl coenzyme A binding protein (ACBP) [429, 430].

ACBP is a four-helix bundle protein that is denatured under acidic conditions [431, 432]. For acid-denatured ACBP (pH 2.3) the secondary C_α chemical shifts suggest the presence of considerable helicity among members of the ensemble of unfolded structures [429, 433], whereas the CD spectrum recorded at pH 2.21 was characteristic of a disordered polypeptide chain with only a slight deviation at 222 nm, indicating a very small amount of residual α-helical structure [430]. Also in this case, the predicted helicity by Agadir (pH 2.4) was found to be considerably smaller compared to the experimentally observed one [433]. The observed residual helicity in acid-denatured ACBP is, in the first place, the result of an equilibrium between an unfolded state and a state with local helix structure formation, strongly shifted toward the unstructured state [433]. Long-range helix to helix interactions were found in the acid-denatured state of ACBP between regions of sequence that in the native state adopt a helical conformation and are organized in a four-helix bundle [433]. It was therefore further proposed that the residual helicity in acid-denatured ACBP may not only be a result of the locally stabilized helicity, but that in addition it might be stabilized by other types of interactions, such as the observed long-range helix to helix interactions [433]. Moreover, it was proposed that those stabilizing effects might cause the discrepancy between the observed and predicted helicity for acid-denatured ACBP [433].

In a similar way, the high helical propensity observed in AF1* could be a combination of transient local helicity and long-range helix to helix interactions. We have shown that no long-range interactions that induce structure take place between the different AF1* segments, i.e. between AF1*a, AF1*b and AF1*c. However helix stabilizing interactions within each of the AF1* segments, i.e. within AF1*a, within AF1*b and/or within AF1*c, may occur, as well as intermolecular long-range helix to helix interactions between two AF1* monomers upon dimerization.

Our results suggest that the high helical propensity of the core Tau-1 region is intrinsic to the AF1* construct. This region seems to populate a helical conformation a large fraction of the time in a dynamic manner (fast forming and breaking of the helix) and is characterized by a fast exchange regime. Although in the concentration range we could span in our experiments no noticeable increase in helicity was observed in the core Tau-1 region upon dimerization, the ^{1}H and ^{15}N chemical shift changes upon concentrating the protein indicate that this region is involved in dimerization. In the AF1* dimer, the core Tau-1 regions of both AR molecules most likely interact via a

coiled-coil interaction. The presence of even a very small population of dimeric AF1* with a more stable coiled-coil interaction could contribute to the high helicity observed in this region.

The Tau-5 region of AF1* is characterized by a more collapsed conformation compared to other regions of AF1*. The three stretches with helical propensity in the Tau-5 region of AF1* likely stabilize each other through transient long-range contacts to form a relatively collapsed conformation (see Fig. 4.37). In addition, our data indicate that dimerization takes place primarily through the Tau-5 region and possibly involves a conformational rearrangement of this region from a relatively collapsed conformation stabilized by helix to helix interactions between the different helical elements in monomeric AF1*c to a head-to-head orientation of these helical elements in the context of the dimer. The more pronounced line broadening and high relaxation rates in this part of the sequence most likely do not only reflect the conformational exchange arising from the equilibrium between more extended and more collapsed conformations in the monomer, but also the additional chemical exchange due to the conformational rearrangement of the Tau-5 region upon dimerization. Furthermore, interactions in the dimer between the regions of sequence with helical propensity could contribute to the observed high helicity in the Tau-5 region. AF1*c is characterized by an intermediate exchange regime.

Even though we cannot exclude a contribution of AF1* dimer in the observed high helical propensity in various regions of AF1*, the dimeric state of AF1* is very low populated at all protein concentrations tested and the helicity of AF1* is almost the same at 25 μM and 390 μM. We therefore conclude that monomeric AF1* displays considerable, but transient, helicity in several stretches of the sequence. The residual helicity in these regions is most likely more pronounced in a fully dimeric state of AF1*. Collectively, our data indicate that monomeric AF1* exists in a dynamic equilibrium between a completely extended conformation and different partially helical conformations. This is in line with a molten globular conformation as suggested before [156]. However, our findings indicate that mainly the AF1*c region is characterized by a molten globule-like behavior whereas the rest of AF1* behaves like an IDP with high intrinsic helical propensity in the core Tau-1 region.

Importantly, the regions of sequence identified to have high helical propensity also display high transverse relaxation rates, are involved in dimerization and coincide with regions previously identified to be crucial for AR transactivation, core Tau-1 and Tau-5.

Tau-1 and Tau-5 are known to be functionally independent units, dominant for the regulation of androgen-dependent and androgen-independent AR transactivation,

respectively [149, 230, 232]. We have shown that they are also structurally independent units (no structure-inducing long-range contacts between Tau-1 and Tau-5) with different conformational properties (core Tau-1 is in fast exchange and in equilibrium between a disordered and a helical conformation, whereas Tau-5 adopts a more collapsed conformation, likely stabilized by helix to helix interactions between the three helical elements within the Tau-5 region, and is possibly undergoing a conformational rearrangement upon dimerization, which is characterized by an intermediate exchange regime).

5

Interaction of AR and RAP74

As a transcription factor the AR interacts with members of the transcription machinery and co-regulatory proteins. AF1 within the N-terminal transactivation domain of AR is indispensable for the protein-protein interactions that trigger this key biological process. However, the molecular mechanisms of these interactions are poorly understood due to the ID nature of AF1. Importantly, the aberrant transactivation activity of AR in late stage PCa is presumably linked to the AF1 in its NTD [150, 155, 232]. Gaining insights into the mechanisms by which AF1 regulates transcription could therefore help to better understand the role of AF1 in the survival of late stage PCa cells and potentially lead to new therapeutic strategies.

Having addressed the conformational properties of the AF1 region of AR, we further aimed to understand in more detail how the NTD interacts with the transcription machinery to activate transcription. In addition, we were interested in how the structural properties of the NTD are altered by the protein-protein interactions that drive this biological process.

It is known that the NTD of AR interacts with RAP74 (RNA polymerase II associated protein 74), subunit 1 of the general transcription factor TFIIF. This interaction has been shown to occur *in vitro*, both by biochemical binding assays [282, 283, 319, 320] and in cellular assays [351] but has remained poorly characterized due to the absence of high resolution information of the NTD of AR. A summary of the information in the literature regarding this interaction is provided in the introduction (see section 1.3.2). This chapter describes the detailed characterization we carried out of this interaction and its biological relevance.

5.1 Protein constructs and peptides used for the NMR study of this interaction

We used several protein constructs and peptides to study the interaction between AR and RAP74, which are summarized here.

5.1.1 AR

For the AR, we used AF1* (AR 142–448), AF1*c (AR 330–448) and several peptides of the binding motif (wild type and mutants) (see table 5.1). In addition, we used a double mutant (E440K/E441K) of AF1*c.

Table 5.1: Peptides of AR used in the study of the interaction with RAP74. Mutated residues are underlined. All peptides were acetylated at the N-terminus (Ac) and amidated at the C-terminus (NH_2) to mimic the peptide bonds that would be present in the context of the protein. Phosphorylated serine residues are denoted by pS.

peptide name	sequence
WT	Ac-^{426}SAAASSSWHTLFTAEEGQLYG^{446}-NH_2
W433A/L436A/F437A	Ac-^{426}SAAASSSAHTAATAEEGQLYG^{446}-NH_2
E440K/E441K	Ac-^{426}SAAASSSWHTLFTAKKGQLYG^{446}-NH_2
pS424	Ac-^{421}GSGpSPSAAASSSWHTLFTAEEGQLYG^{446}-NH_2
pS430	Ac-^{426}SAApSSSWHTLFTAEEGQLYG^{446}-NH_2
pS431	Ac-^{426}SAASpSSWHTLFTAEEGQLYG^{446}-NH_2
pS432	Ac-^{426}SAASSpSWHTLFTAEEGQLYG^{446}-NH_2
S430E/S431E/S432E	Ac-^{426}SAAAEEEWHTLFTAEEGQLYG^{446}-NH_2
pS430/pS432	Ac-^{426}SAApSSpSWHTLFTAEEGQLYG^{446}-NH_2
C-Cys	Ac-^{426}SAAASSSWHTLFTAEEGQLYG^{446}C-NH_2

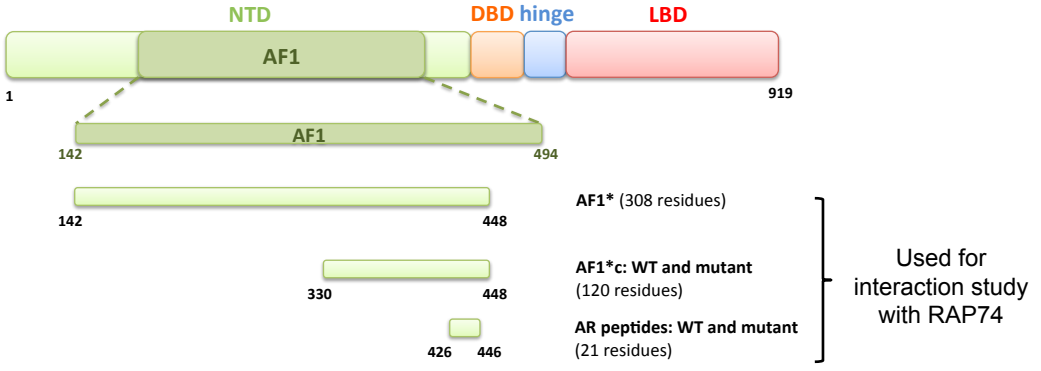


Figure 5.1: Schematic representation of the AR constructs used for the interaction study with RAP74.

5.1.2 RAP74

Both the N-terminal 136 amino acids and the C-terminal 155 amino acids of RAP74 were found to interact with AR, with the C-terminal residues presenting the principal binding site [320]. As described in the introduction, the N-terminal region of RAP74 (residues 2–172) forms a heterodimer with the N-terminal region of RAP30 (residues 2–119), the other subunit of TFIIF. This heterodimer is present in the RNA polymerase II–TFIIF complex both in human [348] and in yeast [349] and is bound to RNA polymerase II. It is therefore unlikely that the N-terminal domain of RAP74 plays an important role in the interaction with the AR.

We used two protein constructs of RAP74 to study the interaction with AR: RAP74-CTD (RAP74 363–517 with a C-terminal His-tag: LEHHHHHH) and RAP74NMR (RAP74 450–517) (see Fig. 5.2 and appendix A.2). RAP74-CTD contains the folded winged helix-turn-helix domain at the C-terminus (residues 449–517) and an unstructured tail at the N-terminus, and corresponds to one of the constructs used for the original interaction studies [320]. The pET-23d plasmid encoding the RAP74-CTD construct [320] was a gift from Prof. Iain McEwan (University of Aberdeen, UK). In RAP74NMR, the unstructured tail was removed, so the construct only contains the folded domain.

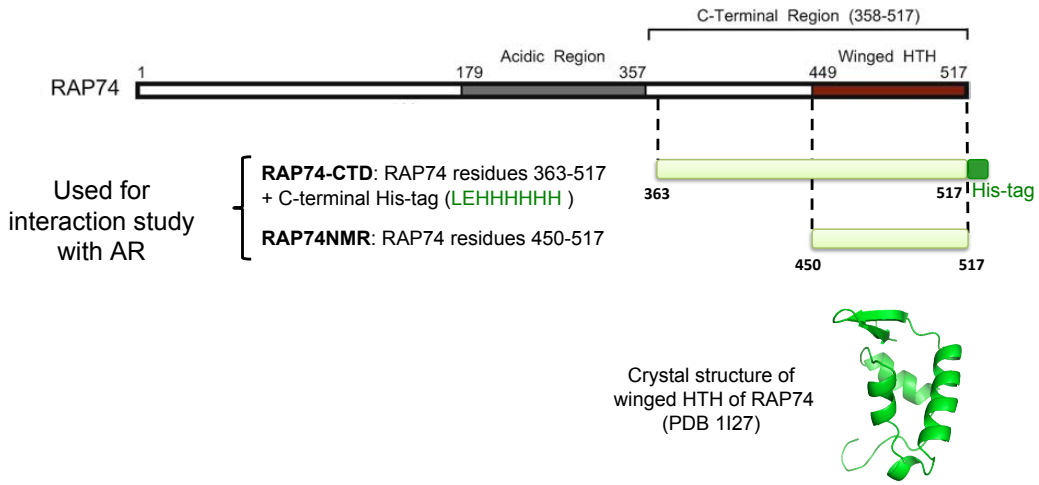


Figure 5.2: Schematic representation of the RAP74 constructs used for the interaction study with AR. The crystal structure of the C-terminal domain of RAP74, showing the winged helix-turn-helix (HTH) domain, is represented by PDB entry 1I27 [350].

5.1.2.1 Cloning of RAP74NMR construct

To obtain RAP74NMR, we recloned the corresponding sequence of RAP74 in Gateway vector pDEST-HisMBP and introduced a TEV cleavage site between the N-terminal HisMBP-tag and the RAP74 sequence, similar to the cloning strategy used for the AR constructs. Consequently, the HisMBP-tag could be cleaved with TEV protease during protein purification. Since the residual glycine residue at the N-terminus of the construct obtained through this methodology coincides with glycine residue 450 of RAP74, the final product after TEV cleavage corresponds to RAP74 450–517, without non-native residues.

5.2 The interaction between AR and RAP74 takes place at the key regulatory ^{433}WHTLF^{437} motif of AR

NMR is a useful tool for the study of molecular interactions. Chemical shift perturbation experiments provide the binding epitope of interacting partners. This is because the residues involved in binding have a different chemical environment in the free and in the bound state, which is translated into chemical shift changes. Residues that are not in the binding site are, by contrast, not affected by the presence of the binding partner, hence their resonances will appear in the same position in the spectrum independent of the presence of the interaction partner.

To identify the binding motif in the AR, we carried out NMR experiments of isotopically labeled ^{15}N-AF1* (NMR visible), which contains the ^{159}P<u>S</u>TL<u>S</u>L^{164} and ^{340}P<u>S</u>TL<u>S</u>L^{345} motifs previously suggested to be important for interaction [320], as well as residues M245, L247 and V249 suggested to form part of the binding surface for TFIIF [353, 434], in the presence of unlabeled RAP74-CTD (NMR silent). As shown in figure 5.3, binding takes place at a 16-amino acid stretch in Tau-5 between residues S431 and G446 (^{431}SSWHTLFTAEEGQLYG^{446}). Interestingly, we did not observe significant chemical shift changes at residues M245, L247 or V249, nor at residues S160, S163, S341 or S344, all previously identified to be important for RAP74 binding [320, 353, 434].

Average ^{1}H and ^{15}N chemical shift changes were calculated using equation 5.1, where $\Delta\delta_{H^N}$ and $\Delta\delta_{N^H}$ are the amide proton and nitrogen chemical shift differences, respectively.

$$\Delta\delta_{av} = \sqrt{(\Delta\delta_{H^N})^2 + \left(\frac{\Delta\delta_{N^H}}{5}\right)^2} \tag{5.1}$$

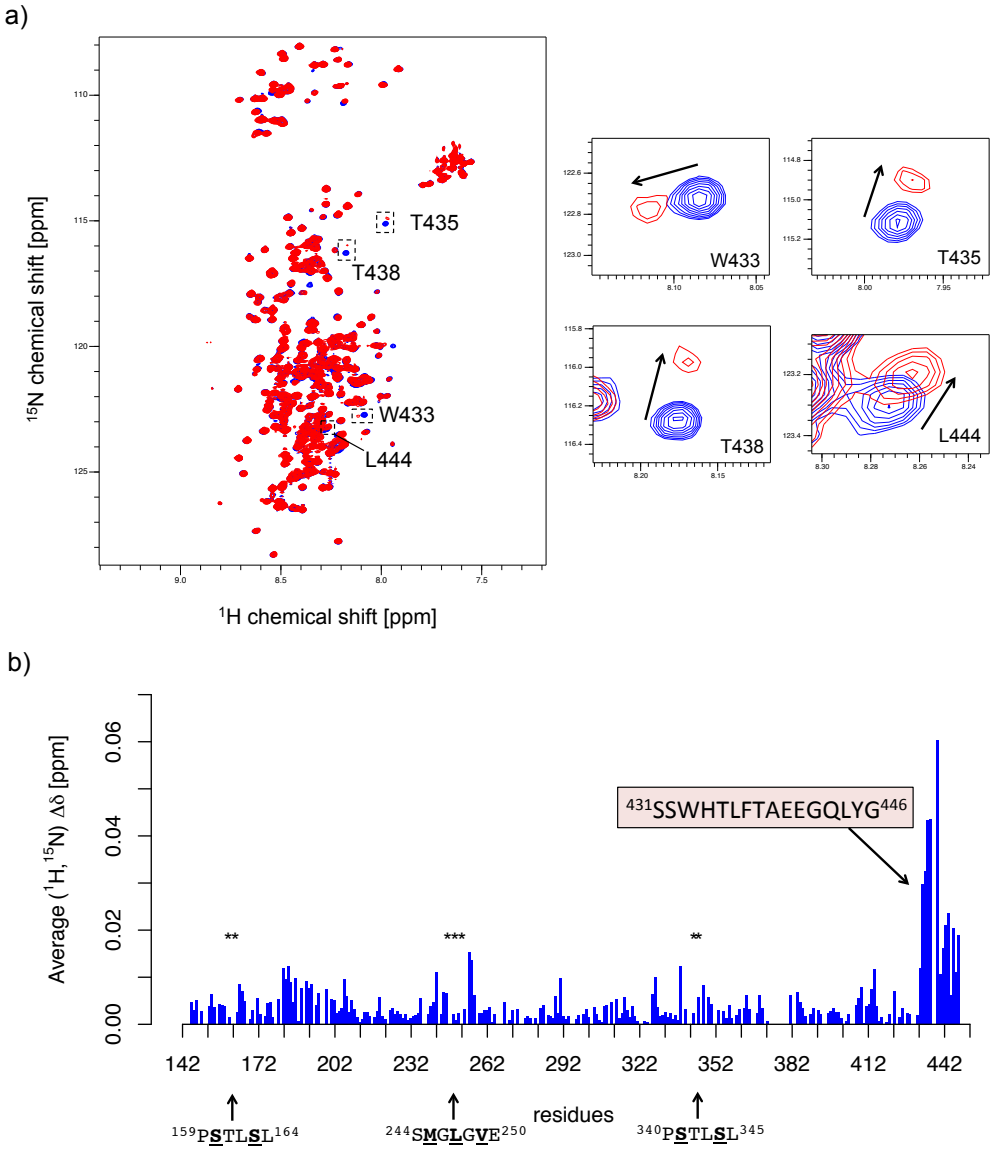


Figure 5.3: a) [^{1}H,^{15}N]-HSQC of free ^{15}N-AF1* (blue) and ^{15}N-AF1* + 3 molar equivalents of RAP74-CTD (red). Blow-ups of residues W433, T435, T438 and L444 are shown. b) Average ^{1}H and ^{15}N chemical shift changes in AF1* upon binding of RAP74-CTD. The binding site is shown and residues S160, S163, M245, L247, V249, S341 and S344 are highlighted (*).

To confirm that the interaction takes place solely at the binding motif and does not involve tertiary structure formation in AR, in addition to looking at the interaction of ^{15}N-AF1* and RAP74-CTD, we examined the interaction of a smaller AR construct, ^{15}N-AF1*c, which contains the identified binding site, and RAP74-CTD. The chemical shift changes in the two AR constructs upon RAP74-CTD binding occur at the same amino acids, indicating that the interaction is local and does not involve tertiary structure formation in AR (see Fig 5.4).

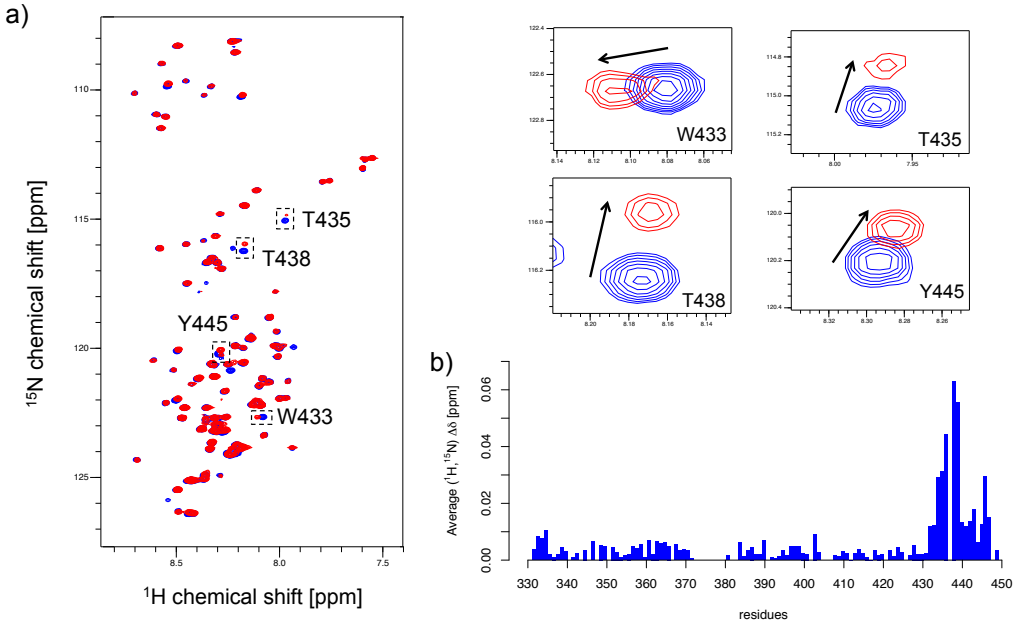


Figure 5.4: a) [^{1}H,^{15}N]-HSQC of free ^{15}N-AF1*c (blue) and ^{15}N-AF1*c + 4 molar equivalents of RAP74-CTD (red). Blow-ups of residues W433, T435, T438 and Y445 are shown. b) Average ^{1}H and ^{15}N chemical shift changes in AF1*c upon binding of RAP74-CTD.

We further investigated whether the unstructured residues at the N-terminus of the RAP74-CTD construct were involved in the interaction. For this purpose, we compared the interaction between ^{15}N-AF1*c and RAP74-CTD with that of ^{15}N-AF1*c and RAP74NMR (see Fig. 5.5). At the same excess of the different RAP74 constructs, we observed comparable chemical shift changes in ^{15}N-AF1*c, indicating that the unstructured tail at the N-terminus of RAP74-CTD does not play a role in the interaction.

Furthermore, the chemical shift changes caused by 10 molar equivalents of RAP74NMR on ^{15}N-AF1* are identical to those caused by the same excess of RAP74NMR on ^{15}N-AF1*c, indicating the affinity of both AR constructs for RAP74NMR is the same and confirming that the binding site is the sequence between S431 and G446 (see Fig. 5.6). For this reason, it was possible to use a peptide of the AR spanning residues 426 to 446 (Ac-^{426}SAAASSSWHTLFTAEEGQLYG^{446}-NH_2) for further interaction studies when

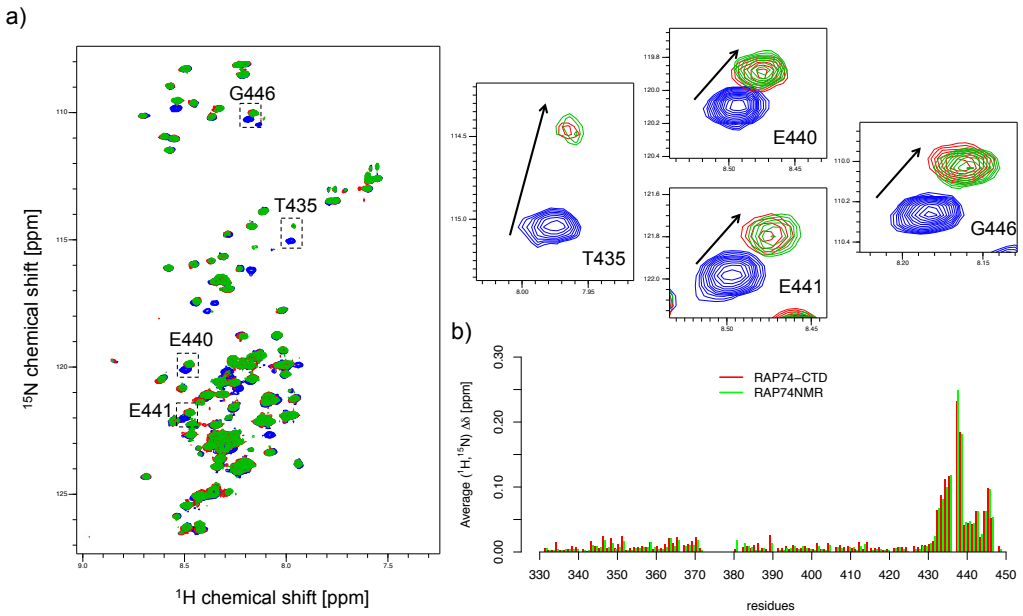


Figure 5.5: a) [^{1}H,^{15}N]-HSQC of free ^{15}N-AF1*c (blue), ^{15}N-AF1*c + 12 molar equivalents of RAP74-CTD (red) and ^{15}N-AF1*c + 12 molar equivalents of RAP74NMR (green). Blow-ups of residues T435, E440, E441 and G446 are shown. b) Average ^{1}H and ^{15}N chemical shift changes in AF1*c upon binding of RAP74-CTD (red) and RAP74NMR (green).

monitoring changes in the RAP74NMR spectra. This peptide is further referred to as wild type (WT) peptide (see table 5.1 on p. 184).

Importantly, the ^{433}WHTLF^{437} motif in the core of the binding site is known to be crucial for the regulation of AR transcriptional activity in CRPC cells [232], i.e. in advanced PCa for which no efficient treatment is currently available. In healthy prostate cells or in early stage PCa cells (that are still dependent on hormone), the activity of AR is hormone-dependent and mediated through the Tau-1 region of AR. However, in late stage PCa cells where the presence of hormone is no longer required to activate AR (CRPC cells), it is instead the Tau-5 region that plays a dominant role in the regulation of AR activity. It was found that for these CRPC cells the role of Tau-5 is fundamentally distinct depending on hormone levels. In the presence of androgen, Tau-5 *inhibits* transactivation (i.e. has a protective role) in androgen-independent cell lines. However, it *enhances* AR transactivation in these cell lines when no hormone is present or at castrate levels of hormone (i.e. $\leq$1 nM[1]) [232]. The core motif within Tau-5 mediating this activity was mapped to ^{433}WHTLF^{437} [232].

In other words, the ^{433}WHTLF^{437} motif enhances aberrant AR transactivation at

[1] According to the FDA, the target testosterone level for androgen deprivation therapy is <1.7 nM [435].

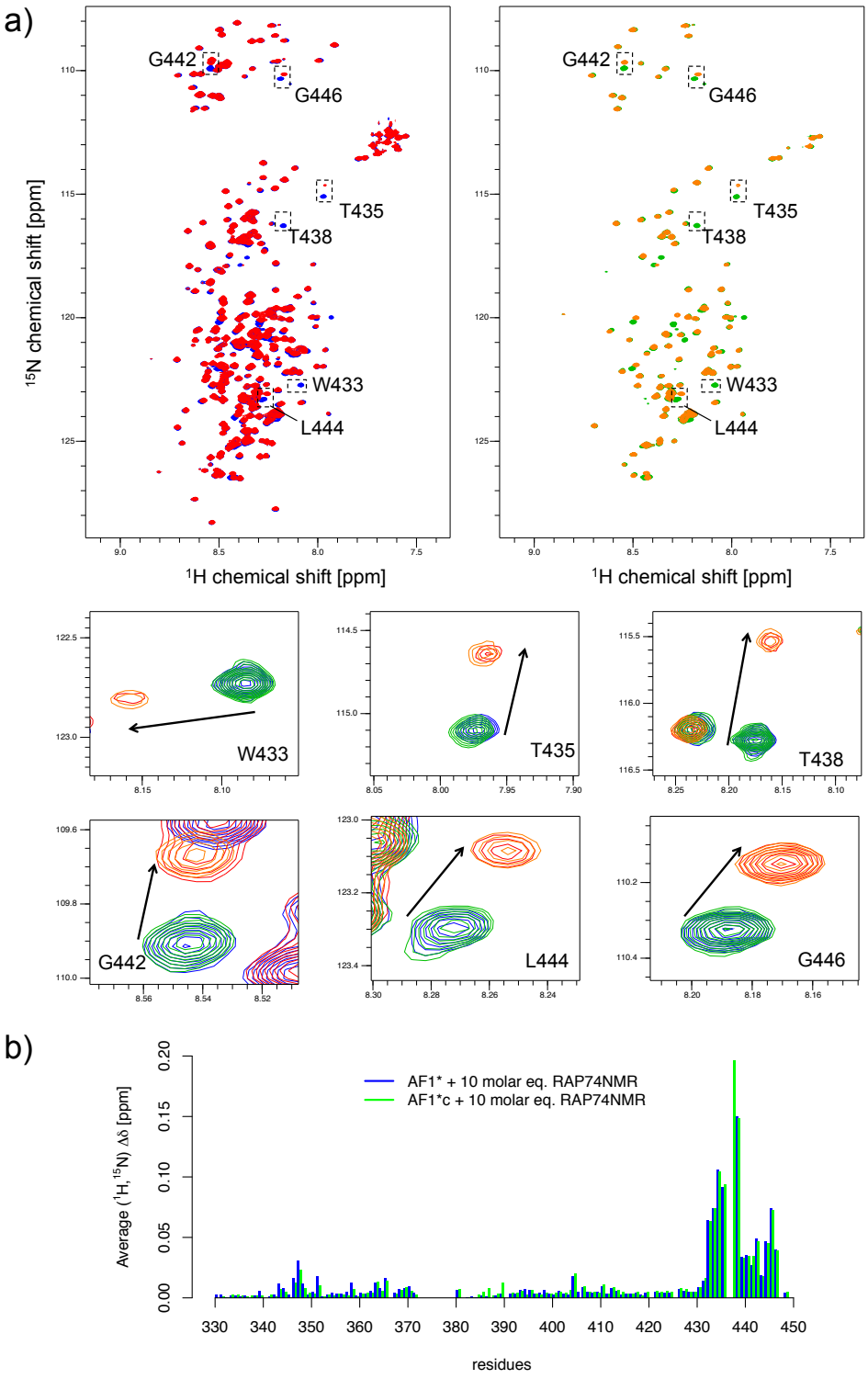


Figure 5.6: a) On the left: [^{1}H,^{15}N]-HSQC of free ^{15}N-AF1* (blue) and ^{15}N-AF1* + 10 molar equivalents of RAP74NMR (red). On the right: [^{1}H,^{15}N]-HSQC of free ^{15}N-AF1*c (green) and ^{15}N-AF1*c + 10 molar equivalents RAP74NMR (orange). Blow-ups of selected residues are shown (the four spectra overlapped). c) Average ^{1}H and ^{15}N chemical shift changes in AF1* (blue) and AF1*c (green) upon binding of RAP74NMR.

castrate levels of hormone in CRPC cells. These are the conditions found in late stage PCa patients undergoing ADT. For these patients, no efficient treatment is currently available due to our lack of understanding of the mechanisms regulating AR activity at this stage of the disease. It is therefore of paramount importance to understand how the AR is activated under *these* conditions. The ^{433}WHTLF^{437} motif has been identified to play a key role in enhancing AR activity in CRPC cells and at castrate levels of hormone and has consequently been suggested to represent an interface through which aberrant AR activity could be targeted for therapy of CRPC [232]. However, the molecular mechanisms underlying the role of this motif remain poorly understood.

We found that the interaction between AR and RAP74, a member of the general transcription machinery, takes place precisely at this key regulatory motif. In addition to characterizing this interaction in more detail, we further investigated its potential role in this aberrant transactivation.

5.2.1 Conformational changes in AR upon binding RAP74

To investigate whether binding of AR to RAP74 caused conformational changes in AR at the binding motif, we recorded C_α chemical shifts, sensitive probes of secondary structure, of AF1*c both in the absence and in the presence of RAP74-CTD. We observed a moderate increase in helicity at the binding motif of AR in the presence of RAP74-CTD (see Fig. 5.7), suggesting folding of this region into a more helical conformation upon binding. Interestingly, the binding site of AR in RAP74 coincides with the region most C-terminal of AF1* with some degree of α-helical propensity. A more detailed description of this observation is given in section 5.10.

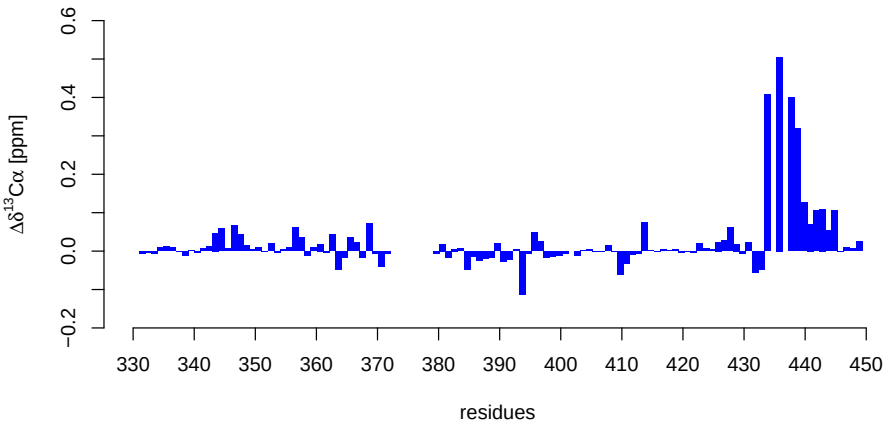


Figure 5.7: Difference in C_α chemical shifts between AF1*c (free) and AF1*c in the presence of 5 molar equivalents of RAP74-CTD (bound).

5.3 Binding site of AR on RAP74 is located in a hydrophobic groove

To identify the binding epitope of AR on RAP74 we reversed our labeling strategy and monitored the changes in the spectrum of ^{15}N-RAP74NMR upon addition of unlabeled AF1*c. In order to do this, we first assigned the RAP74NMR protein.

5.3.1 Backbone assignment of RAP74NMR

For this purpose we recorded 3D triple resonance experiments at 298 K of RAP74NMR in 20 mM sodium phosphate buffer at pH 6.5. RAP74NMR does not contain cysteine residues so it was not necessary to use a reducing agent like TCEP. The combination of 298 K and pH 6.5 yielded good quality spectra for RAP74NMR. At this temperature, the correlation time did not cause severe line broadening, and at pH 6.5 the effects of solvent exchange were limited. The assigned [^{1}H,^{15}N]-HSQC of RAP74NMR is shown in figure 5.8.

A complete list of the RAP74NMR backbone assignment (H^N, N^H, CO, C_α) can be found in appendix A.8.

5.3.2 Determination of the binding epitope

We next monitored the changes in the HSQC spectrum of ^{15}N-RAP74NMR upon addition of 10 molar equivalents of unlabeled AF1*c (see Figs. 5.9a and b). The spectra were recorded at 298 K and pH 7.4. These experiments revealed that AR binds a well-defined hydrophobic surface formed by helix 2 and helix 3 of the winged helix-turn-helix domain formed by the C-terminus of RAP74 (see Fig. 5.9c).

As shown in figure 5.10 the binding groove of the C-terminal domain of RAP74 is hydrophobic and the protein domain is highly positively charged in general.

We next probed the changes on the spectrum of ^{15}N-RAP74NMR caused by 10 molar equivalents of the AR peptide spanning residues 426–446 (WT peptide, see table 5.1, p. 184) at 298 K and pH 7.4 (see Fig. 5.11). As expected, the peptide caused chemical shift changes in the same amino acids as the longer AF1*c construct. Therefore, the peptide could be used for further interaction studies.

These interaction experiments were recorded at pH 7.4 because the core of the AR binding motif contains histidine residue 434 (^{433}W<u>H</u>TLF^{437}). The typical pKa of his-

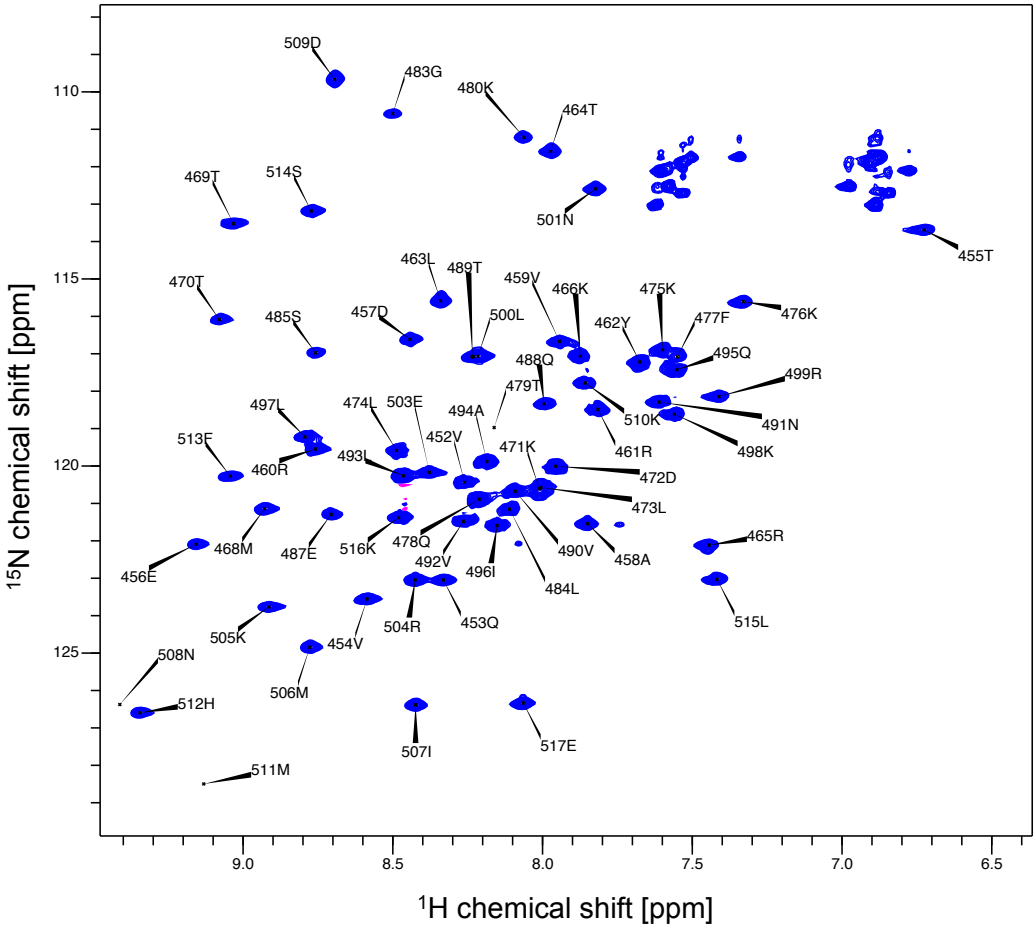


Figure 5.8: The [^{1}H,^{15}N]-HSQC of RAP74NMR (RAP74 450–517) with assignment of the backbone amide proton and nitrogen, in 20 mM sodium phosphate at pH 6.5 and 298 K.

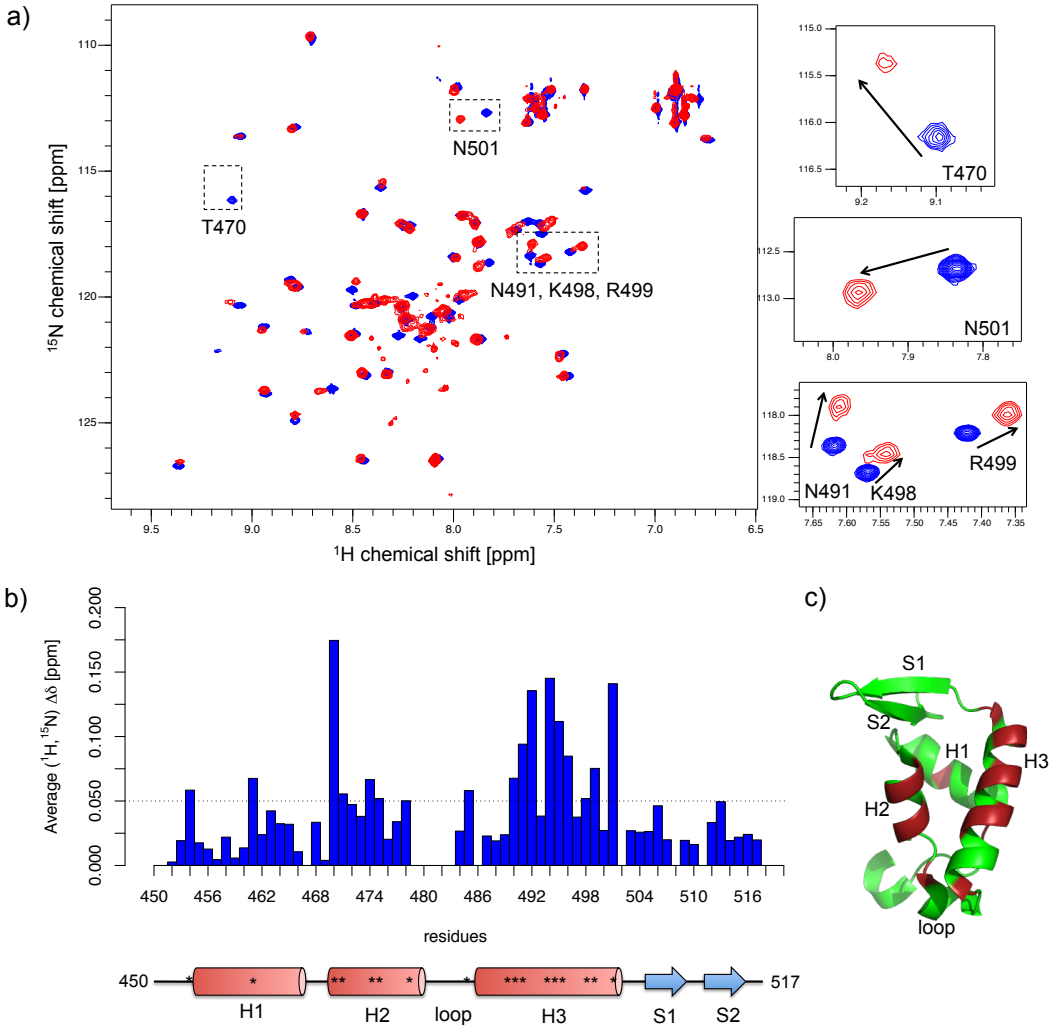


Figure 5.9: a) [^{1}H,^{15}N]-HSQC of free ^{15}N-RAP74NMR (blue) and ^{15}N-RAP74NMR + 10 molar equivalents of AF1*c (red) at pH 7.4. b) Average ^{1}H and ^{15}N chemical shift changes in RAP74NMR upon binding of AF1*c, with an indication of the structural elements in RAP74NMR beneath the sequence. The residues for which the average ^{1}H and ^{15}N chemical shift changes are larger than 0.05 ppm are indicated with the symbol *. c) Mapping of chemical shift perturbations on the structure of RAP74NMR (PDB 1I27), indicating the residues of RAP74NMR that are affected by binding of AR. RAP74NMR residues for which the average ^{1}H and ^{15}N chemical shift changes are larger than 0.05 ppm are colored in dark red (V454, R461, T470, L474, K475, Q478, S485, V490, N491, V492, A494, Q495, I496, K498, R499 and N501).

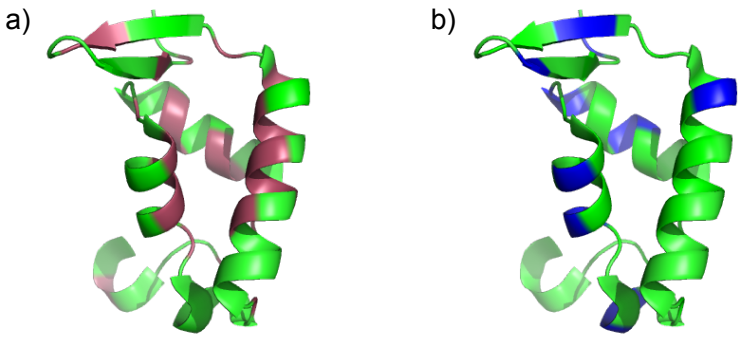


Figure 5.10: The hydrophobic residues (raspberry) and residues with positively charged side chains (blue) of RAP74NMR are shown on its structure (PDB 1I27), figures a and b respectively.

tidine side chains is ca. 6, but this value depends on the chemical environment of the histidine side chain in the context of the protein [410]. Since the surface of RAP74NMR is highly positively charged a change in the protonation status of AR H434 could have a considerable effect on the interaction between both proteins. The interaction experiments of ^{15}N-AF1*c and RAP74NMR as well as those of ^{15}N-RAP74NMR and AF1*c or AR WT peptide were performed at pH 7.4, at which the histidine side chain is expected to be uncharged. However, at this pH value the resonances of RAP74 residues E456, T479, K480, G483, S485, E487 and N508 are broadened due to solvent exchange. It is therefore not possible to determine whether they are part of the binding site. We therefore repeated the interaction experiment at pH 6.5, at which the resonances corresponding to these residues are distinguishable. As shown in figure 5.12, the chemical shift changes in RAP74NMR caused by binding of the WT peptide are almost identical at pH 6.5 and pH 7.4. These results show that RAP74 residues E456, T479, K480, G483, S485, E487 and N508 are not located directly in the binding site and in addition confirm that the protonation status of H434 of the AR is the same at pH 6.5 and pH 7.4.

5.3.3 Conformational changes in RAP74NMR upon binding of AR

To investigate whether interaction causes changes in the conformational properties of RAP74NMR, we monitored the C_α chemical shifts of RAP74NMR in the absence and presence of 4 molar equivalents of the AR peptide. Since the C_α chemical shifts do not change upon binding, we conclude that the interaction does not significantly alter the structural properties of RAP74 (see Fig. 5.13).

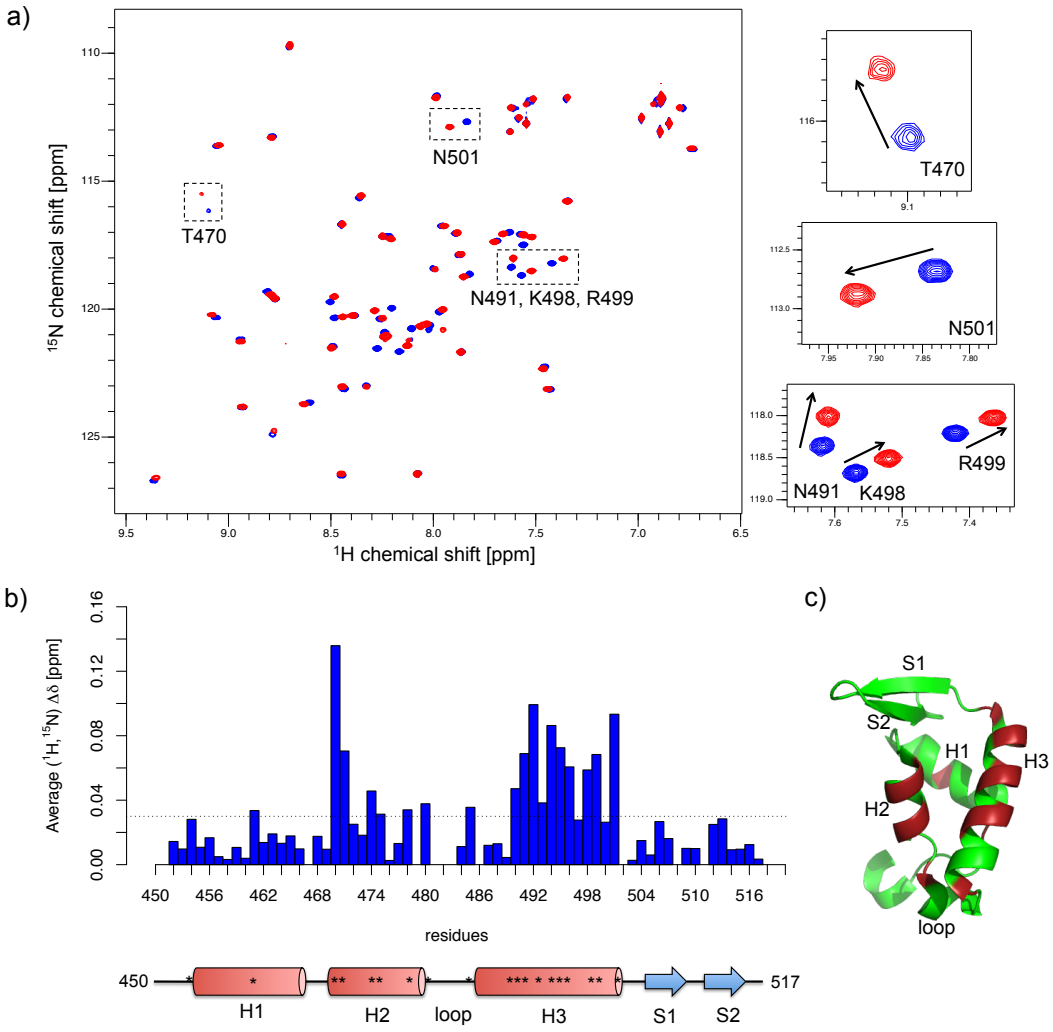


Figure 5.11: a) [^{1}H,^{15}N]-HSQC of free ^{15}N-RAP74NMR (blue) and ^{15}N-RAP74NMR + 10 molar equivalents WT peptide (red) at pH 7.4. b) Average ^{1}H and ^{15}N chemical shift changes in RAP74NMR upon binding of the peptide (at pH 7.4), with an indication of the structural elements in RAP74NMR beneath the sequence. The residues for which the average ^{1}H and ^{15}N chemical shift changes are larger than 0.03 ppm are indicated with the symbol *. c) Mapping of chemical shift perturbations on the structure of RAP74NMR (PDB 1I27), indicating the residues of RAP74NMR that are affected by binding of AR. RAP74NMR residues for which the average ^{1}H and ^{15}N chemical shift changes are larger than 0.03 ppm are colored in dark red (R461, T470, K471, L474, K475, Q478, K480, S485, V490, N491, V492, L493, A494, Q495, I496, K498, R499 and N501).

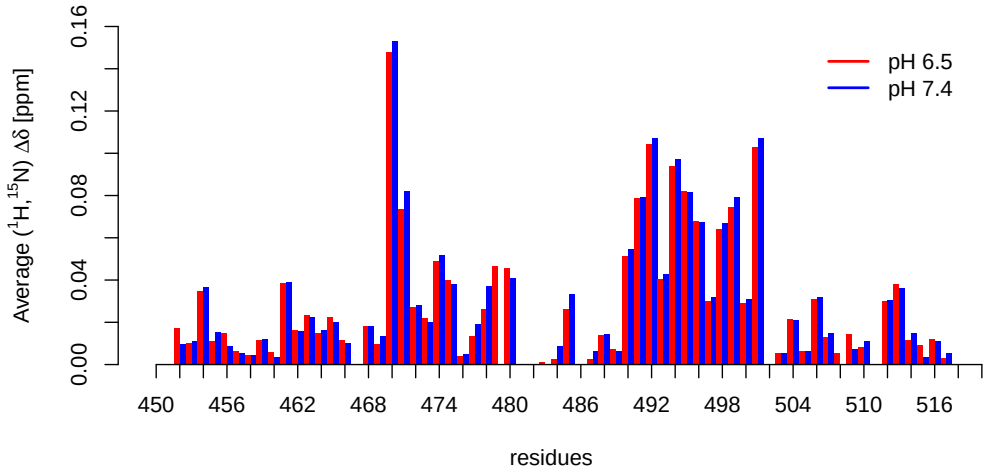


Figure 5.12: Average ^{1}H and ^{15}N chemical shift changes in RAP74NMR upon binding of WT peptide, at pH 6.5 (red) and at pH 7.4 (blue).

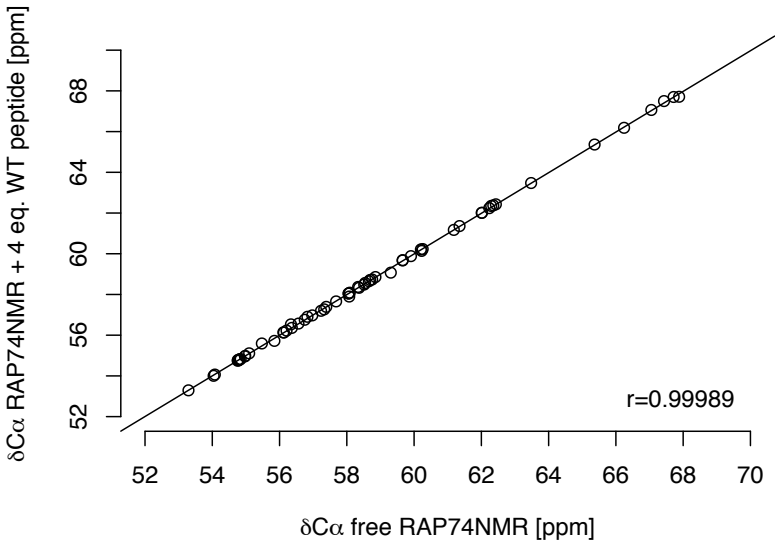


Figure 5.13: Correlation plot of the C_α chemical shifts in free RAP74NMR and in RAP74NMR in the presence of 4 molar equivalents AR peptide. The correlation coefficient r is indicated.

5.4 Binding affinity

To measure the binding affinity of the interaction, we performed NMR titration experiments in which gradually increasing amounts of unlabeled RAP74NMR protein were added to isotopically (^{15}N) labeled AF1*c (see Fig. 5.14). The changes in average ^{1}H and ^{15}N chemical shift with concentration were fitted to equation 5.2, which yielded the binding affinity, K_D.

$$\Delta\delta = \frac{\Delta\delta_{max}}{2} \cdot \left(1 + \frac{[B]_0}{[A]_0} + \frac{K_D}{[A]_0} - \sqrt{\left(1 + \frac{[B]_0}{[A]_0} + \frac{K_D}{[A]_0}\right)^2 - 4 \cdot [B]_0}\right) \tag{5.2}$$

Where $[A]_0$ and $[B]_0$ are the initial concentrations of ^{15}N-AF1*c and RAP74NMR, respectively, and $\Delta\delta$ is the change in average ^{1}H and ^{15}N chemical shift at a given concentration $[B]_0$. The maximum difference in average ^{1}H and ^{15}N chemical shift (between the free and the bound state), $\Delta\delta_{max}$, was treated as a parameter and also fitted, since it was not possible to reach the fully bound state at the highest concentration of RAP74NMR added.

The binding affinity was determined to be 966 ± 39 μM, at 278 K. A global fitting was performed using the residues in the binding site (W433, H434, T435, F437, T438, A439, E440, E441, G442, Q443, L444, Y445 and G446). A single best fit binding affinity K_D and its associated error were estimated by bootstrapping 10% of the data (100 independent calculations). Saturation ($\Delta\delta_{max}$) was fitted individually for each residue. L436 was not used as it is in an overlapped region of the spectrum. The binding curve resulting from the global fitting is shown for each of the residues in the binding site in appendix A.9.

The order of magnitude of the affinity was confirmed by an orthogonal titration in which AR WT peptide was added to ^{15}N-labeled RAP74NMR (see Fig. 5.15). These experiments were also carried out at pH 7.4 to avoid potential different charge states of the histidine side chain of H434 in both titrations. As a consequence of measuring at pH 7.4, we could not observe some peaks of RAP74 because their resonances are not visible at this pH (residues E456, T479, K480, G483, S485, E487 and N508). However, it is not necessary to observe peaks for all residues in a titration experiment as for one binding event the resonances of all residues in the binding site are expected to undergo chemical shift changes that will yield the same binding affinity. The titration experiments were, nevertheless, recorded at 298 K instead of 278 K to obtain good spectral quality. The value for the affinity obtained from this titration (at 298 K) is consequently expected to

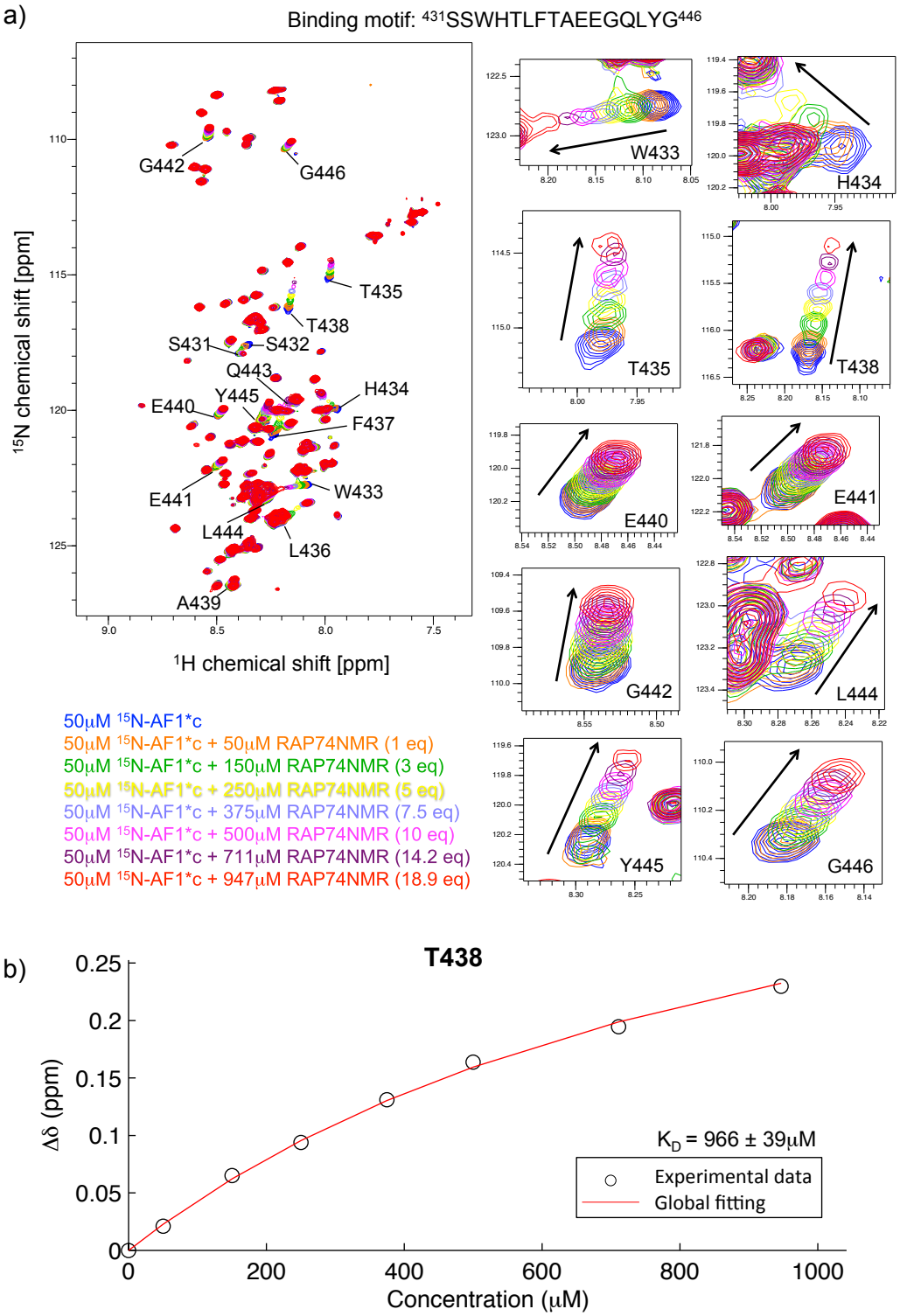


Figure 5.14: a) [^{1}H,^{15}N]-HSQC spectra of ^{15}N-AF1*c with increasing concentrations of RAP74NMR, at 278 K and pH 7.4, to determine the binding affinity (K_D). b) Fitting of the binding curve and corresponding K_D value (shown for AR residue T438). $\Delta\delta$ on the y-axis is the average ^{1}H and ^{15}N chemical shift difference and the concentration of added RAP74NMR is indicated on the x-axis.

be different, although of the same order of magnitude, compared to the one determined from the titration in which the shifts in AF1*c were monitored (at 278 K).

The binding affinity derived from this titration was 1749 ± 60 μM, at 298 K. Also in this case a global fitting was performed for the residues in the binding site (T470, K471, L474, V490, N491, V492, A494, Q495, I496, K498, R499 and N501) and a single best fit K_D and its associated error were estimated by bootstrapping 10% of the data (100 independent calculations). Saturation ($\Delta\delta_{max}$) was fitted individually for each residue. The residues used for the fitting were those for which the average ^{1}H and ^{15}N chemical shift was >0.04 ppm when 10 molar equivalents WT peptide were added to 50 μM RAP74NMR. The binding curve resulting from the global fitting is shown for each of the residues in the binding site in appendix A.9. Consequently, both titrations demonstrated that the interaction between AR and RAP74 has an affinity in the millimolar range, which is extremely low by conventional standards.

Weak protein-protein interactions are known to occur in biology and are often involved in transient processes, such as electron transport, transcription, cell-cell adhesion and enzyme catalysis [436–440]. In these regulatory processes it is essential the dissociation rate is high enough to allow a rapid turnover of the complex [441]. Given that the binding affinity, K_D, is given by the ratio of the dissociation rate constant (k_{off}) and the association rate constant (k_{on}) (see equation 5.3), this has consequences on the specificity and/or the strength of the interaction.

$$K_D = \frac{k_{off}}{k_{on}} \tag{5.3}$$

In order to form short-lived complexes (high k_{off}), either the specificity will be limited (higher k_{on}) or the interaction will be weaker (higher K_D). Even though limited specificity can be advantageous, e.g. for broad substrate specificity during enzyme catalysis [438], specificity needs to be high enough for appropriate signaling. Hence, in general transient interactions are relatively weak. They are often mediated by linear motifs interacting with globular binding partners as this offers advantages to maintain high specificity [305, 442]. In part, these weak interactions remain specific because of the cellular location which excludes the interaction with other partners that are not present [442]. In addition, the flanking regions of linear motifs may also influence the specificity. Analysis of the relative contributions of linear motifs and their flanking regions to the interaction-binding energy has shown that flanking regions contribute 21% of the binding energy on average [443]. Furthermore, the ID nature of linear motifs prior to binding has also been proposed to play a key role in maintaining high specificity in transient interactions [444]. In addition, these weak and transient complexes usually take place in the context of large multisubunit assemblies, in which the simultaneous

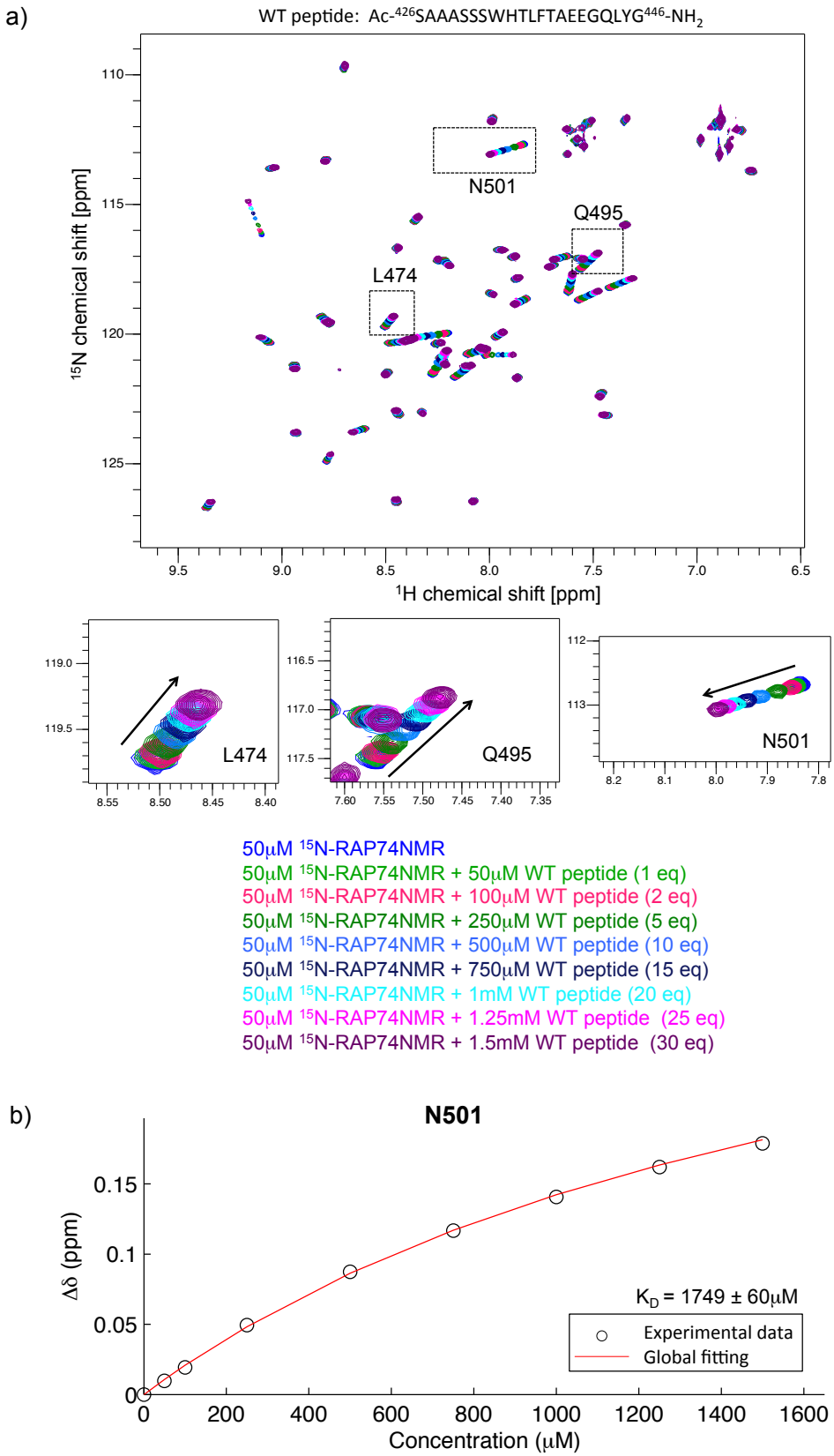


Figure 5.15: a) [^{1}H,^{15}N]-HSQC spectra of ^{15}N-RAP74NMR with increasing concentrations of WT peptide, at 298 K and pH 7.4, to determine the binding affinity (K_D). b) Fitting of the binding curve and corresponding K_D value (shown for RAP74 residue N501). $\Delta\delta$ on the y-axis is the average ^{1}H and ^{15}N chemical shift difference and the concentration of added WT peptide is indicated on the x-axis.

existence of multiple weak interactions gives an inherent stability to the complexes [305, 445, 446].

Our results show that the interaction between the AR and RAP74 is mediated through the ^{433}WHTLF^{437} linear motif. It takes place at initiation of transcription when a complex composed of a multitude of partners is assembled at the promoter. When this interaction takes place in a cell, the AR and RAP74 are both tethered to DNA. As a consequence, the local concentration of both proteins is higher than we can achieve in the *in vitro* experiments where the protein constructs are free to diffuse in solution. It is therefore expected that the binding affinity in the biological milieu is higher than the one determined *in vitro* between both protein constructs. In addition, the interaction between the ^{433}WHTLF^{437} motif of AR and RAP74 likely enhances transcription of AR target genes. Therefore, it is transient by necessity as RNA polymerase II slides over the DNA after transcription initiation, thereby carrying RAP74, which is associated to the polymerase, away from the AR. The affinity between AR and RAP74 is therefore expected to be weak, even in the biological context, in line with typically observed weak and transient interactions of protein-protein interactions involved in transcription [447].

However, the binding affinities of weak and transient protein-protein interactions that have been described and shown to have functional relevance are typically in the high micromolar range [305], and rarely in the millimolar range [439, 441]. It is therefore likely that other factors influence the interaction of AR and RAP74 in the biological context, such as the presence of co-regulators or other binding partners, or PTMs.

5.5 Structural model based on similarity with the RAP74-FCP1 complex

Due to the low binding affinity and the transient nature of the AR-RAP74 interaction, it is challenging to determine the structure of the complex by experimental techniques. However, the RAP74 surface that binds to AR is the same as that which binds to IDRs of FCP1 (transcription factor IIF (TFIIF)-associated CTD phosphatase), a specific phosphatase that dephosphorylates RNA polymerase II at the end of transcription [346].

FCP1 is a 961-residue serine phosphatase that is involved in transcription. For initiation of transcription to occur, the C-terminal domain of the largest subunit of RNA polymerase II must be extensively phosphorylated [448–450]. At the end of elongation, specific dephosphorylation of this subunit of RNA polymerase II by FCP1

recycles the protein rendering it competent to start another cycle of transcription [347, 451]. The interaction between RAP74 and FCP1 strongly stimulates the phosphatase activity of FCP1 [452]. Interestingly, the residues of FCP1 that interact with RAP74 are also disordered in the apo state and adopt a helical conformation when bound to RAP74. There are two regions of sequence of FCP1 known to interact with RAP74 (see Fig. 5.16). One is located at the C-terminus of FCP1 (residues 937–961) and one in the central part of the protein (residues 579–600). For the complex between the C-terminal domain of FCP1 (cterFCP1) and the C-terminal domain of RAP74 (cterRAP74) both a co-crystal structure (PDB code 1J2X, [453]) and an NMR structure (PDB code 1ONV, [454]) are available. An α-helix is formed in FCP1 between residues E945 and M961 in both structures, and the binding affinity has been determined to be $0.5 \pm 0.1\,\mu$M [455]. The interaction between the central region of FCP1 (centFCP1) and the C-terminal domain of RAP74 was described shortly after [456]. It was later found that phosphorylation of threonine residue 584 of FCP1 enhanced the interaction marginally with an increase in binding affinity from $2.1 \pm 0.9\,\mu$M to $0.7 \pm 0.1\,\mu$M [455, 457]. The NMR structure of the complex between the central region of FCP1 with phosphorylated T584 (centFCP1-PO4) and the C-terminal domain of RAP74 is deposited with PDB code 2K7L and shows a helix between FCP1 residues D587 and V598 in the complex [455].

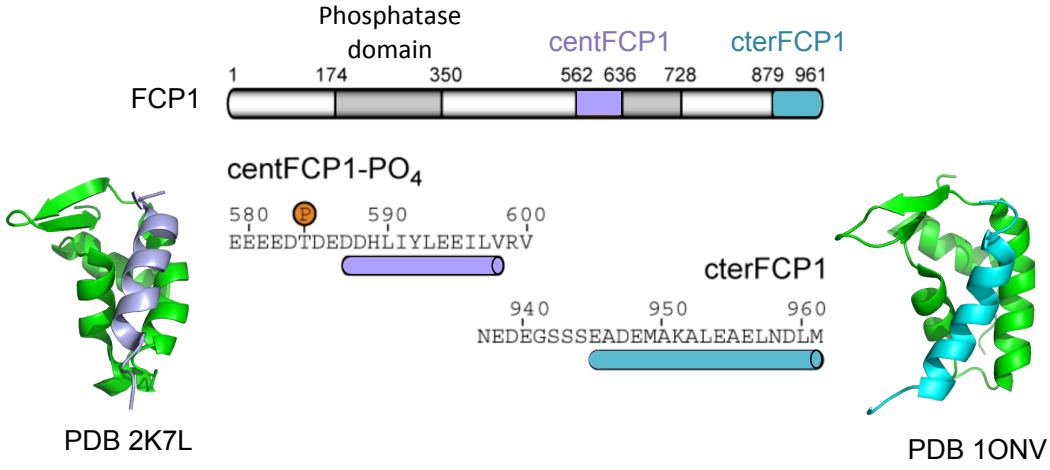


Figure 5.16: Schematic representation of FCP1. The sequence of the regions interacting with the C-terminal domain of RAP74 is indicated as well as the residues forming a helical conformation upon binding (colored cylinder). The structures of both complexes are shown (PDB 1ONV (cterRAP74 + cterFCP1) [454] and PDB 2K7L (cterRAP74 + centFCP1-PO4) [455]). The C-terminal domain of RAP74 is shown in green, and the peptides corresponding to cterFCP1 and centrFCP1-PO4 are shown in cyan and purple, respectively. Figure adapted from [455].

In general, the interaction mode of the two FCP1 motifs for RAP74 is the same and involves the folding of the FCP1 interaction motif into an amphipathic helix that docks

on the hydrophobic surface formed by helix 2 and helix 3 of RAP74. The interaction is further stabilized by two electrostatic contacts between negatively charged residues of FCP1 and positively charged residues on the RAP74 surface (see Fig. 5.17). Importantly, the C-terminal region of FCP1 retains significant flexibility after binding to RAP74, also in the binding motif, as shown both experimentally [458] and with molecular dynamics simulations [459], and in agreement with the high B-factors obtained for the crystal structure of the complex [453]. In addition, nascent α-helical structure has been observed for the C-terminal residues involved in RAP74 binding in the unbound state of FCP1 [460]. Similar studies for the centFCP1-RAP74 complex have not been reported.

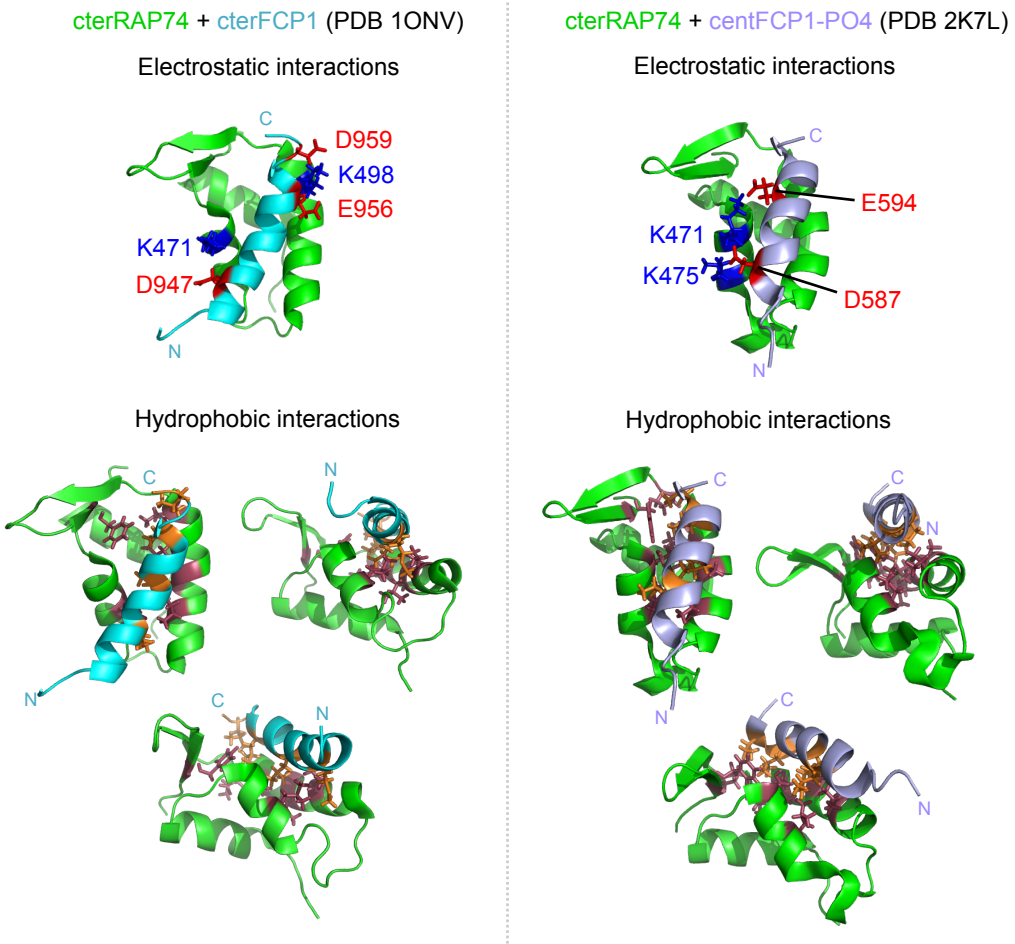


Figure 5.17: The hydrophobic and electrostatic contacts stabilizing the interaction of both RAP74-FCP1 complexes are indicated. Positively charged residues of RAP74 key for interaction are shown in blue, negatively charged residues on the FCP1 peptides that form salt bridges with these residues are shown in red. The hydrophobic groove on the RAP74 surface is shown in raspberry and the hydrophobic residues on the FCP1 peptides are shown in orange. The orientation of the peptides is indicated by N (N-terminus) and C (C-terminus).

Even though the interaction between both of these FCP1 regions with RAP74 is several magnitudes of order stronger than the one we observed between RAP74 and AR, comparison of the sequences revealed similarities between the binding motif in AR and the two FCP1 motifs known to interact with the C-terminal domain of RAP74 (see Fig. 5.18). Given that it is difficult to determine the structure of transient complexes, these similarities can be used to propose a structural model for the AR-RAP74 interaction.

cterFCP1: ^{943}SSEADEMAKALEAELNDLM^{961}
centFCP1: ^{583}DTDEDDHLIYLEEILVRVH^{601}
AR: ^{428}AASSSWHTLFTAEEGQLYG^{446}

Figure 5.18: Similarities in the sequence of the FCP1 motifs that bind RAP74 and the AR motif that binds the same epitope. Hydrophobic residues are shown in green, negatively charged residues are shown in red. Those negatively charged residues of FCP1 that were identified to form salt bridges with positively charged residues in RAP74 are underlined.

The AR binding motif contains several hydrophobic residues and two negatively charged glutamate residues towards the C-terminus. If this sequence would adopt a helical conformation, as suggested by the observed increase in helical propensity (see Fig. 5.7), it would form an amphipathic helix similar to the ones formed by the FCP1 sequences in complex with RAP74. This would yield a helical element with two negatively charged side chains on the same side of the helix as the hydrophobics (as shown in Fig. 5.19). This helix could potentially dock in the same hydrophobic groove offered by helix 2 and 3 of the RAP74 surface and form salt bridges with the positively charged RAP74 surface.

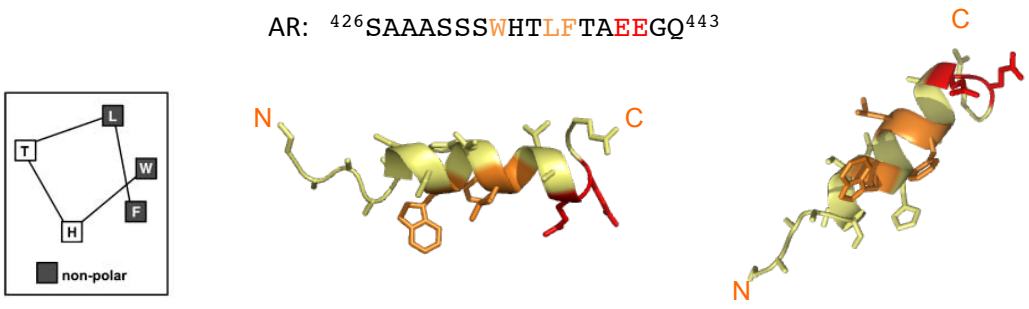


Figure 5.19: The binding motif of AR represented in a helical conformation. Hydrophobic residues are shown in orange and negatively charged glutamate residues E440 and E441 are shown in red. The orientation of the peptide is indicated by N (N-terminus) and C (C-terminus).

5.6 Mutation of key residues for binding disrupts the interaction *in vitro*

To validate this model for the interaction between AR and RAP74 we designed AR mutations to destabilize the complex. If the binding motif of AR adopts a helical conformation upon binding RAP74NMR, a triple hydrophobic mutant (W433A/L436A/F437A) of AR should disrupt the interaction with the hydrophobic surface of RAP74NMR.

To see whether this is the case, we measured HSQC spectra of ^{15}N-RAP74NMR in the absence and presence of 10 molar equivalents of a triple mutant peptide of AR (W433A/L436A/F437A) and compared the chemical shift changes with those caused by the same excess of WT peptide on ^{15}N-RAP74NMR (see Fig. 5.20). The results show that the triple hydrophobic mutant does not bind to RAP74NMR, indicating that the hydrophobic residues W433, L436 and F437 are important for the interaction with RAP74NMR.

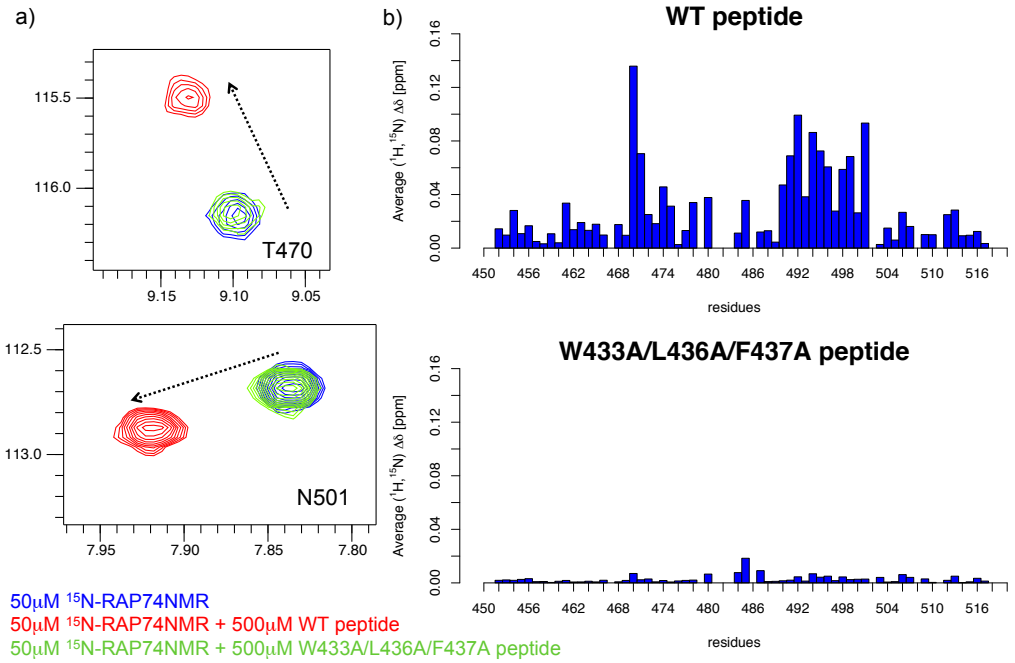


Figure 5.20: a) Blow-up of residues T470 and N501 from [^{1}H,^{15}N]-HSQC of free ^{15}N-RAP74NMR (blue), ^{15}N-RAP74NMR + 10 molar equivalents WT peptide (red) and ^{15}N-RAP74NMR + 10 molar equivalents W433A/L436A/F437A peptide (light green). b) Average ^{1}H and ^{15}N chemical shift changes in RAP74NMR upon binding of the WT peptide (top) and triple mutant peptide (bottom).

Furthermore, the negatively charged side chains of the glutamate residues E440 and E441 might form salt bridges with positively charged residues on the RAP74NMR interaction surface. In order to verify whether this interaction occurs, we designed a

double mutant that reverses both charges (E440K/E441K). We found that the chemical shift changes in the HSQC spectrum of ^{15}N-RAP74NMR with 10 molar equivalents of mutant peptide (E440K/E441K) were negligible compared to those caused by the same excess of WT peptide (see Fig. 5.21). This indicates that these negatively charged residues play a role in the interaction with RAP74NMR.

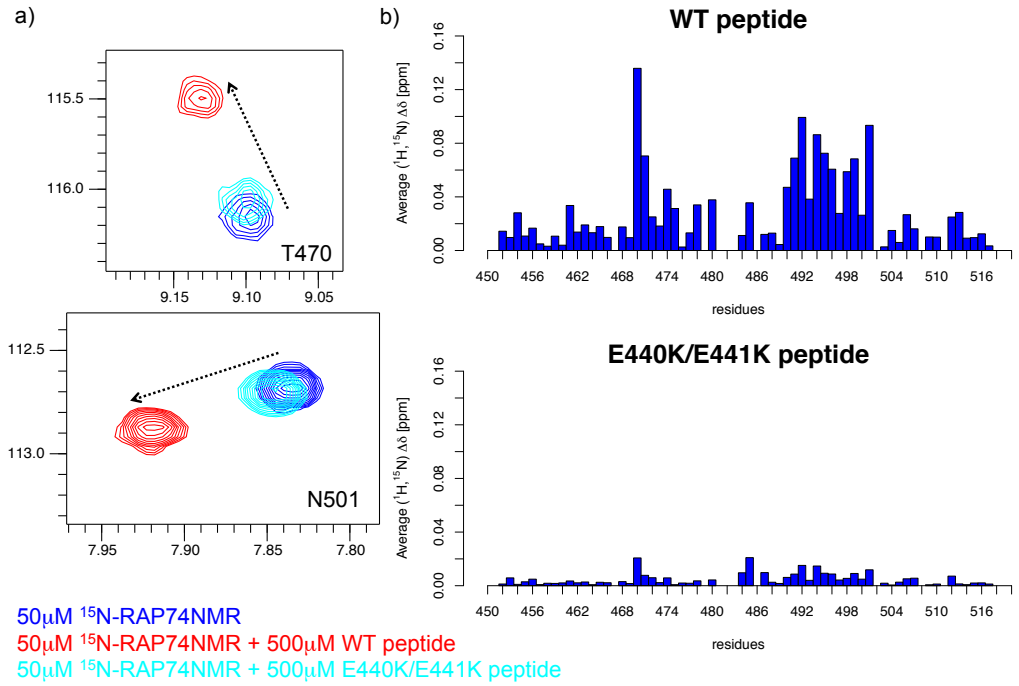


Figure 5.21: a) Blow-up of residues T470 and N501 from [^{1}H,^{15}N]-HSQC of free ^{15}N-RAP74NMR (blue), ^{15}N-RAP74NMR + 10 molar equivalents WT peptide (red) and ^{15}N-RAP74NMR + 10 molar equivalents E440K/E441K peptide (cyan). b) Average ^{1}H and ^{15}N chemical shift changes in RAP74NMR upon binding of the WT peptide (top) and charge reversal mutant peptide (bottom).

This result was corroborated by the reverse experiment in which we compared the effect of 10 molar equivalents RAP74NMR on the chemical shifts of the ^{15}N-labeled double charge reversal mutant (E440K/E441K) of AF1*c with that of the same excess of RAP74NMR on WT ^{15}N-AF1*c (see Fig. 5.22). Also in this case, the chemical shifts observed for the AR mutant are negligible compared to those for the WT sequence. Therefore, our results show that negatively charged glutamate residues E440 and E441 are also critical for the interaction with RAP74NMR.

The results of both mutants are compatible with the proposed structural model. Binding of AR and RAP74 seems to involve hydrophobic side chains of AR (W433, L436, F437) interacting with a hydrophobic surface of RAP74 and an acidic anchoring point of AR (E440, E441) making electrostatic contacts with positively charged residues

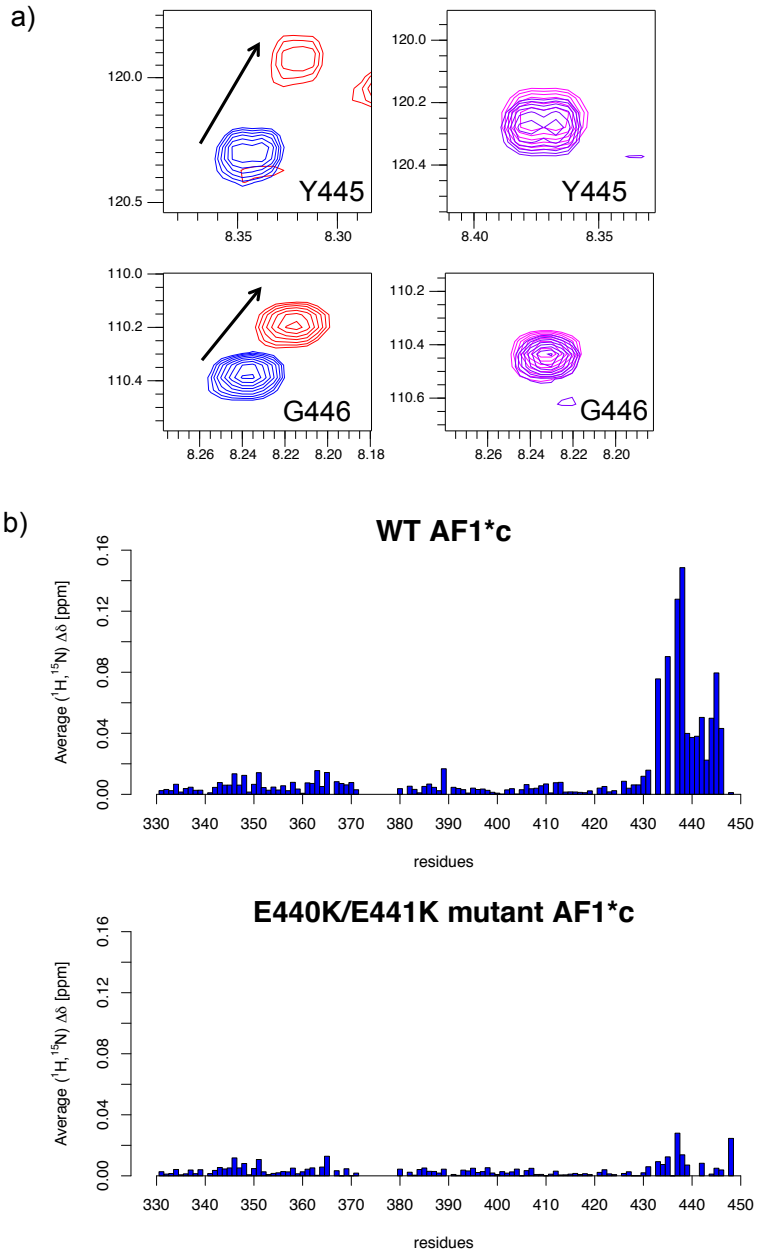


Figure 5.22: a) Blow-up of residues Y445 and G446 from [^{1}H,^{15}N]-HSQC of free ^{15}N-AF1*c (blue) and ^{15}N-AF1*c + 10 molar equivalents RAP74NMR (red), and from [^{1}H,^{15}N]-HSQC of free mutant E440K/E441K ^{15}N-AF1*c (mauve) and mutant E440K/E441K ^{15}N-AF1*c + 10 molar equivalents RAP74NMR (magenta). c) Average ^{1}H and ^{15}N chemical shift changes in WT AF1*c (top) and mutant AF1*c (bottom) upon binding of the RAP74NMR.

of RAP74.

5.7 Orientation of AR peptide when bound to RAP74

To determine the orientation of the AR when it binds to the C-terminal domain of RAP74, we attached an MTSL spin label to a C-terminal cysteine residue that we introduced in the peptide for this purpose (C-Cys peptide, see table 5.1, p. 184). HSQC spectra of RAP74NMR were acquired in the presence of spin labeled AR peptide, both in the oxidized and in the reduced state and the ratio of the peak intensities in both states, $I_{ratio} = I_{ox}/I_{red}$, was calculated. These experiments were carried out at 298 K and at pH 6.5 to obtain good spectral quality. In this case it is important to observe signals for as many peaks as possible to confidently monitor which residues are affected by the spin label, in contrast to the titration experiments for which it was not crucial to observe peaks for all residues. As the MTSL spin label has an effect on the peak intensity of nearby nuclei, this approach should allow us to identify those residues of RAP74NMR that are close to the C-terminus of the AR peptide. The peaks corresponding to RAP74 residues V454, E456, T470 and S485 in the HSQC spectrum measured for the oxidized MTSL label were completely quenched, indicating that these residues are affected most by the presence of the spin label. These residues are marked with a * symbol in figures 5.23 and 5.24.

The experimental PRE data was further analyzed, together with Dr. Santiago Esteban Martín (IRB Barcelona), using SCOOBE, a Monte Carlo (MC) software that uses NMR and small-angle X-ray scattering (SAXS) data as restraints in molecular simulations [461, 462]. SCOOBE builds models of proteins in agreement with experimental data and can be used to calculate the PRE profile of RAP74NMR that would be obtained for a specific position of the MTSL label on the interacting peptide. Or, the other way around, to calculate the position of the MTSL label corresponding to a specific experimentally obtained PRE profile.

Since there is no structural complex available for the interaction of RAP74NMR and the AR peptide, it is not possible to back-calculate the expected PRE profile for the MTSL label placed in the C-terminal cysteine of the AR C-Cys peptide. Comparison of such a back-calculated PRE profile with the experimentally obtained PRE profile would allow to confirm the orientation of the AR peptide when bound to RAP74NMR. Our data, however, suggest that the binding mode of RAP74NMR and the AR is similar to that of the C-terminal domain of RAP74 and the FCP1 interaction motifs. We therefore used the available RAP74-FCP1 complexes to guide the analysis.

First, the available X-ray and NMR complexes of RAP74 and FCP1 peptides were analyzed. These include complexes deposited under PDB codes 1ONV (cterRAP74 + cterFCP1, NMR, [454]), 1J2X (cterRAP74 + cterFCP1, X-ray, [453]) and 2K7L (cterRAP74 + centFCP1-PO4, NMR, [455]). We placed the MTSL label on each amino acid position in the bound FCP1 peptides, one by one, to scan every position. We then computed the expected PRE I_{ratio} for each of the positions and compared them to the experimentally measured data for the interaction of RAP74NMR with the AR C-Cys peptide. Based on the binding affinity for the WT peptide and RAP74NMR (K_D 1.7 mM) and the amount of MTSL labeled peptide (470 μM) and RAP74NMR protein (50 μM) in the experiments, the percentage of bound peptide was set to 20% for the calculations. We found that an MTSL label placed on tyrosine residue 592 of FCP1 in the 2K7L complex best reproduced the measured PRE data (see Fig. 5.23, rmsd 0.24).

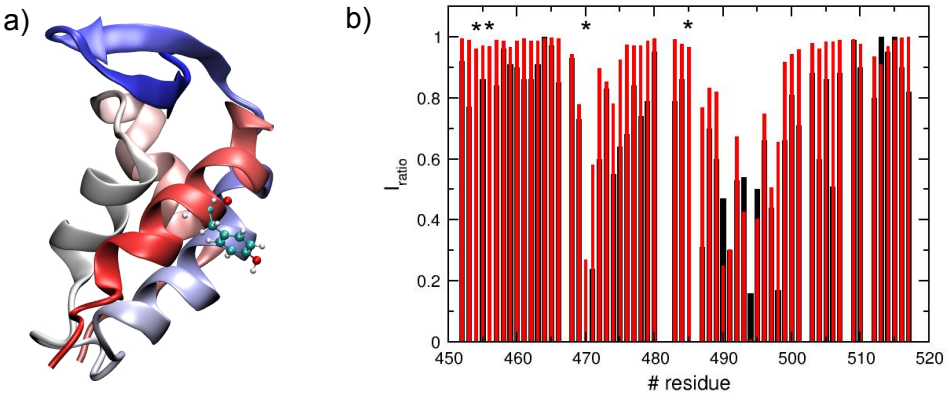


Figure 5.23: a) The structure of the complex between the C-terminal domain of RAP74 and centFCP1-PO4 peptide, PDB 2K7L, [455]. Tyrosine residue 592 on the FCP1 peptide is shown in sticks. b) I_{ratio} of the experimental values obtained for the AR C-Cys peptide with an MTSL label on the C-terminal cysteine and RAP74NMR (black) and calculated values for an MTSL label at position Y592 in the FCP1 peptide for PDB 2K7L (red). Residues V454, E456, T470 and S485 are indicated with the symbol *.

We next performed a MC search using RAP74NMR and the AR C-Cys peptide. The peptide was modeled as a fully folded helix, and only rotations and translations were performed on RAP74NMR and the peptide (rigid body MC search). The MTSL label was attached to the C-terminal cysteine residue and was modeled using ten conformers to account explicitly for structural heterogeneity in the side chain. Although we included the complete peptide in the calculations (percentage bound 20%), in reality the only information encoded in the data is the distance from the MTSL labeled residue to the RAP74NMR protein. In the absence of additional experimental restraints it is, therefore, not possible to determine the structure of the complex. Moreover, the fact that the peptide is not fully folded in its bound state (helical content for AR residues

430–444 is estimated to be 16% for 20% bound peptide) renders the use of PREs to determine the structure of the complex unreliable (e.g. the MTSL label can be located far from the binding site of the peptide). Therefore, we considered it is only reasonable to determine the position of the MTSL label, as this is what is directly measured and shown in the figures.

We performed calculations with one, two, and three peptides to account for multiple binding sites. The corresponding quality of the calculated PRE I_{ratio} as compared to the experimentally measured data (rmsd) was 0.20, 0.17 and 0.15 (see Fig. 5.24). Ensembles with more than three peptides did not improve the fitting. For structure calculation with one peptide, the position of the label matched the one determined from the 2K7L complex (site 1, shown in green) (see Fig. 5.24a). However, when two or three peptides were used in the calculations the position slightly changed, with one site shifted towards the β-strands (site 1, shown in green) and other sites shifted downward (sites 2, shown in purple, and 3, shown in orange) (see Figs. 5.24c and e). Using more than one peptide allowed to better reproduce the details of the observed PRE profile, such as the effect of the MTSL label on the N-terminal residues of RAP74NMR and the slightly larger PRE I_{ratio} values in the middle region (compare Fig. 5.24b, d and f). From our analysis it appears that more than one site needs to be accounted for to accurately describe the location of the PRE label.

This observation could be related to the low binding affinity that possibly allows the peptide to weakly bind the RAP74 surface unspecifically in different orientations. Alternatively, it could indicate a so-called "fuzzy" complex in which several conformations are present in the bound state and/or in which parts of the binding motif adopt a folded conformation upon binding whereas other regions remain disordered in the complex [297].

Importantly, no information on the relative population of each of these sites can be derived from the calculations, as their populations cannot be isolated from changes in the distance of the label to the protein. However, chemical shift analysis (see Fig. 5.9) clearly mapped the binding site closely to the position referred to as site 1 and shown in green in figure 5.24. When one peptide is used in the PRE analysis, the C-terminal residue of the AR C-Cys peptide is predominantly located at site 1, i.e. at a similar position as Y592 in the RAP74-FCP1 complex. A sequence alignment of both proteins in which the C-terminal cysteine residue in the AR C-Cys peptide is aligned with Y592 of FCP1 is represented in figure 5.25. However, because the predominant position of the MTSL label (site 1, shown in green) shifts towards the β-sheet of RAP74NMR when two or three peptides are used in the analysis, the data suggest that the C-terminus of the C-Cys peptide is located close to the β-sheet and in the binding groove. This

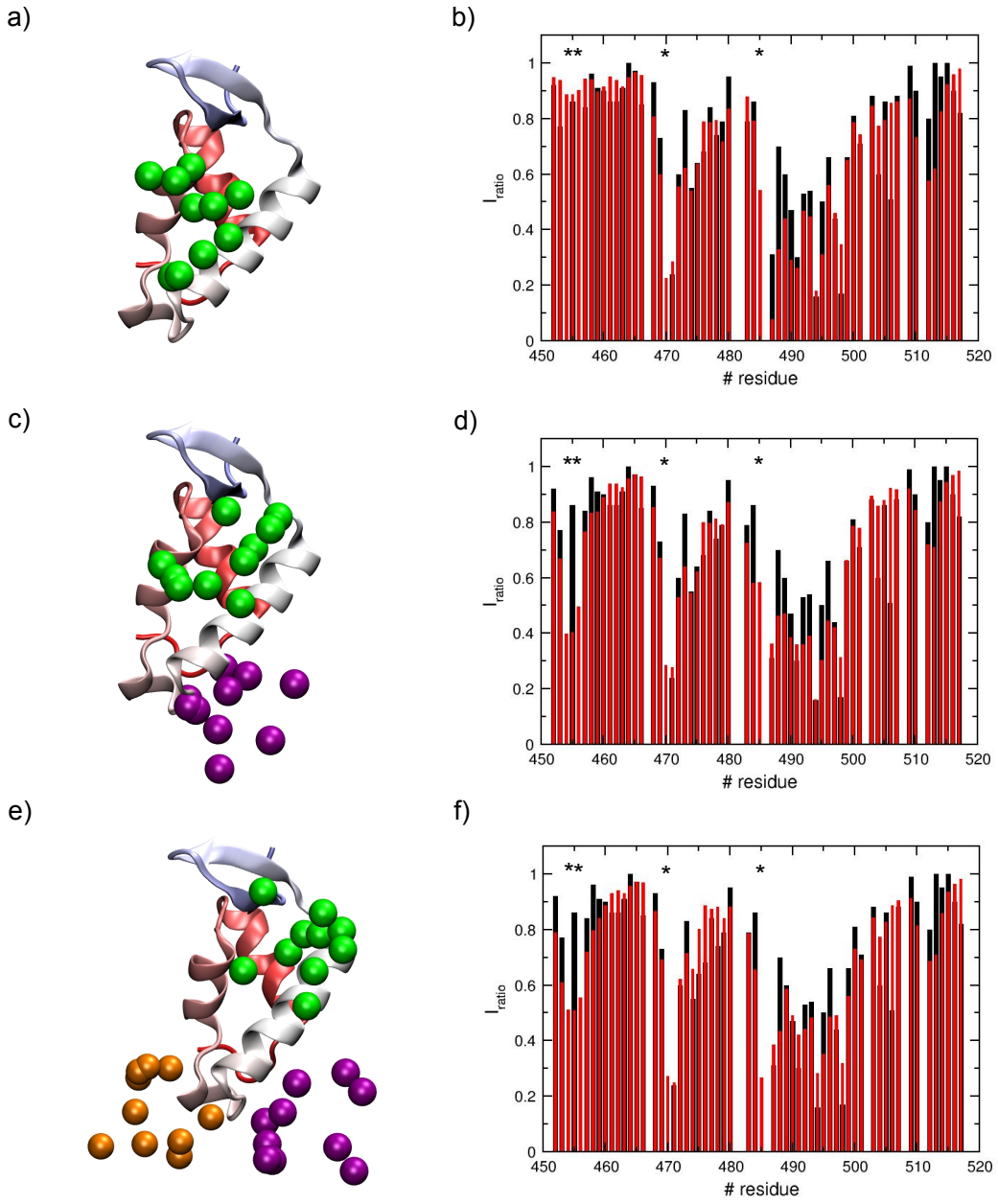


Figure 5.24: a) The position of the MTSL label of AR C-Cys with respect to the C-terminal domain of RAP74 is shown as green spheres for the calculation using one peptide. Ten conformers were used for the calculation to account explicitly for the structural heterogeneity in the side chain, resulting in ten distinct positions of the MTSL label (ten spheres). b) I_{ratio} of the experimental values (black) and calculated values (red) obtained for the AR C-Cys peptide with an MTSL label on the C-terminal cysteine and RAP74NMR, calculated using one peptide (rmsd 0.20). c) As in a) but calculated using two peptides, and ten conformers. The position of the MTSL label on one peptide is represented by green spheres, and on the other peptide by purple spheres (ten conformers of the peptide shown). d) As in b) but calculated using two peptides (rmsd 0.17). e) As in a) but calculated using three peptides, and ten conformers. The position of the MTSL label on the first peptide is represented by green spheres, on the second peptide by purple spheres and on the third peptide by orange spheres (ten conformers of the peptide shown). f) As in b) but calculated using three peptides (rmsd 0.15). Residues V454, E456, T470 and S485 are indicated with the symbol *.

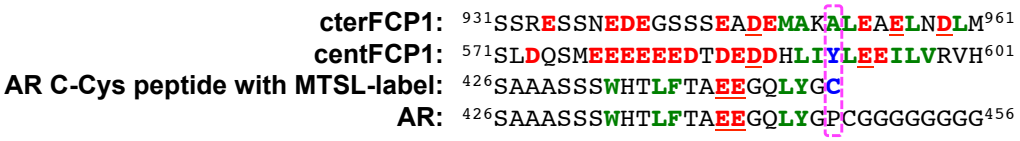


Figure 5.25: Sequence alignment of the AR and FCP1 in which the C-terminal cysteine residue in the AR C-Cys peptide is aligned with tyrosine residue 592 of FCP1. In full-length AR, this position corresponds to proline residue 447. The sequence corresponding to the C-terminal interaction motif of FCP1 is also shown according to the same alignment. Hydrophobic residues are shown in green, negatively charged residues are shown in red. Those negatively charged residues of FCP1 that were identified to form salt bridges with positively charged residues in RAP74 are underlined, as are glutamate residues 440 and 441 of the AR. Tyrosine residue 592 of FCP1 and the C-terminal cysteine residue of the AR C-Cys peptide to which the MTSL label was attached are shown in blue.

favors an orientation of the AR peptide with the C-terminus close to the RAP74NMR β-sheet and the N-terminus close to the loop between helices 2 and 3. Notably, this is the same orientation of the FCP1 peptides when interacting with RAP74.

We further note that the calculations made do not provide any information on the number of AR peptides binding to RAP74NMR. For example, a single bound peptide with a disordered C-terminal part could be responsible for the complete PRE profile, or alternatively a single peptide binding to three sites could result in the observed PRE profile, or three peptides binding simultaneously to the protein, etc.

Interestingly, one of the RAP74NMR residues most affected by the spin labeled peptide is positively charged (K498) which suggests it forms salt bridges with the glutamate residues on the C-terminus of the AR interaction motif (E440 and E441). This would also be compatible with an orientation of the AR peptide in the RAP74NMR binding groove where the AR C-terminus is located close to the RAP74NMR β-sheet.

Taken together, our results indicate that the AR and RAP74 form a transient complex with a similar binding mode to that of both FCP1 interaction epitopes and RAP74. The interaction requires folding of the AR binding motif into a more helical conformation that docks onto the hydrophobic surface offered by helix 2 and helix 3 on the RAP74 C-terminal domain and is stabilized by electrostatic contacts between glutamate residues E440 and E441 of AR and the positively charged RAP74.

5.8 Interaction of the ^{433}WHTLF^{437} motif of AR with RAP74 activates transcription in PC-3 cells

Our finding that the ^{433}WHTLF^{437} motif of AR interacts with the RAP74 C-terminal domain taken together with the previous identification of this motif as the region responsible for enhanced AR transcriptional activity at castrate hormone levels in CRPC cells [232] led to the hypothesis that this interaction enables aberrant transactivation in late stage PCa. To test this hypothesis, it is necessary to determine the effect of the interaction between AR and RAP74 on transcription in CRPC cells at castrate hormone levels. As explained in the introduction, transcription can in some cases be activated through direct contact between transcription factors (AR) and the general transcription machinery (RAP74). If the interaction between AR and RAP74 indeed activates transcription, disrupting this interaction in a cell by mutating key residues for binding is expected to decrease or even abolish transcription.

In collaboration with the group of Prof. Iain McEwan (University of Aberdeen, UK) we performed luciferase-based reporter gene assays to assess the transcription efficiency of full-length AR (wild type and mutants designed to disrupt the AR-RAP74 interaction) in human cancer cells. A schematic representation of these functional assays is shown in figure 5.26. The cells are co-transfected with a (wild type or mutant) AR expression plasmid (or empty plasmid) and a luciferase reporter plasmid. The latter is designed to contain DNA response elements that are recognized by the AR, upstream of a luciferase gene. Binding of the AR to its response elements, upon activation by androgenic hormone, will consequently activate transcription of luciferase instead of that of a natural AR target gene. Transcriptional activity can then be quantified by measuring luminescence emitted by the produced luciferase upon addition of its substrate, luciferin.

In order to compare the transcriptional activity of wild type AR to that of triple hydrophobic (W433A/L436A/F437A) and double charge reversal (E440K/E441K) mutants of full-length AR, plasmids with the full-length AR bearing mutations in these key residues for binding to RAP74 were generated. The constructs were cloned by Dr. Marianela Masin (IRB Barcelona).

The functional assays were carried out in PC-3 cells. This is a human CRPC cell line that does not express endogenous AR. As explained above it is crucial to perform these functional assays in a CRPC cell line and at castrate hormone levels (i.e. $\leq$1 nM). Importantly, because PC-3 cells do not contain endogenous AR, the luciferase read-out is reporting exclusively on the AR (wild type or mutant) introduced in the

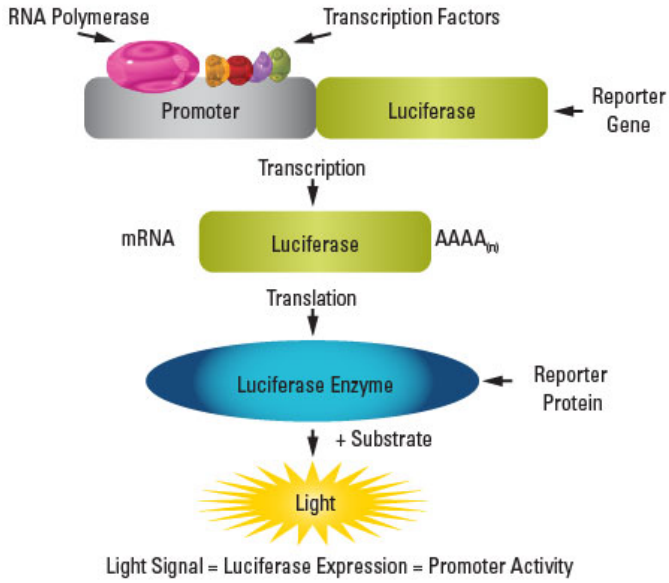


Figure 5.26: Luciferase-based reporter gene assays. See main text for a more detailed explanation. Source: http://www.piercenet.com/method/luciferase-reporters.

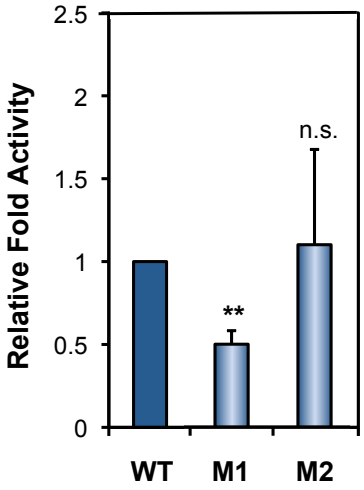
system by transfection. RAP74, as subunit 1 of general transcription factor TFIIF, is endogenously expressed in the cells.

We compared the transcriptional activity of the wild type AR to that of both AR mutants in PC-3 cells at 1.0 nM R1881, a synthetic androgen[2] (see Fig. 5.27). The pPSA-luciferase reporter plasmid was used for our experiments [337].

The charge reversal mutation, which as indicated above shows considerably weaker *in vitro* affinity compared to the wild type, caused a significant decrease in transactivation (**, $p<0.01$). For the triple hydrophobic mutant no significantly reduced transcription was observed. However, the importance of these hydrophobic residues for transcriptional activity is supported by findings by Dehm *et al.*, who reported that the same triple hydrophobic mutant of full-length AR (W433A/L436A/F437A) abolishes Tau-5 mediated transactivation in C4-2 CRPC cells in the absence of hormone [232]. Combined, these results suggest that disruption of the interaction between AR and RAP74 in CRPC cells reduces transcription efficiency, supporting our hypothesis.

Taken together, our data indicate that the key regulatory ^{433}WHTLF^{437} motif in the N-terminal transactivation domain of AR activates transcription by binding directly to the C-terminal domain of RAP74, which forms part of the general transcription machinery. This is a novel transactivation mechanism of the AR N-terminal transactivation domain that involves the ^{433}WHTLF^{437} motif, which is known to be crucial for aber-

[2] http://www.nursa.org (NURSA = Nuclear Receptor Signaling Atlas)



WT: ^{421}GSGSPSAAASSS**WHTLF**TA**EE**GQLYG^{446} (in full-length AR)
M1: ^{421}GSGSPSAAASSSWHTLFTA**KK**GQLYG^{446} (in full-length AR)
M2: ^{421}GSGSPSAAASSS**A**HT**AA**TAEEGQLYG^{446} (in full-length AR)

Figure 5.27: Transcriptional activity of the wild type (WT) and mutant (E440K/E441K (=M1) and W433A/L436A/L437A (=M2)) AR in PC-3 cells. Cells were transfected in triplicate with either empty vector, AR-WT, AR-M1 or AR-M2 together with a pPSA-luciferase reporter gene and treated with 0 or 1.0 nM R1881 for 24 hours. The mean fold activation of the mutant receptors in plotted relative to WT (1) (± standard deviation) for at least three independent experiments. **, $p<0.01$, n.s. not significant (student t-test).

rant transactivation in late stage PCa. Importantly, PC-3 cells are CRPC cells, and the capacity of the ^{433}WHTLF^{437} motif to activate transactivation by interacting with RAP74 in these cells demonstrates that the interaction between these two partners enables *aberrant* transactivation at castrate levels of hormone in currently incurable CRPC. Targeting this interaction might therefore lead to new therapeutic approaches for CRPC in the future.

5.9 Possible regulation of this transactivation mechanism through PTMs

It is difficult to reconcile the clear functional effect observed in the cellular assays described above with the weak millimolar binding affinity observed *in vitro*. The functional assays revealed that the interaction is not as weak in the cells as in the NMR tube, thereby suggesting that one or several elements that influence the binding were not present in our *in vitro* experiments.

The comparison of the AR and FCP1 binding motifs (see Fig. 5.28) demonstrated a notable difference between the two interaction partners of RAP74, namely the presence of a large number of negatively charged side chains in FCP1 N-terminal to the hydrophobic residues, which are absent in AR. In addition, inspection of the complexes between FCP1 and RAP74 revealed that some of these residues are involved in salt bridge interactions that stabilize the complex (see Fig 5.28).

The interaction between the FCP1 motifs and RAP74 is stabilized by two electrostatic contacts, involving negatively charged residues in FCP1 both N-terminal and C-terminal to the hydrophobic residues in the core of the interaction site. In contrast, in the AR binding motif the only negatively charged side chains are located C-terminal to the hydrophobic residues, which indicates AR interaction with RAP74 is not stabilized by electrostatic contacts at both sides of the hydrophobic residues, but instead only by one anchoring point formed by E440 and E441. Instead of aspartates or glutamates, in AR there are three consecutive serine residues (S430, S431 and S432) N-terminal to the hydrophobic residues in the motif. The phosphorylation of one or several of these residues would introduce negative charges and, potentially, increase the affinity of the complex.

In addition, when we compared the chemical shift changes in RAP74NMR upon interaction with AR WT peptide and those reported for interaction of the C-terminal domain of RAP74 and cterFCP1 [463], we found that both peptides overall cause changes in the same residues of RAP74, but with a clear difference in the loop region between

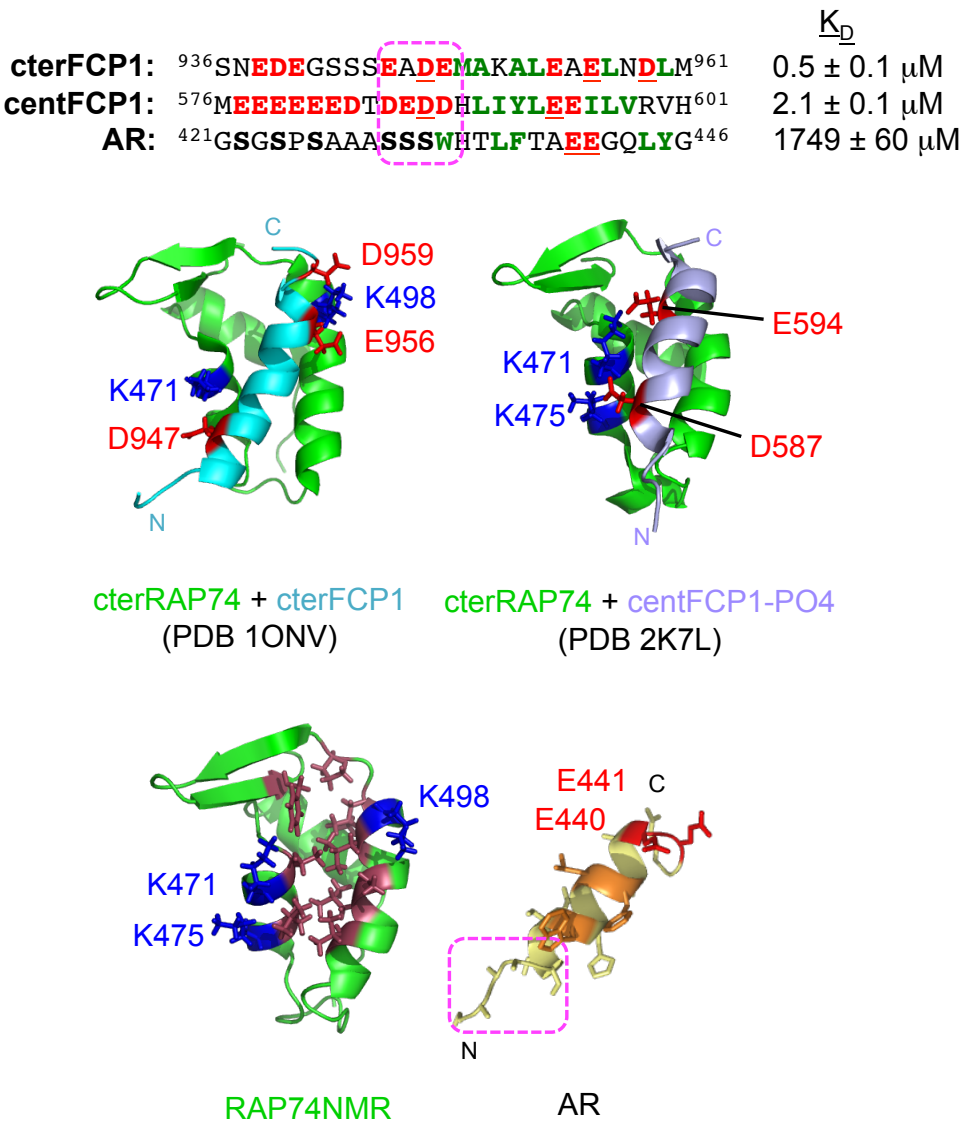


Figure 5.28: Comparison of the RAP74 binding motif of AR with those of FCP1 (C-terminal and central). The binding affinities for each of the complexes is indicated. The NMR structures of the complexes are shown below (cterRAP74 + cterFCP1: PDB 1ONV, [454], cterRAP74 + centFCP1-PO4: PDB 2K7L, [455]). Positively charged RAP74 residues key for binding are highlighted in blue and negatively charged FCP1 residues key for binding are shown in red, and are underlined in the sequences shown above. The apo structure of the C-terminal domain of RAP74 is represented by PDB structure 1I27 [350] with hydrophobic residues in the binding pocket shown in raspberry, and the positively charged residues that are key for binding of FCP1 peptides shown in blue. The peptide of the AR binding motif is represented in a helical conformation with hydrophobic residues marked in orange and negatively charged glutamate residues E440 and E441 in red, and underlined in the sequence shown above. The orientation of the peptides is indicated with N (N-terminus) and C (C-terminus).

helix 2 and helix 3 (see Fig. 5.29). To be able to compare with the reported data, the chemical shift difference was calculated based on the ^{1}H, ^{15}N and ^{13}CO chemical shifts of free RAP74 and RAP74 in complex with the peptide, according to equation 5.4 [463, 464]. In this equation $\Delta\delta_{N^H}$ corresponds to the amide nitrogen chemical shift change, $\Delta\delta_{CO}$ to the carbonyl chemical shift change and $\Delta\delta_{H^N}$ to the amide proton chemical shift change. The reported data correspond to 1 mM ^{15}N-cterRAP74 + 1 mM cterFCP1 peptide, in 20 mM sodium phosphate buffer with 1 mM EDTA and were measured at pH 6.5 and 300 K [463]. The experimental data for the interaction of RAP74NMR and WT peptide were obtained from 300 μM ^{15}N-RAP74NMR + 1.2 mM WT peptide, in 20 mM sodium phosphate buffer and were measured at pH 6.5 and 298 K. Due to the difference in affinity for both complexes, the data for the two peptides are plotted on separate axes in figure 5.29.

$$\Delta\delta = \sqrt{(0.17 \cdot \Delta\delta_{N^H})^2 + (0.39 \cdot \Delta\delta_{CO})^2 + (\Delta\delta_{H^N})^2} \tag{5.4}$$

The RAP74NMR residues in this loop region are notably affected by the cterFCP1 peptide, whereas they experience smaller chemical shift changes upon addition of the AR peptide. One possibility is that this discrepancy is related to the fact that the cterFCP1 peptide is stabilized in the complex by two charge clamps, one on each side of the hydrophobic motif, whereas the AR peptide is only attached by one electrostatic anchoring point between negatively charged E440 and E441 and positively charged side chains of RAP74NMR. Phosphorylation of one or several of the serine residues N-terminal to the hydrophobics could create an additional electrostatic anchoring point at the N-terminal part of the AR binding motif, further stabilizing the complex. This observed discrepancy in chemical shift changes for the residues in the loop region of RAP74 is also in agreement with the proposed orientation of the AR peptide when it is bound to RAP74, i.e. with the N-terminus closer to the loop region between helix 2 and 3 and the C-terminus closer to the β-sheet of the C-terminal domain of RAP74.

To date, none of these three serine residues of AR have been reported to be phosphorylated ([83, 465], and see figure 1.21 in the introduction p. 50). Nevertheless, NetPhos v. 2.0, a predictor of phosphorylation sites in eukaryotic proteins[3], predicts that S430 has a high probability of being phosphorylated (phosphorylation potential 0.906). In contrast, S431 and S432 are predicted to have a phosphorylation potential of only 0.129 and 0.362, respectively. Notably, of the twelve reported phosphorylated serine residues in AR [83, 465], ten are predicted by NetPhos v. 2.0, seven of these with a phosphorylation potential larger than 0.9.

[3] http://www.cbs.dtu.dk/services/NetPhos/; [466]

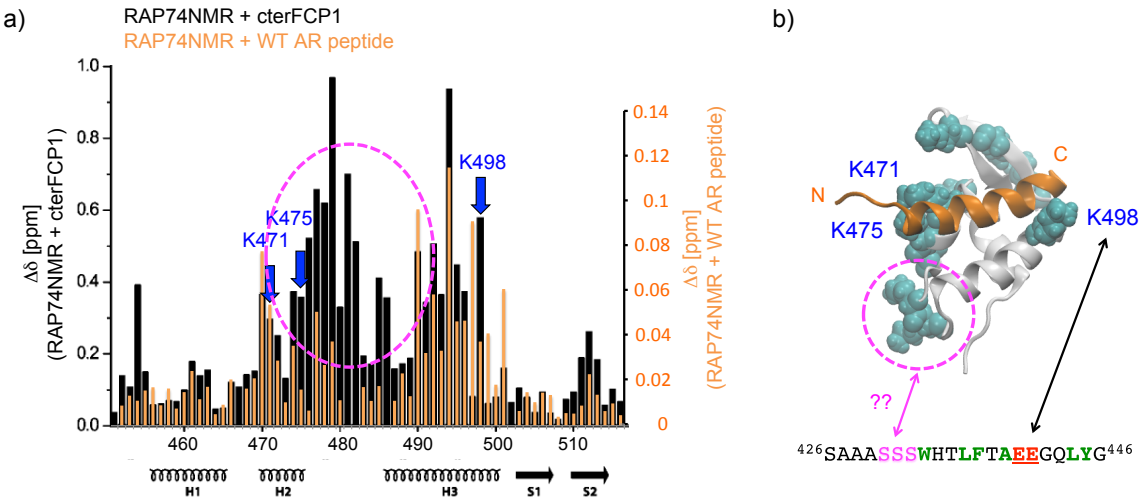


Figure 5.29: a) Chemical shift changes in the C-terminal domain of RAP74 caused by the binding of cterFCP1 peptide (black) or AR peptide (orange). The chemical shift difference calculated is a combination of ^{1}H, ^{15}N and ^{13}CO chemical shifts, according to equation 5.4. The structural elements in the C-terminal domain of RAP74 are indicated below the plot. b) A structural model of the AR-RAP74 complex is represented as the final structure of a coarse grain simulation carried out by Dr. Santiago Esteban Martín (IRB Barcelona), with positively charged residues of RAP74NMR shown in blue space filling, and the AR peptide in ocher. Putative electrostatic contacts are indicated with arrows.

It is important to note that the fact that these residues have not been reported to be phosphorylated does not necessarily mean they cannot be phosphorylated in specific cell lines and under a specific set of conditions. In a recent review [465] it was emphasized that there may be considerable differences in the AR phosphorylation pattern between androgen-dependent cells and androgen-independent CRPC cells. Most phosphorylation studies on AR have, however, been performed using full-length AR in androgen-dependent cell lines. Little is known about the PTMs of full-length AR in CRPC cell lines or of AR splice variants lacking the LBD [465].

The AR motif that binds with RAP74 is fairly conserved in different species (see Fig. 5.30). This is not only true for the hydrophobic residues and glutamates E440 and E441 that we identified to be key residues for binding, but also for the serines N-terminal to the hydrophobic residues.

The serine residue closest to the binding motif reported to be phosphorylated is S424 [467]. This is also the sole residue identified to date to be phosphorylated within the Tau-5 region [465]. However, no studies have identified a role or the kinase responsible for S424 phosphorylation [465].

To validate the hypothesis that the phosphorylation of one or several of the serine residues N-terminal to the hydrophobic residues in the binding motif increases the

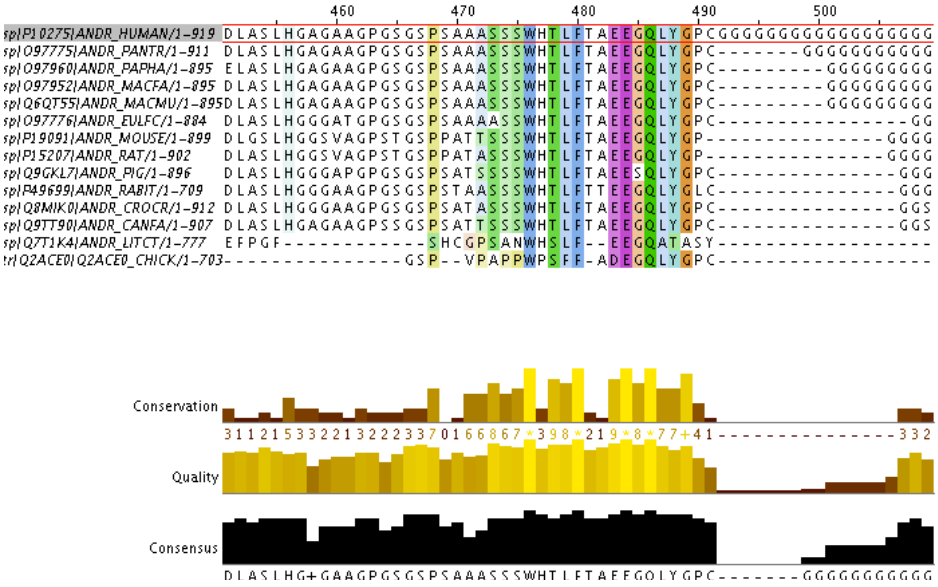


Figure 5.30: Sequence alignment of the AR of different species, colored by conservation.

binding affinity, we investigated the interaction of RAP74NMR and several peptides that were phosphorylated at those specific serines (pS430, pS431 and pS432, see table 5.1 on p. 184). Although S424 is not directly located in the binding site of AR for RAP74, we also examined how the phosphorylation of this residue affects the interaction (pS424, see table 5.1 on p. 184). For this purpose, we measured HSQC experiments of RAP74NMR in the presence of 10 molar equivalents of each of the peptides, including the WT peptide (unphosphorylated), and analyzed the chemical shift differences caused by each of them (see Fig. 5.31). The introduction of a phosphate group at S424 had little effect, whereas phosphorylation of S430, S431 or S432 considerably increased the binding affinity between AR and RAP74 as evidenced by the larger chemical shift changes caused by these peptides compared to the WT peptide.

The resonances of RAP74NMR affected by phosphorylated peptides pS424, pS430, pS431 and pS432, undergo chemical shift changes in the same direction as for the wild type peptide along a straight line. The fact that they are larger for the phosphorylated peptides indicates that the binding event is the same for both WT and phosphorylated peptides but that there is a difference in binding affinity. Because the increase in affinity obtained by phosphorylating S424 was marginal compared to that of the other phosphopeptides, we did not explore the effect of the phosphorylation at S424 on RAP74 binding in more detail.

Contrary to what we expected, introduction of a phosphate group at S430, S431 or S432 did not cause larger chemical shift changes in the loop region of RAP74NMR

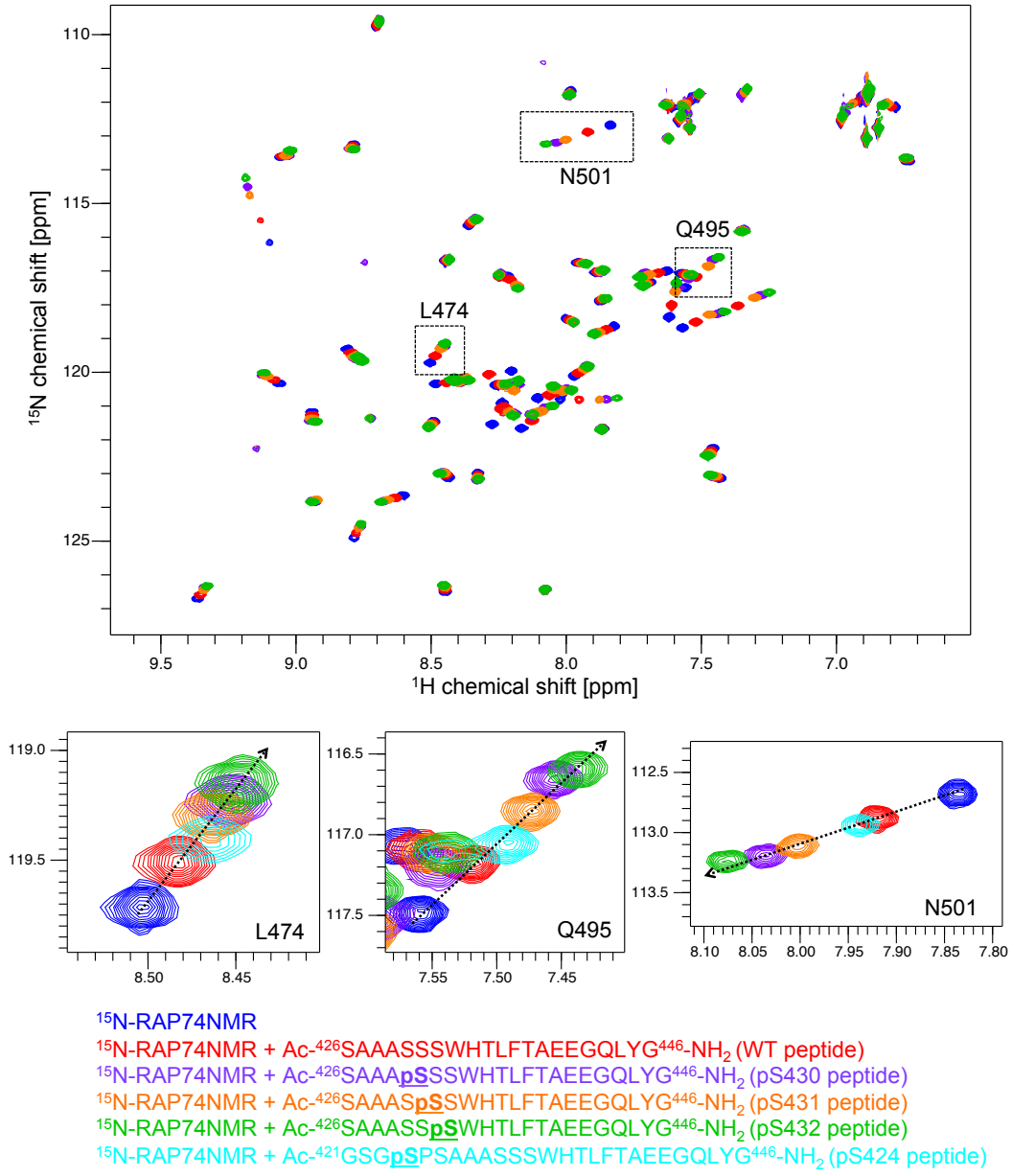


Figure 5.31: [^{1}H,^{15}N]-HSQC of free ^{15}N-RAP74NMR (blue), ^{15}N-RAP74NMR + 10 molar equivalents WT peptide (red), ^{15}N-RAP74NMR + 10 molar equivalents pS430 peptide (mauve), ^{15}N-RAP74NMR + 10 molar equivalents pS431 peptide (orange), ^{15}N-RAP74NMR + 10 molar equivalents pS432 peptide (green), and ^{15}N-RAP74NMR + 10 molar equivalents pS424 peptide (cyan). Blow-ups of RAP74 residues L474, Q495 and N501 are shown.

(residues 476–485) as shown in figure 5.32. We anticipated that phosphorylation of one or several of these residues would create an additional electrostatic anchoring point N-terminal to the hydrophobic core of the binding motif, similar to for the RAP74-FCP1 interaction. Although the addition of negative charges in the AR peptide enhanced the binding affinity, no larger chemical shift changes were observed for positively charged residues of the RAP74 loop region.

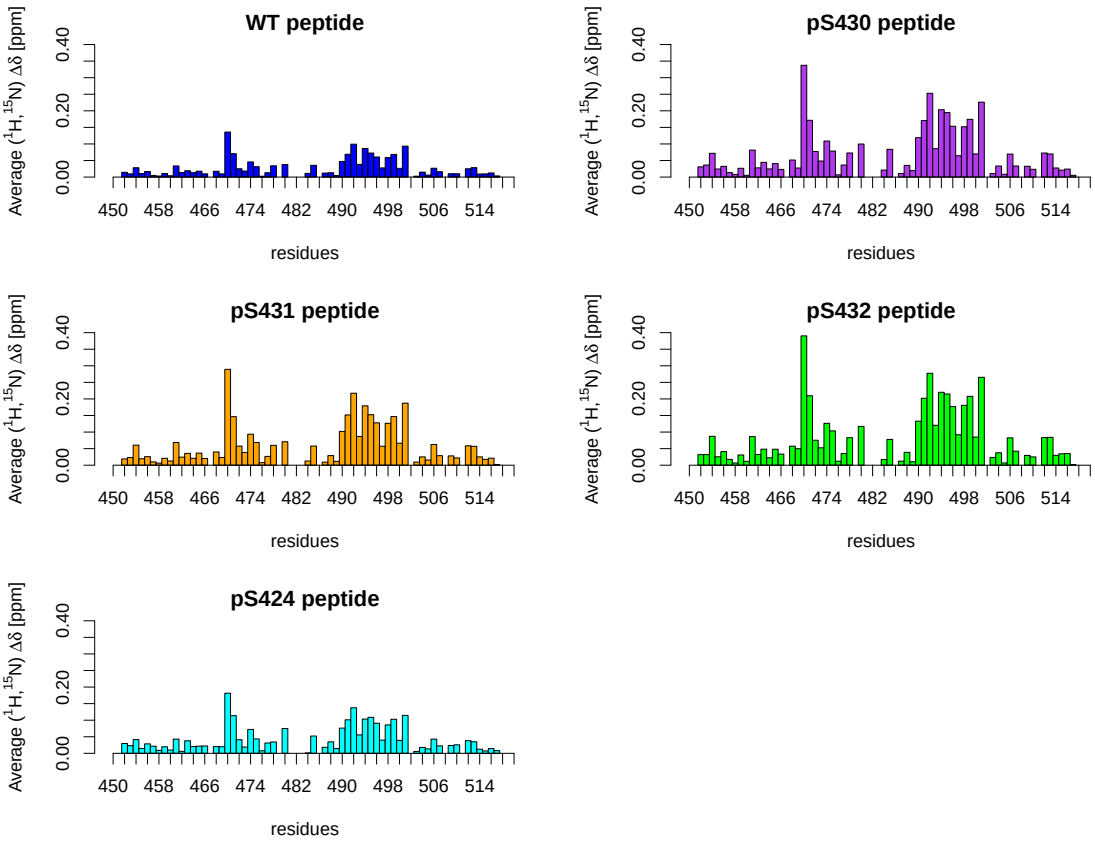


Figure 5.32: Average ^{1}H and ^{15}N chemical shift changes in RAP74NMR upon addition of 10 molar equivalents WT peptide (blue), pS430 peptide (mauve), pS431 peptide (orange), pS432 peptide (green) or pS424 peptide (cyan). No larger chemical shift perturbations are observed in the loop region of RAP74NMR (residues 476–485) for the phosphorylated peptides compared to the WT peptide.

The phosphorylation of S432 increased the binding affinity with RAP74 to the largest extent, evidenced by the fact that the corresponding peptide caused the largest chemical shift changes (see Fig. 5.31). We determined the binding affinity of the pS432 peptide by titrating it into ^{15}N-RAP74NMR, at 298 K and pH 7.4 (see Fig. 5.33). The same experimental conditions as those for measuring the binding affinity of WT peptide and RAP74NMR were used to be able to compare both values.

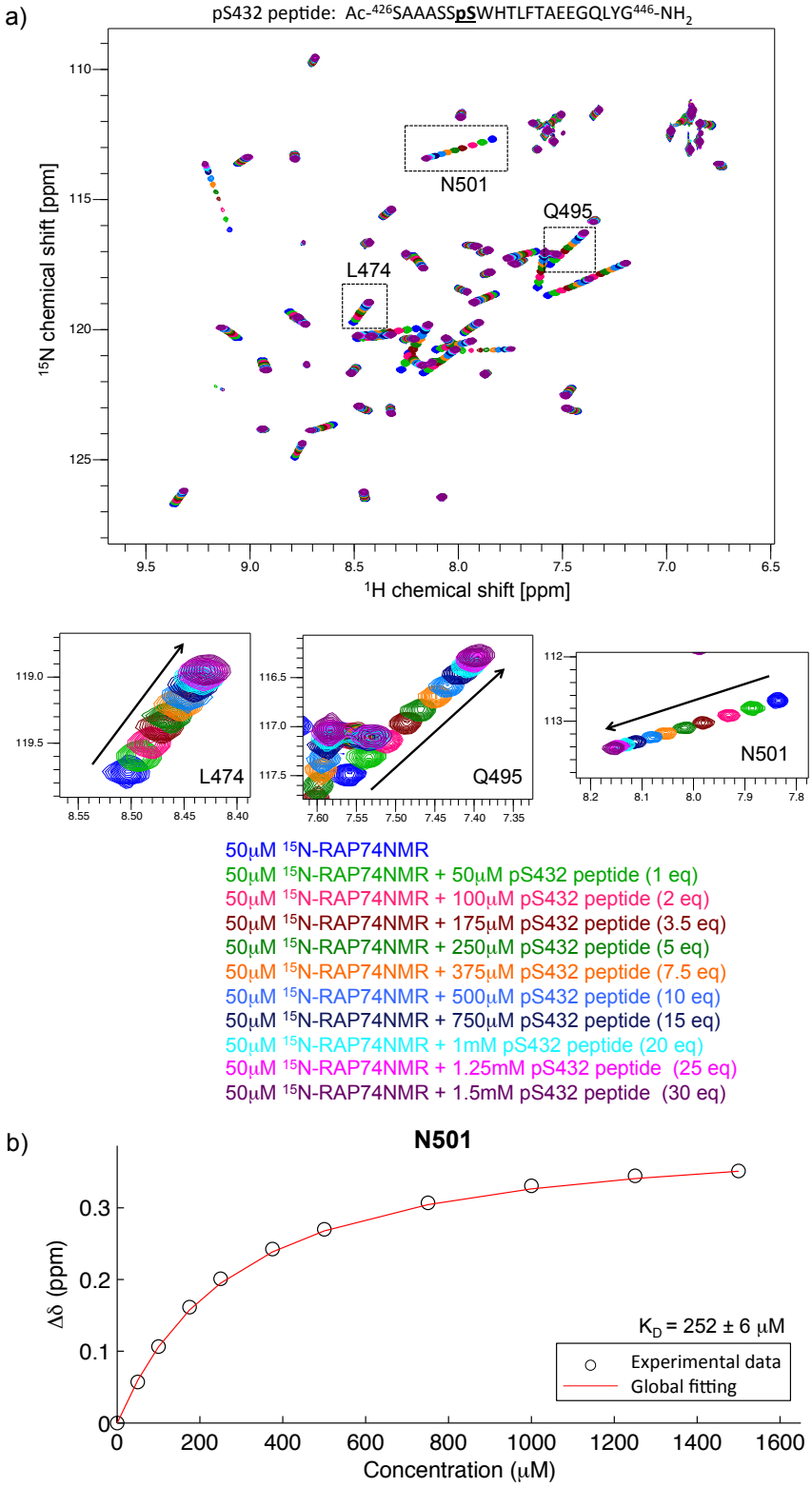


Figure 5.33: a) [^{1}H,^{15}N]-HSQC spectra of ^{15}N-RAP74NMR with increasing concentrations of pS432 peptide, at 298 K and pH 7.4, to determine the binding affinity (K_D). b) Fitting of the binding curve and corresponding K_D value (shown for N501). $\Delta\delta$ on the y-axis is the average ^{1}H and ^{15}N chemical shift difference and the concentration of added pS432 peptide is indicated on the x-axis.

The binding affinity of the pS432 peptide for RAP74NMR was found to be 252 ± 6 μM. A global fitting was performed for the residues in the binding site, which are the same ones as for the titration with the WT peptide, and a single best fit K_D and its associated error were estimated by bootstrapping 10% of the data (100 independent calculations). Saturation ($\Delta\delta_{max}$) was fitted individually for each residue. The binding curve resulting from the global fitting is shown for each of the residues in the binding site in appendix A.9. Consequently, the phosphorylation of this residue caused an increase in the affinity of AR and RAP74 of approx. one order of magnitude.

This increase in binding affinity cannot be attributed to the creation of an additional electrostatic anchoring point as we initially anticipated (see Fig. 5.32). It has been shown that phosphorylation of threonine or serine can affect the helix-forming propensity of proteins [468, 469]. The effect of phosphorylation on α-helical conformation has been reported to be either stabilizing [469–472] or destabilizing [468, 469, 473, 474]. Interestingly, phosphorylation seems to have a stabilizing effect when it occurs at the N-termini of regions with α-helical propensity [469, 472]. To determine whether phosphorylation of any of the serines modulated the intrinsic propensity of the AR binding motif to stabilize an α-helical conformation, we measured far UV-CD spectra of the unphosphorylated peptide and each of the phosphorylated peptides (see Fig. 5.34).

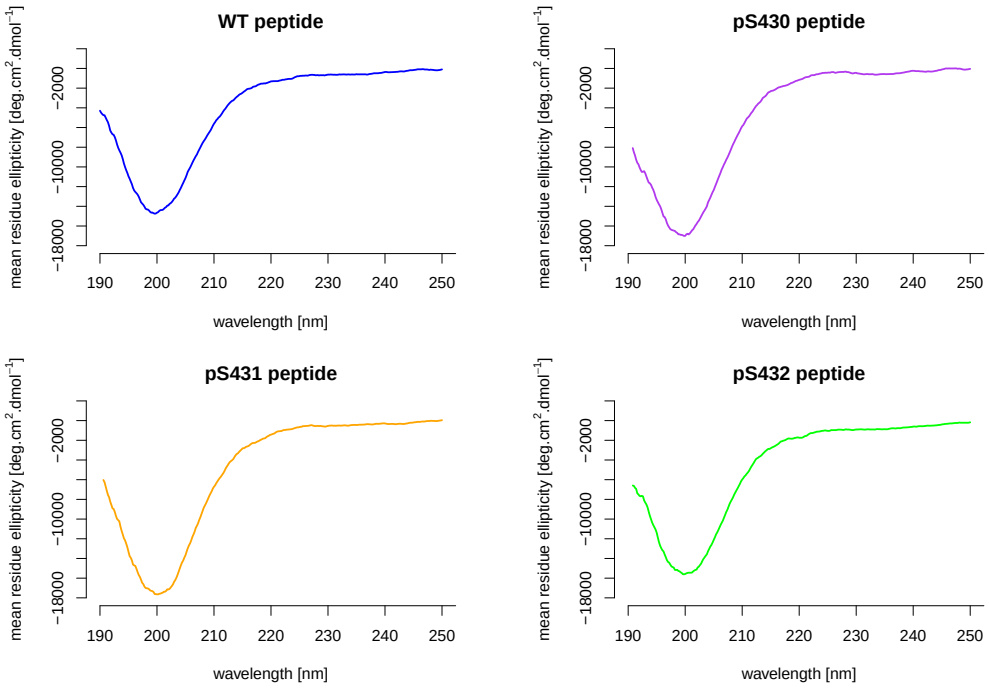


Figure 5.34: Far UV-CD spectra of WT AR peptide (unphosphorylated), pS430 peptide, pS431 peptide and pS432 peptide.

The CD spectra show that phosphorylation at position 430, 431 or 432 has no measurable effect on the helical propensity of the AR peptide of the binding motif, indicating that the role of the phosphate group in increasing the binding affinity is not to induce helicity at the interaction site. This suggests that the underlying reason for the higher affinity of the phosphorylated peptides is an overall increase in electrostatic attraction between both binding partners after introduction of a phosphate group at S430, S431 or S432.

Given that phosphorylation of any of these three serines increased the binding affinity, although to different extents, the exact position of the phosphorylated residue does not seem to be critical. From a structural point of view this is also comprehensible if the main role of the phosphate group is to increase overall electrostatic attraction. If this is indeed the case, phosphorylation of more than one of the three serines simultaneously could increase the binding affinity even further. To test this notion, a triple phosphomimetic mutant peptide of AR (S430E/S431E/S432E, see table 5.1 on p. 184) was titrated into ^{15}N-RAP74NMR to determine its binding affinity, at 298 K and pH 7.4 (see Fig. 5.35).

This triple phosphomimetic mutant of AR was found to have a binding affinity of $166 \pm 2\,\mu$M. Again, a global fitting was performed for the residues in the binding site, which are the same ones as for the titration with the WT peptide, and a single best fit K_D and its associated error were estimated by bootstrapping 10% of the data (100 independent calculations). Saturation ($\Delta\delta_{max}$) was fitted individually for each residue. The binding curve resulting from the global fitting is shown for each of the residues in the binding site in appendix A.9. Interestingly, the binding affinity only increased to some degree compared to that of any of the single phosphorylation mutants at the same serine residues.

Although phosphomimetic mutants are widely used to study the effect of phosphorylations, there are considerable differences between those mutants and phosphorylated amino acids. Glutamate and phosphoserine display both structural and electrostatic differences as can been seen in figure 5.36. The predominant phosphorylation state of a phosphoserine in a protein at physiological pH results in net charge -2, [475] whereas a glutamate residue only carries one negative charge at this pH. Given that the main contribution of phosphorylation in increasing the affinity of AR for RAP74 seems to be related to electrostatic attraction, mimicking the charge state of the amino acids involved correctly is highly relevant.

For most eukaryotic proteins, phosphosites tend to occur at dense clusters of serine/threonine residues. In fact, for proteins with two or more phosphosites, 54% of all

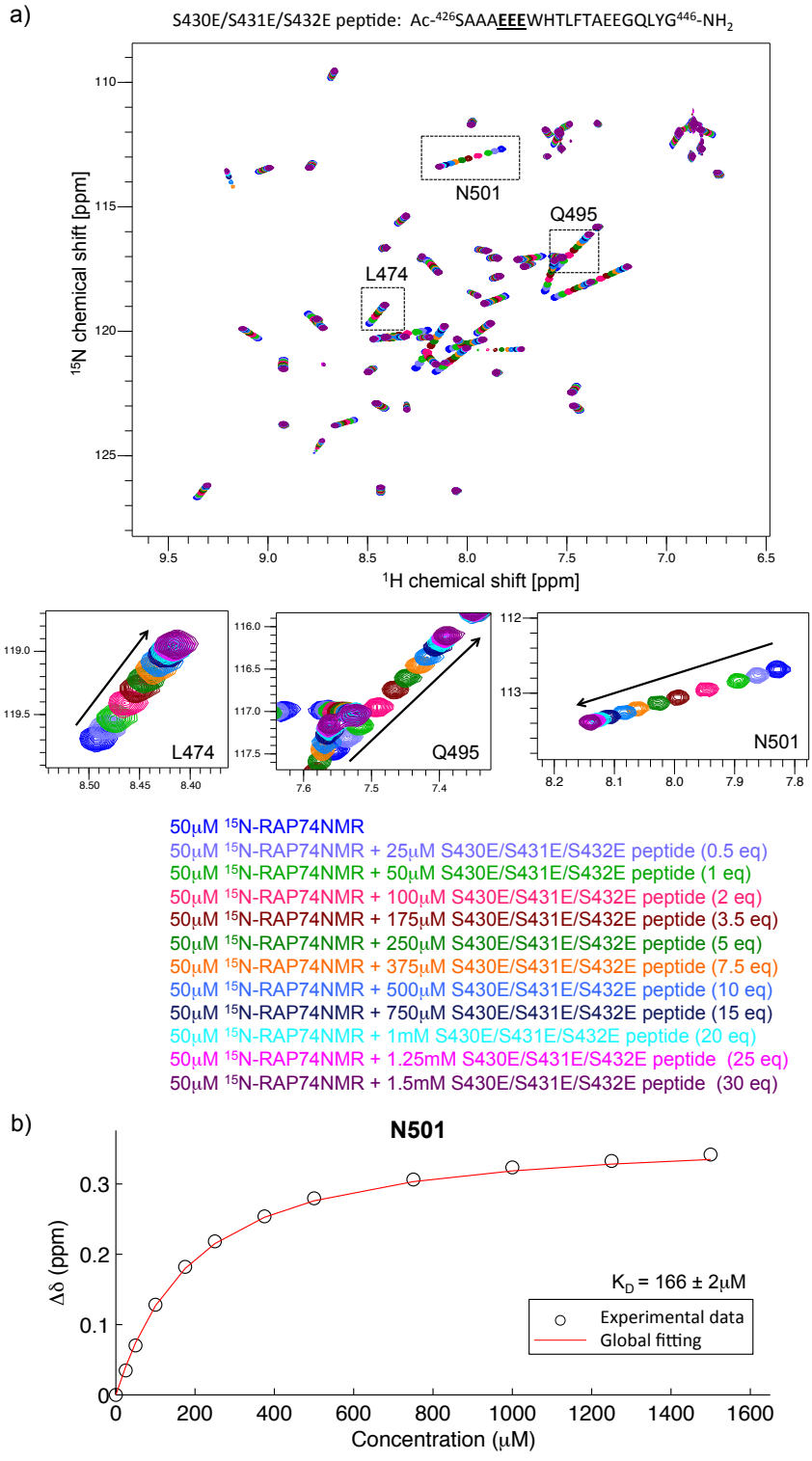


Figure 5.35: a) [^{1}H,^{15}N]-HSQC spectra of ^{15}N-RAP74NMR with increasing concentrations of S430E/S431E/S432E peptide, at 298 K and pH 7.4, to determine the binding affinity (K_D). b) Fitting of the binding curve and corresponding K_D value (shown for N501). $\Delta\delta$ on the y-axis is the average ^{1}H and ^{15}N chemical shift difference and the concentration of added S430E/S431E/S432E peptide is indicated on the x-axis.

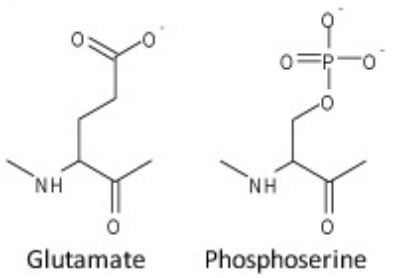


Figure 5.36: Comparison of the amino acid glutamate, which is widely used as phosphomimetic, and phosphoserine.

pS/pT sites are within four amino acids of another site and most likely to occur at amino acids at positions i and i+2 in humans [476]. On the basis of these observations, together with the data in figure 5.31, which suggest that the phosphorylation of S430 and of S432 have the largest effect on the binding affinity, we hypothesized that the simultaneous phosphorylation of S430 and S432 has a synergistic effect, thereby further increasing the binding affinity.

To test this hypothesis, we performed an NMR titration of a double phosphorylated peptide of AR (pS430/pS432, see table 5.1 on p. 184) and RAP74NMR to determine its binding affinity (see Fig. 5.37). We found that the simultaneous phosphorylation of S430 and S432 further increased the binding affinity to $81 \pm 4\,\mu$M. A global fitting was performed for the residues in the binding site, which are the same ones as for the titration with the WT peptide, and a single best fit K_D and its associated error were estimated by bootstrapping 10% of the data (100 independent calculations). Saturation ($\Delta\delta_{max}$) was fitted individually for each residue. The binding curve resulting from the global fitting is shown for each of the residues in the binding site in appendix A.9.

The binding affinity of this mutant was corroborated by microscale thermophoresis (MST). Since the peptide of the AR binding motif contains a tryptophan (W433), a tyrosine (Y445) and a phenylalanine residue (F437), the peptide has intrinsic fluorescence. The C-terminal domain of RAP74 does not contain any tryptophan residues, and only one tyrosine residue (Y462) which is located at the opposite side of the binding interface with AR (see Fig. 5.38). There are also two phenylalanine residues present in RAP74 (F477 and F513), but the contribution of phenylalanine to fluorescence is negligible compared to that of tryptophan and tyrosine. Since we can make use of the intrinsic fluorescence of the AR peptide, this technique offers the possibility to determine the binding affinity without the need to immobilize either of the binding partners and without the need to label either of them, at room temperature and pH 7.4.

As shown in figure 5.39, we monitored the changes in fluorescence when adding different concentrations of RAP74NMR (non-fluorescent) to a constant concentration

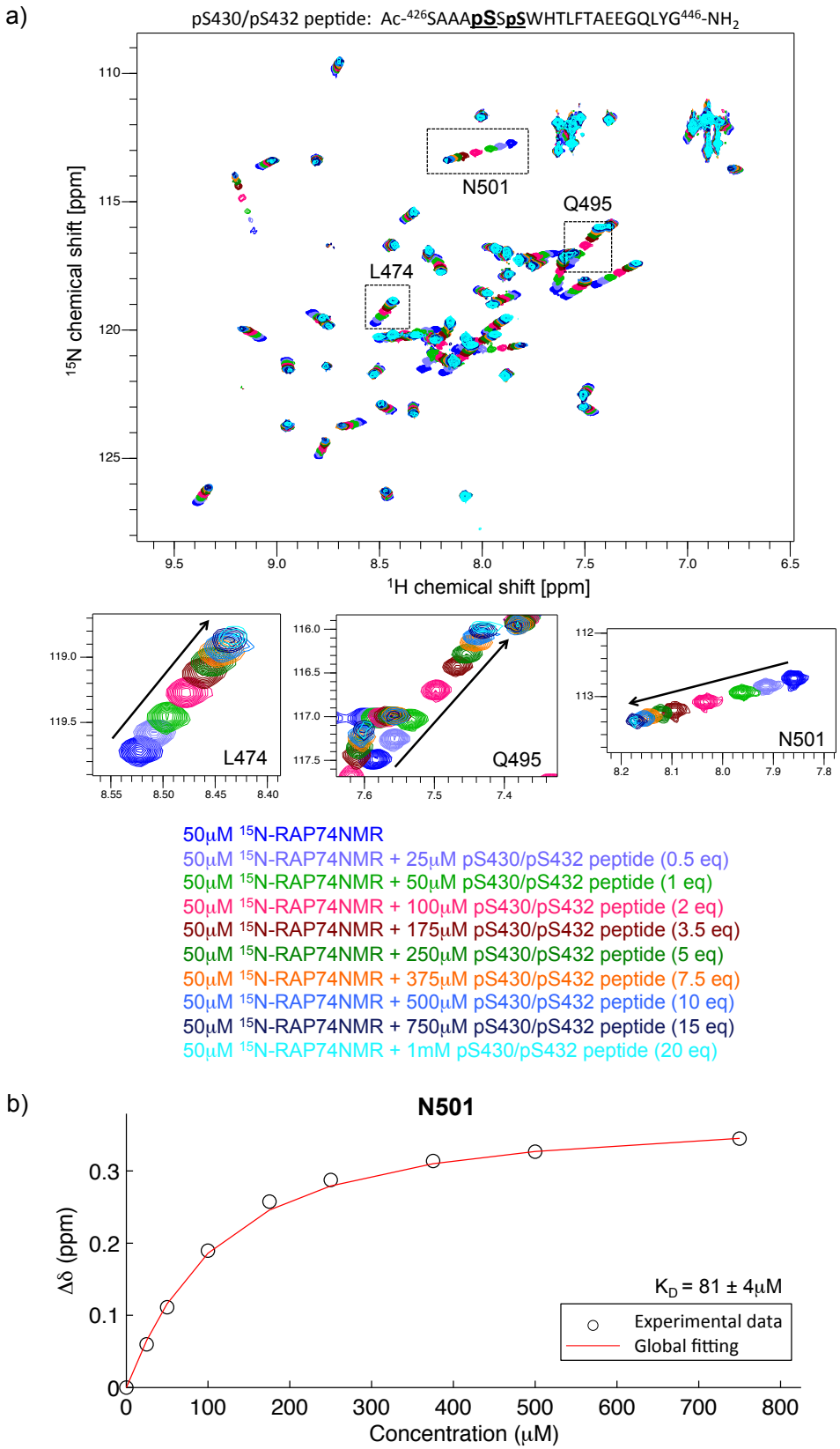


Figure 5.37: a) [^{1}H,^{15}N]-HSQC spectra of ^{15}N-RAP74NMR with increasing concentrations of pS430/pS432 peptide, at 298 K and pH 7.4, to determine the binding affinity (K_D). b) Fitting of the binding curve and corresponding K_D value (shown for N501). $\Delta\delta$ on the y-axis is the average ^{1}H and ^{15}N chemical shift difference and the concentration of added pS430/pS432 peptide is indicated on the x-axis.

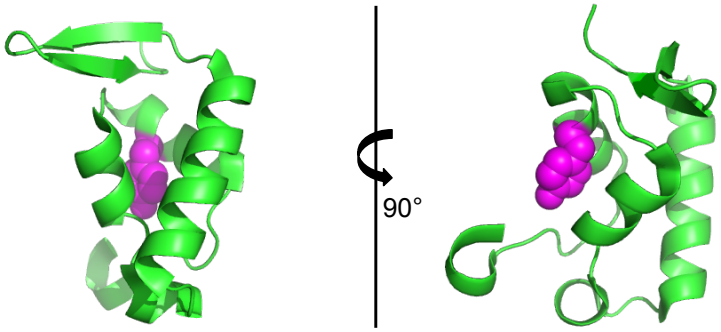


Figure 5.38: The C-terminal domain of RAP74 represented by PDB structure 1I27 ([350]). Tyrosine residue Y462 is located on the opposite side to where AR-binding takes place, and is shown in magenta.

of AR peptide (fluorescent). A binding affinity of 34.6 μM was obtained. This value is compatible with the affinity obtained by NMR.

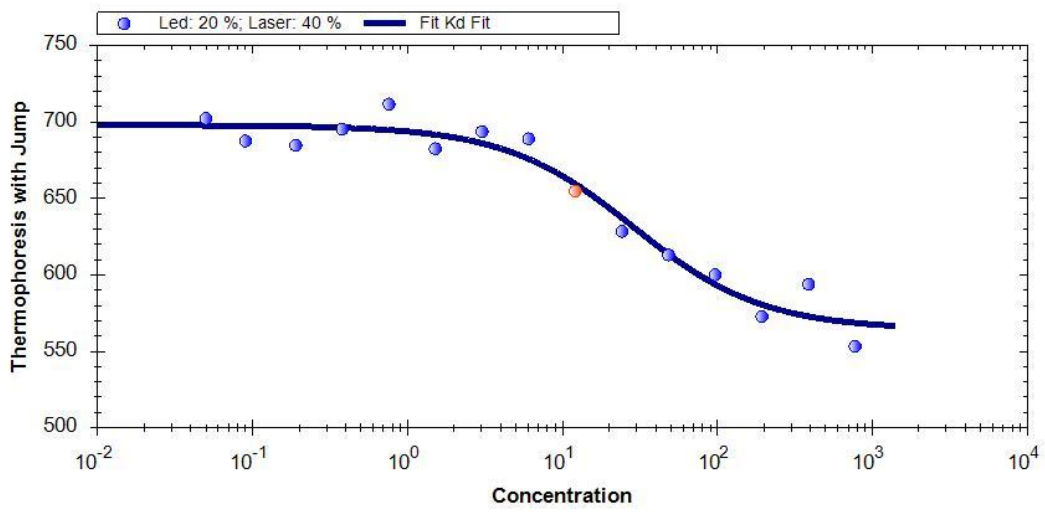


Figure 5.39: Label-free microscale thermophoresis experiment to obtain the binding affinity between RAP74NMR and the pS430/pS432 peptide. We kept the concentration of peptide constant (5 μM) while the concentration of the non-fluorescent RAP74NMR protein was varied between 1565 μM and 48 nM. The concentration of RAP74NMR is plotted on the x-axis in μM and the measured thermophoresis is plotted on the y-axis.

The binding affinity was obtained by fitting of the experimental data to equation 5.5, where $[A]_0$ and $[B]_0$ are the initial concentrations of AR pS430/pS432 peptide and RAP74NMR, respectively. The maximum difference in thermophoresis (between the free and the bound state), $\Delta thermo_{max}$, was treated as a parameter and also fitted.

$$\Delta thermo = \frac{\Delta thermo_{max}}{2} \cdot \left(1 + \frac{[B]_0}{[A]_0} + \frac{K_D}{[A]_0} - \sqrt{\left(1 + \frac{[B]_0}{[A]_0} + \frac{K_D}{[A]_0}\right)^2 - 4 \cdot [B]_0}\right) \tag{5.5}$$

Even though the double phosphorylation of serine residues 430 and 432 caused a marked increase in binding affinity compared to that of the WT AR sequence (1749 $\pm$ 60 μM to 35 μM–81 μM), a micromolar affinity in this range is still considered a low affinity. Nevertheless, we propose that an *in vitro* binding affinity of less than 100 μM is compatible with a weak and transient interaction between AR and RAP74 in the context of the cell, considering both proteins are tightly associated to the DNA when binding occurs in the biological context.

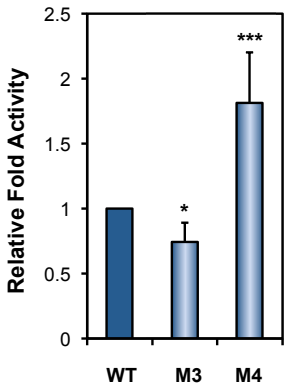
In addition, it is important to keep in mind that the ^{433}WHTLF^{437} motif only adopts a promoting role in AR transactivation under very specific conditions (CRPC cells, castrate levels of hormone, full-length AR) [232]. If this aberrant transactivation is, in part, mediated through the interaction of the ^{433}WHTLF^{437} motif with RAP74, as suggested by our data, this interaction might occur only in specific cell lines and under specific conditions rather than being a general mechanism.

Our results demonstrate that the phosphorylation of one or several of the serine residues N-terminal to the hydrophobic residues of the ^{433}WHTLF^{437} motif of AR strongly enhances the binding between AR and RAP74 *in vitro*. This finding suggests that the phosphorylation of these serine residues is one mechanism to activate transactivation mediated by AR-RAP74 interaction in cells. However, as mentioned above, it is not known whether the corresponding serine residues are phosphorylated in AR in cells.

The effect of phosphorylation in PC-3 cells was studied using reporter gene assays in which serine residues S430, S431 and S432 of the full-length AR had been mutated simultaneously to glutamate (phosphomimetic) or alanine (phosphorylation-resistant) residues (see Fig. 5.40).

If the phosphorylation of these serine residues in AR is critical for high affinity binding between AR and RAP74, the mutation of these serines to alanines is expected to result in a decrease in affinity and therefore less efficient transactivation. However, the mutation of the three serines to glutamates could result either in increased affinity, and therefore enhanced transactivation, or in no considerable change in affinity, which would indicate that one or several of the serine residues under study were endogenously phosphorylated.

We found that the triple mutation of serine to alanine resulted in a significantly decreased transactivation compared to the wild type (*, $p<0.05$), whereas the triple glutamate mutant significantly enhanced transactivation (***, $p<0.001$). The results imply that one or several of the three serine residues immediately N-terminal to the ^{433}WHTLF^{437} motif were phosphorylated in PC-3 cells, but not all three simultaneously. This further suggests that the phosphorylation of S430, S431, S432 or a combination of these could (in part) account for the discrepancy between the functional effects observed in PC-3 cells and the weak binding affinity observed *in vitro* for the wild type AR and RAP74. These results further show that the presence of negative charges at all three serine residues results in significantly increased transcription compared to the WT, indicating that combined phosphorylation of these three residues, if it would occur, might enhance transactivation even more.



WT: ^{421}GSGSPSAAA**SSS**WHTLFTAEEGQLYG^{446} (in full-length AR)
M3: ^{421}GSGSPSAAA**AAA**WHTLFTAEEGQLYG^{446} (in full-length AR) → phosphorylation-resistant
M4: ^{421}GSGSPSAAA**EEE**WHTLFTAEEGQLYG^{446} (in full-length AR) → phosphomimetic

Figure 5.40: Transcriptional activity of the wild type (WT), phosphorylation-resistant (S430A/S431A/S432A) and phosphomimetic (S430E/S431E/S432E) mutant AR in PC-3 cells. Cells were transfected in triplicate with either empty vector, WT, or mutant AR together with a pPSA-luciferase reporter gene and treated with R1881 (0 and 1.0 nM) for 24 hours. The mean transcriptional activity of the mutant receptors is plotted relative to WT (1) (± standard deviation) for at least three independent experiments. *, $p<0.05$, ***, $p<0.001$ (student t-test).

Overall, our data indicate that the phosphorylation of serine residues near to the ^{433}WHTLF^{437} motif is one possible mechanism by which AR transactivation is regulated, since the introduction of negatively charged groups where the serines are located enhanced the binding affinity by several orders of magnitude *in vitro* as well as significantly increased the transcription activity in cellular assays in CRPC PC-3 cells. The interaction is likely only biologically relevant when it is promoted, either by PTMs or through other mechanisms. Possibly this transactivation mechanism becomes essential

for the survival of CRPC cells and does not occur in healthy cells. If this would be the case, preventing the phosphorylation of these serines is a potential therapeutic strategy to selectively target aberrant transactivation in late stage PCa patients.

To understand if phosphorylation occurs in specific PCa cell lines, and if so, which residues are phosphorylated, we are investigating the phosphorylation pattern of AR constructs in several cell lysates by NMR in collaboration with Dr. Philipp Selenko (FMP, Berlin). It is improbable phosphorylation only occurs at S430 and S432 when AR binds to RAP74 in the biological context, due to the peculiar sequence nearby the three consecutive serine residues, which is unlikely to be recognized by a (common) kinase. The direct phosphorylation of S430, S431 or S432 is therefore improbable. Instead, a serine residue closeby (such as S424) could be phosphorylated by a prime kinase, which in turn might trigger the phosphorylation of nearby serine residues (such as S430, S431, S432, S422 and/or S426). Such phosphorylation cascades have been described to take place in other regulatory events [477] and in the case of AR would result in a highly negatively charged sequence N-terminal to the hydrophobics in the core motif for binding to RAP74, much like the highly acidic regions found N-terminal to the hydrophobic core motifs of FCP1 for RAP74-binding. The transient, and often reversible nature of PTMs allows them to play a major role as regulators of cellular processes [305], but at the same time hinders their experimental detection. We therefore anticipate it will be challenging to determine the phosphorylation pattern of the AR binding motif in the biological milieu.

It is likely that the *in vitro* binding affinity between AR and RAP74 would be further increased to the low micromolar range if the AR phosphorylation pattern corresponds to that present in the biological context.

5.10 Fuzzy complex or one well-defined bound conformation?

We could not determine an experimental structure for the AR-RAP74 complex because of the weak binding affinity (order of magnitude millimolar). In addition, it is not clear whether there is only one bound conformation or whether a “fuzzy” complex is formed with not one but several bound conformations and/or in which parts of the binding motif adopt a folded conformation upon binding whereas other regions remain disordered in the complex [297].

The way the line broadening changes during a titration between two binding partners allows to determine whether a single bound conformation exists or whether instead

multiple conformations are present in the bound state. During any titration between two binding partners line broadening is expected to occur as a result of the chemical exchange between the free and the bound states. When saturation is reached, only the bound state is populated and consequently the peaks are expected to become sharp again. This is true if there is one bound conformation. However, if multiple bound conformations exist, an additional contribution to the line broadening is caused by the conformational exchange in the bound state. Hence, when such a complex is formed, the peaks are not expected to become sharp again when saturation is reached. Therefore gradually increasing line broadening with increasing concentration of peptide is indicative of conformational exchange due to structural heterogeneity in the bound state.

For the titration of RAP74NMR with the WT peptide, due to the low binding affinity, only 40% of saturation was reached at the highest concentration of peptide added (30 molar equivalents). Therefore, it is not clear whether the observed line broadening was merely due to chemical exchange between the free and the bound states or whether there was an additional contribution from conformational exchange. However, much higher populations of the bound state were reached with the phosphorylated or phosphomimetic peptides (with 30 molar equivalents: 85% with pS432, 90% with S430E/S431E/S432E and 95% with pS430/pS432). For these peptides, we observed that the lines of most of the resonances gradually broadened upon addition of peptide and did not become sharp again, even when >80% of saturation was reached (see Fig. 5.41). This unequivocally shows that the RAP74 resonances are in intermediate exchange during the titration with AR peptide, indicating that more than one conformation exists in the bound state.

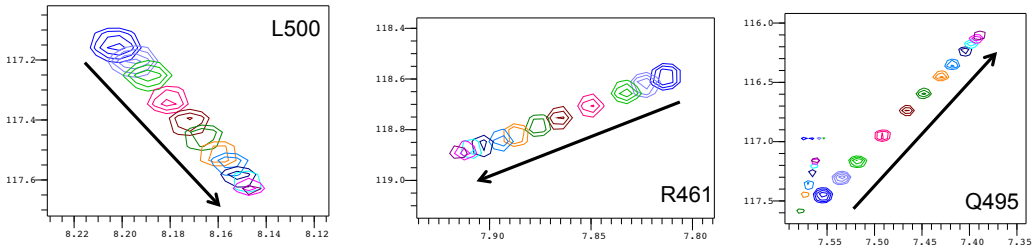


Figure 5.41: Blow-up of residues L500, R461 and Q495 from the RAP74NMR titration with the AR S430E/S431E/S432E peptide, showing that the peaks become broader at increasing concentrations of peptide. At the highest concentration of peptide added (30 molar equivalents) 90% of the bound state is populated. The color code is the same as the one used in figure 5.35.

Given that the trajectories along which the peaks move during the titrations are straight lines, the AR peptide is expected to bind to the RAP74 interaction groove with comparable affinity for the various bound conformations.

Only with the double phosphorylated peptide, which binds with the highest affinity, we observed that for some residues the peaks start to deviate from the straight trajectory at high peptide concentrations (see Fig. 5.42a). This may be related to a another binding event in which a second peptide molecule might bind to the saturated complex of RAP74-peptide. Such an event is likely unspecific and not biologically relevant as it is only observed for saturation levels higher than 90% and presumably due to the excess of peptide added to an already saturated complex. Since this secondary binding event is only observed at such high saturation levels, the binding affinity is considerably lower than that of the specific binding event under study. When we mapped the residues that show this behavior, we found they are located at one part of the RAP74NMR protein (see Fig. 5.42b), suggesting these residues form part of the binding site for this additional binding event.

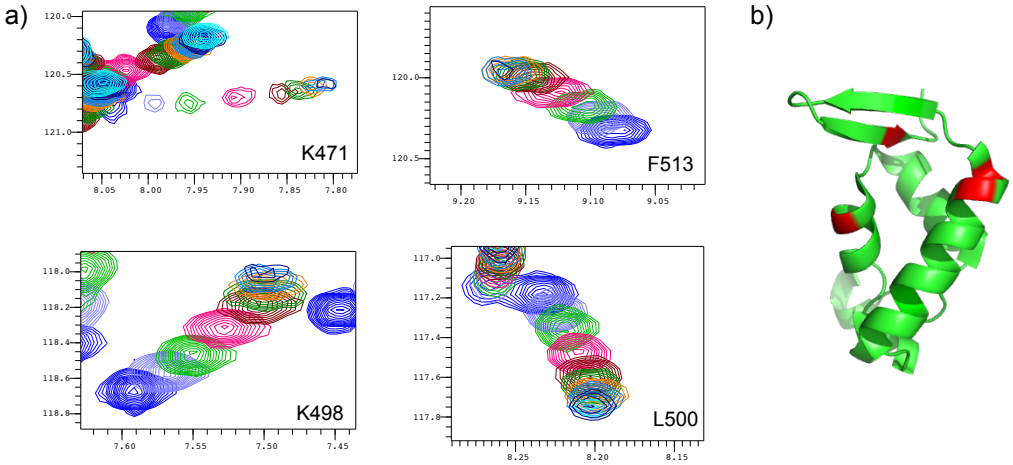


Figure 5.42: a) Blow-up of residues K471, K498, L500 and F513 from the RAP74NMR titration with pS430/pS432 peptide, showing a trajectory with two components. The peaks start to deviate from a straight line at $\geq$15 molar equivalents of peptide, *i.e.* at 90% saturation. b) Mapping of the residues for which the trajectory deviates from a straight line at high peptide concentrations on the structure of the RAP74 C-terminal domain. These residues (K471, K498, L500 and F513) are shown in red on the PDB structure 1I27, [350].

The phosphorylation pattern of the AR when it binds to RAP74 in the biological context is likely different from the one used in our *in vitro* experiments. The structurally heterogenous nature of the bound state could be in part related to this. A non-optimal distribution of the negative charges flanking the hydrophobic core of the binding motif could result in an ambiguous orientation of the peptide in the binding groove, or could fail to lock the peptide in one stable conformation in the complex. The possibility that one stable bound conformation exists with the naturally phosphorylated AR sequence cannot be excluded.

Our data support a more helical conformation of the AR binding site in the complex,

without folding into a stable helix upon binding. Notably, large-amplitude local dynamics for the FCP1 backbone have been described for the FCP1 binding motif when it is in complex with RAP74 [458], indicating that also this complex is not characterized by a fully formed stable helix of the FCP1 binding motif in the complex. This observation was consistent both with molecular dynamics simulations [459] and with the very high B-factors observed crystallographically [453] for an FCP1 peptide in complex with RAP74.

5.11 Multiple roles of the ^{433}WHTLF^{437} motif

As discussed in the introduction, other roles of the ^{433}WHTLF^{437} motif in transactivation have been described. The N/C interaction between the AR NTD and LBD, which is required for AR transactivation of some but not all androgen-regulated genes [191, 244], is believed to take place primarily by the interaction of the ^{23}FQNLF^{27} motif and the AF2 region of the LBD and to be further stabilized by a secondary interaction involving the ^{433}WHTLF^{437} motif. It is not clear whether the ^{433}WHTLF^{437} motif interacts with the AF2 region or instead in a region of the LBD located outside of AF2 [243–245]. In general, the contribution of the ^{433}WHTLF^{437} motif to the N/C interaction remains unclear [184, 207]. The ^{433}WHTLF^{437} motif has also been reported to interact with histone acetyltransferase p300 as part of an interaction network that leads to upregulated AR transcription [246], but no structural details are available for this interaction. We have shown, for the first time, that the ^{433}WHTLF^{437} motif can bind directly to the general transcription machinery and thereby activate transcription.

It is therefore possible that the regulation of transcriptional activity mediated by the ^{433}WHTLF^{437} motif involves the interaction of this region with multiple binding partners. Possibly, the contribution of each of these binding events depends on the conditions in the cell, favoring one interaction over the others. Or instead several interactions could occur in parallel resulting in both direct and indirect interactions with the transcription machinery leading to an overall enhanced transactivation.

In order to compare the interaction of the ^{433}WHTLF^{437} motif with the C-terminal domain of RAP74 to that with the AR LBD, we inspected this second interaction in more detail. Importantly, there is no structure available of the wild type AR sequence including the ^{433}WHTLF^{437} motif in complex with the LBD. However, the AR peptide ^{20}RGAFQNLFQSV^{30} containing the ^{23}FQNLF^{27} motif is crystallized in complex with the LBD and deposited in the PDB (1XOW, [191]). Assuming a similar binding mode for the ^{433}WHTLF^{437} motif at the same binding site, this structure can be used as a model. In addition, the X-ray structure obtained from a non-native WxxLF containing

peptide optimized for binding to the AR LBD by phage display (WQALF) is deposited under PDB code 1T74 [195].

Two clusters of charged residues flanking the AF2 play a dominant role in the recruitment and orientation of AR binding partners containing a FxxLF or LxxLL binding motif (see Fig. 5.43a) [238]. Furthermore, it was found that positively charged residues are predominantly present amino-terminal to the core sequence, whereas negatively charged residues predominate carboxy-terminal to the core sequence [238]. This charge distribution lies at the origin of the correct orientation of both binding partners. Examining the AR sequence around the ^{23}FQNLF^{27} motif, two residues with positively charged side chains are found amino-terminal to ^{23}FQNLF^{27} (K17 and R20) and both one positively and one negatively charged residue are found carboxy-terminal to it (R31 and E32) (see Fig. 5.44a). The ^{23}FQNLF^{27}, when bound to the AR LBD, is hydrogen bonded by conserved charge clamps E897 (helix 12) and K720 (helix 3) (see Fig. 5.43b), which are part of the charged patches flanking AF2 [191].

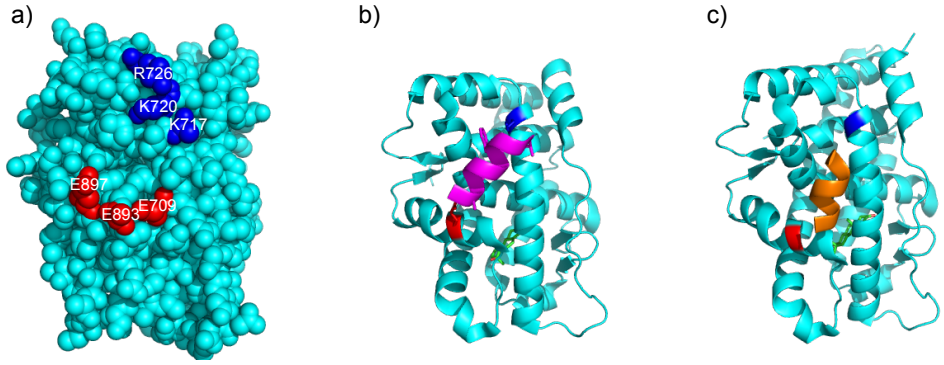


Figure 5.43: a) A space-filled model of the DHT-bound human AR LBD based on PDB structure 2AMA [192]. Positively charged residues K717, K720 and R726 flanking AF2 are shown in blue, negatively charged residues E709, E893 and E897 are shown in red. b) ^{23}FQNLF^{27} motif bound to the LBD, based on PDB structure 1XOW [191]. The ^{23}FQNLF^{27} containing peptide is shown in magenta and charge clamps E897 and K720 on the LBD are shown in red and blue, respectively. The hormone R1881 is shown in sticks. c) WxxLF motif bound to the LBD, based on PDB structure 1T74 [195]. The WxxLF containing peptide is shown in orange and charge clamps E897 and K720 on the LBD are shown in red and blue, respectively. The hormone DHT is shown in sticks.

a) ^{15}PSKTYRGA**FQNLF**QSVREVI^{34}
b) ^{425}PSAAASSS**WHTLF**TAEEGQL^{444}
c) SRWQA**LF**DDGTDTSR

Figure 5.44: AR sequence around the ^{23}FQNLF^{27} motif (a) and around the ^{433}WHTLF^{437} motif (b). The non-native sequence of the peptide containing a WxxLF motif crystallized in complex to the AR LBD (PDB 1T74, [195]) (c).

Around the ^{433}WHTLF^{437} motif two negatively charged residues are found carboxy-terminal to the motif (E440 and E441) and no charged residues are found amino-terminal to the motif (see Fig. 5.44b). However, the presence of these two negative charges does not seem to be sufficient to promote a stable interaction with the LBD given that no structure of the complex has been obtained. The sequence of a WxxLF containing peptide that was crystallized in complex with the LBD was optimized by phage display. This resulted in one positively charged residue amino-terminal to the core motif (position -1 to W433) and three negatively charged residues carboxy-terminal to it (at positions +5, +6 and +9 to W433) (see Fig. 5.44c). None of these charges is present in the wild type sequence. Even though these charges improve the interaction of this peptide with the AF2 binding pocket, the resulting complex (see Fig. 5.43c) shows mediocre binding affinity and an ill-defined helix in the interaction motif [195].

Taken together, the ^{433}WHTLF^{437} containing sequence of wild type AR is not optimal to interact with the LBD, in agreement with the fact that this interaction was proposed to be a secondary binding event and that it is not well understood where on the LBD binding takes place.

The triple hydrophobic mutation (W433A/L436A/F437A) and double charge reversal mutation (E440K/E441K) that destabilize the AR-RAP74 interaction, are also expected to have a destabilizing effect on the interaction of the ^{433}WHTLF^{437} motif with the LBD. The former mutant would weaken the interaction with the AF2 hydrophobic residues and the charge reversal mutant would cause repulsion with the positively charged cluster flanking AF2 (K717, K720 and R726). Consequently, the reduced transcriptional activity observed for both of these mutants in the reporter gene assays could be due to a synergistic effect if both the interaction of the ^{433}WHTLF^{437} motif with RAP74 and with the LBD are weakened.

As discussed above, the surface of the RAP74 C-terminal domain is also not optimal to accommodate the wild type AR sequence containing the ^{433}WHTLF^{437} motif. However, we have shown that the introduction of negative charges at residues N-terminal to the core motif strongly enhances the binding affinity. The same mutations would, by contrast, greatly weaken the interaction with the LBD as this would cause repulsion with the cluster of negatively charged residues (E709, E893 and E897). Therefore, we believe that the significantly increased transcriptional activity observed for the S430E/S431E/S432E mutant is directly related to the enhanced interaction between AR and RAP74.

Despite the potential of the ^{433}WHTLF^{437} motif to interact with both protein domains, these interactions seem to require a stimulus to take place with high enough

affinity to be of biological relevance. This is in agreement with the notion that the ^{433}WHTLF^{437} motif selectively enhances transcription under a specific set of conditions. We have shown that one possible mechanism to active transcription through the direct interaction of the ^{433}WHTLF^{437} motif with the C-terminal domain of RAP74 is the phosphorylation of serine residues N-terminal to the core motif. It is likely other mechanisms exist that promote this interaction, or instead the interaction of the ^{433}WHTLF^{437} motif with other partners, such as the LBD or p300.

5.12 Summary

We have shown that the interaction of the AR and RAP74 takes place at the ^{433}WHTLF^{437} motif in the Tau-5 region of AR, which is a key motif for AR transactivation in the absence of hormone. Furthermore, our results demonstrate that this interaction activates transcription in CRPC PC-3 cells at castrate levels of hormone. This is a novel transactivation mechanism of the AR NTD.

Importantly, our data indicate the AR-RAP74 interaction activates *aberrant* AR transactivation and possibly requires the phosphorylation of several residues N-terminal to the ^{433}WHTLF^{437} motif to occur with a biologically relevant affinity. This supports the notion that AR-RAP74 interaction is not a general mechanism for AR transactivation, but instead is a mechanism that can be activated under specific conditions to enhance AR transactivation. Possibly the interaction between AR and RAP74 only takes place in CRPC cells at low levels of hormone, when specific kinases are upregulated and phosphorylate residues N-terminal to the ^{433}WHTLF^{437} motif to promote the interaction.

Due to the low binding affinity of WT AR and RAP74, we could not determine an experimental structure of the complex. However, mutation of key residues for the interaction (W433, L436, F437, E440 and E441) suggests that the binding mode is similar to that of the well-characterized complexes of RAP74 and FCP1 motifs. Our results show that the AR binding motif becomes more helical upon interacting with RAP74, whereas the conformational properties of the AR residues outside of the binding site do not change significantly. Similarly, interaction induces no notable conformational changes in the C-terminal domain of RAP74. Our results suggest a "fuzzy" complex with structural heterogeneity in the bound state in which the AR binding motif adopts a more helical conformation without folding into a well-defined stable helix. Binding involves aromatic side chains (W433, L436 and F437) of AR interacting with the hydrophobic surface of RAP74 formed by helix 2 and 3, and a single acidic anchoring point of AR (E440, E441) making electrostatic contacts with positively charged

residues of RAP74. Phosphorylation of residues N-terminal to the ^{433}WHTLF^{437} motif further increases electrostatic attraction between both proteins. The phosphorylation pattern of the AR when it binds to RAP74 in the biological context remains elusive. It is, however, likely not to be identical to the one used in our *in vitro* experiments. Therefore, we cannot exclude the possibility that the structural heterogeneity of the complex and/or the not well-defined helical conformation of the AR binding motif in the complex is in part due to the lack of the phosphorylation pattern that is present in the biological context.

Surprisingly, our results are not in line with the previously reported data regarding this interaction. We found no significant interaction at AR residues previously suggested to form part of the binding surface (M245, L247 and V249) [353, 434], nor at residues suggested to have an indirect effect on the binding by altering the structural flexibility of the domain (S160, S163, S341 and S344) [320, 353, 434]. Furthermore, we showed by two orthogonal NMR titrations that the binding affinity of WT AR and RAP74 is in the order of magnitude millimolar. This is in stark contrast to the reported affinity of $0.63 \pm 0.08\,\mu$M [352]. Induced α-helix structure in AR AF1 upon binding of RAP74 has been reported [282]. Furthermore, it has been suggested that this increased structure facilitated the interaction with SRC-1 [352]. We have observed only a modest and local increase in helical propensity, at the binding motif of AR upon binding to RAP74.

6 Interaction of AR and EPI-001

The NTD is indispensable for the transactivation of the AR both in the presence and in the absence of hormone [147, 148]. The latter allows androgen-independent proliferation of CRPC cells. Consequently, the AR NTD is one of the main targets for CRPC.

Recently, several small molecules that bind to the AR NTD and selectively block its transactivation were identified (sintokamide A [139], EPI-001 [134] and niphatenone B [159], see also paragraph 1.1.4.1 in the introduction). They are lead compounds for the development of a new class of drugs for the treatment of CRPC. Interestingly, both EPI-001 and niphatenone B were found to bind irreversibly to the AF1 in the AR NTD [159, 160].

Presumably the interaction of such small compounds with the NTD alters the protein-protein interactions between AF1 and the transcription machinery, thereby interfering with the function of the NTD in AR transactivation. In order to understand how the NTD can be targeted by small molecules and how these influence its function in transcription, we studied the interaction between AF1* and EPI-001 in detail, using a combination of NMR spectroscopy and mass spectrometry.

6.1 EPI-001 as a drug candidate for CRPC

EPI-001 (see Fig. 6.1) is a recently discovered irreversible inhibitor of the AR that causes the regression of CRPC in cell lines and in an animal model of this disease [134]. It belongs to the novel class of AR inhibitors that interact with the NTD, unlike all other AR inhibitors that target the LBD. From all the members of this new class of inhibitors, EPI-001 is of particular interest as EPI-001 and its analogs are, at present,

the only known inhibitors of the constitutively active AR splice variants that have been shown to target the AR NTD [134, 160].

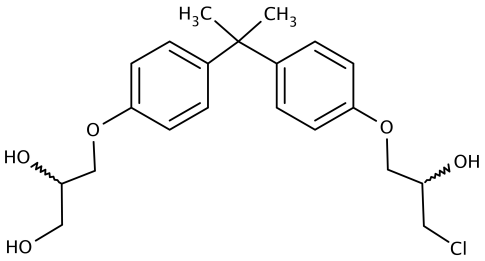


Figure 6.1: Chemical structure of EPI-001.

The interaction of EPI-001 and AF1 of AR was proposed to take place in a two-step mechanism, involving the formation of an encounter complex stabilized by non-covalent interactions, followed by a covalent reaction between the small molecule and the protein (see Fig. 6.2) [160]. First, the secondary alcohol of the chlorohydrin functionality is positioned next to a basic site in AF1 allowing the formation of an intermediate epoxide upon deprotonation of the secondary alcohol. The epoxide is then opened by a nucleophilic attack of a reactive side chain of AF1, resulting in a covalent bond between EPI-001 and AF1 [160].

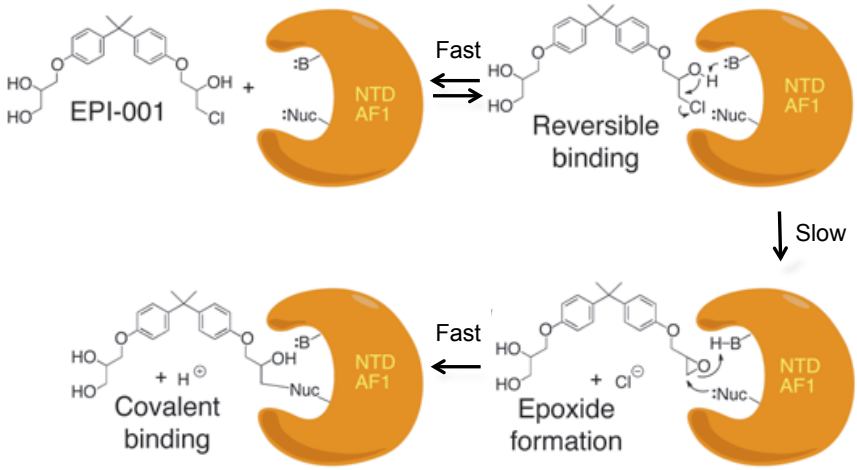


Figure 6.2: Proposed model by Myung *et al.* of the chemical mechanism for the selective covalent binding of EPI-001 analogs to the AR NTD. Adapted from [160].

EPI-001 was reported to specifically block the transactivation of the AR NTD without attenuating the transcriptional activity of related steroid hormone receptors [134]. This specificity is likely determined by the non-covalent step in the interaction mechanism, directing the EPI-001 specifically to the AR NTD and preventing irreversible binding to other nucleophiles present in the cell.

With the aim of elucidating the mechanism of action of EPI-001, that represents a novel class of AR inhibitors, and particularly to gain insight in the molecular basis for the specificity of EPI-001 for the AR NTD, we extensively studied the non-covalent interaction between AF1* and EPI-001 and further looked into the subsequent covalent binding.

6.2 Limited solubility of EPI-001 in aqueous buffer systems

The relatively hydrophobic character of EPI-001 due to its two aromatic rings limits its solubility in aqueous systems. This is commonly observed for small molecules that are lead compounds for drug discovery. To overcome the solubility limitations the interaction between such molecules and their targets is usually studied using co-solvent systems in which a small volume percentage of organic solvent is added to the aqueous buffer in which the protein is present. The organic solvent maintains the small compound in solution.

We have mainly used 1,4-dioxane as co-solvent (0.5% v/v) to study the interaction of AR with EPI-001. To assess the solubility of EPI-001 in this co-solvent system, the absorbance at 275 nm was monitored for different concentrations of the compound. As can be seen in figure 6.3, EPI-001 is soluble up to 250 μM in 95.5% aqueous buffer and 0.5% 1,4-dioxane. In the rest of the document 1,4-dioxane will be referred to as dioxane.

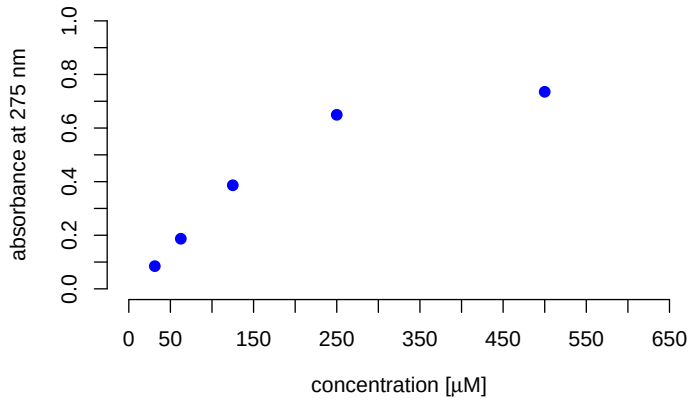


Figure 6.3: Solubility test for EPI-001 in aqueous buffer with 0.5% (v/v) dioxane as co-solvent. The absorbance at 275 nm was measured for different concentrations of EPI-001 in this co-solvent system.

Since several experiments were carried out at 278 K and lasted for 4 hours, the solubility of 250 μM EPI-001 in 0.5% dioxane after 4 hours at 278 K was also assessed. The sample was still optically clear after 4 hours at low temperature, suggesting the solubility of the compound is not affected by lowering the temperature for this amount of time.

6.3 Non-covalent interaction between AF1* and EPI-001

To study the reversible interaction between EPI-001 and AF1 of AR, NMR experiments were performed of the AF1* construct (AR 142–448) in the presence of the small compound. These experiments were conducted at 278 K, because at this temperature the irreversible reaction of EPI-001 with AR is sufficiently slow to detect the formation of the encounter complex, stabilized by non-covalent interactions between these two molecules, recently put forward by Myung *et al.* [160]. Indeed, MS analysis of a sample containing ^{15}N-labeled AF1* and 10 molar equivalents EPI-001 incubated at 278 K for 4 hours, the duration of the NMR experiment, confirmed that no covalent reaction had occurred under these conditions (see Fig. 6.4). The theoretical average mass of ^{15}N-AF1* is 31578 Da, but because the yield of isotopic labeling with ^{15}N is not 100% the mass detected for ^{15}N-AF1* in this experiment was 31575 Da. Covalent interaction of ^{15}N-AF1* with EPI-001 would result in the loss of one proton (M=1.01 Da) from ^{15}N-AF1* and the loss of a chlorine atom (M=35.45 Da) from the EPI-001 molecule (M=394.89 kDa). An adduct of ^{15}N-AF1* with one EPI-001 molecule would thus have a molecular weight of $31575 - 1.01 + (394.89 - 35.45) = 31933$ Da. This molecular weight is not observed, indicating that no covalent reaction has taken place. Furthermore, measuring the NMR experiments at 278 K offers the additional advantage that the rate of solvent exchange of the amide protons is minimized at low temperature [404], leading to favorable spectral properties.

The compound solubility limited us to work at relatively low concentrations of EPI-001. To be able to measure HSQC spectra with 10 molar equivalents of small compound relative to protein, we worked with samples of 25 μM AF1* (^{15}N-labeled) and 250 μM EPI-001 in 0.5% dioxane-d_8. Deuterated dioxane (dioxane-d_8) was used for the NMR experiments to minimize the dioxane signal in the spectra, which facilitated the data acquisition and processing.

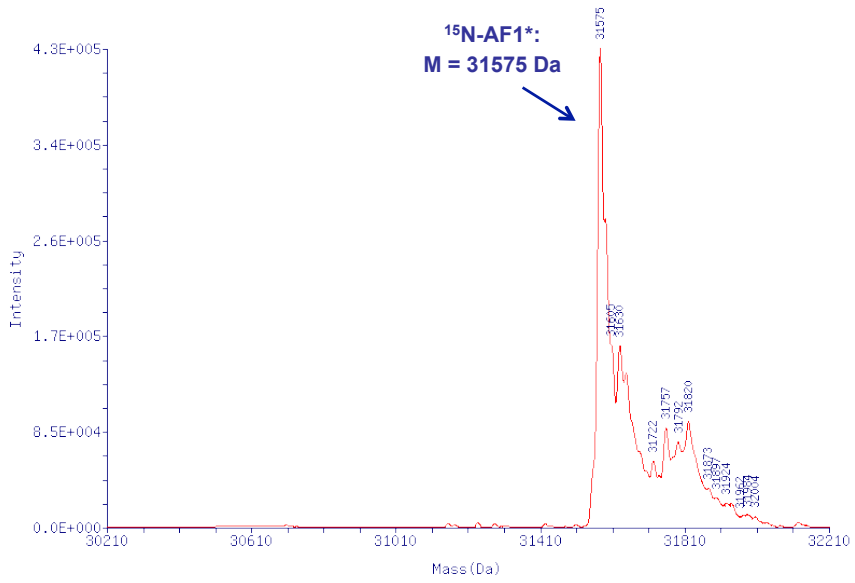


Figure 6.4: Mass spectrometry analysis of a mixture of ^{15}N-labeled AF1* and 10 molar equivalents EPI-001 (0.5% dioxane-d$_8$) after 4 hours incubation at 278 K. ^{15}N-AF1* is detected at molecular weight 31575 Da. The absence of a species with molecular weight corresponding to ^{15}N-AF1* covalently bound to one or several EPI-001 molecules shows that no covalent reaction has taken place under these incubation conditions.

6.3.1 EPI-001 specifically interacts with the Tau-5 region of AF1*

To map the binding epitope of EPI-001 on the AR NTD, we carried out [^{1}H,^{15}N]-HSQC experiments of 25 μM ^{15}N-AF1* + 250 μM EPI-001 in 0.5% dioxane-d$_8$. Because the addition of even a small volume percentage of organic solvent also affects the protein backbone chemical shifts, the spectrum of this sample was compared to that of a sample containing 25 μM ^{15}N-AF1* in the presence of 0.5% dioxane-d$_8$.

As can be seen in figure 6.5, the effect of the organic solvent on the resonances from the protein backbone is larger than the chemical shift changes originating from the interaction of AF1* with the small compound (compare Fig. 6.5a to Fig. 6.5b). Importantly, the residues to which dioxane causes perturbations are evenly distributed over the AF1* sequence, whereas EPI-001 causes chemical shift changes only to a specific set of residues. The chemical shift changes due to the addition of dioxane to AF1* increased approximately twofold when the volume percentage dioxane was changed from 0.5% to 1% (see Figs. 6.5b, c and d). Consequently, the volume percentage of organic solvent had to be kept as low as possible to minimize the effect of the organic solvent on the protein backbone resonances. We therefore prepared most of the samples using a co-solvent system of 99.5% aqueous buffer and 0.5% dioxane. As shown before the

solubility of EPI-001 in this co-solvent system is limited to 250 μM. As a consequence, it was not possible to use higher concentrations of EPI-001 when using this co-solvent system. An advantage of working at low volume percentages of organic solvent is a reduced probability of competition of the organic solvent with the compound to bind to the protein. In addition, organic solvents can reduce hydrophobic interactions between a protein and a small compound, so also in this regard it is preferable to work with a low solvent percentage.

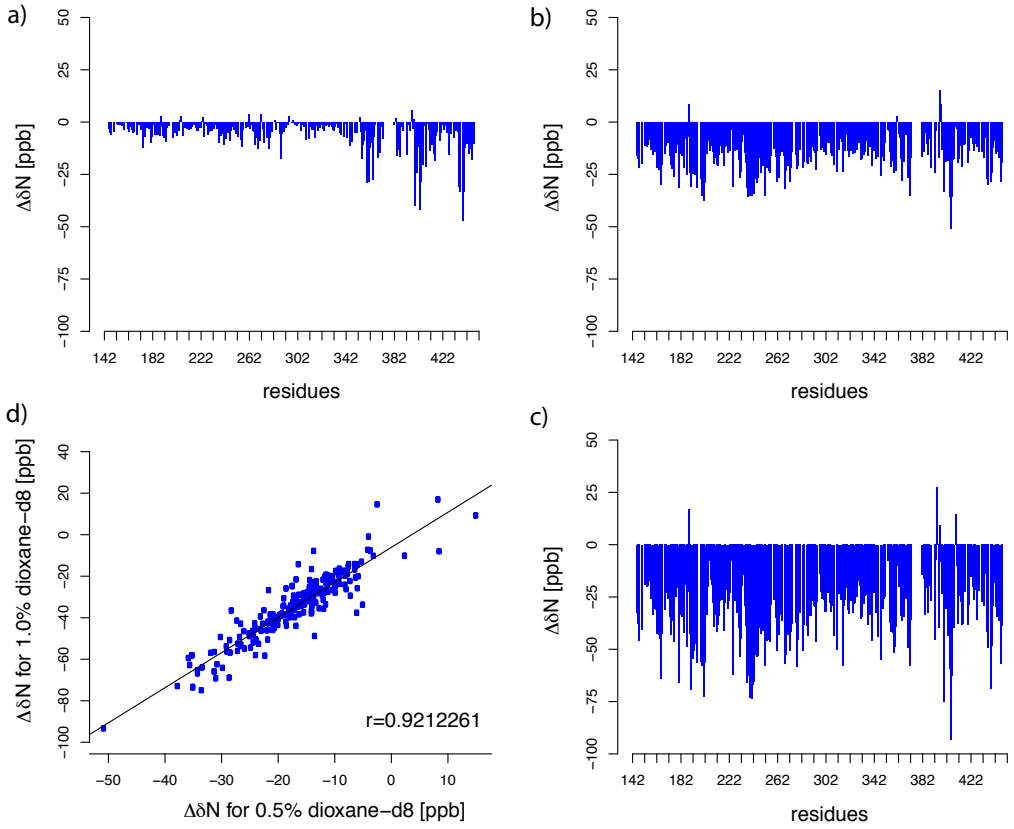


Figure 6.5: a) The ^{15}N chemical shift changes of AF1* caused by interaction with 10 molar equivalents of EPI-001 as a function of residue number, corrected for the effect of the organic solvent on the protein chemical shifts. Spectrum of 25 μM AF1* + 250 μM EPI-001 in 0.5% dioxane-d$_8$ was compared with the spectrum 25 μM AF1* + 0.5% dioxane-d$_8$. b) The ^{15}N chemical shift changes of AF1* caused by addition of 0.5% dioxane-d$_8$ as a function of residue number. Spectrum of 25 μM AF1* + 0.5% dioxane-d$_8$ was compared with the spectrum 25 μM AF1*. c) The ^{15}N chemical shift changes of AF1* caused by addition of 1% dioxane-d$_8$ as a function of residue number. Spectrum of 25 μM AF1* + 1% dioxane-d$_8$ was compared with the spectrum 25 μM AF1*. d) Correlation plot of the ^{15}N chemical shift changes of AF1* caused by 0.5% and 1% of dioxane-d$_8$. The correlation coefficient r is indicated.

An analysis of the ^{1}H and ^{15}N chemical shift changes of AF1* upon addition of EPI-001 indicates that the small molecule interacts specifically with residues 345 to 448 of the AF1* construct, which are located within transactivation unit 5 (Tau-5)

of AR (see Fig. 6.6, in which the lower panel of Fig. 6.6c is a repetition of Fig. 6.5a). As explained before, the Tau-5 region of AR plays a dominant role in the androgen-independent transactivation of AR which takes place in CRPC cells. Interestingly, the ^{433}WHTLF^{437} motif that is known to be important for this aberrant activity in the absence of hormone [232] and that we identified to be the binding site for interaction with RAP74 is located in one of the regions that experience the largest chemical shift changes.

Contrary to what we expected, the interaction with EPI-001 affected many residues of AF1* instead of taking place at a well-defined binding site consisting of a smaller number of amino acids. The amino acids interacting with EPI-001 are located in three contiguous regions of sequence (see Fig. 6.6) that correspond to regions with relatively high helical propensity and that are involved in dimerization (see Fig. 6.7).

That EPI-001 binding causes chemical shift perturbations in three regions of sequence can be due to the existence of three independent interaction sites, as has been reported for other systems [312], or to the existence of a single interaction site involving residues far in sequence that are structurally connected. Another possibility is a redistribution of the conformational ensemble of AF1* upon interaction.

The observation that the amino acids affected by interaction with EPI-001 are involved in dimerization and have some helical propensity suggests that interaction is related to the structural properties of AF1*, as previously suggested based on *in vitro* binding experiments monitored by fluorescence spectroscopy [160].

To investigate the nature of the non-covalent interactions that stabilize the complex between AF1* and EPI-001 we analyzed the identity of the residues experiencing the largest ^{15}N chemical shift perturbations. Ten residues, all of them in the three regions defined above, show perturbations larger that 25 ppb (1 ppb = 0.001 ppm) (see Fig. 6.6d). Those in the region centered around position 360 (Q358, R360, Y363) are charged or polar, those in the region centered around position 400 (A398, A402, Q403) are polar or of intermediate hydrophobicity, whereas those in the region centered around position 435 (H434, T435, F437, T438) are large, polar and hydrophobic. ^{15}N chemical shifts in the protein backbone are strongly affected by the properties of the side chain of the previous residue in the sequence [478]. Adding these residues to the list of perturbed positions –S359 and Y362 to the first region; A397 and A401 to the second one; W433, L436 to third one– reveals the motif centered around position 435 to be markedly more hydrophobic that the other two. These three regions of sequence have different residue compositions and physico-chemical properties, suggesting that the ^{15}N chemical shift changes observed upon EPI-001 binding are due to a redistribution of

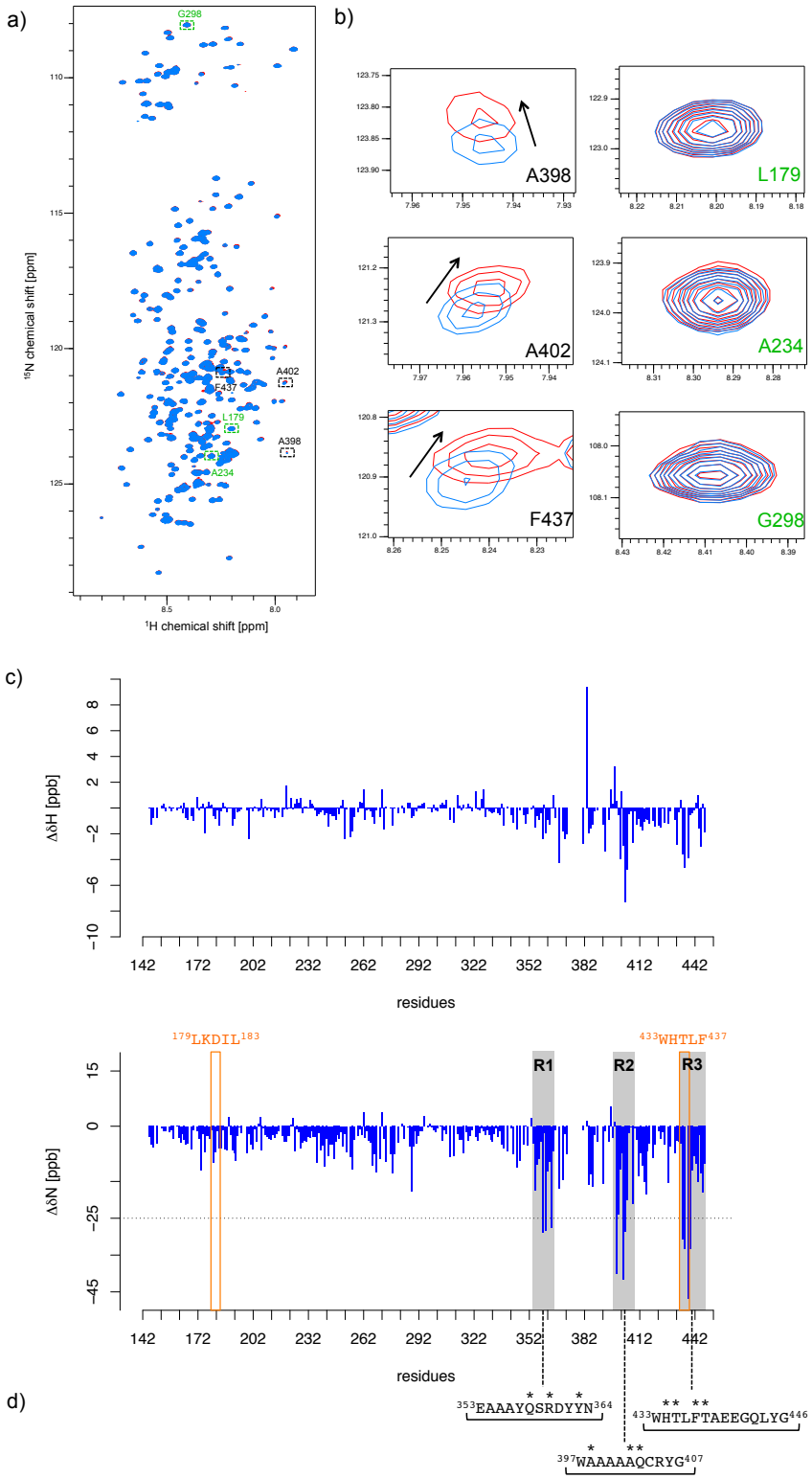


Figure 6.6: a) The [^{1}H,^{15}N]-HSQC spectrum of AF1* in the absence (blue) and presence (red) of 10 molar equivalents EPI-001. Both samples contained 0.5% (v/v) dioxane-d_8. b) Blow-ups of selected residues interacting with EPI-001 (A398, A402 and F437) and selected residues that do not interact with EPI-001 (L179, A234 and G298). c) The ^{1}H chemical shift changes (top) and the ^{15}N chemical shift changes (bottom) of AF1* caused by interaction with EPI-001 as a function of residue number. Relevant functional regions of AR are indicated. The dotted line at ±25 ppb indicates the cut-off value used for data analysis (see main text). d) Sequences of the three regions of sequence (R1, R2 and R3) that experience ^{15}N chemical shift perturbations with an indication with the symbol * of the residues experiencing perturbations larger than ±25 ppb (see main text).

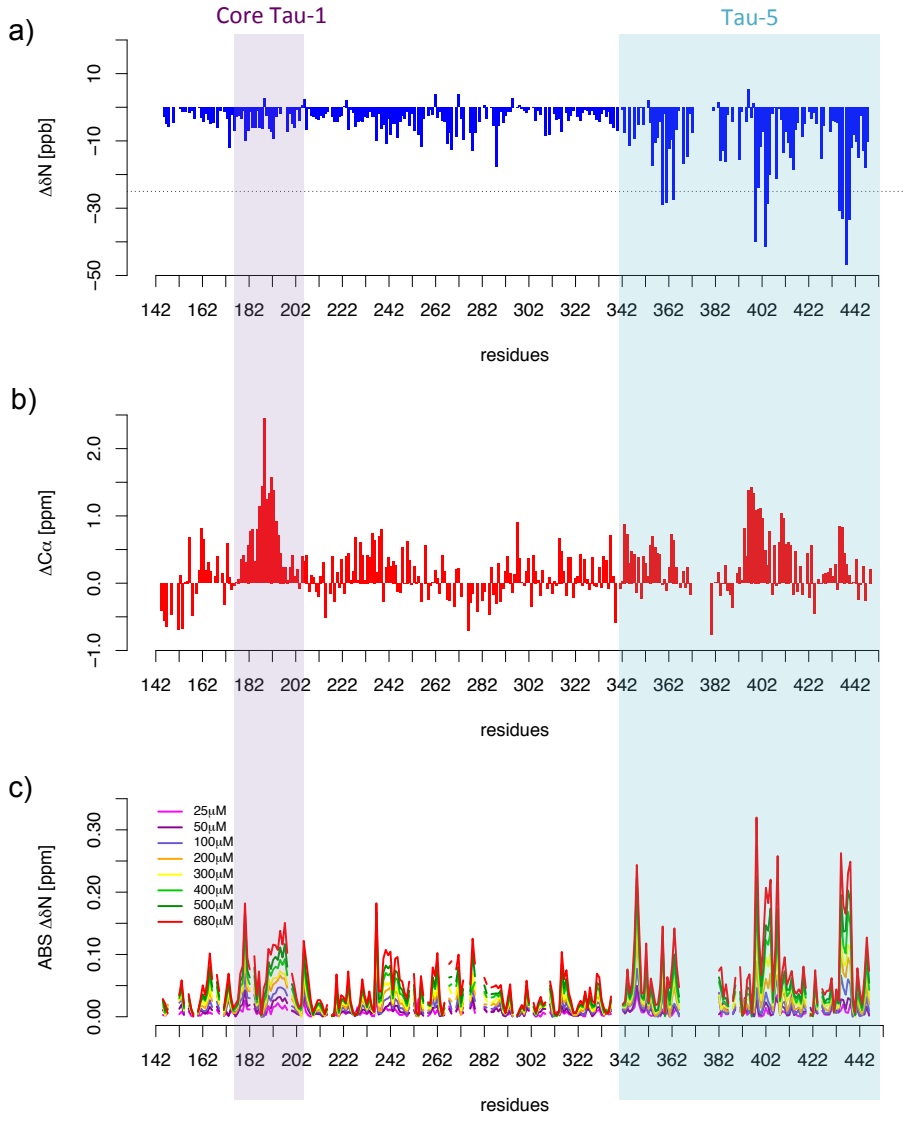


Figure 6.7: a) The ^{15}N chemical shift changes of AF1* caused by interaction with EPI-001 as a function of residue number. Relevant functional regions of AR are indicated. The dotted line at ±25 ppb indicates the cut-off value used for data analysis (see main text). b) Secondary C_α chemical shifts of AF1* calculated with the random coil values from Wang & Jardetzky [413], indicative of the helical propensity of AF1*. c) Absolute value of the difference in ^{15}N chemical shift for AF1* at different concentrations, relative to its chemical shifts at 10 μM, indicating the residues that are involved in dimerization.

the conformational ensemble of the AF1* construct.

Furthermore, the observed chemical shift changes are surprisingly small in size (order of magnitude ppb). This could be related to an unusually low binding affinity, but this would be in stark contrast to the clear biological effects observed in cellular and *in vivo* studies of the compound administered in low doses (10 μg/mL = 25.3 μM in cellular assays [134], and 50 mg/kg body weight = 0.13 mmol/kg body weight in mice studies [134]). However, taken together with the observation that many amino acids are affected by interaction with EPI-001, these small chemical shift changes seem to suggest, instead, that EPI-001 interacts with a low populated conformation of AF1*. Interaction could then slightly shift the equilibrium towards the conformational state with which EPI-001 interacts, causing small chemical shift changes in a large region of the protein.

Contrary to what was the case for the resonances of residues corresponding to Tau-5 we could not detect significant chemical shift changes in the resonances of residues corresponding to Tau-1, the most important transactivation unit in AR [157], even though it contains a region with particularly high helical propensity corresponding to residues 176–202 (see Fig. 6.7). That EPI-001 interacts with Tau-5 but not with Tau-1 suggests that its interaction with the NTD is specific and with a relatively well-defined conformational state of this domain.

In spite of being small in size, the chemical shift changes observed upon interaction with EPI-001 are confidently distinguishable from the noise (see Fig. 6.8) and highly reproducible (see Fig. 6.9). We determined the experimental uncertainty of the chemical shift measurements by recording [^{1}H,^{15}N]-HSQC spectra of three individually prepared but otherwise identical NMR samples and comparing the chemical shifts (see Fig. 6.8), similar to Bruun *et al.* [433]. The average standard deviation of the chemical shift measurements was $\pm$0.5 ppb for ^{1}H and $\pm$2.4 ppb for ^{15}N obtained as an average over all peaks in the spectra.

Both the sample preparation and the NMR experiments (data acquisition, data processing, data analysis and chemical shift referencing) were carried out extremely carefully to minimize the experimental error originating from different sources. This is detailed under paragraph 3.6.4 in the Methodology chapter and summarized in table 6.1.

Only cross peaks for which the EPI-001-induced ^{15}N chemical shift change was larger than $\pm$25 ppb were considered. These chemical shifts are well above the confidence levels of the experimental average standard deviation ($\pm$2.4 ppb) and the peak picking (digital resolution is 2.4 ppb in the ^{1}H dimension and 10.7 ppb in the ^{15}N dimension).

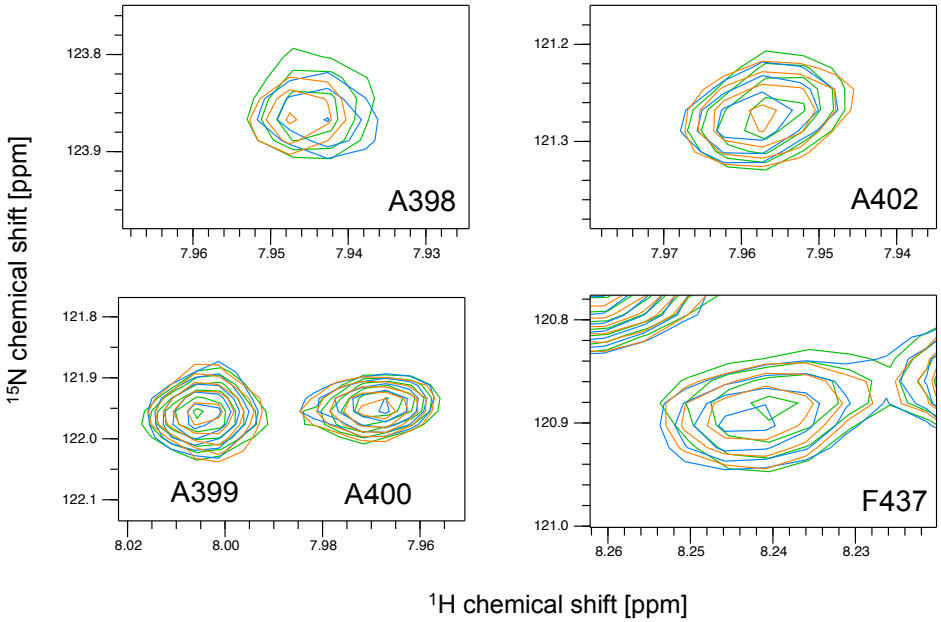


Figure 6.8: Experimental uncertainty of chemical shift measurements. Blow-ups of selected residues A398, A399, A400, A402 and F437 from the HSQC spectra of three individually prepared but otherwise identical samples (25 μM ^{15}N-AF1* + 0.5% dioxane-d_8). The resonances of these amino acids are affected by interaction with EPI-001 (see Fig. 6.6b) and by the presence of 0.5% dioxane, but as can be seen here their peak position in the presence of 0.5% dioxane is highly reproducible, indicating that the additional perturbation observed upon adding EPI-001 is reliable.

Table 6.1: Experimentally determined uncertainty of chemical shift measurements and the digital resolution of the NMR experiments after processing of the data with relevant acquisition and processing parameters.

	^{1}H	^{15}N
Experimental standard deviation in ppb	±0.5	±2.4
Number of points/increments	2048	512
Sweep width in Hz	8013	1784
Linear prediction (LP)	none	1024
Zero filling (ZF)	4096	2048
Digital resolution after LP and ZF in ppb	2.4	10.7

Since chemical shifts are highly sensitive to changes in the chemical environment of the nuclei observed [479] small chemical shift perturbations of this size can contain valuable information. Comparably small chemical shift changes, in combination with site-specific mutations, have been used in a study by Bruun *et al.* to detect long-range interactions between the site of mutation and residues far away in sequence of which the chemical shifts were slightly perturbed by introduction of the mutation [433]. In this particular case, chemical shift changes for ^{15}N, $^{13}C_{\alpha}$ and ^{13}CO larger than ± 15 ppb were included in the analysis, for experimental uncertainties similar to ours (± 1.5 ppb for ^{15}N, ± 1.4 ppb for $^{13}C_{\alpha}$ and ± 0.7 ppb for ^{13}CO) [433]. Interestingly, these long-range effects were typically observed in a contiguous set of residues, similar to the chemical shift perturbations observed for the interaction between AF1* and EPI-001. In this work, the authors showed that chemical shift changes of protein backbone residues of order of magnitude ppb can report on a minor redistribution of the conformational ensemble [433]. This gives further support to the notion that the small chemical shift changes observed when EPI-001 binds to AF1* reflect a slight shift in the conformational ensemble towards the state EPI-001 binds.

To evaluate the reproducibility of the chemical shift changes observed, we monitored the chemical shift perturbations in the HSQC spectra of two sets of samples prepared from a different protein batch and a different EPI-001 stock solution of 50 mM in 100% dioxane-d_8. As shown in figure 6.9, the chemical shift changes obtained from the two sets of samples were highly comparable.

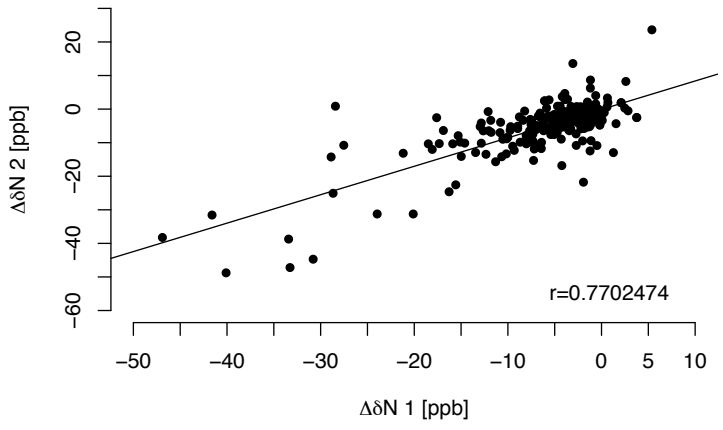


Figure 6.9: Reproducibility of the ^{15}N chemical shift changes. Correlation plot of the ^{15}N chemical shift changes of AF1* caused by interaction with 10 molar equivalents of EPI-001 as a function of residue number, for two different protein batches (expression and purification) and two separately prepared stock solutions of 50 mM EPI-001 in 100% dioxane-d_8. The correlation coefficient r is indicated.

Furthermore, the chemical shift perturbations increase when increasing concentrations of EPI-001 are added to AF1* (see Fig. 6.10). This further indicates that the interaction between AF1* and EPI-001 is specific and involves the residues located in the C-terminal region of AF1*.

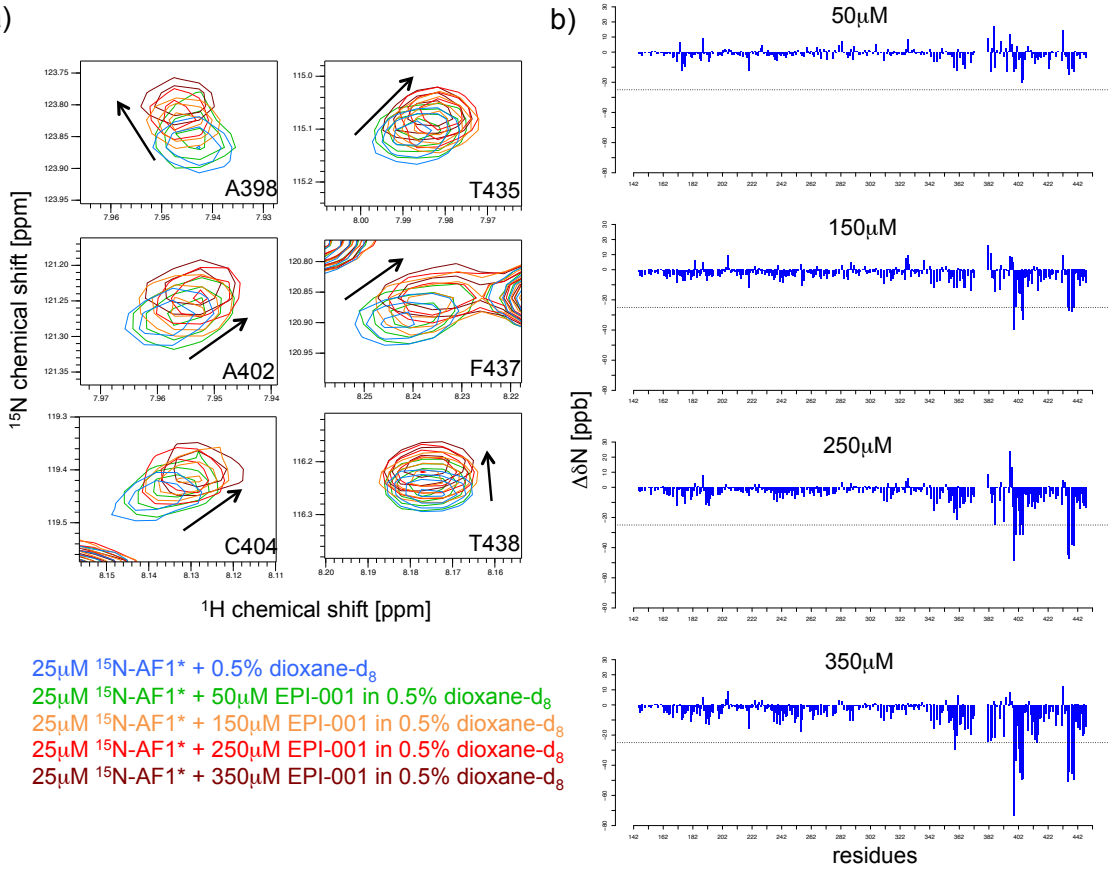


Figure 6.10: a) Blow-ups of residues A398, A402, C404, T435, F437 and T438 from HSQC spectra of ^{15}N-AF1* with increasing concentrations of EPI-001. All samples contained 0.5% (v/v) dioxane-d$_8$. b) The ^{15}N chemical shift changes of AF1* caused by interaction with increasing concentrations of EPI-001 as a function of residue number. From top to bottom the concentrations of EPI-001 are 50 μM, 150 μM, 250 μM and 350 μM.

To assess possible effects of the organic solvent on the interaction between AF1* and EPI-001, the HSQC experiment was also performed using a different co-solvent system. When a mixture of 99.5% aqueous buffer and 0.5% DMSO-d$_6$ was used, chemical shift changes were observed in the same residues of AF1* as when 0.5% dioxane-d$_8$ was used as co-solvent (see Fig. 6.11). This shows that the interaction of AF1* and EPI-001 is independent of the organic solvent used as a co-solvent.

An EPI-001 analog in which the chlorine group is replaced by a hydroxyl group, rendering the molecule symmetric, was found to have no effect on the transcriptional

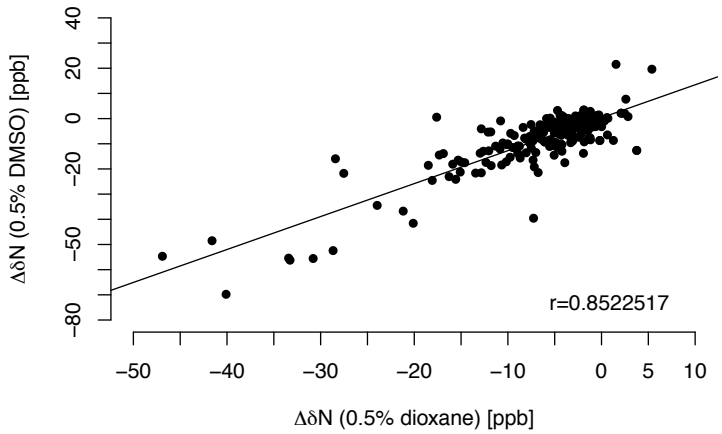


Figure 6.11: Correlation plot of the ^{15}N chemical shift changes of AF1* caused by interaction with 10 molar equivalents of EPI-001 as a function of residue number, using a co-solvent system with 0.5% dioxane-d_8 or using a co-solvent system with 0.5% DMSO-d_6. The correlation coefficient r is indicated.

activity of AR in cellular assays and was used as a control compound [134]. The molecular structure from this diol analog of EPI-001 is compared to that of EPI-001 in figure 6.12.

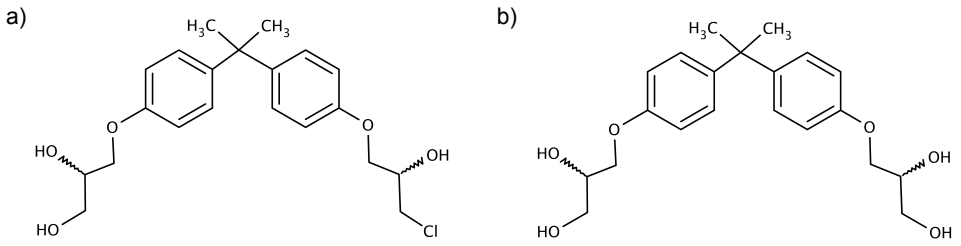


Figure 6.12: a) Molecular structure of EPI-001. b) Molecular structure of an inactive EPI-001 analog, referred to as the diol analog of EPI-001. The chlorine functionality of EPI-001 is replaced by a hydroxyl group, which leads to a symmetric molecule.

In order to confirm that the inactive diol analog does not interact reversibly with AF1*, as suggested by the fluorescence experiments by Andersen *et al.* [134], we compared the ^{15}N chemical shift changes caused on AF1* by 10 molar equivalents of EPI-001 and of the diol analog (co-solvent 0.5% dioxane-d_8). To ensure that the concentration of both small molecules added to the protein samples was the same, the small compound signals in the 1D proton spectra of both samples, which were recorded immediately before and after the HSQC, were integrated. When the integration of the multiplet signal centered around 7.28 ppm in the 1D spectrum of EPI-001 (for the assignment of the proton spectrum of EPI-001, see Fig. 6.16a) was set to 100% (250 μM), the corre-

sponding signal of the inactive diol was found to be 103% (258 μM).

We found that the addition of 10 molar equivalents of the inactive compound caused chemical shift perturbations in AF1* resonances below the established threshold of 25 ppb which can therefore not be confidently distinguished from the noise (see Fig. 6.13), supporting the findings from Andersen *et al.* Nevertheless, the residues that are affected most seem to be in the same regions of sequence identified to interact with EPI-001. Closer inspection of these data therefore seems to suggest that the inactive diol analog also interacts reversibly with AF1*, although with a lower affinity than EPI-001.

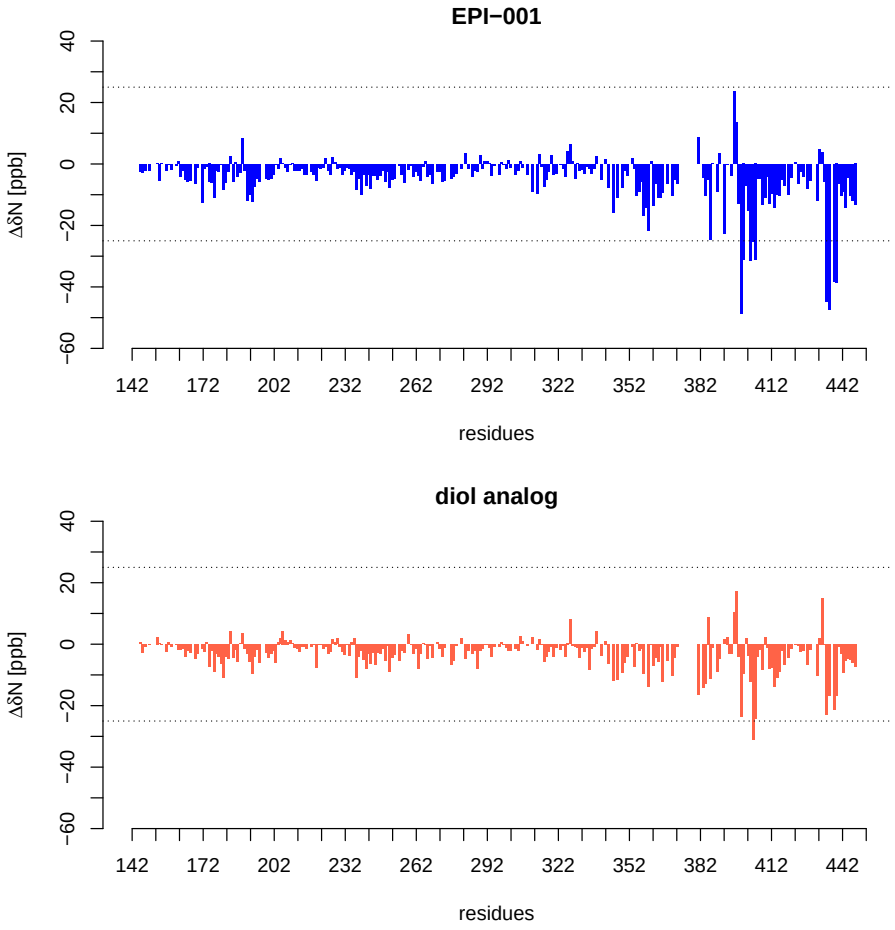


Figure 6.13: The ^{15}N chemical shift changes of AF1* caused by interaction with 10 molar equivalents of EPI-001 as a function of residue number (top). The ^{15}N chemical shift changes of AF1* caused by interaction with 10 molar equivalents of the inactive diol analog of EPI-001 as a function of residue number (bottom).

Therefore the origin of the inactivity of this compound could be twofold. On the one hand the absence of the chlorine functional group impedes the formation of a covalent

bond (-OH is a considerably worse leaving group than -Cl) and on the other hand the reversible interaction to position the small molecule in the proximity of a specific region of AF1* prior to covalent binding is of lower affinity for the diol analog compared to the active EPI-001.

Since the EPI-001 compound contains two chiral centers four stereoisomers exist: EPI-002 (2R,20S), EPI-003 (2S,20R), EPI-004 (2R,20R) and EPI-005 (2S,20S) (see Fig. 6.14). The same nomenclature for the stereoisomers is used as in Myung *et al.* [160]. To investigate whether the interaction with AF1* is stereoselective, the four stereoisomers were synthesized individually by Àlex de la Fuente in collaboration with Prof. Antoni Riera (IRB Barcelona and University of Barcelona) and the chemical shift perturbations caused on the resonances of ^{15}N-AF1* by addition of 10 molar equivalents of each of them were analyzed (see Fig. 6.15). Also for these experiments 0.5% dioxane-d_8 was used as co-solvent.

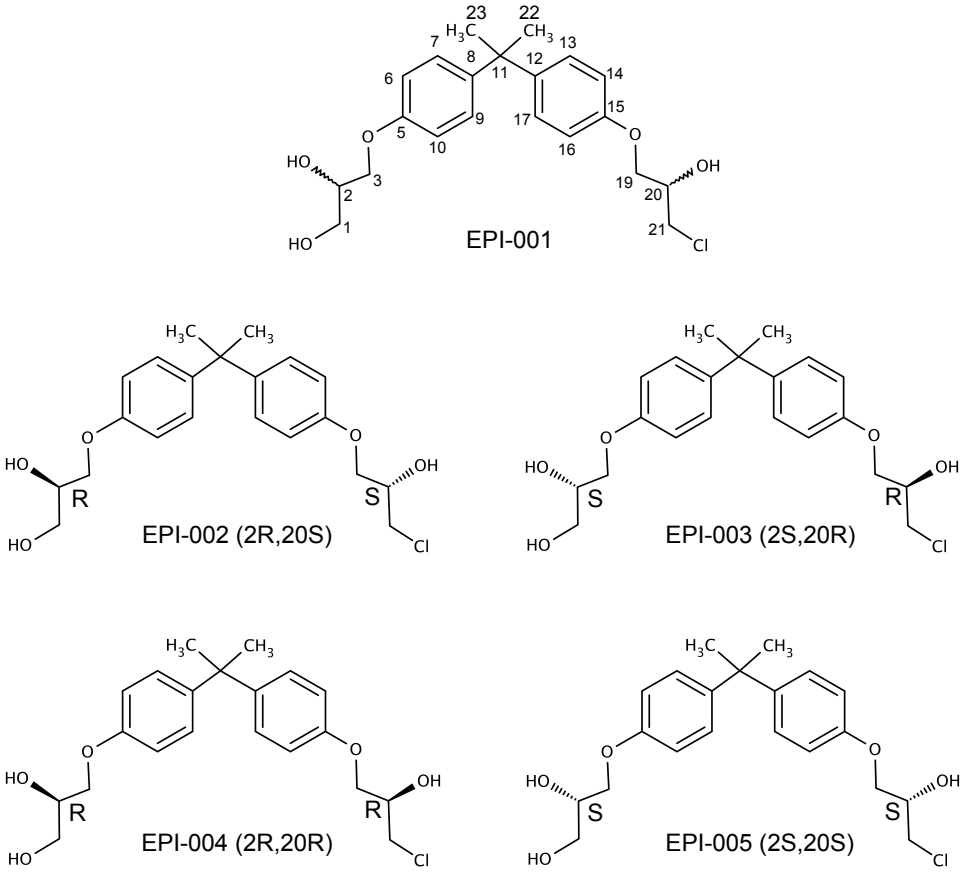


Figure 6.14: Chemical structure of EPI-001 and its four stereoisomers. The atom numbering used for the nomenclature is indicated. The chiral centers are located at positions 2 and 20.

In agreement with the *in vivo* results reported by Myung *et al.* [160] we found that

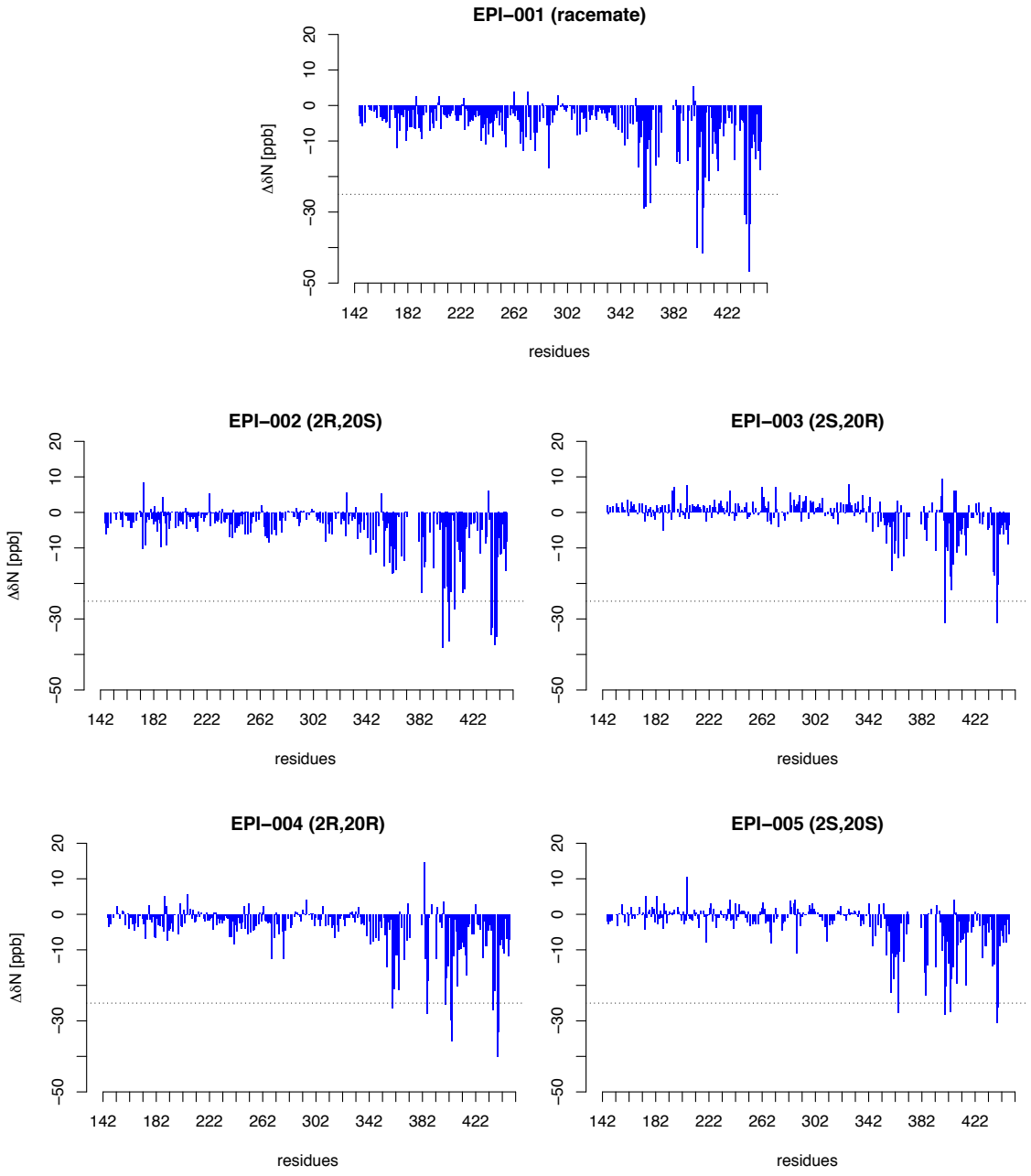


Figure 6.15: Influence of the chirality of EPI-001 on the non-covalent interaction with AF1*. The ^{15}N chemical shift changes of AF1* caused by interaction with 10 molar equivalents racemic EPI-001, EPI-002 (2R,20S), EPI-003 (2S,20R), EPI-004 (2R,20R) or EPI-005 (2S,20S) as a function of residue number. The dotted line at ±25 ppb indicates the cut-off value used for data analysis (see main text). In all samples 0.5% dioxane-d_8 was used as a co-solvent.

each of the four stereoisomers reversibly interacts with AF1* and causes chemical shift changes of similar size in the same resonances of the AF1* construct that are affected by interaction with the racemate. That the chemical shift changes are comparable in size for all stereoisomers and for the racemate indicates that no specific stereochemistry is required for the interaction to take place.

6.3.2 Confirmation of non-covalent interaction from the ligand perspective

To further confirm that a reversible interaction between AF1* and EPI-001 takes place and to investigate the nature of the complex we monitored the effect of AF1* on the resonances of EPI-001 in a 1D proton NMR spectrum. Addition of increasing substoichiometric amounts (5, 10 and 25 μM) of AF1* to 100 μM EPI-001 leads to small but significant and increasing chemical shift changes as well as to increases in line width, as expected for a low molecular weight compound transiently associating with a macromolecule [378] (see Fig. 6.16 and table 6.2).

Table 6.2: The chemical shift changes observed, in Hz, relative to the resonances of 100 μM EPI-001 in the absence of AF1* for each of the EPI-001 resonances corresponding to the aromatics (the multiplet signal centered around 7.28 ppm ($m_{7.28}$, corresponding to **8** in the assignment) and the multiplet signal centered around 6.95 ppm ($m_{6.95}$, corresponding to **7** in the assignment)) and the methyl group (s, corresponding to **9** in the assignment) of EPI-001 (the resonances corresponding to the aliphatics of EPI-001 are not shown but they follow the same trend). The resolution for the ^{1}H chemical shift was 0.20 Hz.

	$m_{7.28}$	$m_{6.95}$	s	molar equivalents
100 μM EPI-001	0	0	0	0
100 μM EPI-001 + 5 μM AF1*	0.32	0.13	0.16	0.05
100 μM EPI-001 + 10 μM AF1*	0.64	0.42	0.40	0.10
100 μM EPI-001 + 25 μM AF1*	1.59	1.17	1.02	0.25

We also carried out saturation transfer difference (STD) experiments [376] to further confirm the binding between these two molecules from the ligand perspective. STD is an NMR technique used to detect ligand binding to a protein. Binding results in a difference spectrum with ligand peaks, whereas if no binding occurs, the difference spectrum does not show any peaks. The technique is explained in more detail under paragraph 3.6.3.6 in the Methodology chapter.

The clear signals of EPI-001 observed in the STD difference spectrum of AF1* in the presence of 10 molar equivalents EPI-001 confirmed that reversible interaction between both molecules is taking place (see Fig. 6.17). The STD difference spectrum of a sample containing only the small compound was recorded to ensure that irradiation at -0.4 ppm

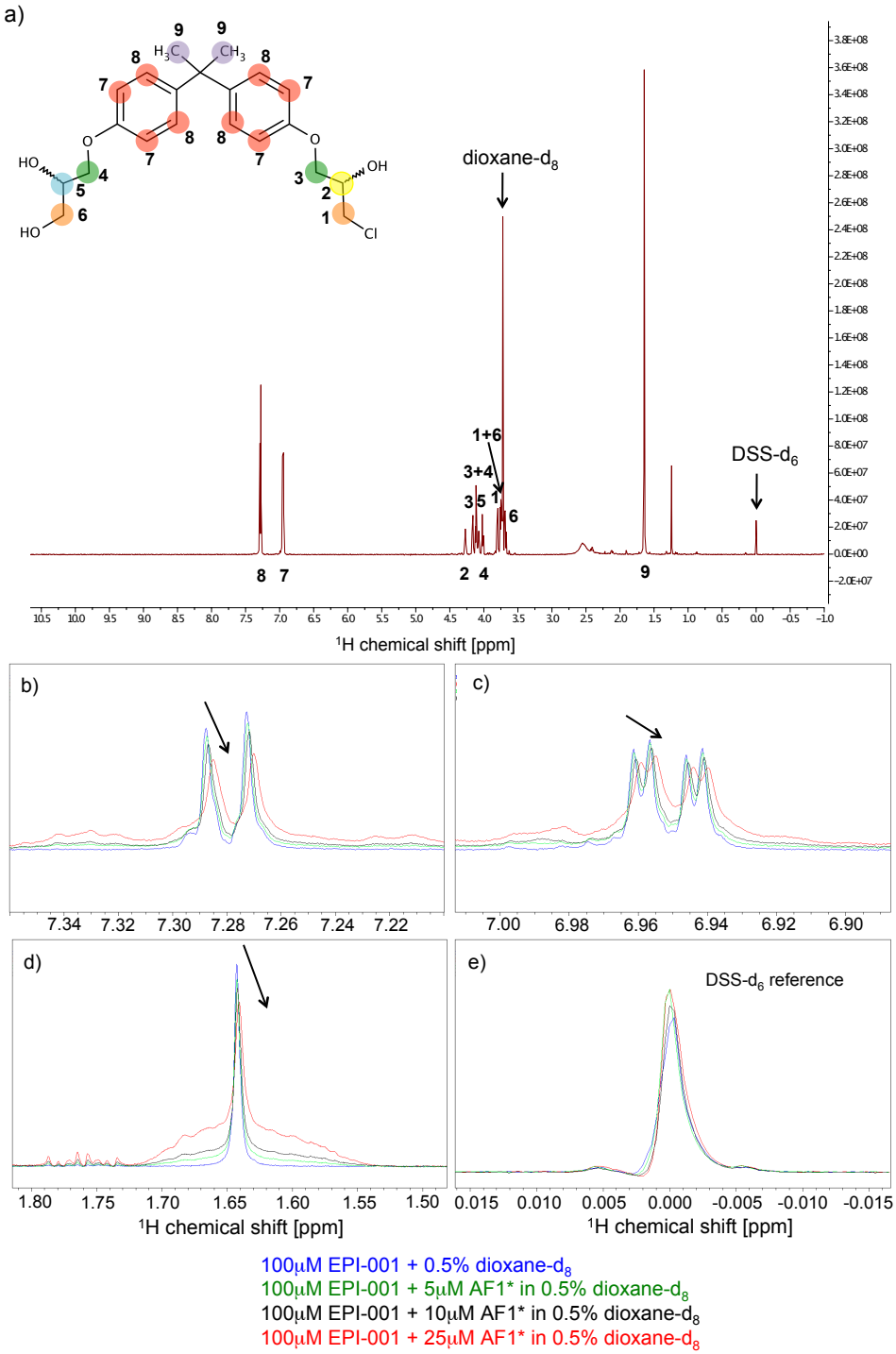


Figure 6.16: Effect of varying concentrations of AF1* on the chemical shifts of EPI-001 in a 1D ^{1}H spectrum. a) Assigned ^{1}H NMR spectrum of EPI-001. b) Blow-up of the multiplet signal centered around 7.28 ppm ($m_{7.28}$, corresponding to **8** in the assignment) for 100 μM EPI-001 in the presence of increasing concentrations of AF1* (see figure legend). c) Idem to b, but for the multiplet signal centered around 6.95 ppm ($m_{6.95}$, corresponding to **7** in the assignment). d) Idem to b, but for the methyl singlet (s, corresponding to **9** in the assignment). e) Blow-up of the DSS-d_6 signal from the different samples, which was used for reference purposes. All samples contained 0.5% (v/v) dioxane-d_8. The resolution was 0.20 Hz, i.e. 0.00034 ppm (600 MHz spectrometer).

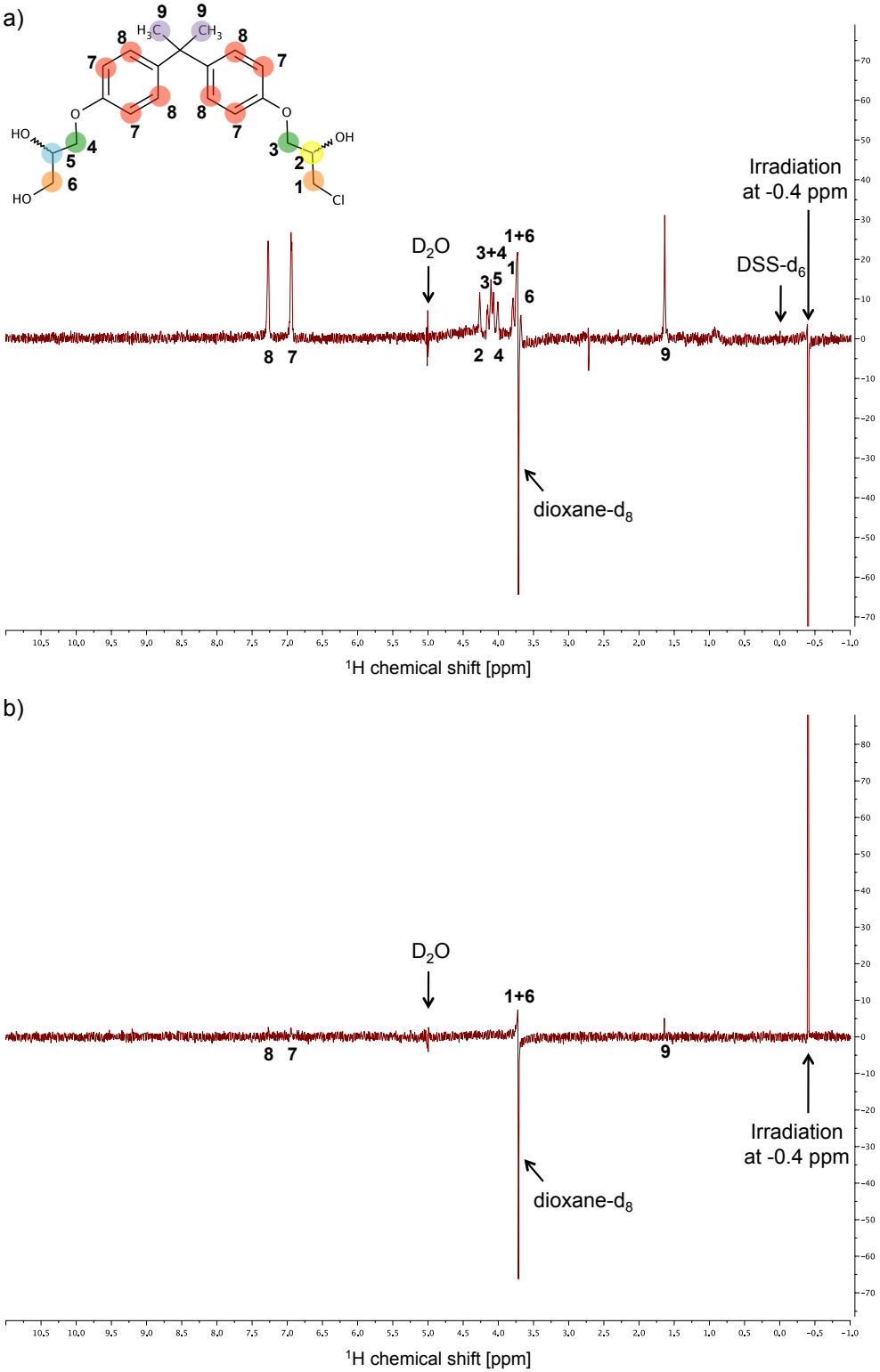


Figure 6.17: STD experiments confirm the non-covalent interaction between AF1* and EPI-001. a) The STD difference spectrum of 25 μM AF1* + 250 μM EPI-001 in 0.5% dioxane-d_8. b) The STD difference spectrum of 250 μM EPI-001 in 0.5% dioxane-d_8 was recorded as a control. The on-resonance was -0.4 ppm (see main text) and the off-resonance was +60 ppm. All spectra were measured at 278 K.

did not saturate EPI-001 directly. As can be seen in figure 6.17b, this is not the case, as the residual signals observed in this spectrum are negligible compared to the signals observed in the STD spectrum of the sample containing both protein and ligand (see table 6.3). This demonstrated that not EPI-001 but instead AF1* was saturated by irradiating at -0.4 ppm and that saturation could be transferred from the protein to the small compound, indicating they interact non-covalently. It is not uncommon to observe negligible residual signals from the ligand in the control experiment, and this is most commonly observed for hydrophobic compounds (personal communication with Dr. Jesús Angulo (University of East Anglia, UK)).

Table 6.3: Analysis of the STD experiments. The STD value (%) for each of the protons is calculated as the ratio between the integrated area under the corresponding signal in the STD difference spectrum and in the STD off-resonance spectrum. The first column indicates the range of chemical shifts over which the integral was taken for each signal in the ^{1}H spectrum of EPI-001. In the second and third column, the STD values (%) are shown for the STD experiment containing 25 μM AF1* and 250 μM EPI-001, and for the control sample containing only 250 μM EPI-001, respectively. The forth column indicates the number used in the assignment for the corresponding protons.

range (ppm)	AF1*+EPI-001 STD (%)	only EPI-001 STD (%)	protons
7.33 .. 7.22	8.95	0.29	**8**
7.00 .. 6.89	10.06	0.55	**7**
4.29 .. 4.24	9.46	0.08	**2**
4.18 .. 4.14	5.51	0.03	**3**
4.13 .. 4.09	5.76	0.04	**3+4**
4.09 .. 4.05	9.61	0.23	**5**
4.04 .. 3.99	6.69	-0.01	**4**
3.82 .. 3.78	6.97	0.14	**1**
3.76 .. 3.73	7.35	1.46	**1+6**
3.70 .. 3.66	3.93	-1.48	**6**
1.69 .. 1.59	6.13	0.29	**9**

These experiments indicate that EPI-001 interacts with a conformational state of AF1* that has relaxation properties similar to those of a globular protein. Together with the observation that incubation with EPI-001 causes chemical shift perturbations in three distinct and well-defined regions of sequence of AF1* (see Fig. 6.6) they give support to the hypothesis that EPI-001 binds to a conformational state of AF1* where these three regions are structurally and/or dynamically related.

6.3.3 Molecular basis for the specificity of EPI-001 for Tau-5

Having established that EPI-001 interacts specifically with the Tau-5 region of AF1*, our next objective was to understand the origin of the specificity for this particular

part of the protein. In figure 6.18 a schematic representation of the various possibilities is given. In this paragraph the experiments performed to answer this question will be explained step by step and the discussion will be guided by figure 6.18.

Taken together, our data seem to indicate that interaction of EPI-001 with AF1* is a cooperative process involving the three parts of sequence with helical propensity in the Tau-5 region, implying that the specificity to interact with this region is related to its structural properties (option B in Fig. 6.18). To exclude the alternative possibility, that is that EPI-001 interacts independently with the three regions of sequence experiencing chemical shift perturbations in AF1* (option A in Fig. 6.18) (see Fig. 6.6), we monitored the effect on the ^{1}H NMR spectrum of EPI-001 of three peptides with sequences corresponding to these regions (see table 6.4 and Fig. 6.19). We found that co-incubation of these under the same conditions of buffer, temperature and concentration that lead to the spectral changes reported in figure 6.16 caused no significant change in the spectrum of EPI-001 (compare the data from Fig. 6.19 to that of 100 μM EPI-001 + 25 μM AF1* in Fig. 6.16). That none of the peptides binds confirms that cooperativity between the residues in the Tau-5 region is required for interaction (i.e. option A in Fig. 6.18 is excluded). Put in a different way, it shows that the specificity of EPI-001 to interact with the Tau-5 region of AR is related to the structural properties of this part of the sequence in the context of the protein. It further shows that EPI-001 specifically recognizes a particular conformation that this region adopts.

Table 6.4: The three peptides used for this study, spanning each one of the regions of sequence of AF1* that experiences chemical shift perturbations upon interaction of AF1* with EPI-001. The cysteine residue in R2' peptide (C404) is underlined (see main text). Note that the R3' peptide is the same peptide as the one used to study the interaction between AR and RAP74, which was defined as WT peptide in table 5.1 on p. 184.

peptide name	sequence
R1' = AR 341–371	Ac-^{341}STLSLYKSGALDEAAAYQSRDYYNFPLALAG^{371}-NH_2
R2' = AR 391–414	Ac-^{391}LDYGSAWAAAAAQ<u>C</u>RYGDLASLHG^{414}-NH_2
R3' = AR 426–446	Ac-^{426}SAAASSSWHTLFTAEEGQLYG^{446}-NH_2

A cooperative model is compatible with interaction of EPI-001 with a conformation of AF1* where the three regions of sequence form a tripartite binding site (option C in Fig. 6.18) or where they undergo a conformational change upon binding (option D in Fig. 6.18). Given that the resonances of about 100 amino acids undergo small chemical shift perturbations upon interaction with EPI-001, it is unlikely that a single EPI-001 molecule interacts with each of these residues simultaneously. Furthermore, the polyproline stretch (AR 372–379) located between the R1 and R2 regions of sequence sterically hinders the interaction of R1 with the other two regions that are affected by EPI-001 interaction, and thus impedes the formation of a collapsed three-helix bundle.

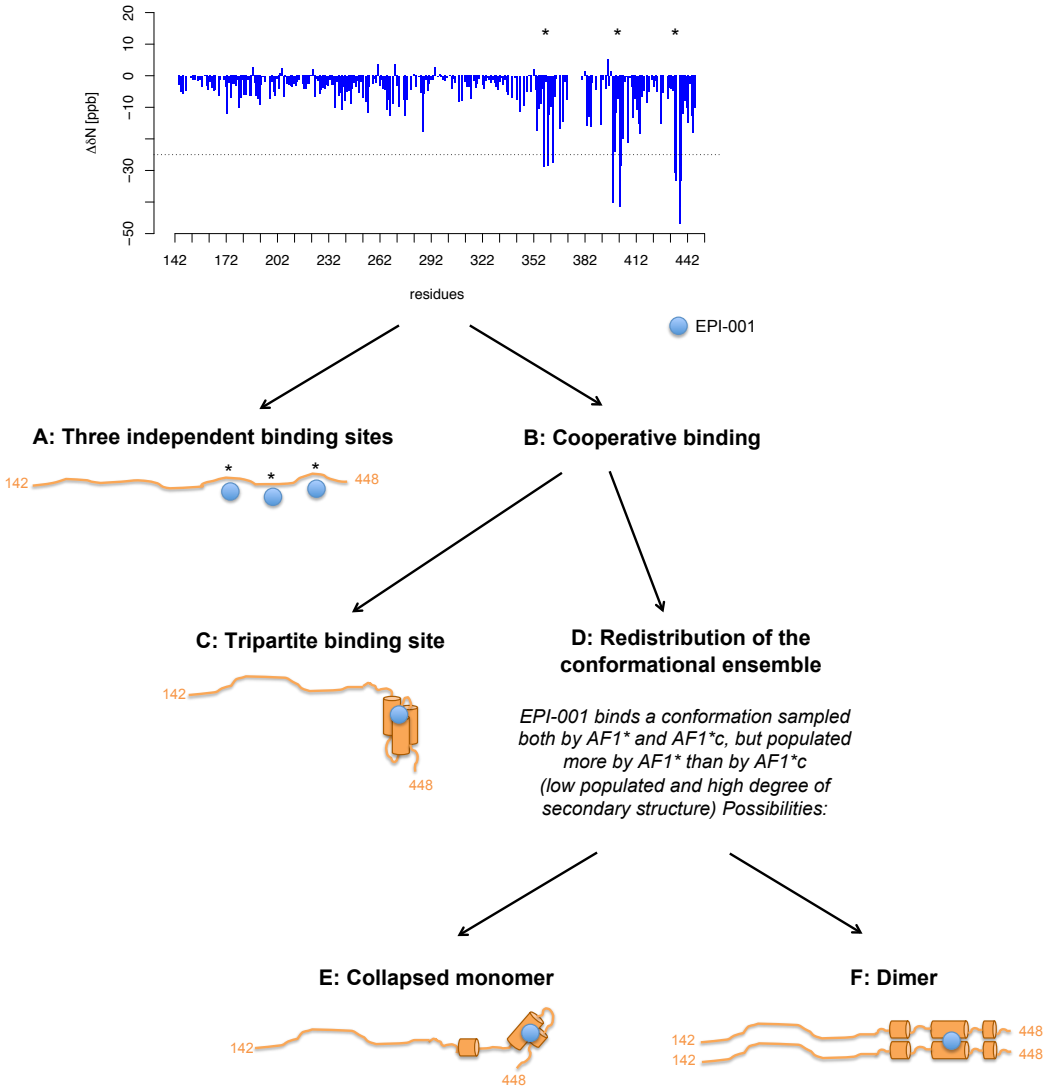


Figure 6.18: Schematic representation of the various possibilities underlying the specificity of EPI-001 for the Tau-5 region of AF1*. Interaction of AF1* and EPI-001 might involve three independent binding sites (option A) or instead be cooperative (option B). In case of cooperativity, the binding site can be tripartite, i.e. consisting of residues far in sequence but that are structurally connected (option C) or instead interaction with EPI-001 can involve the redistribution of the AF1* conformational ensemble (option D). A redistribution of the conformational ensemble may be related to different more extended and more collapsed conformations of the AF1* monomer (option E) or involve the monomer-dimer equilibrium of AF1* (option F). The AF1* regions in which chemical shift perturbations are observed upon addition of EPI-001 are indicated with the symbol *.

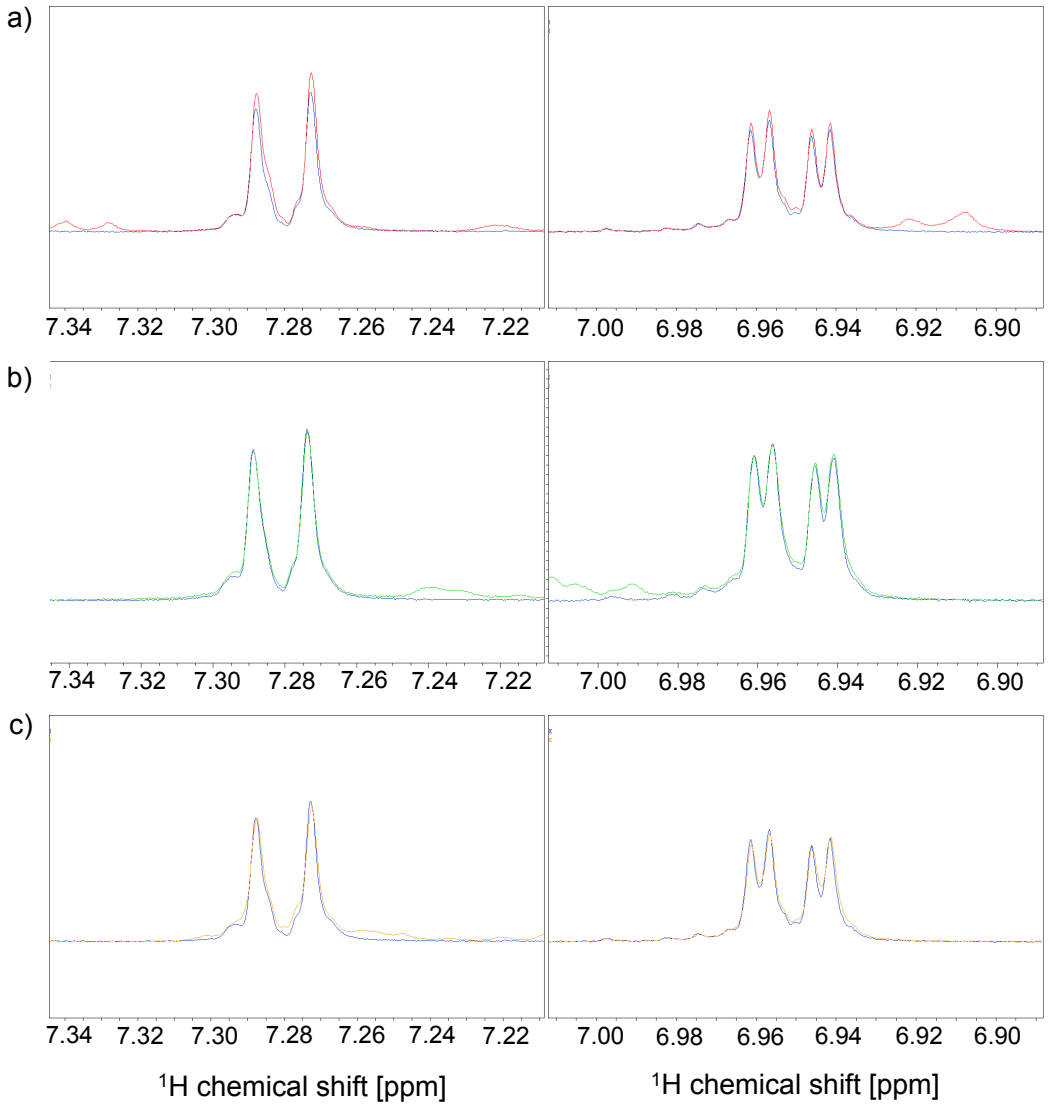


Figure 6.19: Effect of 0.25 molar equivalents of R1', R2' and R3' peptides on the chemical shifts of EPI-001 in a 1D ^{1}H spectrum. a) Blow-up of the multiplet signal centered around 7.28 ppm ($m_{7.28}$, corresponding to **8** in the assignment) (left) and the multiplet signal centered around 6.95 ppm ($m_{6.95}$, corresponding to **7** in the assignment) (right) for 100 μM EPI-001 in absence (blue) and presence (red) of 25 μM R1' peptide. b) Idem to a, but for 100 μM EPI-001 in absence (blue) and presence (green) of 25 μM R2' peptide. c) Idem to a, but for 100 μM EPI-001 in absence (blue) and presence (orange) of 25 μM R3' peptide. All samples contained 0.5% (v/v) dioxane-d_8. Both the blank and the sample for R2' peptide contained in addition 1% DMSO-d_6. This was necessary due to the low solubility of this peptide (only one charged residues for a 24-amino acid peptide). The resolution was 0.20 Hz, i.e. 0.00034 ppm (600 MHz spectrometer).

It is therefore also difficult to imagine a collapsed state of the Tau-5 region in which residues from the three contiguous regions are close in space to offer a tripartite binding site for EPI-001 interaction (i.e. option C in Fig. 6.18 is unlikely to occur). Therefore, we believe that interaction with EPI-001 causes a slight redistribution of the AF1* conformational ensemble, favoring the conformation EPI-001 is binding.

Our next goal was to identify which was the conformation Tau-5 adopts that EPI-001 can specifically recognize. Given that the interaction of EPI-001 requires a specific conformation of the Tau-5 region that is low populated in the conformational ensemble of AF1* in the absence of the small compound, this conformation is likely to be a relatively well-defined state with a fairly high structural content. Possible AF1* conformations that EPI-001 can interact with include one in which the Tau-5 region is fairly collapsed and possibly stabilized by helix to helix interactions between the various stretches of sequence with helical propensity (option E in Fig. 6.18), and a conformation in which the Tau-5 region of two monomers is brought together in a head-to-head orientation (option F in Fig. 6.18).

We have shown in chapter 4 that the conformational properties of the residues between A330 and C448 are the same in the context of the AF1*c protein construct (AR 330–448) and of the AF1* construct (AR 142–448). It is therefore reasonable to assume that the conformation adopted by the Tau-5 region and recognized specifically by EPI-001 is self-contained within the protein region spanning residues 330–448 (see Fig. 6.6c).

We therefore anticipated that the molecular mechanism by which EPI-001 interacts with AF1* would be the same as for the shorter protein construct AF1*c. In order to test this hypothesis, we monitored the effect of AF1*c on the ^{1}H chemical shifts of EPI-001 under the same experimental conditions as used for AF1* (see Fig. 6.20 and table 6.5). Contrary to what we expected, addition of 0.25 molar equivalents of AF1*c to EPI-001 caused smaller chemical shift changes in the EPI-001 resonances than those obtained for the same protein-to-ligand ratio for AF1* and EPI-001 (1.12 Hz for AF1*c vs. 1.59 Hz for AF1* for the multiplet signal centered around 7.28 ppm ($m_{7.28}$, corresponding to **8** in the assignment)).

These observations confirm that the conformation EPI-001 specifically interacts with is also sampled by AF1*c, but the smaller perturbations of the EPI-001 ^{1}H resonances indicate that the small molecule interacts with a lower affinity with AF1*c than with AF1*. That the affinity of EPI-001 to interact with AF1*c is lower than that for AF1* either indicates that a region of AF1* located outside residues 330–448 is necessary for interaction, or that the conformation specifically recognized by EPI-001 is populated to

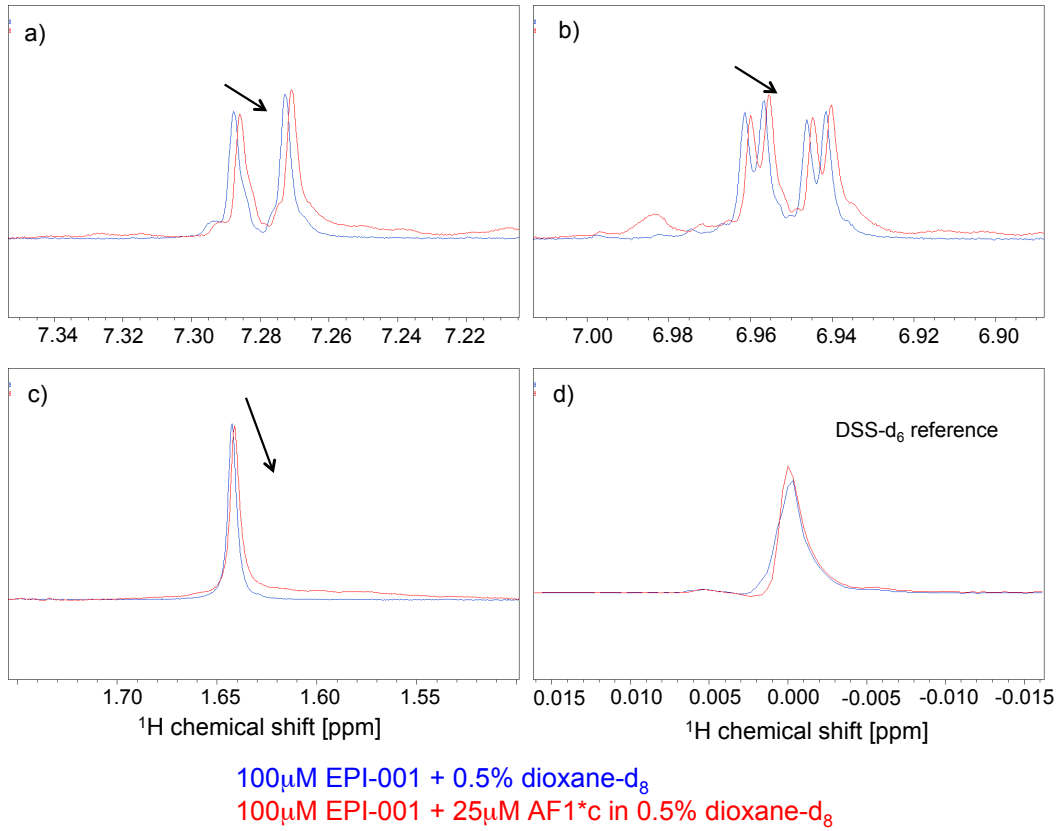


Figure 6.20: Effect of 0.25 molar equivalents of AF1*c on the chemical shifts of EPI-001 in a 1D ^{1}H spectrum. a) Blow-up of the multiplet signal centered around 7.28 ppm ($m_{7.28}$, corresponding to **8** in the assignment) for 100 μM EPI-001 in absence (blue) and presence (red) of 25 μM AF1*c. b) Idem as a) but for the multiplet signal centered around 6.95 ppm ($m_{6.95}$, corresponding to **7** in the assignment). c) Idem as a) but for the methyl singlet (s, corresponding to **9** in the assignment). d) Blow-up of the DSS-d$_6$ signal from the different samples, which was used for reference purposes. All samples contained 0.5% (v/v) dioxane-d$_8$. The resolution was 0.20 Hz, i.e. 0.00034 ppm (600 MHz spectrometer).

Table 6.5: The chemical shift changes observed, in Hz, relative to the resonances of 100 μM EPI-001 in the absence of AF1*c for each of the EPI-001 resonances corresponding to the aromatics (the multiplet signal centered around 7.28 ppm ($m_{7.28}$, corresponding to **8** in the assignment) and the multiplet signal centered around 6.95 ppm ($m_{6.95}$, corresponding to **7** in the assignment)) and the methyl group (s) of EPI-001 (the resonances corresponding to the aliphatics of EPI-001 are not shown but they follow the same trend). The resolution for the ^{1}H chemical shift was 0.20 Hz.

	$m_{7.28}$	$m_{6.95}$	s	molar equivalents
100 μM EPI-001	0	0	0	0
100 μM EPI-001 + 25 μM AF1*c	1.12	0.85	0.77	0.25

a lower extent by AF1*c than by AF1* at a given concentration. The former possibility can be excluded because no significant chemical shift changes were observed in residues of AF1* outside of residues 330–448 upon interaction with EPI-001.

Since the conformational properties of AR 330–448 are the same in the context of AF1* and AF1*c, it is expected that the various more extended and more collapsed conformational states of the monomer are populated to the same extent in both protein ensembles. Therefore, if interaction of EPI-001 takes place with a more collapsed monomeric state, we would anticipate that this conformation is equally populated in the ensemble of both proteins and hence interaction of EPI-001 with each of these proteins under the same conditions of buffer, temperature and protein-to-ligand ratio is expected to cause chemical shift changes of identical size. Our data therefore suggest that EPI-001 does not bind to a more collapsed state of the monomer (i.e. option E in Fig. 6.18 is unlikely).

We have shown in chapter 4 that the population of the dimer is smaller for AF1*c than for AF1* at the same protein concentration, and that the conformational properties of the AR 330–448 region in the dimer are the same in the context of AF1* and AF1*c. Our results are therefore compatible with a specific interaction of EPI-001 with the Tau-5 region when it adopts the conformation it has in the dimer (i.e. option F in Fig. 6.18 is compatible with the data).

Furthermore, the resonances of the AF1* residues that are affected by the interaction with EPI-001 shift in the same direction, both in the ^{1}H and ^{15}N dimension, upon addition of the small molecule as when the concentration of AF1* is increasing (see Fig. 6.21). Since at higher concentrations of AF1* the dimer is higher populated than at lower concentrations, this observation gives further support to the hypothesis that EPI-001 interacts specifically with the AF1* dimer and that interaction shifts the monomer-dimer equilibrium slightly towards the dimer (in agreement option F in Fig. 6.18).

If EPI-001 indeed interacts specifically with the dimer conformation of the Tau-5 region, this further explains why the observed chemical shift changes upon interaction are so small because the AF1* dimer is extremely low populated at 25 μM AF1*. As a consequence, the small chemical shift changes observed report only on an apparent binding affinity, whereas the actual binding affinity –for EPI-001 interaction with the AF1* dimer– is (expected to be) considerably higher (i.e. stronger binding), in agreement with the observed biological effects by Andersen *et al.* [134].

In conclusion, our data collectively show that the molecular basis for the specificity of EPI-001 for the Tau-5 region of the AR is the recognition by EPI-001 of a particular conformation sampled by the Tau-5 region, possibly the one Tau-5 adopts when it is in

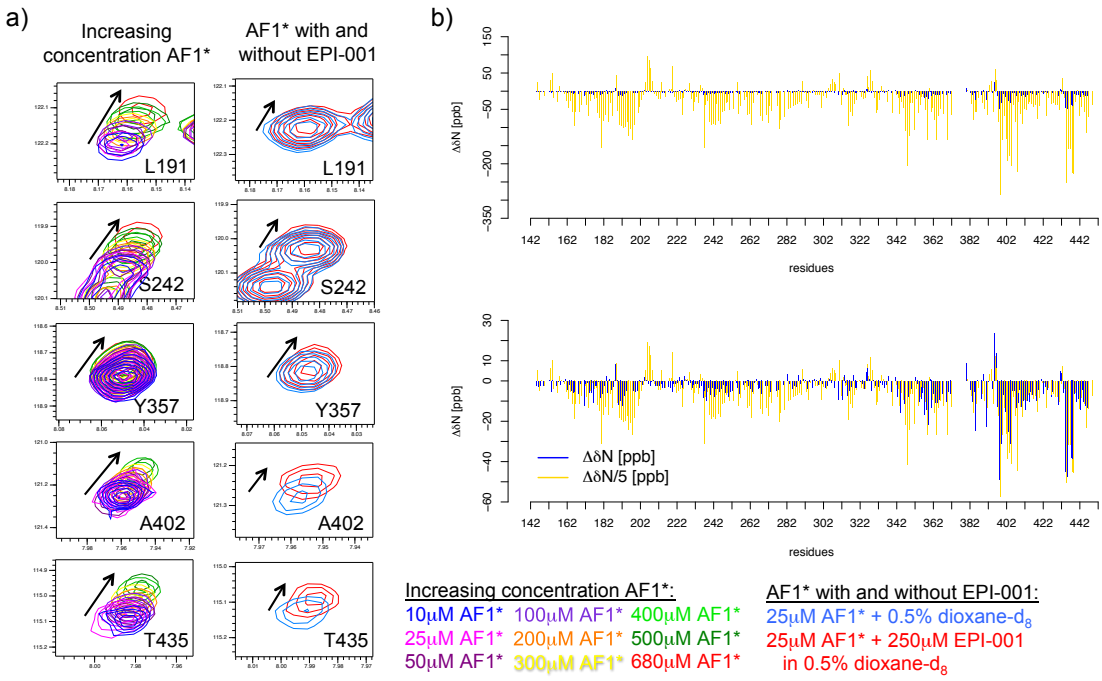


Figure 6.21: a) Blow-ups of selected residues from the HSQC spectra of AF1* at increasing concentration (left) and from the HSQC spectra of AF1* in the absence and presence of 10 molar equivalents EPI-001 (right). b) The ^{15}N chemical shift changes of AF1* caused by interaction with EPI-001 as a function of residue number (blue) overlaid with the ^{15}N chemical shift changes for AF1* at different concentrations relative to its chemical shifts at 25 μM (gold) (top). Same plots, but with the ^{15}N chemical shift changes for AF1* at different concentrations relative to its chemical shifts at 25 μM scaled (/5) (bottom).

the dimer state.

6.4 Covalent binding of AF1* and EPI-001

Evidence for the irreversible binding between EPI-001 and AR AF1 both *in vivo* and *in vitro* was provided by Myung *et al.*, although the binding site and the stoichiometry of the reaction remain elusive [160]. The chlorohydrin moiety of EPI-001 is highly susceptible to covalent reactions with nucleophilic amino acids such as cysteine, potentially through the formation of an intermediate epoxide which can be opened by a nucleophilic attack of an amino acid side chain as suggested in the proposed reaction mechanism by Myung *et al.* (see Fig. 6.2). The AF1* protein construct contains eight cysteine residues (C175, C238, C265, C282, C288, C325, C404 and C448) and the entire NTD domain (AR 1–559) contains an additional four (C123, C129, C518 and C559). For an irreversible inhibitor to be effective as a drug, high specificity for the target is required. To determine where covalent binding takes place, we used a combination of trypsin digestion and mass spectrometry. To promote covalent reaction, mixtures of AF1* and EPI-001 were incubated at 310 K for several hours. As shown in figure 6.22, covalent reaction between AF1* and 10 molar equivalents EPI-001 has taken place after 16 hours incubation at 310 K. At different time points, the reaction was stopped by addition of an excess of 2-iodoacetamide, which binds covalently to any remaining free nucleophiles in the protein (i.e. reactive side chains that have not reacted with EPI-001). Subsequently, the samples were subjected to trypsin digestion followed by mass spectrometry analysis of the obtained peptides. This approach allows the identification of the binding site(s) of EPI-001 on AF1*. These experiments were performed by Dr. Christopher Phang (IRB Barcelona) in collaboration with the IRB MS core facility.

The results of trypsin digestion and following mass spectrometry analysis for 25 μM AF1* + 250 μM EPI-001 (0.5% dioxane) after incubation at 310 K for different times are shown in figure 6.23. EPI-001 was only found to react with cysteine residues and not with any other amino acids with a nucleophilic side chain. Contrary to what we expected, EPI-001 did not react specifically with a particular cysteine residue, but could instead react with several cysteines present in AF1* with no particular preference to interact with cysteines in more disordered or more ordered parts of the sequence. Interestingly, we observed that only one EPI-001 molecule bound per protein molecule in all cases, although the identity of the cysteine residue could differ. At 10 molar equivalents of EPI-001, out of the eight cysteine residues, two were found to react in the first four hours of incubation (C265 and C404) (see Fig 6.23b and c), and when incubation was extended to 24 hours four additional cysteines were observed to form

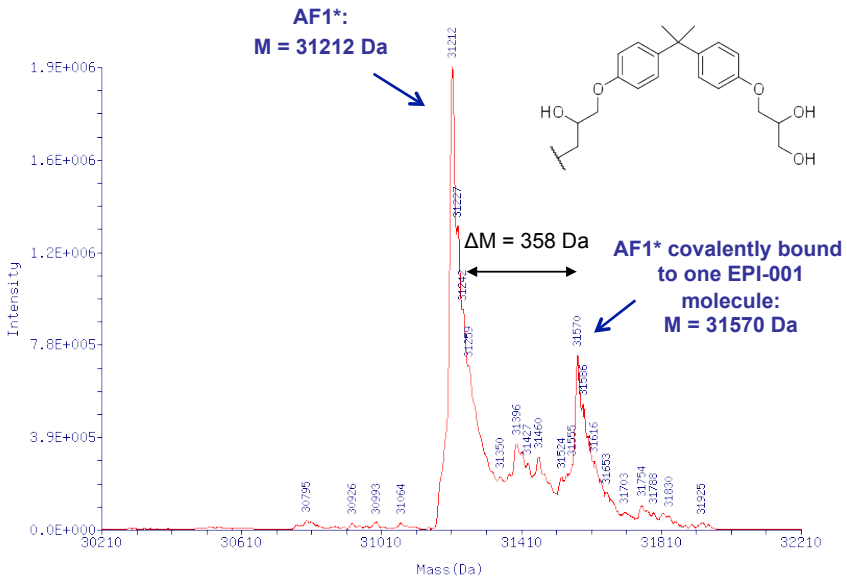


Figure 6.22: Mass spectrometry analysis of a mixture of AF1* and 10 molar equivalents EPI-001 (0.5% dioxane-d_8) after 16 hours incubation at 310 K. The presence of a species with molecular weight corresponding to AF1* covalently bound to one EPI-001 molecule (M=31570 Da) shows that covalent reaction has taken place under these incubation conditions.

a covalent bond with AF1* (C238, C282, C288 and C325) (see Fig 6.23c). The extent of EPI-001 covalent binding for a given cysteine residue was quantified as the ratio between the number of identified peptides after trypsin digestion corresponding to a given cysteine and the number of those same peptides to which EPI-001 was covalently bound.

However, at 10 molar equivalents of EPI-001, the small compound will not only interact specifically (and reversibly) with the Tau-5 region of AF1*, but due to the excess it will also be able to undergo unspecific interaction with the rest of the protein. Consequently, if any selectivity of EPI-001 to react covalently with a specific nucleophile exists, it will be lost under experimental conditions of a protein-to-ligand ratio of 1:10. In addition, the population of AF1* dimer is very small at 25 μM. Therefore, if our hypothesis that EPI-001 specifically recognizes the Tau-5 conformation when it is in the dimer is correct, the state EPI-001 interacts with will be higher populated at increased protein concentrations.

We therefore repeated the incubation experiment at 310 K at a protein-to-ligand ratio of 1:1.25, i.e. 200 μM AF1* + 250 μM EPI-001 (0.5% dioxane), which does not promote unspecific interaction between the two molecules and increases the concentration of the AF1* state EPI-001 is putatively binding. As can be seen in figure 6.24,

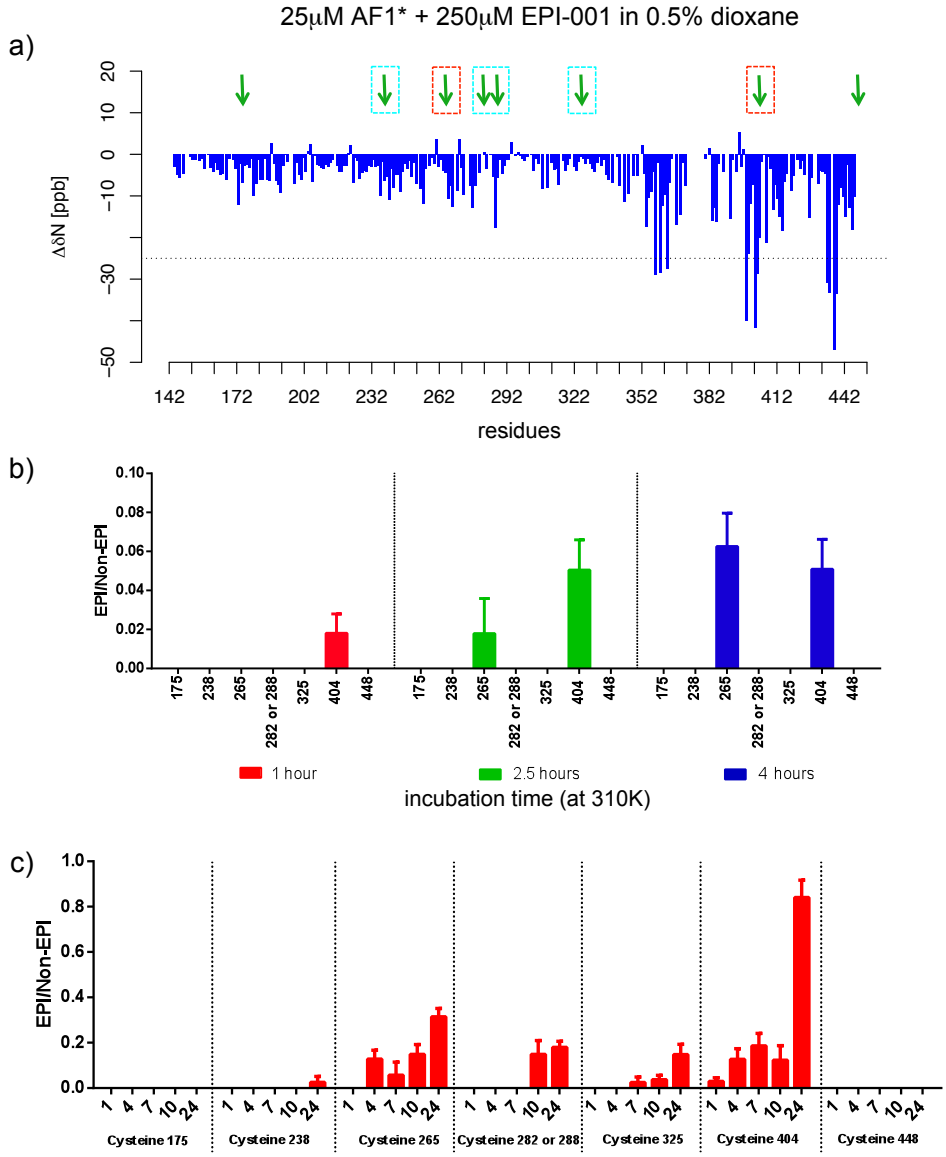


Figure 6.23: Study of the irreversible binding between AF1* and EPI-001 at protein-to-ligand ratio 1:10. a) The ^{15}N chemical shift changes of AF1* caused by interaction with 10 molar equivalents of EPI-001 as a function of residue number. The arrows indicate the position of the eight cysteine residues present in AF1* (C175, C238, C265, C282, C288, C325, C404 and C448). Residues that interacted covalently with EPI-001 under these experimental conditions in 4 hours or less are boxed in red, those that interacted covalently between 4 hours and 24 hours are boxed in cyan. b) and c) Results from trypsin digestion followed by mass spectrometry analysis of 25 μM AF1* + 250 μM EPI-001 (0.5% dioxane), indicating for each of the cysteine residues whether covalent reaction with EPI-001 had occurred after different times of incubation at 310 K . The ratio between the number of identified peptides after trypsin digestion corresponding to a given cysteine and the number of those same peptides to which EPI-001 was covalently bound is given on the y-axis of the plots in b and c.

under these experimental conditions C404 is the kinetically preferred residue to react irreversibly with EPI-001. This suggests that the irreversible reaction between AF1* and EPI-001 takes place specifically at C404, which is located in the center of the helical region R2 in the Tau-5 part of the sequence of AF1*.

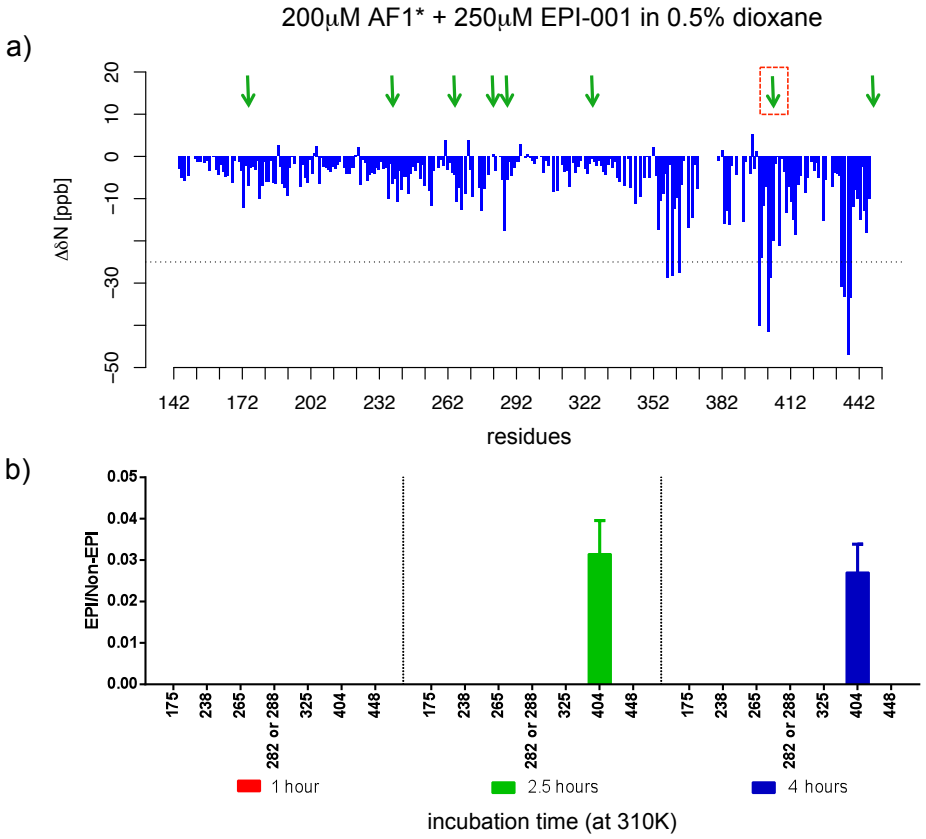


Figure 6.24: Study of the irreversible binding between AF1* and EPI-001 at protein-to-ligand ratio 1:1.25. a) The ^{15}N chemical shift changes of AF1* caused by interaction with 10 molar equivalents of EPI-001 as a function of residue number. The arrows indicate the position of the eight cysteine residues present in AF1* (C175, C238, C265, C282, C288, C325, C404 and C448). The arrow corresponding to cysteine residue 404 is boxed, indicating that this residue preferentially reacted with EPI-001 under the experimental conditions used. b) Results from trypsin digestion followed by mass spectrometry analysis of 200 μM AF1* + 250 μM EPI-001 (0.5% dioxane), indicating for each of the cysteine residues whether covalent reaction with EPI-001 had occurred after different times of incubation at 310 K. The ratio between the number of identified peptides after trypsin digestion corresponding to a given cysteine and the number of those same peptides to which EPI-001 was covalently bound is given on the y-axis.

Taken together with our data obtained for the non-covalent step in the interaction, these results suggest that the interaction of EPI-001 and AF1* involves, first, the specific reversible recognition of the Tau-5 region, followed by the specific irreversible binding of the small compound to the nucleophile present in this region of the protein, cysteine 404. This provides a rationale for the specificity of EPI-001 for, not only

the NTD of AR, but more specifically for the Tau-5 region of AF1 and C404 for the irreversible interaction, at least in our *in vitro* system.

Even though the effect of other parts of the NTD and/or of the other domains of the AR on the structural properties of the NTD when bound to DNA is not known, it is possible that the small population of dimer we observe in AF1* will also be present in the fully matured dimer of the NTD when the AR is bound to DNA. It is therefore possible that the irreversible binding of EPI-001 also (first) takes place at C404 *in vivo*, and that the molecular determinants underlying the specificity of EPI-001 to interact with C404 are the same as for the *in vitro* system we have studied.

6.5 Speculation on the mode of action of EPI-001

EPI-001 represents a novel class of AR inhibitors that target the AR NTD directly. This is of extreme importance because the NTD has been identified to be essential for hormone-independent AR transcriptional activity, which allows late stage PCa cells to survive. It is therefore essential to understand the mechanism of action of this class of compounds to rationally design new lead compounds for drug discovery with improved properties, especially given the fact that the members of this class seem to be irreversible binders.

The results we have obtained using the AR protein construct spanning residues 142 to 448, overlapping well with AF1, give insights into the molecular mechanism of the interaction between these two molecules in this *in vitro* system, which is possibly similar to the mechanism of interaction *in vivo*.

Target specificity is crucial for irreversible inhibitors. We showed that, in our *in vitro* model, prior to reacting, EPI-001 specifically recognizes transactivation unit 5 in the transactivation domain of the AR and that this specificity is due to a particular conformation adopted by the Tau-5 region, possibly the one in the dimer. The subsequent specific covalent reaction with C404 is directly related to the positioning of EPI-001 in the proximity of this nucleophile in the reversible formation of the encounter complex, facilitating the specific interaction with this cysteine residue. Irrespective of C404 being the nucleophile where irreversible binding to AF1 takes place or not *in vivo*, we propose several mechanisms by which irreversible binding of EPI-001 to AF1 can influence the functional properties of the AR NTD in transcription.

Covalent binding of EPI-001 to any nucleophilic side chain of AF1* may influence the helical propensity, both locally and in regions that form long-range helix to helix interactions with the binding site. Such interactions might occur between regions far

in sequence, for instance between regions with helical propensity in AF1*c, or between two AR monomers in a dimeric conformation. Local (de)stabilization of the helical propensity due to EPI-001 binding may interfere directly with protein-protein interactions between the binding site of AR and co-regulatory proteins. C404 is located in the center of the region with highest helical propensity in the C-terminal part of AF1* (between residues 391 and 413, see Fig. 4.27 in chapter 4, p. 145). It is therefore likely that AR interacts with a still unidentified binding partner through this sequence of AR and that covalent binding of EPI-001 to this residue influences this interaction.

Furthermore, covalent binding of EPI-001 might also affect the interaction of other parts of the AR, distant from the residue to which EPI-001 binds, and their binding partners. A change in local helical propensity due to covalent binding could be reflected in the helical propensity of other parts of the sequence that form long-range contacts with the binding site, which in turn could influence the interaction of these regions with their respective binding partners. Helix to helix interactions most likely exist between region R2 and region R3 in the C-terminal region of AF1*, which are reflected in the higher degree of collapse of this part of AF1* in comparison with the entire AF1* protein. Mutation of arginine 405, the residue next to cysteine 404, to serine and subsequent phosphorylation has been reported to have an effect on the interaction of the ^{433}WHTLF^{437} motif with co-activator p300 [246], which further supports the existence of long-range interactions between these two parts of the sequence. Therefore, covalent binding of EPI-001 to C404 could potentially affect the interaction between the ^{433}WHTLF^{437} motif of AR and its binding partners, including RAP74, p300 and the AF2 region of AR located in the LBD (N/C interaction). All of these interactions modulate the transcriptional activity of AR, so enhancing or inhibiting them will affect the regulatory function of the NTD in transcription.

In addition, (de)stabilization of the helix to helix interactions that stabilize the dimer could occur following EPI-001 binding. Our data suggest that interaction with EPI-001 slightly shifts the equilibrium towards the dimer. Given that C404 is located in a region of the sequence through which the AF1* dimer is stabilized, covalent binding of EPI-001 to C404 might also influence the dimerization state of the AF1*. Potentially, this could disfavor the interaction of the sequence around C404 with a binding partner.

6.6 Summary

We have shown that EPI-001 specifically recognizes three contiguous regions of sequence in AF1* located between residues 345 and 448. Our results suggest that the interaction of EPI-001 and AF1* involves, first, specific reversible recognition of the Tau-5 region,

followed by the specific irreversible binding of the small compound to the nucleophile present in this region of the protein, cysteine 404. The molecular mechanism underlying the specificity of EPI-001 for the Tau-5 region in our *in vitro* system is the recognition by the small compound of a particular conformation adopted by the Tau-5 region, possibly the one this region adopts when it is in the dimer state. Possibly the molecular determinants of the interaction between EPI-001 and the AR NTD *in vivo* are the same as for the *in vitro* system we have studied.

7 General discussion

In this thesis we have studied the conformational properties of the AF1* region of the AR NTD at atomic resolution for the first time. We further investigated the molecular mechanisms by which it activates AR transcription and studied its interaction with EPI-001, a recently discovered irreversible AR inhibitor, at a molecular level.

We found that the NTD of AR displays high helical propensity in several regions of sequence, and that these coincide with regions that are considered crucial for transcriptional activity in Tau-1 and Tau-5. We observed that AF1* has a low but clear propensity to self-associate in a dimer with a head-to-head orientation and that AF1* interacts with the C-terminal domain of RAP74 via its ^{433}WHTLF^{437} motif, located in the Tau-5 region of AR and crucial for aberrant transcriptional activation in CRPC. Our data suggest that phosphorylation of several residues N-terminal to the ^{433}WHTLF^{437} motif is required for this interaction to take place. Importantly, we showed that the interaction between AR and RAP74 activates transcription in PC-3 CRPC cells at low levels of hormone. Finally, we determined that EPI-001 specifically recognizes a conformational state populated by the Tau-5 region, possibly the conformation that Tau-5 adopts in the dimer. In this chapter, a working model of the role of the NTD in AR transcriptional activation based on the obtained data from the three projects is proposed. This model is summarized in figure 7.6 and will be explained stepwise in this chapter.

7.1 Possible formation of an NTD dimer in DNA-bound AR

Our data show that, *in vitro*, the AF1* construct of the AR NTD has a low propensity to dimerize. This allows us to propose the existence of a dimeric state of AR NTD *in vivo*, which the NTD would adopt in a specific biological context.

The low population of dimer present in samples of AF1* limited its characterization. However, our results indicate that AF1* dimerizes in a head-to-head orientation and that this is a cooperative process in which regions with helical propensity in the monomer stabilize each other through helix to helix interactions. We anticipate that in the entire NTD (559 residues) additional stretches with helical propensity will be present that contribute to the stabilization of the dimer. Therefore, the propensity of the NTD (559 residues) to dimerize is likely to be higher than that of the AF1* region (308 residues).

In addition, in WT full-length AR, the NTD is linked to the DBD that binds as a dimer to AREs on DNA [210]. Therefore binding of the DBDs of two AR molecules to two adjacent AREs on DNA increases the local concentration of the NTD substantially compared to when AR is free to diffuse in solution. Given that the AF1* region has a propensity to self-interact and that this propensity is likely to be higher for the entire NTD domain, the tethering of both NTDs to the DNA upon DNA binding of AR is expected to facilitate this interaction considerably, resulting in the dimerization of the NTDs in a head-to-head orientation. Therefore, it is possible that an NTD dimer exists in the biological milieu when AR is bound to the DNA (see Fig. 7.1).

Even though it is well-established that AR dimerization is required for its transactivation activity [200], the nature of the AR dimer remains elusive and a potential role of the NTD in this process has not been described [155]. We propose here that the NTD can contribute to AR dimerization.

7.2 The nature of the NTD dimer

The conformational properties of such an NTD dimer in DNA-bound AR are likely to be influenced by many factors, including the residues of the NTD outside of the AF1* region, the DBD and the LBD, the DNA sequence to which AR binds, the presence of co-regulators, the presence of hormone and PTMs. Since it is not known what the effect of each of these factors is, we can only assume that the NTD adopts a dimer conformation in DNA-bound AR similar to that of dimeric AF1* in solution. Nevertheless,

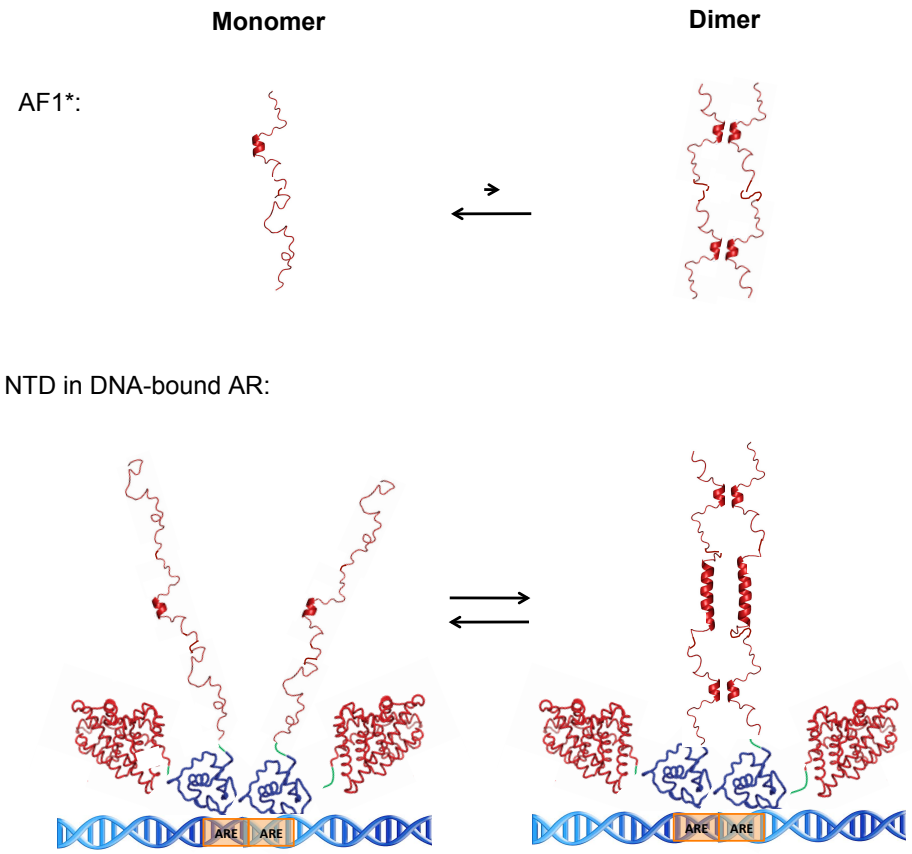


Figure 7.1: Schematic representation of how AF1* dimerization could lead to NTD dimerization in a head-to-head orientation for DNA-bound AR. We propose that in the DNA-bound context the monomer-dimer equilibrium of the NTD is shifted more towards the dimer conformation compared to the monomer-dimer equilibrium of AF1* in solution, due to the increased local concentration of the NTD in DNA-bound AR. The DBD is indicated as a blue globular domain and the LBD as a red globular domain, the DNA is shown in blue and androgen response elements as orange boxes.

if the main interactions stabilizing the NTD dimer conformation are cooperative and between transient secondary structure elements in the monomers, it is possible that the mechanism for dimerization is relatively simple and involves the NTD dimerizing like a "zipper" from the DBD to the N-terminus of the protein. Potentially this only represents the early stages of dimerization and a more mature conformation of the NTD dimer could for instance be stabilized by tertiary contacts. Notably, the formation of a zipper between domains of two monomers is commonly observed in transcription factors when they are bound to DNA. In the bZIP (basic-region leucine zipper) family of transcription factors, leucine zipper structural motifs serve as dimerization domains to form homo- or heterodimers [480]. The resulting structure is a left-handed parallel dimeric coiled-coil [480]. Dimerization of these transcription factors through the leucine zipper domains is required for their function in transactivation [294, 480]. There are many structural variations of the zipper. For instance, dimerization can also occur through basic helix-loop-helix leucine zipper (bHLHZip) domains. One of the best-studied examples of such an interaction is that of the heterodimer formed between the c-Myc oncoprotein and the developing controlling factor Max [294]. The individual monomers are disordered and undergo coupled folding and binding upon dimerization [308]. Regardless of the structural details of dimeric NTD, according to our model the NTD dimer is characterized by two features: a higher structural content than in the NTD monomer and a head-to-head orientation.

On the one hand, a higher helical content in the NTD dimer compared to the monomer is in agreement with the previous observation that binding of AR to the AREs caused increased structure in the otherwise ID NTD [154, 259].

On the other hand, the current view on the dimerization of full-length AR when the protein is bound to DNA is compatible with a head-to-head orientation of the NTD dimer [7, 155]. Even though it is uncertain and even unlikely that dimerization of the AR LBD occurs [155], a head-to-head orientation of the DBD domains was observed in the crystal structure of the DNA-bound (rat AR) DBD dimer (PDB 1R4I, [210]). Furthermore, in the crystal structure the residues N-terminal to the DBD domain, connecting the DBD to the NTD, are located in flexible regions (see Fig. 7.2). This potentially allows the interaction of the two NTD domains in a head-to-head orientation as suggested by our model.

Two recently proposed models of the DNA-bound AR homodimer [155] are based on the open conformation of the PPAR-RXR heterodimer [221] and the closed HNF-4α homodimer [219] (see Fig. 7.3). Both models propose a head-to-head orientation of the AR DBDs and LBDs, with contacts between the DBDs. The positions of the LBDs are quite uncertain due to the lack of LBD dimerization and the undefined role of the

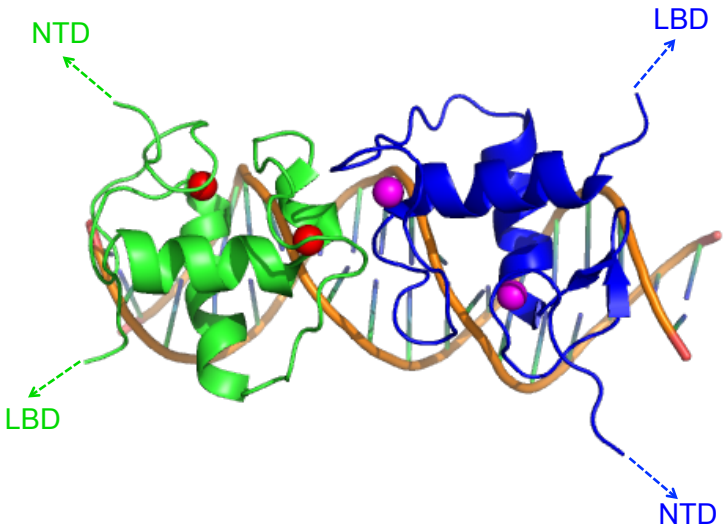


Figure 7.2: Top view of the crystal structure of the AR DBD dimer on DNA (PDB 1R4I, [210]). The DBD domains of both AR molecules are shown in green and blue. The Zn^{2+} atoms are shown as red and magenta spheres. The position of the NTD and LBD domains, which are not included in the crystallized complex, are indicated by arrows. Green arrows correspond to the NTD and LBD domains of the AR DBD shown in green, and blue arrows correspond to the NTD and LBD domains of the AR DBD shown in blue.

NTD in the DNA binding complex [155].

Importantly, the structures of (nearly) full-length NRs bound to DNA that are available all show (homo- or hetero-) dimerization of the LBDs [218–221, 481]. LBD dimerization is well-described for several NRs (VDR, TR, PPAR, RAR and RXR) and is said to facilitate dimerization via the DBDs [155]. However, there is no clear evidence LBD dimerization occurs for SRs (GR, MR, PR and AR), and it is even unlikely that such LBD-LBD interactions ever occur for these proteins [155]. Instead, for SRs (AR, MR, PR-B and ERα) interactions between the N-terminal and C-terminal domains (N/C interactions) have been reported [236, 268–272, 482].

For AR this N/C interaction is well-characterized and is regulated in a spatiotemporal way [7]. However, it is still unclear whether N/C interaction takes place when AR is bound to DNA. The results from van Royen *et al.* strongly suggested that in DNA-bound AR the N/C interaction is lost [278]. It was proposed that abolishing of the N/C interaction upon DNA binding allows interaction with co-activators to initiate transcription [7, 278]. Data from Klokk *et al.*, on the other hand, indicate that there are significant intramolecular N/C interactions in AR in its DNA-bound and transcriptionally active state [280]. Intermolecular N/C interactions were also observed in DNA-bound AR, but to a lesser extent (approximately 15% of the observed N/C interactions) [280]. Possibly this depends on the context including factors like the

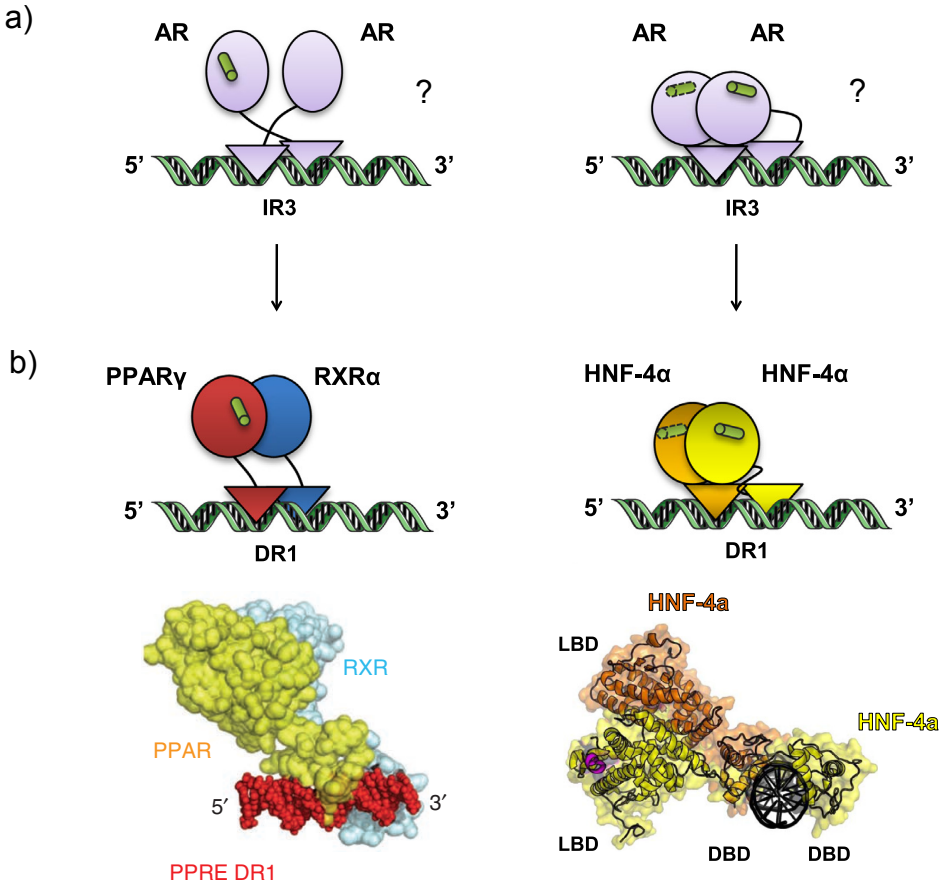


Figure 7.3: Current view on the dimerization of AR in the DNA binding complex. a) Hypothetical full-size binding mode of the AR based on the open conformation of PPAR-RXR (left) and the closed HNF-4α homodimer (right), taken from [155]. DBDs are depicted as triangles, LBDs as circles and co-activators as green cylinders. IR3 is a response element composed of two inverted half-sites, spaced by three nucleotides. b) Similar cartoon structures of the open conformation of the PPAR-RXR heterodimer (top left) and HNF-4α homodimer (top right) (taken from [155]), with the solution structure of PPAR-RXR (bottom left) [221] and X-ray structure of the HNF-4α homodimer (bottom right) [219]. DR1 is a response element composed of a direct repeat of two half-sites, spaced by one nucleotide.

promoter sequence [244, 273] and the presence of hormone as the N/C interaction is ligand-induced [483]. If the N/C interaction occurs when AR is bound to the DNA, this will undoubtably affect the interdomain communication within AR, but it does not necessarily exclude the existence of an NTD homodimer as one possible state. This will be further discussed later in this chapter. If, however, this N/C interaction does not occur in DNA-bound AR, it is very likely the NTD domains of both AR molecules, tethered to the DNA, would interact in a head-to-head orientation given their propensity to self-interact.

We therefore propose that in the absence of an N/C interaction when AR is bound to DNA, AR dimerizes in a head-to-head orientation and that the head-to-head dimerization of the NTD domain further stabilizes this dimer.

7.3 The NTD dimer is suggested to be weak and in equilibrium with monomer

Even though the NTD might adopt a dimeric conformation when AR is bound to DNA, we propose that in DNA-bound AR the NTD is not predominantly present as a dimer, but that instead it is in equilibrium between a monomeric and a dimeric state. This hypothesis is supported by the following observations:

Our NMR data obtained at different concentrations of AF1* and AF1*c (see chapter 4) indicate that the affinity of these protein constructs for themselves is low, with a dissociation constant in the order of ca. 1 mM for the dimer. Therefore, even if the full-length NTD is tethered to the DNA via the DBD, it is unlikely that an NTD dimer would self-associate with an affinity as high as low micromolar to nanomolar and yield a fully populated dimer.

Furthermore, no electron density was observed for the NTD region in crystals obtained in an attempt to crystallize full-length AR in the presence of the hormone, DNA elements and peptides of co-activator motifs [287], indicating that the NTD did not adopt a stable globular-like conformation.

In spite of the limited structural information available for complexes between the AR NTD and its binding partners, it is known that at least some of these interactions require a monomeric state of the NTD. Binding of the ^{23}FQNLF^{27} motif of AR to the LBD involves the folding-upon-binding of this region of sequence into a helix (PDB 1XOW, [191]). The binding mode is highly similar to that of LxxLL motifs in various co-activators that bind to the LBD, and requires a monomeric state of the NTD.

Interaction of the ^{433}WHTLF^{437} motif with the LBD presumably happens through a similar mechanism, although there is no structure available for the wild type WHTLF sequence and the LBD. We have shown that the ^{433}WHTLF^{437} motif interacts with the C-terminal domain of RAP74. The structural model we propose involves folding of the ^{433}WHTLF^{437} motif into a more helical conformation that docks in a hydrophobic pocket of RAP74. This binding mode also implies a monomeric conformation for the NTD.

Moreover, it has become increasingly clear in recent years that ID is essential for the regulation of complex cellular processes like transcription [151]. This underlines the importance of monomeric NTD, which is ID, for the function of AR. Recently, the importance of ID for allostery in nuclear receptors, and transcription factors in general, was reviewed by Hilser and Thompson [154]. They describe disorder as the key to the NTD functions because it optimizes allosteric couplings between the NR domains and their binding partners. Since the allosteric properties of NRs are essential to their function in the regulation of cellular processes the presence of ID domains is crucial. In addition to being important for allostery, ID also facilitates interaction of a particular motif with different binding partners depending on context [305, 484]. In the NTD of AR several motifs have been described to interact with multiple binding partners. This undoubtably is facilitated by the ID nature of the NTD when it is in a monomeric conformation. ID offers many other advantages over a globular structure in the regulation and fine-tuning of cellular processes. It allows, for instance, transient interactions to occur with weak affinities but high specificity [305].

7.4 Possible implications of dimerization for the biological function of AR NTD

Potentially a monomer-dimer equilibrium of the AR NTD has implications for its biological function(s). The NTD dimer could be a transcriptionally active or inactive conformation. Based on the arguments above, it is possible that the dimer state is an inactive state of the AR. Nevertheless, many of the NTD interaction partners have been identified by techniques like co-immunoprecipitation or yeast-two-hybrid (Y2H) and for many of these complexes no structural information is available. We cannot exclude the possibility that the NTD interacts as a dimer with some of its binding partners. A dimer conformation is expected to have a higher secondary structure content compared to the monomer, and this increase in structure could create new binding surfaces for particular partners. Alternatively, the NTD could be transcriptionally active as a monomer and adopt an auto-inhibitory dimer conformation to inhibit transcriptional

activity when it is not required. This notion is supported by our findings that RAP74 binds to monomeric AR NTD to activate transcription. Furthermore, our data suggest that interaction of EPI-001 with AF1* may stabilize the AF1* dimer. If we assume that in DNA-bound AR the NTD dimer is similarly stabilized by EPI-001 interaction, this is a further indication that dimeric NTD is a transcriptionally inactive state since EPI-001 inhibits AR transcriptional activity.

The reverse correlation between AR activity and the length of the polyQ stretch in the AR NTD [249, 250, 273] is a further indication that the NTD dimer is an inactive conformation. As explained in the introduction, the polyQ stretch in AR is involved in SBMA. In healthy individuals the polyQ repeat contains 17–26 residues, whereas in SBMA patients repeats of 40–52 glutamine residues were found [248]. When the number of glutamine residues in the repeat is expanded over the threshold, the polyQ region aggregates into amyloid fibrils. The longer the polymorphic polyQ stretch in the protein, the more aggregation prone AR and hence the higher its affinity to self-interact. We therefore deduce that the NTD dimer will be more stable for longer polyQ lengths. The observation that AR molecules with longer polyQ stretches show a decreased transcriptional activity could therefore be related to a higher stabilization of the NTD dimer for longer polyQ repeats. Furthermore, a shorter polyQ repeat is correlated with an increased risk for PCa. [254–256] Similarly, a shorter polyQ stretch could be related to a less stable dimer favoring monomeric NTD. If the latter is indeed the transcriptionally active state of the NTD this could lead to an increased transactivation activity.

Shorter polyG stretches are related to an increased AR activity [251] and an increased risk for PCa [256]. Potentially shortening the polyG stretch reduces the flexibility of the region of sequence linking Tau-1 and Tau-5 to the DBD, thus preventing efficient dimerization of the NTD of two DNA-bound AR molecules. This would further indicate that a lower population of dimer is associated with increased AR activity, supporting the notion that the NTD is a transcriptionally inactive state.

We have shown that several regions of sequence of the NTD that have been identified to interact with co-regulatory proteins are involved in dimerization (see chapter 4), including the ^{179}LKDIL^{183} motif, the ^{183}LSEASTMQLL^{192} region (binding site for TAB2 as a component of an NCoR co-repressor complex, [242]), the ^{234}AKELCKAVSVSMGL^{247} ANTS sequence to which CHIP binds [247] and the ^{433}WHTLF^{437} motif. Possibly in the dimer conformation the NTD "hot spots" for interaction with binding partners are preferentially involved in dimer-dimer contacts and are therefore not available to interact with co-regulatory proteins. Activation of transcription would then require (local) destabilization of the dimer to allow interaction with co-regulators.

Extrapolation of our data to the biological setting suggests that EPI-001 interaction may stabilize the NTD dimer of DNA-bound AR. If the NTD dimer is indeed a transcriptionally inactive state, stabilization of the NTD dimer by EPI-001 interaction might contribute to the mode of action of this small compound (see Fig. 7.6). On the one hand, covalent binding of EPI-001 to the AR NTD could interfere with protein-protein interactions between the AR NTD and co-regulatory proteins (that bind either directly to the binding site of EPI-001 in AR or to regions in long-range contact with the binding site of EPI-001 in AR), as discussed in chapter 6. On the other hand, interaction of EPI-001 with the AR NTD could stabilize a transcriptionally inactive dimer of the NTD and thereby directly inhibit AR transcriptional activity. Possibly both mechanisms contribute to the mode of action of EPI-001.

Stabilization of the NTD dimer by EPI-001 would further explain how this small molecule interferes with the interaction of AR and RAP74. According to our model, interaction with EPI-001 depletes monomeric NTD, the NTD conformation to which RAP74 binds (see Fig. 7.6). This view might be extended to other protein-protein interactions that EPI-001 affects, and is consistent with the fact that EPI-001 has been reported to weaken several protein-protein interactions between the AR NTD and co-regulators or members of the transcription machinery. [485]

If transcriptional activity of AR indeed requires a monomeric conformation of the NTD to interact with co-regulatory proteins, the formation of an irreversible NTD dimer would abolish the functional role of the NTD in transcriptional activation. Given that the EPI-001 molecule only contains one chlorohydrin moiety it can only react with AF1 via one covalent bond. Therefore it is unlikely that interaction of AR and EPI-001 creates an irreversible NTD dimer. However, molecules designed to simultaneously bind to the AF1 region of two AR monomers in the dimer conformation could potentially lock the NTD in an inactive dimer conformation. A candidate for such a compound could be the dichlorohydrin analog of EPI-001, which is a symmetric molecule with two chlorohydrin moieties (see Fig. 7.4). Due to the extremely low solubility of this compound in aqueous buffers, it is unlikely that such a compound would be included in a high-throughput screening assay and identified as a lead compound. However, a better understanding of the molecular mechanism of the interaction of AF1 and EPI-001 aids in the rational design of new compounds which has led us to this hypothesis. Under the assumption that the dimer of the NTD is a transcriptionally inactive conformation, we therefore propose that molecules designed to create an irreversible dimer of the NTD might have a substantial potential to prevent aberrant transactivation of AR in late-stage PCa. Since, according to our model, the NTD has conformational properties similar to those of a globular protein when AR is bound to the DNA, it can be thought

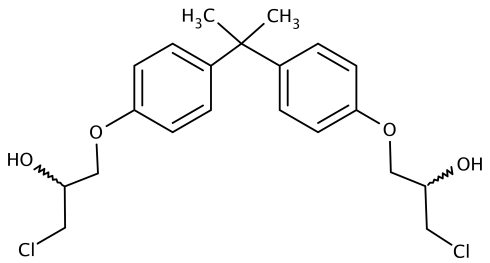


Figure 7.4: Molecular structure of the dichlorohydrin analog of EPI-001. The diol functionality of EPI-001 is replaced by a chlorohydrin group, which leads to a symmetric molecule with the potential to form two covalent bonds.

of as a "conventional" drug target instead of an IDP in this context. Potentially the presence of "druggable surfaces" in the NTD dimer when the protein is bound to DNA facilitates the design of small molecules that specifically interact with this domain of AR.

At first sight there is no clear common mechanism for the different members of the new class of AR inhibitors to which EPI-001 belongs. Sintokamide A is also chlorinated so it could potentially bind covalently to the AR, but it has not been reported to do so [139]. Niphatenone B on the other hand does not contain chlorine atoms but is described to bind irreversibly [159]. Potentially what these compounds have in common is that they specifically recognize a conformation of the NTD dimer and stabilize the inactive NTD dimer conformation by interaction. This would lead to a decrease in the transcriptional activity mediated through the NTD. Like EPI-001, these small molecules could have a combined mode of action in which they do not only stabilize the NTD dimer conformation of DNA-bound AR, but in addition interfere with protein-protein interactions between the NTD and co-regulatory proteins via different mechanisms.

Taken together, we propose that monomeric NTD is a transcriptionally active state and the NTD dimer a transcriptionally inactive state. Furthermore, we suggest that the mode of action of small molecules that specifically inhibit AR transactivation mediated through its NTD involves the stabilization of such a transcriptionally inactive NTD dimer.

7.5 Modulation of the NTD monomer-dimer equilibrium to regulate transcription

The transcriptional activity of nuclear receptors, and transcription factors in general, is well-regulated to react quickly to environmental changes. A monomer-dimer equi-

librium of the AR NTD could represent an additional layer of regulation for AR transcription. Possibly this is a more general mechanism that can also occur for other NRs. In turn, this monomer-dimer equilibrium can potentially be influenced by a multitude of factors to further fine-tune specific responses under particular conditions in the biological milieu. Possible factors that affect the monomer-dimer equilibrium include disorder-to-order transitions, allosteric mechanisms, long-range contacts, the presence or absence of hormone, PTMs, the nature of the DNA sequence that is bound, binding of small molecules and the presence or absence of the LBD and/or the N/C interaction.

Disorder-to-order transitions of regions of the NTD upon dimerization could induce structural changes locally and/or in remote regions due to allosteric mechanisms and/or long-range contacts, which could in turn change the affinity of such regions for their binding partners. PTMs could directly affect the structural properties of the NTD, but could also modulate the binding specificity of linear motifs and thereby affect the interaction of a specific region of the NTD with its binding partner. Depending on the DNA sequence that is bound by the DBDs, the orientation the DBD dimer adopts tends to be different. Typically, when NRs bind to a DNA sequence with direct repeats the DBDs dimerize in a head-to-tail orientation (e.g. VDR DBD bound to a DR3 element, [486]) and when they bind to indirect repeats they usually adopt a head-to-head orientation (e.g. ER DBD and GR DBD bound to IR3 elements, [208, 209]). This could have consequences for the orientation of the NTD and LBD domains attached to the dimerized DBDs, which potentially impedes NTD-NTD interaction to form an NTD dimer in some cases.

7.5.1 Expected effect of N/C interaction on the monomer-dimer equilibrium of the NTD

In the absence of an interaction between the NTD and LBD domains of AR (N/C interaction) when the protein is bound to DNA, our data suggest that the model presented thus far is representative of the DNA binding complex (see Fig. 7.6). As mentioned before in this chapter it is, however, not clear from literature whether N/C interaction is lost or maintained when AR binds to DNA [278, 280].

The N/C interaction takes place primarily through the ^{23}FQNLF^{27} motif of AR and the AF2 region in the LBD, but also the ^{179}LKDIL^{183} and ^{433}WHTLF^{437} motifs have been described to be involved in the N/C interaction [233, 243]. Importantly, the ^{23}FQNLF^{27} motif binds to AF2 in an androgen-dependent manner [243]. Both homo- or heterodimers have been observed for other steroid receptors and nuclear receptors when they are bound to DNA [155]. The existence of an NTD-LBD “heterodimer”

in DNA-bound AR in addition to an NTD-NTD "homodimer" is therefore plausible. Furthermore, the flexibility of the regions of sequence connecting the DBD to the NTD and to the LBD, observed in the crystal structure of the DBD dimer of AR (PDB 1R4I, [210], see Fig. 7.2) potentially allows not only interaction of the two NTD domains, but also both intra- and intermolecular NTD-LBD interactions. Taken together, it is possible that intra- and/or intermolecular N/C interactions occur when full-length AR is bound to DNA, and it is even likely to happen in the presence of hormone.

This would create an equilibrium between monomeric NTD, homodimeric NTD, and both intra- and intermolecular NTD-LBD dimers (see Fig. 7.5a). Given that at least two of the NTD motifs that are involved in N/C interaction (^{23}FQNLF^{27}, ^{179}LKDIL^{183} and ^{433}WHTLF^{437}) are also important for the homodimerization of the NTD (see chapter 4), it is likely there will be competition between NTD-NTD and NTD-LBD interactions. The binding affinity of the ^{23}FQNLF^{27} motif and the LBD of AR was determined to be low micromolar ($1.2 \pm 0.2\,\mu$M by ITC [238], and $9.2 \pm 0.4\,\mu$M by fluorescence polarization [191]). Our data suggest that the affinity of AF1*, and by extension the NTD, to self-interact is in the order of low millimolar, i.e. weaker binding. Consequently, in the presence of hormone N/C interaction is anticipated to be favored over NTD-NTD interaction. Therefore, when the hormone is present, we expect that NTD-LBD "heterodimers" (possibly both intra- and intermolecular) will be highly populated, potentially in equilibrium with monomeric NTD, and that the NTD-NTD "homodimer" will be less populated (see Fig. 7.5a).

However, in the absence of hormone the ^{23}FQNLF^{27} motif does not bind to AF2, which is expected to disfavor N/C interactions considerably. This results in an equilibrium between monomeric NTD and NTD-NTD "homodimer" with only a small population of NTD-LBD "heterodimer", if any (see Fig. 7.5b). In other words, this reflects the interdomain communications described so far in our model of the DNA binding complex of AR when the N/C interaction does not take place.

7.6 Possible role of NTD-NTD "homodimer" in (prevention of) aberrant AR transactivation

Even though it is not clear whether N/C interaction occurs in the physiological DNA binding complex of AR, it is clear that in the absence of hormone or in the absence of the LBD, no N/C interaction can take place. Consequently, we anticipate an equilibrium between monomeric NTD and NTD-NTD "homodimer" in DNA-bound AR for full-length AR in the absence of hormone and for splice variants lacking the LBD. Im-

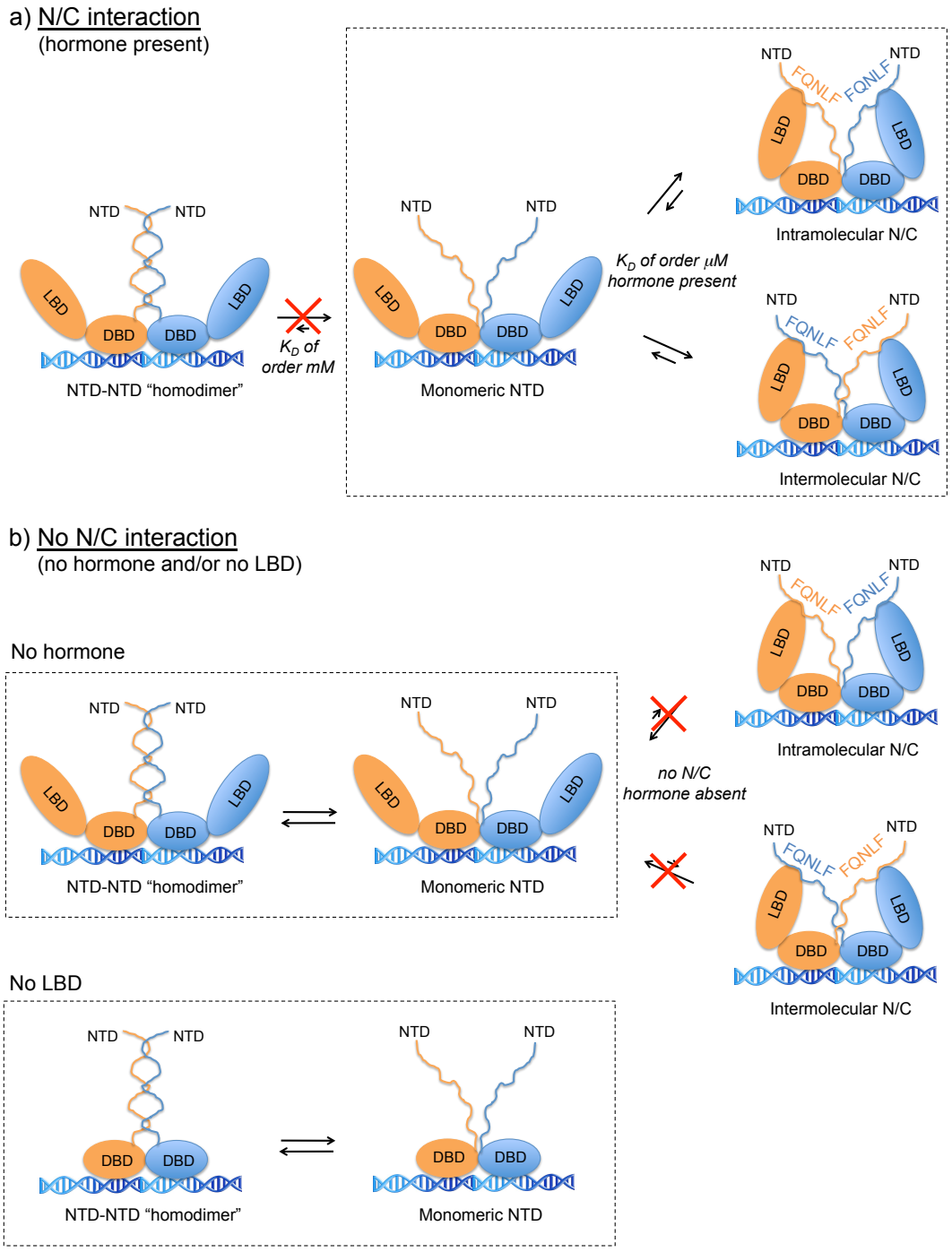


Figure 7.5: a) Schematic representation of potential interdomain interactions of DNA-bound AR when the N/C interaction can take place. b) Schematic representation of potential interdomain interactions of DNA-bound AR when the N/C interaction does not occur.

portantly, AR should not be transcriptionally active in the absence of hormone because physiological transcription is stimulated as a response to androgens. Therefore transcription activation of full-length AR in the absence of hormone or by constitutively active splice variants is considered aberrant. In both cases, the NTD-NTD "homodimer" is present in the DNA binding complex according to our model. Furthermore, the above discussion suggests that this dimer plays a protective role to inhibit AR transactivation when it is not required.

Importantly, this model implies that the interdomain communication in DNA-bound AR depends on the presence or absence of the N/C interaction (see Fig. 7.5). Distinct interdomain communication of AR in the DNA binding complex could result in a different availability of the various NTD regions for interaction with co-regulators and/or the transcription machinery. In turn, this could be related to the dominant role of Tau-1 in androgen-dependent AR transactivation and of Tau-5 in mediating hormone-independent AR transactivation. Moreover, distinct co-regulators might be recruited to the AR transcription complex in both cases, leading to a different protein assembly and altogether changing the mechanisms by which the NTD activates transcription.

Potentially this interdomain communication is distinct in healthy cells (possibly predominantly N/C interaction as hormone is present) and in hormone-independent cancer cells (mainly NTD-NTD "homodimerization"). If the protein assembly on the DNA is indeed different for physiological and aberrant AR transactivation, a better understanding of the components of these complexes and their interactions would allow selective targeting of aberrant transactivation. Our model suggests that the NTD-NTD "homodimer" that would be predominantly formed in the context of aberrant AR transactivation is a transcriptionally inactive conformation. Therefore trapping the NTD in this conformation might lead to blocking aberrant transactivation. Furthermore, the NTD-NTD "homodimer" has, according to our model, globular-like properties and hence might be "druggable" via conventional methods, e.g. with small molecules obtained through rational design.

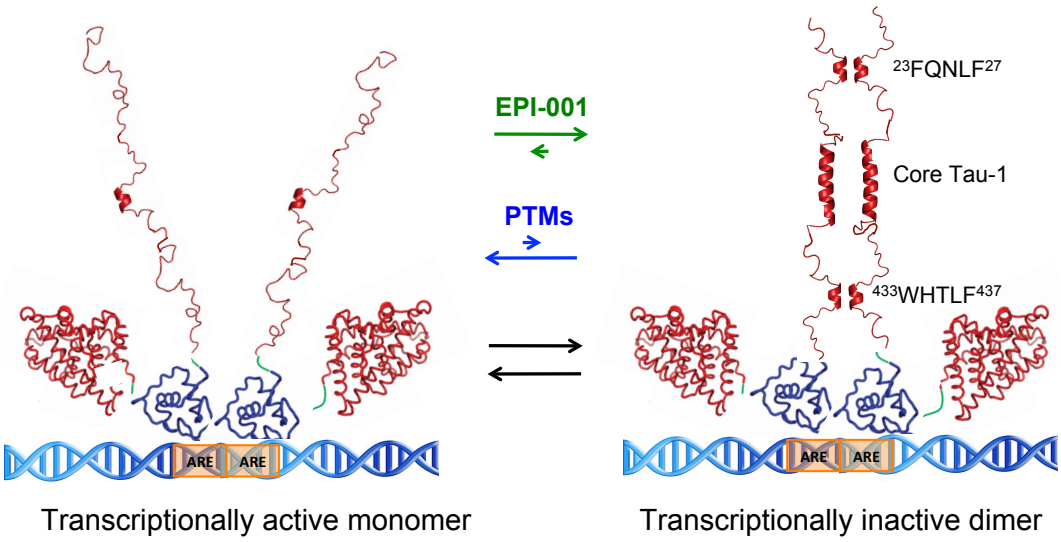


Figure 7.6: Working model of the role of the NTD in AR transcriptional activation based on the obtained data from the three projects. This model is suggested to be representative for full-length AR bound to DNA in the absence of hormone, for splice variants lacking the LBD, and potentially for physiological AR transactivation that does not require N/C interaction.
We propose that in the absence of N/C interaction the NTD domain of DNA-bound AR exists in an equilibrium between a monomeric transcriptionally active state and an auto-inhibitory homodimeric conformation. The dimer is suggested to be stabilized by cooperative helix to helix interactions between regions of sequence with helical propensity in the NTD monomer, including ^{23}FQNLF^{27}, ^{179}LKDIL^{183} (in core Tau-1) and ^{433}WHTLF^{437} (in Tau-5). We further propose that the equilibrium between these two conformations can be shifted by several factors like PTMs and interaction with small molecules. Shifting the equilibrium towards the monomer enhances AR transactivation and stabilization of the dimer inhibits transactivation. Interaction with co-activators or members of the transcription machinery might require (local) destabilization of the dimer conformation to expose key regions for interaction. RAP74 interacts with the monomeric ^{433}WHTLF^{437} motif of AR to activate transcription. Our data suggest that PTMs are a switch to enable this aberrant transactivation to allow the ^{433}WHTLF^{437} motif to interact with RAP74. Binding of small molecules, like EPI-001, may also affect the monomer-dimer equilibrium of the NTD and therefore the capacity of the NTD to activate transcription. We suggest that small molecules that stabilize the dimer might inhibit AR transactivation, similar to EPI-001.
For simplicity, no interaction between the NTD and LBD (N/C interaction) is assumed here. Since it is not clear whether N/C interaction occurs in DNA-bound AR the expected effect of N/C interaction on this model is discussed separately in figure 7.5a. The DBD is indicated as a blue globular domain and the LBD as a red globular domain, the DNA is shown in blue and androgen response elements as orange boxes. Yellow spheres with a - sign indicate phosphorylated residues. Positively charged residues of RAP74 are shown in blue space filling.

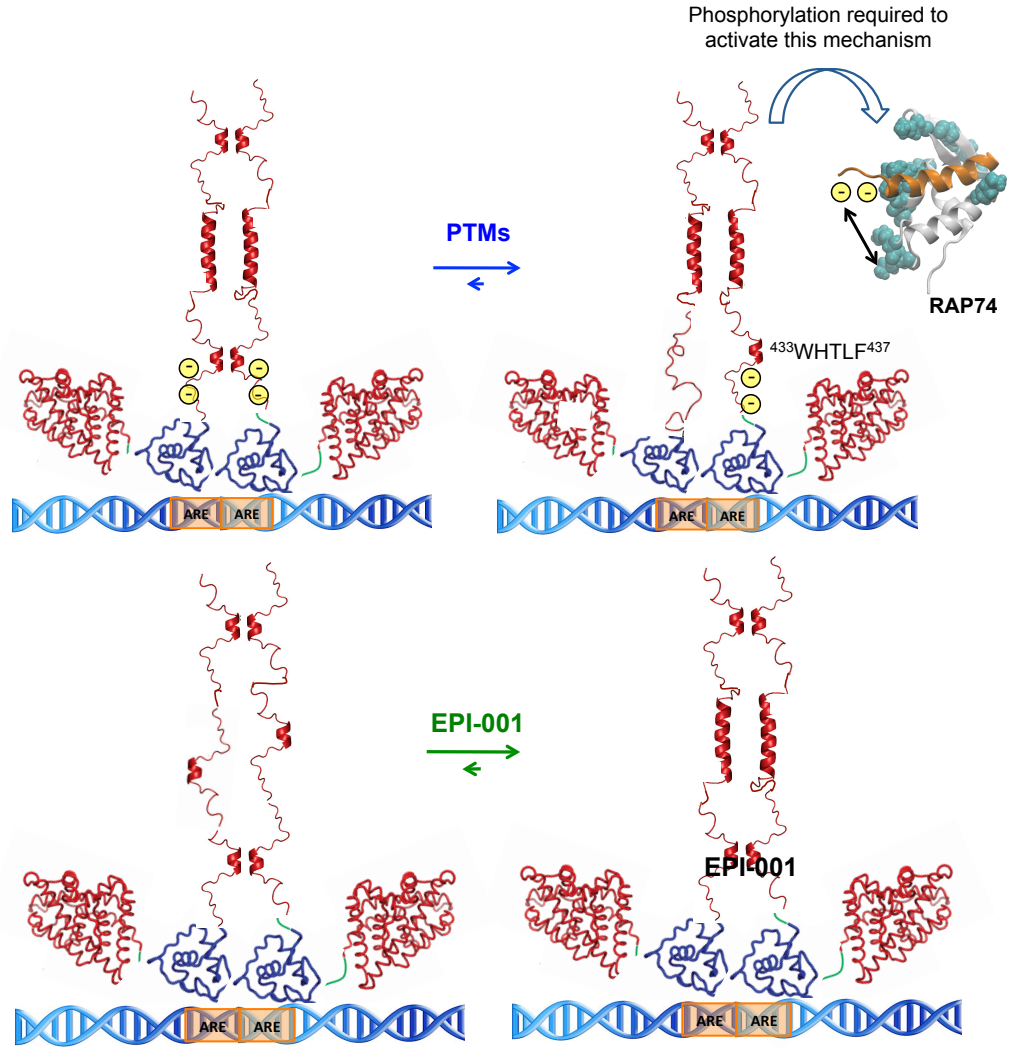


Figure 7.7: Figure 7.6 continued.

8 Conclusions

8.1 Conformational properties of AF1*/NTD

1. We found that, *in vitro*, AF1* has a low tendency to self-associate into a dimer with a head-to-head orientation. The AF1* dimer is mainly stabilized by helix to helix interactions between regions with high helical propensity in the AF1* monomer. Possibly dimerization of the entire NTD occurs in a similar way, *in vivo*, when AR is bound to DNA. This might stabilize the dimeric DNA-bound state of AR and play a role in the regulation of the AR transcriptional activity.

2. AF1* is mainly monomeric under the conditions in which we studied its conformational properties. Under these conditions it displays a dynamic equilibrium between a completely unfolded state and different partially helical conformations. Unusually high helicity was observed in regions known to be important for function (core Tau-1 and Tau-5). This could be in part due to a low populated AF1* dimer with higher secondary structure content compared to the monomer.

3. Regions of AF1* with higher secondary structure propensity and less flexibility are also involved in dimerization of AF1* and coincide with regions previously identified to be functional (core Tau-1 and Tau-5).

4. Tau-1 and Tau-5 are known to be functionally independent units, dominant for the regulation of androgen-dependent and androgen-independent AR transactivation, respectively. We have shown that they are also structurally independent units (no long-range contacts between Tau-1 and Tau-5). Instead, our data suggest a potential equilibrium for both Tau-1 and Tau-5 between a monomeric disordered state and a more globular-like homodimer (head-to-head orientation) in the context of DNA-bound AR. Possibly in such a dimeric state the key motifs

for interaction with co-regulatory proteins and/or the transcription machinery are not available.

8.2 Interaction of AR and RAP74

1. AF1* of AR interacts with the C-terminal domain of RAP74 at the ^{433}WHTLF^{437} motif, a key regulatory motif for ligand-independent AR transactivation in CRPC. This motif is located in the Tau-5 transactivation unit of AR.

2. Interaction of the ^{433}WHTLF^{437} motif of AR with RAP74 enables transactivation in CRPC cells. This is a novel transactivation mechanism of the AR N-terminal transactivation domain.

3. Interaction can be disrupted *in vitro* by mutation of key binding residues (W433, L436, F437, E440, E441). In addition, mutation of these residues in cellular assays leads to lower transcriptional activity.

4. Post-translational modifications N-terminal to the binding motif might be required to promote the interaction and consequently to trigger this transactivation mechanism. This supports the notion that interaction of AR with RAP74 is not a general mechanism but instead is one possible mechanism to enable aberrant AR transactivation in CRPC at castrate levels of hormone.

5. Phosphorylation of S430, S431 or S432 increases the binding affinity and transcriptional activity but does not create a new binding site on the RAP74 surface.

8.3 Interaction of AR and EPI-001

1. EPI-001 interacts specifically and reversibly with the C-terminal region of AF1*, which is located in the Tau-5 region. The compound specifically recognizes a particular low populated conformation of the Tau-5 region, possibly the one Tau-5 adopts in the AF1* dimer.

2. *In vitro*, irreversible binding of EPI-001 to AF1* occurs first at cysteine 404, a nucleophile located in the region of sequence that EPI-001 specifically recognizes in the reversible step of the interaction.

3. Interaction of EPI-001 with AR might take place in a similar manner *in vivo*. Our data suggest that EPI-001 interaction might stabilize a dimeric and tran-

scriptionally inactive conformation of the NTD in DNA-bound AR. This could contribute to the mode of action of this small molecule.

9 Summary (Spanish)

9.1 Introducción

9.1.1 Cáncer de próstata y el receptor androgénico

El cáncer de próstata (PCa) es el segundo tipo de cáncer más común en hombres después del cáncer de pulmón.[1] Alrededor de 1.1 millones de hombres en todo el mundo se les diagnosticó PCa durante el año 2012. El crecimiento del cáncer de próstata depende de la relación de células proliferantes y de células en proceso de morirse. Los andrógenos, las hormonas sexuales masculinas, son los principales reguladores de esta relación y estimulan la proliferación e inhibición de la apoptosis celular. Por ello, el PCa depende esencialmente de la estimulación de los andrógenos para el crecimiento y la supervivencia celular [1]. El receptor de andrógeno (AR) es un receptor de hormonas nuclear que es activado por las hormonas andrógenas y la proteína mediante la cual los efectos fisiológicos de los andrógenos se producen [2]. El AR es necesario para el desarrollo normal de la próstata, su crecimiento y fisiología pero también tiene un papel muy importante en la progresión del PCa.

Como todos los miembros de la superfamilia de receptores nucleares, el AR es un factor de transcripción activado por la unión de sus ligandos naturales, la testosterona o la dihidro-testosterona (DHT) que tiene un efecto aún más potente que la testosterona. El AR está compuesto por cuatro dominios: el dominio N-terminal (NTD) que se piensa que es intrínsicamente desordenado (residuos 1–559), dominio de unión a DNA (DBD) que contiene dos dedos de zinc (560–622), la región bisagra (623–670) y el dominio de unión de ligando (LBD) que une tanto testosterona como DHT (residuos 671–919) (ver figura 1.2 en pág. 2).

[1] Fuente: http://globocan.iarc.fr

En ausencia de hormonas andrógenas, el AR está presente en el citoplasma acomplejado con chaperonas moleculares. La unión de los andrógenos al LBD produce un cambio conformacional que genera la disociación del complejo produciendo la formación del sitio de unión destinado a la unión de co-reguladores en el LBD del AR y permite la dimerización del AR. También expone la señal de localización nuclear, permitiendo la translocación del AR al núcleo. En el núcleo y ya como dímero, el AR reconoce y se une a los secuencias específicas del DNA (AREs, elementos de respuesta de andrógenos) a través del DBD regulando la transcripción de numerosos genes involucrados en el desarrollo de la reproducción de los hombres (ver fig. 1.3 en pág. 4). Este proceso recibe el nombre del eje de señalización andrógeno.

El tratamiento del PCa depende del estadio de la enfermedad. Si el PCa está localizado y todavía se encuentra en una etapa temprana puede ser tratado mediante operación (prostatectomía, extracción quirúrgica parcial o total de la glándula de la próstata) y/o radioterapia. La terapia básica para la enfermedad avanzada o metastásica es una combinación de terapias por privación de andrógenos (ADT), por vía quirúrgica o castración química, y la administración de antagonistas del AR. La ADT tiene como objetivo la reducción de los niveles de testosterona en el suero para bloquear el eje de señalización de andrógenos, sin embargo los antagonistas del AR son anti-andrógenos que compiten con él ligando natural del AR y se unen preferentemente a la proteína sin activarla. La combinación de ambas terapias recibe el nombre de ablación de andrógeno máxima. Aunque la ADT suele ser satisfactoria al inicio y causa una regresión de los tumores del PCa, la progresión de la enfermedad es inevitable hasta un punto en el cual se hace resistente a la terapia de ablación andrógena, que comúnmente recibe el nombre del cáncer de próstata resistente a la castración (CRPC). Actualmente no existe un tratamiento efectivo para CRPC y, aunque la esperanza de vida de los pacientes se alarga algunos meses, desgraciadamente no consiguen curarse nunca completamente. La media de supervivencia de los pacientes después del desarrollo de resistencia es de unos 12 meses [16].

Durante la progresión de la enfermedad, las células tumorales se adaptan a las condiciones de privación de andrógenos y a la presencia de antagonistas del AR mediante diversos mecanismos que llevan a una hipersensibilización del AR a niveles muy bajos de testosterona. Estos mecanismos permiten la activación general del AR mediante la unión de hormonas alternativas o incluso antagonistas, permiten la activación del AR de manera independiente de su ligando o incluso evitando la activación del eje de señalización de andrógenos mediante la activación de rutas de señalización celular alternativas esenciales para la supervivencia celular. La combinación de estas estrategias mantiene la actividad del AR y activa la transcripción de los genes dependientes del

AR a pesar de la terapia de ablación de andrógeno máxima. En conclusión, el AR es un regulador esencial para el crecimiento, diseminación y supervivencia de células cancerígenas en tumores resistentes a la castración y por tanto, representa la diana terapéutica más prometedora [13, 145].

En los últimos años cinco nuevos fármacos han sido aprobados para el tratamiento de pacientes con CRPC, dos de ellos actúan directamente sobre el eje de señalización andrógeno: el acetato de abiraterona y la enzalutamida. El acetato de abiraterona es un inhibidor selectivo e irreversible de una encima clave para la síntesis de andrógenos en las glándulas adrenales que reduce además los niveles de andrógenos circulantes en la sangre. Este hecho resulta importante dado que la ADT tiene como diana la producción de testosterona en los testículos. La enzalutamida es un antagonista de AR perteneciente a la siguiente generación de fármacos que ha sido diseñado para evitar la conversión antagonista-agonista que tienen los antagonistas del AR de primera generación. Se une fuertemente al LBD del AR y inhibe la señalización de andrógeno posterior. A pesar de su superioridad sobre los fármacos de primera generación, ya han aparecido tumores de PCa resistentes tanto al acetato de abiraterona como a la enzalutamida a través de varios mecanismos, incluyendo la sobre-expresión de variantes de empalme del AR sin LBD. Estos dos fármacos aprobados recientemente alargan tan sólo un par de meses la supervivencia de los pacientes sin evitar el temido desenlace. En conclusión, a pesar de los avances realizados en los últimos años, aún no es posible curar la enfermedad una vez llegada la etapa de CRPC.

Estos nuevos fármacos están basados en las mismas estrategias de los tratamientos de primera generación como la inhibición de la síntesis de andrógenos y la inhibición del AR por unión competitiva de los antagonistas del AR. Ya que los mecanismos de resistencia emergentes en parte se parecen a aquellos que fueron adquiridos por los fármacos de primera generación también estos nuevos fármacos fallan a la hora de curar la enfermedad. Este hecho subraya la necesidad de adquirir un mejor conocimiento de los mecanismos que llevan a la activación del AR aberrante en pacientes CRPC para el desarrollo de nuevas estrategias terapéuticas.

El NTD de AR juega un papel crucial en la actividad constitutiva del AR en tumores de pacientes CRPC. La transactivación del NTD es independiente de los andrógenos y por tanto evita las estrategias actuales de tratamiento dirigidos al eje de señalización de andrógenos (por inhibición de la síntesis de andrógenos o por unión competitiva de antagonistas al AR previniendo la unión del ligando natural). Para activar la transcripción de sus genes, el AR a través del NTD necesita el contacto de la maquinaria de transcripción general bien por una interacción directa o bien vía indirecta de co-reguladores. La prevención de estas interacciones proteína-proteína, esenciales para

la transcripción representa una oportunidad para bloquear el eje de señalización de andrógenos.

Recientemente se han identificado mediante ensayos de cribado de alto rendimiento tres compuestos que interaccionan específicamente con el NTD de AR inhibiendo la actividad transcripcional de este mismo en células de PCa: sintokamida A, EPI-001 y nifatenona B (ver Fig. 1.8 en pág. 21). Actualmente su mecanismo de acción es desconocido. Las propiedades conformacionales del NTD han obstaculizado el diseño racional de pequeñas moléculas con el objetivo de interaccionar con este dominio o con las interacciones proteína-proteína que median la activación de la transcripción. A diferencia del DBD y el LBD, cuyas estructuras están bien caracterizadas, la predicción de la estructura del NTD sugiere una estructura intrínsecamente desordenada confirmada por los datos experimentales [281, 283]. El desorden intrínseco en proteínas está caracterizado por la ausencia de estructura tanto secundaria como terciaria bien definida a pesar de la función biológica de dichas regiones. En los últimos años se ha demostrado que la presencia de estas regiones tienen un papel crucial en varios procesos celulares incluidos la transcripción [151, 152].

Una visión molecular más detallada del papel del NTD en la transactivación de AR podría llevar potencialmente a la revelación del mecanismo de acción de los pequeños compuestos que tienen como objetivo el NTD. Esto podría permitir el diseño racional de compuestos más potentes en el futuro.

9.1.2 Propiedades estructurales y funcionales del dominio N-terminal del AR

Tanto la naturaleza intrínsecamente desordenada del NTD como su tamaño (559 aminoácidos) han impedido la caracterización estructural mediante técnicas de cristalografía de rayos X o mediante espectroscopía de resonancia magnética nuclear (NMR). Mediante experimentos de biología molecular se han identificado varias regiones en el NTD como cruciales para la función de transactivación mediante interacciones proteína-proteína con proteínas reguladoras. El NTD acomoda la función de activación 1 (AF1), función de activación más potente del AR, vagamente definida entre los residuos 141 al 494 [149, 199]. AF1 contiene dos unidades de transactivación, Tau-1 (residuos 102–371) y Tau-5 (residuos 361–537) [149, 232]. Tau-1 es esencial para la transactivación del AR dependiente de andrógeno mientras que Tau-5 regula la transactivación de AR independiente del andrógeno [149, 232]. El dominio central que media la actividad de Tau-1 está localizado en la parte central de Tau-1 (residuos 174–204), y contiene un motivo lineal ^{179}LKDIL^{183} que se superpone con el motivo ^{183}LSEASTMQLL^{192}

(Lx_7LL) [157]. De modo similar, se sabe que el motivo ^{433}WHTLF^{437} es la secuencia central que regula la actividad de Tau-5, responsable de aproximadamente el 50% de la actividad transcripcional del AR independiente de andrógeno en células CRPC [232]. Un tercer motivo lineal, ^{23}FQNLF^{27}, puede unirse a AF2 en el LBD del AR de manera dependiente del ligando para formar la interacción cabeza-cola que, dependiendo del contexto, es necesaria para la actividad transcripcional completa del AR [244, 273].

Está demostrado mediante dicroísmo circular (CD) y espectroscopía de infrarrojos basada en la transformada de Fourier (FTIR) que AF1 del AR contiene entre un 13–16% de estructura secundaria helicoidal [281, 282]. Sin embargo, adopta una conformación más helicoidal en presencia del osmolito natural N-óxido de trimetilamina (TMAO), o un disolvente hidrófobo como el trifluoroetanol (TFE) [281]. Además se ha observado un incremento de estructura helicoidal en el NTD por unión de DNA en el DBD [259] y mediante interacción del NTD del AR con RAP74, una subunidad de un factor de transcripción general [281, 282]. Además el incremento del contenido helicoidal facilita la posterior interacción con el co-activador SRC-1 [282].

Los modelos actuales para el NTD y en particular el AF1 del AR, sugieren que existe un abanico de conformaciones con una estructura limitada a nivel secundario y terciario que son necesarias para la unión de las proteínas co-reguladoras (ver fig. 1.17 en pág. 43). La interacción con otras proteínas puede inducir una conformación más helicoidal o estabilizar una conformación en particular de todas las posibles que facilitaría además la interacción con otras proteínas en la formación del complejo transcripcional del AR.

9.1.3 Interacción del NTD del AR con la máquina de transcripción general

Como ocurre con otros factores de transcripción la activación de la transcripción por el AR depende del reclutamiento de la polimerasa de RNA II (pol II) y de los factores de transcripción generales (GTFs: TFIIA, TFIIB, TFIID, TFIIE, TFIIF y TFIIH) al promotor y/o a los sitios potenciadores de los correspondientes genes. Se sabe que el AR interacciona directamente con la máquina de transcripción general y también mediante interacciones indirectas a través de los co-reguladores. Son bien conocidas las interacciones directas entre el AR y TFIIF [319], TFIIH [321], el factor positivo de transcripción de elongación b (pTEF-b) [324] y la segunda más grande subunidad de pol II (RBP2) [329]. TFIIF es un factor de transcripción general que se asocia muy fuertemente con pol II para facilitar su unión al promotor donde se forma el complejo de pre-iniciación (PIC). Contiene dos subunidades, las proteínas asociadas a la RNA polimerasa II 30 (RAP30) y 74 (RAP74) [318]. La región de interacción entre un

constructo del NTD de AR que contiene los residuos 143–494 y TFIIF se localiza en el dominio C-terminal de RAP74, es decir, RAP74-CTD (residuos 363–517). Aunque el sitio de unión de ambas proteínas es desconocido, los residuos M245, L247 y V249 (y los residuos de alrededor) de AR formarían parte del sitio de unión, mientras que los residuos S160 y S163, que están en la secuencia ^{159}P<u>S</u>TL<u>S</u>L^{164}, tendrían un efecto indirecto sobre la unión al alterar la flexibilidad estructural del NTD de AR [320, 353]. Además, se sabe que los residuos hidrófobos V490, L493 y L497 en la hélice 3 del motivo hélice-giro-hélice de RAP74-CTD rompen la unión selectiva de la región de AR contenida entre los residuos 143–494 [352].

9.2 Objetivos

El NTD del AR es una diana terapéutica prometedora para el CRPC. Sin embargo, la naturaleza intrínsecamente desordenada de este dominio ha imposibilitado el diseño racional de fármacos que tienen como diana este dominio. Los principales objetivos de este doctorado fueron por tanto, primero, la caracterización de las propiedades conformacionales del NTD intrínsecamente desordenado del AR y en particular de la región de AF1, y segundo, el estudio de las interacciones de este dominio a resolución atómica con algunos de los componentes de la transcripción y con potenciales fármacos.

1. Obtener una mejor visión de los mecanismos moleculares por los cuales este dominio intrínsecamente desordenado posibilita la función del AR y produce la enfermedad. Hemos realizado la caracterización por resonancia magnética nuclear de las propiedades conformacionales de este dominio midiendo los parámetros de NMR que permiten la asignación del esqueleto proteico de AF1 que permiten describir sus características dinámicas y de estructura secundaria. Esta información ha sido completada con otras técnicas biofísicas como dicroísmo circular.

2. Investigar el papel del NTD del AR en la activación transcripcional a nivel molecular. Hemos estudiado su interacción con RAP74, una subunidad del factor de transcripción general IIF (TFIIF) y se ha descrito que interacciona con el dominio NTD del AR [319, 320]. Hemos realizado experimentos de perturbación de los desplazamientos químicos que no sólo muestran el sitio de unión exacto sino que también producen información adicional como la afinidad de la unión y los cambios estructurales debidos a la interacción.

3. Explicar cómo el NTD es la diana de pequeñas moléculas. Para ello hemos probado a estudiar el mecanismo de interacción de EPI-001, una molécula pequeña que recientemente se ha descrito como un potente inhibidor de AR que interac-

ciona con el dominio NTD de AR y produce la regresión del proceso cancerígeno en células de CRPC [134].

9.3 Resultados y discusión

9.3.1 Propiedades conformacionales del NTD de AR

Para estudiar las propiedades conformacionales del dominio de transactivación N-terminal del AR y especialmente de AF1, hemos diseñado un constructo (AF1*) que permita el estudio por NMR de la región de AR comprendida entre los residuos 142–448 (308 residuos). AF1* incluye las unidades Tau-1 y Tau-5 excluyendo la región de poli-glicinas (residuos 449-472) y la región de poli-glutaminas (residuos 58–78). La asignación del esqueleto de AF1* se consiguió mediante una estrategia *divide y vencerás* basado en la asignación de tres secuencias más pequeñas con una superposición de diez residuos: AF1*a (AR 142–275), AF1*b (AR 265–340) y AF1*c (AR 330–448). Esta estrategia fue elegida para evitar la superposición de picos en los espectros de NMR de AF1* debido al tamaño del constructo (559 aminoácidos) y la baja dispersión de los desplazamientos químicos del ^{1}H característicos de proteínas desordenadas.

Los desplazamientos químicos y el ensanchamiento de la resonancia en los espectros [^{1}H,^{15}N]-HSQC de AF1* medidos a diferentes concentraciones indican que AF1* tiene baja tendencia a oligomerizar. Sin embargo, no fuimos capaces de detectar oligomerización mediante ningún otro tipo de técnica, excepto con un gel nativo. Además, los desplazamientos químicos observados por NMR resultaron pequeños para un amplio rango de concentraciones (10 μM–680 μM). De esta manera, nuestros datos sugieren que AF1* tiene una baja tendencia a oligomerizar y la especie dimérica es la única especie de oligomerización observada en concentraciones menores a 680 μM. Además nuestros datos sugieren que las regiones con una tendencia helicoidal en el monómero están estabilizadas por interacciones hélice-hélice y que la dimerización se produce en orientación cabeza-cabeza.

Nuestros resultados por tanto indican que bajo las condiciones experimentales utilizadas para determinar las propiedades conformacionales de AF1* (concentración de proteína, disolución amortiguadora, temperatura, pH) la proteína está principalmente como monómero y solamente una pequeña parte se encuentra como dímero. La asignación del esqueleto de AF1* mostró que existe un equilibrio dinámico entre una conformación completamente extendida y conformaciones parcialmente helicoidales. Las regiones con mayor tendencia helicoidal están localizadas entre los residuos 176–202, residuos 204–212, residuos 229–239, residuos 353–364, residuos 391–413 y residuos 427–

436 (ver fig. 4.27 en pág. 145). Además se observaron también dos regiones con valores ΔC_α negativos entre los residuos 142–154 y los residuos 268–286, regiones ricas en prolinas que sugieren que estas regiones poseen cierta tendencia a adoptar preferentemente una conformación extendida en lugar de una lámina beta, como la poli-prolina II. La tendencia helicoidal asociada a los residuos 176–202 y los residuos 391–413 es considerablemente más alta que la esperada para proteínas intrínsecamente desordenadas (ver fig. 4.26 en pág. 143). Resulta notorio que los residuos 176–202 corresponden a la región central de Tau-1 (residuos 174–204) y residuos 391–413 están localizados en la región Tau-5. Para determinar si una mayor tendencia helicoidal está relacionada con interacciones de largo alcance entre estas regiones, se compararon los cambios ΔC_α para AF1*c y AF1* (ver figs. 4.34 y 4.35 en pág. 152). Estos resultados demostraron que la tendencia helicoidal asociada a los residuos 391–413 es indistinguible en el contexto de AF1*c y en el contexto de un constructo más largo, AF1*, indicando que la helicidad en esta región no es estabilizada por contactos de largo alcance con regiones ausentes en AF1*c como los residuos 176–202. Sin embargo, las interacciones hélice-hélice de las regiones con tendencia helicoidal en el constructo AF1*c podrían contribuir a la mayor tendencia helicoidal observada en este constructo y a una conformación más colapsada de AF1*c en comparación con el resto de AF1*, tal y como indican los experimentos de cromatografía de exclusión por tamaño (SEC) (ver fig. 4.37 en pág. 154). Además, la mayor tendencia helicoidal de los residuos 176–202 podría estar relacionada con la dimerización. Siendo que los residuos 171–215 deberían tener *a priori* mayor tendencia a formar estructura del tipo hélice superenrollada ("coiled-coil") se midieron los cambios de C_α de AF1* a 25 μM y 390 μM (ver fig. 4.38 en pág. 156). El resultado observado indica que la tendencia helicoidal de los residuos 176–202 es indistinguible a cualquier concentración, indicando que dicha tendencia es intrínseca a esta región y que no es originada por la interacción entre hélices alfa debido a la dimerización de AF1*. Además de las interacciones estabilizantes de la hélice en cada uno de los "segmentos" del AF1* (AF1*a, AF1*b y/o AF1*c), las interacciones de largo alcance entre hélices entre dos monómeros de AF1*, debidas a la dimerización, podrían contribuir en última instancia a la mayor tendencia helicoidal observada de estas secuencias.

Es importante señalar que las regiones identificadas con una mayor tendencia helicoidal, que muestran además velocidades de relajación transversales más altas, están involucradas en la dimerización y coinciden con regiones que son cruciales para la transactivación del AR: las regiones centrales de Tau-1 y Tau-5.

Los resultados sugieren que la mayor tendencia a formar hélices de la región central de Tau-1 es intrínseca al constructo AF1*. Esta región parece que puebla una conformación helicoidal durante una fracción considerable de tiempo de una manera

dinámica (formación y ruptura rápida de la hélice) y está caracterizada por un régimen de intercambio rápido. Aunque no se observaron incrementos en helicidad importantes en los rangos de concentraciones de estudio en la región central de Tau-1 debido a la dimerización, los cambios químicos de ^{1}H y ^{15}N al concentrar la proteína indican que esta región está involucrada en la dimerización. En el dímero AF1*, la región central de Tau-1 de ambos monómeros de AR interaccionan muy probablemente mediante interacciones del tipo hélice superenrollada ("coiled-coil"). La presencia de una población muy pequeña de la especie dimérica de AF1* con una interacción más estable podría contribuir a la mayor helicidad observada en esta región.

Las tres regiones helicoidales de la región Tau-5 de AF1* podrían estabilizarse entre ellas a través de contactos de largo alcance transitorios que formen una conformación relativamente colapsada (ver fig. 4.37). Además nuestros datos indican que la dimerización tiene lugar primariamente a través de la región Tau-5 y que posiblemente se produce mediante una reorganización de la conformación de esta región desde una conformación relativamente colapsada, estabilizada por interacciones hélice-hélice entre ciertos motivos helicoidales del monómero AF1*c, a una interacción entre estas hélices en una orientación cabeza-cabeza del dímero. El pronunciado ensanchamiento de las líneas y las velocidades de relajación más altas en esta parte de la secuencia reflejan más que probablemente el intercambio conformacional presente, debido al equilibrio entre conformaciones más extendidas y más colapsadas del monómero, pero también existe intercambio químico adicional debido a la reorganización conformación de la región Tau-5 tras la dimerización. Además las interacciones en el dímero entre las regiones con una tendencia helicoidal podrían contribuir a la mayor helicidad observada en la región Tau-5. AF1*c se caracteriza por un régimen de intercambio intermedio.

Aunque no se puede excluir una contribución del dímero AF1* en las regiones de más tendencia helicoidal de AF1*, el estado dimérico de AF1* está muy poco poblado en el rango de concentraciones estudiado y la helicidad de AF1* es prácticamente la misma tanto a 25 μM como a 390 μM. Por tanto, los resultados muestran que la especie monomérica de AF1* muestra una considerable, aunque transitoria, helicidad en varios tramos de su secuencia. Probablemente, la helicidad residual en estas regiones es mayor en el estado dimérico.

En conclusión, se ha estudiado por primera vez las propiedades conformacionales de (o más bien una parte del) dominio de transactivación intrínsecamente desordenado de un receptor nuclear a resolución atómica. La región del AR estudiada (142–448 = AF1*) incluye el AF1 original que es necesario y suficiente para la completa actividad de AR y la actividad intrínseca de NTD de AR [157].

La baja tendencia del constructo de AF1* a formar el estado dimérico en la orientación cabeza-cabeza en disolución sugiere una posible interacción cabeza-cabeza del NTD en el dímero de AR cuando esté unido al DNA. Si dicho dímero existiera *in vivo*, podría contribuir a la estabilidad del homodímero de AR al unirse a DNA y regular la transcripción.

9.3.2 Interacción de AR y RAP74

RAP74 es la subunidad grande del factor de transcripción general TFIIF. Experimentos bioquímicos de unión ya demostraron la interacción de RAP74 con la región AF1 de AR [319, 320]. En este doctorado hemos estudiado la interacción entre AF1* (residuos 142–448) y el dominio C-terminal de RAP74 (residuos 450–517) mediante NMR.

Los experimentos de perturbación de desplazamientos químicos revelaron que la interacción entre AR y RAP74 tiene lugar en el motivo ^{433}WHTLF^{437} de la región Tau-5, una región esencial para la transactivación de AR en ausencia de hormona. Tras la unión de RAP74 al motivo de unión de AR, este último adquiere mayor helicidad mientras que la interacción no produce cambios observables en RAP74. El AR se une a RAP74 en un surco hidrófobo formado por las hélices 2 y 3 de RAP74 que forman una estructura hélice-giro-hélice.

La afinidad de la unión calculada mediante titraciones con NMR resultó ser del rango millimolar y por tanto resulta imposible producir una estructura experimental del complejo. Sin embargo, basándonos en la similitud estructural del motivo de unión de AR y los dos motivos resueltos de FCP1 que interaccionan con el mismo dominio C-terminal de RAP74, se realizó un modelo estructural. FCP1 (fosfatasa de dominio C-terminal asociada a TFIIF) es una fosfatasa específica que defosforila la polimerasa de RNA II al final de la transcripción. El modo de interacción de los dos motivos FCP1 con RAP74 es el mismo y requiere que FCP1 presente una hélice anfipática que se acomoda en la superficie hidrófoba formada por las hélices 2 y 3 de RAP74. La interacción es estabilizada gracias a dos contactos electrostáticos entre los residuos negativamente cargados de FCP1 y los positivamente cargados en la superficie de RAP74. En el caso del motivo de interacción de AR, están presentes tanto residuos hidrofóbicos como residuos negativamente cargados que hacen pensar que el modo de unión de AR y RAP74 podría ser similar a los motivos de FCP1.

Con el objetivo de confirmar este modelo para la interacción de AR y RAP74, se estudió el efecto de mutaciones que desestabilizaron el complejo. Tanto la mutación simultánea de los residuos hidrófobos W433, L436 y F437 por alaninas y la mutación doble de carga opuesta E440K/E441K rompieron la interacción de AR y RAP74 *in*

vitro, demostrando la importancia de estos residuos en la interacción. En base a estos resultados, proponemos que la unión de AR y RAP74 implica cadenas aromáticas (W433, L437 y F437) por parte de AR y la superficie hidrófoba de RAP74 formada por las hélices 2 y 3 así como interacciones electrostáticas entre los residuos ácidos (E440, E441) de AR y los residuos cargados positivamente de RAP74. Además nuestros resultados sugieren la formación de un complejo transitorio con cierta heterogeneidad estructural, en el que el motivo de unión de AR adopta una conformación más helicoidal, pero sin llegar a formar una hélice suficientemente estable y definida.

Para determinar si la interacción entre AR y RAP74 activa la transcripción en células de cáncer humanas, se realizaron experimentos con luciferasa como gen reportador en colaboración con el Prof. Iain McEwan (Universidad de Aberdeen, Reino Unido). Los resultados mostraron que la interacción entre AR y RAP74 activa la transcripción en células PC-3 de pacientes CRPC a niveles hormonales de castración y que la misma mutación doble de carga opuesta (E440K/E441K) realizada en el AR entero causaba una reducción importante de la activación de la transcripción. A pesar de que las mutaciones de los residuos hidrófobos (W433A/L436A/F437A) no tuvieron el efecto esperado en la transactivación, la importancia de estos residuos para la actividad transcripcional ha sido demostrada por los experimentos realizados por Dehm *et al.* en los que la misma mutación hidrófoba en el AR entero elimina la transactivación mediada por Tau-5 en células C4-2 de pacientes CRPC en ausencia de hormona [232].

Una comparación más profunda de los motivos de unión de AR y FCP1 demuestra la existencia de un gran número de residuos cargados negativamente presentes en el lado N-terminal de los residuos hidrófobos en ambos motivos de interacción de FCP1 que, en contraposición, no se encuentran en AR. Sin embargo, el AR contiene varias serinas en la región análoga. Estos cambios en la secuencia de AR podrían indicar que la presencia de modificaciones post-transduccionales como fosforilaciones de una o varias de estas serinas en la parte N-terminal del motivo ^{433}WHTLF^{437} aumentarían la afinidad de la interacción entre AR y RAP74. Aunque ninguna fosforilación en dichas serinas (S430, S431 y S432) ha sido documentada previamente, hemos descubierto que, *in vitro*, la fosforilación de cualquiera de éstas incrementa la afinidad de la unión entre AR y RAP74 en comparación con la forma AR no modificada. Es importante decir que una triple mutación de serina a glutamato (S430/S431/S432) mostró tan sólo un pequeño incremento de la afinidad comparada con cualquiera de las monofosforilaciones. Sin embargo, la fosforilación simultánea de las serinas S430 y S432 mostró un efecto sinérgico que incrementó la afinidad de la unión hasta $81 \pm 4\,\mu$M. Este resultado sería compatible con una interacción débil y temporal entre el AR y RAP74 en la célula teniendo en cuenta que ambas proteínas están fuertemente asociadas cuando

el AR está unido al DNA.

Para estudiar el efecto de la fosforilación en AR entero en células PC-3 se utilizó un gen reportador con mutaciones simultáneas de las serinas S430, S431 y S432 a glutamato (fosfomimético) o a alanina (resistente a fosforilación). Los resultados mostraron que la triple mutación de las serinas a alaninas redujo notablemente la transactivación en comparación con la forma nativa, mientras que la triple mutación de las serinas a glutamatos incrementó en gran medida la transactivación. Estos resultados sugieren que la fosforilación de los residuos de la parte N-terminal del motivo ^{433}WHTLF^{437} podría favorecer la interacción entre AR y RAP74 *in vivo*.

Nuestros datos indican que la interacción entre AR y RAP74 activa la transactivación de AR de manera incorrecta en células CRPC y a niveles de hormona de castración. Además sugieren también que la fosforilación de varios residuos de la parte N-terminal del motivo ^{433}WHTLF^{437} podría ser necesaria para que la interacción entre RAP74 y AR se produjera con una afinidad biológicamente relevante. De ser cierta esta hipótesis, se demostraría que la interacción entre RAP74 y AR no es un mecanismo general de la transactivación de AR sino un mecanismo que puede ser activado en condiciones específicas para aumentar la transactivación de AR. La interacción entre AR y RAP74 tiene lugar posiblemente en células CRPC a niveles hormonales bajos cuando ciertas quinasas específicas fosforilan uno o varios de los residuos N-terminales del motivo ^{433}WHTLF^{437} para facilitar la interacción.

El patrón de fosforilación de AR cuando se une a RAP74 en condiciones biológicas es aún desconocido. Aunque es probable que no sea tal y como sugerimos en nuestros experimentos *in vitro*, no podemos excluir la posibilidad de que la heterogeneidad estructural del complejo y/o la conformación helicoidal no bien definida del motivo de unión de AR en el complejo con RAP74, sean debidas en parte por la ausencia del patrón de fosforilación adecuado.

9.3.3 Interacción de AR y EPI-001

Con el objetivo de elucidar el mecanismo de acción de EPI-001 y, en particular, la base molecular de la especificidad de EPI-001 por el NTD de AR, hemos estudiado la interacción no-covalente de AF1* y EPI-001 mediante NMR así como la unión covalente posterior mediante espectrometría de masas. EPI-001 se presenta como una prometedora clase de inhibidores de AR que se une a AR a través de NTD.

Inicialmente medimos el espectro de [^{1}H,^{15}N]-HSQC de AF1* en presencia de 10 equivalentes molares de EPI-001 para determinar el sitio de unión de EPI-001 en AF1*.

En contra de lo esperado, la interacción con EPI-001 involucra muchos residuos de AF1* (residuos 345–448) en lugar de restringirse a una región bien definida por un número menor de residuos. Los aminoácidos que interaccionan con EPI-001 están localizados en las regiones contiguas de la secuencia de la región Tau-5 que corresponden a regiones con una tendencia helicoidal relativamente alta y que están involucrados en la dimerización. Las perturbaciones de los desplazamientos químicos causados por la unión de EPI-001 entre estas regiones pueden ser debidas a: i) la existencia de tres sitios de interacción independientes, ii) la existencia de un único sitio de interacción que contenga residuos lejos en la secuencia pero cercanos en el espacio (sitio de unión tripartito), o iii) la interacción de AF1* con EPI-001 causa una redistribución del equilibrio de conformaciones de AF1*.

Para determinar si EPI-001 interacciona independientemente con las regiones experimentando perturbaciones de los desplazamientos químicos en AF1*, seguimos el efecto en el espectro de NMR ^{1}H de EPI-001 de tres péptidos con secuencias correspondientes a estas tres regiones. Ninguno de los péptidos interaccionó con EPI-001, indicando que la cooperatividad entre los residuos en la región Tau-5 es necesaria para la interacción. Las tres regiones de la secuencia de AF1* presentaban cambios en los desplazamientos químicos debidos a la interacción con EPI-001 que podrían indicar la formación de un sitio de unión tripartito. Sin embargo, es improbable que EPI-001 pueda contactar simultáneamente con un número tan grande de residuos (alrededor de 100 residuos) para llevar a cabo pequeñas perturbaciones de los desplazamientos químicos. Además es también improbable que la región Tau-5 adopte una conformación colapsada en la que las tres regiones con una tendencia helicoidal formen un haz de tres hélices, debido a la presencia de una región de poli-prolinas, motivos considerados relativamente rígidas [423], entre la primera y la segunda región helicoidal. Por consiguiente, nuestros datos sugieren que la interacción de AF1* con EPI-001 causa una redistribución del equilibrio de conformaciones de AF1*. Los cambios en los desplazamientos químicos observados en AF1* debidos a la unión de EPI-001 son pequeños (orden de magnitud de ppb). Si añadimos el hecho de que muchos aminoácidos están afectados por la interacción con EPI-001, estos pequeños cambios en los desplazamientos químicos sugieren que EPI-001 interacciona con una conformación de AF1* poco poblada. Esta interacción podría desplazar ligeramente el equilibrio hacia un estado conformacional con el que EPI-001 podría interaccionar.

La interacción de EPI-001 con AF1*c, que contiene las tres regiones que muestran perturbaciones en los desplazamientos químicos debidos a la interacción con EPI-001, causó cambios en los desplazamientos químicos incluso más pequeños en el espectro de NMR ^{1}H de EPI-001 comparado con la interacción de EPI-001 con AF1*. Esto indica

que el estado conformacional de la región Tau-5 con la que interacciona EPI-001 es mayor para AF1* que para AF1*c. Esto es compatible con la interacción específica de EPI-001 con la región Tau-5 cuando ésta adopta la conformación que tendría en el dímero, ya que esta conformación está poblada en mayor medida por AF1* que por AF1*c a una concentración dada.

En general, nuestros resultados muestran que la base molecular de la especificidad de EPI-001 por la región Tau-5 de AR es el reconocimiento de una conformación particular de la región Tau-5 por EPI-001, siendo posiblemente la conformación que adopta en el estado de dimerización.

Esta misma interacción fue comprobada desde el punto de vista del ligando, tanto por las perturbaciones en los desplazamientos químicos como el ensanchamiento de línea de las señales de EPI-001 en los espectros de NMR ^{1}H de EPI-001 a concentraciones cada vez mayores de AF1* y mediante NMR de diferencia de transferencia de saturación (STD).

Además el análisis por espectrometría de masas reveló que la cisteína 404 es la preferida cinéticamente (de las ocho cisteínas de AF1*) en la unión irreversible con EPI-001. Es importante señalar que la cisteína 404 se encuentra en el centro de la segunda región helicoidal que muestra perturbaciones pequeñas del desplazamiento químico en los experimentos de interacción no-covalente con EPI-001.

En conclusión, hemos demostrado que EPI-001 reconoce específicamente regiones contiguas en la secuencia de AF1* localizadas entre los residuos 345 y 448. Además estos resultados sugieren que la interacción de EPI-001 con AF1* involucra, primero, el reconocimiento reversible y específico de la región Tau-5 seguida de la unión irreversible y específica del compuesto a un nucleófilo presente en esta misma región de la proteína, la cisteína 404. El mecanismo molecular subyacente de la especificidad de EPI-001 por la región Tau-5 en nuestro sistema *in vitro* es aparentemente el reconocimiento de una conformación en particular adoptada por esta región, posiblemente la conformación adoptada en el estado dimérico, por el compuesto. Es posible que los determinantes moleculares de la interacción entre EPI-001 y el NTD de AR *in vivo* sean los mismos que los encontrados en nuestro sistema *in vitro*.

9.4 Conclusiones

9.4.1 Propiedades conformacionales de AF1*/NTD

1. Hemos encontrado que, *in vitro*, AF1* (AR 142–448) posee una tendencia re-

ducida pero notable a asociarse a sí misma formando un estado dimérico en una orientación cabeza-cabeza. Es posible que la dimerización del dominio NTD completo ocurra de manera similar *in vivo* cuando el AR se encuentre unido al DNA.

2. AF1* de AR es una proteína intrínsecamente desordenada con una helicidad excepcional en regiones esenciales para su función. Este hecho podría ser debido en parte a un estado dimérico de AF1* de baja población con un contenido más elevado de estructura secundaria que el monómero.

3. Las regiones de AF1* con una tendencia mayor de estructura secundaria y menor flexibilidad también están involucradas en la dimerización de AF1* y coinciden con las regiones que fueron previamente identificadas como funcionales (regiones centrales de Tau-1 y Tau-5).

4. Se sabe que Tau-1 y Tau-5 son unidades funcionalmente independientes, importantes para la regulación de la transactivación de AR tanto dependiente como independiente de andrógenos respectivamente. Hemos demostrado que además son unidades estructuralmente independientes (no se han encontrado interacciones de largo alcance que generen estructura entre Tau-1 y Tau-5).

9.4.2 Interacción de AR y RAP74

1. La interacción del motivo ^{433}WHTLF^{437} de AR, localizado en la región Tau-5, permite la transactivación por unión con RAP74 en células CRPC. Este mecanismo de transactivación del NTD de AR es hasta la fecha totalmente nuevo.

2. Nuestros datos sugieren que no es un mecanismo general de activación mas sí un mecanismo que permite la transactivación aberrante en células de pacientes CRPC activado por la fosforilación de residuos en el lado N-terminal del motivo ^{433}WHTLF^{437}.

9.4.3 Interacción de AR y EPI-001

1. Hemos demostrado que, *in vitro*, EPI-001 se une covalentemente a AF1* preferentemente en la cisteína 404.

2. El mecanismo de reconocimiento específico de EPI-001 por C404 de AF1* *in vitro* se basa en el reconocimiento de una conformación poco poblada de Tau-5 y es posible que esta conformación adoptada para el reconocimiento de EPI-001 sea la misma que la que está presente en el estado dimérico de AF1*.

A

Appendix

A.1 AR sequence and residue numbering used in this thesis, according to Uniprot

The AR sequence deposited in the Uniprot database (entry P10275, isoform 1), corresponding to 919 residues for the full-length protein, and with 21 glutamine residues in the polymeric glutamine stretch (between residues 58 and 78) and 24 glycine residues in the polymorphic glycine stretch (between residues 449 and 472). The residue numbering corresponding to this AR sequence is used throughout the thesis.

```
        10          20          30          40          50          60
MEVQLGLGRV YPRPPSKTYR GAFQNLFQSV REVIQNPGPR HPEAASAAPP GASLLLLQQQ

        70          80          90         100         110         120
QQQQQQQQQQ QQQQQQQQET SPRQQQQQQG EDGSPQAHRR GPTGYLVLDE EQQPSQPQSA

       130         140         150         160         170         180
LECHPERGCV PEPGAAVAAS KGLPQQLPAP PDEDDSAAPS TLSLLGPTFP GLSSCSADLK

       190         200         210         220         230         240
DILSEASTMQ LLQQQQQEAV SEGSSSGRAR EASGAPTSSK DNYLGGTSTI SDNAKELCKA

       250         260         270         280         290         300
VSVSMGLGVE ALEHLSPGEQ LRGDCMYAPL LGVPPAVRPT PCAPLAECKG SLLDDSAGKS

       310         320         330         340         350         360
TEDTAEYSPF KGGYTKGLEG ESLGCSGSAA AGSSGTLELP STLSLYKSGA LDEAAAYQSR

       370         380         390         400         410         420
DYYNFPLALA GPPPPPPPPH PHARIKLENP LDYGSAWAAA AAQCRYGDLA SLHGAGAAGP

       430         440         450         460         470         480
GSGSPSAAAS SSWHTLFTAE EGQLYGPCGG GGGGGGGGGG GGGGGGGGGG GGEAGAVAPY

       490         500         510         520         530         540
GYTRPPQGLA GQESDFTAPD VWYPGGMVSR VPYPSPTCVK SEMGPWMDSY SGPYGDMRLE

       550         560         570         580         590         600
TARDHVLPID YYFPPQKTCL ICGDEASGCH YGALTCGSCK VFFKRAAEGK QKYLCASRND

       610         620         630         640         650         660
CTIDKFRRKN CPSCRLRKCY EAGMTLGARK LKKLGNLKLQ EEGEASSTTS PTEETTQKLT

       670         680         690         700         710         720
VSHIEGYECQ PIFLNVLEAI EPGVVCAGHD NNQPDSFAAL LSSLNELGER QLVHVVKWAK

       730         740         750         760         770         780
ALPGFRNLHV DDQMAVIQYS WMGLMVFAMG WRSFTNVNSR MLYFAPDLVF NEYRMHKSRM

       790         800         810         820         830         840
YSQCVRMRHL SQEFGWLQIT PQEFLCMKAL LLFSIIPVDG LKNQKFFDEL RMNYIKELDR

       850         860         870         880         890         900
IIACKRKNPT SCSRRFYQLT KLLDSVQPIA RELHQFTFDL LIKSHMVSVD FPEMMAEIIS

       910
VQVPKILSGK VKPIYFHTQ
```

A.2 Sequences of protein constructs and fusion tags used in this thesis

The AR protein constructs and RAP74 protein constructs used in this PhD. Non-native amino acids are shown in grey.

AF1*a = AR 142-275

GGLPQQLPAPPDEDDSAAPSTLSLLGPTFPGLSSCSADLKDILSEASTMQLLQQQQQEAV
SEGSSSGRAREASGAPTSSKDNYLGGTSTISDNAKELCKAVSVSMGLGVEALEHLSPGEQ
LRGDCMYAPLLGVPP

AF1*b = AR 265-340

GCMYAPLLGVPPAVRPTPCAPLAECKGSLLDDSAGKSTEDTAEYSPFKGGYTKGLEGESL
GCSGSAAAGSSGTLELP

AF1*c = AR 330-448

GAAGSSGTLELPSTLSLYKSGALDEAAAYQSRDYYNFPLALAGPPPPPPPPHPHARIKLE
NPLDYGSAWAAAAAQCRYGDLASLHGAGAAGPGSGSPSAAASSSWHTLFTAEEGQLYGPC

AF1* = AR 142-448

GGLPQQLPAPPDEDDSAAPSTLSLLGPTFPGLSSCSADLKDILSEASTMQLLQQQQQEAV
SEGSSSGRAREASGAPTSSKDNYLGGTSTISDNAKELCKAVSVSMGLGVEALEHLSPGEQ
LRGDCMYAPLLGVPPAVRPTPCAPLAECKGSLLDDSAGKSTEDTAEYSPFKGGYTKGLEG
ESLGCSGSAAAGSSGTLELPSTLSLYKSGALDEAAAYQSRDYYNFPLALAGPPPPPPPPH
PHARIKLENPLDYGSAWAAAAAQCRYGDLASLHGAGAAGPGSGSPSAAASSSWHTLFTAE
EGQLYGPC

RAP74-CTD = RAP74 363-517 + C-terminal His-tag

MAKKKTPPKRERKPSGGSSRGNSRPGTPSAEGGSTSSTLRAAASKLEQGKRVSEMPAAKR
LRLDTGPQSLSGKSTPQPPSGKTTPNSGDVQVTEDAVRRYLTRKPMTTKDLLKKFQTKKT
GLSSEQTVNVLAQILKRLNPERKMINDKMHFSLKELEHHHHHH

RAP74NMR = RAP74 450-517

GDVQVTEDAVRRYLTRKPMTTKDLLKKFQTKKTGLSSEQTVNVLAQILKRLNPERKMIND
KMHFSLKE

The sequences of the N-terminal fusion tags encoded by destination vectors pDEST17, pDEST15 and pDEST-HisMBP are given below.

N-terminal His-tag encoded by pDEST17 vector:

MSYYHHHHHHLESTSLYKKAGS

N-terminal GST-tag encoded by pDEST15 vector:

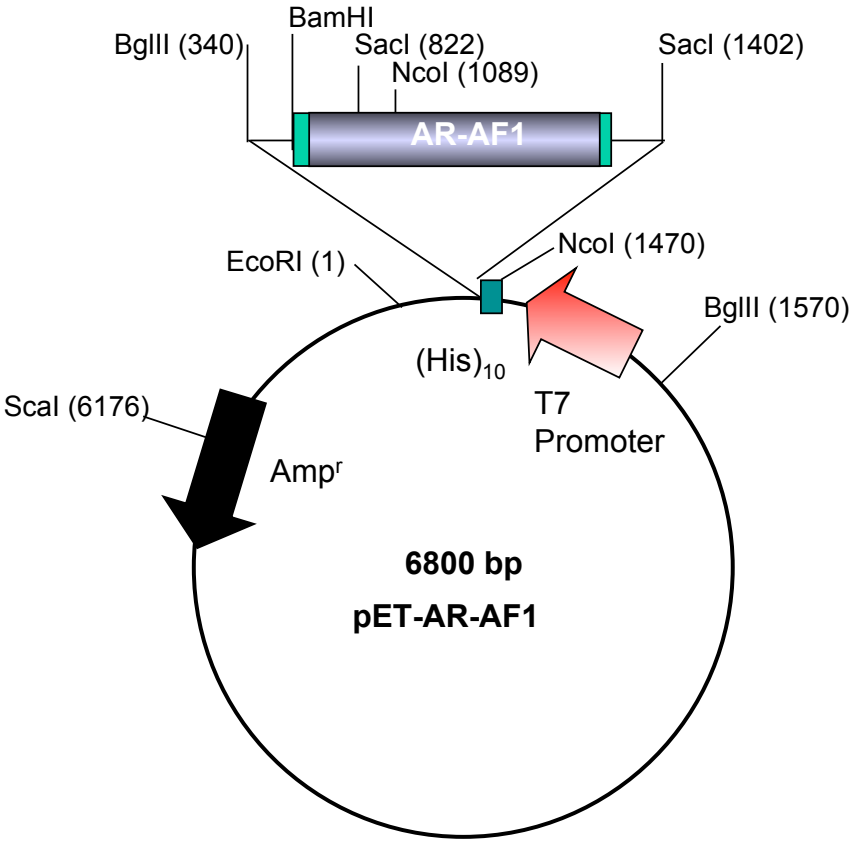
MSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFELGLEFPNLPYYID
GDVKLTQSMAIIRYIADKHNMLGGCPKERAEISMLEGAVLDIRYGVSRIAYSKDFETLKV
DFLSKLPEMLKMFEDRLCHKTYLNGDHVTHPDFMLYDALDVVLYMDPMCLDAFPKLVCFK
KRIEAIPQIDKYLKSSKYIAWPLQGWQATFGGGDHPPKSDLVPRPWSNQTSLYKKAGS

N-terminal HisMBP-tag encoded by pDEST-HisMBP vector:

MKIHHHHHHEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAAT
GDGPDIIFWAHDRFGGYAQSGLLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEAL
SLIYNKDLLPNPPKTWEEIPALDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAFKYEN
GKYDIKDVGVDNAGAKAGLTFLVDLIKNKHMNADTDYSIAEAAFNKGETAMTINGPWAWS
NIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGINAASPNKELAKEFLENYLLTDEGLEAV
NKDKPLGAVALKSYEEELAKDPRIAATMENAQKGEIMPNIPQMSAFWYAVRTAVINAASG
RQTVDEALKDAQTNSITSLYKKAGS

A.3 Vector map of pET-AR-AF1 and pET-19bm

Vector maps of the pET-AR-AF1 vector, kindly provided by Prof. Iain McEwan (University of Aberdeen, UK), and the pET-19bm on which it was based. The pET-AR-AF1 vector contains the AR DNA sequence corresponding to AR residues 143–494 (numbering used in this thesis, corresponds to residues 142–485 in the numbering used by Prof. McEwan and Prof. Brinkmann).

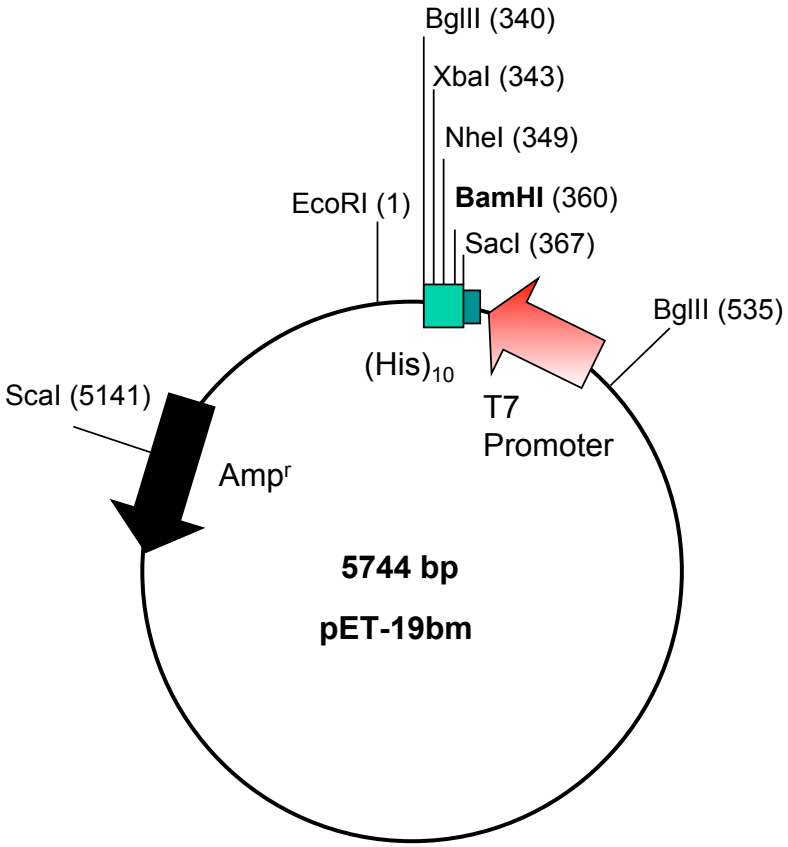


AR-AF1 was cloned as a PCR product, digested with BamHI and BglII enzymes, into the BamHI site of **pET-19bm** This destroyed the 5' BamHI site and retained a BamHI site at the 3'

Plasmid Name: pET-AR-AF1
Construction: Iain McEwan
Based on pET-19bm. Amino acids 142 to 485 of the AR-NTD were amplified by PCR. After digestion with BamHI and BglII the PCR product was cloned into pET-19bm digested with BamHI.

```
5' ARN142 GCGCGCAGATCTCTGCCGCAGCAGCTGCCAGC
3' ARC485 GCGCGCGGATCCGCTTTCCTGGCCCGCCAGCCCC
```

Sequencing: AR primers
Purpose: Bacterial expression plasmid. The AR-NTD, amino acids 142 to 485 (numbering for AR cDNA from Brinkmann laboratory) are fused to an N-terminal 10x histidine tag and synthesised in response to IPTG induction.



Plasmid pET-19bm is a variant of plasmid pET-19b (Invitrogen), with a different set of cloning sites introduced by Tony Wright (Karolinska Institute). The following double stranded oligonucleotide was inserted between NdeI and BamHI of pEt-19b. The original BamHI site is destroyed.

```
              SacI                    NheI
5′ TATGGCGAGCTCCGGATCCTAATAGCTAGCTCTAGATCT        3′
3′   ACCGCTCGAGGCCTAGGATTATCGATCGAGATCTAGACTAG    5′
                  BamHI              XbaI
```

Plasmid Name: pET-19bm
Construction: Tony Wright
Based on pET-19b (Novagen). The sequence below was cloned into the NdeI (5') and BamHI (3') sites to create new multicloning sites.

NdeI SacI BamHI NheI XbaI BamHI

```
TATGGCGAGCTCCGGATCCTAATAGCTAGCTCTAGATCT
  ACCGCTCGAGGCCTAGGATTATCGATCGAGATCTAGACTAG
```

Purpose: Bacterial expression plasmid. Recombinant proteins are fused to an N-terminal histadine tag and synthesised in response to IPTG induction.

A.4 Experimental conditions for NMR studies on IDPs from literature

This table is taken from [406] and lists more than 30 reported experimental conditions that have been used for NMR studies on IDPs, illustrating the compromise between temperature and pH to obtain good spectral quality for IDPs.

Table S1: List of sample conditions that were used to obtain heteronuclear NMR data of Intrinsically Disordered Proteins

BMRB entry	Protein	Temp (K)	pH	Comment	Reference
6968	Alpha-synuclein	288	6.8	Assigned by ^{13}C direct alone	9
N/D	Alpha-synuclein	283	7.4		10,11,12,1
N/D	Alpha-synuclein	288	7.4		13,14,3,15
N/D	Alpha-synuclein	263	7.4		16
N/D	Alpha-synuclein	258	7.4		17
15298	Beta-synuclein	283	6.5		18
N/D	Beta-synuclein	283	7.4		2
7244	Gamma-synuclein	278	7.4		19
N/D	Tau	273	7.4		20
N/D	Tau	293	6.8		21
N/D	Tau	293	6.8	In intact *Xenopus* oocytes and homogenized lysates	22
15136	Endosulfine alpha	283	7.4		23
15225	NS5A-domain 2 hepatitis C virus	298	6.5	Significant line broadening in the glycine region	24
15379	Sup35 NM domain	300	5	DMSO 95%	25
15397	Transcriptional activation domain 1 of ACTR	305	6.7	Free and in complex with CBP	26
15398	Nuclear coactivator binding domain of CBP	305	6.7	Free and in complex with ACTR	26
15409	Cytoplasmic domain of the T cell receptor zeta chain	288	6.7		27
N/D	Cytoplasmic domain of the T cell receptor zeta chain	288	6.2	N/A	28
15662	Murine amelogenin	298	3		29
4286	Unstructured Yeast Vesicular SNARE Snc1	288	4.5		30
4287	Unstructured Yeast Target Membrane SNARE Sso	288	4.5		30
4873	Bovine pancreatic trypsin inhibitor	271	4.6	Peptide fragment	31
4922	Unfolded apoplastocyanin	308	6	5 mM beta-Mercapto-EtOH	32

Table S1 (continued): List of sample conditions that were used to obtain heteronuclear NMR data of Intrinsically Disordered Proteins

BMRB entry	Protein	Temp (K)	pH	Comment	Reference
7159	Two-disulfide variant of hen lysozyme	298	3.8	Partially folded disulfide intermediate	33
N/D	Securin	283	7.2	Employed ^{13}C detection	34
N/D	Porcine amelogenin	283	3.8		35
N/D	hepatitis C virus core protein	278	6.6		36
N/D	FlgM	290	5.0	Substantial signal loss in isolation, which could be recovered with glucose crowding or in cell	37
7279	HMGA1	298	6.0		38
5076	Cold Shock Domain of the human YB-1 protein	303	6.7		39
4767	Apocytochrome b562	293	5.1		40

A.5 Backbone assignment of AF1*b (AR 265–340)

The data were recorded at 200 μM AF1*b, in 20 mM sodium phosphate buffer and 1 mM TCEP, at pH 7.4 and 278 K.

		H	N	C	CA	CB
	G	#N/A	#N/A	#N/A	#N/A	#N/A
265	C	#N/A	#N/A	174.81	58.57	28.42
266	M	8.696	124.15	175.66	55.62	32.68
267	Y	8.489	122.42	174.89	57.58	38.79
268	A	8.258	128.08	174.72	50.29	18.38
269	P	#N/A	#N/A	176.84	62.92	32.07
270	L	8.450	122.33	177.43	55.09	42.12
271	L	8.366	123.38	177.77	54.94	42.44
272	G	8.479	109.85	173.65	45.05	#N/A
273	V	8.077	121.26	174.33	59.76	32.54
274	P	#N/A	#N/A	#N/A	#N/A	#N/A
275	P	#N/A	#N/A	176.52	62.65	32.09
276	A	8.478	124.60	177.80	52.29	19.19
277	V	8.313	120.47	175.97	62.01	32.74
278	R	8.614	127.22	174.12	53.80	30.19
279	P	#N/A	#N/A	176.71	62.82	32.14
280	T	8.524	118.34	172.93	60.02	69.68
281	P	#N/A	#N/A	176.72	63.19	32.14
282	C	8.592	120.35	173.97	58.29	28.23
283	A	8.517	128.12	175.31	50.64	18.05
284	P	#N/A	#N/A	176.96	62.84	32.02
285	L	8.483	122.81	177.43	55.11	42.14
286	A	8.444	125.13	177.69	52.50	19.33
287	E	8.514	120.40	176.34	56.46	30.35
288	C	8.563	121.76	174.79	58.64	27.88
289	K	8.683	125.01	177.11	56.53	32.85
290	G	8.635	110.64	174.06	45.18	#N/A
291	S	8.364	115.89	174.62	58.23	63.92
292	L	8.475	124.28	177.46	55.36	42.14
293	L	8.291	122.44	177.06	55.00	42.38
294	D	8.300	121.37	176.22	54.16	41.09
295	D	8.472	122.54	176.94	54.45	40.79
296	S	8.439	116.44	175.01	59.36	63.62
297	A	8.289	125.56	178.47	53.04	18.87
298	G	8.407	108.00	174.22	45.35	#N/A
299	K	8.130	120.84	176.84	56.06	33.21
300	S	8.649	117.87	175.16	58.29	63.81
301	T	8.423	116.21	174.82	62.03	69.47
302	E	8.488	122.76	176.25	56.64	30.18
303	D	8.492	121.88	176.62	54.42	41.03
304	T	8.216	114.70	174.57	61.70	69.63
305	A	8.392	126.44	177.77	52.80	19.16
306	E	8.399	119.99	176.10	56.52	30.28
307	Y	8.345	121.99	175.36	57.91	38.77
308	S	8.139	120.05	172.65	55.34	63.88
309	P	#N/A	#N/A	176.57	63.22	31.91
310	F	8.173	119.63	175.95	57.68	39.17
311	K	8.194	123.93	176.67	56.25	32.89
312	G	7.991	109.54	174.36	45.35	#N/A
313	G	8.292	108.73	173.75	45.00	#N/A
314	Y	8.236	120.45	176.01	58.04	38.95
315	T	8.212	117.83	173.83	61.62	69.91
316	K	8.371	124.58	176.95	56.63	33.01
317	G	8.592	110.94	174.17	45.12	#N/A
318	L	8.338	121.92	177.83	55.13	42.25
319	E	8.712	121.67	177.16	57.06	29.87
320	G	8.509	109.96	174.28	45.27	#N/A
321	E	8.381	120.89	176.71	56.56	30.32
322	S	8.541	117.09	174.73	58.35	63.52

323	L	8.469	124.46	178.13	55.35	42.14
324	G	8.538	109.54	174.19	45.31	#N/A
325	C	8.351	119.04	175.02	58.41	28.28
326	S	8.661	118.81	175.17	58.89	63.64
327	G	8.622	111.48	174.32	45.33	#N/A
328	S	8.324	116.03	174.46	58.35	63.89
329	A	8.490	126.28	177.69	52.61	18.99
330	A	8.346	123.48	177.80	52.34	18.86
331	A	8.405	123.80	178.53	52.81	18.84
332	G	8.493	108.54	174.41	45.22	#N/A
333	S	8.340	115.74	174.92	58.34	63.90
334	S	8.612	118.03	175.08	58.95	63.79
335	G	8.534	110.98	174.17	45.29	#N/A
336	T	8.177	114.41	174.41	61.75	69.86
337	L	8.439	125.07	177.06	55.12	42.32
338	E	8.476	122.59	175.90	56.09	30.16
339	L	8.426	125.92	174.41	52.73	41.45
340	P	#N/A	#N/A	#N/A	#N/A	#N/A

A.6 Backbone assignment of AF1*c (AR 330–448)

The data were recorded at 290 μM AF1*c, in 20 mM sodium phosphate buffer and 1 mM TCEP, at pH 7.4 and 278 K.

		H	N	C	CA	CB
	G	#N/A	#N/A	#N/A	#N/A	#N/A
330	A	#N/A	#N/A	177.82	52.39	19.33
331	A	8.693	124.30	178.47	52.75	18.92
332	G	8.572	108.95	174.43	45.20	#N/A
333	S	8.370	115.83	174.94	58.34	63.94
334	S	8.630	118.07	175.08	58.69	63.79
335	G	8.549	111.03	174.24	45.31	#N/A
336	T	8.171	114.44	174.43	61.84	69.82
337	L	8.429	125.08	177.07	55.16	42.30
338	E	8.472	122.66	176.02	56.08	30.33
339	L	8.492	125.45	175.28	52.94	41.57
340	P	#N/A	#N/A	177.10	63.18	32.06
341	S	8.579	116.10	175.37	58.34	63.71
342	T	8.326	116.47	174.81	62.40	69.45
343	L	8.225	123.92	177.61	55.60	42.20
344	S	8.349	116.64	174.68	58.51	63.48
345	L	8.244	124.08	177.20	55.46	42.19
346	Y	8.175	120.54	175.78	58.03	38.72
347	K	8.236	124.00	176.17	56.06	33.13
348	S	8.449	117.45	175.21	58.62	63.73
349	G	8.570	111.43	173.98	45.25	#N/A
350	A	8.204	123.75	178.08	52.65	19.37
351	L	8.390	121.14	177.39	55.37	42.14
352	D	8.316	121.09	176.68	54.25	40.89
353	E	8.458	122.28	176.97	57.53	30.02
354	A	8.328	123.64	178.40	53.19	18.75
355	A	8.091	122.20	178.25	53.11	18.80
356	A	8.128	122.18	178.08	53.00	18.85
357	Y	8.048	118.78	176.10	58.41	38.39
358	Q	8.103	121.44	175.76	55.79	29.42
359	S	8.301	116.65	174.85	58.59	63.65
360	R	8.360	122.76	175.94	56.42	30.56
361	D	8.334	120.61	176.10	54.30	40.80
362	Y	8.009	120.28	175.64	58.68	38.57
363	Y	8.002	119.97	175.33	58.41	38.55
364	N	8.018	119.38	174.13	53.03	38.87
365	F	8.048	121.30	173.62	55.93	38.82
366	P	#N/A	#N/A	176.97	63.27	31.94
367	L	8.350	122.29	177.38	55.20	42.32
368	A	8.360	124.93	177.62	52.44	19.10
369	L	8.260	121.60	177.14	54.97	42.33
370	A	8.360	124.97	177.72	52.38	19.43
371	G	8.211	108.51	170.96	44.32	#N/A
372	P	#N/A	#N/A	#N/A	#N/A	#N/A
373	P	#N/A	#N/A	#N/A	#N/A	#N/A
374	P	#N/A	#N/A	#N/A	#N/A	#N/A
375	P	#N/A	#N/A	#N/A	#N/A	#N/A
376	P	#N/A	#N/A	#N/A	#N/A	#N/A
377	P	#N/A	#N/A	#N/A	#N/A	#N/A
378	P	#N/A	#N/A	#N/A	#N/A	#N/A
379	P	#N/A	#N/A	176.33	62.67	31.86
380	H	8.427	121.34	174.21	54.43	30.59
381	P	#N/A	#N/A	176.92	63.62	31.97
382	H	8.855	119.79	175.41	56.03	30.70
383	A	8.287	124.91	177.60	52.61	19.22
384	R	8.515	120.83	176.32	56.14	30.68
385	I	8.259	122.67	175.94	61.11	38.65
386	K	8.491	126.28	176.05	56.18	32.96
387	L	8.421	125.08	177.14	54.83	42.29

388	E	8.550	122.08	175.81	56.48	30.36
389	N	8.608	120.42	173.51	50.92	38.89
390	P	#N/A	#N/A	177.25	63.69	32.20
391	L	8.215	119.89	177.25	55.35	41.64
392	D	8.026	119.87	176.18	54.05	40.95
393	Y	8.086	121.18	176.78	58.79	38.30
394	G	8.455	109.62	174.93	45.65	#N/A
395	S	8.228	116.11	175.54	59.38	63.50
396	A	8.538	125.88	179.09	53.95	18.30
397	W	8.183	120.01	177.22	58.90	28.93
398	A	7.943	123.76	178.85	53.69	18.60
399	A	7.998	121.90	178.93	53.62	18.41
400	A	7.965	121.89	179.00	53.62	18.38
401	A	8.118	122.04	178.75	53.43	18.35
402	A	7.951	121.17	178.43	53.34	18.66
403	Q	8.013	117.74	176.37	56.34	29.05
404	C	8.132	119.44	174.75	59.31	27.65
405	R	8.299	122.87	176.17	56.56	30.52
406	Y	8.308	120.59	176.60	58.31	38.65
407	G	8.362	110.13	173.97	45.38	#N/A
408	D	8.248	120.63	176.86	54.39	41.03
409	L	8.311	122.74	177.94	56.02	41.87
410	A	8.306	123.18	178.61	53.43	18.72
411	S	8.110	113.86	175.03	58.87	63.43
412	L	8.070	123.33	177.53	55.68	42.07
413	H	8.215	118.75	176.06	56.36	30.62
414	G	8.331	109.85	174.09	45.32	#N/A
415	A	8.337	123.90	178.42	52.77	19.15
416	G	8.492	108.26	174.03	45.18	#N/A
417	A	8.187	123.83	177.48	52.51	19.39
418	A	8.376	123.09	177.87	52.31	19.40
419	G	8.227	108.11	172.04	44.49	#N/A
420	P	#N/A	#N/A	177.85	63.54	32.01
421	G	8.701	110.11	174.45	45.23	#N/A
422	S	8.311	115.66	175.14	58.66	63.92
423	G	8.598	110.94	173.82	45.06	#N/A
424	S	8.280	116.92	172.94	56.41	63.30
425	P	#N/A	#N/A	177.16	63.50	32.02
426	S	8.450	115.93	174.56	58.34	63.78
427	A	8.417	126.37	177.78	52.75	18.97
428	A	8.297	123.01	177.79	52.73	18.92
429	A	8.298	123.03	178.13	52.71	18.95
430	S	8.285	114.78	174.89	58.52	63.69
431	S	8.389	117.76	174.74	58.62	63.69
432	S	8.357	117.50	174.49	58.65	63.58
433	W	8.085	122.62	176.14	57.77	29.36
434	H	7.933	119.87	175.40	56.63	30.59
435	T	7.971	115.00	174.29	62.35	69.62
436	L	8.200	124.00	176.93	55.33	42.32
437	F	8.230	120.76	175.79	57.66	39.69
438	T	8.170	116.20	174.01	61.50	70.09
439	A	8.436	126.40	178.02	52.79	19.11
440	E	8.494	120.10	176.95	56.89	30.16
441	E	8.501	122.00	177.30	57.18	30.10
442	G	8.541	109.84	174.17	45.42	#N/A
443	Q	8.147	119.60	175.81	55.69	29.28
444	L	8.271	123.21	177.01	55.25	42.41
445	Y	8.291	120.18	175.89	57.43	38.96
446	G	8.182	110.23	171.34	44.49	#N/A
447	P	#N/A	#N/A	176.43	63.32	32.04

448	C	8.190	123.88	178.80	59.45	29.05

A.7 Backbone assignment of AF1* (AR 142–448)

The data were recorded at 340 μM AF1*, in 20 mM sodium phosphate buffer and 1 mM TCEP, at pH 7.4 and 278 K.

		H	N	C	CA	CB
	G	#N/A	#N/A	#N/A	#N/A	#N/A
142	G	#N/A	#N/A	#N/A	#N/A	#N/A
143	L	#N/A	#N/A	#N/A	#N/A	#N/A
144	P	#N/A	#N/A	176.85	63.13	31.99
145	Q	8.621	120.93	175.83	55.78	29.69
146	Q	8.570	122.46	175.63	55.47	29.52
147	L	8.572	126.14	175.11	53.06	41.50
148	P	#N/A	#N/A	176.19	62.77	32.01
149	A	8.512	126.11	175.12	50.30	18.00
150	P	#N/A	#N/A	#N/A	#N/A	#N/A
151	P	#N/A	#N/A	176.66	62.90	32.03
152	D	8.551	120.83	176.43	54.44	41.06
153	E	8.494	121.47	176.19	56.46	30.46
154	D	8.530	121.81	176.22	54.31	41.19
155	D	8.486	122.74	176.84	54.33	40.84
156	S	8.430	116.76	174.45	59.15	63.71
157	A	8.262	125.54	177.22	52.15	19.21
158	A	8.254	124.96	175.59	50.61	17.96
159	P	#N/A	#N/A	177.23	63.18	32.09
160	S	8.657	116.43	175.35	58.48	63.72
161	T	8.378	116.75	174.82	62.32	69.48
162	L	8.258	124.21	177.65	55.57	42.18
163	S	8.377	116.82	174.76	58.42	63.54
164	L	8.352	124.56	177.44	55.29	42.10
165	L	8.201	121.85	177.58	54.99	42.37
166	G	8.205	109.56	171.69	44.58	#N/A
167	P	#N/A	#N/A	177.13	63.10	32.22
168	T	8.276	114.26	173.88	61.73	69.91
169	F	8.394	123.54	174.05	55.58	39.33
170	P	#N/A	#N/A	177.36	63.52	32.02
171	G	8.163	108.73	174.12	45.13	#N/A
172	L	8.255	121.81	177.80	55.20	42.48
173	S	8.548	117.02	174.68	58.32	63.77
174	S	8.509	118.02	174.43	58.39	63.75
175	C	8.483	121.02	174.78	58.41	28.26
176	S	8.544	118.86	174.60	58.72	63.68
177	A	8.516	126.31	177.50	52.91	19.16
178	D	8.338	119.26	176.49	54.44	41.00
179	L	8.198	122.90	177.60	55.46	42.09
180	K	8.318	121.12	176.66	56.81	32.77
181	D	8.348	121.02	176.53	54.53	40.96
182	I	8.119	121.43	176.86	61.86	38.41
183	L	8.418	125.10	177.93	55.52	42.01
184	S	8.351	116.95	175.19	58.70	63.81
185	E	8.586	123.32	177.18	57.64	29.92
186	A	8.419	123.96	179.07	53.71	18.77
187	S	8.396	115.26	175.88	59.65	63.38
188	T	8.267	117.20	175.61	64.08	69.20
189	M	8.307	121.49	177.43	57.09	32.24
190	Q	8.266	120.72	177.25	57.41	28.75
191	L	8.161	122.18	178.49	56.58	41.90
192	L	8.147	121.28	178.53	56.45	41.91
193	Q	8.242	119.77	177.24	57.11	28.79
194	Q	8.315	120.61	177.03	57.10	28.95
195	Q	8.412	120.51	177.08	56.84	28.99
196	Q	8.418	120.93	176.65	56.63	28.99
197	Q	8.403	120.96	176.63	56.50	29.25
198	E	8.472	122.02	176.57	56.91	30.13
199	A	8.350	124.92	178.13	52.75	18.96

200	V	8.214	119.59	176.64	62.50	32.63
201	S	8.489	119.77	174.85	58.39	63.81
202	E	8.667	123.70	177.28	57.09	30.10
203	G	8.590	110.09	174.58	45.31	#N/A
204	S	8.353	115.88	175.19	58.44	63.98
205	S	8.624	118.24	175.08	58.73	63.64
206	S	8.465	117.87	175.31	58.85	63.74
207	G	8.498	111.02	174.27	45.45	#N/A
208	R	8.170	120.79	176.39	56.11	30.90
209	A	8.494	125.54	178.04	52.74	19.01
210	R	8.462	120.77	176.67	56.30	30.70
211	E	8.560	122.34	176.47	56.53	30.19
212	A	8.542	125.60	177.96	52.68	19.10
213	S	8.452	115.44	175.15	58.62	63.74
214	G	8.493	111.03	173.44	44.94	#N/A
215	A	8.230	125.02	175.66	50.59	18.05
216	P	#N/A	#N/A	177.34	63.09	32.05
217	T	8.466	114.84	174.68	61.87	69.84
218	S	8.488	118.08	174.66	58.08	63.98
219	S	8.585	118.50	174.85	58.55	63.76
220	K	8.474	123.23	176.43	56.62	32.78
221	D	8.295	120.71	175.92	54.53	41.03
222	N	8.335	119.07	175.03	53.14	38.66
223	Y	8.309	121.06	176.15	58.32	38.31
224	L	8.318	124.54	177.71	55.14	42.01
225	G	7.913	108.91	174.61	45.38	#N/A
226	G	8.333	108.75	174.50	45.21	#N/A
227	T	8.274	113.67	174.78	61.82	69.89
228	S	8.543	118.58	174.74	58.31	63.83
229	T	8.399	116.82	174.63	62.05	69.77
230	I	8.307	123.53	176.43	61.36	38.57
231	S	8.517	120.41	174.50	58.38	63.79
232	D	8.526	123.08	176.50	54.77	40.89
233	N	8.474	119.30	175.58	53.65	38.57
234	A	8.294	123.94	178.53	53.46	18.86
235	K	8.248	119.73	177.29	57.17	32.76
236	E	8.314	121.14	177.02	56.85	29.99
237	L	8.302	122.95	177.67	55.64	42.13
238	C	8.316	119.80	174.88	59.05	27.69
239	K	8.378	124.08	176.36	56.52	32.91
240	A	8.394	125.57	177.95	52.67	19.14
241	V	8.287	120.12	176.45	62.47	32.64
242	S	8.489	119.82	174.80	58.34	63.77
243	V	8.402	122.42	176.50	62.54	32.70
244	S	8.481	119.23	174.80	58.53	63.55
245	M	8.533	122.68	176.77	55.54	32.61
246	G	8.460	109.68	174.08	45.33	#N/A
247	L	8.298	121.49	178.11	55.16	42.41
248	G	8.628	110.08	174.34	45.26	#N/A
249	V	8.098	119.08	176.63	62.62	32.55
250	E	8.707	124.39	176.39	56.91	29.83
251	A	8.358	124.93	177.76	52.55	19.14
252	L	8.219	121.13	177.74	55.26	42.22
253	E	8.388	120.94	176.20	56.83	30.24
254	H	8.378	119.86	175.00	55.95	30.35
255	L	8.199	123.48	177.17	54.75	42.57
256	S	8.624	118.86	172.85	56.56	63.11
257	P	#N/A	#N/A	177.93	64.04	32.00
258	G	8.600	109.08	174.51	45.35	#N/A
259	E	8.210	120.82	176.79	56.74	30.44

260	Q	8.507	121.24	176.08	56.01	29.29
261	L	8.379	123.31	177.50	55.28	42.08
262	R	8.424	121.87	176.99	56.31	30.95
263	G	8.494	109.78	173.97	45.40	#N/A
264	D	8.388	120.32	176.46	54.32	41.02
265	C	8.358	119.27	174.68	58.83	27.91
266	M	8.474	122.44	175.78	55.72	32.43
267	Y	8.218	121.18	174.90	57.54	38.80
268	A	8.213	127.69	174.80	50.37	18.34
269	P	#N/A	#N/A	176.88	62.98	32.05
270	L	8.442	122.20	177.47	55.14	42.08
271	L	8.364	123.27	177.78	55.04	42.30
272	G	8.469	109.75	173.64	45.06	#N/A
273	V	8.073	121.23	174.33	59.75	32.55
274	P	#N/A	#N/A	#N/A	#N/A	#N/A
275	P	#N/A	#N/A	176.52	62.67	32.07
276	A	8.479	124.61	177.82	52.29	19.18
277	V	8.317	120.51	175.99	62.02	32.78
278	R	8.621	127.27	174.11	53.83	30.23
279	P	#N/A	#N/A	176.71	62.83	32.19
280	T	8.540	118.48	172.93	60.07	69.73
281	P	#N/A	#N/A	176.71	63.20	32.23
282	C	8.601	120.44	173.97	58.26	28.23
283	A	8.535	128.22	175.32	50.63	18.02
284	P	#N/A	#N/A	176.93	62.86	32.07
285	L	8.484	122.91	177.42	55.14	42.25
286	A	8.450	125.20	177.69	52.47	19.24
287	E	8.516	120.46	176.37	56.52	30.37
288	C	8.575	121.83	174.80	58.62	27.89
289	K	8.685	125.00	177.12	56.57	32.87
290	G	8.628	110.59	174.05	45.23	#N/A
291	S	8.365	115.90	174.65	58.21	63.98
292	L	8.475	124.28	177.47	55.38	42.11
293	L	8.290	122.47	177.09	55.13	42.11
294	D	8.304	121.43	176.23	54.07	41.10
295	D	8.481	122.58	176.95	54.43	41.00
296	S	8.440	116.47	175.02	59.37	63.72
297	A	8.291	125.58	178.49	53.06	18.90
298	G	8.407	108.00	174.23	45.36	#N/A
299	K	8.132	120.84	176.85	56.10	33.19
300	S	8.649	117.87	175.17	58.32	63.79
301	T	8.423	116.23	174.79	62.03	69.62
302	E	8.488	122.79	176.26	56.64	30.22
303	D	8.496	121.86	176.63	54.44	41.04
304	T	8.218	114.73	174.59	61.75	69.72
305	A	8.393	126.45	177.76	52.80	19.15
306	E	8.401	120.05	176.11	56.56	30.32
307	Y	8.348	121.99	175.36	57.94	38.79
308	S	8.145	120.06	172.63	55.38	63.92
309	P	#N/A	#N/A	176.56	63.23	31.89
310	F	8.171	119.65	175.96	57.70	39.18
311	K	8.197	123.91	176.67	56.27	32.90
312	G	7.990	109.53	174.37	45.36	#N/A
313	G	8.290	108.73	173.74	45.08	#N/A
314	Y	8.236	120.44	176.01	58.07	38.95
315	T	8.211	117.77	173.85	61.67	70.00
316	K	8.369	124.55	176.95	56.63	32.95
317	G	8.591	110.90	174.19	45.13	#N/A
318	L	8.333	121.90	177.84	55.14	42.21
319	E	8.711	121.66	177.16	57.09	29.90

320	G	8.508	109.93	174.29	45.32	#N/A
321	E	8.382	120.90	176.73	56.58	30.35
322	S	8.538	117.05	174.75	58.41	63.71
323	L	8.467	124.41	178.13	55.43	42.18
324	G	8.535	109.52	174.19	45.34	#N/A
325	C	8.347	119.03	175.03	58.43	28.26
326	S	8.660	118.76	175.17	58.82	63.71
327	G	8.618	111.49	174.32	45.39	#N/A
328	S	8.325	116.04	174.52	58.40	63.90
329	A	8.489	126.24	177.71	52.60	19.08
330	A	8.337	123.41	177.81	52.47	19.07
331	A	8.393	123.73	178.52	52.80	19.09
332	G	8.488	108.44	174.44	45.30	#N/A
333	S	8.337	115.73	174.95	58.44	63.92
334	S	8.609	118.05	175.11	58.77	63.74
335	G	8.530	110.97	174.22	45.33	#N/A
336	T	8.169	114.36	174.43	61.85	69.83
337	L	8.421	125.08	177.07	55.17	42.24
338	E	8.471	122.62	176.02	56.10	30.27
339	L	8.493	125.49	175.27	52.85	41.58
340	P	#N/A	#N/A	177.10	63.19	32.08
341	S	8.580	116.10	175.39	58.32	63.70
342	T	8.328	116.45	174.83	62.39	69.45
343	L	8.224	123.95	177.62	55.64	42.22
344	S	8.353	116.66	174.70	58.53	63.47
345	L	8.242	124.10	177.23	55.46	42.16
346	Y	8.176	120.51	175.79	58.08	38.69
347	K	8.237	123.99	176.19	56.06	33.13
348	S	8.447	117.42	175.21	58.63	63.72
349	G	8.569	111.49	173.99	45.23	#N/A
350	A	8.205	123.76	178.08	52.70	19.39
351	L	8.390	121.09	177.37	55.33	42.27
352	D	8.318	121.13	176.66	54.28	40.97
353	E	8.464	122.30	176.96	57.50	30.06
354	A	8.331	123.63	178.39	53.23	18.72
355	A	8.094	122.20	178.25	53.05	18.84
356	A	8.132	122.18	178.07	52.99	18.85
357	Y	8.053	118.79	176.09	58.43	38.43
358	Q	8.106	121.46	175.76	55.78	29.46
359	S	8.306	116.68	174.85	58.58	63.68
360	R	8.364	122.78	175.93	56.46	30.59
361	D	8.337	120.63	176.08	54.33	40.83
362	Y	8.014	120.26	175.64	58.66	38.56
363	Y	8.006	120.02	175.32	58.33	38.56
364	N	8.022	119.43	174.13	53.01	38.96
365	F	8.056	121.30	173.63	55.88	38.81
366	P	#N/A	#N/A	176.96	63.29	31.91
367	L	8.353	122.28	177.37	55.08	42.32
368	A	8.365	124.90	177.62	52.43	19.12
369	L	8.261	121.66	177.14	55.07	42.32
370	A	8.363	124.99	177.73	52.34	19.44
371	G	8.217	108.44	170.96	44.31	#N/A
372	P	#N/A	#N/A	#N/A	#N/A	#N/A
373	P	#N/A	#N/A	#N/A	#N/A	#N/A
374	P	#N/A	#N/A	#N/A	#N/A	#N/A
375	P	#N/A	#N/A	#N/A	#N/A	#N/A
376	P	#N/A	#N/A	#N/A	#N/A	#N/A
377	P	#N/A	#N/A	#N/A	#N/A	#N/A
378	P	#N/A	#N/A	#N/A	#N/A	#N/A
379	P	#N/A	#N/A	176.35	62.76	31.93

380	H	8.430	121.36	174.17	54.42	30.54
381	P	#N/A	#N/A	176.92	63.63	31.98
382	H	8.860	119.81	175.40	56.04	30.67
383	A	8.296	124.99	177.60	52.63	19.24
384	R	8.520	120.83	176.32	56.20	30.67
385	I	8.265	122.69	175.96	61.14	38.77
386	K	8.501	126.34	176.06	56.13	33.03
387	L	8.426	125.10	177.15	54.78	42.30
388	E	8.554	122.15	175.82	56.46	30.42
389	N	8.615	120.53	173.53	50.95	38.89
390	P	#N/A	#N/A	177.24	63.66	32.11
391	L	8.222	119.94	177.25	55.36	41.61
392	D	8.030	119.90	176.19	54.07	40.96
393	Y	8.086	121.19	176.77	58.77	38.29
394	G	8.458	109.64	174.92	45.59	#N/A
395	S	8.230	116.08	175.53	59.39	63.46
396	A	8.540	125.82	179.09	53.94	18.36
397	W	8.187	120.00	177.21	58.93	28.96
398	A	7.947	123.78	178.84	53.62	18.59
399	A	8.004	121.94	178.91	53.62	18.41
400	A	7.969	121.89	178.98	53.63	18.46
401	A	8.120	121.99	178.74	53.48	18.34
402	A	7.956	121.20	178.44	53.31	18.73
403	Q	8.017	117.74	176.37	56.32	29.08
404	C	8.134	119.46	174.75	59.33	27.71
405	R	8.299	122.90	176.19	56.55	30.36
406	Y	8.314	120.63	176.60	58.33	38.64
407	G	8.362	110.12	173.97	45.33	#N/A
408	D	8.253	120.67	176.85	54.45	41.00
409	L	8.315	122.70	177.92	56.08	41.86
410	A	8.310	123.19	178.61	53.45	18.74
411	S	8.113	113.86	175.04	58.87	63.45
412	L	8.072	123.35	177.53	55.71	42.08
413	H	8.220	118.77	176.06	56.35	30.62
414	G	8.335	109.83	174.09	45.34	#N/A
415	A	8.340	123.87	178.42	52.80	19.11
416	G	8.492	108.34	174.05	45.21	#N/A
417	A	8.190	123.87	177.47	52.52	19.37
418	A	8.379	123.13	177.89	52.40	19.43
419	G	8.231	108.13	172.04	44.49	#N/A
420	P	#N/A	#N/A	177.85	63.55	32.02
421	G	8.705	110.12	174.47	45.23	#N/A
422	S	8.312	115.69	175.13	58.63	63.95
423	G	8.601	110.92	173.82	45.04	#N/A
424	S	8.283	116.94	172.94	56.41	63.38
425	P	#N/A	#N/A	177.17	63.51	31.99
426	S	8.456	115.95	174.57	58.33	63.74
427	A	8.422	126.38	177.78	52.77	19.05
428	A	8.298	123.05	177.78	52.67	19.08
429	A	8.298	123.04	178.12	52.71	19.07
430	S	8.291	114.75	174.89	58.51	63.71
431	S	8.394	117.77	174.75	58.51	63.71
432	S	8.363	117.55	174.49	58.53	63.65
433	W	8.088	122.63	176.14	57.77	29.34
434	H	7.937	119.91	175.40	56.60	30.62
435	T	7.977	115.05	174.32	62.39	69.67
436	L	8.210	123.95	176.95	55.23	42.17
437	F	8.233	120.77	175.79	57.69	39.69
438	T	8.172	116.21	174.01	61.51	70.05
439	A	8.435	126.39	178.02	52.75	19.15

440	E	8.496	120.14	176.95	56.91	30.28
441	E	8.506	121.98	177.30	57.15	30.08
442	G	8.545	109.87	174.17	45.42	#N/A
443	Q	8.150	119.66	175.82	55.71	29.30
444	L	8.274	123.25	177.01	55.25	42.41
445	Y	8.294	120.19	175.89	57.43	38.98
446	G	8.186	110.27	171.35	44.50	#N/A
447	P	#N/A	#N/A	176.43	63.36	32.03
448	C	8.194	123.88	178.81	59.49	29.10

A.8 Backbone assignment of RAP74NMR (RAP74 450–517)

The data were recorded in 20 mM sodium phosphate buffer, at pH 6.5 and 298 K.

		H	N	C	CA
450	G	#N/A	#N/A	#N/A	#N/A
451	D	#N/A	#N/A	176.25	54.76
452	V	8.261	120.44	175.33	62.43
453	Q	8.330	123.05	175.42	55.10
454	V	8.586	123.54	174.29	62.37
455	T	6.718	113.70	174.63	58.75
456	E	9.159	122.09	177.66	60.25
457	D	8.443	116.60	177.94	57.40
458	A	7.848	121.55	179.58	54.98
459	V	7.941	116.68	177.38	67.72
460	R	8.757	119.55	178.32	60.22
461	R	7.813	118.47	179.45	59.67
462	Y	7.674	117.23	178.54	61.35
463	L	8.340	115.58	178.64	56.57
464	T	7.971	111.57	175.81	63.48
465	R	7.440	122.13	176.72	58.36
466	K	7.873	117.07	#N/A	54.52
467	P	#N/A	#N/A	178.19	62.24
468	M	8.929	121.14	176.06	54.76
469	T	9.034	113.53	175.97	60.21
470	T	9.081	116.07	176.52	67.89
471	K	8.008	120.65	178.72	59.92
472	D	7.955	119.99	179.47	57.30
473	L	8.005	120.50	178.35	58.06
474	L	8.488	119.59	179.43	58.54
475	K	7.597	116.90	178.83	58.84
476	K	7.326	115.61	177.69	55.85
477	F	7.548	117.12	175.74	58.08
478	Q	8.212	120.90	176.90	56.90
479	T	8.155	118.99	176.98	56.76
480	K	8.062	111.20	#N/A	62.03
481	K	#N/A	#N/A	#N/A	#N/A
482	T	#N/A	#N/A	176.37	62.02
483	G	8.501	110.57	173.90	45.58
484	L	8.107	121.13	177.58	54.07
485	S	8.758	116.96	#N/A	57.55
486	S	#N/A	#N/A	176.05	62.01
487	E	8.707	121.28	178.88	59.66
488	Q	7.996	118.33	178.59	58.51
489	T	8.229	117.09	175.51	67.46
490	V	8.089	120.69	177.91	67.07
491	N	7.609	118.28	177.98	56.35
492	V	8.263	121.47	178.60	66.25
493	L	8.463	120.26	178.07	58.06
494	A	8.184	119.90	181.03	55.48
495	Q	7.549	117.45	179.05	58.67
496	I	8.151	121.60	178.09	65.37
497	L	8.793	119.21	179.49	58.38
498	K	7.557	118.61	178.84	59.31
499	R	7.408	118.13	178.47	58.72
500	L	8.206	117.06	176.46	56.20
501	N	7.821	112.60	#N/A	53.40
502	P	#N/A	#N/A	176.65	62.29
503	E	8.377	120.15	175.94	57.23
504	R	8.424	123.05	175.98	54.80
505	K	8.916	123.75	174.84	54.81
506	M	8.777	124.83	176.07	54.03
507	I	8.423	126.38	176.02	61.18
508	N	9.411	126.36	174.26	54.85

509	D	8.694	109.65	174.92	56.13
510	K	7.856	117.77	177.00	54.80
511	M	9.146	128.52	174.83	56.84
512	H	9.349	126.57	173.28	54.98
513	F	9.046	120.29	175.59	56.13
514	S	8.771	113.18	171.98	57.69
515	L	7.413	123.01	175.57	53.31
516	K	8.480	121.39	175.90	56.37
517	E	8.064	126.33	#N/A	58.09

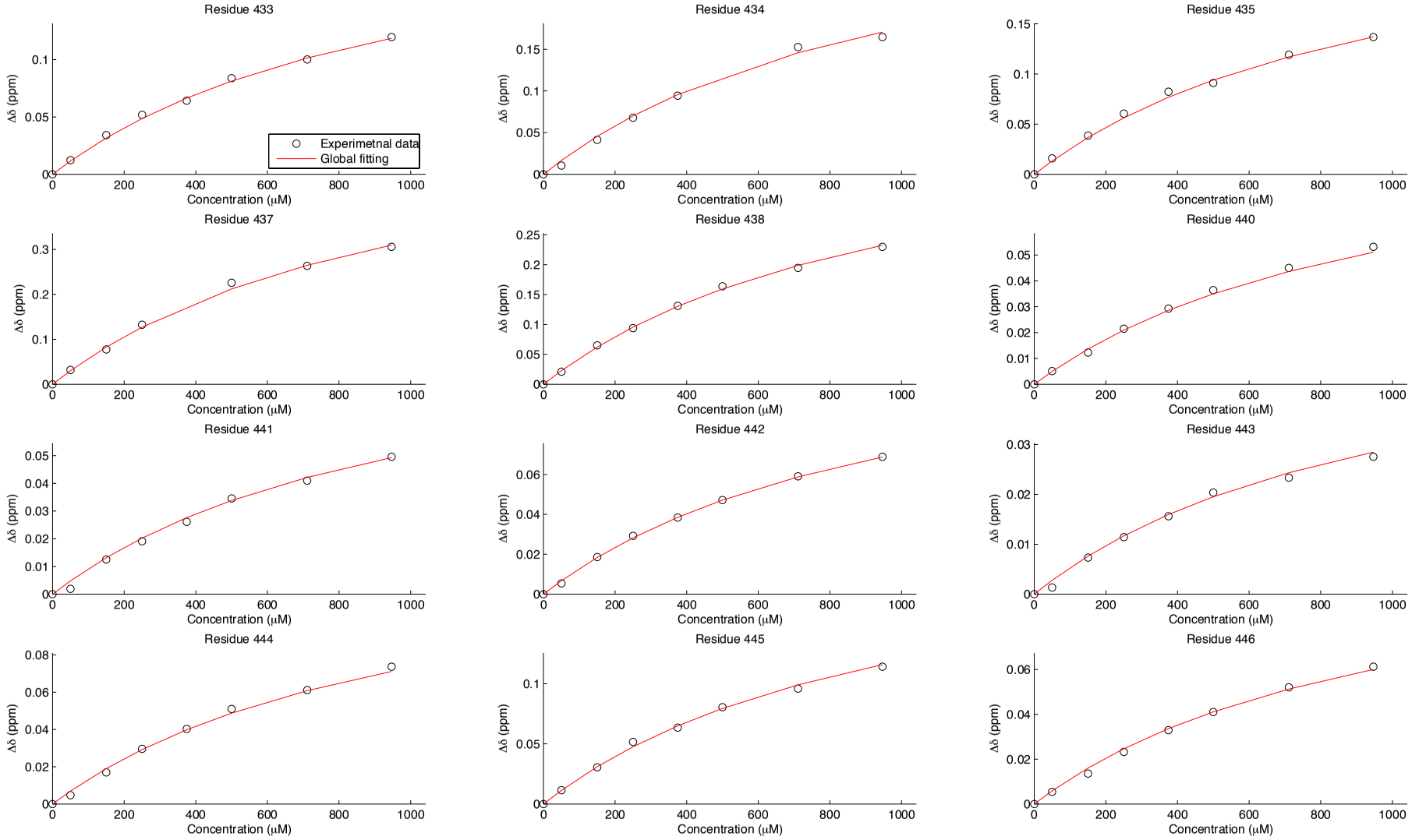
A.9 Global fitting

Results from the global fitting for each of the titrations, using 10% bootstrap (100 independent calculations).

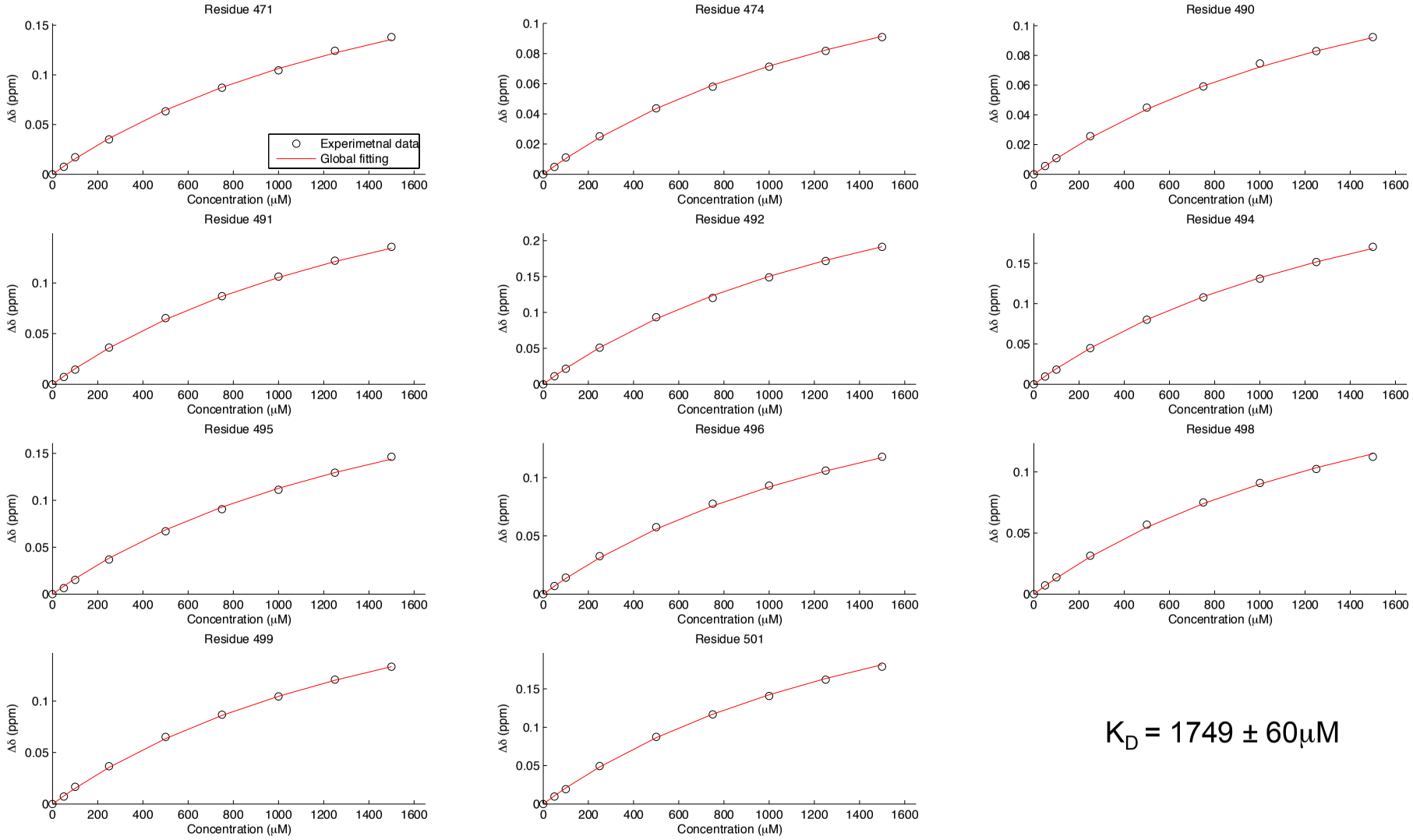
To determine the affinity from the titration of ^{15}N-AF1*c and RAP74NMR, a global fitting was performed for the residues in the binding site (AR residues: W433, H434, T435, F437, T438, A439, E440, E441, G442, Q443, L444, Y445 and G446). The peak corresponding to L436 was overlapped in the spectra and is therefore not used for the fitting. On the x-axis the concentration of RAP74NMR added to 50 μM ^{15}N-AF1*c is indicated, and on the y-axis the average ^{1}H and ^{15}N chemical shift difference between free ^{15}N-AF1*c and ^{15}N-AF1*c in the presence of RAP74NMR. Black circles indicate the experimentally obtained data, whereas the red curve corresponds to the global fitting.

For the titrations of ^{15}N-RAP74NMR and the AR peptides (WT, pS432, S430E/S431E/S432E and pS430/pS432) global fitting was performed for the residues in the binding site (RAP74 residues: T470, K471, L474, V490, N491, V492, A494, Q495, I496, K498, R499 and N501). On the x-axis the concentration of peptide added to 50 μM ^{15}N-RAP74NMR is indicated, and on the y-axis the average ^{1}H and ^{15}N chemical shift difference between free ^{15}N-RAP74NMR and ^{15}N-RAP74NMR in the presence of peptide. Black circles indicate the experimentally obtained data, whereas the red curve corresponds to the global fitting.

Titration of ^{15}N-AF1*c and RAP74NMR (K_D = 966 ± 39μM)

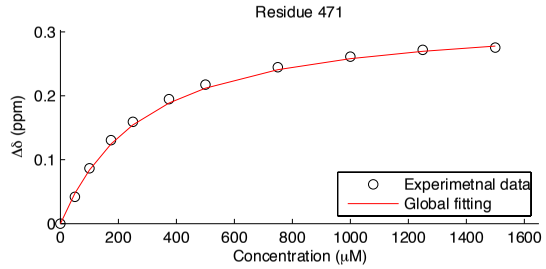


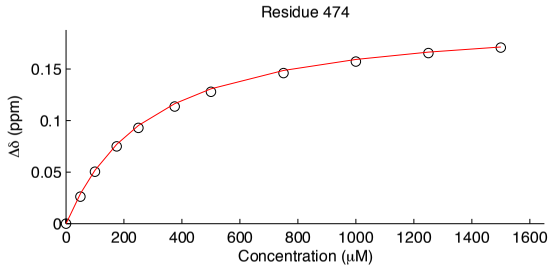
Titration of ^{15}N-RAP74NMR and WT peptide

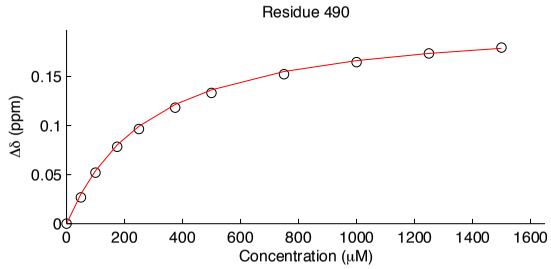


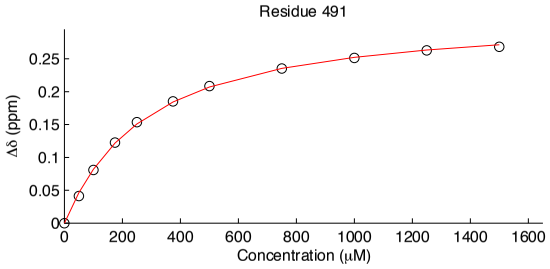
K_D = 1749 ± 60μM

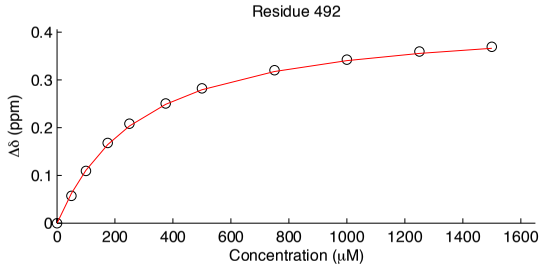
Titration of ^{15}N-RAP74NMR and pS432 peptide

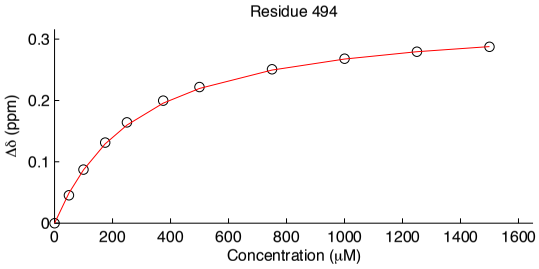


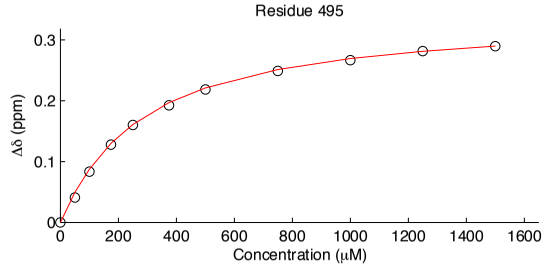


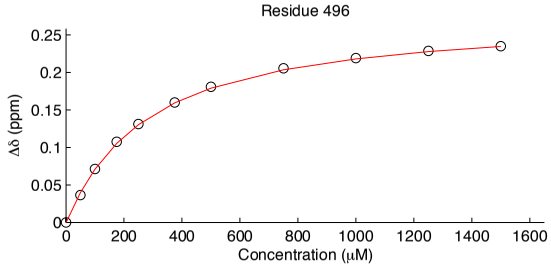


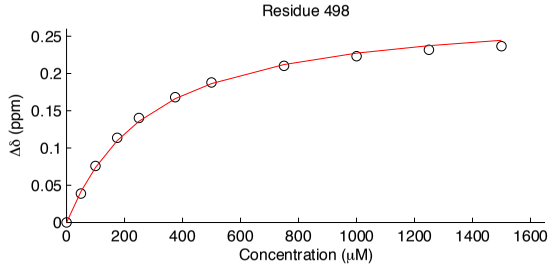


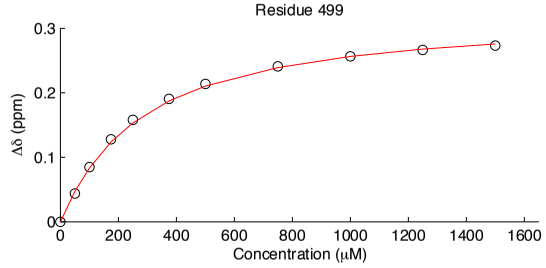


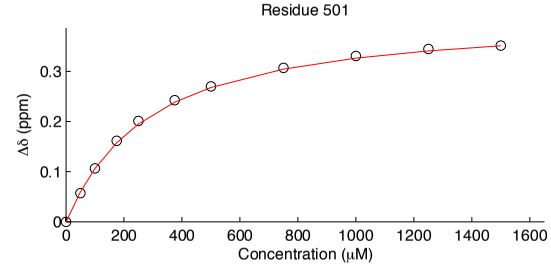






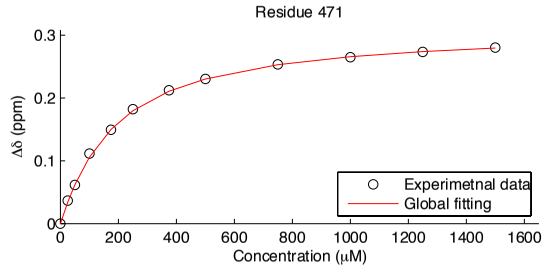


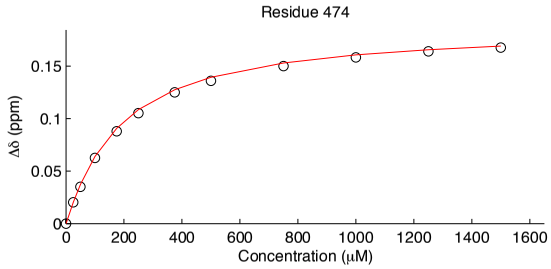


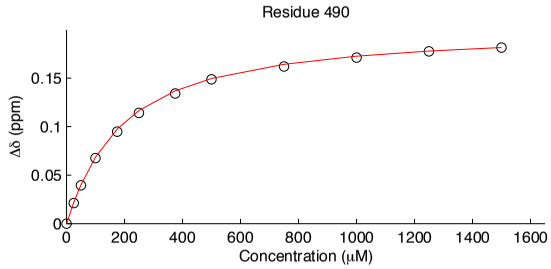


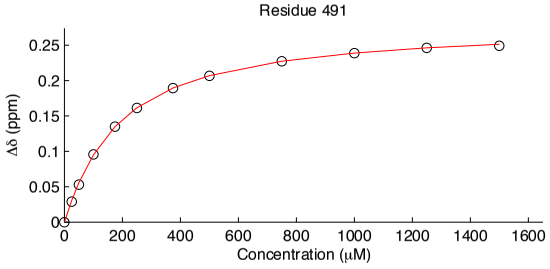
K_D = 252 ± 6μM

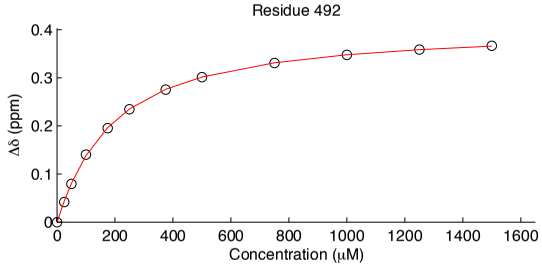
Titration of ^{15}N-RAP74NMR and S430E/S431E/S432E peptide

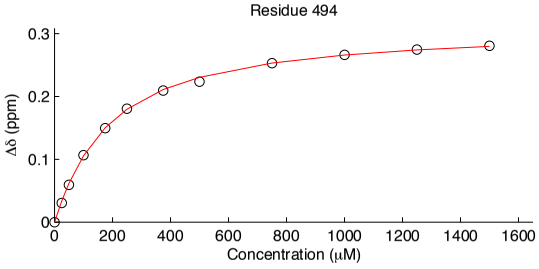


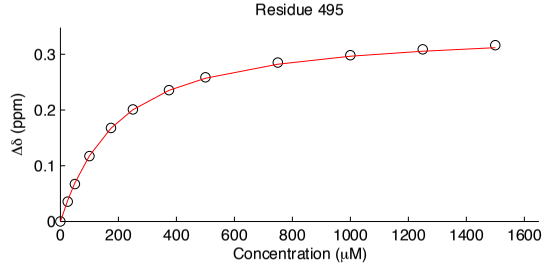


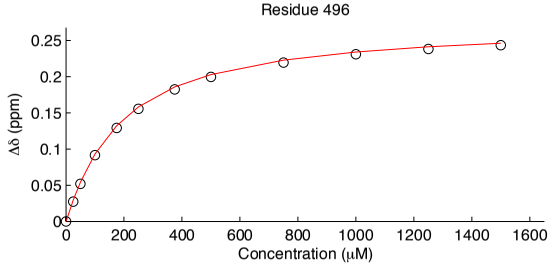


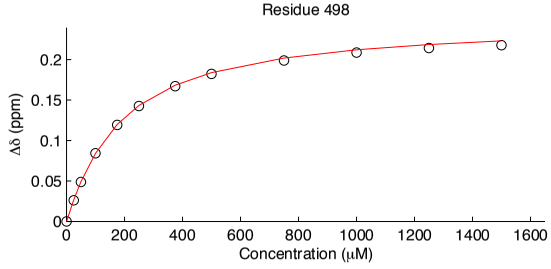


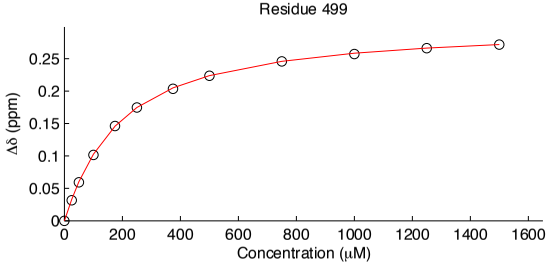


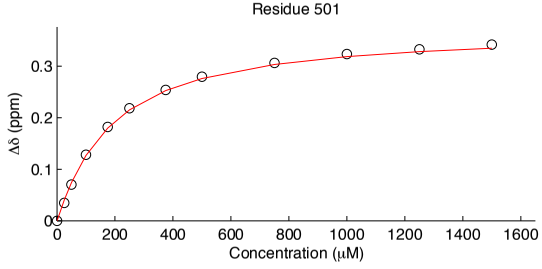






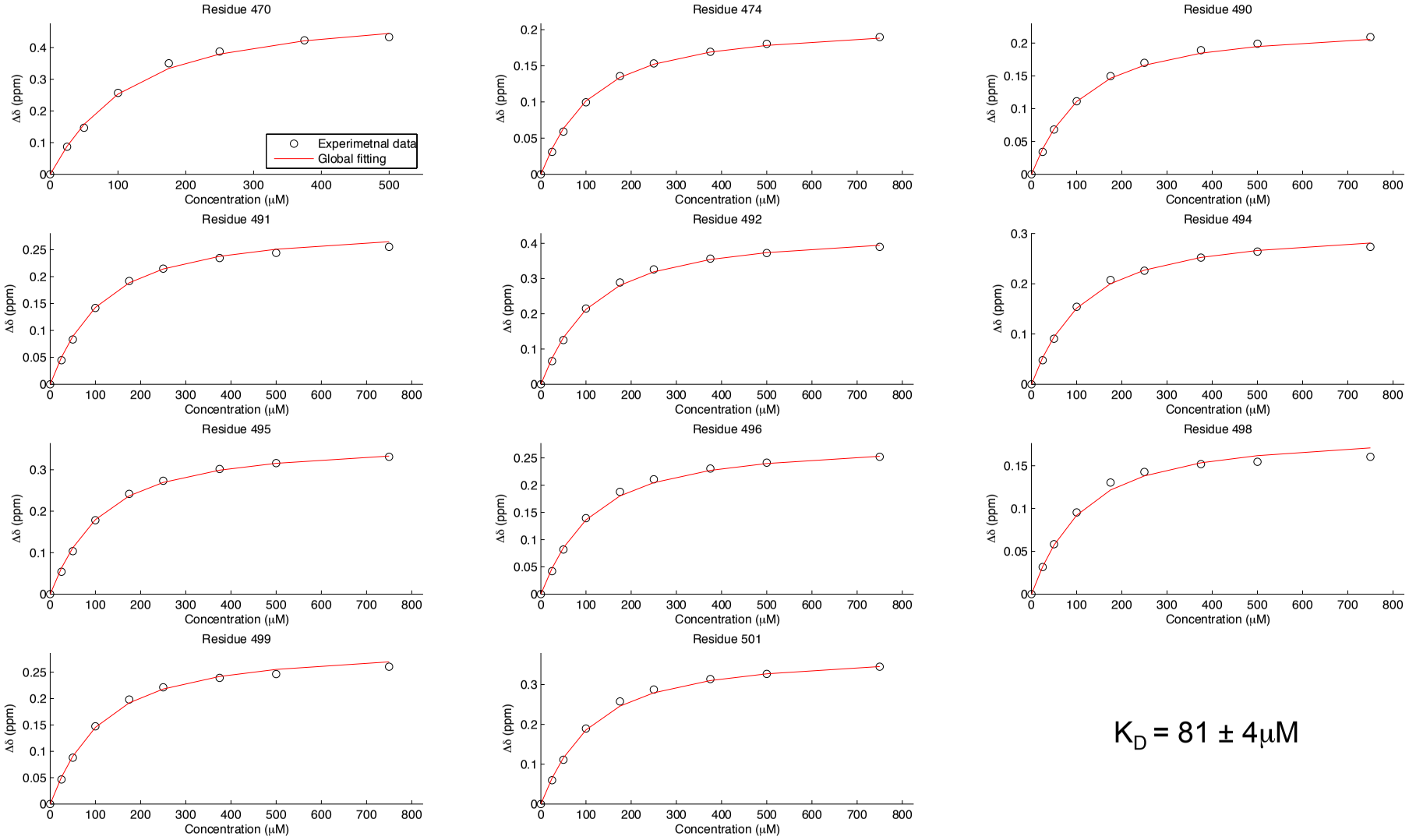






K_D = 166 ± 2μM

Titration of ^{15}N-RAP74NMR and pS430/pS432 peptide



$K_D = 81 \pm 4\mu M$

Bibliography

[1] B J Feldman and D Feldman. The development of androgen-independent prostate cancer. *Nat. Rev. Cancer*, 1(1):34–45, October 2001.

[2] Howard C Shen and Gerhard A Coetzee. The Androgen Receptor: Unlocking the Secrets of Its Unique Transactivation Domain. pages 301–319. Elsevier, 2005.

[3] H E MacLean, G L Warne, and J D Zajac. Localization of functional domains in the androgen receptor. *J. Steroid Biochem. Mol. Biol.*, 62(4):233–242, July 1997.

[4] H Suzuki, T Ueda, T Ichikawa, and H Ito. Androgen receptor involvement in the progression of prostate cancer. *Endocr. Relat. Cancer*, 10(2):209–216, June 2003.

[5] G Jenster, P E de Ruiter, H A van der Korput, G G Kuiper, J Trapman, and A O Brinkmann. Changes in the abundance of androgen receptor isotypes: effects of ligand treatment, glutamine-stretch variation, and mutation of putative phosphorylation sites. *Biochemistry*, 33(47):14064–14072, November 1994.

[6] Z X Zhou, J A Kemppainen, and E M Wilson. Identification of three proline-directed phosphorylation sites in the human androgen receptor. *Molecular Endocrinology*, 9(5):605–615, May 1995.

[7] Martin E van Royen, Wiggert A van Cappellen, Carola de Vos, Adriaan B Houtsmuller, and Jan Trapman. Stepwise androgen receptor dimerization. *Journal of Cell Science*, 125(Pt 8):1970–1979, April 2012.

[8] S R Denmeade, X S Lin, and J T Isaacs. Role of programmed (apoptotic) cell death during the progression and therapy for prostate cancer. *Prostate*, 28(4):251–265, April 1996.

[9] N J McKenna, R B Lanz, and B W O'Malley. Nuclear receptor coregulators: cellular and molecular biology. *Endocrine Reviews*, 20(3):321–344, June 1999.

[10] Himisha Beltran, Tomasz M Beer, Michael A Carducci, Johann de Bono, Martin Gleave, Maha Hussain, William K Kelly, Fred Saad, Cora Sternberg, Scott T Tagawa, and Ian F Tannock. New therapies for castration-resistant prostate cancer: efficacy and safety. *Eur. Urol.*, 60(2):279–290, 2011.

[11] Charles Huggins and Clarence V Hodges. Studies on prostatic cancer. *Cancer Research*, 1:v297, 1941.

[12] Jill A Trendel. The hurdle of antiandrogen drug resistance: drug design strategies. *Expert Opin. Drug Discov.*, 8(12):000–000, November 2013.

[13] Isabel Heidegger, Petra Massoner, Iris E Eder, Andreas Pircher, Renate Pichler, Friedrich Aigner, Jasmin Bektic, Wolfgang Horninger, and Helmut Klocker. Journal of Steroid Biochemistry and Molecular Biology. *J. Steroid Biochem. Mol. Biol.*, 138:248–256, November 2013.

[14] P Ferrari, G Castagnetti, G Ferrari, B Baisi, and A Dotti. Combination treatment versus LHRH alone in advanced prostatic cancer. *Urol. Int.*, 56 Suppl 1:13–17, 1996.

[15] Nicolas Mottet, Joaquim Bellmunt, Michel Bolla, Steven Joniau, Malcolm Mason, Vsevolod Matveev, Hans-Peter Schmid, Theo Van der Kwast, Thomas Wiegel, Filiberto Zattoni, and Axel Heidenreich. EAU Guidelines on Prostate Cancer. Part II: Treatment of Advanced, Relapsing, and Castration-Resistant Prostate Cancer. *European Urology*, 59(4):572–583, April 2011.

[16] Joanne Edwards and John M S Bartlett. The androgen receptor and signal-transduction pathways in hormone-refractory prostate cancer. Part 2: Androgen-receptor cofactors and bypass pathways. *BJU International*, 95(9):1327–1335, June 2005.

[17] Jorge A Garcia and Brian I Rini. Castration-resistant prostate cancer: many treatments, many options, many challenges ahead. *Cancer*, 118(10):2583–2593, May 2012.

[18] Joanne Edwards and John M S Bartlett. The androgen receptor and signal-transduction pathways in hormone-refractory prostate cancer. Part 1: Modifications to the androgen receptor. *BJU International*, 95(9):1320–1326, June 2005.

[19] W K Kelly and H I Scher. Prostate specific antigen decline after antiandrogen withdrawal: the flutamide withdrawal syndrome. *J. Urol.*, 149(3):607–609, 1993.

[20] E J Small and S Srinivas. The antiandrogen withdrawal syndrome. Experience in a large cohort of unselected patients with advanced prostate cancer. *Cancer*, 76(8):1428–1434, 1995.

[21] M E Taplin, G J Bubley, Y J Ko, E J Small, M Upton, B Rajeshkumar, and S P Balk. Selection for androgen receptor mutations in prostate cancers treated with androgen antagonist. *Cancer Res.*, 59(11):2511–2515, 1999.

[22] L G Gomella, M Ismail, and F E Nathan. Antiandrogen withdrawal syndrome with nilutamide. *J. Urol.*, 157(4):1366, 1997.

[23] S D Huan, R G Gerridzen, J C Yau, and D J Stewart. Antiandrogen withdrawal syndrome with nilutamide. *Urology*, 49(4):632–634, 1997.

[24] P T Nieh. Withdrawal phenomenon with the antiandrogen casodex. *J. Urol.*, 153(3 Pt 2):1070–2– discussion 1072–3, 1995.

[25] E J Small and P R Carroll. Prostate-specific antigen decline after casodex withdrawal: evidence for an antiandrogen withdrawal syndrome. *Urology*, 43(3):408–410, 1994.

[26] Vincenzo Pagliarulo, Sergio Bracarda, Mario A Eisenberger, Nicolas Mottet, Fritz H Schröder, Cora N Sternberg, and Urs E Studer. Contemporary role of androgen deprivation therapy for prostate cancer. *European Urology*, 61(1):11–25, January 2012.

[27] Ying Ying Sung and Edwin Cheung. Androgen receptor co-regulatory networks in castration-resistant prostate cancer. *Endocr. Relat. Cancer*, 21(1):R1–R11, February 2014.

[28] Amit Bahl, Susan Masson, Alison Birtle, Simon Chowdhury, and Johann de Bono. Second-line treatment options in metastatic castration-resistant prostate cancer: a comparison of key trials with recently approved agents. *Cancer Treatment Reviews*, 40(1):170–177, February 2014.

[29] C Dumontet and B I Sikic. Mechanisms of action of and resistance to antitubulin agents: microtubule dynamics, drug transport, and cell death. *J. Clin. Oncol.*, 17(3):1061–1070, March 1999.

[30] Amanda J O'Neill, Maria Prencipe, Catherine Dowling, Yue Fan, Laoighse Mulrane, William M Gallagher, Darran O'Connor, Robert O'Connor, Aoife Devery, Claire Corcoran, Sweta Rani, Lorraine O'Driscoll, John M Fitzpatrick, and R William G Watson. Characterisation and manipulation of docetaxel resistant prostate cancer cell lines. *Mol. Cancer*, 10:126, 2011.

[31] Joelle El-Amm, Nihar Patel, Ashley Freeman, and Jeanny B Aragon-Ching. Metastatic Castration-Resistant Prostate Cancer: Critical Review of Enzalutamide. *Clin Med Insights Oncol*, 7:235–245, 2013.

[32] Srikala S Sridhar, Stephen J Freedland, Martin E Gleave, Celestia Higano, Peter Mulders, Chris Parker, Oliver Sartor, and Fred Saad. Castration-Resistant Prostate Cancer: From New Pathophysiology to New Treatment. *European Urology*, 65(2):289–299, February 2014.

[33] Medha S Darshan, Matthew S Loftus, Maria Thadani-Mulero, Benjamin P Levy, Daniel Escuin, Xi Kathy Zhou, Ada Gjyrezi, Chantal Chanel-Vos, Ruoqian Shen, Scott T Tagawa, Neil H Bander, David M Nanus, and Paraskevi Giannakakou. Taxane-induced blockade to nuclear accumulation of the androgen receptor predicts clinical responses in metastatic prostate cancer. *Cancer Res.*, 71(18):6019–6029, 2011.

[34] Matthew D Galsky, Argyris Dritselis, Peter Kirkpatrick, and William K Oh. Cabazitaxel. *Nat Rev Drug Discov*, 9(9):677–678, 2010.

[35] Gerhardt Attard, Alison H M Reid, David Olmos, and Johann S de Bono. Antitumor activity with CYP17 blockade indicates that castration-resistant prostate cancer frequently remains hormone driven. *Cancer Res.*, 69(12):4937–4940, 2009.

[36] Alison Hm Reid, Gerhardt Attard, Elaine Barrie, and Johann S de Bono. CYP17 inhibition as a hormonal strategy for prostate cancer. *Nat Clin Pract Urol*, 5(11):610–620, 2008.

[37] R B Montgomery, E A Mostaghel, R Vessella, D L Hess, T F Kalhorn, C S Higano, L D True, and P S Nelson. Maintenance of Intratumoral Androgens in Metastatic Prostate Cancer: A Mechanism for Castration-Resistant Tumor Growth. *Cancer Research*, 68(11):4447–4454, June 2008.

[38] Eleni Efstathiou, Mark Titus, Dimitra Tsavachidou, Vassiliki Tzelepi, Sijin Wen, Anh Hoang, Arturo Molina, Nicole Chieffo, Lisa A Smith, Maria Karlou, Patricia Troncoso, and Christopher J Logothetis. Effects of abiraterone acetate on androgen signaling in castrate-resistant prostate cancer in bone. *Journal of Clinical Oncology*, 30(6):637–643, February 2012.

[39] Elahe A Mostaghel. Abiraterone in the treatment of metastatic castration-resistant prostate cancer. *Cancer Manag Res*, 6:39–51, 2014.

[40] Joaquin Mateo, Alan Smith, Michael Ong, and Johann S Bono. Novel drugs targeting the androgen receptor pathway in prostate cancer. *Cancer Metastasis Rev*, January 2014.

[41] X Y Zhao, P J Malloy, A V Krishnan, S Swami, N M Navone, D M Peehl, and D Feldman. Glucocorticoids can promote androgen-independent growth of prostate cancer cells through a mutated androgen receptor. *Nat Med*, 6(6):703–706, 2000.

[42] Gerhardt Attard, Alison H M Reid, Timothy A Yap, Florence Raynaud, Mitch Dowsett, Sarah Settatree, Mary Barrett, Christopher Parker, Vanessa Martins, Elizabeth Folkerd, Jeremy Clark, Colin S Cooper, Stan B Kaye, David Dearnaley, Gloria Lee, and Johann S de Bono. Phase I clinical trial of a selective inhibitor of CYP17, abiraterone acetate, confirms that castration-resistant prostate cancer commonly remains hormone driven. *J. Clin. Oncol.*, 26(28):4563–4571, 2008.

[43] Charles J Ryan, Matthew R Smith, Lawrence Fong, Jonathan E Rosenberg, Philip Kantoff, Florence Raynaud, Vanessa Martins, Gloria Lee, Thian Kheoh, Jennifer Kim, Arturo Molina, and Eric J Small. Phase I clinical trial of the CYP17 inhibitor abiraterone acetate demonstrating clinical activity in patients with castration-resistant prostate cancer who received prior ketoconazole therapy. *J. Clin. Oncol.*, 28(9):1481–1488, 2010.

[44] Z Culig, A Hobisch, M V Cronauer, A C Cato, A Hittmair, C Radmayr, J Eberle, G Bartsch, and H Klocker. Mutant androgen receptor detected in an advanced-stage prostatic carcinoma is activated by adrenal androgens and progesterone. *Molecular Endocrinology*, 7(12):1541–1550, 1993.

[45] Juliet Richards, Ai Chiin Lim, Colin W Hay, Angela E Taylor, Anna Wingate, Karolina Nowakowska, Carmel Pezaro, Suzanne Carreira, Jane Goodall, Wiebke Arlt, Iain J McEwan, Johann S de Bono, and Gerhardt Attard. Interactions of abiraterone, eplerenone, and prednisolone with wild-type and mutant androgen receptor: a rationale for increasing abiraterone exposure or combining with MDV3100. *Cancer Research*, 72(9):2176–2182, April 2012.

[46] Biswajyoti Sahu, Marko Laakso, Päivi Pihlajamaa, Kristian Ovaska, Ievgenii Sinielnikov, Sampsa Hautaniemi, and Olli A Jänne. FoxA1 specifies unique androgen and glucocorticoid receptor binding events in prostate cancer cells. *Cancer Res.*, 73(5):1570–1580, 2013.

[47] Elahe A Mostaghel, Brett T Marck, Stephen R Plymate, Robert L Vessella, Stephen Balk, Alvin M Matsumoto, Peter S Nelson, and R Bruce Montgomery. Resistance to CYP17A1 inhibition with abiraterone in castration-resistant prostate cancer: induction of steroidogenesis and androgen receptor splice variants. *Clin. Cancer Res.*, 17(18):5913–5925, September 2011.

[48] Chris Tran, Samedy Ouk, Nicola J Clegg, Yu Chen, Philip A Watson, Vivek Arora, John Wongvipat, Peter M Smith-Jones, Dongwon Yoo, Andrew Kwon, Teresa Wasielewska, Derek Welsbie, Charlie Degui Chen, Celestia S Higano, Tomasz M Beer, David T Hung, Howard I Scher, Michael E Jung, and Charles L Sawyers. Development of a second-generation antiandrogen for treatment of advanced prostate cancer. *Science*, 324(5928):787–790, 2009.

[49] James D Joseph, Nhin Lu, Jing Qian, John Sensintaffar, Gang Shao, Dan Brigham, Michael Moon, Edna Chow Maneval, Isan Chen, Beatrice Darimont, and Jeffrey H Hager. A clinically relevant androgen receptor mutation confers resistance to second-generation antiandrogens enzalutamide and ARN-509. *Cancer Discov*, 3(9):1020–1029, September 2013.

[50] Manav Korpal, Joshua M Korn, Xueliang Gao, Daniel P Rakiec, David A Ruddy, Shivang Doshi, Jing Yuan, Steve G Kovats, Sunkyu Kim, Vesselina G Cooke, John E Monahan, Frank Stegmeier, Thomas M Roberts, William R Sellers, Wenlai Zhou, and Ping Zhu. An F876L mutation in androgen receptor confers genetic and phenotypic resistance to MDV3100 (enzalutamide). *Cancer Discov*, 3(9):1030–1043, 2013.

[51] Minna D Balbas, Michael J Evans, David J Hosfield, John Wongvipat, Vivek K Arora, Philip A Watson, Yu Chen, Geoffrey L Greene, Yang Shen, and Charles L Sawyers. Overcoming mutation-based resistance to antiandrogens with rational drug design. *Elife*, 2(0):e00499–e00499, 2013.

[52] Rong Hu, Changxue Lu, Elahe A Mostaghel, Srinivasan Yegnasubramanian, Meltem Gurel, Clare Tannahill, Joanne Edwards, William B Isaacs, Peter S Nelson, Eric Bluemn, Stephen R Plymate, and Jun Luo. Distinct transcriptional programs mediated by the ligand-dependent full-length androgen receptor and its splice variants in castration-resistant prostate cancer. *Cancer Res.*, 72(14):3457–3462, 2012.

[53] Yingming Li, Siu Chiu Chan, Lucas J Brand, Tae Hyun Hwang, Kevin A T Silverstein, and Scott M Dehm. Androgen receptor splice variants mediate enzalutamide resistance in castration-resistant prostate cancer cell lines. *Cancer Research*, 73(2):483–489, January 2013.

[54] Clare McCourt, Pamela Maxwell, Roberta Mazzucchelli, Rodolfo Montironi, Marina Scarpelli, Manuel Salto-Tellez, Joe M O'Sullivan, Daniel B Longley, and David J J Waugh. Elevation of c-FLIP in castrate-resistant prostate cancer antagonizes therapeutic response to androgen receptor-targeted therapy. *Clin. Cancer Res.*, 18(14):3822–3833, 2012.

[55] Amit Bahl. Metastatic castration-resistant prostate cancer. Part 1: the challenges of the disease and its treatment. *Eur J Oncol Nurs*, 17 Suppl 1:S1–6, September 2013.

[56] Hein Van Poppel and Laurence Klotz. Gonadotropin-releasing hormone: an update review of the antagonists versus agonists. *Int J Urol*, 19(7):594–601, 2012.

[57] Noel W Clarke. Landmarks in non-hormonal pharmacological therapies for castration-resistant prostate cancer. *BJU Int.*, 110 Suppl 1(s1):14–22, 2012.

[58] Thomas A Gardner, Bennett D Elzey, and Noah M Hahn. Sipuleucel-T (Provenge) autologous vaccine approved for treatment of men with asymptomatic or minimally symptomatic castrate-resistant metastatic prostate cancer. *Hum Vaccin Immunother*, 8(4):534–539, 2012.

[59] O Harris Ford, Christopher W Gregory, Desok Kim, Andrew B Smitherman, and James L Mohler. Androgen receptor gene amplification and protein expression in recurrent prostate cancer. *J. Urol.*, 170(5):1817–1821, 2003.

[60] M J Linja, K J Savinainen, O R Saramäki, T L Tammela, R L Vessella, and T Visakorpi. Amplification and overexpression of androgen receptor gene in hormone-refractory prostate cancer. *Cancer Res.*, 61(9):3550–3555, 2001.

[61] P A Koivisto and H J Helin. Androgen receptor gene amplification increases tissue PSA protein expression in hormone-refractory prostate carcinoma. *J. Pathol.*, 189(2):219–223, October 1999.

[62] Charlie D Chen, Derek S Welsbie, Chris Tran, Sung Hee Baek, Randy Chen, Robert Vessella, Michael G Rosenfeld, and Charles L Sawyers. Molecular determinants of resistance to antiandrogen therapy. *Nat Med*, 10(1):33–39, January 2004.

[63] J Edwards, N S Krishna, K M Grigor, and J M S Bartlett. Androgen receptor gene amplification and protein expression in hormone refractory prostate cancer. *Br J Cancer*, 89(3):552–556, August 2003.

[64] T Visakorpi, E Hyytinen, P Koivisto, M Tanner, R Keinänen, C Palmberg, A Palotie, T Tammela, J Isola, and O P Kallioniemi. In vivo amplification of the androgen receptor gene and progression of human prostate cancer. *Nature Genetics*, 9(4):401–406, 1995.

[65] Kati K Waltering, Merja A Helenius, Biswajyoti Sahu, Visa Manni, Marika J Linja, Olli A Jänne, and Tapio Visakorpi. Increased expression of androgen

receptor sensitizes prostate cancer cells to low levels of androgens. *Cancer Res.*, 69(20):8141–8149, 2009.

[66] Jeff Holzbeierlein, Priti Lal, Eva LaTulippe, Alex Smith, Jaya Satagopan, Liying Zhang, Charles Ryan, Steve Smith, Howard Scher, Peter Scardino, Victor Reuter, and William L Gerald. Gene expression analysis of human prostate carcinoma during hormonal therapy identifies androgen-responsive genes and mechanisms of therapy resistance. *Am. J. Pathol.*, 164(1):217–227, 2004.

[67] N Mitsiades. A Road Map to Comprehensive Androgen Receptor Axis Targeting for Castration-Resistant Prostate Cancer. *Cancer Research*, 73(15):4599–4605, July 2013.

[68] Alison Egan, Yan Dong, Haitao Zhang, Yanfeng Qi, Steven P Balk, and Oliver Sartor. Cancer Treatment Reviews. *Cancer Treatment Reviews*, 40(3):426–433, April 2014.

[69] Nicholas Mitsiades, Clifford C Sung, Nikolaus Schultz, Daniel C Danila, Bin He, Vijay Kumar Eedunuri, Martin Fleisher, Chris Sander, Charles L Sawyers, and Howard I Scher. Distinct patterns of dysregulated expression of enzymes involved in androgen synthesis and metabolism in metastatic prostate cancer tumors. *Cancer Research*, 72(23):6142–6152, December 2012.

[70] Jonathan L Wright, Erika M Kwon, Elaine A Ostrander, R Bruce Montgomery, Daniel W Lin, Robert Vessella, Janet L Stanford, and Elahe A Mostaghel. Expression of SLCO transport genes in castration-resistant prostate cancer and impact of genetic variation in SLCO1B3 and SLCO2B1 on prostate cancer outcomes. *Cancer Epidemiol. Biomarkers Prev.*, 20(4):619–627, 2011.

[71] Bruce Gottlieb, Lenore K Beitel, Abbesha Nadarajah, Miltiadis Paliouras, and Mark Trifiro. The androgen receptor gene mutations database: 2012 update. *Hum. Mutat.*, 33(5):887–894, May 2012.

[72] Shahriar Koochekpour. Androgen receptor signaling and mutations in prostate cancer. *Nature Publishing Group*, 12(5):639–657, August 2010.

[73] M E Taplin, G J Bubley, T D Shuster, M E Frantz, A E Spooner, G K Ogata, H N Keer, and S P Balk. Mutation of the androgen-receptor gene in metastatic androgen-independent prostate cancer. *N. Engl. J. Med.*, 332(21):1393–1398, May 1995.

[74] J Veldscholte, C A Berrevoets, C Ris-Stalpers, G G Kuiper, G Jenster, J Trapman, A O Brinkmann, and E Mulder. The androgen receptor in LNCaP cells

contains a mutation in the ligand binding domain which affects steroid binding characteristics and response to antiandrogens. *J. Steroid Biochem. Mol. Biol.*, 41(3-8):665–669, 1992.

[75] Jos Veldscholte, Marleen M Voorhorst-Ogink, Joan Bolt-de Vries, Henri C J van Rooij, Jan Trapman, and Eppo Mulder. Unusual specificity of the androgen receptor in the human prostate tumor cell line LNCaP: High affinity for progestagenic and estrogenic steroids. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1052(1):187–194, 1990.

[76] Karine Steketee, Leon Timmerman, Angelique C J Ziel-van der Made, Paul Doesburg, Albert O Brinkmann, and Jan Trapman. Broadened ligand responsiveness of androgen receptor mutants obtained by random amino acid substitution of H874 and mutation hot spot T877 in prostate cancer. *Int. J. Cancer*, 100(3):309–317, 2002.

[77] S McDonald, L Brive, D B Agus, H I Scher, and K R Ely. Ligand responsiveness in human prostate cancer: structural analysis of mutant androgen receptors from LNCaP and CWR22 tumors. *Cancer Res.*, 60(9):2317–2322, 2000.

[78] Mara P Steinkamp, Orla A O'Mahony, Michele Brogley, Haniya Rehman, Elizabeth W Lapensee, Saravana Dhanasekaran, Matthias D Hofer, Rainer Kuefer, Arul Chinnaiyan, Mark A Rubin, Kenneth J Pienta, and Diane M Robins. Treatment-dependent androgen receptor mutations in prostate cancer exploit multiple mechanisms to evade therapy. *Cancer Research*, 69(10):4434–4442, May 2009.

[79] M J Wallén, M Linja, K Kaartinen, J Schleutker, and T Visakorpi. Androgen receptor gene mutations in hormone-refractory prostate cancer. *J. Pathol.*, 189(4):559–563, 1999.

[80] B J Patel, A J Pantuck, A Zisman, K H Tsui, S H Paik, R Caliliw, S Sheriff, L Wu, J B deKernion, C L Tso, and A S Belldegrun. CL1-GFP: an androgen independent metastatic tumor model for prostate cancer. *J. Urol.*, 164(4):1420–1425, 2000.

[81] H Thakkar, X F Chen, F Tyan, S Gim, and H Robinson. Prosurvival function of Akt/protein kinase B in prostate cancer cells. Relationship with TRAIL resistance. *Journal of Biological Chemistry*, 2002.

[82] Colin W Hay and Iain J McEwan. The impact of point mutations in the human androgen receptor: classification of mutations on the basis of transcriptional activity. *PLoS ONE*, 7(3):e32514, 2012.

[83] Daniel Gioeli and Bryce M Paschal. Molecular and Cellular Endocrinology. *Molecular and Cellular Endocrinology*, 352(1-2):70–78, April 2012.

[84] Liesbeth Clinckemalie, Dirk Vanderschueren, Steven Boonen, and Frank Claessens. Molecular and Cellular Endocrinology. *Molecular and Cellular Endocrinology*, 358(1):1–8, July 2012.

[85] Kelly Coffey and Craig N Robson. Regulation of the androgen receptor by post-translational modifications. *J. Endocrinol.*, 215(2):221–237, 2012.

[86] H Poukka, U Karvonen, O A Janne, and J J Palvimo. Covalent modification of the androgen receptor by small ubiquitin-like modifier 1 (SUMO-1). *Proc. Natl. Acad. Sci. U.S.A.*, 97(26):14145–14150, December 2000.

[87] Maofu Fu, Mahadev Rao, Chenguang Wang, Toshiyuki Sakamaki, Jian Wang, Dolores Di Vizio, Xueping Zhang, Chris Albanese, Steven Balk, Chawnshang Chang, Saijun Fan, Eliot Rosen, Jorma J Palvimo, Olli A Jänne, Selen Muratoglu, Maria Laura Avantaggiati, and Richard G Pestell. Acetylation of androgen receptor enhances coactivator binding and promotes prostate cancer cell growth. *Molecular and Cellular Biology*, 23(23):8563–8575, December 2003.

[88] Luke Gaughan, Ian R Logan, Susan Cook, David E Neal, and Craig N Robson. Tip60 and histone deacetylase 1 regulate androgen receptor activity through changes to the acetylation status of the receptor. *Journal of Biological Chemistry*, 277(29):25904–25913, May 2002.

[89] Kristin R Lamont and Donald J Tindall. Minireview: Alternative activation pathways for the androgen receptor in prostate cancer. *Mol. Endocrinol.*, 25(6):897–907, June 2011.

[90] Irina U Agoulnik and Nancy L Weigel. Coactivator selective regulation of androgen receptor activity. *Steroids*, 74(8):669–674, 2009.

[91] Xin Yuan and Steven P Balk. Mechanisms mediating androgen receptor reactivation after castration. *Urol. Oncol.*, 27(1):36–41, 2009.

[92] Nima Sharifi. Mechanisms of androgen receptor activation in castration-resistant prostate cancer. *Endocrinology*, 154(11):4010–4017, November 2013.

[93] Hannelore V Heemers and Donald J Tindall. Androgen receptor (AR) coregulators: a diversity of functions converging on and regulating the AR transcriptional complex. *Endocrine Reviews*, 28(7):778–808, December 2007.

[94] Barbara Comuzzi, Constanze Nemes, Stefan Schmidt, Zerina Jasarevic, Michele Lodde, Armin Pycha, Georg Bartsch, Felix Offner, Zoran Culig, and Alfred Hobisch. The androgen receptor co-activator CBP is up-regulated following androgen withdrawal and is highly expressed in advanced prostate cancer. *J. Pathol.*, 204(2):159–166, 2004.

[95] Barry S Taylor, Nikolaus Schultz, Haley Hieronymus, Anuradha Gopalan, Yonghong Xiao, Brett S Carver, Vivek K Arora, Poorvi Kaushik, Ethan Cerami, Boris Reva, Yevgeniy Antipin, Nicholas Mitsiades, Thomas Landers, Igor Dolgalev, John E Major, Manda Wilson, Nicholas D Socci, Alex E Lash, Adriana Heguy, James A Eastham, Howard I Scher, Victor E Reuter, Peter T Scardino, Chris Sander, Charles L Sawyers, and William L Gerald. Integrative genomic profiling of human prostate cancer. *Cancer Cell*, 18(1):11–22, 2010.

[96] Marika J Linja, Kati P Porkka, Zhikang Kang, Kimmo J Savinainen, Olli A Jänne, Teuvo L J Tammela, Robert L Vessella, Jorma J Palvimo, and Tapio Visakorpi. Expression of androgen receptor coregulators in prostate cancer. *Clin. Cancer Res.*, 10(3):1032–1040, February 2004.

[97] Maggie C Louie, Hong Qiong Yang, Ai-Hong Ma, Wei Xu, June X Zou, Hsing-Jien Kung, and Hong-Wu Chen. Androgen-induced recruitment of RNA polymerase II to a nuclear receptor-p160 coactivator complex. *Proc. Natl. Acad. Sci. U.S.A.*, 100(5):2226–2230, March 2003.

[98] S Yeh and C Chang. Cloning and characterization of a specific coactivator, ARA70, for the androgen receptor in human prostate cells. *Proc. Natl. Acad. Sci. U.S.A.*, 93(11):5517–5521, 1996.

[99] Z Culig, H Klocker, G Bartsch, and A Hobisch. Androgen receptors in prostate cancer. *Endocr. Relat. Cancer*, 9(3):155–170, 2002.

[100] Cynthia A Heinlein and Chawnshang Chang. Androgen receptor in prostate cancer. *Endocrine Reviews*, 25(2):276–308, April 2004.

[101] S Yeh, H C Chang, H Miyamoto, H Takatera, M Rahman, H Y Kang, T H Thin, H K Lin, and C Chang. Differential induction of the androgen receptor transcriptional activity by selective androgen receptor coactivators. *Keio J Med*, 48(2):87–92, 1999.

[102] V M Gelfanov, G S Burgess, S Litz-Jackson, A J King, M S Marshall, H Nakshatri, and H S Boswell. Transformation of interleukin-3-dependent cells without participation of Stat5/bcl-xL: cooperation of akt with raf/erk leads to p65 nuclear

factor kappaB-mediated antiapoptosis involving c-IAP2. *Blood*, 98(8):2508–2517, 2001.

[103] H Miyamoto, S Yeh, G Wilding, and C Chang. Promotion of agonist activity of antiandrogens by the androgen receptor coactivator, ARA70, in human prostate cancer DU145 cells. *Proc. Natl. Acad. Sci. U.S.A.*, 95(13):7379–7384, 1998.

[104] H Miyamoto, S Yeh, H Lardy, E Messing, and C Chang. Delta5-androstenediol is a natural hormone with androgenic activity in human prostate cancer cells. *Proc. Natl. Acad. Sci. U.S.A.*, 95(19):11083–11088, 1998.

[105] Kalipso Halkidou, Vincent J Gnanapragasam, Piyush B Mehta, Ian R Logan, Mark E Brady, Susan Cook, Hing Y Leung, David E Neal, and Craig N Robson. Expression of Tip60, an androgen receptor coactivator, and its role in prostate cancer development. *Oncogene*, 22(16):2466–2477, 2003.

[106] Meng-Lei Zhu and Natasha Kyprianou. Androgen receptor and growth factor signaling cross-talk in prostate cancer cells. *Endocr. Relat. Cancer*, 15(4):841–849, 2008.

[107] Jennifer D Wu, Kathy Haugk, Libby Woodke, Peter Nelson, Ilsa Coleman, and Stephen R Plymate. Interaction of IGF signaling and the androgen receptor in prostate cancer progression. *J. Cell. Biochem.*, 99(2):392–401, 2006.

[108] Z Culig, A Hobisch, M V Cronauer, C Radmayr, J Trapman, A Hittmair, G Bartsch, and H Klocker. Androgen receptor activation in prostatic tumor cell lines by insulin-like growth factor-I, keratinocyte growth factor, and epidermal growth factor. *Cancer Res.*, 54(20):5474–5478, 1994.

[109] Joanne Edwards, N Sarath Krishna, Caroline J Witton, and John M S Bartlett. Gene amplifications associated with the development of hormone-resistant prostate cancer. *Clin. Cancer Res.*, 9(14):5271–5281, 2003.

[110] Robert E Bakin, Daniel Gioeli, Robert A Sikes, Eric A Bissonette, and Michael J Weber. Constitutive activation of the Ras/mitogen-activated protein kinase signaling pathway promotes androgen hypersensitivity in LNCaP prostate cancer cells. *Cancer Research*, 63(8):1981–1989, April 2003.

[111] Cécile Rochette-Egly. Nuclear receptors: integration of multiple signalling pathways through phosphorylation. *Cell. Signal.*, 15(4):355–366, April 2003.

[112] S M Powell, V Christiaens, D Voulgaraki, J Waxman, F Claessens, and C L Bevan. Mechanisms of androgen receptor signalling via steroid receptor coactivator-1 in prostate. *Endocr. Relat. Cancer*, 11(1):117–130, 2004.

[113] Takeshi Ueda, Nicholas Bruchovsky, and Marianne D Sadar. Activation of the androgen receptor N-terminal domain by interleukin-6 via MAPK and STAT3 signal transduction pathways. *J. Biol. Chem.*, 277(9):7076–7085, March 2002.

[114] N Sato, M D Sadar, N Bruchovsky, F Saatcioglu, P S Rennie, S Sato, P H Lange, and M E Gleave. Androgenic induction of prostate-specific antigen gene is repressed by protein-protein interaction between the androgen receptor and AP-1/c-Jun in the human prostate cancer cell line LNCaP. *J. Biol. Chem.*, 272(28):17485–17494, July 1997.

[115] Hui-Kuan Lin, Yueh-Chiang Hu, Lin Yang, Saleh Altuwaijri, Yen-Ta Chen, Hong-Yo Kang, and Chawnshang Chang. Suppression versus induction of androgen receptor functions by the phosphatidylinositol 3-kinase/Akt pathway in prostate cancer LNCaP cells with different passage numbers. *J. Biol. Chem.*, 278(51):50902–50907, December 2003.

[116] M T Spiotto and T D Chung. STAT3 mediates IL-6-induced growth inhibition in the human prostate cancer cell line LNCaP. *Prostate*, 42(2):88–98, 2000.

[117] M M Ittmann. Chromosome 10 alterations in prostate adenocarcinoma (review). *Oncol. Rep.*, 5(6):1329–1335, 1998.

[118] Paramita M Ghosh, Shazli Malik, Roble Bedolla, and Jeffrey I Kreisberg. Akt in prostate cancer: possible role in androgen-independence. *Curr. Drug Metab.*, 4(6):487–496, 2003.

[119] N M Corcoran and A J Costello. Interleukin-6: minor player or starring role in the development of hormone-refractory prostate cancer? *BJU International*, 91(6):545–553, April 2003.

[120] Linda B Mora, Ralf Buettner, John Seigne, Jose Diaz, Nazeel Ahmad, Roy Garcia, Tammy Bowman, Robert Falcone, Rita Fairclough, Alan Cantor, Carlos Muro-Cacho, Sandy Livingston, James Karras, Julio Pow-Sang, and Richard Jove. Constitutive activation of Stat3 in human prostate tumors and cell lines: direct inhibition of Stat3 signaling induces apoptosis of prostate cancer cells. *Cancer Res.*, 62(22):6659–6666, 2002.

[121] T Chen, L H Wang, and W L Farrar. Interleukin 6 activates androgen receptor-mediated gene expression through a signal transducer and activator of transcription 3-dependent pathway in LNCaP prostate cancer cells. *Cancer Res.*, 60(8):2132–2135, 2000.

[122] J Trachtenberg and G Blackledge. Looking to the future: advances in the management of hormone-refractory prostate cancer. *European Urology Supplements*, 2002.

[123] Natalie Blaszczyk, Bassam A Masri, Nasrin R Mawji, Takeshi Ueda, Gavan McAlinden, Clive P Duncan, Nicholas Bruchovsky, Hans-Udo Schweikert, Doris Schnabel, Edward C Jones, and Marianne D Sadar. Osteoblast-derived factors induce androgen-independent proliferation and expression of prostate-specific antigen in human prostate cancer cells. *Clin. Cancer Res.*, 10(5):1860–1869, March 2004.

[124] Zoran Culig and Georg Bartsch. Androgen axis in prostate cancer. *J. Cell. Biochem.*, 99(2):373–381, 2006.

[125] Ian G Mills. PERSPECTIVES. *Nature Publishing Group*, 14(3):187–198, March 2014.

[126] Kexin Xu, Hermela Shimelis, Douglas E Linn, Richeng Jiang, Xi Yang, Feng Sun, Zhiyong Guo, Hege Chen, Wei Li, Hegang Chen, Xiangtian Kong, Jonathan Melamed, Shengyun Fang, Zhen Xiao, Timothy D Veenstra, and Yun Qiu. Regulation of androgen receptor transcriptional activity and specificity by RNF6-induced ubiquitination. *Cancer Cell*, 15(4):270–282, 2009.

[127] Scott M Dehm, Lucy J Schmidt, Hannelore V Heemers, Robert L Vessella, and Donald J Tindall. Splicing of a novel androgen receptor exon generates a constitutively active androgen receptor that mediates prostate cancer therapy resistance. *Cancer Research*, 68(13):5469–5477, July 2008.

[128] Zhiyong Guo, Xi Yang, Feng Sun, Richeng Jiang, Douglas E Linn, Hege Chen, Hegang Chen, Xiangtian Kong, Jonathan Melamed, Clifford G Tepper, Hsing-Jien Kung, Angela M H Brodie, Joanne Edwards, and Yun Qiu. A novel androgen receptor splice variant is up-regulated during prostate cancer progression and promotes androgen depletion-resistant growth. *Cancer Res.*, 69(6):2305–2313, 2009.

[129] Emma Hörnberg, Erik Bovinder Ylitalo, Sead Crnalic, Henrik Antti, Pär Stattin, Anders Widmark, Anders Bergh, and Pernilla Wikström. Expression of androgen receptor splice variants in prostate cancer bone metastases is associated with castration-resistance and short survival. *PLoS ONE*, 6(4):e19059, 2011.

[130] Rong Hu, Thomas A Dunn, Shuanzeng Wei, Sumit Isharwal, Robert W Veltri, Elizabeth Humphreys, Misop Han, Alan W Partin, Robert L Vessella, William B

Isaacs, G Steven Bova, and Jun Luo. Ligand-independent androgen receptor variants derived from splicing of cryptic exons signify hormone-refractory prostate cancer. *Cancer Res.*, 69(1):16–22, 2009.

[131] Shihua Sun, Cynthia C T Sprenger, Robert L Vessella, Kathleen Haugk, Kathryn Soriano, Elahe A Mostaghel, Stephanie T Page, Ilsa M Coleman, Holly M Nguyen, Huiying Sun, Peter S Nelson, and Stephen R Plymate. Castration resistance in human prostate cancer is conferred by a frequently occurring androgen receptor splice variant. *J. Clin. Invest.*, 120(8):2715–2730, August 2010.

[132] Xiaotun Zhang, Colm Morrissey, Shihua Sun, Melanie Ketchandji, Peter S Nelson, Lawrence D True, Funda Vakar-Lopez, Robert L Vessella, and Stephen R Plymate. Androgen receptor variants occur frequently in castration resistant prostate cancer metastases. *PLoS ONE*, 6(11):e27970, 2011.

[133] Z Yu, S Chen, A G Sowalsky, O S Voznesensky, E A Mostaghel, P S Nelson, C Cai, and S P Balk. Rapid Induction of Androgen Receptor Splice Variants by Androgen Deprivation in Prostate Cancer. *Clinical Cancer Research*, January 2014.

[134] Raymond J Andersen, Nasrin R Mawji, Jun Wang, Gang Wang, Simon Haile, Jae-Kyung Myung, Kate Watt, Teresa Tam, Yu Chi Yang, Carmen A Bañuelos, David E Williams, Iain J McEwan, Yuzhou Wang, and Marianne D Sadar. Regression of castrate-recurrent prostate cancer by a small-molecule inhibitor of the amino-terminus domain of the androgen receptor. *Cancer Cell*, 17(6):535–546, June 2010.

[135] Bo Cao, Xichun Liu, Jing Li, Shuang Liu, Yanfeng Qi, Zhenggang Xiong, Allen Zhang, Thomas Wiese, Xueqi Fu, Jingkai Gu, Paul S Rennie, Oliver Sartor, Benjamin R Lee, Clement Ip, Lijuan Zhao, Haitao Zhang, and Yan Dong. 20(S)-protopanaxadiol-aglycone downregulation of the full-length and splice variants of androgen receptor. *Int. J. Cancer*, 132(6):1277–1287, March 2013.

[136] Jing Li, Bo Cao, Xichun Liu, Xueqi Fu, Zhenggang Xiong, Li Chen, Oliver Sartor, Yan Dong, and Haitao Zhang. Berberine suppresses androgen receptor signaling in prostate cancer. *Mol. Cancer Ther.*, 10(8):1346–1356, August 2011.

[137] Xuesen Li, Zhongbo Liu, Xia Xu, Christopher A Blair, Zheng Sun, Jun Xie, Michael B Lilly, and Xiaolin Zi. Kava components down-regulate expression of AR and AR splice variants and reduce growth in patient-derived prostate cancer xenografts in mice. *PLoS ONE*, 7(2):e31213, 2012.

[138] Tetsuo Mashima, Sachiko Okabe, and Hiroyuki Seimiya. Pharmacological targeting of constitutively active truncated androgen receptor by nigericin and suppression of hormone-refractory prostate cancer cell growth. *Mol. Pharmacol.*, 78(5):846–854, November 2010.

[139] Marianne D Sadar, David E Williams, Nasrin R Mawji, Brian O Patrick, Thamrin Wikanta, Ekowati Chasanah, Hari Eko Irianto, Rob Van Soest, and Raymond J Andersen. Sintokamides A to E, chlorinated peptides from the sponge Dysidea sp. that inhibit transactivation of the N-terminus of the androgen receptor in prostate cancer cells. *Org. Lett.*, 10(21):4947–4950, November 2008.

[140] Shinichi Yamashita, Kuo-Pao Lai, Kun-Lung Chuang, Defeng Xu, Hiroshi Miyamoto, Tatsuo Tochigi, See-Tong Pang, Lei Li, Yoichi Arai, and Hsing-Jien Kung. ASC-J9 suppresses castration-resistant prostate cancer growth through degradation of full-length and splice variant androgen receptors. *Neoplasia (New York, NY)*, 14(1):74, 2012.

[141] W C Bell, R B Myers, T O Hosein, D K Oelschlager, and W E Grizzle. The response of extracellular signal-regulated kinase (ERK) to androgen-induced proliferation in the androgen-sensitive prostate cancer cell line, LNCaP. *Biotech Histochem*, 78(1):11–16, 2003.

[142] Zheng Fu, Peter C Smith, Lizhi Zhang, Mark A Rubin, Rodney L Dunn, Zhi Yao, and Evan T Keller. Effects of raf kinase inhibitor protein expression on suppression of prostate cancer metastasis. *J. Natl. Cancer Inst.*, 95(12):878–889, 2003.

[143] Michelle A Schultz, Sharika S Hagan, Amrita Datta, Yiguo Zhang, Michael L Freeman, Suresh C Sikka, Asim B Abdel-Mageed, and Debasis Mondal. Nrf1 and Nrf2 Transcription Factors Regulate Androgen Receptor Transactivation in Prostate Cancer Cells. *PLoS ONE*, 9(1):e87204, January 2014.

[144] Eleni Efstathiou, Maria Karlou, Sijin Wen, Anh Hoang, Curtis A Pettaway, Louis L Pisters, Sankar Maity, Patricia Troncoso, and Christopher J Logothetis. Integrated Hedgehog signaling is induced following castration in human and murine prostate cancers. *Prostate*, 73(2):153–161, 2013.

[145] R Ferraldeschi, C Pezaro, V Karavasilis, and J de Bono. Abiraterone and novel antiandrogens: overcoming castration resistance in prostate cancer. *Annu. Rev. Med.*, 64(1):1–13, 2013.

[146] Guido Jenster, Hetty AGM van der Korput, Cor van Vroonhoven, Theo H van der Kwast, Jan Trapman, and Albert O Brinkmann. Domains of the human androgen

receptor involved in steroid binding, transcriptional activation, and subcellular localization. *Molecular Endocrinology*, 5(10):1396–1404, 1991.

[147] Steven N Quayle, Nasrin R Mawji, Jun Wang, and Marianne D Sadar. Androgen receptor decoy molecules block the growth of prostate cancer. *Proc. Natl. Acad. Sci. U.S.A.*, 104(4):1331–1336, January 2007.

[148] Gang Wang and Marianne D Sadar. Amino-terminus domain of the androgen receptor as a molecular target to prevent the hormonal progression of prostate cancer. *J. Cell. Biochem.*, 98(1):36–53, 2006.

[149] G Jenster, H A van der Korput, J Trapman, and A O Brinkmann. Identification of two transcription activation units in the N-terminal domain of the human androgen receptor. *J. Biol. Chem.*, 270(13):7341–7346, March 1995.

[150] Derek N Lavery and Iain J McEwan. Structure and function of steroid receptor AF1 transactivation domains: induction of active conformations. *Biochem. J.*, 391(3):449, November 2005.

[151] H Jane Dyson and Peter E Wright. Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol*, 6(3):197–208, March 2005.

[152] Peter Tompa. Intrinsically disordered proteins: a 10-year recap. *Trends Biochem. Sci.*, 37(12):509–516, December 2012.

[153] Monika Fuxreiter, Peter Tompa, István Simon, Vladimir N Uversky, Jeffrey C Hansen, and Francisco J Asturias. Malleable machines take shape in eukaryotic transcriptional regulation. *Nat Chem Biol*, 4(12):728–737, December 2008.

[154] V J Hilser and E B Thompson. Structural Dynamics, Intrinsic Disorder, and Allostery in Nuclear Receptors as Transcription Factors. *Journal of Biological Chemistry*, 286(46):39675–39682, November 2011.

[155] Christine Helsen and Frank Claessens. Looking at nuclear receptors from a new angle. *Molecular and Cellular Endocrinology*, 382(1):97–106, January 2014.

[156] Iain J McEwan. Intrinsic disorder in the androgen receptor: identification, characterisation and drugability. *Mol. BioSyst.*, 8(1):82–90, January 2012.

[157] L Callewaert. Interplay between Two Hormone-Independent Activation Domains in the Androgen Receptor. *Cancer Research*, 66(1):543–553, January 2006.

[158] Dennis J van de Wijngaart, Hendrikus Jan Dubbink, Martin E van Royen, Jan Trapman, and Guido Jenster. Molecular and Cellular Endocrinology. *Molecular and Cellular Endocrinology*, 352(1-2):57–69, April 2012.

[159] Labros G Meimetis, David E Williams, Nasrin R Mawji, Carmen A Bañuelos, Aaron A Lal, Jacob J Park, Amy H Tien, Javier Garcia Fernandez, Nicole J de Voogd, Marianne D Sadar, and Raymond J Andersen. Niphatenones, glycerol ethers from the sponge Niphates digitalis block androgen receptor transcriptional activity in prostate cancer cells: structure elucidation, synthesis, and biological activity. *J. Med. Chem.*, 55(1):503–514, January 2012.

[160] Jae-Kyung Myung, Carmen A Bañuelos, Javier Garcia Fernandez, Nasrin R Mawji, Jun Wang, Amy H Tien, Yu Chi Yang, Iran Tavakoli, Simon Haile, Kate Watt, Iain J McEwan, Stephen Plymate, Raymond J Andersen, and Marianne D Sadar. An androgen receptor N-terminal domain antagonist for treating prostate cancer. *J. Clin. Invest.*, 123(7):2948–2960, July 2013.

[161] Edward P Gelmann. Molecular biology of the androgen receptor. *J. Clin. Oncol.*, 20(13):3001–3015, July 2002.

[162] C A Quigley, A De Bellis, K B Marschke, M K el Awady, E M Wilson, and F S French. Androgen receptor defects: historical, clinical, and molecular perspectives. *Endocrine Reviews*, 16(3):271–321, June 1995.

[163] D B Lubahn, D R Joseph, M Sar, J Tan, H N Higgs, R E Larson, F S French, and E M Wilson. The human androgen receptor: complementary deoxyribonucleic acid cloning, sequence analysis and gene expression in prostate. *Molecular Endocrinology*, 2(12):1265–1275, December 1988.

[164] C J Brown, S J Goss, D B Lubahn, D R Joseph, E M Wilson, F S French, and H F Willard. Androgen receptor locus on the human X chromosome: regional localization to Xq11-12 and description of a DNA polymorphism. *Am. J. Hum. Genet.*, 44(2):264–269, February 1989.

[165] C S Chang, J Kokontis, and S T Liao. Molecular cloning of human and rat complementary DNA encoding androgen receptors. *Science*, 240(4850):324–326, April 1988.

[166] D B Lubahn, D R Joseph, P M Sullivan, H F Willard, F S French, and E M Wilson. Cloning of human androgen receptor complementary DNA and localization to the X chromosome. *Science*, 240(4850):327–330, April 1988.

[167] J Trapman, P Klaassen, G G Kuiper, J A van der Korput, P W Faber, H C van Rooij, A Geurts van Kessel, M M Voorhorst, E Mulder, and A O Brinkmann. Cloning, structure and expression of a cDNA encoding the human androgen receptor. *BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS*, 153(1):241–248, May 1988.

[168] W D Tilley, M Marcelli, J D Wilson, and M J McPhaul. Characterization and expression of a cDNA encoding the human androgen receptor. *Proc. Natl. Acad. Sci. U.S.A.*, 86(1):327–331, January 1989.

[169] Peter E Lonergan and Donald J Tindall. Androgen receptor signaling in prostate cancer development and progression. *J Carcinog*, 10:20, 2011.

[170] Scott M Dehm and Donald J Tindall. Alternatively spliced androgen receptor variants. *Endocr. Relat. Cancer*, 18(5):R183–R196, August 2011.

[171] C M Wilson and M J McPhaul. A and B forms of the androgen receptor are expressed in a variety of human tissues. *Molecular and Cellular Endocrinology*, 120(1):51–57, June 1996.

[172] Isabelle Ahrens-Fath, Oliver Politz, Christoph Geserick, and Bernard Haendler. Androgen receptor function is modulated by the tissue-specific AR45 variant. *FEBS J.*, 272(1):74–84, January 2005.

[173] Shuji Hirata, Tomoko Shoda, Junzo Kato, and Kazuhiko Hoshi. Isoform/variant mRNAs for sex steroid hormone receptors in humans. *Trends Endocrinol. Metab.*, 14(3):124–129, April 2003.

[174] Ute M Liegibel, Ulrike Sommer, Irma Boercsoek, Ulrike Hilscher, Angelika Bierhaus, Hans-Udo Schweikert, Peter Nawroth, and Christian Kasperk. Androgen receptor isoforms AR-A and AR-B display functional differences in cultured human bone cells and genital skin fibroblasts. *Steroids*, 68(14):1179–1187, December 2003.

[175] Nigel C Bennett, Robert A Gardiner, John D Hooper, David W Johnson, and Glenda C Gobe. The International Journal of Biochemistry & Cell Biology. *International Journal of Biochemistry and Cell Biology*, 42(6):813–827, June 2010.

[176] C D Foradori, M J Weiser, and R J Handa. Non-genomic actions of androgens. *Frontiers in Neuroendocrinology*, 29(2):169–181, May 2008.

[177] Jin Li and Farook Al-Azzawi. Mechanism of androgen receptor action. *Maturitas*, 63(2):142–148, June 2009.

[178] I E Eder, Z Culig, T Putz, C Nessler-Menardi, G Bartsch, and H Klocker. Molecular biology of the androgen receptor: from molecular understanding to the clinic. *European Urology*, 40(3):241–251, September 2001.

[179] Jacob Raber. Androgens, apoE, and Alzheimer's disease. *Sci Aging Knowledge Environ*, 2004(11):re2, March 2004.

[180] Angela Rizk, Jennifer Robertson, and Jacob Raber. Behavioral performance of tfm mice supports the beneficial role of androgen receptors in spatial learning and memory. *Brain Research*, 1034(1-2):132–138, February 2005.

[181] Christian J Pike, Thuy-Vi V Nguyen, Martin Ramsden, Mingzhong Yao, M Paul Murphy, and Emily R Rosario. Androgen cell signaling pathways involved in neuroprotective actions. *Hormones and Behavior*, 53(5):693–705, May 2008.

[182] Di Wu, Grace Lin, and Andrea C Gore. Age-related changes in hypothalamic androgen receptor and estrogen receptor α in male rats. *J. Comp. Neurol.*, 512(5):688–701, February 2009.

[183] I J McEwan. The Nuclear Receptor Superfamily, 2009.

[184] Frank Claessens, Sarah Denayer, Nora Van Tilborgh, Stefanie Kerkhofs, Christine Helsen, and Annemie Haelens. Diverse roles of androgen receptor (AR) domains in AR-mediated signaling. *Nuc R Sig*, 6:e008, 2008.

[185] P M Matias, P Donner, R Coelho, M Thomaz, C Peixoto, S Macedo, N Otto, S Joschko, P Scholz, A Wegg, S Bäsler, M Schäfer, U Egner, and M A Carrondo. Structural evidence for ligand specificity in the binding domain of the human androgen receptor. Implications for pathogenic gene mutations. *J. Biol. Chem.*, 275(34):26164–26171, August 2000.

[186] John S Sack, Kevin F Kish, Chihuei Wang, Ricardo M Attar, Susan E Kiefer, Yongmi An, Ginger Y Wu, Julie E Scheffler, Mark E Salvati, and Stanley R Krystek. Crystallographic structures of the ligand-binding domains of the androgen receptor and its T877A mutant complexed with the natural agonist dihydrotestosterone. *Proc. Natl. Acad. Sci. U.S.A.*, 98(9):4904–4909, 2001.

[187] W Bourguet, P Germain, and H Gronemeyer. Nuclear receptor ligand-binding domains: three-dimensional structures, molecular interactions and pharmacological implications. *Trends in Pharmacological Sciences*, 21(10):381–388, October 2000.

[188] Laszlo Nagy and John W R Schwabe. Mechanism of the nuclear receptor molecular switch. *Trends Biochem. Sci.*, 29(6):317–324, June 2004.

[189] D Moras and H Gronemeyer. The nuclear receptor ligand-binding domain: structure and function. *Current Opinion in Cell Biology*, 10(3):384–391, June 1998.

[190] P S Danielian, R White, J A Lees, and M G Parker. Identification of a conserved region required for hormone dependent transcriptional activation by steroid hormone receptors. *The EMBO Journal*, 11(3):1025–1033, March 1992.

[191] Bin He, Robert T Gampe, Adam J Kole, Andrew T Hnat, Thomas B Stanley, Gang An, Eugene L Stewart, Rebecca I Kalman, John T Minges, and Elizabeth M Wilson. Structural basis for androgen receptor interdomain and coactivator interactions suggests a transition in nuclear receptor activation function dominance. *Molecular Cell*, 16(3):425–438, November 2004.

[192] Karine Pereira de Jésus-Tran, Pierre-Luc Côté, Line Cantin, Jonathan Blanchet, Fernand Labrie, and Rock Breton. Comparison of crystal structures of human androgen receptor ligand-binding domain complexed with various agonists reveals molecular determinants responsible for binding affinity. *Protein Sci.*, 15(5):987–999, May 2006.

[193] D M Heery, E Kalkhoven, S Hoare, and M G Parker. A signature motif in transcriptional co-activators mediates binding to nuclear receptors. *Nature*, 387(6634):733–736, June 1997.

[194] Hendrikus J Dubbink, Remko Hersmus, Chandra S Verma, Hetty A G M van der Korput, Cor A Berrevoets, Judith van Tol, Angelique C J Ziel-van der Made, Albert O Brinkmann, Ashley C W Pike, and Jan Trapman. Distinct recognition modes of FXXLF and LXXLL motifs by the androgen receptor. *Molecular Endocrinology*, 18(9):2132–2150, September 2004.

[195] Eugene Hur, Samuel J Pfaff, E Sturgis Payne, Hanne Grøn, Benjamin M Buehrer, and Robert J Fletterick. Recognition and accommodation at the androgen receptor coactivator binding interface. *Plos Biol*, 2(9):E274, September 2004.

[196] Yong Li, Millard H Lambert, and H Eric Xu. Activation of Nuclear Receptors. *Structure*, 11(7):741–746, July 2003.

[197] E Estebanez-Perpina, L A Arnold, P Nguyen, E D Rodrigues, E Mar, R Bateman, P Pallai, K M Shokat, J D Baxter, R K Guy, P Webb, and R J Fletterick. A surface on the androgen receptor that allosterically regulates coactivator binding. *Proceedings of the National Academy of Sciences*, 104(41):16074–16079, October 2007.

[198] Cynthia A Heinlein and Chawnshang Chang. Androgen receptor (AR) coregulators: an overview. *Endocrine Reviews*, 23(2):175–200, April 2002.

[199] J A Simental, M Sar, M V Lane, F S French, and E M Wilson. Transcriptional activation and nuclear targeting signals of the human androgen receptor. *J. Biol. Chem.*, 266(1):510–518, January 1991.

[200] C I Wong, Z X Zhou, M Sar, and E M Wilson. Steroid requirement for androgen receptor dimerization and DNA binding. Modulation by intramolecular interactions between the NH2-terminal and steroid-binding domains. *J. Biol. Chem.*, 268(25):19004–19012, September 1993.

[201] K Umesono and R M Evans. Determinants of target gene specificity for steroid/thyroid hormone receptors. *Cell*, 57(7):1139–1146, June 1989.

[202] J Zilliacus, A P Wright, J Carlstedt-Duke, and J A Gustafsson. Structural determinants of DNA-binding specificity by steroid receptors. *Molecular Endocrinology*, 9(4):389–400, April 1995.

[203] K Dahlman-Wright, A Wright, J A Gustafsson, and J Carlstedt-Duke. Interaction of the glucocorticoid receptor DNA-binding domain with DNA as a dimer is mediated by a short segment of five amino acids. *J. Biol. Chem.*, 266(5):3107–3112, February 1991.

[204] Qianben Wang, Wei Li, X Shirley Liu, Jason S Carroll, Olli A Jänne, Erika Krasnickas Keeton, Arul M Chinnaiyan, Kenneth J Pienta, and Myles Brown. A hierarchical network of transcription factors governs androgen receptor-dependent prostate cancer growth. *Molecular Cell*, 27(3):380–392, August 2007.

[205] Charles E Massie, Boris Adryan, Nuno L Barbosa-Morais, Andy G Lynch, Maxine G Tran, David E Neal, and Ian G Mills. New androgen receptor genomic targets show an interaction with the ETS1 transcription factor. *EMBO Rep*, 8(9):871–878, September 2007.

[206] Eric C Bolton, Alex Y So, Christina Chaivorapol, Christopher M Haqq, Hao Li, and Keith R Yamamoto. Cell- and gene-specific regulation of primary target genes by the androgen receptor. *Genes & Development*, 21(16):2005–2017, August 2007.

[207] M M Centenera, J M Harris, W D Tilley, and L M Butler. Minireview: The Contribution of Different Androgen Receptor Domains to Receptor Dimerization and Signaling. *Molecular Endocrinology*, 22(11):2373–2382, August 2008.

[208] J W Schwabe, L Chapman, J T Finch, and D Rhodes. The crystal structure of the estrogen receptor DNA-binding domain bound to DNA: how receptors discriminate between their response elements. *Cell*, 75(3):567–578, November 1993.

[209] B F Luisi, W X Xu, Z Otwinowski, L P Freedman, K R Yamamoto, and P B Sigler. Crystallographic analysis of the interaction of the glucocorticoid receptor with DNA. *Nature*, 352(6335):497–505, August 1991.

[210] Paul L Shaffer, Arif Jivan, D Eric Dollins, Frank Claessens, and Daniel T Gewirth. Structural basis of androgen receptor binding to selective androgen response elements. *Proc. Natl. Acad. Sci. U.S.A.*, 101(14):4758–4763, April 2004.

[211] Annemie Haelens, Guy Verrijdt, Leen Callewaert, Valerie Christiaens, Kris Schauwaers, Ben Peeters, Wilfried Rombauts, and Frank Claessens. DNA recognition by the androgen receptor: evidence for an alternative DNA-dependent dimerization, and an active role of sequences flanking the response element on transactivation. *Biochem. J.*, 369(Pt 1):141–151, January 2003.

[212] E Schoenmakers, P Alen, G Verrijdt, B Peeters, G Verhoeven, W Rombauts, and F Claessens. Differential DNA binding by the androgen and glucocorticoid receptors involves the second Zn-finger and a C-terminal extension of the DNA-binding domains. *Biochem. J.*, 341 (Pt 3):515–521, August 1999.

[213] E Schoenmakers, G Verrijdt, B Peeters, G Verhoeven, W Rombauts, and F Claessens. Differences in DNA binding characteristics of the androgen and glucocorticoid receptors can determine hormone-specific responses. *J. Biol. Chem.*, 275(16):12290–12297, April 2000.

[214] A Haelens, G Verrijdt, L Callewaert, B Peeters, W Rombauts, and F Claessens. Androgen-receptor-specific DNA binding to an element in the first exon of the human secretory component gene. *Biochem. J.*, 353(Pt 3):611–620, February 2001.

[215] Z X Zhou, M Sar, J A Simental, M V Lane, and E M Wilson. A ligand-dependent bipartite nuclear targeting signal in the human androgen receptor. Requirement for the DNA-binding domain and modulation by NH2-terminal and carboxyl-terminal sequences. *J. Biol. Chem.*, 269(18):13115–13123, May 1994.

[216] R M Evans. The steroid and thyroid hormone receptor superfamily. *Science*, 240(4854):889–895, May 1988.

[217] A Haelens, T Tanner, S Denayer, L Callewaert, and F Claessens. The Hinge Region Regulates DNA Binding, Nuclear Translocation, and Transactivation of the Androgen Receptor. *Cancer Research*, 67(9):4514–4523, April 2007.

[218] Vikas Chandra, Pengxiang Huang, Yoshitomo Hamuro, Srilatha Raghuram, Yongjun Wang, Thomas P Burris, and Fraydoon Rastinejad. Structure of the intact PPAR-γ–RXR-α nuclear receptor complex on DNA. *Nature*, pages 350–356, October 2008.

[219] Vikas Chandra, Pengxiang Huang, Nalini Potluri, Dalei Wu, Youngchang Kim, and Fraydoon Rastinejad. Multidomain integration in the structure of the HNF-4a nuclear receptor complex. *Nature*, 495(7441):394–398, March 2013.

[220] Igor Orlov, Natacha Rochel, Dino Moras, and Bruno P Klaholz. Structure of the full human RXR/VDR nuclear receptor heterodimer complex with its DR3 target DNA. *The EMBO Journal*, 31(2):291–300, January 2012.

[221] Natacha Rochel, Fabrice Ciesielski, Julien Godet, Edelmiro Moman, Manfred Roessle, Carole Peluso-Iltis, Martine Moulin, Michael Haertlein, Phil Callow, Yves Mély, Dmitri I Svergun, and Dino Moras. Common architecture of nuclear receptor heterodimers on DNA direct repeat elements with different spacings. *Nat. Struct. Mol. Biol.*, 18(5):564–570, May 2011.

[222] C Wang and T Uchida. [Androgen receptor gene mutations in prostate cancer]. *Nippon Hinyokika Gakkai Zasshi*, 88(5):550–556, May 1997.

[223] Jocelyn Céraline, Marion D Cruchant, Eva Erdmann, Philippe Erbs, Jean-Emmanuel Kurtz, Brigitte Duclos, Didier Jacqmin, Dominique Chopin, and Jean-Pierre Bergerat. Constitutive activation of the androgen receptor by a point mutation in the hinge region: a new mechanism for androgen-independent growth in prostate cancer. *Int. J. Cancer*, 108(1):152–157, January 2004.

[224] M Fu, C Wang, A T Reutens, J Wang, R H Angeletti, L Siconolfi-Baez, V Ogryzko, M L Avantaggiati, and R G Pestell. p300 and p300/cAMP-response element-binding protein-associated factor acetylate the androgen receptor at sites governing hormone-dependent transactivation. *J. Biol. Chem.*, 275(27):20853–20860, July 2000.

[225] Maofu Fu, Chenguang Wang, Jian Wang, Xueping Zhang, Toshiyuki Sakamaki, Y G Yeung, Chawnshang Chang, Torsten Hopp, Suzanne A W Fuqua, Ellis Jaffray, Ron T Hay, Jorma J Palvimo, Olli A Jänne, and Richard G Pestell. Androgen receptor acetylation governs trans activation and MEKK1-induced apoptosis without affecting in vitro sumoylation and trans-repression function. *Molecular and Cellular Biology*, 22(10):3373–3388, May 2002.

[226] Monzy Thomas, Nahid Dadgar, Abhishek Aphale, Jennifer M Harrell, Robin Kunkel, William B Pratt, and Andrew P Lieberman. Androgen receptor acetylation site mutations cause trafficking defects, misfolding, and aggregation similar to expanded glutamine tracts. *J. Biol. Chem.*, 279(9):8389–8395, February 2004.

[227] Soyoung Ko, Jungmi Ahn, Chung S Song, Soyoung Kim, Katarzyna Knapczyk-Stwora, and Bandana Chatterjee. Lysine methylation and functional modulation

of androgen receptor by Set9 methyltransferase. *Mol. Endocrinol.*, 25(3):433–444, March 2011.

[228] Luke Gaughan, Jacqueline Stockley, Nan Wang, Stuart R C McCracken, Achim Treumann, Kelly Armstrong, Fadhel Shaheen, Kate Watt, Iain J McEwan, Chenguang Wang, Richard G Pestell, and Craig N Robson. Regulation of the androgen receptor by SET9-mediated methylation. *Nucleic Acids Research*, 39(4):1266–1279, March 2011.

[229] Iain J McEwan, Derek Lavery, Katharina Fischer, and Kate Watt. Natural disordered sequences in the amino terminal domain of nuclear receptors: lessons from the androgen and glucocorticoid receptors. *Nuc R Sig*, 5:e001, 2007.

[230] N L Chamberlain, D C Whitacre, and R L Miesfeld. Delineation of two distinct type 1 activation functions in the androgen receptor amino-terminal domain. *J. Biol. Chem.*, 271(43):26772–26778, October 1996.

[231] V Christiaens. Characterization of the Two Coactivator-interacting Surfaces of the Androgen Receptor and Their Relative Role in Transcriptional Control*. *Journal of Biological Chemistry*, 277(51):49230–49237, October 2002.

[232] S M Dehm, K M Regan, L J Schmidt, and D J Tindall. Selective Role of an NH2-Terminal WxxLF Motif for Aberrant Androgen Receptor Activation in Androgen Depletion Independent Prostate Cancer Cells. *Cancer Research*, 67(20):10067–10077, October 2007.

[233] P Alen, F Claessens, G Verhoeven, W Rombauts, and B Peeters. The androgen receptor amino-terminal domain plays a key role in p160 coactivator-stimulated gene transcription. *Molecular and Cellular Biology*, 19(9):6085–6097, September 1999.

[234] C L Bevan, S Hoare, F Claessens, D M Heery, and M G Parker. The AF1 and AF2 domains of the androgen receptor interact with distinct regions of SRC1. *Molecular and Cellular Biology*, 19(12):8383–8392, December 1999.

[235] E Langley, Z X Zhou, and E M Wilson. Evidence for an anti-parallel orientation of the ligand-activated human androgen receptor dimer. *J. Biol. Chem.*, 270(50):29983–29990, December 1995.

[236] E Langley, J A Kemppainen, and E M Wilson. Intermolecular NH2-/carboxyl-terminal interactions in androgen receptor dimerization revealed by mutations that cause androgen insensitivity. *J. Biol. Chem.*, 273(1):92–101, January 1998.

[237] Karine Steketee, Cor A Berrevoets, Hendrikus J Dubbink, Paul Doesburg, Remko Hersmus, Albert O Brinkmann, and Jan Trapman. Amino acids 3-13 and amino acids in and flanking the 23FxxLF27 motif modulate the interaction between the N-terminal and ligand-binding domain of the androgen receptor. *European Journal of Biochemistry*, 269(23):5780–5791, December 2002.

[238] Bin He and Elizabeth M Wilson. Electrostatic modulation in steroid receptor recruitment of LXXLL and FXXLF motifs. *Molecular and Cellular Biology*, 23(6):2135–2150, March 2003.

[239] C J Burd, C E Petre, H Moghadam, E M Wilson, and K E Knudsen. Cyclin D1 binding to the androgen receptor (AR) NH2-terminal domain inhibits activation function 2 association and reveals dual roles for AR corepression. *Molecular Endocrinology*, 19(3):607–620, March 2005.

[240] Suxia Bai, Bin He, and Elizabeth M Wilson. Melanoma antigen gene protein MAGE-11 regulates androgen receptor function by modulating the interdomain interaction. *Molecular and Cellular Biology*, 25(4):1238–1257, February 2005.

[241] Feng Jin, Frank Claessens, and Joseph D Fondell. Regulation of androgen receptor-dependent transcription by coactivator MED1 is mediated through a newly discovered noncanonical binding motif. *Journal of Biological Chemistry*, 287(2):858–870, January 2012.

[242] Ping Zhu, Sung Hee Baek, Eliot M Bourk, Kenneth A Ohgi, Ivan Garcia-Bassets, Hideki Sanjo, Shizuo Akira, Paul F Kotol, Christopher K Glass, Michael G Rosenfeld, and David W Rose. Macrophage/cancer cell interactions mediate hormone resistance by a nuclear receptor derepression pathway. *Cell*, 124(3):615–629, February 2006.

[243] B He, J A Kemppainen, and E M Wilson. FXXLF and WXXLF sequences mediate the NH2-terminal interaction with the ligand binding domain of the androgen receptor. *J. Biol. Chem.*, 275(30):22986–22994, July 2000.

[244] B He. Dependence of Selective Gene Activation on the Androgen Receptor NH2- and COOH-terminal Interaction. *Journal of Biological Chemistry*, 277(28):25631–25639, May 2002.

[245] B He, N T Bowen, J T Minges, and E M Wilson. Androgen-induced NH2- and COOH-terminal Interaction Inhibits p160 coactivator recruitment by activation function 2. *J. Biol. Chem.*, 276(45):42293–42301, November 2001.

[246] William H Lagarde, Amanda J Blackwelder, John T Minges, Andrew T Hnat, Frank S French, and Elizabeth M Wilson. Androgen receptor exon 1 mutation causes androgen insensitivity by creating phosphorylation site and inhibiting melanoma antigen-A11 activation of NH2- and carboxyl-terminal interaction-dependent transactivation. *Journal of Biological Chemistry*, 287(14):10905–10915, March 2012.

[247] Bin He, Suxia Bai, Andrew T Hnat, Rebecca I Kalman, John T Minges, Cam Patterson, and Elizabeth M Wilson. An androgen receptor NH2-terminal conserved motif interacts with the COOH terminus of the Hsp70-interacting protein (CHIP). *J. Biol. Chem.*, 279(29):30643–30653, July 2004.

[248] A R La Spada, E M Wilson, D B Lubahn, A E Harding, and K H Fischbeck. Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. *Nature*, 352(6330):77–79, July 1991.

[249] N L Chamberlain, E D Driver, and R L Miesfeld. The length and location of CAG trinucleotide repeats in the androgen receptor N-terminal domain affect transactivation function. *Nucleic Acids Research*, 22(15):3181–3186, August 1994.

[250] R A Irvine, H Ma, M C Yu, R K Ross, M R Stallcup, and G A Coetzee. Inhibition of p160-mediated coactivation with increasing androgen receptor polyglutamine length. *Human Molecular Genetics*, 9(2):267–274, January 2000.

[251] Dacheng Ding, Lihua Xu, Mani Menon, G Prem Veer Reddy, and Evelyn R Barrack. Effect of GGC (glycine) repeat length polymorphism in the human androgen receptor on androgen action. *Prostate*, 62(2):133–139, February 2005.

[252] Leen Callewaert, Valerie Christiaens, Annemie Haelens, Guy Verrijdt, Guido Verhoeven, and Frank Claessens. Implications of a polyglutamine tract in the function of the human androgen receptor. *BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS*, 306(1):46–52, June 2003.

[253] Grant Buchanan, Miao Yang, Albert Cheong, Jonathan M Harris, Ryan A Irvine, Paul F Lambert, Nicole L Moore, Michael Raynor, Petra J Neufing, Gerhard A Coetzee, and Wayne D Tilley. Structural and functional consequences of glutamine tract variation in the androgen receptor. *Human Molecular Genetics*, 13(16):1677–1692, August 2004.

[254] E Giovannucci, M J Stampfer, K Krithivas, M Brown, D Dahl, A Brufsky, J Talcott, C H Hennekens, and P W Kantoff. The CAG repeat within the androgen receptor gene and its relationship to prostate cancer. *Proc. Natl. Acad. Sci. U.S.A.*, 94(7):3320–3323, April 1997.

[255] S A Ingles, R K Ross, M C Yu, R A Irvine, G La Pera, R W Haile, and G A Coetzee. Association of prostate cancer risk with genetic polymorphisms in vitamin D receptor and androgen receptor. *J. Natl. Cancer Inst.*, 89(2):166–170, January 1997.

[256] R A Irvine, M C Yu, R K Ross, and G A Coetzee. The CAG and GGC microsatellites of the androgen receptor gene are in linkage disequilibrium in men with prostate cancer. *Cancer Research*, 55(9):1937–1940, May 1995.

[257] K Sircar, B Gottlieb, C Alvarado, A Aprikian, L K Beitel, M Alam-Fahmy, L Begin, and M Trifiro. Androgen receptor CAG repeat length contraction in diseased and non-diseased prostatic tissues. *Prostate Cancer Prostatic Dis.*, 10(4):360–368, 2007.

[258] M L Cutress, H C Whitaker, I G Mills, M Stewart, and D E Neal. Structural basis for the nuclear import of the human androgen receptor. *Journal of Cell Science*, 121(7):957–968, March 2008.

[259] Jacqueline Brodie and Iain J McEwan. Intra-domain communication between the N-terminal and DNA-binding domains of the androgen receptor: modulation of androgen response element DNA binding. *Journal of Molecular Endocrinology*, 34(3):603–615, June 2005.

[260] Christoph Geserick, Hellmuth-Alexander Meyer, Karina Barbulescu, and Bernard Haendler. Differential modulation of androgen receptor action by deoxyribonucleic acid response elements. *Molecular Endocrinology*, 17(9):1738–1750, September 2003.

[261] R Kumar, I V Baskakov, G Srinivasan, D W Bolen, J C Lee, and E B Thompson. Interdomain signaling in a two-domain fragment of the human glucocorticoid receptor. *J. Biol. Chem.*, 274(35):24737–24741, August 1999.

[262] Suzanne E Wardell, Stanley C Kwok, Lori Sherman, Robert S Hodges, and Dean P Edwards. Regulation of the amino-terminal transcription activation domain of progesterone receptor by a cofactor-induced protein folding mechanism. *Molecular and Cellular Biology*, 25(20):8792–8808, October 2005.

[263] N Greenfield, V Vijayanathan, T J Thomas, M A Gallo, and T Thomas. Increase in the stability and helical content of estrogen receptor alpha in the presence of the estrogen response element: analysis by circular dichroism spectroscopy. *Biochemistry*, 40(22):6646–6652, June 2001.

[264] J R Wood, G L Greene, and A M Nardulli. Estrogen response elements function as allosteric modulators of estrogen receptor conformation. *Molecular and Cellular Biology*, 18(4):1927–1934, April 1998.

[265] C M Klinge, S C Jernigan, S L Smith, V V Tyulmenkov, and P C Kulakosky. Estrogen response element sequence impacts the conformation and transcriptional activity of estrogen receptor alpha. *Molecular and Cellular Endocrinology*, 174(1-2):151–166, March 2001.

[266] M A Loven, J R Wood, and A M Nardulli. Interaction of estrogen receptors alpha and beta with estrogen response elements. *Molecular and Cellular Endocrinology*, 181(1-2):151–163, July 2001.

[267] Guo-Zhen Liu, Hua Wang, and Zhengxin Wang. Identification of a highly conserved domain in the androgen receptor that suppresses the DNA-binding domain-DNA interactions. *J. Biol. Chem.*, 278(17):14956–14960, April 2003.

[268] Jyotsna B Pippal, Yizhou Yao, Fraser M Rogerson, and Peter J Fuller. Structural and functional characterization of the interdomain interaction in the mineralocorticoid receptor. *Mol. Endocrinol.*, 23(9):1360–1370, September 2009.

[269] Glenn S Takimoto, Lin Tung, Hany Abdel-Hafiz, Michael G Abel, Carol A Sartorius, Jennifer K Richer, Britta M Jacobsen, David L Bain, and Kathryn B Horwitz. Functional properties of the N-terminal region of progesterone receptors and their mechanistic relationship to structure. *J. Steroid Biochem. Mol. Biol.*, 85(2-5):209–219, June 2003.

[270] M J Tetel, P H Giangrande, S A Leonhardt, D P McDonnell, and D P Edwards. Hormone-dependent interaction between the amino- and carboxyl-terminal domains of progesterone receptor in vitro and in vivo. *Molecular Endocrinology*, 13(6):910–924, June 1999.

[271] W L Kraus, E M McInerney, and B S Katzenellenbogen. Ligand-dependent, transcriptionally productive association of the amino- and carboxyl-terminal regions of a steroid hormone nuclear receptor. *Proc. Natl. Acad. Sci. U.S.A.*, 92(26):12314–12318, December 1995.

[272] R Métivier, G Penot, G Flouriot, and F Pakdel. Synergism between ERalpha transactivation function 1 (AF-1) and AF-2 mediated by steroid receptor coactivator protein-1: requirement for the AF-1 alpha-helical core and for a direct interaction between the N- and C-terminal domains. *Molecular Endocrinology*, 15(11):1953–1970, November 2001.

[273] Leen Callewaert, Guy Verrijdt, Valerie Christiaens, Annemie Haelens, and Frank Claessens. Dual function of an amino-terminal amphipatic helix in androgen receptor-mediated transactivation through specific and nonspecific response elements. *J. Biol. Chem.*, 278(10):8212–8218, March 2003.

[274] C A Berrevoets, P Doesburg, K Steketee, J Trapman, and A O Brinkmann. Functional interactions of the AF-2 activation domain core region of the human androgen receptor with the amino-terminal domain and with the transcriptional coactivator TIF2 (transcriptional intermediary factor2). *Molecular Endocrinology*, 12(8):1172–1183, August 1998.

[275] T Ikonen, J J Palvimo, and O A Janne. Interaction between the amino- and carboxyl-terminal regions of the rat androgen receptor modulates transcriptional activity and is influenced by nuclear receptor coactivators. *J. Biol. Chem.*, 272(47):29821–29828, November 1997.

[276] S M Dehm and D J Tindall. Androgen Receptor Structural and Functional Elements: Role and Regulation in Prostate Cancer. *Molecular Endocrinology*, 21(12):2855–2863, July 2007.

[277] Jiwen Li, Junjiang Fu, Charalambos Toumazou, Ho-Geun Yoon, and Jiemin Wong. A role of the amino-terminal (N) and carboxyl-terminal (C) interaction in binding of androgen receptor to chromatin. *Molecular Endocrinology*, 20(4):776–785, April 2006.

[278] Martin E van Royen, Sónia M Cunha, Maartje C Brink, Karin A Mattern, Alex L Nigg, Hendrikus J Dubbink, Pernette J Verschure, Jan Trapman, and Adriaan B Houtsmuller. Compartmentalization of androgen receptor protein-protein interactions in living cells. *J. Cell Biol.*, 177(1):63–72, April 2007.

[279] Fred Schaufele, Xavier Carbonell, Martin Guerbadot, Sabine Borngraeber, Mark S Chapman, Aye Aye K Ma, Jeffrey N Miner, and Marc I Diamond. The structural basis of androgen receptor activation: intramolecular and intermolecular amino-carboxy interactions. *Proc. Natl. Acad. Sci. U.S.A.*, 102(28):9802–9807, July 2005.

[280] Tove I Klokk, Piotr Kurys, Cem Elbi, Akhilesh K Nagaich, Anindya Hendarwanto, Thomas Slagsvold, Ching-Yi Chang, Gordon L Hager, and Fahri Saatcioglu. Ligand-specific dynamics of the androgen receptor at its response element in living cells. *Molecular and Cellular Biology*, 27(5):1823–1843, March 2007.

[281] James Reid, Sharon M Kelly, Kate Watt, Nicholas C Price, and Iain J McEwan. Conformational analysis of the androgen receptor amino-terminal domain

involved in transactivation. Influence of structure-stabilizing solutes and protein-protein interactions. *J. Biol. Chem.*, 277(22):20079–20086, May 2002.

[282] Raj Kumar, Russell Betney, Jianquan Li, E Brad Thompson, and Iain J McEwan. Induced α-Helix Structure in AF1 of the Androgen Receptor upon Binding Transcription Factor TFIIF †. *Biochemistry*, 43(11):3008–3013, March 2004.

[283] Derek N Lavery and Iain J McEwan. Structural characterization of the native NH2-terminal transactivation domain of the human androgen receptor: a collapsed disordered conformation underlies structural plasticity and protein-induced folding. *Biochemistry*, 47(11):3360–3369, March 2008.

[284] V S Pande and D S Rokhsar. Is the molten globule a third phase of proteins? *Proc. Natl. Acad. Sci. U.S.A.*, 95(4):1490–1494, February 1998.

[285] M Ohgushi and A Wada. 'Molten-globule state': a compact form of globular proteins with mobile side-chains. *FEBS Lett.*, 164(1):21–24, November 1983.

[286] A Keith Dunker, Marc S Cortese, Pedro Romero, Lilia M Iakoucheva, and Vladimir N Uversky. Flexible nets. The roles of intrinsic disorder in protein interaction networks. *FEBS J.*, 272(20):5129–5148, October 2005.

[287] X Edward Zhou, Kelly Suino-Powell, Phumzile L Ludidi, Donald P McDonnell, and H Eric Xu. Expression, purification and primary crystallographic study of human androgen receptor in complex with DNA and coactivator motifs. *PROTEIN EXPRESSION AND PURIFICATION*, 71(1):21–27, May 2010.

[288] Rita Pancsa and Peter Tompa. Structural disorder in eukaryotes. *PLoS ONE*, 7(4):e34687, 2012.

[289] Peter Tompa and Alan Fersht. *Structure and Function of Intrinsically Disordered Proteins*. CRC Press, December 2010.

[290] Ishwar Radhakrishnan, Gabriela C Pérez-Alvarado, H Jane Dyson, and Peter E Wright. Conformational preferences in the Ser 133-phosphorylated and non-phosphorylated forms of the kinase inducible transactivation domain of CREB. *FEBS Lett.*, 430(3):317–322, 1998.

[291] J P Richards, H P Bächinger, R H Goodman, and R G Brennan. Analysis of the structural properties of cAMP-responsive element-binding protein (CREB) and phosphorylated CREB. *J. Biol. Chem.*, 271(23):13716–13723, June 1996.

[292] I Radhakrishnan, G C Pérez-Alvarado, D Parker, H J Dyson, M R Montminy, and P E Wright. Solution structure of the KIX domain of CBP bound to the

transactivation domain of CREB: a model for activator:coactivator interactions. *Cell*, 91(6):741–752, December 1997.

[293] Stephen J Demarest, Maria Martinez-Yamout, John Chung, Hongwu Chen, Wei Xu, H Jane Dyson, Ronald M Evans, and Peter E Wright. Mutual synergistic folding in recruitment of CBP/p300 by p160 nuclear receptor coactivators. *Nature*, 415(6871):549–553, 2002.

[294] Satish K Nair and Stephen K Burley. X-ray structures of Myc-Max and Mad-Max recognizing DNA. Molecular bases of regulation by proto-oncogenic transcription factors. *Cell*, 112(2):193–205, January 2003.

[295] Peter E Wright and H Jane Dyson. Linking folding and binding. *Current Opinion in Structural Biology*, 19(1):31–38, February 2009.

[296] P Andrew Chong, Barish Ozdamar, Jeffrey L Wrana, and Julie D Forman-Kay. Disorder in a target for the smad2 mad homology 2 domain and its implications for binding and specificity. *J. Biol. Chem.*, 279(39):40707–40714, September 2004.

[297] Peter Tompa and Monika Fuxreiter. Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions. *Trends Biochem. Sci.*, 33(1):2–8, January 2008.

[298] T A Graham, D M Ferkey, F Mao, D Kimelman, and W Xu. Tcf4 can specifically recognize beta-catenin using alternative conformations. *Nat. Struct. Biol.*, 8(12):1048–1052, December 2001.

[299] H Jane Dyson and Peter E Wright. Coupling of folding and binding for unstructured proteins. *Current Opinion in Structural Biology*, 12(1):54–60, February 2002.

[300] Amrita Mohan, Christopher J Oldfield, Predrag Radivojac, Vladimir Vacic, Marc S Cortese, A Keith Dunker, and Vladimir N Uversky. Analysis of molecular recognition features (MoRFs). *J. Mol. Biol.*, 362(5):1043–1059, October 2006.

[301] Sonja A Dames, Maria Martinez-Yamout, Roberto N De Guzman, H Jane Dyson, and Peter E Wright. Structural basis for Hif-1 alpha /CBP recognition in the cellular hypoxic response. *Proc. Natl. Acad. Sci. U.S.A.*, 99(8):5271–5276, April 2002.

[302] Vincent J Hilser and E Brad Thompson. Intrinsic disorder as a mechanism to optimize allosteric coupling in proteins. *Proc. Natl. Acad. Sci. U.S.A.*, 104(20):8311–8315, May 2007.

[303] Miles A Pufall, Gregory M Lee, Mary L Nelson, Hyun-Seo Kang, Algirdas Velyvis, Lewis E Kay, Lawrence P McIntosh, and Barbara J Graves. Variable control of Ets-1 DNA binding by multiple phosphates in an unstructured region. *Science*, 309(5731):142–145, July 2005.

[304] B A Shoemaker, J J Portman, and P G Wolynes. Speeding molecular recognition by using the folding funnel: the fly-casting mechanism. *Proc. Natl. Acad. Sci. U.S.A.*, 97(16):8868–8873, August 2000.

[305] James R Perkins, Ilhem Diboun, Benoit H Dessailly, Jon G Lees, and Christine Orengo. Transient Protein-Protein Interactions: Structural, Functional, and Network Properties. *Structure/Folding and Design*, 18(10):1233–1243, October 2010.

[306] Vladimir N Uversky, Christopher J Oldfield, and A Keith Dunker. Intrinsically disordered proteins in human diseases: introducing the D2 concept. *Annu. Rev. Biophys.*, 37:215–246, 2008.

[307] Swasti Raychaudhuri, Sucharita Dey, Nitai P Bhattacharyya, and Debashis Mukhopadhyay. The Role of Intrinsically Unstructured Proteins in Neurodegenerative Diseases. *PLoS ONE*, 4(5):e5566, May 2009.

[308] Steven J Metallo. Intrinsically disordered proteins are potential drug targets. *Current Opinion in Chemical Biology*, 14(4):481–488, August 2010.

[309] Yugong Cheng, Tanguy LeGall, Christopher J Oldfield, James P Mueller, Ya-Yue J Van, Pedro Romero, Marc S Cortese, Vladimir N Uversky, and A Keith Dunker. Rational drug design via intrinsically disordered protein. *Trends in Biotechnology*, 24(10):435–442, October 2006.

[310] Lyubomir T Vassilev, Binh T Vu, Bradford Graves, Daisy Carvajal, Frank Podlaski, Zoran Filipovic, Norman Kong, Ursula Kammlott, Christine Lukacs, Christian Klein, Nader Fotouhi, and Emily A Liu. In vivo activation of the p53 pathway by small-molecule antagonists of MDM2. *Science*, 303(5659):844–848, February 2004.

[311] Ariele Viacava Follis, Dalia I Hammoudeh, Huabo Wang, Edward V Prochownik, and Steven J Metallo. Brief Communication. *Chemistry & Biology*, 15(11):1149–1155, November 2008.

[312] Dalia I Hammoudeh, Ariele Viacava Follis, Edward V Prochownik, and Steven J Metallo. Multiple independent binding sites for small-molecule inhibitors on the oncoprotein c-Myc. *J. Am. Chem. Soc.*, 131(21):7390–7401, June 2009.

[313] Hayriye V Erkizan, Yali Kong, Melinda Merchant, Silke Schlottmann, Julie S Barber-Rotenberg, Linshan Yuan, Ogan D Abaan, Tsu-hang Chou, Sivanesan Dakshanamurthy, Milton L Brown, Aykut Üren, and Jeffrey A Toretsky. A small molecule blocking oncogenic protein EWS-FLI1 interaction with RNA helicase A inhibits growth of Ewing's sarcoma. *Nat Med*, 15(7):750–756, July 2009.

[314] Thomas L Kukar, Thomas B Ladd, Maralyssa A Bann, Patrick C Fraering, Rajeshwar Narlawar, Ghulam M Maharvi, Brent Healy, Robert Chapman, Alfred T Welzel, Robert W Price, Brenda Moore, Vijayaraghavan Rangachari, Bernadette Cusack, Jason Eriksen, Karen Jansen-West, Christophe Verbeeck, Debra Yager, Christopher Eckman, Wenjuan Ye, Sarah Sagi, Barbara A Cottrell, Justin Torpey, Terrone L Rosenberry, Abdul Fauq, Michael S Wolfe, Boris Schmidt, Dominic M Walsh, Edward H Koo, and Todd E Golde. Substrate-targeting gamma-secretase modulators. *Nature*, 453(7197):925–929, June 2008.

[315] M Marcelli, M Ittmann, S Mariani, R Sutherland, R Nigam, L Murthy, Y Zhao, D DiConcini, E Puxeddu, A Esen, J Eastham, N L Weigel, and D J Lamb. Androgen receptor mutations in prostate cancer. *Cancer Research*, 60(4):944–949, February 2000.

[316] G Buchanan, N M Greenberg, H I Scher, J M Harris, V R Marshall, and W D Tilley. Collocation of androgen receptor gene mutations in prostate cancer. *Clin. Cancer Res.*, 7(5):1273–1281, May 2001.

[317] Eija-R Hyytinen, Kyllikki Haapala, James Thompson, Ilkka Lappalainen, Mikko Roiha, Immo Rantala, Heikki J Helin, Olli A Jänne, Mauno Vihinen, Jorma J Palvimo, and Pasi A Koivisto. Pattern of somatic androgen receptor gene mutations in patients with hormone-refractory prostate cancer. *Lab. Invest.*, 82(11):1591–1598, November 2002.

[318] Mary C Thomas and Cheng-Ming Chiang. The General Transcription Machinery and General Cofactors. *Critical Reviews in Biochemistry and Molecular Biology*, 41(3):105–178, January 2006.

[319] I J McEwan and J Gustafsson. Interaction of the human androgen receptor transactivation function with the general transcription factor TFIIF. *Proc. Natl. Acad. Sci. U.S.A.*, 94(16):8485–8490, August 1997.

[320] James Reid, Ian Murray, Kate Watt, Russell Betney, and Iain J McEwan. The androgen receptor interacts with multiple regions of the large subunit of general transcription factor TFIIF. *J. Biol. Chem.*, 277(43):41247–41253, October 2002.

[321] D K Lee, H O Duan, and C Chang. From androgen receptor to the general transcription factor TFIIH. Identification of cdk activating kinase (CAK) as an androgen receptor NH(2)-terminal associated coactivator. *J. Biol. Chem.*, 275(13):9308–9313, March 2000.

[322] Dong Kun Lee and Chawnshang Chang. Molecular communication between androgen receptor and general transcription machinery. *J. Steroid Biochem. Mol. Biol.*, 84(1):41–49, January 2003.

[323] D Chen, T Riedl, E Washbrook, P E Pace, R C Coombes, J M Egly, and S Ali. Activation of estrogen receptor alpha by S118 phosphorylation involves a ligand-dependent interaction with TFIIH and participation of CDK7. *Molecular Cell*, 6(1):127–137, July 2000.

[324] D K Lee, H O Duan, and C Chang. Androgen receptor interacts with the positive elongation factor P-TEFb and enhances the efficiency of transcriptional elongation. *J. Biol. Chem.*, 276(13):9978–9984, March 2001.

[325] J W Conaway, A Shilatifard, A Dvir, and R C Conaway. Control of elongation by RNA polymerase II. *Trends Biochem. Sci.*, 25(8):375–380, August 2000.

[326] J Q Svejstrup, P Vichi, and J M Egly. The multiple roles of transcription/repair factor TFIIH. *Trends Biochem. Sci.*, 21(9):346–350, September 1996.

[327] A Dvir, J W Conaway, and R C Conaway. Mechanism of transcription initiation and promoter escape by RNA polymerase II. *Current Opinion in Genetics & Development*, 11(2):209–214, April 2001.

[328] D Reines, J W Conaway, and R C Conaway. The RNA polymerase II general elongation factors. *Trends Biochem. Sci.*, 21(9):351–355, September 1996.

[329] Dong Kun Lee, Mei Li, and Chawnshang Chang. The second largest subunit of RNA polymerase II interacts with and enhances transactivation of androgen receptor. *BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS*, 302(1):162–169, February 2003.

[330] Yongfeng Shang, Molly Myers, and Myles Brown. Formation of the androgen receptor transcription complex. *Molecular Cell*, 9(3):601–610, March 2002.

[331] Myles C Hodgson, Inna Astapova, Anthony N Hollenberg, and Steven P Balk. Activity of androgen receptor antagonist bicalutamide in prostate cancer cells is independent of NCoR and SMRT corepressors. *Cancer Research*, 67(17):8388–8395, September 2007.

[332] Guoqing Liao, Liuh-Yow Chen, Aihua Zhang, Aparna Godavarthy, Fang Xia, Jagadish Chandra Ghosh, Hui Li, and J Don Chen. Regulation of androgen receptor activity by the nuclear receptor corepressor SMRT. *J. Biol. Chem.*, 278(7):5052–5061, February 2003.

[333] Shinta Cheng, Sabrina Brzostek, Suzanne R Lee, Anthony N Hollenberg, and Steven P Balk. Inhibition of the dihydrotestosterone-activated androgen receptor by nuclear receptor corepressor. *Molecular Endocrinology*, 16(7):1492–1501, July 2002.

[334] Cor A Berrevoets, Arzu Umar, Jan Trapman, and Albert O Brinkmann. Differential modulation of androgen receptor transcriptional activity by the nuclear receptor co-repressor (N-CoR). *Biochem. J.*, 379(Pt 3):731–738, May 2004.

[335] Myles C Hodgson, Inna Astapova, Shinta Cheng, Larissa J Lee, Manon C Verhoeven, Eunis Choi, Steven P Balk, and Anthony N Hollenberg. The androgen receptor recruits nuclear receptor CoRepressor (N-CoR) in the presence of mifepristone via its N and C termini revealing a novel molecular mechanism for androgen receptor antagonists. *J. Biol. Chem.*, 280(8):6511–6519, February 2005.

[336] K B Cleutjens, C C van Eekelen, H A van der Korput, A O Brinkmann, and J Trapman. Two androgen response regions cooperate in steroid hormone regulated activity of the prostate-specific antigen promoter. *J. Biol. Chem.*, 271(11):6379–6388, March 1996.

[337] K B Cleutjens, H A van der Korput, C C van Eekelen, H C van Rooij, P W Faber, and J Trapman. An androgen response element in a far upstream enhancer region is essential for high, androgen-regulated activity of the prostate-specific antigen promoter. *Molecular Endocrinology*, 11(2):148–161, February 1997.

[338] Zhigang Kang, Olli A Jänne, and Jorma J Palvimo. Coregulator recruitment and histone modifications in transcriptional regulation by the androgen receptor. *Molecular Endocrinology*, 18(11):2633–2648, November 2004.

[339] Qianben Wang, Jason S Carroll, and Myles Brown. Spatial and temporal recruitment of androgen receptor and its coactivators involves chromosomal looping and polymerase tracking. *Molecular Cell*, 19(5):631–642, September 2005.

[340] F Gaiser, S Tan, and T J Richmond. Novel dimerization fold of RAP30/RAP74 in human TFIIF at 1.7 A resolution. *J. Mol. Biol.*, 302(5):1119–1127, October 2000.

[341] C M Groft, S N Uljon, R Wang, and M H Werner. Structural homology between the Rap30 DNA-binding domain and linker histone H5: implications for preinitiation complex assembly. *Proc. Natl. Acad. Sci. U.S.A.*, 95(16):9117–9122, August 1998.

[342] F Robert, M Douziech, D Forget, J M Egly, J Greenblatt, Z F Burton, and B Coulombe. Wrapping of promoter DNA around the RNA polymerase II initiation complex induced by TFIIF. *Molecular Cell*, 2(3):341–351, September 1998.

[343] C Yong, H Mitsuyasu, Z Chun, S Oshiro, N Hamasaki, and S Kitajima. Structure of the human transcription factor TFIIF revealed by limited proteolysis with trypsin. *FEBS Lett.*, 435(2-3):191–194, September 1998.

[344] S M Fang and Z F Burton. RNA polymerase II-associated protein (RAP) 74 binds transcription factor (TF) IIB and blocks TFIIB-RAP30 binding. *J. Biol. Chem.*, 271(20):11703–11709, May 1996.

[345] B Q Wang and Z F Burton. Functional domains of human RAP74 including a masked polymerase binding domain. *J. Biol. Chem.*, 270(45):27035–27044, November 1995.

[346] J Archambault, G Pan, G K Dahmus, M Cartier, N Marshall, S Zhang, M E Dahmus, and J Greenblatt. FCP1, the RAP74-interacting subunit of a human protein phosphatase that dephosphorylates the carboxyl-terminal domain of RNA polymerase IIO. *J. Biol. Chem.*, 273(42):27593–27601, October 1998.

[347] E J Cho. Opposing effects of Ctk1 kinase and Fcp1 phosphatase at Ser 2 of the RNA polymerase II C-terminal domain. *Genes & Development*, 15(24):3319–3329, December 2001.

[348] Yuan He, Jie Fang, Dylan J Taatjes, and Eva Nogales. Structural visualization of key steps in human transcription initiation. *Nature*, 495(7442):481–486, February 2013.

[349] K Murakami, H Elmlund, N Kalisman, D A Bushnell, C M Adams, M Azubel, D Elmlund, Y Levi-Kalisman, X Liu, B J Gibbons, M Levitt, and R D Kornberg. Architecture of an RNA Polymerase II Transcription Pre-Initiation Complex. *Science*, 342(6159):1238724–1238724, November 2013.

[350] K Kamada, J De Angelis, R G Roeder, and S K Burley. Crystal structure of the C-terminal domain of the RAP74 subunit of human transcription factor IIF. *Proc. Natl. Acad. Sci. U.S.A.*, 98(6):3115–3120, March 2001.

[351] Miltiadis Paliouras, Naif Zaman, Rose Lumbroso, Laurie Kapogeorgakis, Lenore K Beitel, Edwin Wang, and Mark Trifiro. Dynamic rewiring of the androgen receptor protein interaction network correlates with prostate cancer clinical outcomes. *Integr Biol (Camb)*, 3(10):1020–1032, October 2011.

[352] Derek N Lavery and Iain J McEwan. Functional Characterization of the Native NH2-Terminal Transactivation Domain of the Human Androgen Receptor: Binding Kinetics for Interactions with TFIIF and SRC-1a†. *Biochemistry*, 47(11):3352–3359, 2008.

[353] R Betney and I J McEwan. Role of conserved hydrophobic amino acids in androgen receptor AF-1 function. *Journal of Molecular Endocrinology*, 31(3):427–439, December 2003.

[354] N S Berrow, D Alderton, S Sainsbury, J Nettleship, R Assenberg, N Rahman, D I Stuart, and R J Owens. A versatile ligation-independent cloning method suitable for high-throughput expression screening applications. *Nucleic Acids Research*, 35(6):e45–e45, March 2007.

[355] René Assenberg, Olivier Delmas, Stephen C Graham, Anil Verma, Nick Berrow, David I Stuart, Raymond J Owens, Hervé Bourhy, and Jonathan M Grimes. Expression, purification and crystallization of a lyssavirus matrix (M) protein. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.*, 64(Pt 4):258–262, April 2008.

[356] F C Neidhardt, P L Bloch, and D F Smith. Culture medium for enterobacteria. *J. Bacteriol.*, 119(3):736–747, September 1974.

[357] B Q Wang, C F Kostrub, A Finkelstein, and Z F Burton. Production of human RAP30 and RAP74 in bacterial cells. *PROTEIN EXPRESSION AND PURIFICATION*, 4(3):207–214, June 1993.

[358] H Jane Dyson and Peter E Wright. Insights into the structure and dynamics of unfolded proteins from nuclear magnetic resonance. *Adv. Protein Chem.*, 62:311–340, 2002.

[359] H Jane Dyson and Peter E Wright. Unfolded Proteins and Protein Folding Studied by NMR. *Chem. Rev.*, 104(8):3607–3622, August 2004.

[360] Tanja Mittag and Julie D Forman-Kay. Atomic-level characterization of disordered protein ensembles. *Current Opinion in Structural Biology*, 17(1):3–14, February 2007.

[361] David Eliezer. Biophysical characterization of intrinsically disordered proteins. *Current Opinion in Structural Biology*, 19(1):23–30, February 2009.

[362] Vladimir N Uversky and A Keith Dunker. *Intrinsically Disordered Protein Analysis*. Volume 2, Methods and Experimental Tools. Springer, July 2012.

[363] D S Wishart, C G Bigam, J Yao, F Abildgaard, H J Dyson, E Oldfield, J L Markley, and B D Sykes. 1H, 13C and 15N chemical shift referencing in biomolecular NMR. *J Biomol NMR*, 6(2):135–140, September 1995.

[364] F Delaglio, S Grzesiek, G W Vuister, G Zhu, J Pfeifer, and A Bax. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR*, 6(3):277–293, November 1995.

[365] Wim F Vranken, Wayne Boucher, Tim J Stevens, Rasmus H Fogh, Anne Pajon, Miguel Llinas, Eldon L Ulrich, John L Markley, John Ionides, and Ernest D Laue. The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins*, 59(4):687–696, June 2005.

[366] James Keeler. *Understanding NMR Spectroscopy*. John Wiley & Sons, May 2013.

[367] A E Ferentz and G Wagner. NMR spectroscopy: a multifaceted approach to macromolecular structure. *Q. Rev. Biophys.*, 33(1):29–65, February 2000.

[368] Stephan Grzesiek and Ad Bax. Improved 3D triple-resonance NMR techniques applied to a 31 kDa protein. *Journal of Magnetic Resonance (1969)*, 96(2):432–440, 1992.

[369] L E Kay, G Y Xu, and T Yamazaki. Enhanced-Sensitivity Triple-Resonance Spectroscopy with Minimal H 2 O Saturation. *Journal of Magnetic Resonance*, 1994.

[370] Stephan Grzesiek and Ad Bax. An efficient experiment for sequential backbone assignment of medium-sized isotopically enriched proteins. *Journal of Magnetic Resonance (1969)*, 99(1):201–207, 1992.

[371] Stephan Grzesiek and Ad Bax. Amino acid type determination in the sequential assignment procedure of uniformly 13C/15N-enriched proteins. *J Biomol NMR*, 3(2):185–204, 1993.

[372] D R Muhandiram and L E Kay. Gradient-enhanced triple-resonance three-dimensional NMR experiments with improved sensitivity. *Journal of Magnetic Resonance*, 1994.

[373] Robert T Clubb, Venkataraman Thanabal, and Gerhard Wagner. A constant-time three-dimensional triple-resonance pulse scheme to correlate intraresidue$^{1}H^{N}$, ^{15}N, and ^{13}C' chemical shifts in ^{15}N-^{13}C-labeled proteins. *Journal of Magnetic Resonance*, pages 213–217, 1992.

[374] Wolfgang Bermel, Ivano Bertini, Luminita Duma, Isabella C Felli, Lyndon Emsley, Roberta Pierattelli, and Paul R Vasos. Complete Assignment of Heteronuclear Protein Resonances by Protonless NMR Spectroscopy. *Angew. Chem. Int. Ed.*, 44(20):3089–3092, May 2005.

[375] Jesus Angulo and Pedro M Nieto. STD-NMR: application to transient interactions between biomolecules-a quantitative approach. *Eur. Biophys. J.*, 40(12):1357–1369, December 2011.

[376] Moriz Mayer and Bernd Meyer. Characterization of ligand binding by saturation transfer difference NMR spectroscopy. *Angew. Chem. Int. Ed.*, 38(12):1784–1788, 1999.

[377] M Mayer and B Meyer. Group epitope mapping by saturation transfer difference NMR to identify segments of a ligand in direct contact with a protein receptor. *J. Am. Chem. Soc.*, 123(25):6108–6117, June 2001.

[378] Christopher A Lepre, Jonathan M Moore, and Jeffrey W Peng. Theory and applications of NMR-based screening in pharmaceutical research. *Chem. Rev.*, 104(8):3641–3676, August 2004.

[379] Tsang-Lin Hwang and A J Shaka. Water suppression that works. Excitation sculpting using arbitrary wave-forms and pulsed-field gradients. *Journal of Magnetic Resonance, Series A*, 112(2):275–279, 1995.

[380] Magnus Kjaergaard, Vytautas Iesmantavicius, and Flemming M Poulsen. The interplay between transient α-helix formation and side chain rotamer distributions in disordered proteins probed by methyl chemical shifts. *Protein Sci.*, 20(12):2023–2034, December 2011.

[381] Sharon M Kelly, Thomas J Jess, and Nicholas C Price. How to study proteins by circular dichroism. *Biochimica et Biophysica Acta (BBA) - Proteins & Proteomics*, 1751(2):119–139, August 2005.

[382] Christoph J Wienken, Philipp Baaske, Ulrich Rothbauer, Dieter Braun, and Stefan Duhr. Protein-binding assays in biological liquids using microscale thermophoresis. *Nature Communications*, 1(7):1–7, October 2010.

[383] Moran Jerabek-Willemsen, Chistoph J Wienken, Dieter Braun, Philipp Baaske, and Stefan Duhr. Molecular interaction studies using microscale thermophoresis. *Assay Drug Dev Technol*, 9(4):342–353, August 2011.

[384] Stefan Duhr and Dieter Braun. Thermophoretic depletion follows Boltzmann distribution. *Physical review letters*, 96(16):168301, 2006.

[385] Stefan Duhr and Dieter Braun. Why molecules move along a temperature gradient. *Proc. Natl. Acad. Sci. U.S.A.*, 103(52):19678–19682, December 2006.

[386] Bin Xue, Roland L Dunbrack, Robert W Williams, A Keith Dunker, and Vladimir N Uversky. PONDR-FIT: A meta-predictor of intrinsically disordered amino acids. *BBA - Proteins and Proteomics*, pages 1–15, January 2010.

[387] Jianlin Cheng, Michael J Sweredoski, and Pierre Baldi. Accurate prediction of protein disordered regions by mining protein structure data. *Data Mining and Knowledge Discovery*, 11(3):213–222, 2005.

[388] J J Ward, J S Sodhi, L J McGuffin, B F Buxton, and D T Jones. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J. Mol. Biol.*, 337(3):635–645, March 2004.

[389] Emily B Askew, Robert T Gampe, Thomas B Stanley, Jonathan L Faggart, and Elizabeth M Wilson. Modulation of androgen receptor activation function 2 by testosterone and dihydrotestosterone. *J. Biol. Chem.*, 282(35):25801–25816, August 2007.

[390] D T Jones. Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.*, 292(2):195–202, September 1999.

[391] Daniel W A Buchan, Federico Minneci, Tim C O Nugent, Kevin Bryson, and David T Jones. Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucleic Acids Research*, 41(Web Server issue):W349–57, July 2013.

[392] Gianluca Pollastri, Darisz Przybylski, Burkhard Rost, and Pierre Baldi. Improving the prediction of protein secondary structure in three and eight classes using recurrent neural networks and profiles. *Proteins*, 47(2):228–235, March 2002.

[393] V Muñoz and L SERRANO. Elucidating the folding problem of helical peptides using empirical parameters. *Nat. Struct. Biol.*, 1(6):399–409, June 1994.

[394] Fatemeh Miri Disfani, Wei-Lun Hsu, Marcin J Mizianty, Christopher J Oldfield, Bin Xue, A Keith Dunker, Vladimir N Uversky, and Lukasz Kurgan. MoRFpred, a computational tool for sequence-based prediction and characterization of

short disorder-to-order transitioning binding regions in proteins. *Bioinformatics*, 28(12):i75–83, June 2012.

[395] George D Rose, Ari R Geselowitz, Glenn J Lesser, Richard H Lee, and Micheal H Zehfus. Hydrophobicity of amino acid residues in globular proteins. *Science*, 229(4716):834–838, 1985.

[396] Chiaki Nishimura, Michael A Lietzow, H Jane Dyson, and Peter E Wright. Sequence Determinants of a Protein Folding Pathway. *J. Mol. Biol.*, 351(2):383–392, August 2005.

[397] Chiaki Nishimura, H Jane Dyson, and Peter E Wright. Consequences of Stabilizing the Natively Disordered F Helix for the Folding Pathway of Apomyoglobin. *J. Mol. Biol.*, 411(1):248–263, August 2011.

[398] H Jane Dyson, Peter E Wright, and Harold A Scheraga. The role of hydrophobic interactions in initiation and propagation of protein folding. *Proc. Natl. Acad. Sci. U.S.A.*, 103(35):13057–13061, August 2006.

[399] Daniel J Felitsky, Michael A Lietzow, H Jane Dyson, and Peter E Wright. Modeling transient collapsed states of an unfolded protein to provide insights into early folding events. *Proceedings of the National Academy of Sciences*, 105(17):6278–6283, April 2008.

[400] A V McDonnell, T Jiang, A E Keating, and B Berger. Paircoil2: improved prediction of coiled coils from sequence. *Bioinformatics*, 22(3):356–358, January 2006.

[401] Marco D Mukrasch, Stefan Bibow, Jegannath Korukottu, Sadasivam Jeganathan, Jacek Biernat, Christian Griesinger, Eckhard Mandelkow, and Markus Zweckstetter. Structural Polymorphism of 441-Residue Tau at Single Residue Resolution. *Plos Biol*, 7(2):e34, 2009.

[402] Julia Wirmer, Wolfgang Peti, and Harald Schwalbe. Motional properties of unfolded ubiquitin: a model for a random coil protein. *J Biomol NMR*, 35(3):175–186, July 2006.

[403] V Uversky and S Longhi. Instrumental Analysis of Intrinsically Disordered Proteins: Assessing Structure and Conformation. *Wiley*, 2010.

[404] Kaare Teilum, Birthe B Kragelund, and Flemming M Poulsen. Application of hydrogen exchange kinetics to studies of protein folding. *Protein folding handbook*, pages 634–672, 2008.

[405] Kurt Wüthrich. *Nmr of Proteins and Nucleic Acids*. John Wiley & Sons Incorporated, February 2008.

[406] Shang-Te Danny Hsu, Carlos W Bertoncini, and Christopher M Dobson. Use of protonless NMR spectroscopy to alleviate the loss of information resulting from exchange-broadening. *J. Am. Chem. Soc.*, 131(21):7222–7223, June 2009.

[407] Vladimir N Uversky. What does it mean to be natively unfolded? *Eur. J. Biochem.*, 269(1):2–12, January 2002.

[408] Igor N Serdyuk, Nathan R Zaccai, and Joseph Zaccai. *Methods in Molecular Biophysics*. Structure, Dynamics, Function. Cambridge University Press, March 2007.

[409] S Schwarzinger, G J Kroon, T R Foss, P E Wright, and H J Dyson. Random coil chemical shifts in acidic 8 M urea: implementation of random coil shift data in NMRView. *J Biomol NMR*, 18(1):43–48, September 2000.

[410] Alan Fersht. *Structure and Mechanism in Protein Science*. A Guide to Enzyme Catalysis and Protein Folding. Macmillan, 1999.

[411] Stephan Schwarzinger, Gerard J A Kroon, Ted R Foss, John Chung, Peter E Wright, and H Jane Dyson. Sequence-Dependent Correction of Random Coil NMR Chemical Shifts. *J. Am. Chem. Soc.*, 123(13):2970–2978, April 2001.

[412] Magnus Kjaergaard, Søren Brander, and Flemming M Poulsen. Random coil chemical shift for intrinsically disordered proteins: effects of temperature and pH. *J Biomol NMR*, 49(2):139–149, 2011.

[413] Yunjun Wang and Oleg Jardetzky. Probability-based protein secondary structure identification using combined NMR chemical-shift data. *Protein Sci.*, 11(4):852–861, April 2002.

[414] Haiyan Zhang, Stephen Neal, and David S Wishart. RefDB: a database of uniformly referenced protein chemical shifts. *J Biomol NMR*, 25(3):173–195, March 2003.

[415] Alfonso De Simone, Andrea Cavalli, Shang-Te Danny Hsu, Wim Vranken, and Michele Vendruscolo. Accurate Random Coil Chemical Shifts from an Analysis of Loop Regions in Native States of Proteins. *J. Am. Chem. Soc.*, 131(45):16332–16333, November 2009.

[416] Kamil Tamiola, Burçin Acar, and Frans A A Mulder. Sequence-specific random coil chemical shifts of intrinsically disordered proteins. *J. Am. Chem. Soc.*, 132(51):18000–18003, December 2010.

[417] Yunjun Wang and Oleg Jardetzky. Investigation of the Neighboring Residue Effects on Protein Chemical Shifts. *J. Am. Chem. Soc.*, 124(47):14075–14084, November 2002.

[418] Magnus Kjaergaard and Flemming M Poulsen. Sequence correction of random coil chemical shifts: correlation between neighbor correction factors and changes in the Ramachandran distribution. *J Biomol NMR*, 50(2):157–165, 2011.

[419] R Bussell and D Eliezer. Residual structure and dynamics in Parkinson's disease-associated mutants of alpha-synuclein. *J. Biol. Chem.*, 276(49):45996–46003, December 2001.

[420] D S Wishart, C G Bigam, A Holm, R S Hodges, and B D Sykes. 1H, 13C and 15N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects. *J Biomol NMR*, 5(1):67–81, January 1995.

[421] D S Wishart and B D Sykes. The 13C chemical-shift index: a simple method for the identification of protein secondary structure using 13C chemical-shift data. *J Biomol NMR*, 4(2):171–180, March 1994.

[422] Joseph A Marsh, Vinay K Singh, Zongchao Jia, and Julie D Forman-Kay. Sensitivity of secondary structure propensities to sequence differences between α- and γ-synuclein: Implications for fibrillation. *Protein Science*, 15(12):2795–2804, December 2006.

[423] Robert B Best, Kusai A Merchant, Irina V Gopich, Benjamin Schuler, Ad Bax, and William A Eaton. Effect of flexibility and cis residues in single-molecule FRET studies of polyproline. *Proceedings of the National Academy of Sciences*, 104(48):18964–18969, November 2007.

[424] T M Logan, Y Thériault, and S W Fesik. Structural characterization of the FK506 binding protein unfolded in urea and guanidine hydrochloride. *J. Mol. Biol.*, 236(2):637–648, February 1994.

[425] H Schwalbe, K M Fiebig, M Buck, J A Jones, S B Grimshaw, A Spencer, S J Glaser, L J Smith, and C M Dobson. Structural and dynamical properties of a denatured protein. Heteronuclear 3D NMR experiments and theoretical simulations of lysozyme in 8 M urea. *Biochemistry*, 36(29):8977–8991, July 1997.

[426] Yi Xue and Nikolai R Skrynnikov. Motion of a disordered polypeptide chain as studied by paramagnetic relaxation enhancements, 15N relaxation, and molecular dynamics simulations: how fast is segmental diffusion in denatured ubiquitin? *J. Am. Chem. Soc.*, 133(37):14614–14628, September 2011.

[427] Mikael Akke. NMR methods for characterizing microsecond to millisecond dynamics in recognition and catalysis. *Current Opinion in Structural Biology*, 12(5):642–647, October 2002.

[428] Hoang D Nguyen, Dung T Dang, Joost L J van Dongen, and Luc Brunsveld. Protein Dimerization Induced by Supramolecular Interactions with Cucurbit[8]uril. *Angew. Chem. Int. Ed.*, 49(5):895–898, 2010.

[429] Kristofer Modig, Vibeke W Jürgensen, Kresten Lindorff-Larsen, Wolfgang Fieber, Henrik G Bohr, and Flemming M Poulsen. Detection of initiation sites in protein folding of the four helix bundle ACBP by chemical shift analysis. *FEBS Lett.*, 581(25):4965–4971, October 2007.

[430] Jens K Thomsen, Birthe B Kragelund, Kaare Teilum, Jens Knudsen, and Flemming M Poulsen. Transient intermediary states with high and low folding probabilities in the apparent two-state folding equilibrium of ACBP at low pH. *J. Mol. Biol.*, 318(3):805–814, May 2002.

[431] K V Andersen and F M Poulsen. Three-dimensional structure in solution of acyl-coenzyme A binding protein from bovine liver. *J. Mol. Biol.*, 226(4):1131–1141, August 1992.

[432] Wolfgang Fieber, Sigridur Kristjansdottir, and Flemming M Poulsen. Short-range, long-range and transition state interactions in the denatured state of ACBP from residual dipolar couplings. *J. Mol. Biol.*, 339(5):1191–1199, June 2004.

[433] Susanne W Bruun, Vytautas Iesmantavicius, Jens Danielsson, and Flemming M Poulsen. Cooperative formation of native-like tertiary contacts in the ensemble of unfolded states of a four-helix protein. *Proceedings of the National Academy of Sciences*, 107(30):13306–13311, July 2010.

[434] I J McEwan. Molecular mechanisms of androgen receptor-mediated gene regulation: structure-function analysis of the AF-1 domain. *Endocr. Relat. Cancer*, 11(2):281–293, June 2004.

[435] E David Crawford and Kyle O Rove. Incomplete testosterone suppression in prostate cancer. *N. Engl. J. Med.*, 363(20):1976, November 2010.

[436] Qamar Bashir, Sandra Scanu, and Marcellus Ubbink. Dynamics in electron transfer protein complexes. *FEBS Journal*, 278(9):1391–1400, March 2011.

[437] P Anton van der Merwe and Simon J Davis. Molecular interactions mediating T cell antigen recognition. *Annu. Rev. Immunol.*, 21:659–684, 2003.

[438] Dietmar Schomburg and Ida Schomburg. Springer Handbook of Enzymes, 2002.

[439] Julia Vaynberg, Tomohiko Fukuda, Ka Chen, Olga Vinogradova, Algirdas Velyvis, Yizeng Tu, Lily Ng, Chuanyue Wu, and Jun Qin. Structure of an ultraweak protein-protein complex and its crucial role in regulation of cell morphology and motility. *Molecular Cell*, 17(4):513–523, February 2005.

[440] Tom Huxford, Dennis Mishler, Christopher B Phelps, De-Bin Huang, Lei Lei Sengchanthalangsy, Ryan Reeves, Carrie A Hughes, Elizabeth A Komives, and Gourisankar Ghosh. Solvent Exposed Non-contacting Amino Acids Play a Critical Role in NF-κB/IκBα Complex Formation. *J. Mol. Biol.*, 324(4):587–597, December 2002.

[441] Peter B Crowley and Marcellus Ubbink. Close Encounters of the Transient Kind: Protein Interactions in the Photosynthetic Redox Chain Investigated by NMR Spectroscopy. *Acc. Chem. Res.*, 36(10):723–730, October 2003.

[442] Amelie Stein, Roland A Pache, Pau Bernadó, Miquel Pons, and Patrick Aloy. Dynamic interactions of proteins in complex networks: a more structured view. *FEBS Journal*, 276(19):5390–5405, August 2009.

[443] Amelie Stein and Patrick Aloy. Contextual Specificity in Peptide-Mediated Protein Interactions. *PLoS ONE*, 3(7):e2524, July 2008.

[444] Huan-Xiang Zhou. Intrinsic disorder: signaling via highlyspecific but short-lived association. *Trends Biochem. Sci.*, 37(2):43–48, February 2012.

[445] Irene M A Nooren and Janet M Thornton. Diversity of protein-protein interactions. *The EMBO Journal*, 22(14):3486–3492, July 2003.

[446] Toby J Gibson. Cell regulation: determined to signal discrete cooperation. *Trends Biochem. Sci.*, 34(10):471–482, October 2009.

[447] Bruce Alberts, Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts, and Peter Walter. *Molecular Biology of the Cell, Fifth Edition.* Garland Science, November 2007.

[448] D L Cadena and M E Dahmus. Messenger RNA synthesis in mammalian cells is catalyzed by the phosphorylated form of RNA polymerase II. *J. Biol. Chem.*, 262(26):12468–12474, September 1987.

[449] H Lu, O Flores, R Weinmann, and D Reinberg. The nonphosphorylated form of RNA polymerase II preferentially associates with the preinitiation complex. *Proc. Natl. Acad. Sci. U.S.A.*, 88(22):10004–10008, November 1991.

[450] T O'Brien, S Hardin, A Greenleaf, and J T Lis. Phosphorylation of RNA polymerase II C-terminal domain and transcriptional elongation. *Nature*, 370(6484):75–77, July 1994.

[451] S Hausmann. Characterization of the CTD Phosphatase Fcp1 from Fission Yeast. PREFERENTIAL DEPHOSPHORYLATION OF SERINE 2 VERSUS SERINE 5. *Journal of Biological Chemistry*, 277(24):21213–21220, April 2002.

[452] R S Chambers, B Q Wang, Z F Burton, and M E Dahmus. The activity of COOH-terminal domain phosphatase is regulated by a docking site on RNA polymerase II and by the general transcription factors IIF and IIB. *J. Biol. Chem.*, 270(25):14962–14969, June 1995.

[453] Katsuhiko Kamada, Robert G Roeder, and Stephen K Burley. Molecular mechanism of recruitment of TFIIF- associating RNA polymerase C-terminal domain phosphatase (FCP1) by transcription factor IIF. *Proc. Natl. Acad. Sci. U.S.A.*, 100(5):2296–2299, March 2003.

[454] Bao D Nguyen, Karen L Abbott, Krzysztof Potempa, Michael S Kobor, Jacques Archambault, Jack Greenblatt, Pascale Legault, and James G Omichinski. NMR structure of a complex containing the TFIIF subunit RAP74 and the RNA polymerase II carboxyl-terminal domain phosphatase FCP1. *Proc. Natl. Acad. Sci. U.S.A.*, 100(10):5688–5693, May 2003.

[455] Ao Yang, Karen L Abbott, Alexandre Desjardins, Paola Di Lello, James G Omichinski, and Pascale Legault. NMR Structure of a Complex Formed by the Carboxyl-Terminal Domain of Human RAP74 and a Phosphorylated Peptide from the Central Domain of the FCP1 Phosphatase. *Biochemistry*, 48(9):1964–1974, March 2009.

[456] Karen L Abbott, Jacques Archambault, Hua Xiao, Bao D Nguyen, Robert G Roeder, Jack Greenblatt, James G Omichinski, and Pascale Legault. Interactions of the HIV-1 Tat and RAP74 proteins with the RNA polymerase II CTD phosphatase FCP1. *Biochemistry*, 44(8):2716–2731, March 2005.

[457] Karen L Abbott, Matthew B Renfrow, Michael J Chalmers, Bao D Nguyen, Alan G Marshall, Pascale Legault, and James G Omichinski. Enhanced binding of RNAP II CTD phosphatase FCP1 to RAP74 following CK2 phosphorylation. *Biochemistry*, 44(8):2732–2745, March 2005.

[458] Chad W Lawrence and Scott A Showalter. Carbon-Detected 15N NMR Spin Relaxation of an Intrinsically Disordered Protein: FCP1 Dynamics Unbound and

in Complex with RAP74. *The Journal of Physical Chemistry Letters*, 3(10):1409–1413, 2012.

[459] Christopher Wostenberg, Sushant Kumar, William G Noid, and Scott A Showalter. Atomistic simulations reveal structural disorder in the RAP74-FCP1 complex. *J Phys Chem B*, 115(46):13731–13739, November 2011.

[460] Chad W Lawrence, Alain Bonny, and Scott A Showalter. Biochemical and Biophysical Research Communications. *BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS*, 410(3):461–465, July 2011.

[461] Santiago Esteban-Martin, Jordi Silvestre-Ryan, Carlos W Bertoncini, and Xavier Salvatella. Identification of fibril-like tertiary contacts in soluble monomeric α-synuclein. *Biophysical Journal*, 105(5):1192–1198, September 2013.

[462] Jordi Silvestre-Ryan, Carlos W Bertoncini, Robert Bryn Fenwick, Santiago Esteban-Martin, and Xavier Salvatella. Average conformations determined from PRE data provide high-resolution maps of transient tertiary interactions in disordered proteins. *Biophysical Journal*, 104(8):1740–1751, April 2013.

[463] Bao D Nguyen, Hung-Ta Chen, Michael S Kobor, Jack Greenblatt, Pascale Legault, and James G Omichinski. Solution structure of the carboxyl-terminal domain of RAP74 and NMR characterization of the FCP1-binding sites of RAP74 and human TFIIB. *Biochemistry*, 42(6):1460–1469, February 2003.

[464] B T Farmer, K L Constantine, V Goldfarb, M S Friedrichs, M Wittekind, J Yanchunas, J G Robertson, and L Mueller. Localizing the NADP+ binding site on the MurB enzyme by NMR. *Nat. Struct. Biol.*, 3(12):995–997, December 1996.

[465] Travis van der Steen, Donald J Tindall, and Haojie Huang. Posttranslational modification of the androgen receptor in prostate cancer. *Int J Mol Sci*, 14(7):14833–14859, 2013.

[466] N Blom, S Gammeltoft, and S Brunak. Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. 294(5):1351–1362, December 1999.

[467] D Gioeli. Androgen Receptor Phosphorylation. REGULATION AND IDENTIFICATION OF THE PHOSPHORYLATION SITES. *Journal of Biological Chemistry*, 277(32):29304–29314, May 2002.

[468] L Szilák, J Moitra, D Krylov, and C Vinson. Phosphorylation destabilizes alphahelices. *Nat. Struct. Biol.*, 4(2):112–114, February 1997.

[469] Charles D Andrew, Jim Warwicker, Gareth R Jones, and Andrew J Doig. Effect of Phosphorylation on α-Helix Stability as a Function of Position. *Biochemistry*, 41(6):1897–1905, February 2002.

[470] Jesús Mendieta, Miguel A Fuertes, Rani Kunjishapatham, Ismael Santa-María, Francisco J Moreno, Carlos Alonso, Federico Gago, Victor Muñoz, Jesús Avila, and Félix Hernández. Phosphorylation modulates the alpha-helical structure and polymerization of a peptide from the third tau microtubule-binding repeat. *Biochimica et Biophysica Acta (BBA) - General Subjects*, 1721(1-3):16–26, January 2005.

[471] L Michel Espinoza-Fonseca, David Kast, and David D Thomas. Molecular Dynamics Simulations Reveal a Disorder-to-Order Transition on Phosphorylation of Smooth Muscle Myosin. *Biophysical Journal*, 93(6):2083–2090, September 2007.

[472] J L Smart and J A McCammon. Phosphorylation stabilizes the N-termini of alpha-helices. *Biopolymers*, 49(3):225–233, March 1999.

[473] Malgorzata Broncel, Sara C Wagner, Kerstin Paul, Christian P R Hackenberger, and Beate Koksch. Towards understanding secondary structure transitions: phosphorylation and metal coordination in model peptides. *Org. Biomol. Chem.*, 8(11):2575–2579, June 2010.

[474] Maxim V Dorovkov, Alla S Kostyukova, and Alexey G Ryazanov. Phosphorylation of Annexin A1 by TRPM7 Kinase: A Switch Regulating the Induction of an α-Helix. *Biochemistry*, 50(12):2187–2193, March 2011.

[475] P V Attwood, P G Besant, and Matthew J Piggott. Focus on phosphoaspartate and phosphoglutamate. *Amino Acids*, 40(4):1035–1051, September 2010.

[476] Regev Schweiger and Michal Linial. Cooperativity within proximal phosphorylation sites is revealed from large-scale proteomics data. *Biol. Direct*, 5(6):1–17, 2010.

[477] Florence Cordier, Alain Chaffotte, Elouan Terrien, Christophe Préhaud, Francois-Xavier Theillet, Muriel Delepierre, Monique Lafon, Henri Buc, and Nicolas Wolff. Ordered phosphorylation events in two independent cascades of the PTEN C-tail revealed by NMR. *J. Am. Chem. Soc.*, 134(50):20533–20543, December 2012.

[478] Stephen Neal, Alex M Nip, Haiyan Zhang, and David S Wishart. Rapid and accurate calculation of protein 1H, 13C and 15N chemical shifts. *J Biomol NMR*, 26(3):215–240, July 2003.

[479] Franc Avbelj, Darko Kocjan, and Robert L Baldwin. Protein chemical shifts arising from alpha-helices and beta-sheets depend on solvent exposure. *Proc. Natl. Acad. Sci. U.S.A.*, 101(50):17394–17397, December 2004.

[480] Dmitry Krylov and Charles R Vinson. Leucine zipper. *eLS*, 2001.

[481] Judit Osz, Maxim V Pethoukhov, Serena Sirigu, Dmitri I Svergun, Dino Moras, and Natacha Rochel. Solution Structures of PPARγ2/RXRα Complexes. *PPAR Research*, 2012(21):1–8, 2012.

[482] B He, J A Kemppainen, J J Voegel, H Gronemeyer, and E M Wilson. Activation function 2 in the human androgen receptor ligand binding domain mediates interdomain communication with the NH(2)-terminal domain. *J. Biol. Chem.*, 274(52):37219–37225, December 1999.

[483] Bin He and Elizabeth M Wilson. The NH(2)-terminal and carboxyl-terminal interaction in the human androgen receptor. *Molecular Genetics and Metabolism*, 75(4):293–298, April 2002.

[484] Christopher J Oldfield, Jingwei Meng, Jack Y Yang, Mary Qu Yang, Vladimir N Uversky, and A Keith Dunker. Flexible nets: disorder and induced fit in the associations of p53 and 14-3-3 with their partners. *BMC Genomics*, 9 Suppl 1:S1, 2008.

[485] Marianne D Sadar. Advances in small molecule inhibitors of androgen receptor for the treatment of advanced prostate cancer. *World J Urol*, August 2011.

[486] Paul L Shaffer and Daniel T Gewirth. Structural basis of VDR-DNA interactions on direct repeat response elements. *The EMBO Journal*, 21(9):2242–2252, May 2002.