Universitat Autònoma de Barcelona

Computational modeling of fluorescent markers for amyloid detection

Francesca Peccati

Universitat Autònoma
de Barcelona

Computational modeling of fluorescent markers for amyloid detection

Francesca Peccati

PhD thesis

PhD program in Chemistry

Supervisors: Mariona Sodupe and Xavier Solans-Monfort

Chemistry Department
Faculty of Sciences

2016

Memòria presentada per aspirar al Grau de Doctor per

Francesca Peccati

Vist i plau,

Mariona Sodupe **Xavier Solans-Monfort**

Bellaterra,

Contents

Chapter 1

Introduction

1.1 Alzheimer's disease

Alzheimer's disease is a neurodegenerative disorder and the first cause of dementia (60 - 80% of cases). It is characterized by a decline in memory, language, and cognitive skills that affects the ability to perform everyday tasks. This decline is linked to the damage or destruction of nerve cells in the areas of the brain that are devoted to these functions. Moreover, in the progress of the disease, destruction of neurons eventually affects also other parts of the brain, such as those related to walking and swallowing. Alzheimer's disease patients in the final stages are bed-bound and require constant care. Eventually, the disease is fatal.[1] Despite the fact that the discovery of this condition dates back to 1906, it was not until 70 years later that its destructive potential was understood. In 1976, it was

estimated that Alzheimer's may rank as the fourth or fifth most common cause of death in the United States.[2]

Dementia is commonly associated to old age, and for this reason it is expected to be burdensome especially for developed countries, where life expectancy is higher. However, it has been shown that dementia is really a global problem, since population ageing is occurring at an unprecedentedly fast rate also in poorer regions.[3] Overall, although nowadays cases are concentrated in the world's richest and most demographically aged countries, already the majority (62%) of people with dementia live in low and middle income countries, where access to social protection, services and care are limited. For this reason, it is expected that by 2050 the global burden of dementia will shift to poorer countries.[3]

The spreading of the disease may be reduced through improvements in population health, but it is estimated that only up to 10% of the cases may be avoided through health policies. In particular, education and other factors that enhance and cognitive skills improve the brain health of those entering old age, and can reduce the incidence of dementia.[3] Regarding Western Europe, although life expectancy at birth continues to show substantial variation between social environments across countries, it has been suggested that the number of people with dementia is stabilizing, despite population ageing. However, dementia care will still remain a a challenge of the utmost importance for many years, in particular due to the fact that the oldest age group (85 years and older), which is the most affected by dementia, is the fastest growing age group in the population.[4]

Despite the lack of a cure for Alzheimer's disease, an early diagnosis is fundamental for a proper care of the patient. In particular, early de-

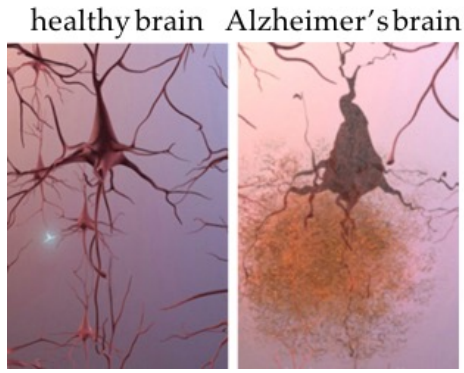


Figure 1.1: Amyloid plaques depositing in the brain of Alzheimer's disease patient. Picture: www.alz.org

tection and diagnosis allows people to access available treatments, build a care team, participate in support services, and enroll in clinical trials. Since early signs such as memory problems, confusion and personality changes may be attributable to a variety of sources, great effort must be put into the correct identification of dementia. The hallmark pathologies of Alzheimer's disease are the progressive accumulation in the brain of the protein fragment amyloid-β (plaques) outside neurons and twisted strands of the protein tau (tangles) inside neurons (Figure 1.1). These changes are responsible for the damage and death of neurons.[5]

The *post mortem* examination of brain tissues and the identification of plaques and tangles ultimately confirms the diagnosis of Alzheimer's disease. Indeed, the diagnosis performed on living patients is based on testing their neurological skills, and remains an hypothesis until amyloid plaques are eventually identified. For this reason, examination of the brain is of fundamental importance, because without access to brain material

only clinical symptoms would be used to distinguish various types of dementia. Moreover, thanks to examination of a large number of brain samples, it was possible to establish a pattern of development of the disease as it spreads from region to region. These measurements are now used as a benchmark for disease severity, allowing changes which occur early in the disease to be studied.[6]

Morover, Alzheimer's disease also results in an increase of oxidative stress, which is the imbalance between the formation and spread of reactive oxygen species, and antioxidant defenses. In particular, it has been shown that amyloid aggregates are bound to copper (II) ions, and these ions retain their redox activity and are able to degrade hydrogen peroxide. Eventually, this leads to the formation of hydroxyl radicals with high oxydizing power, which result in additional brain damage.[7–9]

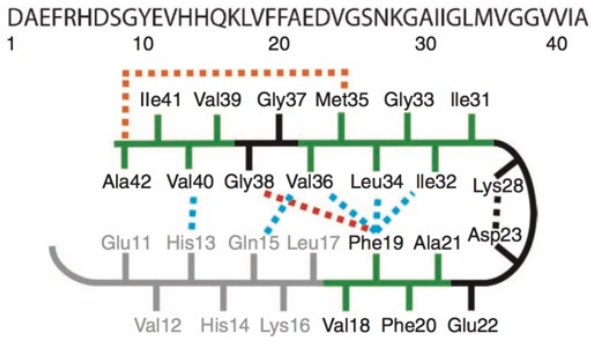
1.2 Amyloid plaques

1.2.1 Structural features of Aβ40 and Aβ42 fibrils

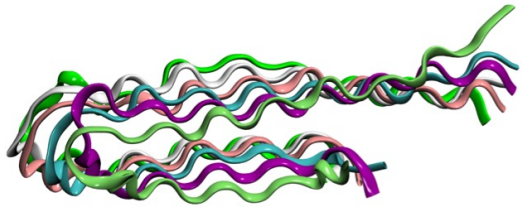
As already mentioned, cerebral plaques and neurofibrillary tangles are important pathological features of Alzheimer's disease. Plaques are fibrillar aggregates of amyloid-β peptide.[10] This hydrophobic peptide, which can be 37 to 42 residues long, is formed from the proteolytic cleavage of the amyloid precursor protein by secretases, and can abnormally accumulate in the brain eventually leading to the deposition of plaques.[10–13] Indeed, according to the *amyloid cascade hypothesis*, the deposition of the

amyloid-β peptide in the brain initiates a complex sequence of events that ultimately leads to Alzheimer's dementia. Thus, amyloid-β accumulation seems to be the central event of the disease. Since several mutations in the amyloid precursor protein have been identified and linked with hereditary Alzheimer's disease, a lot of interest has been devoted to the study of this process. Two predicted cleavages, one in the extracellular domain (β-secretase cleavage) and the other in the transmembrane region (γ-secretase cleavage), are necessary to release the amyloid-β from the precursor molecule.[11] Nearly 90% of secreted amyloid-β ends in residue 40, whereas Aβ42 accounts for roughly the remainig 10%. Only minor amounts of shorter amyloid-β peptides such (37 - 38 residues) are produced. Interestingly, some mutations in the amyloid precursor protein increase the production of Aβ42, thus increasing the Aβ42/Aβ40 ratio.[11,14]

Figure 1.2a shows the sequence and structure of the monomer unit in Aβ40 and Aβ42 fibrils. Solid state NMR measurements have provided precious insight into inter-residue interactions. In particular, for Aβ40 fibrils, residues 1 to 10 are unstructured, and for this reason often missing in crystallographic structures, while residues 11 to 40 have a cross-β architecture, with two parallel β-sheets connected by a loop involving residues 23 to 29. Moreover, side chain packing has been observed between PHE19 and ILE32, LEU4 and VAL36 and between GLN15 and VAL36 as well as between HIS13 and VAL40 (blue dashed lines).[12,15,16] Concerning Aβ42 fibrils, residues 1 to 17 may be unstructured (in gray), with residues 18 to 42 again forming a cross-β motif. Molecular contacts have been reported within the monomer unit of Aβ42 fibrils between PHE19 and GLY38 (red dashed line) and between MET35 and ALA42 (orange dashed line).[17] In both Aβ40 and Aβ42, the turn conformation is sta-



(a) Aminoacidic sequence.



(b) Protofilament: $(A\beta 40)_n$.

Figure 1.2: Sequence and structure of the monomer unit in Aβ40 and Aβ40 fibrils; representation of structural constraints in Aβ40 and Aβ42 fibrils. Picture: ref[12].

bilized by hydrophobic interactions (green residues) and by a salt bridge between ASP23 and LYS28 (black dashed line).[12] Aβ42 is less soluble than Aβ40, and is predominant in the plaques characteristic of Alzheimer's disease, even though Aβ40 is much more abundant than Aβ42.[18,19] Recently, it has been shown that *in vitro* Aβ40 and Aβ42 can form mixed fibrils. This suggests that, in *in vivo* conditions, there is some mechanism that favors Aβ42 deposition.[20] However, understanding the aggregation and toxicity mechanisms of amyloid-β species is a complex and still open problem.[21–25]

In amyloid fibrils, the single Aβ40 and Aβ42 peptides arrange regularly giving rise to protofilaments, the basic units of the fibrils, which are constituted by two parallel β sheets resulting from the lateral aggregation of the peptides (Figure 1.2b). The further aggregation of two or more protofilament units ultimately leads to the final structure of the fibril.

Much less information is available concerning amyloid-β oligomers, which represent an earlier stage of aggregation compared to amyloid-β fibrils and have been linked to neuronal disfunction. The structure and neurotoxicity of Aβ40 and Aβ42 are still to be explored. However, interestingly, it has been shown that protofibrillar aggregates, oligomers, might be more toxic than fully formed fibrils.[12,26–28]

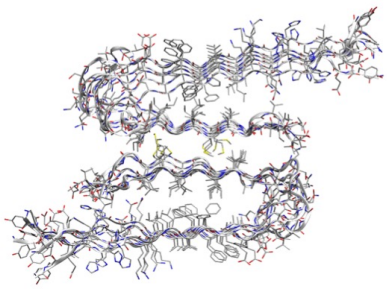
1.2.2 Polymorphism

The last decade has seen several advances in the determination of Aβ fibril structures at the atomic level. This is particularly challenging mainly due to their inherent polymorphism, *i.e.* architectures being strongly af-

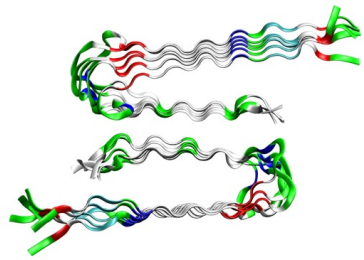
fected by the aggregation conditions, and by their scarce solubility. Both magnetic (NMR, EPR) and optical (IR, Raman) techniques have been employed in the determination of the secondary and tertiary structures of amyloid-β deposits.[17,29–34] Indeed, since the initial observation of the cross-β architecture of amyloid fibrils by X-ray christallography, many more structural traits have been identified as typical of amyloid fibrils. Amyloid fibrils have typical widths of $5 \div 15\ nm$, and are several μm long. Concerning Aβ40 fibers, whose binding properties have been investigated throughout this thesis, both 2-fold and 3-fold symmetry structures have been reported (Figure 1.3).[16,29,31] More recently (in 2016), also a 2-fold Aβ42 fibril structure has been reported.[33,34] Due to the higher availability of Aβ40 fibrillar structures over Aβ42 ones, Aβ40 fibril models have been employed in this thesis.

An interesting point about these fibrils is that the same monomer, Aβ40, or Aβ42, can aggregate into fibrils with a variety of morphologies. In all the structures that have been proposed, the basic unit is the profilament $(A\beta40)_n$ or $(A\beta42)_n$ (Figure 1.2b), and the different morphologies arise from different arrangements of protofilament units. In particular, it has been shown the predominant morphology in Aβ40 fibril samples obtained *in vitro* can be affected by subtle changes in growth conditions, such as the presence or absence of gentle agitation of the peptide solution during fibril deposition.[30] Agitation leads to 2-fold symmetry fibrils, in which protofilaments associate laterally to form striated ribbons (Figures 1.3a and 1.3b). Quiescent growth, on the other hand, leads to fibrils with a three-fold symmetry that are less prone to lateral aggregation (Figures 1.3c and 1.3d).

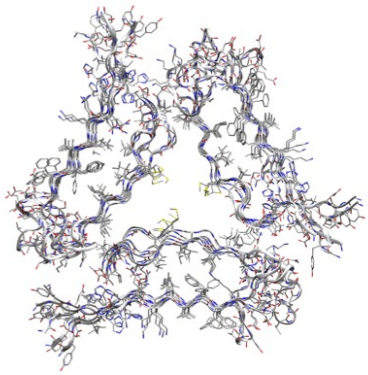
(a) PDB code: 2LMN; residues 9-40.



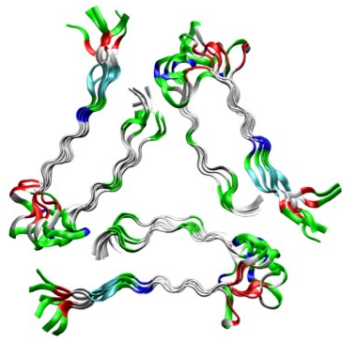
(b) Nature of residues



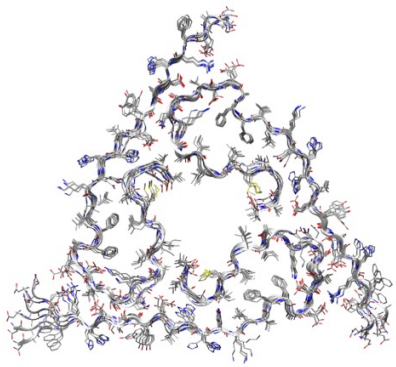
(c) PDB code: 2LMQ; residues 9-40.



(d) Nature of residues



(e) PDB code: 2M4J; residues 1-40.



(f) Nature of residues

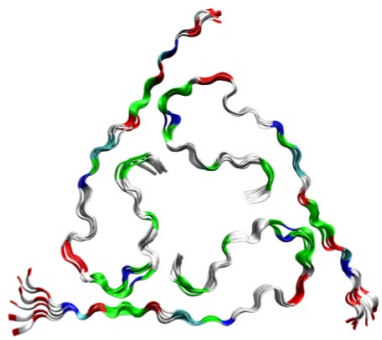


Figure 1.3: Representation of the models of amyloid-β fibrils that have been studied. Color scale for the residues: white = nonpolar, green = polar, blue = basic, red = acidic.

Concerning the examination of *ex vivo* amyloid plaques (Figures 1.3e and 1.3f), there are structural features that distinguish these fibrils from those obtained *in vitro*.[31] Comparing the structures reported in Figures 1.3c and 1.3e, it is evident that in the case of the *ex vivo* fibril, 1.3e, there is a twist in residues 19 to 23 which alters the regularity of the cross-β architecture that allows the formation of contacts that had not been observed in *in vitro* fibrillar structures. However, this structure alone cannot be taken as a reference for *in vivo* amyloid-β fibrils. This is due to the fact that is was observed that two Alzheimer's disease patients with different clinical histories both had a single predominant type of amyloid-β fibril structure, but this structure was different in the two cases. This indicates that structural variations from patient to patient are relevant, and depend on the different aggregation conditions.[31] Under this assumption, and with the aim of detecting amyloid-β fibrils in the brain of living patients, understanding these morphological variations from patient to patient is the basis for the development of structure specific markers.[31]

1.3 Amyloid-β detection

1.3.1 Nuclear methods

Currently, the method of choice for *in vivo* detection of amyloid-β fibrils is positron emission tomography (PET).[35–38] This technique involves the use of a radioactive isotope, which is injected into the body. In PET, the gamma rays emitted indirectly by a positron emitting radioactive isotope, which is bound to a chemical with known physiological properties,

are detected.[39] Since radioactive isotopes do not occur naturally in biological molecules, the synthesis of PET tracers incorporating these isotopes is challenging. Moreover, even if this technique is not invasive, it involves exposure to radiation. A related technique is SPECT, single photon emission computed tomography, in which tomographic data are acquired with a rotating gamma camera, which allows the construction of three-dimensional images.[40] The main difference between PET and SPECT is the kind of tracer that is used: while in PET a positron emitting isotope is used, and the gamma rays that are detected result from their annihilation with electrons, in SPECT the direct emission of gamma rays is observed.[39]

Pittsburgh compound-B is a successful amyloid-imaging positron emission tomography tracer. Since its first human application, PIB has been shown to be retained in areas of the brain known to contain large amounts of amyloid deposits in Alzheimer's disease.[37] Moreover, PIB is selective for fibrillar amyloid-β (K_d=1.4 nM). In contrast, it does not bind appreciably to soluble amyloid-β species, such as oligomers. This selectivity towards the pleated sheet structure of amyloid fibrils originates from the aromatic rod-like strucuture of PIB, and will be discussed in greater detail along with the properties of the related fluorescent amyloid marker thioflavin-T, with which it shares a common architecture (Scheme 1.1).[35,36,41,42]

Another PET tracer of interest, related to Alzheimer's disease but not to amyloid fibrils, is the glucose analog 2-[^{18}F]-fluoro-2-deoxy-d-glucose, which can be used to monitor brain glucose metabolism. This is important because a progressive reduction in glucose metabolism occurs years in advance of clinical symptoms Alzheimer's disease. During disease pro-

gression, glucose metabolism continues to decline, and this deterioration is associated with decreasing patient scores in cognitive tests. [43,44]

Even if it is the most common imaging technique for Alzheimer's disese, PET presents serious limitations in terms of high cost, toxicity of the radioactive isotopes, and practical limitations associated with the experimental setup and need to use the radioactive isotopes within the short time of their half-lives. For these reasons, there is a strive towards the development of alternative techniques with better performance both in terms of cost and safety. [45]

1.3.2 Fluorescence imaging

Fluorescence spectroscopy is promising alternative to nuclear techniques for amyloid-β imaging. It allows an early diagnosis of Alzheimer's disease and enables the real time visualization of biomolecules in living systems. [45,46] In general, optical imaging is a versatile technique commonly used in clinical practice, for instance for the visualization of tumors. [47,48] Recently, fluorescence microscopy and imaging have undergone a significant development due to the increasing availability of probes, that allow the study of several phenomena, including amyloid-β deposition. A parallel development has been observed in fluorescent imaging techniques, which greatly enhanced resolution and sensitivity. Of particular interest is the use of light in the near-infrared region ($600 \div 900\ nm$). This is because penetration depth increases with the decreasing wavelength of radiation, and near-infrared light has a penetration of several centimeters. [47] However, already after a few millimeters, near-infrared

light becomes diffuse due to elastic scattering, with a consequent loss of resolution. Therefore, macroscopic fluorescence imaging largely depends on spatially resolving and quantifying bulk signals from specific fluorescent entities reporting molecular activity.[47]

Several imaging modalities are available. In planar imaging, the tissue is illuminated with a plane wave, an expanded light beam, and fluorescence signals emitted toward a camera are collected. A simple and noninvasive mode to perform planar fluorescence imaging measurements is epi-illumination, in which the source shines light onto tissue surface and collects emitted light from the same side of tissue. This method, however, presents some disadvantages, including it inability to resolve depth and the fact that superficial fluorescence activity may shield more underlying optical activity. An alternative mode is trans-illumination, in which light is shined through the tissue and either the relative attenuation of light or fluorescence emission is recorded. An advantage of trans-illumination over epi-illumination is that in the first case the whole volume is sampled, while it is difficult to quantify tissue penetration in epi-illumination mode.[47]

With more sophisticated techniques, actual three-dimensional images can be generated. In optical tomography, it possible to reconstruct the internal distribution of a chromophore in a given sample. The principle of optical tomography is that tissue is illuminated at different points, and the collected light is used in combination with a mathematical formulation that describes photon propagation in tissues. A fascinating opportunity provided by optical tomography is the possibility of combining it with other imaging modalities, such as magnetic resonance imaging.[47] Moreover, several techniques are available that allow the increase of spatial res-

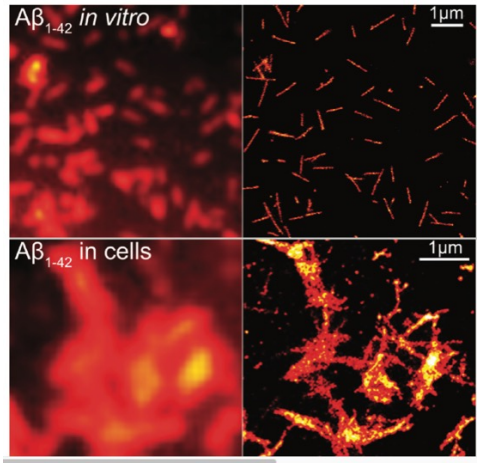


Figure 1.4: Traditional (left) and super-resolution (right) fluorescence images of Aβ42 fibrils grown *in vitro* and intracellular. Picture: ref[49].

olution of fluorescence images.[50,51] For instance, it has been shown that commercially available dyes, normally excited in the near-infrared range, can be used to obtain a near nanometer resolution. This increased spatial resolution is achieved by sequentially photoswitching individual fluorophores between a fluorescent state and a dark state. The advantage of doing so is that a temporal separation of the individual fluorophores that are emitting at the given time from the total ensemble of emitters is possible. An example of this is reported in Figure 1.4.[49–51] The fluorescence images indicate that while the resolution provided by traditional fluorescence imaging is not sufficient to provide information on the morphology of amyloid-β fibrils, with super-resolution methods this is possible.

1.3.3 Fluorescent markers

Fluorescence imaging involves the use of probes, which can be either nanoparticles or organic molecules, which undergo a variation of their fluorescence response depending of the environment and may be used to identify specifically a target, which in this case is the amyloid-β deposits typical of Alzheimer's disease. In particular, this thesis in focused on the use of organic markers, or dyes. In order to be suitable candidates for *in vivo* application, organic fluorescent markers need to fulfill a series of requirement, in terms of both optical properties and biocompatibility.[45,46] The main requirements are presented in the following list.

- Fluorescence emission in the near-infrared region;
- ability to rapidly enter the brain after intravenous injection;
- specific interaction with amyloid-β species;
- variation of the fluorescence properties when bound to amyloid-β species compared to aqueous solution.

The requirement on the fluorescence emission wavelength is related to the fact that near-infrared radiation is particularly suited for *in vivo* application to its tissue penetration, safety and the fact that it involves minimal photodamage and interference from the autofluorescence of biological matter.[45,52] The latter originates from the fact that cells contain molecules which can become fluorescent when excited with radiation falling in the visible and ultraviolet frequency range. Emission from these systems, which are mainly mitochondria, lysosomes, aromatic aminoacids

and lipopigments, is called autofluorescence. Morover, additional autofluorescence may originate from the extracellular matrix, for instance collagen or elastine. For this same reason of interference originating from living tissues, it is desirable to red shift also the absorption wavelength of the marker, for the excitation to be as selective as possible, *i.e.* ideally exciting only the dye.[52]

Regarding the ability of markers to enter the brain of patients, the fluorescent molecule essentially needs to cross the blood-brain barrier, which separates the brain from the systemic blood circulation and maintains the homeostasis of the central nervous system.[53] This is essential, since a strict homeostasis is of paramount importance for optimal brain functioning. In particular, the blood-brain barrier is a complex system formed by specialized endothelial cells that cling to brain capillaries and maintain this protection.[53] Crossing the blood-brain barrier is an open challenge for the treatment of several diseases affecting the central nervous system. The structural requirements for molecules to cross this barrier are a certain lipophilicity and the lack of ionic groups, and a small size, with an upper bound of 600 Da. Obviously, fluorescent markers should also present the least possible toxicity for the body.[46]

Concerning selectivity, fluorescent markers should be able to interact preferentially with the hydrophobic grooves arising from the cross-β structure of amyloid fibrils, which is encouraged by an hydrophobic or aromatic linear rod-like architecture. This maximizes favorable contacts between the marker and apolar residues belonging to several peptides. This means that the selectivity is strictly related to the global structure of the marker, and its interaction is with a whole class of residues, the apo-

lar ones, rather than being driven by specific marker-residues interaction. These hydrophobic residues concur in defining a channel-like or groove-like binding pocket for the dye. These structural features explain why binding is more favorable with amyloid-β fibril than with, for instance, globular proteins, which do not present these channel-like binding pockets. [46,54]

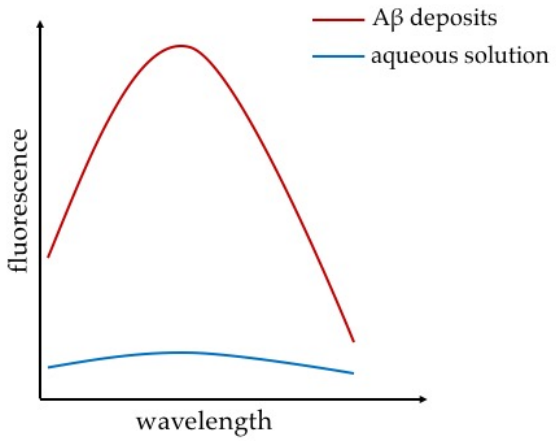


Figure 1.5: Fluorescence enhancement upon binding to amyloid-β fibrils of fluorescent markers.

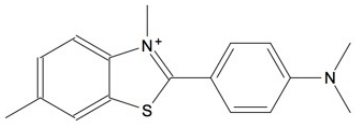
A last fundamental requirement is for the fluorescence properties of the marker to be modified by the interaction with amyloid-β fibrils. Commonly, a fluorescence enhancement is observed for the bound marker compared to the free one, but the effect can also be opposite (Figure 1.5). The modification of the quantum yield of fluorescence may also be accompanied by a displacement of the emission maximum. [46,54–58] This different behavior in water and bound to amyloid-β fibrils is essential to distinguish

between the two situations and allows the detection of signal coming exclusively from the bound dye.[46]

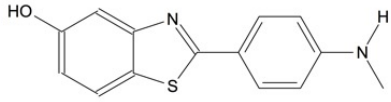
Many different families of organic dyes have been proposed over the years with different properties and performances. The following section will be devoted to a brief and non-exhaustive presentation of the features of some common markers for amyloid-β fibrils detection (Scheme 1.1 and 1.2).

Thioflavin-T

Thioflavin-T (Scheme 1.1) is regarded as the golden standard for *in vitro* and *post mortem* amyloid staining. Due to its success, this molecule has been widely studied both experimentally and computationally.[59–61]



Thioflavin-T



Pittsburgh compound B

Scheme 1.1: Molecular structures of thioflavin-T and Pittsburgh compound B.

This molecule is composed of three rigid units: a benzothiazole ring, a benzene ring and a dimethylamino group, linked by single bonds that allow torsional motion. This is important because torsion are strictly linked

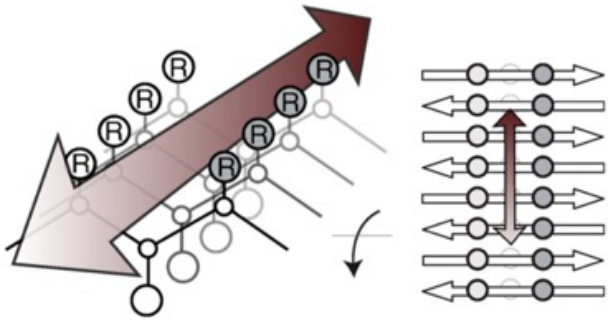


Figure 1.6: Schematic representation of the channel-like binding mode of thioflavin-T on the β-sheet of a amyloid fibril. Picture: ref[62].

to the photophysical properties of the marker. Indeed, calculations show that thioflavin-T behaves as a molecular rotor: while in the planar conformation it emits fluorescence, rotation around the single bond linking the benzothiazole to the phenyl ring yields a dark, nonfluorescent, twisted internal charge transfer excited state. This point is particularly interesting because it allows the interaction with amyloid-β fibrils to tune fluorescence emission. The free molecule in solution, indeed, will spontaneously twist, yielding the dark state, while binding to amyloid fibrils constraints the molecule to a planar conformation, which enhances fluorescence emission. Experimentally, a 1000-fold increase is observed, which leads to an optimal imaging contrast.[59–61]

In this regard, it is crucial to investigate the nature of the biding between fibrils and thioflavin-T. As reported in Figure 1.6, the cross-β architecture of amyloid fibrils involves the formation of superficial grooves, defined by the side chains of the β-sheet. Molecular dynamics simulations indicate that due to its aromatic, rod-like, architecture, thioflavin-T can

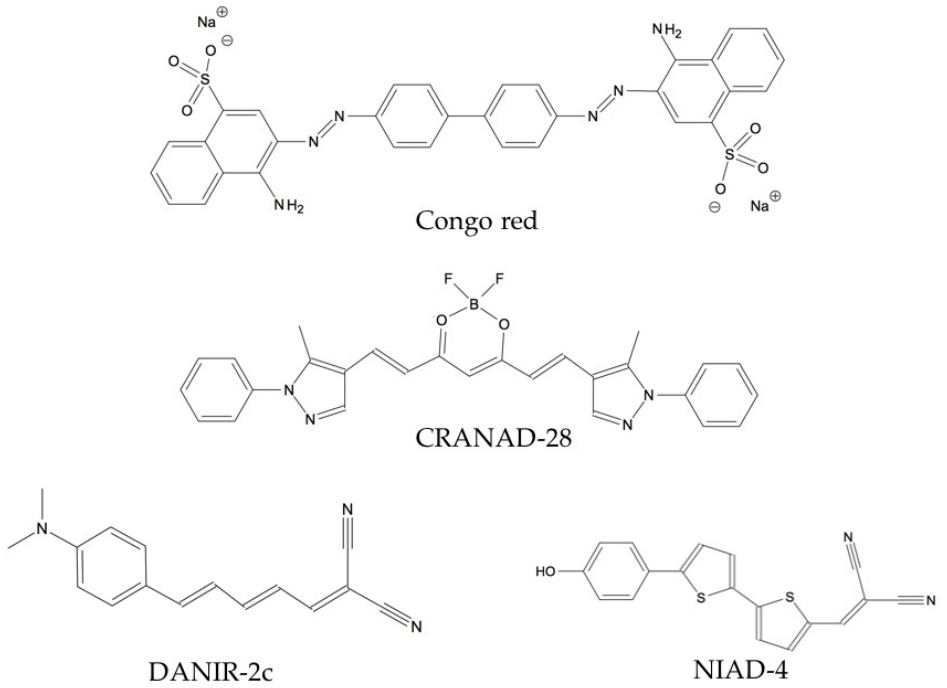
be accommodated in these channels, parallel to the long axis of the fibril. This kind of interaction is particularly favorable because it maximizes the number of side chains which can interact with the marker.[62] A similar binding is expected for all fluorescent markers that share this linear aromatic/conjugated structure, as will be discussed in Chapters 3, 4 and 5. Remarkably, thioflavin-T shares a common structure with Pittsburgh compound-B, which has similar binding properties and is employed in PET imaging (Scheme 1.1).

The 1000-fold fluorescence enhancement observed when thioflavin-T is interacting with amyloid-β fibrils has made this molecule the reference for amyloid staining *in vitro*. This fluorescence enhancement is due to the fact that upon binding the molecules is constrained to a roughly planar geometry, which prevents the rotation yielding a dark state and thus maximizes fluorescence emission. Unfortunately, thioflavin-T bears a positive charge, which makes it difficult for this marker to cross the blood-brain barrier. For this reason, it cannot be employed for *in vivo* applications. Moreover, the emission wavelength falls at $\sim$ 480 nm, which is far below the desirable 600 $\div$ 900 nm range.[59–61]

Congo red

Congo red is a diazo dye built around a central biphenyl unit, with an emission at 614 nm. It has been known to stain amyloid-β plaques since the 1960s (Scheme 1.2).[63] Similarly to thioflavin-T, molecular dynamics simulations indicate that the primary binding mode of congo red on amyloid-β fibrils is in the grooves formed by the β-sheets, and upon

binding a restriction of the torsional rotation is observed.[64] The interaction seems to involve both electrostatic and hydrophobic interactions.[65]



Scheme 1.2: Molecular structures of common amyloid markers.

Moreover, it has been reported experimentally that Congo red may inhibit the formation of amyloid fibrils. A likely mechanism of inhibition involves Congo red binding preferentially to partially folded states, stabilizing them and therefore inhibiting the formation of fibrils.[65]

Unfortunately, under physiological conditions, Congo Red penetrability through the blood-brain barrier is limited due to its molecular size

and electrostatic charge.[66] However, its derivative (trans,trans)-1-bromo-2,5-bis-(3-hydroxycarbonyl-4-hydroxy)styrylbenzene, BSB, can be employed for *in vivo* applications. Remarkably, this marker is able to distinguish between amyloid fibrils comprising peptides based on the critical carboxyl terminus of the Alzheimer's disease amyloid protein Aβ42 and those comprising the critical region of the type II diabetes pancreatic amyloid protein.[67,68]

Curcumin based markers

Curcumin is a natural yellow compound produced by some plants which served as reference for the design and synthesis of molecules of interest for Alzheimer's disease's diagnosis and treatment. As already mentioned, along with the deposition of amyloid-β fibrils, an important pathological hallmark of the disease is the increased oxidative stress. This originates from brain inflammation and is linked to the excessive production of amyloid-β.[69] Curcumin ((1E,6E)-1,7-bis(4-hydroxy- 3-methoxyphenyl)-1,6-heptadiene-3,5-dione) has been shown to have an antioxidant and anti-amyloidogenic effect.[70,71]

Indeed, polymerization assays of both Aβ40 and Aβ42 in presence of increasing concentrations of curcumin indicate that this molecule interacts with amyloid-β and inhibits fibril aggregation. Moreover, it has been shown to be able to destabilize preexisting fibrils, even though not enough to produce a depolymerization.[71] Furthermore, curcumin presents a fluorescence emission which is sensitive to solvent polarity and falls at 524 nm in acetonitrile.[72] This fluorescence emission is too low to be useful in flu-

orescence imaging *in vivo*, but the excellent properties of curcumin can be employed synthesizing derivatives with improved optical performances.

In this regard, an interesting family of markers is that of CRANADs (Scheme 1.2). Among these markers, CRANAD-58 and CRANAD-28 seem particularly promising.[55–57,73] Both markers have been designed specifically to bind the hydrophobic fragment Aβ13-20 of the amyloid-β peptide. CRANAD-58 can bind both insoluble amyloid-β aggregates and soluble species (oligomers, dimers, monomers), and in all cases the binding is accompanied by a fluorescence enhancement. A blue-shift is observed in the fluorescence maximum of the bound marker compared to the free marker in aqueous solution, and, interestingly, small differences are observed among the spectra recorded for the molecule bound to different amyloid species. Thus, a single molecule could help discriminate between different types of amyloid-β aggregates.[57] CRANAD-28, on the other hand, has a peculiar spectral response to binding to amyloid fibrils: a significant reduction of fluorescence emission is observed upon binding to amyloid fibrils. This behavior is exactly opposite to that of most fluorescent markers that have been proposed until now, which undergo a fluorescence enhancement upon binding. This phenomenon is not unprecedented but its origin is still unknown.[74,75]

A further point that makes CRANAD-28 particularly interesting is that it may be employed for two-photon imaging.[57] The key difference between two-photon absorption and a standard fluorescence imaging technique lies in the process of light absorption: two photons are absorbed simultaneously (within less than 1 fs), and combine their energies to promote the electronic transition. For this reason, low energy light (700 $\div$

1000 nm) may be employed to excite transitions in the visible range of the electromagnetic spectrum. As already mentioned, near-infrared light is convenient in terms of penetration depth and interference with the autofluorescence of biological matter.[76,77]

Along with this possible two-photon spectroscopy application, CRANAD-28 also presents promising properties in terms of inhibition of the crosslinking of amyloid-β. The mechanisms underlying amyloid aggregation in Alzheimer's disease are not fully uncovered,[21,23–25] but it is known that aggregation and crosslinking may be favored by metals such as copper.[21,78] The anticrosslinking effect of CRANAD-28 both alone and in presence of copper ions was assessed, and results indicate that CRANAD-28 is effective in both conditions. It has been hypothesized that the pyrazole moiety of the marker can coordinate with copper, competing with histidines 13 and 14 of the amyloid peptide, and that this reduced availability of copper ions may be the fundamental reason behind the reduced crosslinking in the presence of copper ions.

An important structural modification of CRANAD markers compared to curcumin is the incorporation of a difluoroboronate into the diketone group. This addition of boron has the effect of red-shifting the emission wavelength, and has been reported for several markers.[79–81] Moreover, this modification has the effect of "locking" the ring and preventing non-radiative deactivation through excited state proton transfer, which is observed in curcumin.[82] The activity of these molecules has not been addressed with computational methods yet, but some studies have been conducted on a related class of markers, the BODIPYs.

Boradiazaindacene markers

Boron-dipyrromethene markers (BODIPYs) share with CRANAD markers the presence of the difluoroboron unit. Since their discovery in 1968, these markers have found a wide range of applications in the fields of imaging and light harvesting.[83] These markers offer many advantages in terms of versatility: their optical and coordination properties can be fine-tuned with an appropriate substitution, leading to a wide range of compounds. For example, they have been employed in dye-synthesized solar cells[84], as green-light synthesizers for lanthanide-based near-infrared emitters in medical diagnosis[85], but also as markers themselves. In particular, far red and near-infrared BODIPYs can act as fluorescent pH probes, but can also be employed for labelling in living systems.[86]

In the field of Alzheimer's disease diagnosis, BAP-1 has been proposed, a BODIPY based probe for the imaging of amyloid-β plaques *in vivo*.[81] Despite the somewhat short emission wavelength of this molecule (648 nm) for *in vivo* applications, BAP-1 is selective towards amyloid-β fibrils. Indeed, a fluorescence enhancement is observed in a solution containing aggregated amyloid-β, while the fluorescence of the isolated molecule is not affected by the presence of bovine serum albumin. Once more, this is strictly dependent on molecular structure, which favors the recognition of the linear, hydrophobic channels that run at the core of amyloid fibrils.

Moreover, an appropriate substitution on the BODIPY boradiazaindacene core allows the control of the photodynamics of these markers.[87] The effect that the molecular structure exerts of the optical properties of BODIPY markers has been also been addressed computationally, by accu-

rate calculation of the absorption and emission spectra of several molecules.[88]

Bithiophene based markers

The first marker of this class, NIAD-4, was proposed in 2005 and represents a milestone in the field of *in vivo* amyloid detection (Scheme 1.2).[45] This molecule is composed of a central bithiophene body with a donor and an acceptor group on either sides of the π system (push-pull) and it undergoes a charge transfer transition. Due to this architecture, this molecule combines two favorable features: i) it readily crosses the blood-brain barrier after intravenous injection and ii) it emits in the near-infrared despite the low molecular weight. Moreover, NIAD-4 undergoes a 400-fold fluorescence increase when bound to amyloid-β fibrils compared to the free molecule in aqueous solution, offering a good imaging contrast for fluorescence imaging.

For this reason, bihiophene-based markers have attracted a lot of attention, culminating with the synthesis of two derivatives, NIAD-11 and NIAD-16, which were designed for improved optical performances.[46,54,89] Indeed, the structural variations introduced in these molecules effectively red-shift both the absorption and emission wavelength, which is now deeper into the near-infrared region. This effect is due to the extension of the conjugated π system for NIAD-11, and to an increase of the charge transfer character of the electronic transition responsible for fluorescence emission for NIAD-16. These modifications, however, also affect significantly the fluorescence increase observed when the marker binds to amy-

loid fibrils, which reduces by a factor of $\sim$ 40 in NIAD-11 compared to NIAD-4.[46,54,89] This information suggests that small structural modifications can have dramatic effects on the properties of a fluorescent molecule.

However, despite the success of these markers, little computational effort has been devoted to uncovering the photophysical processes behind their performance.[90] Chapter 3 is devoted to the study of their photophysics as isolated molecules, while Chapter 5 describes their interaction with amyloid-β fibrils, providing not only an explanation for their effectiveness as amyloid markers, but also general conclusions on the relationship between structural, optical and binding properties of conjugated push-pull markers, which could serve the purpose of designing new markers with tailored performances.[91,92]

Conjugated double bond markes

The conjugated double bond markes of the DANIR family have a push-pull architecture, with the donor and acceptor groups separated by a variable number of conjugated double bonds.[58] The simple linear architecture of these markers allows them to cross the blood brain barrier, while their donor-acceptor nature results in an emission wavelength in the near-infrared region for DANIR-2c (Scheme 1.2). These features make this family of markers very interesting for *in vivo* application.[58,93] Binding assays *in vitro* indicated that DANIR-2c undergoes a $\sim$ 10-fold fluorescence enhancement in presence of Aβ42 fibrils, while a lesser enhancement is observed in presence of bovine serum albumin, which is a globular protein. Again, this partial selectivity is due to the linear, hydrophobic, architecture

of these molecules, which interacts effectively with the apolar, channel-like, motifs of the cross-β structure of amyloid fibrils, and less favorably with bovine serum albumin, which does not offer the same kind of binding pocket. Despite the promising properties, these molecules have not been studied enough to provide a complete understanding of their performance, and for this reason have been chosen for a computational analysis in this thesis.

More recently, a second class of markers with the same architecture has been proposed, in which the single phenyl ring is replaced by a naphtalene.[93] This substitution is accompanied by a remarkable improvement of the optical properties, with a red-shift of the emission wavelength recorded in dichloromethane and bound to amyoid-β fibrils.[93] Moreover, a pronounced increase in the quantum yield of fluorescence is observed comparing these new markers with the parent molecules. Finally, one of these markers undergoes a $\sim$ 700-fold fluorescence enhancement upon binding to amyloid-β fibrils. This increase, which is comparable to that of thioflavin-T (1000-fold), suggests that this molecule will attract attention for *in vivo* application.[93]

The optical and amyloid staining properties of DANIR markers have been studied and will be presented in Chapter 4, while their binding properties will be presented in Chapter 5.

1.4 Aims

This thesis is aimed at providing a complete description of the activity of fluorescent markers for amyloid-β detection using computational techniques. Previous studies have focused either on the deactivation processes of the isolated marker or on its interaction with amyloid deposits. This is due to the fact that these two phenomena require quite different computational approaches. The calculation of the excited states of the markers requires methods based on quantum mechanics, such as TDDFT and CASPT2, and is affordable only for small systems, typically the isolated molecule. The study of the binding to amyloid fibrils, on the other hand, involves calculations on much larger systems, which are performed with the methods of molecular mechanics.

This thesis tries to combine the two aspects, describing both the photophysics and binding of two classes of amyloid markers. In particular, it is crucial to understand how does the interaction with amyloid fibrils affect the relative efficiency of the decay pathways of the markers, or, in other words, how does this interaction affect the fluorescence emission of the molecule (Chapters 3, 4 and 5). Since amyloid-β fibril structures have become available only recently, this is a novelty in the field.[17,29–34]

Indeed, in general it is understood that the interaction with amyloid-β deposits results in a spatial constrain of the molecule, accompanied by a marked reduction of its flexibility. For this reason, it is hypothesized that the fluorescence enhancement that is commonly observed when these molecules bind to amyloid fibrils is due to a reduced efficiency of non-radiative decay processes, which are normally activated by torsional mo-

tions, due to spatial confinement. These processes, however, had not been described in detail before these studies.

Overall, for each class of fluorescent markers, the subsequent approach has been followed:

- calculation of the absorption spectrum and emission wavelengths of the marker in solution;
- exploration of the nonradiative deactivation pathways of the molecules;
- identification of the binding poses of the markers on models of amyloid fibrils.

A second part of this thesis has been devoted to surface hopping simulation of an artificial nucleobase. These results are presented in Chapter 6 and show how for small systems dynamics methods represent a valid alternative for the study of excited states to the static methods presented in the first part of the thesis. Dynamic simulations elucidate the relationship between dynamics, structure and deactivation processes of isolated molecules. Due to the high computational cost of surface hopping simulations, a simple system has been studied, 5-bromouracil. A brief introduction on nucleobases will be presented in Chapter 6.

Chapter 2

Computational methods

This Chapter presents an overview of the computational methods employed in this thesis, divided in four main sections: i) *electronic structure methods*, aimed at solving the electronic Schrödinger's equation, ii) *force field methods*, where the calculation of the electronic energy is bypassed by parametrizing the potential energy of the system under study as a function of nuclear coordinates, iii) *molecular dynamics* simulations, which are used to sample the configuration space of the system, and iv) the *protein energy landscape exploration* (PELE) method, which is used to identify the binding poses for a system composed of a protein and a small ligand.

2.1 Electronic structure methods

2.1.1 Foundations

Schrödinger equation is the foundation of quantum mechanics. It is a partial differential equation describing how the wavefunction of a given system evolves over time.

$$\hat{H}\Psi(\boldsymbol{q},t) = -\frac{\hbar}{i}\frac{\partial\Psi}{\partial t} \tag{2.1}$$

The Hamiltonian operator $\hat{H}$ contains a kinetic and a potential term. According to Born's probabilistic interpretation, the squared modulus of the wavefunction $\mid \Psi(\boldsymbol{q},t) \mid^2$ is the probability distribution function, meaning that the probability of a particle's having coordinates between $\boldsymbol{q}$ and $d\boldsymbol{q}$ at the time t is given by $\mid \Psi(\boldsymbol{q},t) \mid^2 d\boldsymbol{q}$.

Most of the computational methods presented in this thesis deal with physical situations is which the probability distribution function does not vary over time (*stationary state*). These methods rely on the time-independent Schrödinger equation, which can be derived from the time-dependent one by a separation of variables. Indeed, the total wavefunction, which depends on both the spatial coordinates of the particle and on time, can be written as the product of a time-independent function (stationary wavefunction) and an exponential function of time.[94,95]

The time-independent Schrödinger equation assumes the form reported in eq. 2.2.

$$\hat{H}\Psi(\boldsymbol{q}) = E\Psi(\boldsymbol{q}) \tag{2.2}$$

This equation describes the stationary states of the system. The electronic structure methods presented in the following sections strive to solve this equation for complex multi-body systems. The exact solutions to Schrödinger equation are not known for most systems of chemical interest. In practice, *approximate* solutions to this equation are calculated.

The Born-Oppenheimer approximation

As mentioned, the core of many *ab initio* methods for the electronic structure calculations is finding an approximate solution to the time- independent Schrödinger equation.

$$\hat{H}\Psi(\boldsymbol{r}, \boldsymbol{R}) = E\Psi(\boldsymbol{r}, \boldsymbol{R}) \tag{2.3}$$

The wavefunction Ψ, eigenfunction of the Hamiltonian operator $\hat{H}$, depends on both nuclear ($\boldsymbol{R}$) and electronic ($\boldsymbol{r}$) coordinates of the system under study. Several approximations need to be introduced to solve this equation, each affecting the quality of the result.

An important approximation that underlies many computational methods is the *Born-Oppenheimer* approximation, which states that the total wavefunction $\Psi(\boldsymbol{r}, \boldsymbol{R})$ may be factorized, *i.e.* it may been written as the product of a nuclear wavefunction χ_n, depending on nuclear coordinates,

and an electronic wavefunction $\Psi_i(\boldsymbol{r};\boldsymbol{R})$ depending on electronic coordinates and, parametrically, on nuclear coordinates.

This separation implies that nuclear and electronic motions are independent. This is not completely true, but is a reasonable approximation, because since nuclei are much more massive than electrons, they move considerably slower, and electrons can readjust their motion almost instantly to match the new nuclear position, *i.e.* that electrons move in a *frozen nuclei* picture. For this reason, the electronic Hamiltonian $\hat{H}_{el}$ may be defined as the sum of an electronic kinetic energy term, $\hat{T}_e$, an electron-nucleus interaction term $\hat{V}_{en}$, an electron-electron interaction term $\hat{V}_{ee}$ and a nucleus-nucleus interaction term $\hat{V}_{nn}$ which is constant for the given nuclear geometry.

$$\hat{H}_{el} = \hat{T}_e + \hat{V}_{en} + \hat{V}_{ee} + \hat{V}_{nn} \tag{2.4}$$

The solution of the corresponding electronic Schrödinger equation

$$\hat{H}_{el}\Psi_i(\boldsymbol{r};\boldsymbol{R}) = U_i(\boldsymbol{R})\Psi_i(\boldsymbol{r};\boldsymbol{R}) \tag{2.5}$$

$$U_i(\boldsymbol{R}) = E_{el,i} + V_{nn} \tag{2.6}$$

$U_i(\boldsymbol{R})$ defines a *potential energy surface* (PES) for nuclear motion, so that the nuclear Schrödinger equation may we written as in eq. 2.7.

$$(\hat{T}_n + U_i(\boldsymbol{R}))\chi_{ni} = E_{tot}\chi_{ni} \tag{2.7}$$

This means that, according to the Born-Oppenheimer approximation, nuclei move on a single potential energy surface which is the solution to the electronic Schrödinger's equation.[94]

When dealing with the study of deactivation processes of excited molecules, more than one potential energy surface E_i is involved, as energy is transferred between different excited states, and the Born-Oppenheimer approximation is no longer valid. In such cases, indeed, the energy differences and *couplings* between different excited states regulate the deexcitation. A non-Born-Oppenheimer treatment of such phenomena is computationally highly demanding, and thus unfeasible for molecules of a certain size. In the so-called *surface hopping* dynamics, which is described in Section 2.2.3, the Born-Oppenheimer approximation is reintroduced by propagating nuclear motion on a single potential excited state energy surface at a time. At each step of the dynamics, the coupling between the current and neighboring states is computed, allowing the system to "hop" between states based on the comparison between the transition probability and a random number, thus effectively propagating nuclear motion on a series of adiabatic surfaces representing different electronic states.

Variational method

The variational method may be viewed as an alternative formulation of Schrödinger's equation, providing a route to approximate its ex-

act solution at the desired degree of accuracy. Given a compact form of Schrödinger's equation

$$\hat{H}\Psi = E\Psi \tag{2.8}$$

it is possible to define a functional $\epsilon[\zeta]$

$$\epsilon[\zeta] = \frac{\langle \zeta \mid \hat{H} \mid \zeta \rangle}{\langle \zeta \mid \zeta \rangle} \tag{2.9}$$

where $\hat{H}$ is the exact Hamiltonian of the system and ζ an arbitrary and "well-behaved" trial function of the system coordinates (quadratically integrable, single-valued, continuous) over the configurational space of the system whose exact wavefunction is Ψ.

If the trial function is exactly the ground-state wavefunction of the system, then $\epsilon[\zeta]$ is the exact energy. If, on the other hand, the trial function is not identical to the exact wavefunction, then the *variational principle* states that $\epsilon[\zeta]$ is an approximation of the exact energy E.[95] From an operational point of view, this provides a way to build an approximate wavefunction for the system, because it translates to finding a function ζ for which the functional

$$\epsilon[\zeta] = \langle \zeta \mid \hat{H} \mid \zeta \rangle \tag{2.10}$$

is stationary, meaning that it differential $\delta\epsilon$ is zero. In practice, the energy computed according to eq. 2.10 is considered an upper bound of the exact

energy E. Interestingly, this is not completely correct because imposing that $E[\zeta]$ is stationary is a necessary but not sufficient condition for $E[\zeta]$ to be a minimum. In most physical applications, however, $E[\zeta]$ actually is a minimum, which grants the success of this theory.[95] Requiring that $E[\zeta]$ is stationary is done with a constrained optimization using the method of the Lagrangian multipliers which will be briefly discussed in the next section in relation to the Hartree-Fock method.

2.1.2 Hartree-Fock method

As mentioned in the previous section, the variational principle provides an operational procedure to find an approximate solution to the electronic Schrödinger equation of a given multi-electron system, *i.e.* by requiring the functional $\langle \zeta \mid \hat{H} \mid \zeta \rangle$ to be stationary. This is done iteratively in the so-called self-consistent field method (SCF). The trial wavefunction ζ of the multi-electron system has to meet, among others, two important requirements: i) due to the probabilistic interpretation of the wavefunction, the trial function must be *normalized*, so that its squared modulus is unitary, and ii) the trial wavefunction must be *antisymmetric* with respect to the exchange of two electrons. In the simplest of these methods, the Hartree-Fock method, the wavefunction is approximated by a single *Slater determinant* wavefunction Φ_{SD}.

$$\Phi_{SD} = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_1(1) & \phi_2(1) & \dots & \phi_N(1) \\ \phi_1(2) & \phi_2(2) & \dots & \phi_N(2) \\ \dots & \dots & \dots & \dots \\ \phi_1(N) & \phi_2(N) & \dots & \phi_N(N) \end{vmatrix} \tag{2.11}$$

The columns ϕ_i of the Slater determinant are orthonormal one-electron wavefunctions (*spin-orbitals*), product of a spatial orbital and a spin function. The electronic Hamiltonian operator for this system, reads $\hat{H}_e = \hat{T}_e + \hat{V}_{ne} + \hat{V}_{ee} + \hat{V}_{nn}$. The nucleus-nucleus repulsion $\hat{V}_{nn}$ does not depend on the electron coordinates and is constant for the given nuclear configuration; the nucleus-electron attraction $\hat{V}_{ne}$ and kinetic $\hat{T}_e$ terms depend on only one electron coordinate, and the electron-electron repulsion $\hat{V}_{ee}$ depends on two electron coordinates.

In matrix form, the energy of a Slater determinant may be written as follows.

$$E = \sum_{i}^{N} \langle \phi_i \mid \hat{h}_i \mid \phi_i \rangle + \frac{1}{2} \sum_{ij}^{N} (\langle \phi_j \mid \hat{J}_i \mid \phi_j \rangle - \langle \phi_j \mid \hat{K}_i \mid \phi_j \rangle) + V_{nn} \quad (2.12)$$

$\hat{h}$ is a one-electron operator $\hat{h}_i = -\frac{1}{2}\nabla_i^2 - \sum_{i=a}^{N_{nuclei}} \frac{Z_a}{|\boldsymbol{R}_a - \boldsymbol{r}_i|}$ which describes the motion of the electron i in the field generated by all nuclei, while $\hat{J}$ and $\hat{K}$ are two-electron operators that account for the electron-electron repulsion. $\hat{J}$ is the *Coulomb operator*, with $\hat{J}_i \mid \phi_j(2)\rangle = \langle \phi_i(1) \mid \frac{1}{|\boldsymbol{r}_i - \boldsymbol{r}_j|} \mid \phi_i(1)\rangle \mid \phi_j(2)\rangle$, and its matrix element J_{ij} represents the classical repulsion between the charge distributions represented by $\phi_i^2(1)$ and $\phi_j^2(2)$. $\hat{K}$ is the *exchange operator*, with $\hat{K}_i \mid \phi_j(2)\rangle = \langle \phi_i(1) \mid \frac{1}{|\boldsymbol{r}_i - \boldsymbol{r}_j|} \mid \phi_j(1)\rangle \mid \phi_i(2)\rangle$. This operator has no classical analogy and permutes the function labels on the right-hand side of the $\frac{1}{|\boldsymbol{r}_i - \boldsymbol{r}_j|}$ term. It has no simple physical interpretation, and results from the antisymmetric character of the Slater determinant and its orbital product form.[94]

The self-consistent field procedure arises from the variational method by performing a constrained optimization of the energy E with the method of the *Lagrangian multipliers*, where the constraint is for the spin-orbitals to be orthonormal. This leads to the pseudo-eigenvalue equation reported in eq. 2.13.

$$\hat{F}_i \phi'_i = \epsilon_i \phi'_i \tag{2.13}$$

In this equation, $\hat{F}_i = \hat{h}_i + \sum_j^N (\hat{J}_j - \hat{K}_j)$ is the one-electron Fock operator, which describes the energy of an independent electron interacting with all the nuclei and with an average field generated by all the other electrons. In this sense, the Hartree-Fock theory is a *mean field* theory. This operator acts on a set of *canonical* molecular orbitals yielding ϵ_i, the orbital energies. This is a pseudo-eigenvalue equation because a specific Fock operator $\hat{F}_i$ may be determined only if all other occupied orbitals are known. From this dependence arises the need for an iterative procedure. From a set of trial molecular orbitals, the Fock operator is constructed and its eigenvalues and eigenfunctions computed. The eigenfunctions are then used to construct a new Fock operator, and so on, until the calculation is converged.[94]

According to the restrictions imposed to the molecular orbitals used for building the Slater determinant, the Hartree-Fock method exists in several variants. As aforementioned, spin-orbitals are the product of a spatial part and a spin function. If there is no restriction on the form of the spatial orbitals associated to α and β electrons, *i.e.* $\phi_{i,\alpha} \neq \phi_{i,\beta}$, the method takes the name of unrestricted Hartree-Fock (UHF); if, on the contrary, it

is required that the spatial orbital for α and β electrons is the same, *i.e.* $\phi_{i,\alpha} = \phi_{i,\beta}$, then the method takes the name of restricted Hartree-Fock (RHF). In the restricted open shell Hartree-Fock method (ROHF), the spatial partof α and β orbitals is the same, but there is a different number of α and β electrons.[94]

Basis sets

In section 2.1.2, the energy of the Slater determinant has been written in matrix form. This is because the most common approach to solve the self-consistent field equation of the Hartree-Fock method is to expand the molecular orbitals ϕ_i in a linear combination of known functions. These known functions, that among other options can be exponentials, Gaussians or plane waves, are chosen according to two criteria: i) that they provide a good representation of the physics of the problem and ii) that the calculation of the one-electron and two-electron integrals defined in eq. 2.12 is computationally efficient.

For non-periodic systems, such as molecules, Gaussian functions are the natural choice, because they give a reasonable description of the physics of the system and their integration is easy. The generic molecular orbital can thus be expressed as $\phi_i = \sum_{\alpha}^{M} c_{\alpha,i}\chi_\alpha$, where χ_α is an *atomic orbital* which is itself a linear combination of *primitive* Gaussian functions. The Hartree-Fock equations reported in eq. 2.13 may be rewritten in terms of atomic orbitals.[94]

$$\hat{F}_i \sum_{\alpha}^{M} c_{\alpha,i}\chi_\alpha = \epsilon_i \sum_{\alpha}^{M} c_{\alpha,i}\chi_\alpha \tag{2.14}$$

After left-multiplication by a specific basis function and integration, the matrix form $\boldsymbol{FC} = \boldsymbol{SC\epsilon}$ represents the so-called *Roothan-Hall equations*, where $F_{\alpha\beta} = \langle \chi_\alpha \mid \boldsymbol{F} \mid \chi_\beta \rangle$ is the Fock matrix element and $S_{\alpha\beta} = \langle \chi_\alpha \mid \chi_\beta \rangle$ the overlap matrix element in the given basis.[94] The Roothan-Hall equations are the operative translation of the Hartree-Fock theory and allow its application for a given system and basis set.

Plane waves represent another common choice of basis functions. They are periodic functions that can be written in terms or complex exponentials or sine-cosine functions, $\phi(x) = Ae^{ikx} + Be^{-ikx}$ or $\phi(x) = Acos(kx) + Bsin(kx)$. Due to their periodic nature, plane waves are particularly suitable for the description of periodic systems.

The Gaussian[96] calculations presented in this thesis have been performed with the split-valence double-zeta basis set 6-31+G(d,p) with diffuse and polarization functions.[97] CP2K calculations use a mixture of Gaussian functions and plane waves.[98] Gaussian functions are used to represent the wavefunction, while plane waves are used to represent the electron density. This dual approach is advantageous in term of the computational effort. A Gaussian functions double zeta basis set has been employed[99] in combination with a pseudopotential.[100,101] The electron density has been described with a plane wave basis set defined by a cutoff of 300 eV. The Molcas[102] calculations presented in this thesis have been performed with a basis set of atomic natural orbitals (ANO).[103–105] In particular, an ANO-S basis set has been employed with contraction [$2s1p$] for

H, [$3s2p1d$] for C, N, O and [$4s3p2d$] for S, corresponding to a double-zeta basis set with polarization functions.

2.1.3 Post-Hartree-Fock methods

Electron correlation

The Hartree-Fock method is the starting point for a wide range of computational techniques. As already mentioned, the Hartree-Fock method (section 2.1.2) is a mean field method because it describes the interaction of each electron with an average field generated by all the remaining electrons. In this sense, it lacks the description of the instant correlation between the single electron and each of the others. The so-called *correlation energy* can be defined as the difference between the true total energy of the system and the Hartree-Fock limit (Hartree-Fock energy extrapolated to the complete basis set limit), and may be divided into two contributions. The first is named *static correlation*, and arises when a system has more than one electronic configurations that have similar weight. Since the Hartree-Fock wavefunction consists of a single determinant, it naturally fails in the description of systems which require the inclusion of more than one configuration for an accurate description. The *dynamical correlation*, on the other hand, is the energy change associated to correlating the actual instantaneous motion of the electrons.[94]

Increasing fractions of correlations energy can be recovered systematically by using methods based either on the variational principle or on perturbation theory, either by expanding the wavefunction in more deter-

minants, or by including higher order perturbation terms respectively. The computational cost of these methods, which are generically labeled as *post-Hartree-Fock* and are presented in the following sections, scales rapidly with the dimension of the system, making calculations burdensome for systems of a certain size. A completely different approach to the problem of recovering the electronic correlation is constituted by the methods based on the density functional theory, which are presented in a separate section.

Configuration interaction method

The configuration interaction (CI) method is based on the Hartree-Fock method, and uses a trial wavefuntion which is a linear combination of determinants. These determinants are built from the Hartree-Fock determinant by keeping the molecular orbitals fixed and exciting the electrons.

$$\Psi_{CI} = a_0\Phi_{HF} + \sum_S a_S\Phi_S + \sum_D a_D\Phi_D + \cdots = \sum_i a_i\Phi_i \tag{2.15}$$

In eq. 2.15, only the coefficients of the linear combination are optimized, while molecular orbitals are kept fixed.[94] Determinants are grouped according to the excitation with respect to the Slater determinant, S being the singly excited, D the doubly excited, and so on. Again, the variation method is employed to obtain a set of working equations in matrix form.

$$(\boldsymbol{H} - E\boldsymbol{I})\boldsymbol{a} = \mathbf{0} \tag{2.16}$$

$\boldsymbol{H}$ is the Hamiltonian matrix with elements $H_{ij} = \langle \Phi_i \mid \hat{H} \mid \Phi_j \rangle$ and $\boldsymbol{a}$ the coefficients vector of elements $(a_0, a_1, a_2, ...)$. In practice, since the eigenfunctions of the Hamiltonian must also be eigenfunctions of the spin operator $\hat{S}^2$, and the excited Slater determinants often are not, the calculations are performed using *configurational state functions*, or, simply, configurations, which are linear combinations of determinants that are eigenfunctions of $\hat{S}^2$. In the *full CI* method, all excited determinants are included in the calculation, and their number increases factorially with the number of electrons. For this reason, only very few systems can be treated. Derivations of the CI method that allow the treatment of larger systems, the so-called *truncated* CI methods, are reported in the following sections.

Multi-configuration self-consistent field method

The main difference between the multi-configuration self-consistent field (MCSCF) and CI method is that while in CI the molecular orbitals used to build the determinants are fixed, and only the coefficients of the determinants are varied, in MCSCF also the molecular orbitals are optimized. The energy is computed with a self-consistent procedure including only a small fraction of configurations, which are selected in order to describe the property of interest. In the complete active space self-consistent filed method (CASSCF), configurations are generated from a full CI calculation within a subset of *active* orbitals and electrons (Figure 2.1).[94]

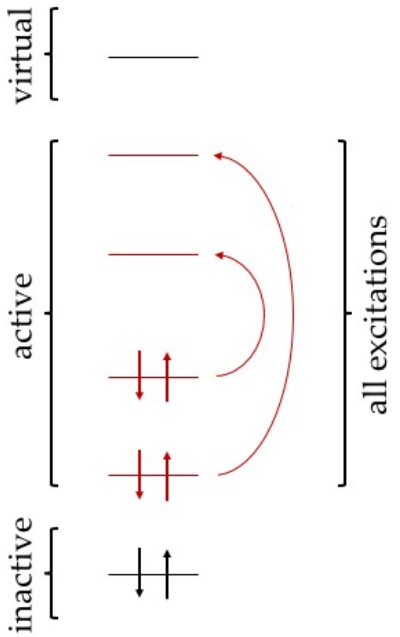


Figure 2.1: Orbital partition of the complete active space self-consistent field method.

The full CI calculation is performed within the active space optimizing the molecular orbitals along with the coefficients of the linear combination. Orbital optimization recovers a large fraction of the static correlation, while in order to recover the dynamical correlation it is much more efficient to increase the number of determinants at fixed molecular orbitals. An important point of this method is the choice of the orbitals that form part of the active space, because it affects significantly the quality of state energies. The main lead in this choice is given by chemical intuition, *i.e.* including the orbitals that are relevant for describing the reaction or process of interest. A second lead comes from the occupation numbers: if one or more of the orbitals that form fart of the active space have an occupation number that is very close to 2 or to 0, then they are redundant and should be exchanged with an inactive or virtual orbital.[94]

Multi-reference configuration interaction method

The multi-reference configuration interaction method (MRCI) is a CI method in which the reference wavefunction used to generate the configuration state functions is no longer the Hartree-Fock wavefunction, but a MCSCF wavefunction. In a MRCISD calculation, for instance, the excitations of one or two electrons out of all the determinants of the MCSCF wavefunction are considered. This often leads to a number of configurations that is too large, and a further truncation may be performed by selecting as references only a subset of determinants.[94]

Excited state calculations

The methods presented in the previous sections may be employed for the calculation of excited states. In this regard, it is interesting to evaluate these methods in terms of excited state properties. In the configuration interaction method, the n-th excited state can be easily generated by using the (n+1)-th eigenvalue from the diagonalization of the CI matrix.[94] The limit of this description is that a full CI is computationally unfeasible for most systems, and for this reason truncated CI methods are employed; when a truncated-CI is used, meaning that only a subset of the molecular orbitals is employed, then the quality of the excited states is reduced. This is due to the fact that the orbitals that are used are normally the Hartree-Fock orbitals of the lowest energy state, and for this reason are biased against the excited states. A higher quality description requires MCSCF-type wavefunctions in which the orbitals are optimized either for a specific state, or for a certain average of desired states. As mentioned, multi-reference methods recover static correlation effects. Accurate calculation of excited state energies, however, requires also the inclusion of dynamic correlation. MRCISD recovers partially both effects.

The problem of size consistency

The problem of the so-called size consistency arises when describing dissociation processes, such as for instance the dissociation of the H_2 dimer. If the two fragments (H_2 molecules) are at large distance, large enough to prevent any kind of interfragment interaction, the energy of the system should be identical to the sum of the energy of the two fragments

calculated separately. If this holds, the method is said to be *size consistent*. While the full CI method is size-consistent, all the truncated CI based methods presented in the previous sections are not. This lack of size consistency is due to the inclusion of only a limited number of determinants in the the wavefunction.[94]

A related property is that of the *size extensivity*, which implies that the energy of a system scales properly with the number of electrons (*e.g.* the study of ionization processes). Truncated CI methods are neither size consistent nor size extensive, and this reflects in their performance in recovering the correlation energy.[94] A different class of methods for the calculation of correlation energy, not variational theory based, is represented by the *Møller-Plesset* methods, which are based on the multi-body perturbation theory.[94,95]

Many-body perturbation theory

In perturbation theory, the Hamiltonian of the system is described as a sum of two terms, an unperturbed part, the zero-th order Hamiltonian $\hat{H}_0$ (Slater determinant), and a perturbation $\hat{H}'$.

$$\hat{H} = \hat{H}_0 + \lambda \hat{H}' \tag{2.17}$$

λ is a parameter that describes the strength of the perturbation acting on the system. The eigenfunctions and eigenvectors of the Hamiltonian may be written in terms of a Taylor expansion of λ, as $\Psi = \lambda^0\Psi_0 + \lambda^1\Psi_1 + \lambda^2\Psi_2 +$

$\cdots$ and $W = \lambda^0 W_0 + \lambda^1 W_1 + \lambda^2 W_2 + \cdots$ respectively. It may be shown that the knowledge of the n-th order wavefunction Ψ_n allows the calculation of the $(2n+1)$-th order correction to the energy W_{2n+1}. From the point of view of the application of this very general theory, the most popular method is the Møller-Plesset (MP), in which the unperturbed Hamiltonian presented in eq. 2.17 is chosen as a sum of the Fock operators of the system. This way, the zero-th order wavefunction is the Hartree-Fock determinant, and the zero-th order energy the sum of the molecular orbital energies. The first-order energy computed with the Møller-Plesset method, $E(MP1)$, is equal to the Hartree-Fock energy, while the second-order energy correction $E(MP2)$ is given in eq. 2.18.

$$E(MP2) = \sum_{i<j}^{occ} \sum_{a<b}^{virt} \frac{\langle \phi_i \phi_j \mid \phi_a \phi_b \rangle - \langle \phi_i \phi_j \mid \phi_b \phi_a \rangle}{\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b} \tag{2.18}$$

A compact notation $\langle \phi_i \phi_j \mid \phi_a \phi_b \rangle = \int \phi_i(1)\phi_j(2) \frac{1}{|\boldsymbol{r}_1 - \boldsymbol{r}_2|} \phi_a(1)\phi_b(2) \rangle$ has been used. The $MP2$ method is able to effectively recover an important amount of the correlation energy (80 - 90%) taking into account only the diexcited configurations obtained from the Hartree-Fock wavefunction. Contribution from the singly excited determinants is null owing to *Brillouin's theorem*, which states that matrix elements between the Hartree-Fock determinant and monoexcited configurations built with canonical orbitals disappear. Møller-Plesset methods efficiently recover increasing fractions of dynamical correlation with the perturbation order. For example, at the second order, $MP2$, *pair correlation* effects are described, while at the third order, $MP3$, also the interaction between electron pairs is included.[94] Since the Møller-Plesset method is not variational, there is no

guarantee that the computed energy is an upper bound of the exact energy. This might seem an inconvenience, but in practice it is not, because in general the interest is not in computing the absolute energy of a system, but rather energy differences. With MP methods, the error in the energy is relatively constant, which allows for error cancellation. This, along with the size extensivity and size consistency of Hartree-Fock based MP methods, is an evident advantage with respect to CI based methods.[94] An expansion of this method, CASPT2, which will be discussed in the following section, involves the perturbative calculation of an energy correction employing a multi-reference wavefunction as zero-th order reference.[106]

Multiconfigurational perturbation theory

As mentioned, calculation of accurate excitation energies requires the inclusion of static and dynamic correlation effects, which can be done with the multiconfigurational perturbation theory methods, among which the CASPT2 method is the most successful.

In the CASPT2 approach, dynamic correlation effects are computed using a CASSCF wavefunction as the zero-th order reference wavefunction. The computational efficiency of such calculations depends strongly on the choice of the zero-th order Hamiltonian, which is built using a Fock-type, one electron operator and reduces to the Møller-Plesset single-reference operator when all orbitals are doubly occupied or unoccupied (corresponding to having no orbitals in the active space). An efficient implementation is achieved when the configuration space, in which the wavefunction is expanded, is decomposed into four subspaces, V_0, V_K,

V_{SD} and $V_{TQ...}$. V_0 is the one-dimensional space spanned by the CASSCF reference function $\mid 0\rangle$; V_K is the space spanned by the orthogonal component to $\mid 0\rangle$ in the restricted full CI subspace ued to generate the CASSCF wavefunction; V_{SD} is the space spanned by all single and double excitation generated from V_0; $V_{TQ...}$ is the space containing the higher excitations not included in V_0, V_K or V_{SD}. This partition leads to a zero-th order Hamiltonian with the following form:

$$\hat{H}_0 = \hat{P}_0\hat{F}\hat{P}_0 + \hat{P}_K\hat{F}\hat{P}_K + \hat{P}_{SD}\hat{F}\hat{P}_{SD} + \hat{P}_{TQ...}\hat{F}\hat{P}_{TQ...} \tag{2.19}$$

which is particularly convenient because only vectors belonging to V_{SD} will contribute to the first-order wavefunction and second order energy. This structure of the zero-th order Hamiltonian, along with the choice of the $\hat{F}$ operator as including only the diagonal terms of a one-electron operator, leads to a block-diagonal structure of the matrix representation of the 2nd order energy correction. Such a block diagonal structure is fundamental because it allows an efficient implementation of this method.[106,107]

From the point of view of applications, CASPT2 presents a series of problems. The first is that it underestimates bond energies, and yields poor excitation energies. This is due to a systematic error in the description of open-shell systems. A correction that preserves the simplicity of the method can be achieved through the so-called *IPEA shift*, a level shift based on the values of electron affinity and ionization potential.[108] This average shift parameter is determined by fitting and for excited states the optimal value is of 0.25 atomic units.[108] A second problem of CASPT2 cal-

culations originates form the presence of intruder states. The solution is to shift the intruder state by adding an arbitrary correction to the expectation value of the Hamiltonian. Two types of level shift have been proposed, a *real* one and an *imaginary* one. The calculations presented in this thesis have been performed with an imaginary shift of 0.10 atomic units.[109]

CASSCF does not provide a good reference wavefunction when strong mixing occurs between the reference state and one or more of the CASSCF states of the secondary space. In such cases, it is convenient to perform a multi-state CASPT2 calculation (MSCASPT2). This approach uses a multi-dimensional reference space and constructs an effective Hamiltonian that allows the CASSCF state to interact via non-diagonal terms.[110]

2.1.4 Time-independent and time-dependent density functional theory

Density functional theory

Density functional methods represent an alternative to post-Hartree-Fock methods for recovering electron correlation. Due to their low computational cost, they have been successful in treating wide variety of systems. In density functional theory (DFT), the electronic wavefunction is replaced by an observable, the electronic density.

$$\rho(\boldsymbol{r}) = N \sum_{s_1} \cdots \sum_{s_N} \int d\boldsymbol{r}_2 \cdots \int d\boldsymbol{r}_N |\Psi(\boldsymbol{r}_1, s_1, \boldsymbol{r}_2, s_2 \cdots \boldsymbol{r}_N, s_N)|^2 \quad (2.20)$$

The fundamental theorems of the density functional theory (Hohenberg-Kohn theorems) affirm the existence of an energy functional of the electronic density ρ, which has a minimum corresponding to the ground state density of the system. In other words, they state that there is a biunivocal correspondence between the ground state energy of a system and its ground state electronic density, and more in general that all observable properties of the given system can be extracted form this density.[94,111] The problem then reduces to find the functional form that links the energy to the density $E[\rho]$.

Among the several distinct approaches to DFT that have been developed, the Kohn-Sham formulation has been particularly successful.[112] This particular method expands the density in a set of molecular orbitals, by analogy with the Hartree-Fock method. The price for introducing the molecular orbitals is an increase of the complexity of the problem ($3N$ variables instead of the three spatial coordinates of the density), but yields good results at a low computational cost.

The Kohn-Sham formulation involves the use of a fictitious noninteracting electron system in such a way that the Hamiltonian operator may be written as follows.

$$\hat{H}_\lambda = \hat{T} + \hat{V}_{ext}(\lambda) + \lambda\hat{V_{ee}} \tag{2.21}$$

In this equation, the λ parameter assumes values between 0 and 1, with 0 corresponding to the system of non-interacting particles, and 1 to the real system. The external potential $\hat{V}_{ext}$ is equal to the nuclei-electrons interaction terms for λ=0, and is adjusted so that it yields the same electronic

density for λ=0, λ=1 and all intermediate values.

The advantage of using a system of non-interacting electron as a reference is explained by looking at the equation that yields the DFT energy.

$$E_{DFT}[\rho] = T_S[\rho] + E_{ne}[\rho] + J[\rho] + E_{xc}[\rho] \tag{2.22}$$

Contributing to the energy (eq. 2.22) are i) T_S, the kinetic energy for the non-interacting electrons system, ii) the nucleus-electron interaction term, E_{ne}, iii) the Coulomb electron-electron interaction J for the non-interacting electrons, and iv) the *exchange-correlation functional*. T_S differs only slightly form the kinetic energy term of the real system, and significantly improves the DFT energy value over the Hartree-Fock one, being an important advantage of this method. The residual difference is incorporated in the exchange-correlation term, which is the core of DFT methods. This terms also includes a potential energy contribution corresponding to the difference between the real electron-electron interaction potential and that of the reference system.[94] A further contribution t this term is the so-called *self-interaction* energy. This is a spurious effect which describes the interaction of the electron with itself, and arises from the Coulomb energy of the Kohn–Sham Hamiltonian. This error affects several properties, including ionization potentials, electron affinity and kinetic barriers.[113]

In DFT calculations, the quality of the results depends on the choice of the functional form for this exchange-correlation functional. A wide variety of functionals is available, and the most common classification for their quality is the so-called *Jacob's ladder*.[114] In *local density approximation* functionals, the first rung of the ladder, the energy depends only on the

electron density; one rung up are the *general gradient approximation* methods, where the energy depends also on the gradient of the density. In the so-called *higher order gradient* methods, the third rung, the energy depends on higher order derivatives of the density. Alternatively, the functional can be taken to depend on the orbital kinetic energy density τ.

$$\tau = \frac{1}{2}\sum_{i}^{occ}|\nabla\phi_i(\boldsymbol{r})|^2 \tag{2.23}$$

At the fourth rung, in the *hybrid generalized gradient approximation* methods, the exchange-correlation functional includes a fraction of the Hartree-Fock exchange. These functionals, which include the famous B3LYP,[115,116] are quite accurate. The development of fifth-rung functionals is ongoing, with different outcomes. For example, *optimized effective potential* methods are based on many-body perturbation theory expansions.[94]

DFT methods do not include the description of dispersion interactions, which often play an important role in determining the structure and properties of chemical systems. A robust and cheap approach to include these effects is the so called DFT-D method, proposed by Grimme, which is available in the two variants DFT-D2 and DFT-D3.[117,118] In this method, the dispersion energy, E_{disp}, is added to the standard DFT energy. In DFT-D2, the dispersion energy consists of a single sum over pair interactions. The DFT-D3 approach is similar, but includes two different contributions to the dispersion energy, a two-body and a three-body term. Moreover, while in DFT-D2 only a R^6 term is included, in the two-body term of DFT-D3 more orders of dispersion are included (eq. 2.24).

$$E_{disp} = \sum_{n=6,8} s_n \sum_{i,j>i}^{N} \frac{C_n^{ij}}{\boldsymbol{R}_{ij}^n} f(\boldsymbol{R}_{ij}) \tag{2.24}$$

C_{ij} are the dispersion coefficients for atom pairs, R_{ij} the interatomic distances and f is a damping function whose role is to define the range of action of dispersion forces: it sets a cutoff value such that for longer interatomic distances than the cutoff, dispersion effects are negligible. The *ab initio* molecular dynamics simulations presented in this thesis were performed with the PBE-D3 functional, a generalized gradient approximation functional with the D3 dispersion correction.[117,118]

Time-dependent density functional theory

Time-dependent density functional theory (TDDFT) is the basis for many well-established methods for the description of the spectroscopic properties of several kinds of systems. The foundation of this method was provided by Runge and Gross, who stated that all observable properties of a time-dependent multi-electron system starting from an initial state Ψ_0 may be extracted from the one-body time-dependent electron density.[119] Given a time-dependent external potential $\hat{V}_{ext}(\boldsymbol{r}, t)$ acting on the system, the response of the electron density to this external potential may be written as a Taylor series, where the zero-th order term is the time-independent density of the ground state ρ_{GS}.

$$\rho(\boldsymbol{r}, t) = \rho_{GS}(\boldsymbol{r}) + \rho_1(\boldsymbol{r}, t) + \rho_2(\boldsymbol{r}, t) + \cdots \tag{2.25}$$

In the so-called *linear response* formalism of TDDFT, only the first-order term of the expansion, ρ_1, is considered. This quantity, ρ_1 is computed from the *density-density linear response function* χ.

$$\chi(\boldsymbol{r}, \boldsymbol{r}', \omega) = \sum_I \left[\frac{\langle \Psi_0 \mid \hat{\rho}(\boldsymbol{r}) \mid \Psi_I \rangle \langle \Psi_I \mid \hat{\rho}(\boldsymbol{r}') \mid \Psi_0 \rangle}{\omega - \Omega_I + i0^+} - \frac{\langle \Psi_0 \mid \hat{\rho}(\boldsymbol{r}') \mid \Psi_I \rangle \langle \Psi_I \mid \hat{\rho}(\boldsymbol{r}) \mid \Psi_0 \rangle}{\omega + \Omega_I + i0^+} \right] \tag{2.26}$$

$\hat{\rho}$ is the density operator $\sum_{i=1}^{N} \delta(\boldsymbol{r} - \hat{\boldsymbol{r}}_i)$, and the sum runs over all excited states.[120] The density-density response function is a function of ω, the frequency of the perturbation, and has *poles* at Ω_I, which are the excitation energies of the system.

So, in order to obtain the excitation energies of a system, it is sufficient to compute the poles of the response function. In practice, this is done by separating the total response function χ into two contributions, one from an auxiliary system of non-interacting electrons (Kohn-Sham system), and one that includes the effects of electron correlation, similarly to what is done in standard DFT. This is done with the aim of computing the exact kinetic energy of the electrons of the auxiliary system, which represents a large fraction of the exact kinetic energy. Electron correlation is included in terms of an *exchange-correlation kernel*, f_{xc}, which is simply the derivative of the exchange correlation potential with respect to the density computed at the ground state density.[120]

$$f_{xc}[\rho_0](\boldsymbol{r}t, \boldsymbol{r}'t') = \left.\frac{\delta V_{xc}[\rho](\boldsymbol{r}, t)}{\delta\rho(\boldsymbol{r}', t')}\right|_{\rho=\rho_0} \tag{2.27}$$

From a practical point of view, the optical response problem is turned into an eigenvalue problem in terms of single particle transitions of the ground-state Kohn-Sham system, yielding the so-called Casida's equations (eq. 2.28).[121,122]

$$\left[\begin{pmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{B}^* & \boldsymbol{A}^* \end{pmatrix} - \Omega_I \begin{pmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{0} & -\mathbf{1} \end{pmatrix}\right] \begin{pmatrix} \boldsymbol{X}_I \\ \boldsymbol{Y}_I \end{pmatrix} = 0 \tag{2.28}$$

Matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ are the Hessians of the electronic energy and include the transitions between the orbitals of the reference system of non-interacting electrons corrected with the exchange-correlation kernel. The *transition vectors* $\boldsymbol{X}_I$ and $\boldsymbol{Y}_I$ are the collective eigenmodes of the time-dependent Kohn-Sham density matrix with excitation energy Ω_I.[121] Transition vectors are useful for analyzing the nature of electronic transitions in terms of occupied and unoccupied orbitals.

The overall performance of TDDFT methods is good. Vertical excitation energies and excited state geometries are in qualitative agreement with those of post-Hartree-Fock methods, but have lower computational cost.[121] For this reason, TDDFT is the method of choice for most photophysical applications. Nevertheless, it presents some important limitations that have to be taken into account, as discussed in the following section.

Practical problems

It is known that TDDFT underestimates by 1 eV or more the energy of charge transfer (CT) transitions. This can be attributed to the long-range behavior of the exchange potential, which is not described correctly in standard DFT functionals.[123] A solution to this problem is splitting the spatial dependence of the potential in two terms, one short-range and another long-range.

$$\frac{1}{r_{12}} = \frac{1 - [\alpha + \beta \cdot erf(\mu r_{12})]}{r_{12}} + \frac{\alpha + \beta \cdot erf(\mu r_{12})}{r_{12}} \tag{2.29}$$

The three parameters, α, β and μ, can be adjusted to fit experimental data, weighting how much of the DFT and Hartree-Fock exchange is incorporated.[123] These functionals take the name of *range-separated*. In particular, the CAM-B3LYP functional[123] has been shown to yield excitation energies in agreement with CASPT2, and for this reason has been used for both DFT and TDDFT calculations presented in this thesis.

A second problem of TDDFT is the description of triplet excitations. While the excitation energy of singlet states usually improves with the increasing fraction of Hartree-Fock exchange, often triplet excitations deteriorate, being severely underestimated.[124]. This happens when there is a problem of triplet instability in the ground state of the system. Computationally, a triplet instability manifests in a negative eigenvalue of the electronic Hessian, indicating that specific orbital rotations of an identified spatial-spin symmetry will lower the energy. This problem is common to all functionals, including range separated ones, and is not easily

cured. It has been shown that the *Tamm-Dancoff approximation* of TDDFT in some cases yields better results than the standard Casida's formalism.[124] The two approaches differ in the excitations that are taken into account. Considering as a reference the standard Casida's equations of eq. 2.28, the Tamm-Dancoff approximation sets matrix $\boldsymbol{B}$ equal to the null matrix. Physically, this corresponds to allow only excitations between occupied-virtual orbital pairs, and to exclude virtual-occupied deexcitations, which in turn are included in the standard TDDFT approach. Despite being a simplification, the Tamm-Dancoff approximation significantly improves triplet excitations over the standard TDDFT method, sometimes at the expense of the quality of singlet excitations.[124]

2.1.5 Solvation models

Solvent effects were included using a *continuum solvent* model. In continuum models, the solute-solvent interaction can be treated in a computationally efficient way by describing the solvent continuosly. The Hamiltonian of the system composed of a solute molecule M and the continuous solvent S may be written as the sum of the Hamiltonian of the solute and a potential term depending on the *solvent response function* Q.[125]

$$\hat{H}^{MS} = \hat{H}^{M} + \hat{V}^{int}[\boldsymbol{r}_M, Q(\boldsymbol{r}, \boldsymbol{r}')] \tag{2.30}$$

Q describes the interaction between solute and solvent, and, according to the specific solvent model which is employed, can include different terms, the main one being electrostatic.

From a practical point of view, the molecule is put in a void cavity within a continuous dielectric medium.[126] This allows the definition of an *apparent surface charge* spread on the cavity surface, whose interaction with the solute completely defines the potential term $\hat{V}^{int}$ of eq. 2.30. In this way, the problem of solvent-solute interaction reduces to solve a classical electrostatic problem: the charge distribution of the solute inside the cavity polarizes the dielectric continuum, which in turn polarizes the charge distribution. This mutual polarization is accounted for with a self-consistent procedure.

Even if this method is conceptually simple, it becomes computationally more demanding as the complexity of the cavity increases. For this reason, the cavity surface is approximated in terms of a set of finite elements, *tesserae*, so that the global interaction may be written as a sum of contributions from the single *tesserae*.[125] Overall, the total solvation free energy is given by the sum of three terms, are reported in eq. 2.31.

$$\Delta G_{sol} = \Delta G_{cav} + \Delta G_{vdw} + \Delta G_{pol} \tag{2.31}$$

where ΔG_{cav} accounts for the formation of the cavity, ΔG_{vdw} is the solute-solvent van der Waals interaction term, and ΔG_{pol} is the solute-solvent electrostatic polarization term.

In general, continuum solvent models offer a robust and well established way to simulate solute-solvent interaction. However, particular care has to be taken when combining the continuum solvent model with an excited state calculation, because in the presence of a time evolution of the solute, the description of the solvent must include the response to

this time-dependent perturbation. At this point, it is useful to point out that the solvent will adapt to the new electronic state of the solute only to a certain degree depending on the nature of the process under study. In the so-called *non-equilibrium* regime, only the fast degrees of freedom are equilibrated with the excited state electronic redistribution, while the slow ones are not. This means that the solvent does not adapt completely to the new state, and for this reason this approach is employed for the study of fast processes, such as vertical excitations. In the *equilibrium* regime, on the other hand, all degrees of freedom of the solvent are equilibrated with the excited state. This implies that the process under study is long enough to allow this relaxation, and for this reason this approach is employed when describing relatively long processes, such as the geometry relaxation in the given excited state.[127] In practice, the change from the non-equilibrium to the equilibrium regime is done by changing the value of ϵ, the permittivity of the medium, from its optical value ϵ_∞, which corresponds to the square of the refraction index, to its static bulk value ϵ_0. This change affects significantly results obtained with polar solvents, for which $\epsilon_\infty << \epsilon_0$.[127]

A further bifurcation is represented by two different approaches to compute the excitation energies of solvated systems, the *state specific* and the *linear response* one. The state-specific approach involves an additional iterative procedure that is not present in the calculation of the isolated molecule. At each step of this iteration, the additional solvent-induced component of the Hamiltonian operator is computed using the first-order density matrix of the state of interest, and the resulting energy is corrected for the work required to polarize the dielectric, until convergence is reached. This approach is quite expensive and presents the further disadvantage that a separate calculation has to be performed for each of the

states of interest. In the linear response approach, on the other hand, excitation energies are determined directly, avoiding the explicit calculation of the excited state wavefunction. Thus, as for the isolated molecule, the whole spectrum of excitation energies can be obtained with a single calculation.[125]

In this thesis, the D-PCM (Dielectric Polarizabile Continuum Model) and SMD (Solvation Model Density) solvent models have been employed in combination with electronic structure methods.[125,128].

Vertical excitation energies have been computed with the linear response non-equilibrium approach, and fluorescence emission wavelengths with the linear-response equilibrium approach. The computed values of excitation and emission wavelengths with the different approaches are reported in Table A.1.

The solvation methods employed in the molecular mechanics calculations that will be presented next are the Generalized Born (GB) and Poisson-Boltzmann (PB) models. The GB model is the least computationally demanding of the two, and originates from the original Born expression of the free energy of a spherical ion surrounded by a continuum dielectric medium, eq. 2.32.

$$\Delta G_{pol} = -166.0\Big(1 - \frac{1}{\epsilon}\Big)\frac{q^2}{\alpha} \tag{2.32}$$

In this equation, α is the ion's Born radius, which is defined as the distance between the center of the ion and the boundary of the dielectric. The fundamental idea underlying the GB model is to extend this simple

equation from an isolated spherical ion to complex solutes considering a sum of terms arising from values of the Born radius that are specific for every atom of the solute.

In the Poisson-Bolzmann model, the polar contribution to the solvation free energy is calculated by subtracting the electrostatic potential computed in vacuum (ϕ_V) from the electrostatic potential calculated in the solvent medium (ϕ_S), according to equation 2.33.

$$\Delta G_{pol} = \frac{1}{2} \sum_i Q_i [\phi_S(\boldsymbol{r}_i) - \phi_V(\boldsymbol{r}_i)] \tag{2.33}$$

where Q_i is the charge of atom i.

2.2 Molecular dynamics simulations

Molecular dynamics simulations describe how a chemical system evolves over time. The core of this kind of simulations lies in the choice of how to describe the interatomic interaction. At this point, there is a main bifurcation between *ab initio* methods and methods based on molecular mechanics.[129] In molecular mechanics methods, the interaction potential is defined based on the force field energy (see below), while in *ab initio* methods the forces acting on the nuclei are computed on-the-fly from electronic structure methods.

In general, molecular dynamics simulations provide useful information on the equilibrium properties of the system under study. Indeed, per-

forming an ideally infinite molecular dynamics simulation corresponds to exploring the full set of microstates (phase space) of the chemical system under study, and this information can be translated into the macroscopic terms by means of statistical mechanics. Statistical mechanics provides a molecular interpretation of the behavior macroscopic systems, and links microscopic phenomena to the observables of thermodynamics. This is done in terms of *ensembles*, which will be defined in the next section.[130]

2.2.1 *Ab initio* molecular dynamics

In *ab initio* molecular dynamics, nuclei are treated classically, their motion being propagated according to Newton's second law.

$$m_A \frac{d^2}{dt^2} \boldsymbol{R}_A(t) = -\nabla E(\boldsymbol{R}_A(t)) \tag{2.34}$$

According to eq. 2.34, the force acting on the nuclei is the gradient of the ground state potential energy surface $\nabla_A E(\boldsymbol{R}(t))$ with the sign changed, which is calculated on-the-fly with electronic structure methods as a function of nuclear coordinates. This represents a tremendous advantage over simulations driven by a potential defined beforehand (molecular mechanics dynamics) because in principle it allows the simulation of any kind of system and does not pose any restriction to the exploration of its phase space.[129]

In practice, the classical equations of motion used to describe the nuclei are integrated by breaking the calculation into *time steps* that are small

enough for the assumption that within each step the force is constant to hold, thus imposing a discretization of the time line. The time step of the molecular dynamics simulation must be chosen to be smaller than the time needed for the fastest event to take place. Commonly, this is the stretching of bonds involving hydrogen, for which a 0.5 - 1 fs timestep is adequate. In practice, after initialization, at each step an electronic structure calculation is performed to compute the forces acting on the nuclei. These forces are then used to update positions and velocities, which effectively corresponds to generating the next step of the simulation, and so on.[129,131]

Several algorithms are available for the numerical integration of the equations of motion, and the discussion of their features is beyond the purpose of this section, but it is important to mention that the choice of the integrator has to meet a certain set of criteria in terms of energy conservation, accuracy, reversibility and, obviously, computational efficiency. While Newton's equations themselves fulfill the requirements of reversibility and energy conservation, their numerical integration fatally introduces a certain amount of error, which has to be controlled.

2.2.2 Thermodynamic ensembles

An ensemble is a large collection of systems constructed so that they are replicas at the macroscopic level of the thermodynamic system whose properties are under study. This is important because the first postulate of statistical mechanics states that, in the limit of an infinite ensemble, the long time average of a mechanical variable (such as pressure, for example) of the thermodynamic system of interest is equal to the ensemble

average of this quantity. Considering the example of pressure as the mechanical variable, its time average is the value of the equilibrium, time-independent, pressure of the system. In microscopic terms, pressure is the force for unit area exerted on the walls of the system by molecular collisions. For this reason, a single macroscopic value of pressure can correspond to a multiplicity of microscopic states.[130]

Several types of thermodynamic ensembles may be employed, and are classified according to the *control variables* that are employed, *i.e.* the variables that define the thermodynamic state. In particular, the molecular dynamics simulations presented in this thesis were performed using the NVE, NVT and NPT ensembles, which will be briefly presented. In an NVE, or microcanonical, ensemble, systems are distributed uniformly over the possible microscopic states consistent with the specified values of number of particles (N), volume (V) and energy (E). This kind of ensemble described an *isolated* system. A canonical ensembl (NVT), on the other hand, describes a situation in which the thermodynamics system of interest, that has fixed values of number of particles (N) and volume (V), is immersed in a heat bath (constant temperature T). This time, the ensemble describes a *closed* system. In the isothermal-isobaric enseble, values of number of particles (N), pressure (P) and temperature (T) are specified.

Each ensemble has a *characteristic function*, a thermodynamic function which is particularly suitable for the description of the system. The characteristic function of the isothermal-isobaric ensemble is Gibbs free energy, the one of the canonical ensemble the Helmholtz free energy, and the one of microcanonical ensemble the product of temperature and entropy.[130]

Depending on the circumstance, one ensemble may be preferred over

the other. For example, for the molecular dynamics simulation of biomolecules both NVT and NPT ensembles may be employed. NPT can be considered as the natural one, due to the fact that it eases comparison to experimental data, which are usually recorded at atmospheric pressure and at constant temperature. Moreover, the isothermal-isobaric ensemble is particularly recommended when the biological system is expected to undergo important conformational modifications along the simulation. If this is not the case, NVT and NPT can be used almost indifferently, since the characteristic functions of the two ensembles, the Helmholtz and Gibbs free energy differ only slightly. Using NVT instead of NPT is convenient in terms of computational cost: NPT is more computationally demanding due to the introduction of a pressure scaling.[132]

Initialization of a molecular dynamics simulation involves choosing the starting geometry and assigning the initial velocities of the particles of the system under study. This is done with a random assignation of velocities, so that the kinetic energy is compatible with the assigned temperature. After initialization, the dynamics evolves according to interparticle interactions.[131]

2.2.3 Surface hopping with arbitrary couplings

This section describes the surface hopping with arbitrary couplings (SHARC) method used for the dynamical study of the photochemistry of 5-bromouracil, presented in Chapter 6.[133] This technique allows the study of the deactivation processes of excited molecules providing useful information about the structural changes that accompany each deactiva-

tion pathway, their branching ratios and kinetic constants. Surface hopping methods describe the motion of the molecule on a subset of its excited potential energy surfaces, and are based on the adiabatic approximation, meaning that nuclear motion is propagated on one energy surface at at time. When a trajectory approaches a crossing point, and the adiabatic approximation loses its validity, a transition probability is computed between the current, *active*, state and other potential energy surfaces, and is compared to a random number. Depending on this comparison, the trajectory can either stay on the same state, or hop to a new state, and in this second case the propagation follows on the new potential energy surface. In this way, it is possible to approximate the more exact motion of the quantum wavepacket with an ensemble of independent classical trajectories, with a significant reduction of the computational cost with respect to more accurate techniques. In practice, surface hopping is analogous to the *ab initio* molecular dynamics presented in the previous section, but while in its standard formulation nuclei only move on the ground state potential energy surface, in surface hopping more potential energy surfaces are explored, one at a time.

Recently, the group of Prof. González developed a surface hopping methodology, called *surface hopping including arbitrary couplings* (SHARC), which can include effects such as the spin-orbit coupling and laser fields, which are not included in standard surface hopping codes. This extends the applicability of this method to a wider range of photophysical effects of interest.[133,134] As all surface hopping programs, SHARC follows this general scheme at each time step:

- electronic properties (such as the electronic Hamiltonian operator

and the nonadiabatic couplings) are computed at the current nuclear geometry;

- the electronic wavefunction is propagated;
- the hopping probabilities are computed and the new active state is determined;
- the gradient of the new active state is used to calculate the new geometry.

The core of this procedure is the propagation of the electronic wavefunction. The electronic equation of motion can be derived from the time-dependent Schrödinger equation expanding the wavefunction in a basis $\sum_\alpha \mid \psi_\alpha\rangle\langle\psi_\alpha \mid$.

$$\frac{d}{dt}\langle\psi_\beta \mid \Psi\rangle = -\sum_\alpha \Big[i\langle\psi_\beta \mid \hat{H}_{el} \mid \psi_\alpha\rangle + \langle\psi_\beta \mid \frac{d}{dt} \mid \psi_\alpha\rangle\Big]\langle\psi_\alpha \mid \Psi\rangle \tag{2.35}$$

The operational formula used to update the coefficients of the wavefunction may be written more conveniently in matrix form.

$$\frac{d}{dt}\boldsymbol{C}^{rep} = -[i\boldsymbol{H}^{rep} + \boldsymbol{T}^{rep}]\boldsymbol{C}^{rep} \tag{2.36}$$

$\boldsymbol{C}^{rep}$ is the wavefunction coefficient vector, $\boldsymbol{H}^{rep}$ the matrix representation of the Hamiltonian in the given basis and $\boldsymbol{T}^{rep}$ the temporal coupling matrix. $\boldsymbol{T}^{rep} = \boldsymbol{v}\boldsymbol{K}^{rep}$, where $\boldsymbol{v}$ is the nuclear velocity vector and $\boldsymbol{K}^{rep}$ the

nonadiabatic coupling matrix. The matrices $\boldsymbol{H}^{rep}$ and $\boldsymbol{K}^{rep}$ have different properties according to the representation, which severely affects the results of the calculation.

In the Molecular Coulomb Hamiltonian (MCH) representation, which includes only the Coulomb and kinetic terms (eq. 2.37), the MCH Hamiltonian is spin independent, and thus shares its eigenstates with the spin operators $\hat{S}^2$, $\hat{S}_z$. In this representation, which is the most common for quantum chemistry programs and, thus, the most natural for on-the-fly dynamics, the Hamiltonian matrix is diagonal, and the elements of the nonadiabatic coupling matrix are non-zero.

$$\hat{H}_{el}^{MCH} = -\sum_i \left[\frac{1}{2}\nabla_i^2 + \sum_A \frac{Z_A}{\boldsymbol{r}_{iA}}\right] + \sum_{i<j} \frac{1}{\boldsymbol{r}_{ij}} + \sum_{A<B} \frac{Z_A Z_B}{\boldsymbol{r}_{AB}} \tag{2.37}$$

When additional terms are included in the Hamiltonian, such as the spin-orbit operator, the Hamiltonian matrix is not diagonal anymore, with off-diagonal couplings delocalized over the potential energy surface. Moreover, the spin-orbit operator lifts the degeneracy between the components of multiplets, which are taken into account as individual states. This represents a further problem because the sum of the transition probabilities into all multiplet components is not independent of the rotation of the molecule in the laboratory frame in the MCH representation.[133,134] A different representation, in which the Hamiltonian matrix is diagonal, can cure both problems (eq 2.38).

$$\boldsymbol{H}^{diag} = \boldsymbol{U}^{\dagger}\boldsymbol{H}\boldsymbol{U} \tag{2.38}$$

Currently, no diagonal gradients are available for ab initio electronic structure calculations. Thus, the SHARC code follows a pragmatic approach combining quantum chemical data in the MCH representation from an external code with surface hopping in the diagonal representation.[133,134] In this work, COLUMBUS was chosen as the external code.[135]

2.2.4 Molecular mechanics dynamics

As mentioned, it is possible to run molecular dynamics simulations with the methods of molecular mechanics, which will be briefly presented in this section.

Force field methods

Molecular mechanics methods also go under the name of *force field* methods. In force field methods, electrons are not treated explicitly. Rather, they are fused with the nucleus into a single unit called *atom type*. Force field methods arise from the observation that all molecules are composed of structural units that keep approximately the same properties over different compounds. An aliphatic C−H bond distance, for instance, will vary only slightly over different molecules with a mean value of 1.09 Å. In this light, it is possible to describe chemical bonds with an harmonic potential whose equilibrium value and force constant depend on both the

two elements that are involved and the type of bond. The latter is due to the fact the same elements can form different kinds of bonds: the C−C, C=C and C≡C bonds, for instance, have different properties and need three different atom types to be described.

Bond angles and torsions can be described with analogous simple functions according to the nature of their composing elements, along with non-bonding interactions. Overall, the generic force field energy of a system can be written as a sum of terms that account for the different interactions.[94]

$$E_{FF} = E_{str} + E_{bend} + E_{tors} + E_{vdw} + E_{el} \tag{2.39}$$

E_{str} is the energy required for stretching a bond between two atoms, E_{bend} the energy required for bending an angle, E_{tors} the energy required to rotate around a bond and E_{vdw} and E_{el} describe the non-bonded atom-atom interactions. The functions that describe this energy are *parametrized* over experimental data or calculations performed at a high level of theory.

This parametrization offers a simple way of computing the potential energy as a function of nuclear coordinates. This approach presents enormous advantages in terms of computational time, being far less demanding than any *ab initio* calculations, but also presents several drawbacks. Firstly, in its standard implementation, it cannot describe chemical reactions, *i.e.* the breaking and formation of bonds. This is due to the fact that the energy is defined in terms of the interactions that exist between the constituent atom types; for this reason, the *topology* of the molecule is defined beforehand, and cannot be changed. To circumvent this problem,

reactive force fields have been developed, where atom types are dynamically changed.[136,137]

Secondly, force fields are usually parametrized over a specific class of systems for which they are meant to be used, and for this reason do not yield good results for compounds out of this class. Despite these limitations, force field methods, also known as *molecular mechanics* methods, are the methods of choice for the study of biological macromolecules. In this thesis, force field methods have been employed to study the binding of organic fluorescent markers to amyloid-β fibrils. In particular, the protein energy landscape exploration program has been used to identify the binding poses, and the molecular mechanics/Poisson-Boltzmann surface area method to evaluate the binding energies. Details of these two methodologies are reported in the following sections.

Molecular mechanics/Poisson-Boltzmann surface area method

The molecular mechanics energies combined with the Poisson Boltzmann or generalized Born model and surface area (MM/PBSA and MM/GBSA) methods are commonly used to estimate the free energy of binding, ΔG_{bind}, of small ligands to biological macromolecules.

$$L + R \rightarrow RL \tag{2.40}$$

The binding free energy for this reaction is computed from the free energy defined in eq. 2.41.

$$G = E_{MM} + G_{pol} + G_{np} - TS \tag{2.41}$$

The first term of the sum is the molecular mechanics energy of the molecule. G_{pol} is the polar contribution to the solvation free energy and can be calculated either by solving the Poisson-Boltzmann equation, or by using the generalized Born model, while G_{np} is the corresponding non-polar contribution which is computed from a linear relation to solvent accessible surface area. The last term is an entropic contribution that is evaluated with a normal mode analysis of vibrational frequencies. The basic idea of this method is to compute the binding free energy as the expected value of the difference between the free energy of the complex, composed of the bound receptor and ligand, and the free energies of the separate components, according to eq. 2.42.[138]

$$\Delta G_{bind} = \langle G_{RL} - G_R - G_L \rangle_{RL} \tag{2.42}$$

In practice, this is done by running a molecular dynamics simulation of the complex, selecting a series of uncorrelated snapshots form the corresponding trajectory, and for each compute both the free energy of the complex and that of the ligand and receptor separately. In principle, three parallel simulations should be run, for both the complex and the separate components, but in practice this is seldom done, and is justified by the assumption that the structural evolution of the ligand and the receptor will not be significantly affected by their mutual interaction. Moreover, the entropic term is often neglected based on the same assumption: if the entropy of the separate components does not differ significantly from that

of the complex, then the entropic contributes to the binding free energy will cancel out. These two simplifications make this method computationally advantageous, and since its first application, it has become increasingly popular.[138,139] The MM/PBSA calculations presented in this thesis have been performed with the AMBER 14 suite,[140,141] using the 99SB force field.[142]

2.3 Protein energy landscape exploration

The *protein energy landscape exploration* (PELE) method combines protein structure algorithms and Metropolis Monte Carlo techniques. This method generates and propagates changes in a system by generating a series of structurally similar local minima and combining them in a trajectory.[143] In our case, the program has been used to explore the energy landscape corresponding to the interaction of a small organic markers with several models of amyloid fibrils, with the aim of identifying a series of plausible binding poses.

Various approaches that go under the name of energy landscape exploration are available, aimed at defining a series of local minima that describe the conformational changes undergone by molecules.[143,144] Among these techniques, a relevant example is represented by normal mode analysis. This method provides insight into the internal motion of proteins with little computational cost. This method, whose first application to biomolecules dates back to the 1980s,[145] is widely used for exploring functional motions of proteins.[146] Normal mode analysis provides information on the equilibrium global motions (low frequency modes) accessible

to the system. This is due to the fact that global modes essentially represent the configurational changes which involve the least energy increase per unitary deformation. The advantages of this technique are its simplicity (it is assumed that the system is stabilized by harmonic potentials) and its robustness: it has been shown that protein structure, and consequently its normal modes, is defined by the network of inter-residue contacts, which is defined by the global shape and topology of the protein, and are insensitive to both local interactions and the specific energy function and parameters of the force field that is employed.[143,147–149]

Alongside energy landscape exploration methods, also protein structure prediction methods based on rotamer libraries are well developed.[143,149,150] These methods are based on the observation that protein side chains tend to assume a limited number of low energy conformations, rotamers.[150] Thus, it is possible to reduce the complexity of the problem defining a protein structure by considering only a small number of rotamers instead of the continuous space of possible side chain conformations. The development of this kind of methodology has been allowed by the growth of the Protein Data Bank.[151] Indeed, not only the number but also the resolution of available structures increases, it becomes easier to recognize and classify side chain conformations according to their frequency. The choice of the combination of rotamers for a given polipeptidic sequence becomes a combinatorial problem, because the search is oriented towards the combination of rotamers which, avoiding steric clashes, establishes the best interresidue network of interactions, yielding a stable conformation.[152,153]

PELE combines energy landscape exploration and protein structure predictions methods in an algorithm composed of four steps. The first

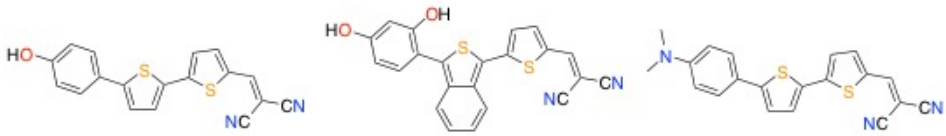
step of the procedure involves a local perturbation of the ligand, with rigid rotation and translations, and torsions deriving from flexibility around dihedral angles of rotable bonds. The algorithm then proceeds by optimally arranging all side chains local to the ligand with a rotamer library side chain optimization. This is followed by a minimization of the region including all residues local to the atoms involved in the first two steps, which may include the backbone of the protein as well, in which the alpha carbons are driven to a new position resulting from small displacements in low frequency anisotropic normal modes.[143,154,155] These calculations are performed using a force field; in particular, for our calculations the OPLS 2005 force field was employed.[156] These three first steps compose a *move* which is accepted or rejected using a Metropolis criterion in the fourth and last step. If it is accepted, a local minimum has been found, and the algorithm is repeated starting from this new geometry. In order to obtain converged results, a number of trajectories has to be run which guarantees that the ligand has explored all binding possibilities. Solvation effects are included with the generalized Born model.[157]

Chapter 3

Bithiophene derivatives

This Chapter presents the results obtained for a family of near-infrared push-pull fluorescent markers named NIADs. In particular, the fluorescent markers NIAD-4, NIAD-11 and NIAD-16 have been studied.[46,54] They all share a common structure, being composed of an electron donating (hydroxyl or N,N-dimethylamino) and an electron withdrawing group (propanedinitrile) linked by a linear polarizable π bridge containing a bithiophene. These common features suggest that the three molecules have similar optical properties, both isolated and interacting with amyloid-β fibrils, but in fact this chapter shows as the small structural differences among the markers deeply affect their performance for amyloid detection.

As mentioned in the Introduction, fluorescent markers for amyloid-β detection need to fulfill a series of stringent requirements, including



Scheme 3.1: Molecular structures of NIAD-4, NIAD-11 and NIAD-16.

emission in the near-infrared region, specific recognition of amyloid deposits, and a "turn on" mechanism triggered by this recognition. NIAD-4 (Scheme 3.1), the first marker of this family to be proposed in 2005,[46] is extraordinarily promising for *in vivo* imaging due to its unique structural and spectral features. Its classical push–pull architecture results in an highly polarizable molecule whose fluorescence emission wavelength falls at 620 nm, in the near-infrared region.[45,46]

This low frequency emission is accompanied by a high blood-brain barrier permeability, which is guaranteed by the low molecular weight (334 Da) and a balance of hydrophobic and hydrophilic groups. Blood-brain barrier permeability may be evaluated with the $LogBB$ index,[158] computed according to the following equation

$$LogBB = -0.0148 \times PSA + 0.152 \times ClogP + 0.139 \tag{3.1}$$

where PSA is the topological polar surface area of the molecule,[159] corresponding to the surface belonging to polar atoms, and $ClogP$ the octanol-water partition coefficient. Values of PSA and $ClogP$ were computed using the Molinspiration properties calculator (www.molinspiration.com).

It has been shown that molecules with values of $LogBB > 0.3$ readily cross the Blood-Brain Barrier, while those with $LogBB <$ - 1 are poorly distributed in the brain. NIAD-4 has a value of $LogBB$ of - 0.10, which falls close to the desired range.

The structural modifications introduced in NIAD-11 and NIAD-16 affect the permeability of the molecule through the blood-brain barrier (Scheme 3.1). The computed values of $LogBB$ for NIAD-11 and NIAD-16 are - 0.27 and 0.24, respectively. Since NIAD-4 has been shown to readily cross the blood-brain barrier and bind to amyloid-β deposits after intravenous injection, it s reasonable to expect that NIAD-16, which has a positive $LogBB$ value, may also able to cross the blood-brain barrier, while NIAD-11, which has a higher negative $LogBB$ value, is expected to show a less pronounced blood-brain barrier permeability.[158]

Additionally, in 2005 Nesterov *et al.* uncovered two important features of NIAD-4: i) a strong binding to amyloid-β aggregates (K_i = 10 nM), and ii) a strong influence of the environment on the fluorescence properties. Indeed, while in aqueous solution NIAD-4 has only trace emission, in the presence of amyloid-β fibrils a $\sim$ 400-fold fluorescence is observed.[46] The strong affinity for amyloid fibrils suggests that NIAD-4 may be selective towards this kind of aggregates, and the significant difference in the emission properties between the free and bound molecule allows high contrast upon detection of amyloid-β.

Concerning the emission properties, NIAD-11 and NIAD-16 were designed with the aim of pushing the emission wavelength deeper into the near-infrared region.[46,52,54,89] This was achieved by extending the conjugated π system in NIAD-11 and by improving the donor capability of

a terminal group in the case of NIAD-16. Moreover, NIAD-16 present a remarkable behavior in terms of lifetimes of the emitting state: when incubated with tissue sections, the distribution of lifetimes depends on the specific tissue substrate, meaning that with a single fluorescent dye it is possible to distinguish a variety of histological targets.[54]

Despite the success of these markers, only few theoretical studies have been addressed to uncover the mechanism underlying their excellent optical properties.[46,90,92,160] This Chapter presents the results that have been obtained for NIAD-4, NIAD-11 and NIAD-16, which explain the photophysical properties of these markers at the molecular level. Since NIAD-4 is the first and most important marker of the family, our theoretical calculations were completed with experimental measurements. In particular, the absorption, fluorescence and fluorescence excitation spectra of the molecule in different solvents were recorded in collaboration with Prof. Jordi Hernando of this University. The performances of NIAD-4, NIAD-11 and NIAD-16 have been compared, with a focus on understanding the relationship between structural modification and optical properties, which provides useful insights for the rational design of markers with tailored properties.

3.1 Absorption and fluorescence

With the aim of investigating the optical properties of the free markers, the absorption and fluorescence spectra of NIAD-4 in several solvents were recorded, and the same absorption spectra were simulated with TD-DFT. NIAD-4 was purchased from AOBIOUS and used without any fur-

ther purification. All spectroscopic experiments were carried out in HPLC or spectroscopy quality solvents. Steady-state UV-vis absorption measurements were recorded on a HP 8453 spectrophotometer.

As aforementioned, NIAD markers are composed of π blocks linked by single bonds, that allow for conformational isomerism, and although the most stable conformation is the one reported in Scheme 3.1, there are several low-lying isomers that need to be taken into account for a proper description of the system. In particular, the three conformers of NIAD-4 reported in Figure 3.1 have been considered.

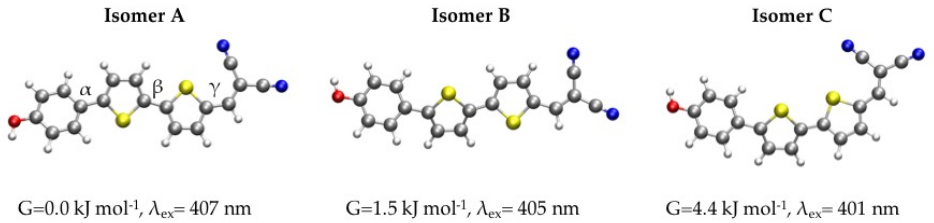


Figure 3.1: Conformational isomers of NIAD-4, with indication of their relative free energies and excitation wavelengths.

α is the torsional angle between the phenyl group and the first thiophene ring, β the one between the two thiophene rings and γ the one between the second thiophene ring and the double bond connected to the malononitrile group (Figure 3.1). Results of ground-state DFT calculations indicate that NIAD-4 has a pronounced flexibility around the α and β angles, with rotational barriers of ~ 12 and ~ 18 $kJ\,mol^{-1}$ respectively. Rotation around γ, on the other hand, has a higher barrier (38 $kJ\,mol^{-1}$), with a much less pronounced flexibility. This means that in aqueous solution there will be a continuous interconversion between the various conformational isomers.

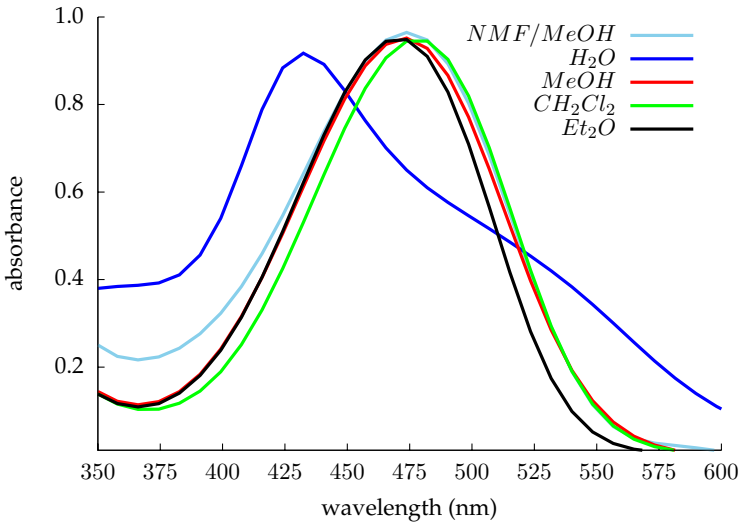
For systems of this kind, the single optimized geometry of the most stable conformational isomer is not a good descriptor of the system, which is better simulated with a collection of structures. For this reason, the absorption spectrum has been computed using a pool of several geometries extracted from molecular dynamics simulations. This strategy is widely accepted for the description of the absorption properties of flexible systems.[161–163]

Ab initio molecular dynamics simulations within the NVT ensemble at T = 300 K and using the PBE-D3[118,164] functional were carried out with the CP2K package,[98] according to what reported in Chapter 2. A 10 ps simulation was conducted on each of the three isomers reported in Figure 3.1, with a timestep of 0.5 fs, and 40 snapshots were sampled from each trajectory. On the resulting 120 geometries, vertical excitations have been computed, both for the isolated molecule and including solvation effects with the PCM method, with the oscillator strengths weighted by the Boltzmann weight BW of the isomer.[126]

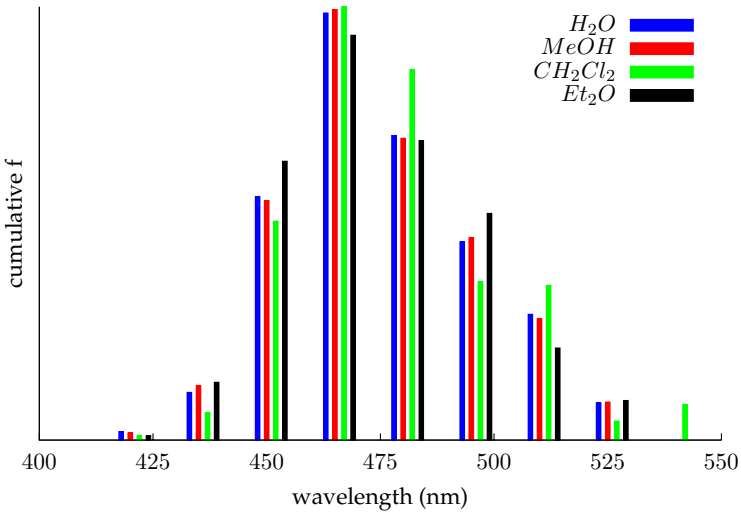
$$BW = e^{-\frac{\Delta G}{RT}} \tag{3.2}$$

The absorption spectra of NIAD-4 in a series of solvents of increasing dielectric constant, diethylether (ϵ = 4), dichloromethane (ϵ = 9), methanol (ϵ = 33) and water (ϵ = 80), are reported in Figure 3.2a. A striking difference is observed in the behavior of the molecule in water compared to other solvents: while in all other solvents the absorption maximum falls in the narrow range 471 $\div$ 480 nm, in water it undergoes a blue-shift of $\sim$ 50 nm, accompanied by the appearance of a second, less intense, peak at $\sim$ 525 nm.

Comparing the computed and absorption spectra reported in Figure 3.2b with the experimental ones of Figure 3.2a, it is evident that our molecular dynamics simulations are not able to reproduce the peculiar shape of the spectrum recorded in water. Indeed, computed spectra are very similar for all solvents, with a maximum around 470 nm, which is in excellent agreement with the experimental values recorded in dichloromethane, methanol and diethylether. The difference between the absorption wavelength computed with this method as the bar plot maximum and the vertical excitation at the optimized S_0 geometry of the most stable conformer can be attributed to two factors: changes in the torsional angles and corresponding rotational barriers derived from using different functionals in molecular dynamics simulations and in excited state calculations, and flexibility effects. Calculated values are in good agreement with experimental data, but they do not predict the $\sim$ 50 nm blue-shift observed in water as compared to other solvents. Moreover, the inclusion of flexibility effects obtained by using a pool of geometries to compute absorption spectra instead of one is able to reproduce correctly the $\sim$ 120 nm band width of the spectra recorded in methanol, dichloromethane and diethylether, but does not yield the double peaked shape of the absorption spectrum recorded in water (Figure 3.2a). This result suggests that other effects may be responsible for the peculiar behavior in water, which will be discussed in the following sections.



(a) Experimental spectra.



(b) Calculated spectra.

Figure 3.2: Experimental and calculated absorption spectra of NIAD-4 (c_{NIAD-4} = 5 μM) in diethylether, dichloromethane, methanol, water and a 1 : 2 mixture of N-methylformamide and methanol.

A similar procedure may be adopted for computing the absorption spectra of NIAD-11 and NIAD-16. The structural modifications of NIAD-11 and NIAD-16 compared to NIAD-4 are the introduction of an additional hydroxyl group, an aromatic ring fused to one of the two thiophene units in NIAD-11 and the substitution of the hydroxyl group of NIAD-4 by an N,N-dimethylamino group in NIAD-16 (Scheme 1).

Table 3.1: Equilibrium values of the torsional angles α, β and γ (°) and free energy rotational barriers around torsional angles α, β and γ ($kJ\,mol^{-1}$) in gas phase and aqueous solution (in parentheses, SMD model).

molecule	α	β	γ
NIAD-4	33 (28)	-14 (-3)	180 (-179)
NIAD-11	-52 (-52)	-21 (-21)	-178 (-179)
NIAD-16	30 (23)	-12 (0)	179 (180)
molecule	TS_α	TS_β	TS_γ
NIAD-4	11.4 (13.1)	19.2 (19.8)	38.6 (45.8)
NIAD-11	6.6 (5.0)	16.5 (14.8)	40.5 (41.7)
NIAD-16	13.2 (17.4)	21.5 (19.0)	41.6 (44.5)

Equilibrium torsional angles for the most stable isomer of NIAD-4, NIAD-11 and NIAD-16 along with the free energy rotational barriers are reported in Table 3.1. The structural modifications do not influence the nature of the most favorable conformer, which has a *transoid* conformation of the two thiophene rings and a *cisoid* conformation between the malononitrile group and the adjacent thiophene.

It can be observed that the structural differences have an effect on the planarity and flexibility of the molecule. The additional hydroxyl group and aromatic ring in NIAD-11 result in a pronounced steric hindrance,

which results in an equilibrium value for dihedral α of 52° and a lower barrier for the corresponding rotation. The remaining angles and barriers are quite similar in the three molecules. Regarding NIAD-16, the substitution of $-$OH by $-N(CH_3)_2$ produces a stronger electron delocalization and, consequently, a marginally more planar geometry that correlates with marginally higher rotational barriers compared to NIAD-4 (Figure 3.1).

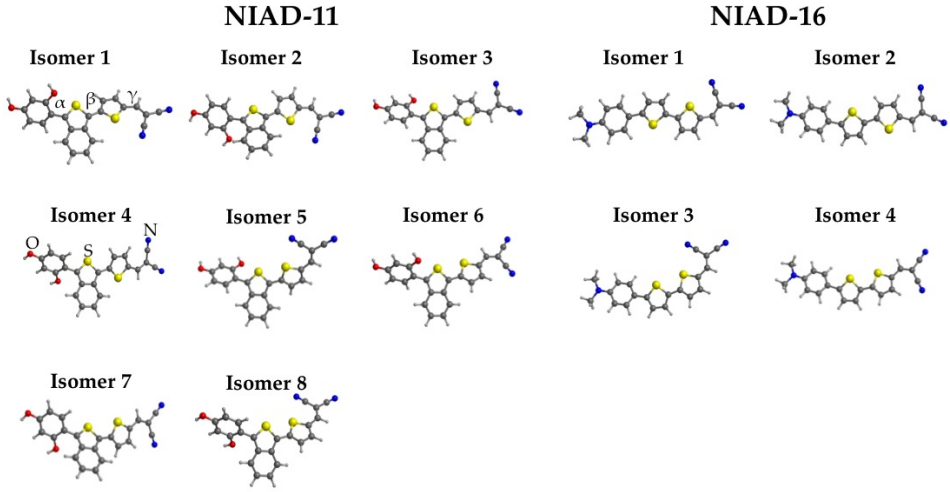


Figure 3.3: Conformational isomers of NIAD-11 and NIAD-16 at the ground state optimized geometry.

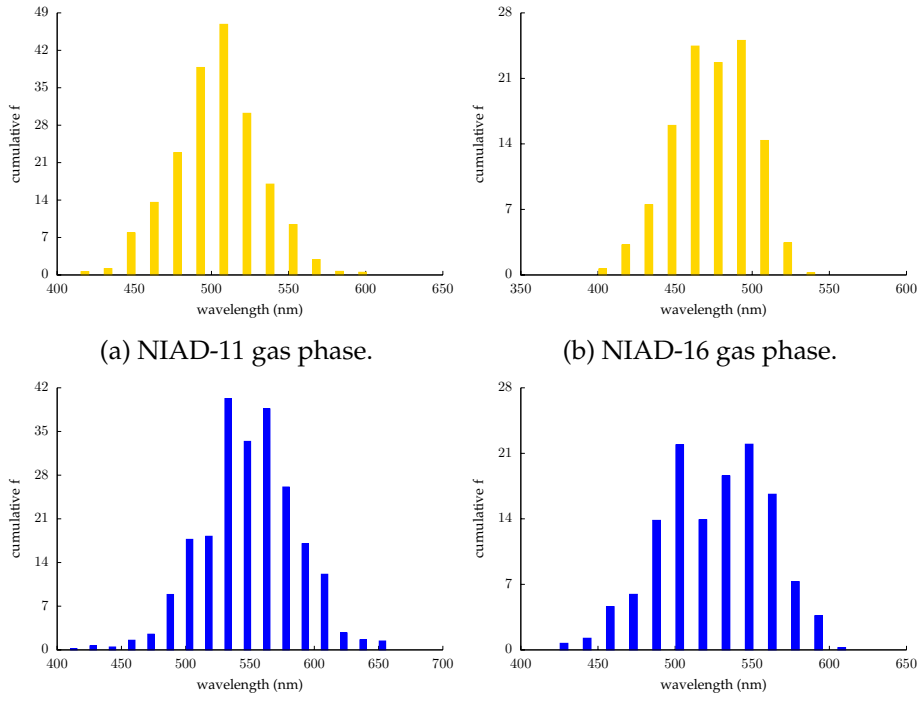
NIAD-11 and NIAD-16 exhibit 8 and 4 low-lying conformational isomers respectively, similarly to what was already described for NIAD-4. These conformers result from rotations around the three single bonds and are shown in Figure 3.3. Energy differences among isomers are small (less than 7 $kJ\,mol^{-1}$, see Figure 3.3 and Table B.5) and their relative stabilities only marginally influenced by the modifications included in NIAD-11 and NIAD-16. This suggests that also for NIAD-11 and NIAD-16 several

conformers are present in significant amounts at thermal equilibrium conditions and thus, they all influence the photophysical properties of these dyes.

Calculated spectra were obtained computing TDDFT vertical excitations in gas phase and in water (SMD model) on 50 equally distributed snapshots sampled from each 10 ps dynamics trajectory (400 total snapshots for NIAD-11 and 200 for NIAD-16) and weighting the contribution of each conformer according to its Boltzmann factor. The resulting spectra are reported in Figure 3.4, while isomer contributions weighted by their Boltzmann factor are reported in Figure B.4. [92] The spectra reflect the large flexibility of NIAD-11 and NIAD-16, with bandwidths of about 120 to 200 nm.

Moreover, the same trends are observed for the three dyes when comparing the vertical excitation of the most stable isomer and the maximum absorption wavelength from molecular dynamics simulations: the latter present a 40 - 50 nm red shift in all cases. As already mentioned for NIAD-4, this difference can be attributed to two factors: changes in torsional angles and corresponding rotational barriers derived from using a different functional in the two calculation and flexibility effects. This flexibility is larger for NIAD-11, which translates to a larger bandwidth for NIAD-11 than for the other two markers. NIAD-4 and NIAD-16, in turn, present very similar bandwidths, in agreement with the similar flexibility suggested by their rotational barriers.

The computed maximum absorption wavelength of NIAD-11 in water is 550 nm and that of NIAD-16 is 525 nm. The computed value for NIAD-11 is in very good agreement with the experimental value of 545 nm



(a) NIAD-11 gas phase.

(b) NIAD-16 gas phase.

(c) NIAD-11 water.

(d) NIAD-16 water.

Figure 3.4: Simulated absorption spectra from *ab initio* molecular dynamics simulations.

registered in PBS (phosphate-buffer saline). NIAD-16, on the other hand, exhibits an experimental absorption maximum at 470 nm, the computed value being significantly larger. This situation is similar to that reported for NIAD-4 (Figures 3.2a and 3.2b), and will be discussed in the following sections.

A further step in the analysis of the dependence of the optical properties of these markers on the environment has been done by recording the

fluorescence emission spectra of NIAD-4 in the same solvents used for the absorption measurements, shown in Figure 3.5.

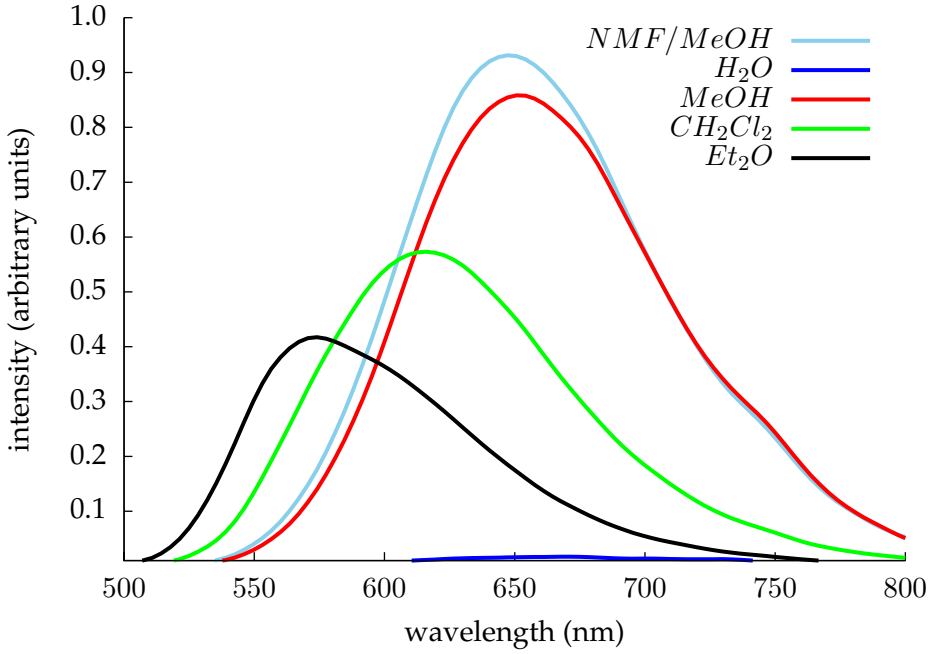


Figure 3.5: Fluorescence spectra of NIAD-4 (c_{NIAD-4} = 5 μM) in diethylether, dichloromethane, methanol, water and a 1:2 mixture of N-methylformamide and methanol.

Fluorescence spectra were recorded using a custom-made spectrofluorometer, where a cw DPSS laser (λ_{exc} = 473 nm, SciTec) was used as excitation source and emitted photons were detected using a ICCD camera (Andor) coupled to a spectrograph. Again, fluorescence spectra show a marked difference between the behavior of the molecule in water and in all other solvents: while in water fluorescence emission is null, in all other

solvents there a significant emission is observed, whose intensity increases with the dielectric constant of the solvent in the order diethylether (ϵ = 4) $<$ dichloromethane (ϵ = 9) $<$ methanol (ϵ = 33). This increase of fluorescence with solvent polarity is accompanied by a red-shift of the emission maximum, which falls at lower energies in highly polar solvents (solvatochromic effect). Understanding these variations requires an analysis of the nature of the electronic excitation undergone by bithiophene markers.

3.2 Charge transfer

TDDFT(CAM-B3LYP) vertical excitations computed at the optimized ground state geometry of NIAD-4, NIAD-11 and NIAD-16 indicate that the first excited singlet state, S_1, has a strong HOMO $\rightarrow$ LUMO character, and involves a *charge transfer* from the donor to the acceptor portion of the molecule, as shown in Figure 3.6. Indeed, while the HOMO is delocalized over the whole π system, the LUMO is polarized on the electron-withdrawing end of the molecules, confirming that a transfer of electron density is produced upon excitation.

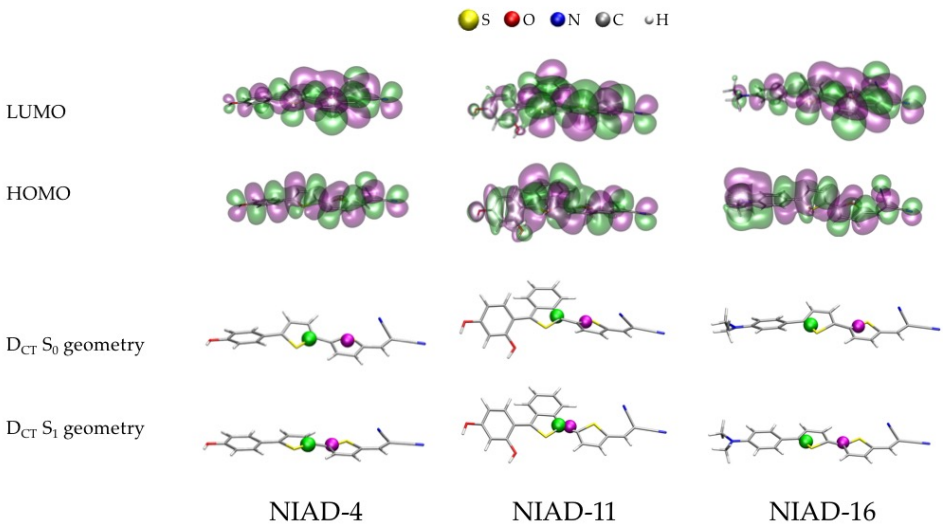


Figure 3.6: Barycenters of charge depletion (green) and increase (pink) and HOMO and LUMO orbitals at the S_0 optimized geometry.

The extent of this transfer can be evaluated using the charge transfer index D_{CT}, proposed by Ciofini and coworkers,[165,166] which is defined as the distance between the barycenters of the regions in which the electron density increases or decreases as a result of the electronic transition. A value of zero of this index indicates that the system under study has no charge transfer character, while increasing values are indicative of an increasing spatial extent of the charge transfer. The calculation uses the density difference $\Delta\rho(\boldsymbol{r})$ between the two electronic states that are involved in the transition, S_1 and S_0 in this case, which is used to define the regions of charge increase $\rho_+(\boldsymbol{r})$ and depletion $\rho_-(\boldsymbol{r})$ as follows.

$$\rho_+(\boldsymbol{r}) = \begin{cases} \Delta\rho(\boldsymbol{r})\ if\ \Delta\rho(\boldsymbol{r}) > 0 \\ 0\ if\ \Delta\rho(\boldsymbol{r}) < 0 \end{cases} \qquad \rho_-(\boldsymbol{r}) = \begin{cases} \Delta\rho(\boldsymbol{r})\ if\ \Delta\rho(\boldsymbol{r}) < 0 \\ 0\ if\ \Delta\rho(\boldsymbol{r}) > 0 \end{cases}$$

For NIAD-4, the distance between the barycenters of $\rho_+(\boldsymbol{r})$ and $\rho_-(\boldsymbol{r})$, the D_{CT} index, has been computed with a custom developed program, and has a value of 3.0 Å at the ground state optimized geometry, confirming the charge transfer character of this marker. Remarkably, when molecular geometry is allowed to relax in the excited state, a reorganization of the π conjugated system is observed, accompanied by a planarization, which results in a significant reduction of the charge transfer index, which for NIAD-4 has a value of 1.8 Å at the S_1 optimized geometry.

A similar situation is observed for NIAD-11 and NIAD-16. The S_1 optimized geometry of NIAD-16 is essentially planar for all conformers. In contrast, for NIAD-11 and regardless of the conformation isomer, the S_1 optimized geometry assumes a value of dihedral α of about 40°, probably to avoid steric clashes between the fused aromatic ring and the adjacent disubstituted phenyl ring. In the resulting geometry, the bithiophene unit and the propanedinitrile group are almost coplanar, while the phenyl ring is twisted (Table B.6 and Figure B.6). Overall, computed emission wavelengths reproduce satisfactorily experimental values, with a little overestimation.[92]

The partial compensation of charge separation observed upon geometry reorganization is confirmed by the values of dipole moment of the molecules. For NIAD-4, at the ground state optimized geometry, S_0 has a

dipole moment of 9.7 D, while S_1 has a dipole moment of 16.3 D; at the S_1 optimized geometry, on the other hand, S_0 has a dipole moment of 12.6 D and S_1 of 15.8 D. This means that i) S_1 has a more pronounced dipolar character than S_0 independently of molecular geometry, which is a further confirmation of the charge transfer character of the molecule, and ii) that geometry relaxation in S_1 lowers the dipole moment of the excited state, partially compensating the charge separation.

Table 3.2: Computed excitation and emission wavelength (nm) and oscillator strength in gas phase and water (in parentheses) of the most stable isomer of NIAD-4, NIAD-11 and NIAD-16, along with charge transfer index D_{CT} (Å) and dipole moments μ (D) of the ground and excited states at the S_0 and S_1 (in square parentheses) optimized geometries of NIAD-4, NIAD-11 and NIAD-16 in gas phase.

molecule	λ_{abs}	f	λ_{em}	f
NIAD-4	407 (442)	1.19 (1.36)	510 (651)	1.36 (1.76)
NIAD-11	469 (509)	0.95 (1.16)	573 (775)	1.02 (1.58)
NIAD-16	432 (469)	1.29 (1.47)	513 (751)	1.47 (2.04)
molecule	D_{CT}	$\mu(S_0)$	$\mu(S_1)$	
NIAD-4	3.0 [1.9]	9.7 [12.6]	16.3 [15.8]	
NIAD-11	2.5 [0.9]	11.7 [14.7]	16.8 [16.2]	
NIAD-16	3.9 [3.1]	12.4 [15.2]	23.3 [22.2]	

The larger dipole moment of S_1 compared to S_0 can be appreciated in the experimental absorption and fluorescence spectra: while in all solvents except water all absorption maxima fall approximately at the same wavelength (Figure 3.2a), emission maxima are red-shifted as the polarity of the solvent increases (Figure 3.5). This is due to the different sensitivity of S_0 and S_1 to the dielectric constant of the environment; while recording

absorption spectra, indeed, the ground state is observed, which has the lower dipole moment, and whose spectroscopic properties are not significantly affected by the dielectric constant of the environment, and this is why the maximum absorption wavelength is not affected by solvent polarity. While recording fluorescence spectra, on the other hand, the excited state is observed, which, having a higher dipole moment than the ground state, is more sensitive to solvent polarity, and becomes more stabilized in solvents of high dielectric constant. This means that the energy difference between S_1 and S_0 becomes increasingly smaller with the increasing polarity of the solvent, which explains the solvatochromic effect.[91] Similar considerations may be drawn for NIAD-11 and NIAD-16.

Comparing the values of D_{CT} of the three molecules, it is evident that NIAD-11 has a less pronounced charge transfer character than NIAD-4, probably due to the participation of the additional aromatic ring to the electron donation, which produces a deviation from planarity owing to sterical hindrance. In contrast, NIAD-16 has a more pronounced charge transfer character than NIAD-4. This may be attributed to the fact that the dimethylamino group is a better electron donor that the hydroxyl group, as already observed in the marginally higher rotational barriers around dihedral angles α, β and γ (Figure 3.1). This observation receives further confirmation by comparing the values of emission wavelength of the three compounds reported in Table 3.2: the fluorescence increase between water and gas phase, which arises form the large dipole moment of the excited state, grows in the order NIAD-11 $<$ NIAD-4 $<$ NIAD-16.

3.3 Deactivation mechanisms

Intersystem crossing

A striking feature of the fluorescence spectra of NIAD-4 is the variation of fluorescence intensity in different solvents. Indeed, fluorescence intensity increases with the polarity of the solvent from diethylether to dichloromethane to methanol, while it is null in water (Figure 3.5). In this regard, it is interestig to explore the nonradiative deactivation pathways of this molecule, which compete with fluorescence emission and can justify the dependence of fluorescence on the environment.

One of the possible explanations for the dependence of fluorescence intensity on solvent polarity is an intersystem crossing involving the emitting state S_1 and triplet T_2, which is the triplet state closest in energy to S_1. The rate constant of intersystem crossing is reported in eq. 3.3, and depends on both the energy difference between the two states, ΔE, and the spin orbit coupling matrix element, H_{SO}. λ is the Marcus reorganization energy.

$$k_{ISC}^{IF} = \frac{2\pi}{\hbar}\langle {}^1\Psi_I^0|\hat{H}_{SO}|{}^3\Psi_F^0\rangle^2 \frac{1}{\sqrt{4\pi\lambda RT}} exp\Big[-\frac{(\Delta E+\lambda)^2}{4\lambda RT}\Big] \qquad (3.3)$$

Indeed, intersystem crossing is an important deactivation pathway for oligothiophenes.[167] In particular, it has been shown that, according to the symmetry selection rules for spin-orbit coupling, ISC has a small prob-

ability to occur in planar, symmetric, conformations and becomes more likely with deviation from planarity.[167]

For NIAD-4, rotation around dihedral α, which has a barrier of $\sim$ 10 $kJ\,mol^{-1}$ in the ground state, cannot be responsible for an intersystem crossing since a preliminary scan of S_1 and T_2 vertical excitation energies around this angle reveal that both states present a maximum for β= 90 °, indicating that the energy gap between the two states is roughly constant for every value of α (Figure 3.7). The rotation around the single bond linking the thiophene ring to the propanedinitrile group, dihedral γ, on the other hand, has not been explored because it involves a much higher rotational barrier, $\sim$ 38 $kJ\,mol^{-1}$ (Table B.1), and therefore is not expected to contribute to molecular flexibility. Indeed, Bae *et al.* indicated that S_1 may present a shallow local minimum for γ = 90°, but separated from the ground state geometry by a significant energy barrier.[90] If this minimum was easily accessible, NIAD-4 would behave as a molecular rotor, like thioflavin-T (see Introduction). In such molecules, after excitation, the molecule evolves from an approximately planar conformation to an internal charge transfer twisted state. This behavior leads to a decreasing fluorescence emission with the increasing polarity of the solvent. A polar solvent, indeed, stabilizes the internal charge transfer state. This is in contrast with our experimental results, which indicate an increase in fluorescence moving from diethylether ($\epsilon = 4$) to methanol ($\epsilon = 33$).

At the DFT optimized geometry of NIAD-4, the torsional angle between the two thiophene rings, β, has a value of - 14° (Table B.1), with a relatively low rotational barrier of $\sim$ 18 $kJ\,mol^{-1}$, indicating that the molecule in solution has a certain flexibility around this angle. For this

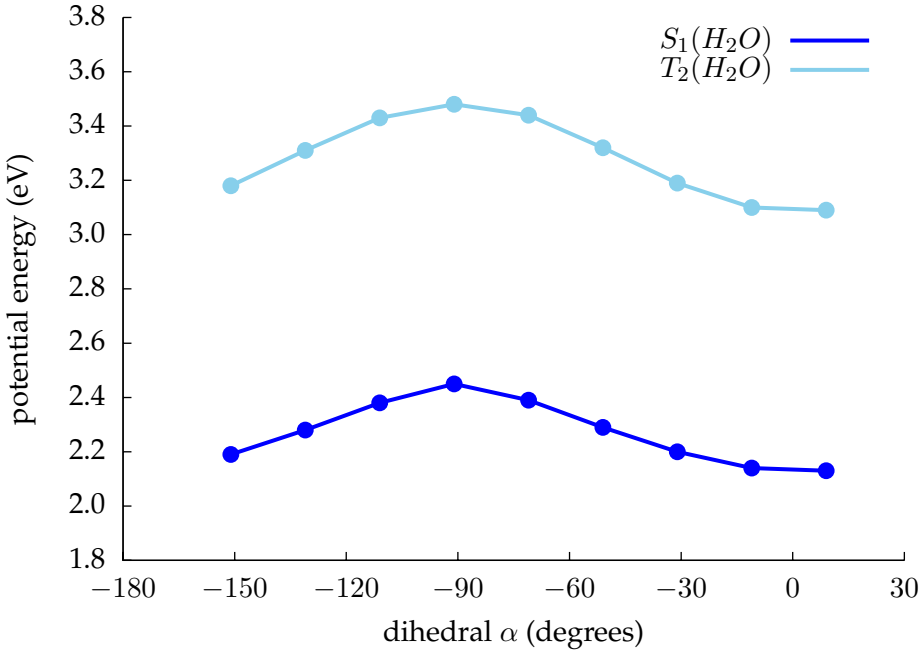


Figure 3.7: Rigid scan around dihedral α of NIAD-4: vertical excitation energies to S_1 and T_2 in water (PCM model).

reason, in order to evaluate the likeliness of an intersystem crossing in different solvents, TDDFT calculations of S_1 and T_2 excitation energies were carried out at the optimized S_1 geometry for several fixed values of dihedral β. Moreover, since TDDFT triplet excitation energies may be too low, as already mentioned in Chapter 2, the curves were corrected by adding a uniform shift corresponding to the difference between MS-CASPT2 and TDDFT values computed at the TDDFT S_1 optimized geometry (Table B.3).[124] An active space of 12 orbitals was chosen describing the π system, and 12 electrons were considered. The uniform correction arising from these calculations is of + 0.14 eV for S_1 and + 0.94 eV for T_2. Corrected values are reported in Figure 3.8 and Table B.2.

Results show that T_2 lies above S_1 regardless of the nature of the solvent. Moreover, S_1 has a maximum and T_2 a minimum at β = 90°, again for all solvents, so that the minimum S_1/T_2 energy gap occurs for β = 90°. This means that the intersystem crossing will become more favorable as deviation from planarity increases. According to eq. 3.3, for an intersystem crossing to be effective, not only the two states need to be close in energy, but they must also show a significant spin-orbit coupling. The spin-orbit coupling between S_1 and T_2 has been computed at the MS-CASPT2 level at four different geometries: the TDDFT optimized S_1 geometry, which is fully planar, and three derived geometries obtained by setting the value of dihedral β to 20, 30 and 40°. The computed spin-orbit coupling values are negligible for the first two geometries and 0.5 and 2.0 cm^{-1} for the 30 and 40° geometries respectively, which mirrors the behavior of oligothiophenes and goes in the same direction as the energy difference. This suggests that intersystem crossing will be negligible for planar conformation and will become increasingly important as deviation

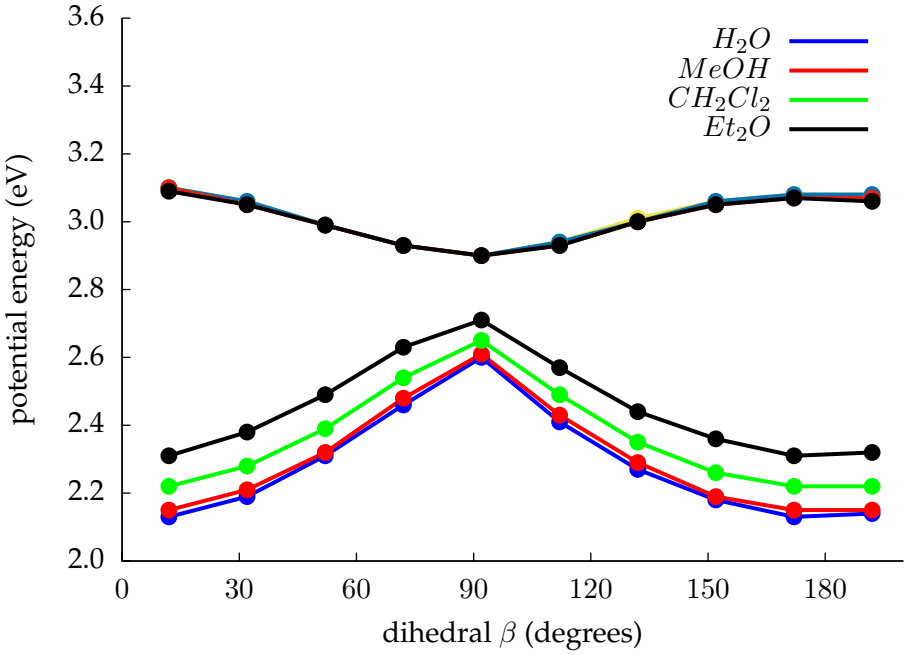


Figure 3.8: Relaxed S_1 energy scan around dihedral β; excitation energies (eV) to S_1 and T_2 computed in gas phase and several solvents (PCM model). Labels refer to Figure 3.1.

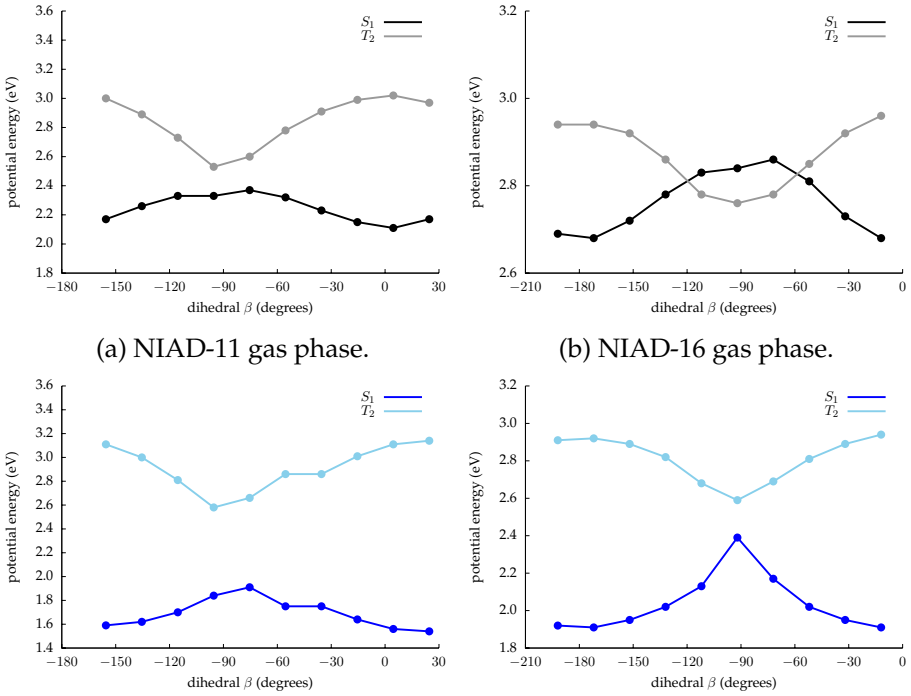
from planarity increases.

Concerning the effect of the dielectric constant of the environment, while the energy of T_2 is essentially unaffected by polarity, the stabilization of S_1 increases with the dielectric constant of the solvent due to its charge transfer character, as already mentioned in section 3.1. This results in higher S_1/T_2 gaps and thus, in less effective intersystem crossing in polar solvents. Since fluorescence and intersystem crossing are competing

deactivation processes, this explains the increasing fluorescence intensity with the increasing dielectric constant of the solvent reported in Figure 3.5: a more effective intersystem crossing results in a lower quantum yield of fluorescence and viceversa. This receives a further confirmation from the calculation of the oscillator strengths at the S_1 geometry optimized in water, methanol, dichloromethane and diethylether. These values are only slightly affected by the nature of the solvent, and for this reason does not reflect the experimental difference among fluorescence intensities.

These calculations are qualitative and do not allow the determination of an intersystem crossing rate constant, however some qualitative conclusions can be drawn. The small values of spin-orbit coupling even at highly torsioned geometries ($\sim 40°$) suggest that fluorescence is more favorable than intersystem crossing in the whole range of solvent polarity that has been explored, and that intersystem crossing can at best tune fluorescence intensity, which decreases with decreasing solvent polarity. These observations, however useful to rationalize the general behavior of NIAD-4 in solution, do not explain the peculiar photophysical properties observed in water. According to the trend observed for the other solvents, indeed, one would expect an absorption maximum in water in the 471 - 480 nm range, and an intense emission peak at $\lambda > 650$ nm. Contrary to what expected, the absorption maximum falls at 429 nm, and the quantum yield of fluorescence is vanishingly small. This peculiar behavior has beeen shown not to depend on solvent polarity, because the same absorption and emission spectra recorded in a $1:2$ mixture of N-methylformamide and methanol, which has a dielectric constant of 85, very close to the value of 80 of water, show just the expected behavior (Figures 3.2a and 3.5). This means that there is a property of water, other than its polarity, which severely affects

the optical properties of NIAD-4. and will be discussed further on.



(a) NIAD-11 gas phase.

(b) NIAD-16 gas phase.

(c) NIAD-11 water.

(d) NIAD-16 water.

Figure 3.9: Relaxed S_1 energy scan around dihedral β; excitation energies (eV) to S_1 and T_2 computed in gas phase and water (SMD model).

The same deactivation through intersystem crossing was explored for NIAD-11 and NIAD-16. Again, the singlet and triplet energies were corrected with the uniform shift computed as the difference between MS-CASPT2 and TDDFT (Table B.3) S_1 and T_2 energies at the S_1 optimized geometry. This correction is of - 0.08 eV for S_1 and + 1.13 eV for T_2 in the case of NIAD-11, and + 0.26 eV for S_1 and + 0.85 eV for T_2 in the case

of NIAD-16. Computed values for NIAD-11 in gas phase indicate that, similarly to NIAD-4, T_2 is higher in energy than S_1 for all geometries, the triplet being an energy minimum and the singlet a maximum for β = 90° (3.9a). This torsioned geometry corresponds to the smallest value of the S_1/T_2 gap and thus, to the geometry for which intersystem crossing is expected to be more efficient. For NIAD-16, however, the computed energy profiles for S_1 and T_2 indicate that the two states cross for a value of the dihedral angle of approximately β=120° (Figure 3.9b). In aqueous solution, and similarly to NIAD-4, the singlet state of both NIAD-11 and NIAD-16 is stabilized due to its large charge transfer character, while the triplet state remains almost unchanged (Figure 3.9c and 3.9d). This suggests that a variation of the fluorescence quantum yield as a function of the polarity of the solvent can be observed for these molecules as well, with fluorescence becoming more intense with the increasing polarity of the environment, as the intersystem crossing involving S_1 and T_2 becomes less effective.[92]

These calculations indicate that the energy of S_1 and T_2 states follow the same trend for the rotation around dihedral β in all three markers. Moreover, our calculations for NIAD-4 indicate that the spin-orbit coupling, which concurs to the efficiency of the intersystem crossing along with the energy difference of the states that are involved, is null for a planar geometry and increases for increasing values of dihedral β. This result, along with the TDDFT scans, indicates that if an intersystem crossing is present in NIAD markers, it will be favored by twisted geometries rather than planar ones.

Besides polarity, also viscosity is expected to affect the quantum yield of fluorescence of NIAD markers in solution. Indeed, since the efficiency

of the intersystem crossing is strictly related to the degree of planarity of the molecule, if the viscosity if the environment increases, molecular flexibility becomes hidered and fluorescence becomes more favorable. Moreover, flexibility may affect the quantum yield of fluorescence also by tuning the efficiency of a second nonradiative deactivation process of the molecule.

Internal conversion

In dyes containing a conjugated π system, the rotation around a double bond may lead to a conical intersection between the excited and the ground state. This is, for instance, the case of retinal.[168,169] In order to explore this possibility for NIAD-4, we analyzed the effect of solvent viscosity on fluorescence independently of polarity effects. To do so, the quantum yield of fluorescence of NIAD-4 was measured in mixtures of glycerol and methanol having approximately the same dielectric constant ($\epsilon = 33 \div 42$) and viscosity ranging from 0.6 to 1285 cP. Fluorescence quantum yields were determined relative to N-N'-bis(*sec*-butyl)-1,6,7,12-tetra-(4-*tert*-butylphenoxy)-perylene-3,4:9,10- tetracarboxylic diimide in dichloromethane.[170] Room temperature viscosities for the mixtures of glycerol and methanol were taken from reference[171]. Results are reported in Figure 3.10, and clearly indicate that the quantum yield of fluorescence increases with viscosity.

For NIAD markers, we expect the ground state and S_1 excited state to become degenerate for a torsion of 90° around the double bond of the propanedinitrile unit (dihedral Φ of Figure 3.11). Due to their intrinsically

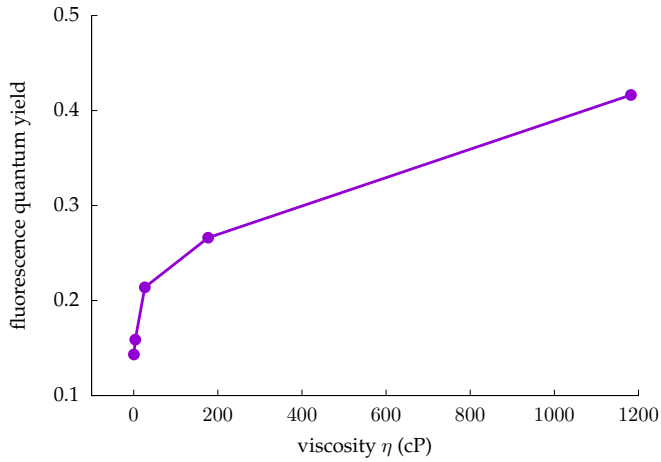


Figure 3.10: Quantum yield of fluorescence of NIAD-4 (c_{NIAD-4} = 5 μM) in solutions of increasing viscosity: pure methanol, mixtures of methanol and glycine 3 : 1, 1 : 1 and 1 : 3, and pure glycerol.

multireference nature, conical intersection geometries cannot be investigated with DFT-based methods, which are single reference. However, it has been shown that in the case of a model system of retinale, TDDFT can reproduce qualitatively results obtained with more accurate multireference methods.[169]

For this reason, the S_1 energy profile for the rotation around dihedral Φ of NIAD-4 has been explored with TDDFT, and is reported in Figure 3.11. Scan points with Φ between 0 and 62° are the result of a geometry optimization fixing the value of Φ, while points with $\Phi > 62°$ are single point S_1 energy calculations on geometries obtained modifying only the value of Φ at the 62° optimized geometry. Once again, it must be stressed that these results are qualitative, but they offer a rough estimation of the

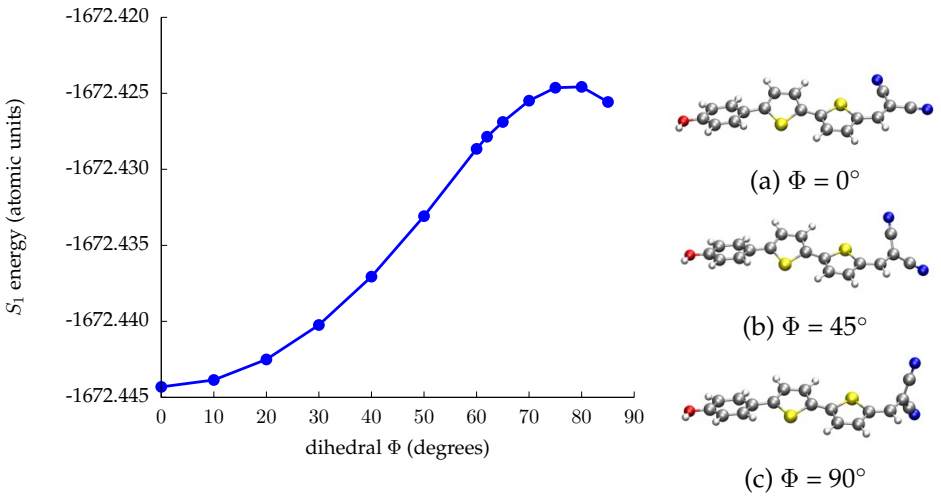
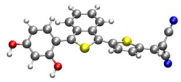


Figure 3.11: S_1 rotational scan around the double bond connecting a thiophene ring with the malononitrile unit of NIAD-4, Φ.

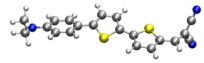
energy barrier for the rotation around Φ. The highest point of our scan, which is located at an 80° geometry, is 52 $kJ\,mol^{-1}$ above the S_1 optimized geometry, and only 24 $kJ\,mol^{-1}$ above the S_1 energy at the ground state optimized geometry (S_0), which corresponds to the S_1 energy of a vertical excitation. These numbers suggest that deactivation through internal conversion is feasible and may account for the variation of the quantum yield of fluorescence as a function of solvent viscosity, along with the intersystem crossing.[91]

Similar calculations were performed for NIAD-11 and NIAD-16. Results are reported in Figure 3.12. The highest point of the scans, Φ=87° is 81 $kJ\,mol^{-1}$ higher in energy than the S_1 optimized geometry for NIAD-11 and 70 $kJ\,mol^{-1}$ for NIAD-16, while it is located only 55 $kJ\,mol^{-1}$ and 44

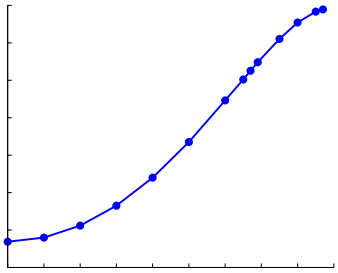
$kJ\,mol^{-1}$ above the S_0 minimum for NIAD-11 and NIAD-16 respectively.



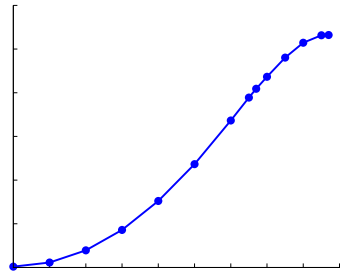
(a) Φ = 90°



(b) Φ = 90°



(c) NIAD-11.



(d) NIAD-16.

Figure 3.12: Rotational scan around the double bond connecting a thiophene ring with the malononitrile unit, Φ.

The propensity of the markers to decay through internal conversion follows the order NIAD-4 > NIAD-16 > NIAD-11. This order is coherent with the previous observation of the electron delocalization properties of the three markers: in both NIAD-11 and NIAD-16, electron delocalization is enhanced compared to that of NIAD-4, in the first case due to the extension of the π system, and in the second due to the increased donor capabilities of $-N(CH_3)_2$ over $-OH$.[92]

These results help rationalize the variation of the quantum yield of fluorescence as a function of solvent viscosity and polarity, but yet they do not explain the optical properties of NIAD-4 in water. In particular, it is left to explain why only in water, among the solvents that have been consid-

ered, NIAD-4 presents only negligible fluorescence emission (Figure 3.5) and a significant blue shift in the absorption spectrum (Figure 3.2a), which is observed for NIAD-16 as well.

3.4 Aggregation

In view of the incapability to disclose the photophysical behavior of NIAD-4 and NIAD-16 in aqueous solution using purely solvent polarity and viscosity arguments, additional factors were investigated. Owing to the organic nature of these bithiophene-based markers, and its likely low solubility in water, the effect of dye aggregation on its optical behavior in aqueous media was considered. Concentration-dependent absorption spectroscopy and fluorescence excitation spectroscopy measurements were performed. Absorption spectra of NIAD-4 in concentrations ranging from 0.48 to 9.43 μM uncovered large spectral variation, which only resembled that expected for the monomeric form of the dye ($\lambda_{max,abs} \sim 475$ nm) at 0.48 μM, as reported in Figure 3.14. Upon adding further amounts of NIAD-4, a new absorption band centered at 430 nm develops and becomes predominant.

This kind of spectral response is typical of co-facial self-assembly driven by π-stacking interactions, as illustrated in Figure 3.15. The corresponding parallel (non-centrosymmetric) and antiparallel (centrosymmetric) π-stacking dimers reported in Figure 3.15 both present blue-shifted absorption compared to the monomer and quenched fluorescence emission.[172] Calculations were performed for two stacked dimeric structures at different interchromophoric distances, ranging from 3.6 to 4.6 Å. The models of

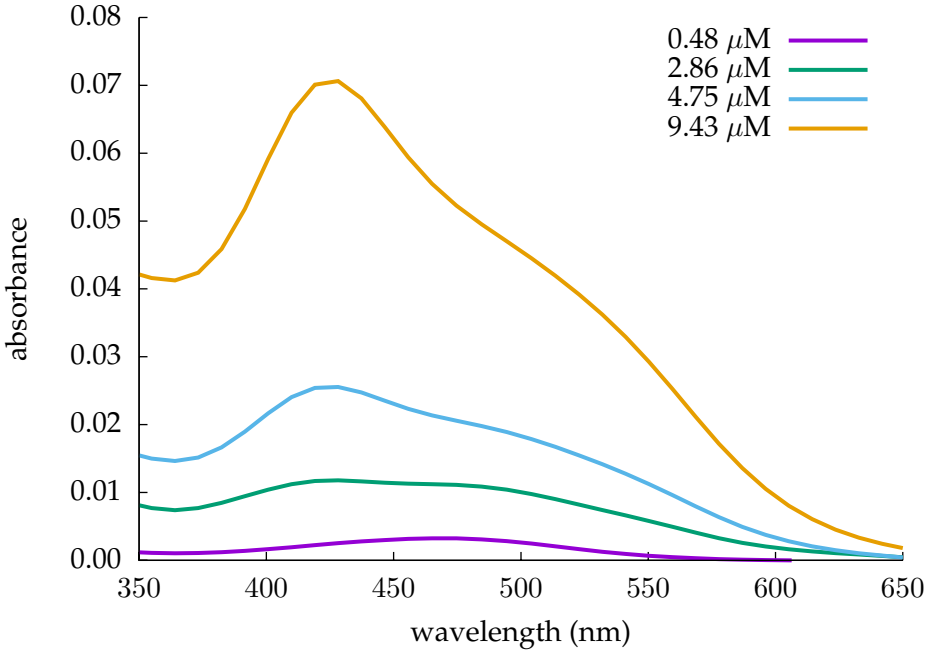


Figure 3.14: Concentration dependent aggregation experiments of NIAD-4 in water.

the non-centrosymmetric and centrosymmetric dimers were constructed from the ground state optimized geometry of NIAD-4 by setting a fixed distance between the monomers (Figure 3.15). In both the parallel and antiparallel dimer model, the π interaction operates a splitting from a single excited level of the monomer to two levels of the dimer, one higher and one lower in energy than that of the monomer. In both dimer models, excitation to the the lower of these levels is forbidden due to symmetry reasons (f = 0), while excitation to the higher has a large oscillation strength ($\sim$ 2.3) and corresponds to the blue-shifted peak observed in the experimental spectrum (Figure 3.2a). Absorption wavelengths and oscillator strength for more interchromophoric distances are reported in Table B.4.

In addition, a weaker red-shifted band appears at around 500 nm in the experimental spectrum, which may be attributed to the formation of cofacially-stacked assemblies with non-zero offset displacement between monomeric units. In such a case, the optical transition to the lowest of the two states would be weakly allowed.[173,174] Concentration dependent absorption spectra recorded in a mixture of methanol and water are reported in Figure B.3. They are similar to the spectra recorded in pure water, but, since NIAD-4 is soluble in methanol, aggregation begins at higher concentrations.

Furthermore, fluorescence excitation spectra of NIAD-4 in methanol and water have been recorded and are reported in Figure 3.16. This technique provides the absorption spectrum of fluorescence emitting species only. In water only a very dim emission is observed, with a quantum yield lesser than 0.01 (Figure 3.17a).

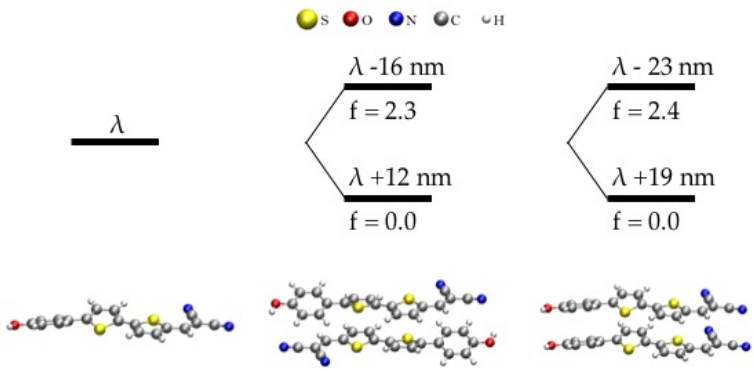
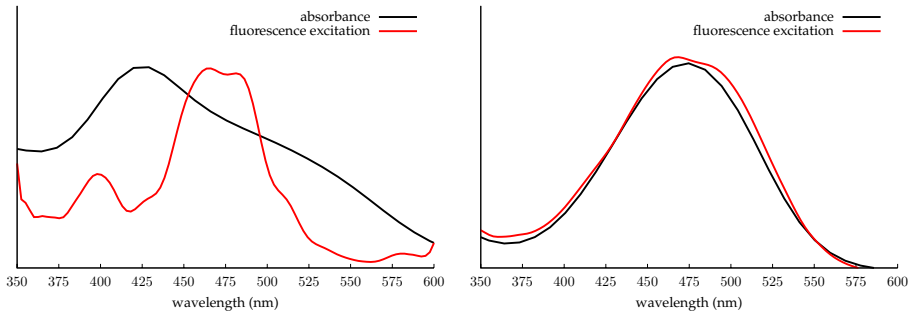


Figure 3.15: Computed optical properties of the centrosymentric and non-centrosymmetric co-facially stacked dimers of NIAD-4 with monomers at a distance of 4.2 Å in water (PCM): absorption wavelength and oscillator strength of the π-stacking dimers of NIAD-4 compared to the monomer.

This fluorescence does not originate from the aggregates, but rather from the little amount of monomeric NIAD-4 species still remaining in solution ($\lambda_{abs} \sim 470$ nm), which confirms that the vast majority of NIAD-4 in water is aggregated. This is in contrast to the nearly identical absorption and fluorescence excitation spectra recorded in methanol (Figure 3.17b), which indicates that in this solvent the species that absorbs and emit radiation is the same. This is due to the fact that methanol is a good solvent for NIAD-4, and aggregation does not take place.

These experimental and computational data allow us to conclude that the spontaneous formation of fluorescence-quenched aggregates in water already at very low concentrations (μM) is the main reason for the lack fluorescence emission in aqueous solution of NIAD-4. This fluoresce quenching is reverted when NIAD-4 binds to amyloid-β fibrils. As will



(a) Water. (b) Methanol.

Figure 3.16: Absorption and fluorescence excitation spectra of NIAD-4 at $c_{NIAD-4} = 5\ \mu M$.

be discussed in detail in Chapter 5, the marker is accommodated into hydrophobic channels defined by the hydrophobic side chains that arise from the cross-β architecture of the fibrils and run parallel to the fibril axis. These hydrophobic channels have the correct size to host only the monomeric form of the marker, which emits fluorescence, and not the fluorescence quenched dimer. Thus, disaggregation of the fluorescence-quenched dimers and diffusion of the emitting monomer into the channels located at the core of amyloid fibrils is the fundamental reason for the fluorescence enhancement observed when NIAD-4 interacts to amyloid fibrils compared to the free marker in aqueous solution.

Regarding NIAD-16, a behavior similar to that of NIAD-4 is expected. As already mentioned, substitution of an $-$OH by an $-\mathrm{N(CH_3)_2}$ group is expected to lower the solubility, making NIAD-16 even more prone to aggregation than NIAD-4. Since no experimental information on the entity of the fluorescence enhancement upon binding to amyloid-β fibrils is avail-

able for NIAD-16, at least to our knowledge, conclusions may be drawn only by comparing our calculations to the experimental values of maximum absorption and fluorescence in saline solution: the 55 nm blue shift of the experimental value from the calculated one suggests aggregation.

For NIAD-11, on the other hand, the calculated[91] and experimental[46] absorption maximum recorded in saline solution are 550 and 545 nm respectively. This agreement seems to indicate that NIAD-11 does not have the same propensity to aggregation as the other two. This can be justified based on structural differences: despite presenting an additional fused aromatic ring on a thiophene, NIAD-11 has one hydroxyl group more than NIAD-4. This additional polar group is expected to improve significantly the solubility of the marker acting as hydrogen bond donor and acceptor. If NIAD-11 does not aggregate in water at the concentrations employed for recording the experimental spectra[54], then the 10-fold fluorescence enhancement upon binding is due to an effect of sterical hindrance of the marker when accommodated in the hydrophobic channels of amyloid fibrils. Here molecular flexibility is reduced, so that the rate of nonradiative deactivation processes that involve torsions, such as intersystem crossing and internal conversion, is reduced, favoring the competing fluorescence emission. Since these nonradiative deactivation processes are not fast enough to cancel out fluorescence, not even in the case of the marker in aqueous solution, the fluorescence enhancement observed for NIAD-11 is significantly smaller than that observed for NIAD-4.

3.5 Final remarks and graphical summary

The photophysical properties of the bithiophene-derived markers NIAD-4, NIAD-11 and NIAD-16 were studied with a combination of computational and experimental techniques. The large discrepancy between calculated and experimental absorption maximum of NIAD-4 and NIAD-16 in water suggests that in this solvent the two markers may undergo aggregation. This has been confirmed in the case of NIAD-4 by performing concentration dependent absorption spectroscopy and fluorescence excitation measurements. As a further confirmation, calculations attributed the experimental maximum absorption to the parallel and antiparallel π-stacked dimers of NIAD-4, whose fluorescence is quenched. In the absence of experimental absorption and fluorescence spectra of NIAD-16 in water, a similar behavior has been hypothesized due to the 50 nm blue shift of the experiental absorption maximum compared to the computed value.

Due to their donor-acceptor structure (donor: hydroxyl / dimethylamino groups; acceptor: propanedinitrile group), all three markers undergo a charge transfer transition, resulting in an emitting excited state S_1 with a high dipole moment. Values of the D_{CT} index evaluating the spatial extent of the charge transfer grow in the order NIAD-11 $<$ NIAD-4 $<$ NIAD-16. A large value of D_{CT} corresponds to a high dipole moment of the excited state. This is confirmed by the experimental fluorescence spectra of NIAD-4 recorded in several solvents, which indicates the presence of a solvatochromic effect, with the fluorescence maximum being red shifted with the increasing solvent polarity. Moreover, also the quantum yield of fluorescence depends on solvent polarity, increasing with the increasing dielectric constant of the environment. This can be attributed, at least

partially, to an intersystem crossing involving the emitting state S_1 and a triplet lying close in energy T_2. The dependence of S_1 energy on solvent polarity leads to a solvent dependent S_1/T_2 energy gap, and thus to a solvent dependent intersystem crossing efficiency; in particular, intersystem crossing is more efficient in solvents of low dielectric constant, and less efficient in highly polar solvents. This correlates well with the experimental observation that the quantum yield of fluorescence increases with solvent polarity. The energy profiles of S_1 and T_2 in water for the three markers indicate that the propensity to undergo intersystem crossing grows in the order NIAD-11 $<$ NIAD-4 $<$ NIAD-16.[92]

Furthermore, also viscosity affects the fluorescence properties of the three markers, with the quantum yield of fluorescence increasing with the increasing viscosity. This can be partially explained in terms of the intersystem crossing, since it becomes more effective as the molecule deviates from planarity, and in a viscous medium flexibility is hindered, but it can also originate from a second nonradiative deactivation process, an internal conversion. Indeed, the rotation around the double bond of the propanedinitrile group leads to a conical intersection geometry , with a an energy barrier increasing in the order NIAD-4 $<$ NIAD-16 $<$ NIAD-11. Again, by hindering molecular flexibility, viscosity opposes the nonradiative deactivation processes, which involve molecular torsions, and favors the competing fluorescence emission.[91,92]

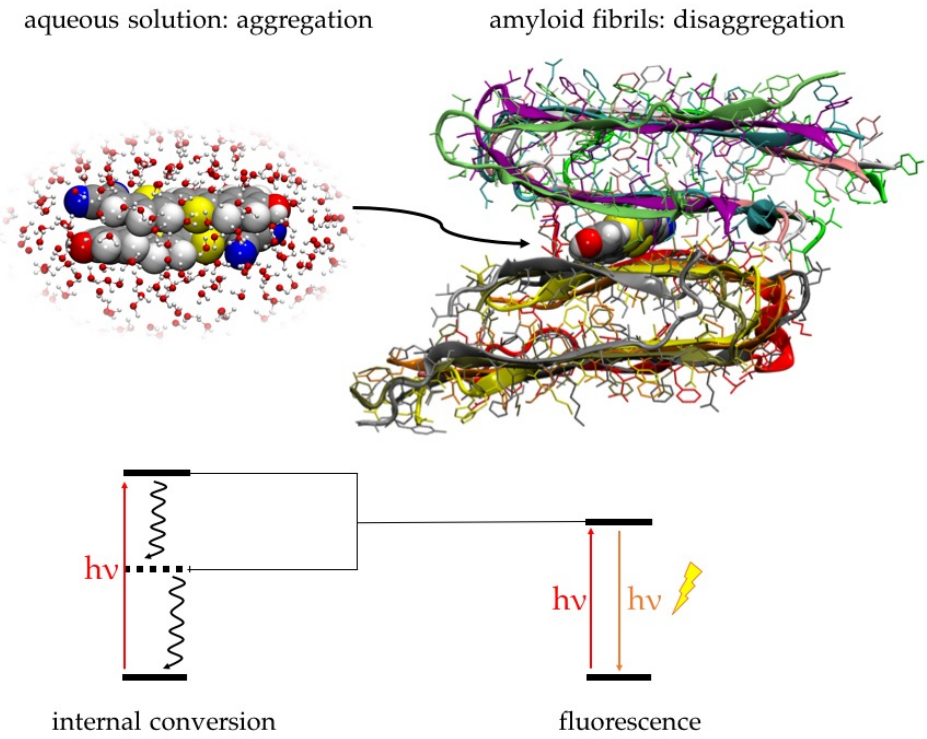


Figure 3.18: Graphical summary

Chapter 4

Conjugated π system markers

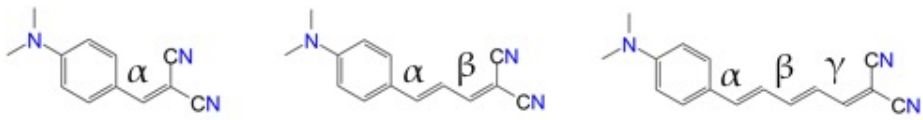
This Chapter covers the results obtained for a class of conjugated π system amyloid markers, DANIRs.[58] These molecules have been proposed in 2014 by Cui *et al.*[58], and have a push-pull architecture. They are composed of an electron donating and an electron withdrawing group linked by a linear polarizable π bridge composed by a benzene ring and a variable number of conjugated double bonds. Our results highlight the importance of the length of the backbone of the molecule in determining its performances as amyloid marker.

The largest molecule of the family, DANIR-2c, emits in the near-infrared and undergoes a fluorescence enhancement in presence of amyloid-β fibrils, which makes it a good candidate for *in vivo* application. Due to its high affinity towards amyloid-β deposits (K_i = 36.9 nM), DANIR-2c has been tested on brain slices of both transgenic mice and an Alzheimer's

disease patient. Results on mice brain slices indicate that DANIR-2c can specifically stain amyloid plaques, while results on human brain slices indicate that it can also stain tangles. Computed values of $LogBB$, which is an indicator of the propensity of the molecule to cross the blood-brain barrier (see Chapter 3), are in the optimal range - 1.0 $\div$ 0.3 for the three markers, being - 0.27, - 0.19 and - 0.11 for DANIR-2a, -2b and -2c respectively. This indicates that DANIR molecules may readily cross the blood-brain barrier after intravenous injection.

A remarkable feature of DANIR markers is the difference among values of the quantum yield of fluorescence in dichloromethane with the number of conjugated double bonds of the system. While compound 2a has a quantum yield of zero, 2b has a value 0.66 and 2c of 4.09. This increase is partially expected due the extension of π system, but not enough to explain the utter lack of fluorescence in smallest molecule compared to the largest one.[58]

This Chapter is aimed at explaining the optical properties of the molecules in both aqueous solution and in the presence of amyloid fibers. First, the absorption and emission properties of the three molecules will be presented. Then, the nature of the spectroscopically active transition will be explored with the aim of understanding the relationship between charge transfer and electron delocalization on the π system. Then, the deactivation processes that compete with fluorescence emission will be presented, and finally, it will be shown how does binding to amyloid-β fibrils affect the properties of the markers.



DANIR-2a DANIR-2b DANIR-2c

Scheme 4.1: Molecular structures of DANIR-2a, DANIR-2b and DANIR-2c.

4.1 Absorption and fluorescence

DANIR makers are composed of an electron donor (N,N- dimethyl-amino) and an acceptor (propanedinitrile) group linked by a π bridge composed of a benzene ring and a variable number of conjugated double bonds (Scheme 4.1), 1, 2 and 3 in DANIR-2a, -2b and -2c respectively. Due to the delocalized π system, the optimized ground state geometry of these molecules is essentially planar, with all dihedral angles being 0 or 180° and bond lengths typical of substituted aromatic rings and conjugated double bonds (Table C.1, Figure C.1).

Table 4.1: Computed rotational barriers ($kJ\,mol^{-1}$) of DANIR-2a, DANIR-2b and DANIR-2c in gas phase and water, in parentheses.

	2a	2b	2c
$\Delta G^{\dagger}\ TS_{\alpha}$	43.9 (42.3)	47.4 (41.8)	35.1 (34.9)
$\Delta G^{\dagger}\ TS_{\beta}$	-	73.5 (58.3)	52.6 (49.2)
$\Delta G^{\dagger}\ TS_{\gamma}$	-	-	59.1 (52.7)

Rotational barriers around the single bonds of the conjugated π sys-

tem α, β and γ are reported in Table 4.1 and are responsible for conformational isomerism in compounds 2b (2 isomers) and 2c (4 isomers). In all cases, the linear isomer reported in Scheme 4.1 is by far the most stable, and thus the most relevant for the discussion of molecular properties (Figure 4.1 and Table C.2). Indeed, in the case of DANIR-2c, the second most stable isomer is 11.8 $kJ\,mol^{-1}$ higher in energy the the linear one, with a corresponding Boltzmann weight of less than 0.01. This indicates indicates that all other conformations will be negligible compared to the linear one.

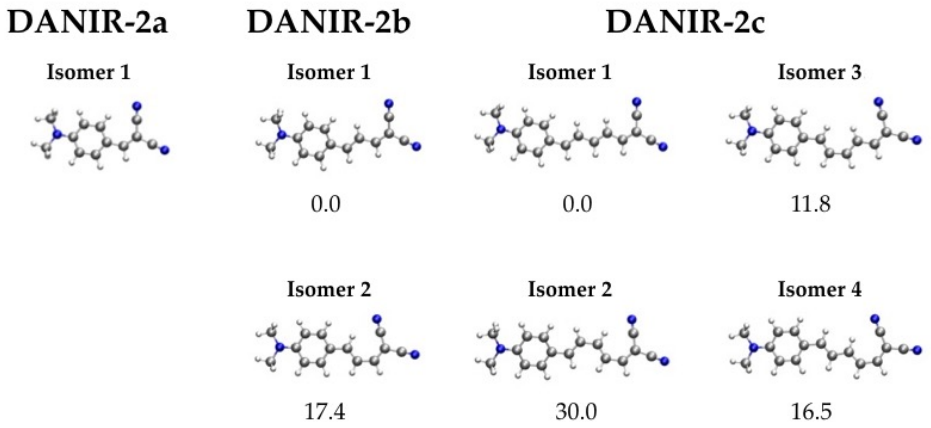
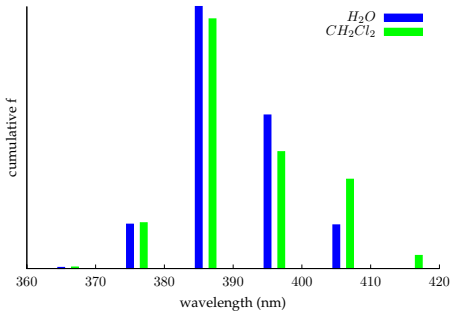


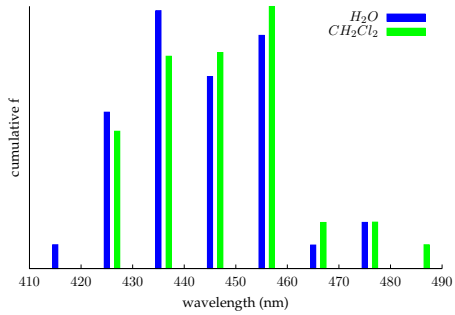
Figure 4.1: Conformational isomers of DANIR-2a, DANIR-2b and DANIR-2c, with indication of relative free energies ($kJ\,mol^{-1}$).

The computed absorption wavelengths in dichloromethane of the most stable isomers are 376, 431 and 473 nm for compounds 2a, 2b and 2c, respectively. Changing the solvent from dichloromethane to water has very little effect on the absorption wavelength but both the values in dichloromethane and water are between 30 and 68 nm larger than the values

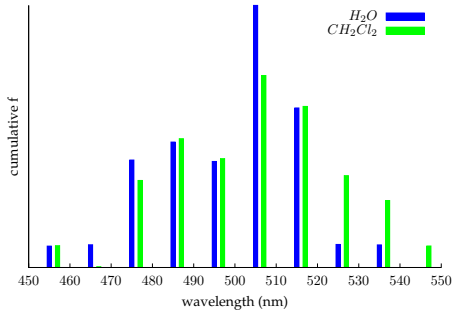
computed in gas phase. The large preference for the linear isomer and the high rotational energy barriers ($\Delta G^{\dagger} > 30\ kJ\,mol^{-1}$) indicate that the molecule will exhibit little flexibility in aqueous solution. Therefore, dye flexibility is expected to affect absorption properties only to a small extent. In any case, an *ab initio* molecular dynamics simulation for each molecule was performed with the aim of reproducing the experimental spectra.



(a) DANIR-2a.



(b) DANIR-2b.



(c) DANIR-2c.

Figure 4.2: Calculated absorption spectra from *ab initio* molecular dynamics simulations in water and dichloromethane of DANIR-2a, DANIR-2b and DANIR-2c.

Table 4.2: Computed absorption and emission wavelength (nm) of the three markers in gas phase, dichloromethane and water from *ab initio* molecular dynamics simulations λ_{max}, vertical excitations computed at the ground state optimized geometry with indication of the oscillator strength in gas phase, dichloromethane and water λ_{abs}; computed emission wavelengths λ_{em} in gas phase, dichloromethane and water; experimental absorption in dichloromethane, emission in PBS (phosphate saline buffer) and corresponding quantum yield of fluorescence are reported for comparison. Wavelengths in nm.

	2a	2b	2c
$\lambda_{abs,gp}$ [f]	336 [0.9]	375 [1.3]	406 [1.8]
λ_{abs,CH_2Cl_2} [f]	376 [1.1]	431 [1.6]	473 [2.0]
$\lambda_{abs,max,CH_2Cl_2}$	385	455	505
$\lambda_{max,expt,CH_2Cl_2}$	433	489	519
λ_{abs,H_2O} [f]	377 [1.1]	425 [1.6]	470 [2.0]
λ_{abs,max,H_2O}	385	435	505
$\lambda_{em,gp}$ [f]	376 [0.8]	411 [1.3]	450 [1.9]
λ_{em,CH_2Cl_2} [f]	433 [1.3]	525 [1.7]	619 [2.2]
λ_{em,H_2O} [f]	450 [1.3]	549 [1.8]	655 [2.3]
$\lambda_{em,expt,PBS}$	487	577	665
Φ_{fl} (%)	0	0.66	4.09

Similarly to what already shown for bithiophene derivatives,[91,161–163] the absorption spectra of DANIR markers have been computed using a collection of snapshots extracted from *ab initio* molecular dynamics simulations. According to what described in Chapter 2, a 10 ps PBE-D[118,164] simulation was performed with the program CP2K[98] for the most stable isomer of DANIR-2a, DANIR-2b and DANIR-2c (Scheme 4.1). Only the linear conformational isomer was considered for each marker because it is by far the most stable in all cases (Figure 4.1 and Table C.2). For each

molecule, 40 snapshots were sampled from the *ab initio* molecular dynamics simulation, and vertical excitations computed with CAM-B3LYP in water and dichloromethane (SMD model[128]). The sum of oscillator strength values is reported in Figure 4.2. Additionally, a longer (20 ps) dynamics has been performed on DANIR-2a and DANIR-2b in order to assess whether simulation time affects the position of the absorption maximum. Results, which are presented in Figure C.4 indicate that no significant difference is observed in the spectra calculated from the longer and shorter trajectory. Absorption maxima are in fair agreement with experimental values. In general, the maximum absorption wavelength is about 10 nm red-shifted with respect to the vertical excitation of the optimized linear isomer in the case of 2a and 2b, and 30 nm in the case of 2c (Table 4.2). This effect is composed of two factors: the different levels of theory used for dynamics simulations and excited state calculations, and molecular flexibility. The latter accounts for the bandwidth, which becomes larger as the number of double bonds increases. Noteworthily, the bandwidth is always smaller than in the case of bithiophene markers, which are much more flexible and for which several isomers contribute to the absorption (Chapter 3).

Regarding the excited state geometry, S_1 minima are also flat for 2b and 2c. However, in the case of 2a, the dihedral angle between the acceptor propanedinitrile group and the aromatic ring assumes a value of 8.2° (Figure C.2). Moreover, in these minima and regardless of the molecule, four C−C bonds of the aromatic ring are longer by 0.002 ÷ 0.026 Å than those of the ground state, while the other two are shorter by 0.004 ÷ 0.012 Å. The C−C bond distances of the double bond conjugated chain also suffer significant variations, the C−C and C=C bonds becoming much more similar

in the case of the S_1 minima (Table C.1). This indicates that there is a close relationship between the electronic transition and aromaticity/degree of electron delocalization of the system, which will be discussed in the following section.[175]

The computed emission wavelengths in gas phase, dichloromethane and water are also reported in Table 4.2. Computed values are in agreement with experimental data, differences between computed and experimental values being smaller than 40 nm. These values indicate a red-shifting of the emission wavelength with increasing solvent polarity (bathochromic shift). This effect is more pronounced for compound 2c, which presents the larger number of double bonds between the aromatic ring and the acceptor group. All these data suggest that the $S_0 \rightarrow S_1$ transition has an important charge transfer character as expected for these donor–acceptor dyes. Increasing the number of double bonds in the spacer chain has two beneficial effects on the emission wavelength: i) the larger number of double bonds moves the emission wavelength to the infrared region as evidenced by the values in gas phase and ii) it increases the charge transfer character of the transition which also moves the emission wavelength to the infrared region with increasing solvent polarity. The nature of the electronic transition responsible for the fluorescence emission will be discussed in the next section.

4.2 Charge transfer and electron delocalization

Due to their architecture, DANIR molecules undergo an intense charge transfer transition which involves the displacement of electron density

from the donor to the acceptor portion of the molecule. This transition corresponds to the first excited singlet state of the three molecules and has large values of the oscillator strength (Table 4.2), increasing with the increasing number of conjugated double bonds of the π bridge. Figure 4.3 shows the main orbitals involved in the transition to the first excited state. In all three molecules, this transition has a strong HOMO–LUMO character. Noteworthy, the HOMO has a larger contribution from the orbitals of the dimethylaniline moiety and the LUMO is more centered in the conjugated double bond chain and cyanide groups, as expected for these push–pull complexes and already observed for bithiophene derivatives.

The charge transfer character of the first excitation of DANIR markers has been analyzed by means of the index D_{CT},[165,166] which provides a simple measure of the spatial extension of the electronic transfer across the molecules, as already mentioned in the previous Chapter. This index is defined as the distance between the barycenters of charges associated with the zones of increase and depletion of the electronic density upon excitation,[165,166] therefore the larger the value of the index, the more separated the barycenters of charge and thus, the more pronounced charge transfer character of the transition.Values of D_{CT} computed for compounds 2a, 2b and 2c at the S_0 and S_1 optimized geometries are reported in Table 4.3. Figure 4.3 shows a representation of the density difference $\Delta\rho(\boldsymbol{r})$ and the barycenters of charge depletion and increase.

Results indicate that, at the S_0 optimized geometry, the value of the D_{CT} index increases steadily with the number of double bonds, confirming that the spacer composed by the conjugated double bonds is effective in red-shifting the energy of the transition not only by increasing the num-

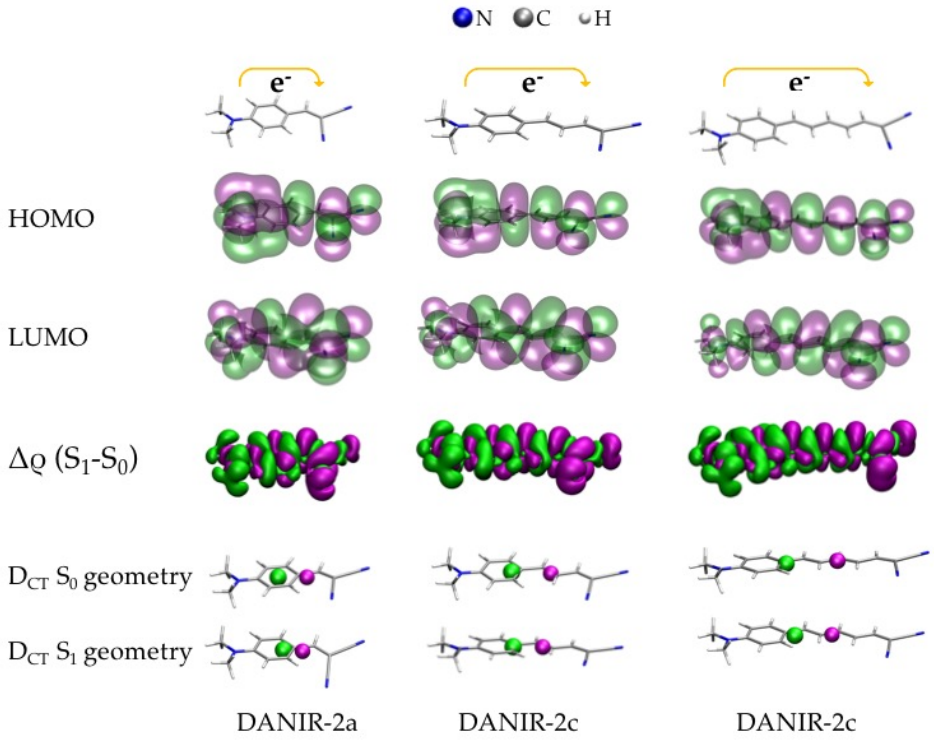


Figure 4.3: HOMO and LUMO orbitals at the S_0 optimized geometry, density difference between S_1 and S_0 at the S_0 optimized geometry, and barycenters of charge depletion (green) and increase (magenta) at the S_0 and S_1 optimized geometry of DANIR-2a, -2b and -2c.

ber of double bonds but also by enlarging the charge transfer character of the transition. The representation of the barycenters of charge depletion (green) and increase (magenta) shows that electron density is transferred from the aromatic ring to the second closest double bond to the electron-withdrawing group. These results confirm that the spatial extent of the

Table 4.3: Dipole moment of S_0 and S_1 at the S_0 and S_1 optimized geometries, values of the aromaticity index HOMA and values of the D_{CT} index (Å) of DANIR-2a, -2b and -2c.

	2a	2b	2c
$\mu S_0/S_0$	11.0	12.4	13.8
$\mu S_1/S_0$	15.5	18.9	22.5
$\mu S_0/S_1$	11.9	13.6	15.6
$\mu S_0/S_1$	14.4	17.6	21.0
HOMA S_0 ring	0.89	0.91	0.92
HOMA S_1 ring	0.67	0.73	0.77
HOMA S_0 double bonds	-	0.72	0.67
HOMA S_1 double bonds	-	0.95	0.97
D_{CT}/S_0	2.0	2.7	3.5
D_{CT}/S_1	1.3	2.1	2.6

charge transfer is inferior to the distance between the donor and acceptor groups[166] and show that the π bridge is an active participant in the transition. On the other hand, the same Table 4.3 and Figure 4.3 indicate that geometry relaxation in S_1 results in a significant reduction of the spatial extent of the charge transfer, because both D_{CT} index and dipole moment of the excited state show a partial compensation of charge separation at the optimized S_1 geometry. This is the same behavior reported for markers of the bithiophene family.

Since, as already mentioned, geometry relaxation in S_1 affects significantly the bond lengths of the delocalized π system, the geometric HOMA index[176] for the six-membered ring, as well as for the conjugated double bond chain at the equilibrium S_0 and S_1 geometries, was computed. This index, which was devised for the study of aromatic systems, has also

been applied for the study of electron delocalizetion in acyclic conjugated molecules.[177] It is defined as follows:

$$HOMA = 1 - \frac{\alpha}{n}\sum_{i-1}^{n}(R_i - R_{ref})^2 \tag{4.1}$$

R_i and R_{ref} are the i-th C$-$C bond length and the C$-$C bond length reference value, respectively, n is the number of C$-$C bonds and α a normalization factor that guarantees that the index value vanishes for fully localized systems. Thus, values of HOMA are close to 1 for fully aromatic systems and close to 0 for for fully localized systems. A R_{ref} value of 1.388 Å and a value of α of 257.7 have been employed according to the literature.[176–178] HOMA index values are reported in Table 4.3, and indicate that at the ground state geometry, the six-membered ring presents a strong aromatic character (HOMA values between 0.89 and 0.92), while the spacer behaves essentially as a normal conjugated double bond chain, with the computed values lying between 0.67 and 0.72.[175] In contrast, relaxation of the S_1 excited state leads to a structure in which the six-membered ring loses part of its aromatic character (HOMA values decrease to around 0.67 – 0.77), while the conjugated bonds of the spacer show larger electron delocalization (HOMA between 0.95 and 0.97). Therefore, the loss of aromaticity in the benzene ring due to electronic excitation is compensated by the delocalization of the electron density in the double bonds of the spacer that stabilizes S_1. That is, at the S_1 optimized geometry the conjugated double bond chain exhibits larger electron delocalization than the six-membered ring.

Overall, the dipole moment, charge transfer transition and HOMA

indexes illustrate that the π bridge is an active participant in the transition and not just a passive spacer between the donor and acceptor groups. This implies that structural modifications of the π bridge, even by keeping the spatial separation between the donor and acceptor constant, may greatly affect the charge transfer transition and the final emission wavelength.

4.3 Deactivation pathways

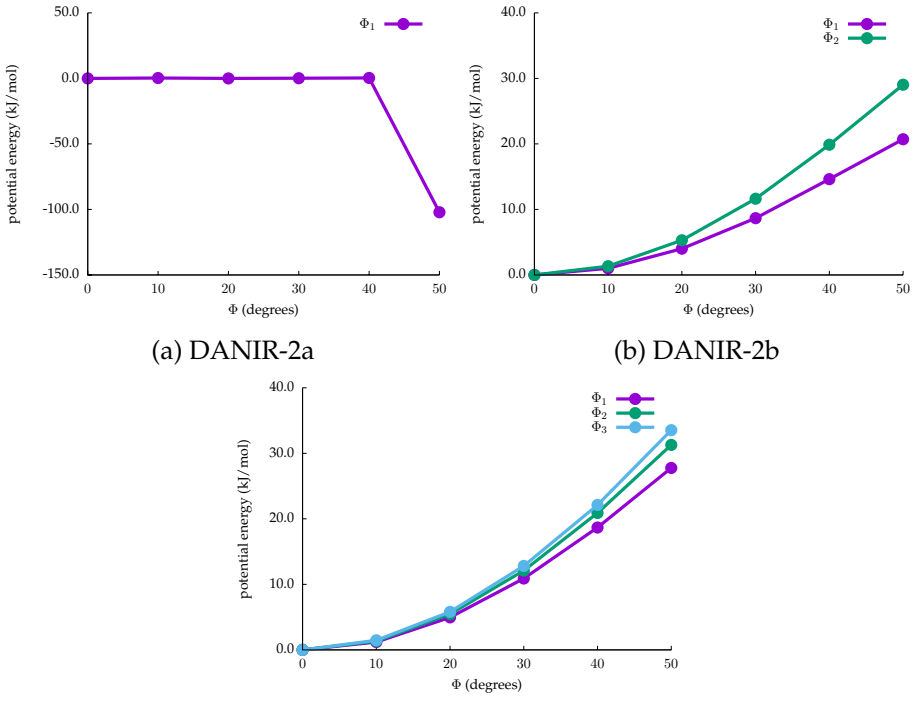
In the previous Chapter, it was shown that bithiophene markers can undergo internal conversion through a conical intersection by torsion around the double bond of the propanedinitrile group. Indeed, in systems containing C=C double bonds, the *trans* to *cis* or *cis* to *trans* isomerization in the S_1 excited state is related to a nonradiative relaxation process involving a conical intersection at a geometry where the double bond is twisted by about 90°.[168,169] The spacer between the donor and acceptor groups in DANIR probes contains a conjugated double bond chain that can undergo *trans* to *cis* isomerization in the S_1 state and thus offer a nonradiative decay pathway back to the ground state. Interestingly, compound 2a presents no fluorescence emission in solution in contrast to 2b and 2c. With the aim of evaluating the feasibility of deactivation through internal conversion, we have performed constrained optimizations along all double bonds Φ_1, Φ_2 and Φ_3 (Figure 4.4) In these scans, the torsional angles defining the rotation around the double bonds are kept fixed, while all other geometrical parameters are allowed to relax in S_1. This methodology was adopted before for the retinal dye.[169] Since TDDFT is not suitable for describing the conical intersection region due to its multireference nature, we confined

DANIR-2a DANIR-2b DANIR-2c

Figure 4.4: Ground state optimized geometries of DANIR-2a, DANIR-2b and DANIR-2c with indication of the torsional angles corresponding to the double bonds Φ_1, Φ_2 and Φ_3.

our exploration to the $0 < \Phi < 50°$ range. This allows a brute estimation of the energy barrier for achieving the conical intersection in the excited states of 2a, 2b and 2c and more importantly a qualitative estimation of the effect of enlarging the spacer. The reader should be aware that Φ_1, Φ_2 and Φ_3 correspond to rotation around double bonds, and are different from dihedrals α, β and γ, which correspond to rotation around single bonds.

Results of relaxed energy scans are reported in Figure 4.5. Rotation around the double bonds in S_1 produce initially a destabilization of this state for DANIR-2b and DANIR-2c (Figure 4.5c and 4.5c). However, in the case of 2a, when the torsional angle reaches a value around 50°, S_1 energy starts decreasing. The energy difference between the *trans* fully optimized geometry of S_1 and the highest point of the scan is very low (less than 1 $kJ\,mol^{-1}$), indicating that the conical intersection can be reached through an essentially barrierless process for 2a (Figure 4.5a). Extending the conjugated double bond chain produces a dramatic change in the feasibility of the C=C double bond rotations. For 2b and 2c, the energy of S_1 keeps



(a) DANIR-2a

(b) DANIR-2b

(c) DANIR-2c

Figure 4.5: S_0 and S_1 energies for the relaxed scans around double bonds Φ_1, Φ_2 and Φ_3 ($kJ\,mol^{-1}$) of DANIR-2a, -2b and -2c.

increasing until at least 50°. In particular, the torsion with the smallest barrier is always that arund Φ_1. This destabilization of S_1 is larger with the increasing number of double bonds, regardless of the considered double bond. This suggests that the nonradiative decay through internal conversion is much less efficient for 2b and 2c, and this correlates with the experimental values of the quantum yield of fluorescence of these molecules in aqueous solution (Table 4.2).

Details of the interaction of DANIR-2c with amyloid-β fibrils will be discussed in Chapter 5, but it is worth mentioning that the fluorescence enhancement observed when DANIR-2c binds to amyloid deposits compared to aqueous solution may originate from steric hindrance. Interactions of the marker with the bulky side chains which define the hydrophobic channels of amyloid fibrils limits the flexibility of the π system of DANIR-2c, and, consequently, also deactivation through internal conversion, which requires highly torsioned geometries. Reducing the efficiency of this nonradiative deactivation favors the competing fluorescence emission, and explains the 12-fold fluorescence enhancement observed in presence of amyloid aggregates.It must be pointed out that this fluorescence enhancement is significantly smaller than that of NIAD-4 (400-fold), which is due to disaggregation.

4.4 Final remarks

The photophysical properties of a class of donor-π-acceptor conjugated amyloid markers, DANIRs, has been studied with TDDFT. The absorption spectra of DANIR-2a, DANIR-2b and DANIR-2c in dichloromethane have been calculated with *ab initio* molecular dynamics simulations. The agreement between calculated and experimental absorption maxima is particularly good for DANIR-2c, which is the most interesting molecule of this family for *in vivo* applications due to its emission wavelength in the near infrared region.

Due to their donor-acceptor structure (donor: dimethylamino groups; acceptor: propanedinitrile group), all three markers undergo a charge trans-

fer transition, resulting in an emitting excited state S_1 with a high dipole moment. Values of the D_{CT} index evaluating the spatial extent of the charge transfer grow in the order DANIR-2a $<$ DANIR-2b $<$ DANIR-2c. The position of the barycenters of charge depletion and increase indicate that the π system composed of a benzene ring and a conjugated double bonds chain is an active participant of the transition rather than as a mere spectator. Moreover, in all cases, relaxation in the excited state S_1 involves a reorganization of electron density over the conjugated π system. The effect of this reorganization can be appreciated in the reduction of the value of D_{CT} index at the S_1 optimized geometry compared to the S_0 optimized geometry. Furthermore, this reorganization is evident in the values of the aromaticity index HOMA at the S_0 and S_1 equilibrium geometry. Geometry relaxation in the excited state, indeed, results in an increase of electron delocalization of the conjugated double bonds chain and, conversely in a reduced aromaticity of the aromatic ring.

Nonradiative deactivation through internal conversion involving a conical intersection between S_1 and S_0 has been explored. As already reported for other conjugated molecules, *trans* to *cis* or *cis* to *trans* isomerization involves a conical intersection for torsions of 90° around the double bonds of the π system. The energy associated to these torsion has been explored by performing relaxed rotational scans around all double bonds of the conjugated chain. Results show that the torsion around the double bond of DANIR-2a in S_1 is essentially barrierless, indicating that for this molecule internal conversion will be the dominant deactivation pathway, while rotational barriers increase with the increasing number of conjuated double bonds of the π system, showing that internal conversion will be less efficient with growing system size. These results explain the experimen-

tal values of quantum yield of fluorescence of DANIR-2a, DANIR-2b and DANIR-2c in water, which is 0 for DANIR-2a and increases with system size: for the smallest marker, only nonradiative deactivation takes place, while for larger molecules, where internal conversion is less favorable, increasing fluorescence emission is observed.

Interaction with amyloid fibers enhances fluorescence emission due to sterical hindrance: when these markers are accommodated into the hydrophobic channels located at the core of amyloid fibers, their flexibility is reduced by contact with the regular arranged side chains of amyloid-β peptides. This has the effect of making internal conversion less favorable, and enhances the competing radiative deactivation through fluorescence emission.

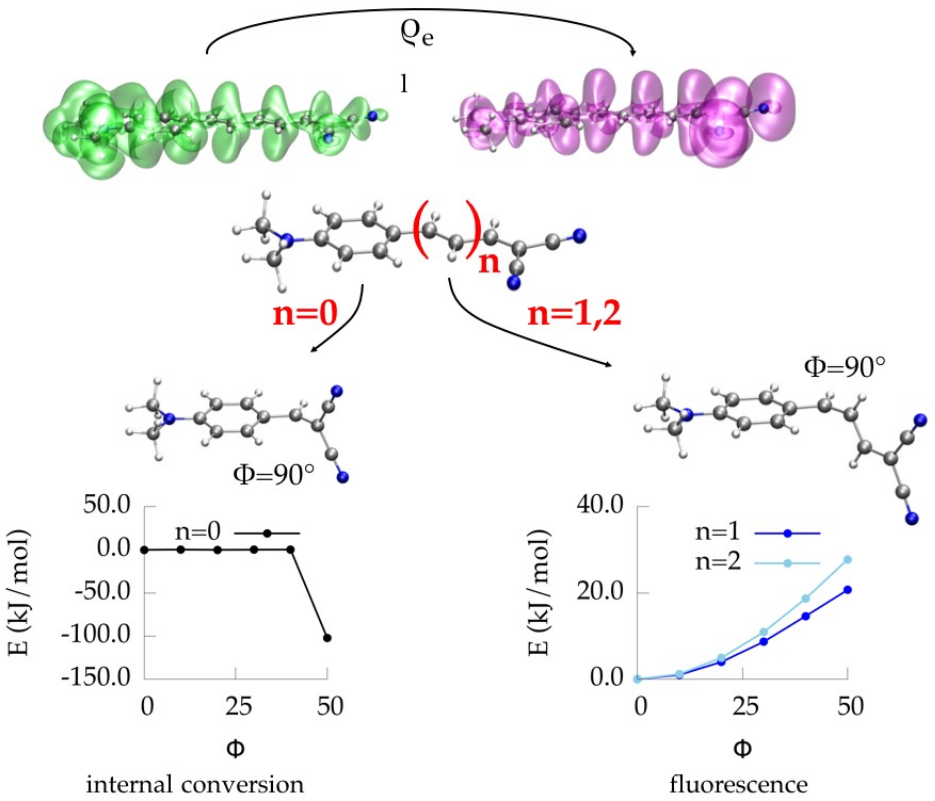


Figure 4.6: Graphical summary

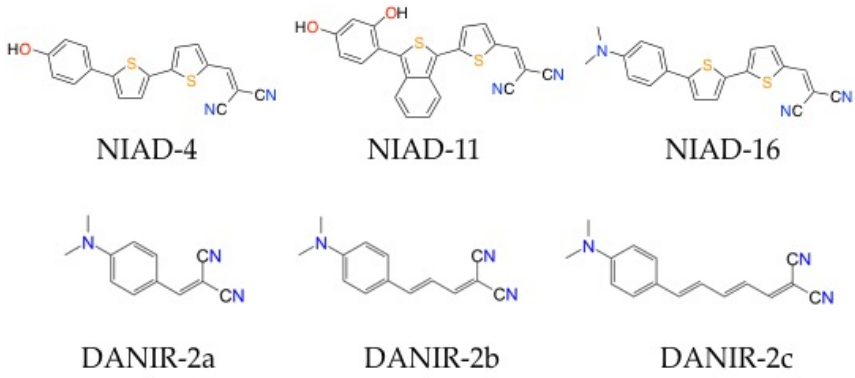
Chapter 5

Binding to amyloid-β fibrils

This Chapter covers the binding of the two classes of fluorescent markers presented in Chapters 3 and 4, NIADs and DANIRs, to amyloid-β fibrils (Scheme 5.1).

As already mentioned in the Introduction, in the last years several models of amyloid-β fibrils have been proposed thanks to advances in solid state NMR spectroscopy and scanning transmission electron microscopy.[16,17,29,31] In this study, we focused on Aβ40 fibrils rather than on Aβ42 ones. This choice was determined by the superior availability of Aβ40 fibrils at the time (2014).

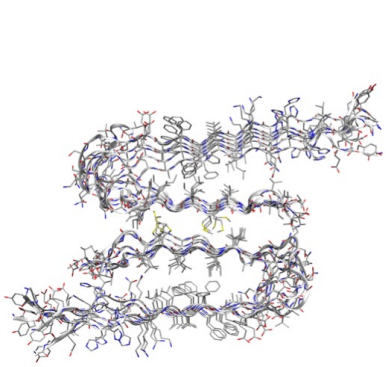
In particular, the three models of Aβ40 fibrils reported in Figure 5.1 (I, II and III) were taken into account for the study of the binding properties of fluorescent markers.[16,29,31]



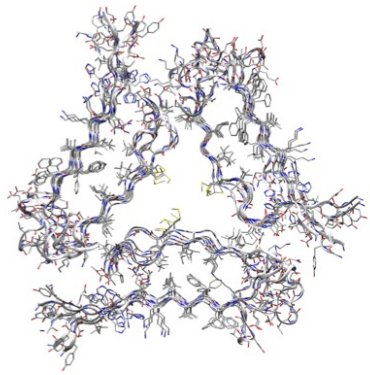
Scheme 5.1: Molecular structures of NIAD and DANIR markers for amyloid detection.

As mentioned, amyloid-β fibrils present a marked conformational diversity. It has been shown that identical polypeptides can fold in multiple conformations, and, remarkably, that the most abundant amyloid architecture may change from patient to patient.[31,32] This variety of structures is due to the fact that aggregation is strongly affected by the environmental conditions, and this complicates the design of structure specific markers. However, all architectures present some common features: i) They are composed of protofilaments, the basic units of these structures, two in the case of I and three in the case of II and III (Figure 5.1); ii) these protofilaments have a cross-β architecture, formed by two parallel β-sheets separated by a loop (Figures 5.1a, 5.1b and 5.1c).

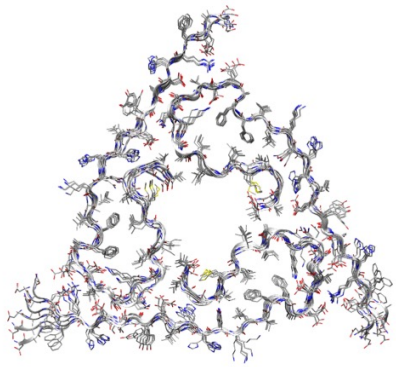
Regarding the nature of the aminoacidic residues, the loops and terminals are composed of charged and polar residues, while the parallel β-sheets contain hydrophobic residues and glycines. Due to this architecture, the regular arrangement of the hydrophobic side chains of the β-



(a) I, PDB code: 2LMN; residues 9-40.



(b) II, PDB code: 2LMQ; residues 9-40.



(c) III, PDB code: 2M4J; residues 1-40.

Figure 5.1: Amyloid-β fibril models that have been considered.

sheets leads to the formation of channels, that run parallel to the fibril axis and can accommodate molecules with an aromatic/conjugated rod-like structure, which interact through π-stacking and dispersion . For this

reason, the structure of amyloid-β fibrils is crucial for determining their binding ability. Indeed, a different conformation of the protofilament, or junctions among protofilaments, modify the relative distances and orientations of the side chains that are responsible for the interaction with the marker, which in turn can severely affect the mutual affinity of the two species.

5.1 Binding site search

Due to their aromatic/conjugated rod-like structure, NIAD and DANIR markers are expected to interact preferentially with hydrophobic residues. In order to identify the binding poses, PELE simulations were performed for NIAD-4, NIAD-11, NIAD-16 and DANIR-2c with each of the three fibril models.[154] For each marker-fibril pair, at least 90 independent trajectories were run, in order to guarantee the full exploration of the binding possibilities.

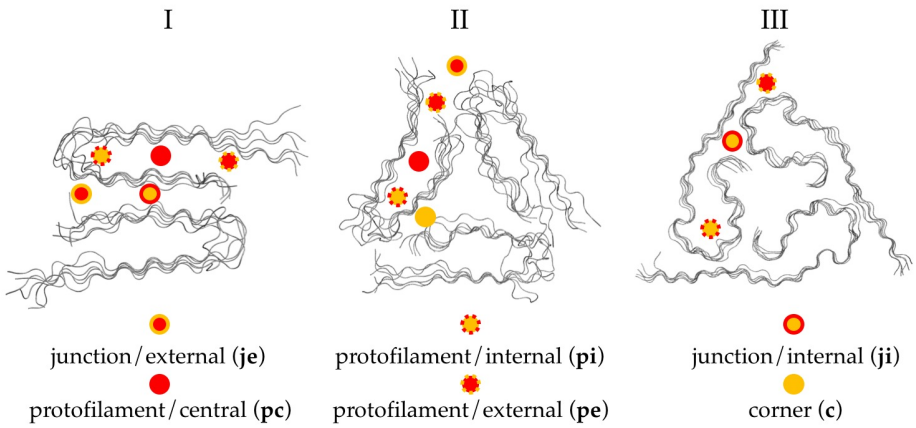
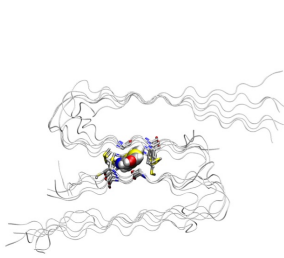
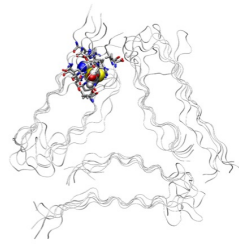


Figure 5.2: Classification of the binding poses for the fibril/marker complexes.

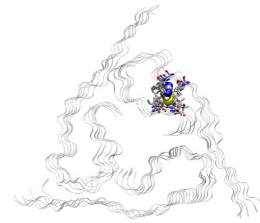
The binding poses will be classified based on the position occupied by the ligand, as shown in Figure 5.2. Markers can be accommodated either at the junction between two protofilaments, at the core of the fibril (**ji**) or in an external position (**je**). Alternatively, they can be accommodated between the β sheets of a single protofilament, close to the loop (**pi**), in the center of the double β sheet (**pc**) or close to the open end of the cross-β motif (**pe**). Depending on the three-dimensional structure of the fibrils, these positions involve interactions with different aminoacidic resides. In addition, fibril II can accommodate ligands also in the corner of the internal cavity defined by two adjacent protofilaments (**c**). A representative binding pose for each marker-fibril pair is reported in Figure 5.3, while a detailed representation of the binding interaction for all cases is reported in Figure 5.4. More binding poses are reported in Figures D.1, D.2, D.3 and D.4.



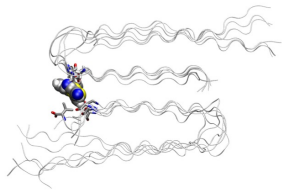
(a) NIAD-4/I-ji.



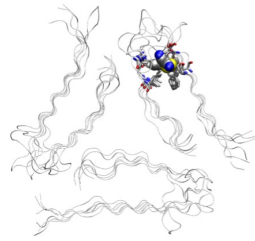
(b) NIAD-4/II-pe.



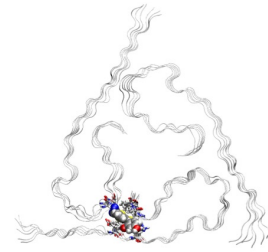
(c) NIAD-4/III-pi.



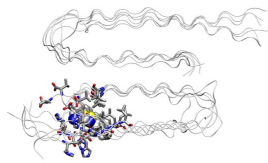
(d) NIAD-11/I-je.



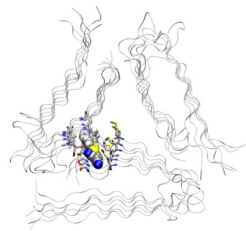
(e) NIAD-11/II-pi.



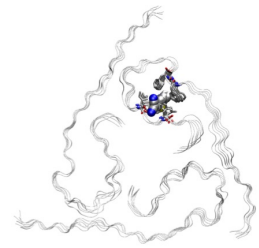
(f) NIAD-11/III-ji.



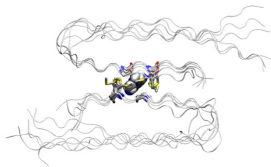
(g) NIAD-16/I-pe.



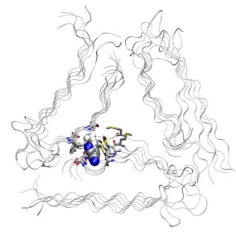
(h) NIAD-16/II-c.



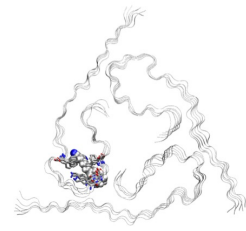
(i) NIAD-16/III-pi.



(j) DANIR-2c/I-ji.

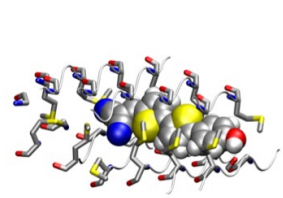


(k) DANIR-2c/II-c.

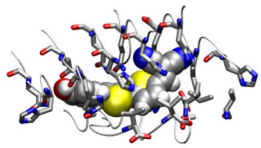


(l) DANIR-2c/III-pi.

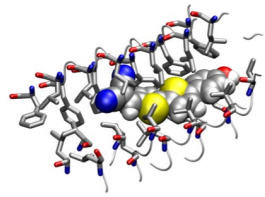
Figure 5.3: Representative binding poses of NIAD-4, NIAD-11, NIAD-16 and DANIR-2c on three models of amyloid fibrils, I, II and III.



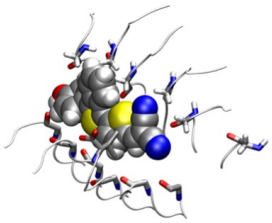
(a) NIAD-4/I-ji.



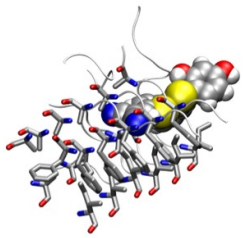
(b) NIAD-4/II-pe.



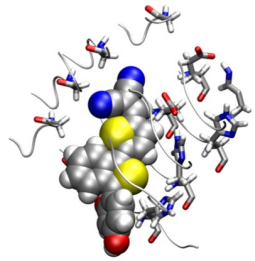
(c) NIAD-4/III-pi.



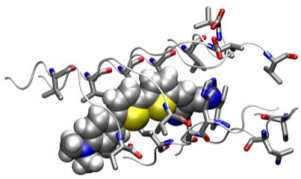
(d) NIAD-11/I-je.



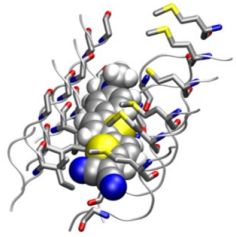
(e) NIAD-11/II-pi.



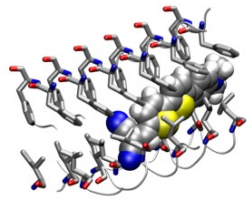
(f) NIAD-11/III-jc.



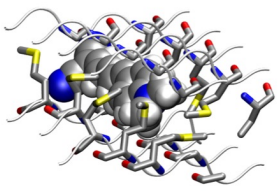
(g) NIAD-16/I-pe.



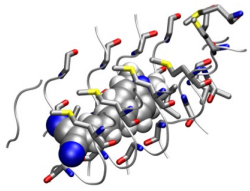
(h) NIAD-16/II-c.



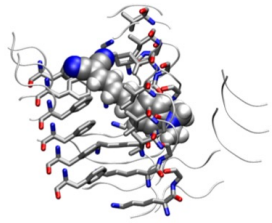
(i) NIAD-16/III-pi.



(j) DANIR-2c/I-ji.



(k) DANIR-2c/II-c.



(l) DANIR-2c/III-pi.

Figure 5.4: Representative binding poses of NIAD-4, NIAD-11, NIAD-16 and DANIR-2c on three models of amyloid fibrils, I, II and III.

Table 5.1 presents the PELE binding energies for each pose, along with the residues involved in the interaction. For NIAD-4, which was the first marker to be studied, a larger number of binding site search trajectories has been run, resulting in the identification of a larger number of structures. This allows a comparison of the relative stabilities of the binding poses involving the same fibril/marker pair.

The binding energies provided by PELE are qualitative, but indicate clearly that binding pose NIAD-4/I-jc (Figure 5.3a) is significantly more stable than the others. This preference has been partially confirmed by performing *ab initio* calculations. For selected binding poses, a cluster has been cut and saturated in a radius of 6 Å around the marker. Afterwards, the aminoacidic side chains and the marker itself have been relaxed by performing a geometry optimization keeping the backbone of the cluster fixed. After optimization, an estimation of the interaction energy has been computed by performing single point PBE-D2 calculations on both the full cluster and the fibril and marker components separately. In general, the computed DFT energy differences follow the same trend as PELE energies, predicting that NIAD-4/I-ji is preferred over NIAD-4/I-pc by ~ 20 $kcal\,mol^{-1}$. A similar difference is observed for NIAD-4/I-pi and NIAD-4/I-pe2, but not for NIAD-4/I-pe. This is due to the fact that our single point DFT calculations do not account for solvation effects, and binding pose NIAD-4/I-pe involves the interaction of NIAD-4 with the C termini of the peptides in an open, flexible area exposed to the solvent. In this condition, our DFT description naturally fails.

Concerning fibril model II, again DFT calculations confirm the PELE energy order, with NIAD-4/II-pe more favorable than NIAD-4/II-c by $\sim$

17 $kcal\,mol^{-1}$. This order of preference for the different binding sites, however, is not constant across the different markers, as it will be explained below.

Comparing the binding properties of different markers, a first consideration that may be drawn is that, despite a very similar structure, bithiophene derived markers NIAD-4, NIAD-11 and NIAD-16 present quite different binding properties. Comparing the most stable binding poses of the three markers of fibril I (Figures 5.3a, 5.3d, 5.3g), indeed, it is evident that only NIAD-4 is accommodated at the core of the fibril, at the juncture of the two protofilaments. In contrast, NIAD-11 and NIAD-16 are located at the exterior of the fibril, NIAD-11 at the juncture of two protofilaments as well, and NIAD-16 at the open end of one cross-β unit.

Due to the architecture of amyloid fibrils, these different binding poses result in markedly different aminoacidic residues involved in the interaction. At the hydrophobic core of fibril I, NIAD-4 occupies a cavity resulting from the presence of glycine residues, and interacts with regular rows of methionines (Figure 5.4a, Table 5.1). A further stabilizing interaction is provided by an hydrogen bond between the hydroxyl group of NIAD-4 and a carbonyl of the backbone (d_{HB} = 2.0 Å). NIAD-11, on the other hand, interacts with the glycines of the loop and terminal valine residues (Figure 5.4d, Table 5.1). In this position, stabilization results essentially from two hydrogen bonds that are formed between the hydroxyl groups of the phenyl ring and carbonyl groups of the backbone ($d_{HB-1,2}$ = 1.8 Å). Overall, the marker remains exposed to the solvent, which, due to the hydrophobic character deriving from the additional benzene ring, results in a binding of NIAD-11 that is expected to be weaker than that of NIAD-4.

Table 5.1: Binding poses of NIAD-4, NIAD-11, NIAD-16 and DANIR-2c on three models of amyloid fibrils, I, II and III: PELE binding energy ΔE_{bind} ($kcal\ mol^{-1}$) and aminoacidic residues involved in the interaction.

binding pose	ΔE_{bind}	residues
NIAD-4/I-ji	-80.6	GLY33, MET35
NIAD-4/I-je	-66.7	GLY29, GLY37, VAL39, VAL40
NIAD-4/I-pi	-57.9	PHE19, ASP23, ALA30, LEU32
NIAD-4/I-pc	-53.9	PHE19, ALA21, VAL34, VAL36
NIAD-4/I-pe	-73.8	GLY9, VAL12, HIS13, VAL40
NIAD-4/I-pe2	-54.9	GLU16, LEU17, VAL37, VAL39
NIAD-4/II-pe	-83.1	HIS13, GLN15, VAL36, VAL40
NIAD-4/II-c	-60.5	GLY33, MET35
NIAD-4/II-pi	-65.2	PHE19, ALA21, ILE32, ALA33
NIAD-4/II-pc	-70.5	GLN15, LEU17, PHE19, LEU34, VAl36, VAL39
NIAD-4/III-pi	-60.4	LEU17, PHE19, LEU34, VAL36
NIAD-4/III-ji	-58.3	GLU11, HIS13, GLY29, GLY38, VAL39
NIAD-4/III-pi2	-53.5	PHE20, LYS28, ISLE 31, LEU34
NIAD-4/III-je	-58.9	GLU3, ARG5, GLU22, VAL24
NIAD-11/I-je	-53.2	GLY29, GLY37, VAL39, VAL40
NIAD-11/I-pe	-38.1	GLU11, VAL39, VAL40
NIAD-11/II-pi	-48.3	PHE19, ALA21, ALA30, ILE32
NIAD-11/II-c	-38.3	GLY33, MET35
NIAD-11/III-jc	-45.7	GLU11, HIS13, GLY38, VAL39
NIAD-11/III-pi	-43.0	PHE19, LYS28, ILE31, LEU34
NIAD-16/I-pe	-58.8	HIS14, LEU17, VAL36, VAL40
NIAD-16/I-je	-56.8	GLY29, GLY37, VAL39, VAL40
NIAD-16/II-c	-56.8	ILE31, GLY33, MET35, GLY38
NIAD-16/II-je	-46.5	GLY9, TYR10, ASP23, VAL24.
NIAD-16/III-pi	-44.9	PHE19, PHE20, ILE31, LEU34
DANIR-2c/I-ji	-60.4	GLY33, MET35
DANIR-2c/I-je	-46.3	GLU11, VAL39, VAL40
DANIR-2c/II-c	-53.2	ILE31, GLY33, MET35, GLY38
DANIR-2c/II-pi	-51.6	PHE19, ALA21, ALA30, ILE32
DANIR-2c/III-pi	-51.9	PHE20, LYS28, ILE31, GLY29
DANIR-2c/III-pi2	-42.6	LEU17, PHE19, LEU34, VAL36

PELE binding energies, however qualitative, confirm this assumption. A more accurate comparison among the binding energies of the markers will be discussed in the following section.

Finally, NIAD-16 is in a position that is intermediate between those of NIAD-4 and NIAD-11: it is accommodated between two parallel β-sheets, like NIAD-4, and interacts with histidine (π-stacking) and leucine residues (dispersion forces). However, being located close to the open end of the cross-β unit, NIAD-16 has a less structured surrounding than NIAD-4. Since terminals are more flexible than the body of the fiber, residues are more disordered, which can be expected to result in a less pronounced stabilization from dispersion interactions. Moreover, being closer to the exterior than NIAD-4, NIAD-16 is more likely to interact with the solvent, water. This is particularly true for the bulky head of the marker, constituted by the N,N-dimethylamino group, which remains outside the fibril, resulting in an interaction which is not expected to be favorable due to the superior hydrophobicity of NIAD-16 over NIAD-4.

A question that may arise is why do NIAD-4 and NIAD-16 have different binding sites on fibril I (5.3). NIAD-11, indeed, is quite different from the other two markers, with the additional hydroxyl group and benzene ring that disrupt the linearity of the structure and offer a pronounced sterical hindrance ($V_{NIAD-11}$ = 329 Å^3) that may prevent entering the ordered channels located at the center of the parallel β-sheets. NIAD-4 and NIAD-16, on the other hand, differ only by the donor group, $-$OH for NIAD-4 and $-\mathrm{N(CH_3)_2}$ for NIAD-16. What seems a small difference has in fact a significant effect on the binding properties of the two molecules, preventing NIAD-16 from accessing the hydrophobic channels that are ac-

cessible to NIAD-4.

This can be explained with two obervations: i) the N,N-dimethylamino group has a markedly superior sterical hindrance compared to the hydroxyl (V_{NIAD-4} = 277 Å^3, $V_{NIAD-16}$ = 315 Å^3),[179] and ii) while the hydroxyl group can provide stabilization by means of the hydrogen bond with a carbonyl of the backbone, the N,N-dimethylamino group does not offer this possibility. Despite being a small difference compared to the overall structure of the marker, this variation is enough to lead the binding process.

Another interesting point about NIAD-11 is that both its absorption and emission wavelengths are red-shifted upon binding to amyloid-β fibrils, by 30 and 15 nm respectively. In both cases, the red shifting may be attributed to a partial planarization, which reduces the HOMO-LUMO gap.

DANIR-2c is accommodated in the same hydrophobic channel as NIAD-4, *i.e.* at the junction of the two protofilaments, between two parallel β-sheets, and is involved in dispersion interactions with methionine residues. DANIR-2c has the same N,N-dimethylamino donor and malononitrile acceptor group as NIAD-16, but can fit at the core of the fibril due to the lesser sterical hindrance offered by the conjugate double bond chain compared to the bithiophene unit ($V_{DANIR-2c}$ = 246 Å^3). While no stabilization by hydrogen bonding has been observed in this case, the linear geometry of the molecule achieves a good stabilization by maximizing the number of residues involved in the binding. Even if only DANIR-2c is of interest for *in vivo* application, due to its fluorescence emission in the near infrared region, a binding site search for DANIR-2a on fibril model I has

been performed, which led to the identification of an analogous binding pose, with DANIR-2a between the hydrophobic β-sheets at the junction of the protofilaments. This seems to indicate that molecules with the same structure, which differ only by the length of the conjugated bridge, have similar biding properties.

Regarding fibril model II, again marked differences are observed in the binding poses of the three bithiophene markers (Figure 5.3). Again, as in the case of fibril I, NIAD-4 seems the one with the most favorable interaction. It is accommodated between two parallel β-sheets, though not too far from the terminals, and interacts with hydrophobic valine and histidine residues, even if also glutamates are present in the active site. Again, further stabilization is provided by an hydrogen bond between the hydroxyl group of NIAD-4 and a carbonyl of the backbone (d_{HB} = 1.9 Å). Also NIAD-11 is accommodated between two parallel β-sheets, close to the loop of one of the protofilaments, and interacts closely with hydrophobic residues of alanine, phenylalanine and isoleucine. However, only a portion of the molecule is buried into the hydrophobic cross-β unit, with the substitued thiophene and donor groups exposed to the solvent. This peculiar behavior may again depend on both i) the deviation from linearity and increased sterical hindrance resulting from the fused benzene ring, which worsens the interaction with the linear amyloid channels and ii) the additional hydroxyl group, which locally increases molecular solubility and favors interaction with the solvent.

As far as NIAD-16 and DANIR-2c are concerned (Figure 5.3h and 5.3k), the marker interacts with two filaments rather than one, and is accomodated in the corner of the triangle defined by the 3-fold symmetry of

the fibril. Also in this case, the binding pocket defined by glycine residues has a channel-like shape, and the markers are protected from exposure to the solvent by the long chains of methionine residues (Figure 5.4k and 5.4h).

Quite surprisingly, binding to the second model of three-fold symmetry amyloid-β fibril, III, presents marked differences from that of model II. This is due to the inherent structure of the two fibrils (Figures 5.1b and 5.1c): while in the case of model II, the cross-β architecture leads to the formation of parallel β-sheets, in the case of model III, protofilaments are arranged so that the internal β-sheet forms a turn. This turn leaves a variable distance between the parallel β units, and a further turn in the terminal residues of each protofilament, which interacts with the loop of another. The key feature of this fibril, as far as binding of small conjugated ligands is concerned, is the variable distance between the β-sheet of each protofilaments, which leads to a multiplicity of hydrophobic channel-like binding pockets with a variety of sizes.

NIAD-4 is buried in a completely apolar pocket defined within a single protofilament by leucine, phenylalanine and valine residues, and is completely shielded from the solvent (Figure 5.3c and 5.4c). NIAD-11, on the other hand, binds at the junction of two protofilaments, and interacts with three β-sheets (Figure 5.3f) via glutamate, histidine, glycine and valine residues. Similarly to what already reported for model II, only a portion of the molecule is buried into the fibril: the phenyl ring bearing the hydroxyl groups is located on the outer part of the fibril, and is exposed to the solvent (Figure 5.4f). Again, the two additional groups of NIAD-11 compared to the parent NIAD-4, the hydroxyl group and fused benzene ring,

result in a deviation from linearity and an increased hydrophilicity of the head group, which concur in favoring interaction with the solvent, with an overall reduction of the binding affinity towards amyloid-β species.

The most stable binding pose of NIAD-16 on fibril models I, II and III allows us to appreciate the aforementioned effect of structure morphology of the fibril on binding properties. Figures 5.3i and 5.4i illustrate that NIAD-16 is completely buried in an apolar channel defined by phenylalanine, isoleucine and leucine belonging to the same protofilament. This is in contrast to what reported for fibril models I and II, where the marker was always accommodated in a more external position, either close to the open side of a cross-β unit (Figure 5.3g), or at the internal surface of the fibril (Figure 5.3h). This difference may be attributed to the different size of the binding pockets of the three fibril models: while in models I and II the β-sheets belonging to each profilament are fully parallel, resulting in smaller binding channels, the turn described by the internal β-sheet in model III yields larger binding pockets, which can harbor the bithiophene body and N,N-dimethylamino terminal group of the marker.

Again, as already reported for model II, binding of DANIR-2c to fibril III is rather similar to that of NIAD-16; DANIR-2c, indeed, is held in an hydrophobic channel defined by phenylalanine, lysine, isoleucine and glycine residues, located close to the loop of a protofilament. Apart from these structural considerations, there are several other aspects that may affect the order of preference for the available binding pockets of different markers. For example, it must be taken into account that the interaction with the marker may alter the structure of the fibril, and disrupt existing stabilizing interactions among side chains. This results in a energy toll that

needs to be paid for binding the marker. For sure, a deep understanding of the binding properties of the markers requires a deeper analysis than that proposed in this thesis.

Overall, however, these results indicate that binding of aromatic/ conjugated rod-like markers to the three models of amyloid fibrils that have been considered is driven by geometric reasons rather than affinity towards a specific aminoacidic sequence. Indeed, while the binding poses provided by the PELE program do not offer quantitative results in terms of binding energy, still they unequivocally indicate that these markers occupy the free hydrophobic channels which run parallel to the fibril axis are delimited by a regular arrangement of non-polar side chains, the dominant interaction being van der Waals/π-stacking type. This is particularly evident in the case of model III, where NIAD-4, NIAD-16 and DANIR-2c occupy three adjacent, essentially equivalent, hydrophobic channels composed of phenylalanine, leucine and valine residues. Moreover, the binding of NIAD-11 shows how deviation from the linear architecture results in a significant reduction of the fibril-marker interaction.

A further confirmation of this geometrical preference is given by the comparison of markers NIAD-16 and DANIR-2c, the only difference between the two being the nature of a portion of the π bridge, which is a bithiophene in the case of NIAD-16 and a conjugated double bond chain in the case of DANIR-2c. Rather than affecting the nature of the interactions that can be established with the fibrils, this change mainly affects the sterical hindrance of the marker, which is crucial in determining the binding pose: while the less sterical hindered DANIR-2c can be harbored at the hydrophobic core of the smallest fibril I (Figure 5.3j), NIAD-16 is re-

jected by this site, and accommodated in the less ordered pocket reported in Figure 5.3g. An estimation of the dimensions of the binding channels of amyloid fibrils reported in Figure 5.3 are presented in Table D.1, and reflect this sterical effect. Indeed, DANIR-2c occupies smaller channels than the more bulky bithiophene markers, which in other words means that the hyrophobic side chains which define these channels can arrange more comfortably around the less sterically hindered DANIR-2c than around bithiophene markers.

5.2 MM/PBSA and MM/GBSA calculations

MM/PBSA and MM/GBSA calculations were performed on the most favorable PELE binding poses of fluorescent markers on the three models of amyloid-β fibrils.[138] Molecular dynamics simulations were carried out with the package Amber14 and the ff99SB protein force field.[142] Ligands were parametrized with the GAFF force field.[180] After 2000 minimization steps, a 50 ps NVT dynamics was performed rising the temperature from 0 to 50° K with a restraint on the backbone of the protein, to avoid strong deformation. A second relaxation was performed in the NPT ensemble, of 1000 ps, with the temperature raising from 50 to 310 K in the first half of this simulation, and kept constant at 310 K for the second half of the simulation. Production was run as 5 independent NVT 300 ps trajectories, and 20 geometry were evenly sampled from the last 200 ps of each trajectory. The resulting 100 structures were used for the free energy calculation. It has been shown that the independent trajectories approach provides results that are more converged than those obtained from a single longer

simulation.[181]

Moreover, we are interested in running relatively short molecular dynamics simulation due to the inherent fragility of the amyloid fibril models that are employed. Considering that the normal length of amyloid fibrils is in the range of $10 \div 10000\ nm$,[182,183] and that model I, II and III are only 6 filaments long ($25 \div 40$ Å), they are expected to present a markedly inferior rigidity and stability compared to the real system. For this reason, a short simulation helps preserving structural features. This can be appreciated considering the evolution of the RMSD of the protein backbone along three kinds of simulations of binding pose NIAD-4/I-ji: i) 1 ns from 5 independent trajectories, as described in the previous paragraph; ii) 1 ns from a single trajectory with the same simulation parameters; iii) a 10 ns simulation from a single trajectory.

In all three cases RMSD was computed on 100 equally sampled frames from each trajectory, and is reported in Figure 5.5. RMSD rises quickly with simulation time, indicating that the fibril model is rapidly altered already at 310 K. A representation of the structure modifications undergone by the fibril along the 10 ns simulation is reported in Figure D.5. With the 5 independent simulations, RMSD indicates that structural changes are relatively small.

PELE, MM/GBSA and MM/PBSA binding energies for several marker-fibril complex are reported in Table 5.2 and D.2. A first general observation is that MM/PBSA predicts lower binding energies than PELE and MM/GBSA. The striking difference between MM/PBSA and MM/GBSA values is due to the description of the interaction with the solvent. Indeed, in the MM/PBSA calculations that were performed, the solute-solvent in-

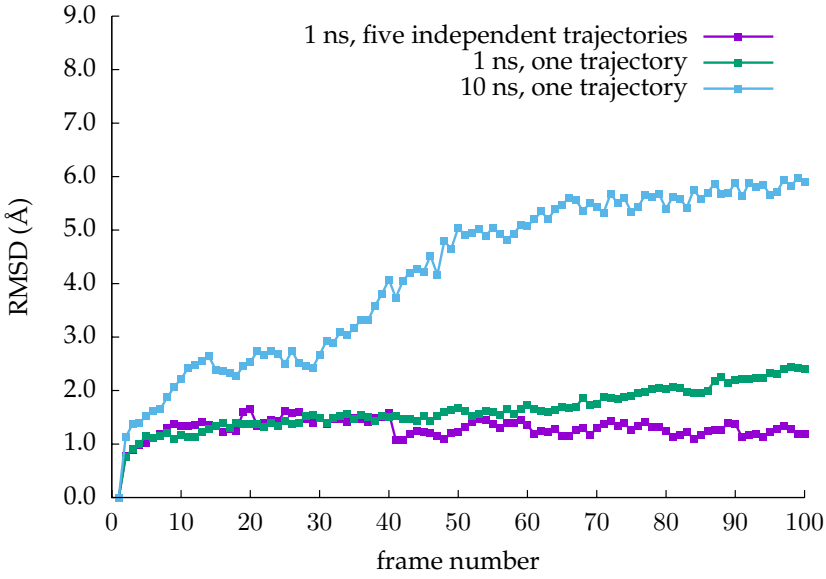


Figure 5.5: RMSD of the protein backbone along several molecular dynamics simulations performed on binding pose I-1: i) 1 ns simulation resulting from 5 independent trajectories; ii) 1 ns simulation from a single trajectory, and iii) 10 ns simulation from a single trajectory.

teraction was modeled as the sum of two terms: a cavity term, linearly proportional to the solvent accessible surface area, and a dispersion term.[132] In the MM/GBSA calculations, on the other hand, the nonpolar contribution to the solvation free energy was modeled as a single term linearly proportional to the solvent accessible surface area.[132] Apparently, this can affect greatly the absolute values of binding free energy. This has been confirmed by repeating the same MM/PBSA calculations with a simpler model of solute-solvent interaction, similar to that employed in the MM/GBSA calculations. These MM/PBSA values, which are reported in

Table 5.2: Binding energies ($kcal\,mol^{-1}$) of selected binding poses of NIAD-4, NIAD-11, NIAD-16 and DANIR-2c on fibril models I, II, III (Figure 5.3).

Binding pose	PELE	MM/GBSA	MM/PBSA
NIAD-4/I-ji	-80.6	-59.3	-13.9
NIAD-4/II-pe	-83.0	-41.9	0.4
NIAD-4/III-pi	-60.4	-51.0	-15.3
NIAD-11/I-je	-53.2	-45.4	-9.7
NIAD-11/II-pi	-48.3	-38.0	-2.0
NIAD-11/III-ji	-45.7	-40.9	10.0
NIAD-16/I-pe	-58.8	-44.5	-4.9
NIAD-16/II-c	-56.8	-43.9	-3.9
NIAD-16/III-pi	-44.9	-52.9	-12.3
DANIR-2c/I-ji	-60.4	-63.3	-14.9
DANIR-2c/II-c	-53.2	-45.7	-9.4
DANIR-2c/III-pi	-51.9	-57.7	12.9

Table D.3, are in good agreement with MM/GBSA values.

The literature provides an experimental value for the dissociation constant K_d of the DANIR-2c Aβ42 complex.[58] Assuming that the binding interaction is similar in Aβ42 and Aβ40 deposits, it is possible to estimate the binding free energy from the equilibrium constant for the dissociation as follows.

$$K_{eq} = \exp -\frac{\Delta G}{RT} \tag{5.1}$$

With an RT value of 0.593 of $kcal\,mol^{-1}$ at room temperature, the experimental dissociation constant of 26.9 nM yields a ΔG of -10.3, which is in satisfactory agreement with the MM/PBSA value of -14.9 The MM/PBSA

values reported in Table 5.2 have been chosen as reference due to their agreement with this value. Despite the fact that PELE and MM/GBSA severely overestimate the strength of the binding interaction, the three methods agree in attributing the most favorable binding energy mainly to interactions that involve nonpolar aliphatic and aromatic residues, where the marker is tightly packed within the hydrophobic channels of amyloid fibrils, while interactions that stray from this configurations are generally penalized.

Some discrepancies between the MM/PBSA and PELE energies may be explained in terms of the description of solvation effects. Indeed, it has been shown that continuum solvation models as the one employed in MM/PBSA yield unsatisfactory binding free energies in systems in which exposure to the solvent is significant. This is due to the lack of the molecular description of the hydration state of the binding site.[184] This may be for instance the case of NIAD-4/I-pe, which has a PELE binding energy of -73.8 $kcal\,mol^{-1}$, indicative of a favorable interaction, but a value of 6.0 $kcal\,mol^{-1}$ from the MM/PBSA calculations (Table D.2).

Some uncertainty remains about binding poses such as NIAD-11/III-ji and DANIR-2c/III-pi, which are favorable according to PELE, but present positive binding energies according to MM/PBSA. This lack of agreement cannot be solved unless with more accurate and expensive methods for free energy calculation, such as thermodynamic integration.

5.3 Final remarks

Overall, the binding site search of markers on three different models of Aβ40 fibrils highlighted how marker recognition is based on geometrical considerations rather than involving specific residues. Indeed, stable binding poses involved the accommodation of the dye molecule within the hydrophobic channels that run parallel to the fibrils axis and are defined by the side chains of aromatic and aliphatic nonpolar residues. This observation is crucial because it informs on the tight structural requirements that fluorescent dyes must fulfill for application in amyloid staining. In this regard, it is exemplary the comparison between NIAD-4 and NIAD-11: the structural differences yield a larger sterical hindrance of NIAD-11 compared to NIAD-4, which is enough to prevent NIAD-11 from fully entering the channels, and favor its interaction with the solvent. A second confirmation of the geometrical basis for the selectivity of amyloid binding channels is given by the observation of the binding properties of NIAD-4 and DANIR-2c. These markers, which present quite different structures but share a linear architecture and similar sterical hindrance, have very similar binding poses. Overall, the rational design of fluorescent amyloid markers must focus on preserving a linear aromatic or conjugated structure and take into account the fact that the addition of polar groups may severely reduce the marker's binding affinity by favoring its interaction with water.

From a computational point of view, our results indicate that the program PELE provides a cost-effective approach to the problem of identifying the binding poses of unknown ligand-receptor complexes. Indeed, the simulations, far from being computationally demanding, provide bind-

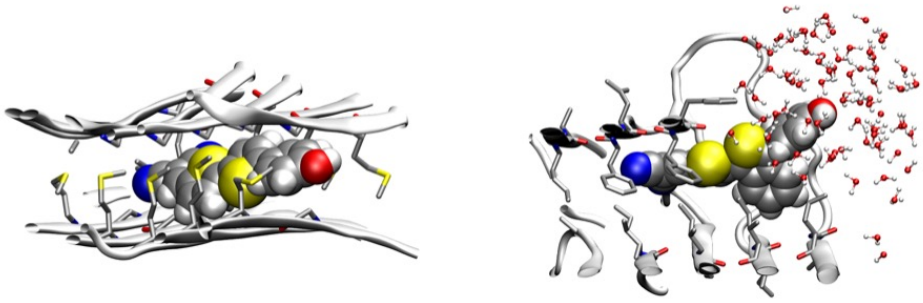


Figure 5.6: Graphical summary

ing energies that are in reasonable agreement with those computed with MM/PBSA.

Chapter 6

Photochemistry of 5-bromouracil

This Chapter is devoted to unravelling the excited state properties of the artificial nuclear base 5-bromouracil by surface hopping dynamics techniques. These results have been obtained during a three months stay in the group of Prof. Leticia González at the University of Vienna. Surface hopping techniques, presented in Chapter 2, are computationally demanding, and this is the main reason why none of the fluorescent markers presented in the previous section has been considered, even if that was the original idea. This is due to the fact that at the time when these calculations were performed, the code SHARC, developed at the González group, only offered an implementation of wavefunction-based methods, whose computational cost scales fast with system size.[133,134] For this reason, the program has been used to study the photochemistry

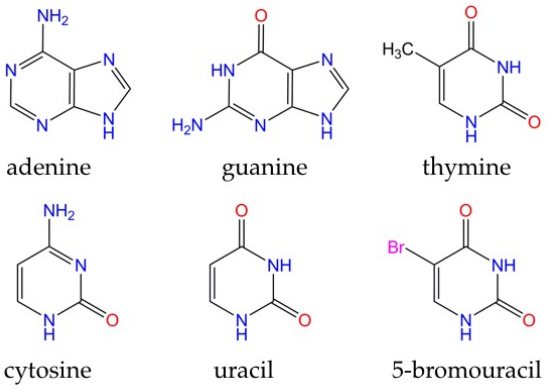
of 5-bromouracil (Scheme 6.3), an artificial nuclear base which has a photosensitization effect on DNA. This choice was motivated by the fact that SHARC has already been applied to a wide variety of natural and artificial nuclear base, with satisfactory results.[185–188]. Before presenting the results, a brief overview of the deactivation mechanisms of natural and artificial nucleobases will be given.

6.1 Natural and artificial nuclear bases

Along with natural nucleobases, adenine, guanine, thymine, cytosine and uracil, artificial ones can be incorporated into DNA (Scheme 6.1). Current endeavors are directed towards the design of unnatural base pairs that can expand the genetic alphabet, and culminated in the creation of semi-synthetic organisms that stably harbor an unnatural base pair in the DNA.[189]

Overall, artificial nucleobases offer a wide range of medical and biochemical applications related with UV absorption. Tackling photophysical processes in systems of such size, involving biological molecules in a complex environment, is a challenge for both experimental and computational techniques.

DNA is extraordinary photostable: even if excited efficiently by UV radiation ($250 \div 280\ nm$), it undergoes very few photoreactions ($< 1\%$). The excitation involves the allowed ${}^1\pi\pi^*$ transitions of the constituent nucleobases, resulting in short-lived ${}^1\pi\pi^*$ excited states, which undergo nonradiative deactivation in the femtosecond time scale.[190] This nonradia-



Scheme 6.1: Molecular structures of natural nucleobases and 5-bromouracil

tive deactivation involves a conical intersection between the $^1\pi\pi^*$ and the ground state, corresponding to deformations of the aromatic rings. These deformations do not affect significantly the energy of the excited state, while they destabilize the ground state due to the loss of electron delocalization.[191–193]

In pyrimidines, a dark $^1\pi\pi^*$ state may mediate the nonradiative deactivation resulting in a cascade of conical intersections.[192,194,195] To a lesser extent, the lowest-energy triplet state is also populated, within a few picoseconds, by an intersystem crossing from the $^1\pi\pi^*$ state, but for natural nucleobases this process has a negligible quantum yield.[194]

Concerning artificial nucleobases, thiopurines such as 6-thioguanine and 6-mercaptopurine are employed as anti-inflammatory, anticancer and immunosuppressive drugs.[196] While canonical DNA and RNA bases show an absorption maximum in the UVC region ($\sim 260\ nm$), thiobases have a

significant absorption in the UVA region ($\sim$ 340 nm). This red-shift is important because it allows for their selective excitation and results in deeper tissue penetration. Thiobases act as photosensitizers by converting the absorbed UVA energy into an unstable triplet state that can interact with oxygen, leading to the formation of reactive oxygen species that in turn cause mutagenic damage to DNA. This photosensitization is a direct consequence of sulfur substitution: unsubstituted thymine has a predominant deactivation pathway through ultrafast internal conversion, as already mentioned, and only a minor pathway leading to the population of T_1. Sulfur substitution greatly enhances intersystem crossing, making it into the dominant deactivation process (heavy atom effect).[196–200] This results in an increased damage of DNA and represents the fundamental reason for the anti-cancer effect of these compounds.

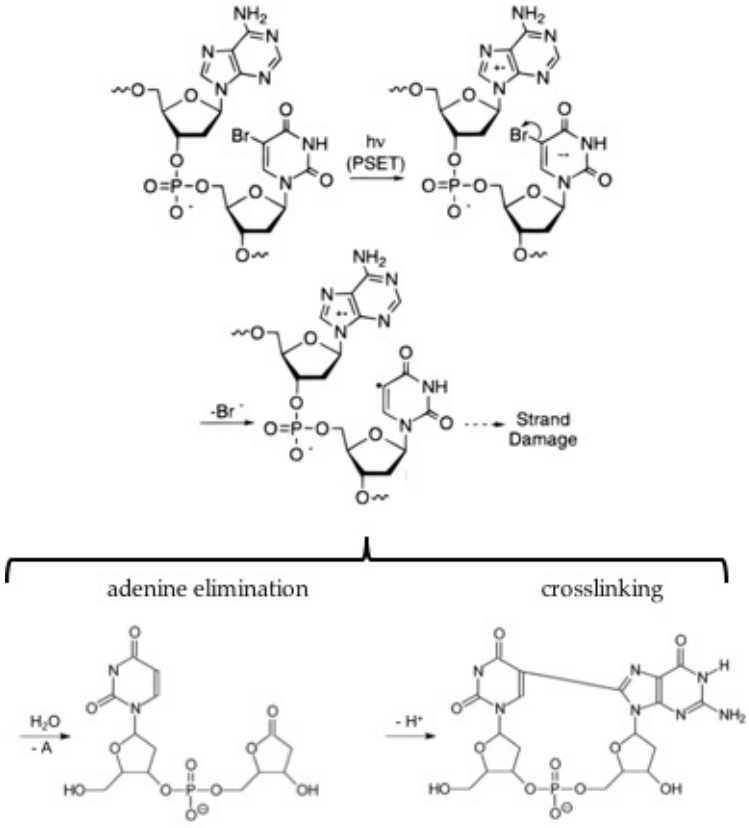
A further application based on the photoreactivity of thiobases is their use as structural probes for nucleic acid structure and nucleic acid-protein interactions; due to the instability of the triplet state, 4-thiouracil and 4-thiothyimine show high crosslinking ability towards pyrimidines and, to a minor extent, towards purines. Analysis of the photocrosslinking patterns can be used to elucidate both single molecule structures and intermolecular adducts, *e.g.* RNA-RNA contacts within the ribosome, conveying precious information on tertiary structures and interactions.[201]

Analogous considerations can be drawn for halogenated artificial nucleobases. 5-fluorouracil and its derivatives have been widely used in the last fifty years in chemotherapy for treating several types of cancer, exerting their effect through a variety of mechanism, including reducing deoxythymidine triphosphate levels, misincorporation of deoxyuridine and

fluorodeoxyuridine triphosphates during DNA replication and disruption of RNA metabolism at several levels.[202] Moreover, 5- fluorouracil undergoes photodegradation after irradiation with UV light, both as pure molecule and in topical anticancer formulations, the main product being the photohydrated form 5-fluoro-6-hydroxyhydrouracil, with the concomitant production of reactive oxygen species representing its underlying mechanism of action as a drug.[203] The exceptional photoreactivity of 5-fluorouracil compared to the related compounds uracil and thymine has been attributed to an energy barrier separating the spectroscopically active $\pi\pi^*$ state from the ground state, which is more pronounced in the halobase than in uracil and thymine, and is responsible for the increased lifetime of the $\pi\pi^*$ state.[204] Despite extensive studies, the photophysical behavior of these compounds has not been uncovered yet. Interestingly, it has been suggested that external properties, such as the hydrogen-bonding power of the solvent, can affect deeply the energy distribution of the excited electronic states of the nucleobase, thus virtually controlling the decay processes of the molecule.[205]

5-bromouracil (Scheme 6.1), an halogenated nucleobase which can be incorporated into DNA replacing tymine, is known for its mutagenic effect, which has been attributed to its tautomeric equilibrium.[206–209] This molecule also has an effect of photosensitization: on UV irradiation, strand breaks are introduced in 5-bromouracil-containing DNA, accompanied by the formation of the photoproduct uracil. This process is initiated by an excited state electron transfer from an adjacent adenine to 5-bromouracil, resulting in the formation of a radical ion pair. The 5-bromouracil radical anion then undergoes cleavage of the $C-$Br bond, with the liberation of a bromide ion, which leaves a highly reactive uracil radical. This radical is

responsible for DNA damage, leading to a variety of processes including DNA crosslinking, adenine elimination and generation of reactive oxygen species (Scheme 6.2).[210–213]



Scheme 6.2: DNA damage resulting form 5-bromuracil incorporation. Picture from ref.[211,212]

Remarkably, it has been shown that DNA sensitization by 5-bromouracil

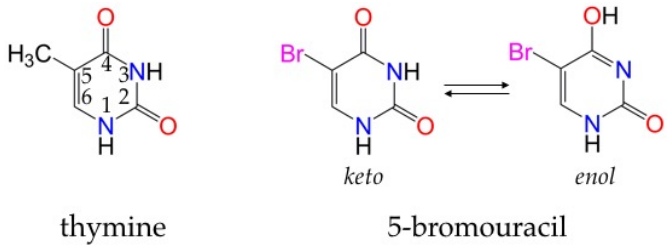
is structure-dependent, with the amount, type and distribution of damage strongly affected by the conformation of DNA. Based on this principle, 5-bromouracil and its derivatives may be employed as tumor radiosensitizers.[214] Moreover, also the isolated 5-bromouracil molecule undergoes cleavage of the $C-$Br bond, in this case homolytic, leading to the formation of the uracil radical and a bromine atom.[209]

The rational design of drugs requires a thorough description of the fundamental molecular processes underlying the reactivity of the bioactive compound. In this light, the description of the deactivation pathways of the isolated 5-bromouracil can provide insight into the structural and energetic aspects that regulate its photoactivity.

6.2 Surface hopping simulations

Among the several techniques that can be employed to describe the photophysical phenomena observed in excited molecules in the fs time scale, dynamic methods such *ab initio* surface hopping[215] simulations represent a particularly suitable tool for isolated nucleobases. As already mentioned in Chapter 2, in surface hopping dynamics nuclear motion is treated classically, the force acting on the nuclei being determined by the gradient of the potential energy, and deactivation processes are explored through a series of independent trajectories. The advantage of such a semiclassical approach is a significant reduction of the computational cost compared to more rigorous methods, which allows the study of large systems with many degrees of freedom. Standard implementations of this method do not include spin-orbit couplings, which are particularly rel-

evant for the description of the photophysics of artificial nuclear bases, especially those containing heavy atoms such as 5-bromouracil. However, the SHARC methodology (Surface Hopping with ARbitrary Couplings), developed in Prof. González group, incorporates couplings arising from both the spin-orbit interaction and laser fields (see Chapter 2).[133,134] This method has been applied to the study of several photosystems, including both natural and artificial nucleobases, and proven to yield accurate and robust results compared with experiments and more accurate computational techniques.[186,188] In this Chapter, SHARC dynamics simulations were performed to study the deactivation processes of the isolated 5-bromouracil molecule. Only the *keto* tautomer has been taken into account, as it has been reported to be the most stable in gas phase (Scheme 6.3).[207]



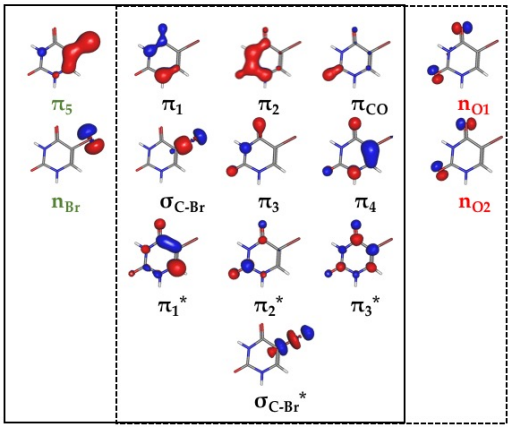
Scheme 6.3: Molecular structures of thymine and 5-bromouracil (*enol* and *keto* forms).

In 2009, Blancafort and coworkers[216] presented a static study on the photophysical properties of 5-bromouracil. According to this work, a correct description of the excitation energies of the molecule with a MS-CASPT 2(16,12) method requires two distinct sets of orbitals, one for the description of $n_O\pi^*$ excitations and one for the description of $n_{Br}\pi^*$ ex-

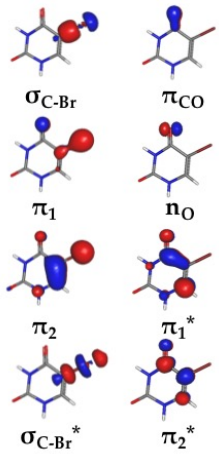
Table 6.1: $^{\alpha}$ Computed MSCASPT2(16,12)/ANO-RCC-VDZP[105,110] excitation energies at the Franck-Condon geometry reported in ref.[216]. Oscillator strength in parentheses; experimental value from ref.[217] in square brackets. Orbital labels refer to Figure 6.1. $^{\beta}$ Computed MRCIS(10,8)/CC-VDZP-DK[135,218,219] excitation energies at the MRCIS(10,8)/CC-VDZP-DK optimized geometry. Oscillator strength in parentheses; experimental value from ref.[217] in square brackets. Orbital labels refer to Figure 6.1.

	$MSCASPT2(16,12)^{\alpha}$		$MRCI(10,8)^{\beta}$	
S_1	5.1 (0.3) [4.7]	$\pi_4 \rightarrow \pi_1^*$	5.0 (0.0)	$n_O \rightarrow \pi_1^*$
S_2	5.1 (0.0)	n_{O1},$n_{O2} \rightarrow \pi_1^*$	6.1 (0.0)	$\pi_2 \rightarrow \sigma CBr^*$
S_3	6.3 (0.0)	$\pi_4 \rightarrow \sigma CBr^*$	6.2 (0.3) [4.7]	$\pi_2 \rightarrow \pi_1^*$
T_1	3.8 [3.4]	$\pi_4 \rightarrow \pi_1^*$	4.0 [3.4]	$\pi_2 \rightarrow \pi_1^*$
T_2	5.1	n_{O1},$n_{O2} \rightarrow \pi_1^*$	4.8	$n_O \rightarrow \pi_1^*$
T_3	5.4	$\pi_4 \rightarrow \sigma CBr^*$	5.5	$\pi_2 \rightarrow \sigma CBr^*$

citations. Based on this scheme, we performed MSCASPT2(16,12)/ANO-RCC-VDZP calculations using the two sets of orbitals reported in Figure 6.1a, solid line, center left, for $n_{Br}\pi^*$ excitations, and dashed line, center right, for $n_O\pi^*$ excitations, with state averaging over 6 singlets and 5 triplets, which guarantees converged results.[102,105,110,216] The IPEA shift was set to 0.25 and no level shift was employed.[108,109] The three first singlet and triplet excitation energies computed at the Franck-Condon geometry taken from the literature.[216] are shown in Table 6.1, and compared with experimental values. The complete excitation energies are reported in Table E.1. The computed MSCAPST2(16,12) energies are in fair agreement with experimental data reported in ref.[217], indicating the presence of a dark $n_O\pi^*$ state energetically close to the spectroscopically active $\pi\pi^*$ state, as already reported for uracil.[185]



(a) MSCASPT2(16,12).



(b) MRCIS(10,8).

Figure 6.1: Molecular orbitals in the orbital spaces of the MSCASPT2(16,12)/ANO-RCC-VDZP and MRCIS(10,8)/CC-VDZP-DK calculations.

COLUMBUS was chosen as the external code for SHARC calculations (for further details see Chapter 2).[135,220] Calculations were performed with the MRCIS method and a relativistic CC-VDZP-DK basis set,[218,219] using as a reference a CASSCF(10,8) calculation with state averaging over 4 singlet and 3 triplet states. The reason for this change of computational code is that the MRCIS(10,8) method that has been employed, which includes only determinants up to the diexcited in the reference function, is significantly cheaper that MSCASPT2(16,12). The change of basis set from the ANO type to the correlation consistent type, is also linked to the cost of the calculations. Even if both basis sets are of the same double zeta quality, indeed, the first has a larger nuber of primitive functions than

the latter, resulting in a higher computational cost. The vertical excitation energies computed at the MRCIS(10,8)/CC-VDZP-DK level of theory at the MRCIS(10,8)/CC-VDZP-DK optimized geometry are reported in Table 6.1, along with the values of oscillator strength. The corresponding molecular orbitals are reported in Figure 6.1b. Excitation energies appear to be in good agreement with experimental values [217] and MSCASPT2 calculations, [216] with the exception of the spectroscopically active $^1\pi\pi^*$ state, whose energy is about 1 eV overestimated. This overestimation introduces a large error in our calculations, which will certainly affect the quality of the results, as will be discussed further on. However, this method describes well the energy of the dark $n_O\pi^*$ state, which, based on what observed for other nuclear bases, is expected to play an important role in the photophysics of the system.

6.3 Absorption spectrum

The ground state geometry of 5-bromouracil was optimized, and its harmonic frequencies computed, at the MP2/ANO-RCC-VDZP [218,219] level of theory with the Gaussian package. [96] The frequencies were then employed to build a Wigner-Ville distribution. This is a a time-frequency analysis technique, which in this case is used to generate a set of 1000 geometries that served to compute the absorption spectrum reported in Figure 6.2. [221] Vibrational frequencies were computed with MP2 and not with MRCI because the interest here is just to obtain a pool of structures, however it was confirmed that the optimized ground state geometries at the two levels of theory are similar.

The absorption spectrum presents a sharp peak centered at 5.9 eV, arising from the spectroscopically active $\pi_2 \rightarrow \pi_1^*$ state, which is S_2 or S_3 according to the geometry. The contributions of S_2 and S_3 to the total spectrum are reported Figure 6.2 as well, as a green and a pink band respectively. Comparison with experimental data[209,217] suffers from two major drawbacks: the first is the lack of solvation effects in the calculated value; the second is the overestimation of the energy of the $\pi\pi^*$ transition at the MRCIS(10,8)/CC-VDZP-DK level of theory. However, the shape of the peak correlates fairly with the EELS spectra reported in ref.[217]. Based on the geometries sampled from the Wigner-Ville distribution, initial conditions for the dynamic simulations were generated selecting an interval of excitation energies of $\pm$ 0.15 eV around the maximum.[222] This interval of 0.30 eV has been chosen to match the amplitude of an experimental laser. Overall, 65 initial conditions were selected starting from both S_2 and S_3.

6.4 Intersystem crossing

Our results show that an important deactivation pathway of 5- bromouracil is ultrafast intersystem crossing involving the dark state S_1 and triplets T_1 and T_2. As already reported for uracil,[185,193] from the spectroscopically active singlet $\pi\pi^*$ state, the molecule readily deactivates to the dark $n_O\pi^*$ S_1 state. From here, in both uracil and 5-bromouracil, intersystem crossing to T_1, which has a $\pi\pi^*$ character, may occur at geometries that differ from the ground state geometry by compressions and elongations along the C=O bonds and deformations of the ring.[185,193] While this

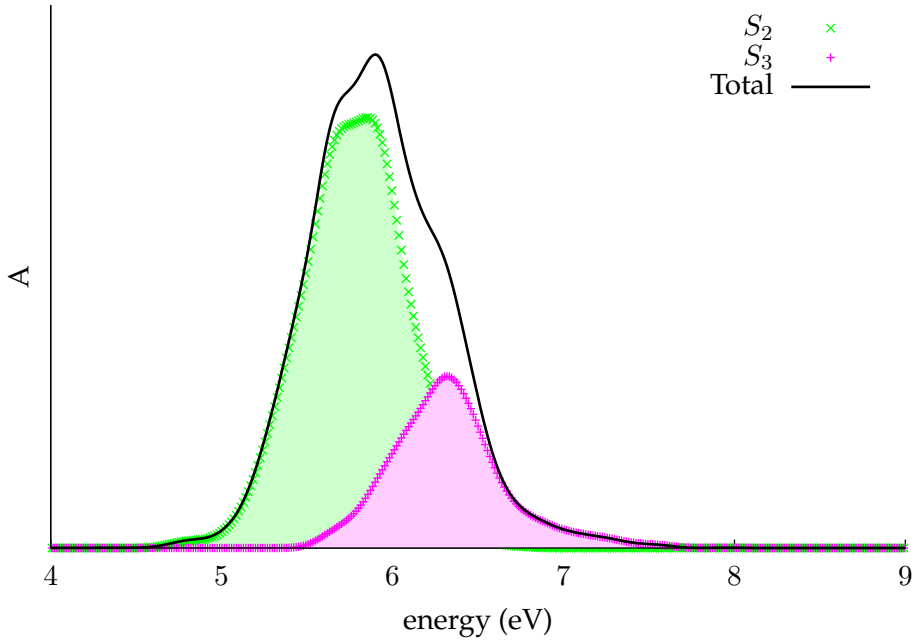


Figure 6.2: Calculated absorption spectrum of the *keto* tautomer of 5-bromouracil at the MRCIS(10,8)/CC-VDZP-DK level of theory. Solid black line: total absorption spectrum; green area: contribution of state S_2; pink area: contribution of state S_3.

process is only marginal in the case of uracil, for 5-bromouracil it accounts for 52% of the trajectories, being an important decay pathway.

This difference between the two molecules can be attributed to the so-called "heavy-atom effect":[223] the presence of a heavy atom such as Br increases the values of spin-orbit coupling between electronic states, which in turn affects the rate of intersystem crossing. This is in agreement with experimental data, which report the presence of an intersystem

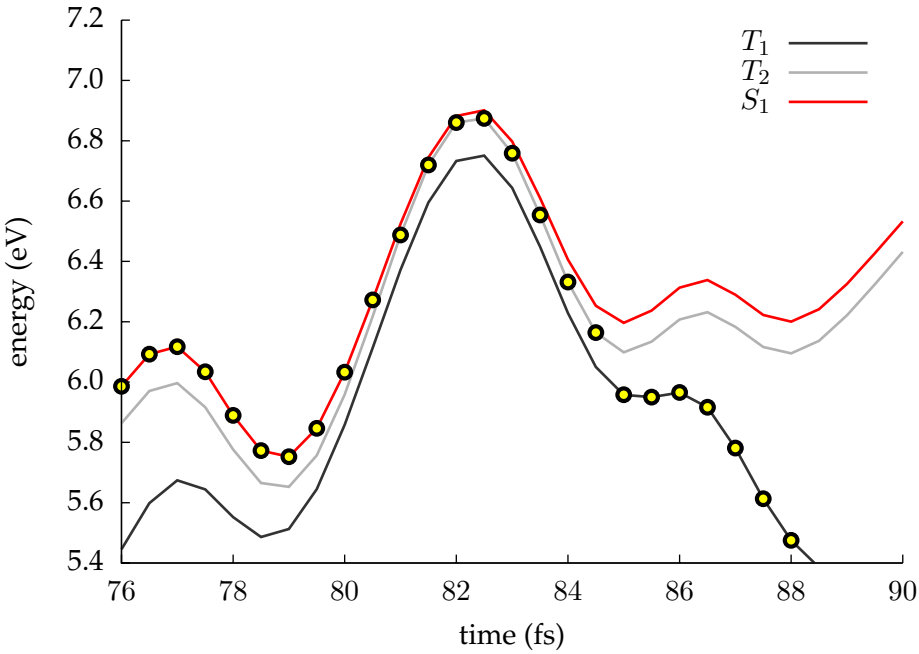


Figure 6.3: Potential energy profiles associated with ultrafast intersystem crossing involving S_1 and T_1. The active state is represented with circles.

crossing to a $^3\pi\pi^*$ state.[213,224] From the point of view of surface hopping trajectories, the intersystem crossing from S_1 to T_1 appears to be mediated by triplet T_2, as reported in Figure 6.3. T_2 has an $n_O\pi^*$ character, and evolves parallel to S_1, which has a similar spatial part, along the simulation. The spin-orbit coupling between $^1n_O\pi^*$ and $^3n_O\pi^*$ is small, while the one between $^1n_O\pi^*$ and $^3\pi\pi^*$ is large. Intersystem crossing takes place at a geometry for which $^1n_O\pi^*$ and $^3\pi\pi^*$ are close in energy, as reported in Figure 6.3. At this geometry, $^3\pi\pi^*$, which is usually much lower in energy than $^3n_O\pi^*$, surpasses it and becomes T_2. After population transfer,

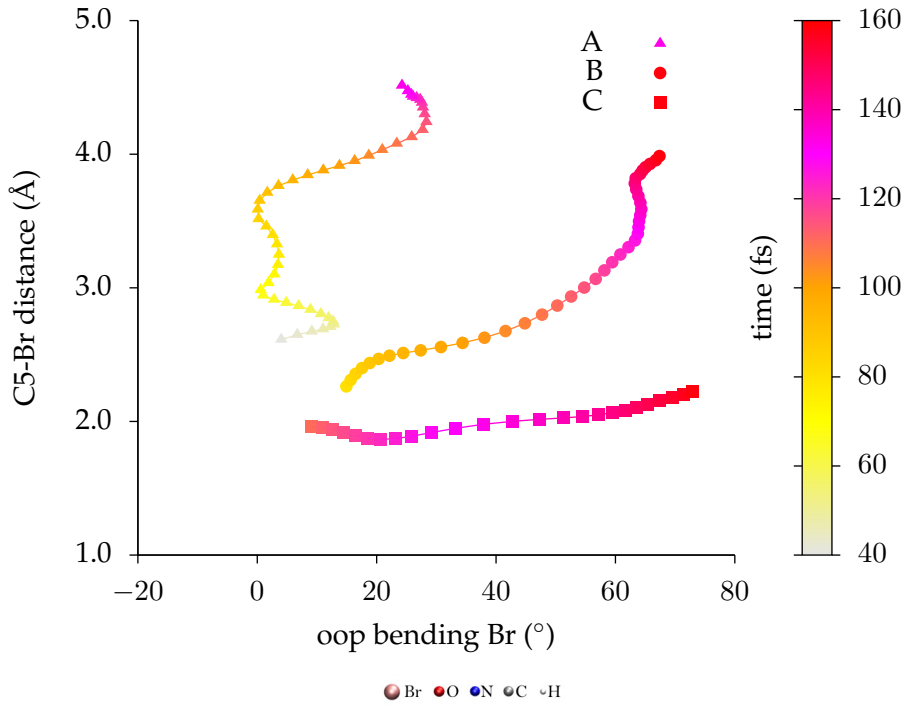
the energy of $^3\pi\pi^*$ rapidly decreases, and the state is labeled again as T_1. Thus, the overall transfer $S_1 \rightarrow T_2 \rightarrow T_1$ is attributed to the $^1n_O\pi^*$ to $^3\pi\pi^*$ intersystem crossing.[225]

Alternatively, intersystem crossing from S_1 can involve exclusively the $^3n_O\pi^*$ state T_2. In this case the molecule evolves in the T_2 state towards the dissociation of the $C5-$Br bond. This process is described in the next section.

6.5 Photolysis and reactant regeneration

As already mentioned, 5-bromouracil can have the $C5-$Br bond cleaved, homolytically, after irradiation with UV light, leading to the formation of a bromine atom and a uracil radical. This process has been investigated experimentally in terms of the quantum yield of 5-bromouracil consumption in solution, which has been shown to be low and affected by both the pH and the nature of the solvent (from Φ=1.8 10^{-3} at pH=6 to 0.012 at pH=10 in aqueous solution, 0.025 in pure methanol).[209,213,224] Cleavage of the $C5-$Br bond is followed by either recombination or hydrogen abstraction from the solvent (Scheme 6.3).[209] Alternatively, the molecule may deactivate through a conical intersection, as reported for other pyrimidines.[192,194,195].

According to what was shown in 2009 by Blancafort and coworkers,[216] there are two internal coordinates that are critical for the description of the deactivation processes of 5-bromouracil: the $C5-$Br distance of the bond that is cleaved, and the out-of-plane bending angle of the



(a) S_0 geometry. (b) A (c) B (d) C

Figure 6.4: Three surface hopping trajectories representative of deactivation processes of 5-bromouracil: photolysis (A and B) and reactant regeneration through a conical intersection (C). The evolution of the trajectories is represented by plotting the values of two internal coordinates: the $C5-$Br distance of the bond that is cleaved, and the out-of-plane bending angle of the bromine atom.

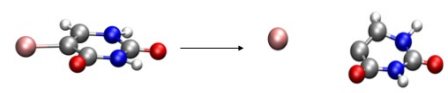
bromine atom. According to their work, 5-bromouracil presents an extended region of degeneracy or near-degeneracy between the spectroscopically active $\pi\pi^*$ state and the ground state along a combination of the two internal coordinates. This results in the two deactivation pathways: i) reactant regeneration through a conical intersection, which involves geometries with a large value of the out-of-plane bending angle of the $C5-$Br bond; ii) photolysis, involving the elongation of the $C5-$Br bond. Between these two limiting cases, the authors hypothesized several intermediate pathways involving different combinations of the relevant internal coordinates. Surface hopping dynamics can complete this static picture and provide insight in the structural and energetic modifications that accompany the different deactivation processes.

Figure 6.4 reports three exemplary trajectories, labeled A, B and C. The structural evolution of 5-bromouracil along the simulations is represented by plotting selected values of the two internal coordinates of interest, the $C5-$Br distance along the y axis, and the out-of-plane bending angle of Br along the x axis. The temporal evolution is reported in terms of a color scale of the points representing the steps of the simulation. Only part of each trajectory is reported for clarity. In all three cases, A, B and C the simulation starts from the Franck-Condon (FC) region, which is located at the bottom-left of the plot, being characterized by small values of both the $C5-$Br bond and the Br bending angle, and evolves to different regions of the plot according to the preferred deactivation process. Trajectories A and B both lead to the photolysis of the molecule as they terminate in the upper part of the plot, which corresponds to large values of the $C5-$Br distance. The difference between the two lies in the combination of the two relevant internal coordinates: while in the case of A

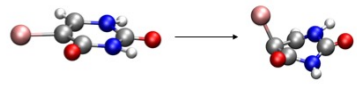
dissociation takes place by pure elongation of the $C5-$Br bond, without significant variations of the bending angle, in the case of B it involves both elongation and bending of the $C5-$Br bond. A and B only represent two examples of trajectories leading to photolysis of 5-bromouracil; analysis of the whole pool of simulations allows us to conclude that photolysis can take place along any combination of the two internal coordinates, effectively covering the whole upper portion of the plot reported in Figure 6.4 and confirming what hypothesized by Blancafort and coworkers.[216] Remarkably, our trajectories indicate that dissociation can take place from both the S_1 $n_O\pi^*$ state and the T_2 $n_O\pi^*$ state.

Trajectory C, on the other hand, is representative of a different deactivation pathway: starting from the Franck-Condon region, it evolves towards geometries characterized by small values of the $C5-$Br distance, indicating that 5-bromouracil does not undergo photolysis, and by large values of the out-of-plane bending angle of Br (lower-right part of the plot). In this region, the molecule can access a conical intersection and deactivate to the ground state. However, our dynamic simulations highlight that the two competing decay pathways, photolysis and reactant regeneration, are not to be considered as separate processes. Since the molecule can easily access virtually any combination of the two internal coordinates from the Franck-Condon region, between the two limiting cases exemplified by the axes of Figure 6.4, the full area of the plot can be explored.[225]

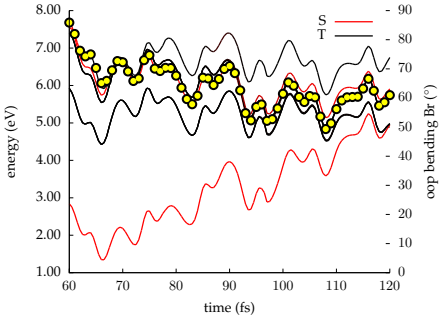
Photolysis and regeneration of the reactant through conical intersection present markedly different energetic features. Figure 6.5 illustrates the energy of the electronic states involved in the process as a function of simulation time for a trajectory leading to the photolysis of the molecule



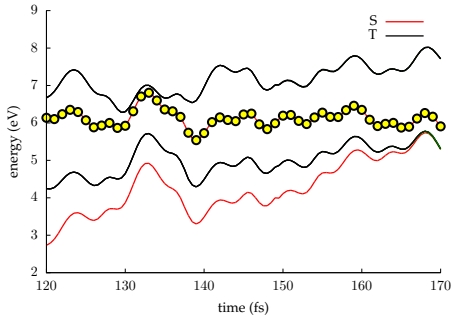
(a) Photolysis (**P**).



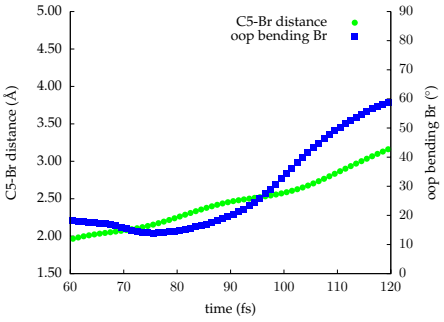
(b) Reactant regeneration (**R**).



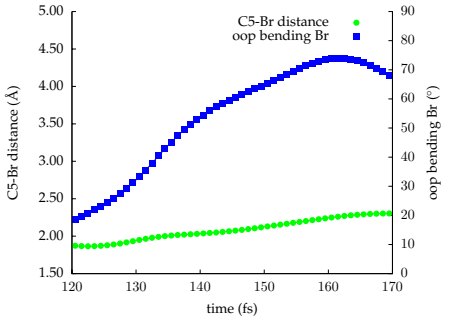
(c) **P** energy profiles.



(d) **R** energy profiles.



(e) **P** structure modifications.



(f) **P** structure modifications.

Figure 6.5: Trajectories leading to photolysis and reactant regeneration. Energy plots: singlet states are represented in red, triplet states in black. The active state is represented with circles.

(Figure 6.5c) and a trajectory leading to the region of conical intersection (Figure 6.5d). In both cases, the active state is represented by yellow circles. The corresponding evolution of the $C5-$Br bond distance and out-of-plane bending angle of Br along the simulation are reported in Figures 6.5e and 6.5f respectively. In the case of photolysis, the energy of the active state smoothly decreases along the simulation as the $C5-$Br distance increases, along with the energies of other low-lying singlet and triplet states. At 110 fs, 5-bromouracil can be considered dissociated (Figure 6.5e, $C5-$Br = 3 Å, green points), and all lower levels are degenerate. The out-of-plane bending angle (blue points) undergoes variations along the dissociation, but is confined to values lesser than 30° as long as Br and C5 are bound. In the case of reactant regeneration, on the other hand, the energy of the active state ($n_O\pi^*$) is not affected by the bending of Br, and keeps oscillating around a constant value. Rather, it is the energy of the ground state that steadily rises towards the conical intersection, as already reported for other pyrimidines.[191] This means that, energetically, photolysis is much more favorable than deactivation through the conical intersection. This energy difference between the two decay pathways is reflected in their relative frequency: while dissociation is observed, reactant regeneration is not observed, even if some trajectories explore the conical intersection region. Trajectories that access the conical intersection region (small values of $C5-$Br bond distance and large values of Br bending angle) eventually deactivate by either photolysis or intersystem crossing to T_1.

Unfortunately, these results cannot be compared straightforwardly to experimental data for a series of reasons: i) first of all, our simulations lack the description of environmental effects. Experimentally, the quan-

tum yield of photolysis is determined by measuring the consumption of 5-bromouracil,[209] or the quantum yield of uracil formation, uracil being the product of hydrogen abstraction by the uracil radical. Besides this reaction, several other processes can take place that involve reactions of the radicals with the solvent, which cannot be described by our calculations.[224] ii) The deactivation mechanism itself may be strongly affected by the solvent. In 1981, Swanson *et al.*[224] showed that photolysis can take place from both singlet and triplet states, but while in the case of the singlet 5-bromouracil reacts via an homolysis of the $C5-$Br bond, in the triplet state 5-bromouracil undergoes an ionic reaction with the solvent (2-propanol) via electron transfer. Our simulation also indicate photolysis from both S_1 and T_2, but with an homolytic mechanism in both cases. iii) Moreover, the solvent exerts a cage effect on 5-bromouracil, promoting recombination after photodissociation; since the product of recombination is the same bromouracil molecule, it is impossible to distinguish between the product of recombination and that of reactant regeneration through the conical intersection. This, in turn, leads to an underestimation of the importance of photolysis. iv) The 1 eV overestimation of the energy of the spectroscopically active $^1\pi\pi^*$ state may affect the computed quantum yield of the deactivation processes.[225]

These are for sure severe problems, that do not allow the understanding of the phenomenona happening in the biological environment. In this sense, the method chosen for these simulations is not the correct one for the system at hand, since not only it does not describe correctly the energy distribution of the states of interest, but it also completely lacks the description of the environment, which in this case is fundamental to determine the photochemical processes under study. These results can be im-

proved by running the more expensive MSCASPT2 simulations, but more importantly including the description of environmental effects, for example through a hybrid QM/MM method that allows the explicit inclusion of at least solvent molecules.

More in general, the process of interest here is the deactivation of 5-bromouracil incorporated into the DNA. This process itself cannot be described correctly by describing only the 5-bromouracil and the solvent, because, as mentioned, in 5-bromuracil containing DNA, photolysis is initiated by an electron transfer. For this reason, a correct description of the phenomenon must also include this initial step.

Despite these problems, our results provide valuable information on the isolated 5-bromouracil molecule. Summarizing, we uncovered the relationship between photolysis intended as pure elongation of the $C5-$Br bond, and decay through the conical intersection intended as pure out-of-plane bending of the $C5-$Br bond, which are not to be considered as separate processes but rather limiting cases of a continuous spectrum of deactivation pathways. Also, our results clearly indicate that dissociation of the $C5-$Br bond from either S_1 or T_2, along with intersystem crossing to T_1, are the main deactivation processes of the molecule and have roughly the same importance.

Moreover, Blancafort and coworkers[216] suggested that for large values of both the $C5-$Br distance and the bending angle, there is a region of near-degeneracy between the $^1\pi\pi^*$ state and the ground state which involves the transfer of Br form $C5$ to $C6$ (atom labels reported in Scheme 6.3). The difference between the $C6-$Br and $C5-$Br bond distance has been computed along the dynamics trajectories, and at no geometry Br is

closer to $C6$ than to $C5$ (Figures E.1 and E.2), indicating that in our calculations no Br trasfer is produced, in disagreement with what previously reported. [216]

6.6 Deactivation pathways

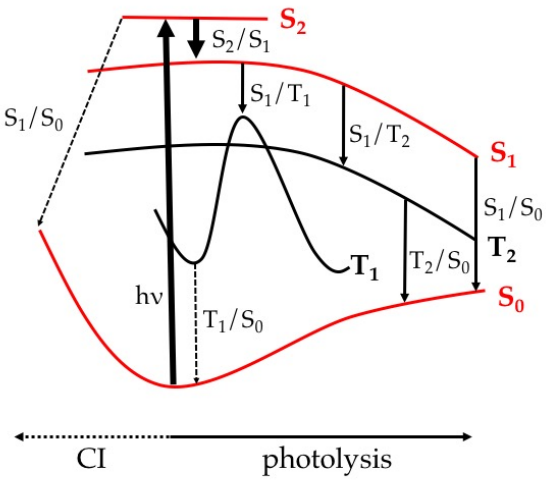


Figure 6.6: Schematic representation of the deactivation pathways of 5-bromouracil; dashed lines refer to processes that have been hypothesized but not observed. CI: conical intersection.

Figure 6.6 summarizes the photophysical and photochemical processes undergone by 5-bromouracil. After the molecule is excited to the bright $^1\pi\pi^*$ state (S_2), it rapidly decays to the dark $^1n_O\pi^*$ state (S_1) via an ultra-

fast internal conversion S_2/S_1; from here, it undergoes either photolysis (S_1/S_0, solid line) or intersystem crossing (S_1/T_1 and S_1/T_2). An internal conversion to the ground state has been hypothesized but not observed (S_1/S_0, dashed line). Intersystem crossing to the $^3n_O\pi^*$ state (S_1/T_2) is followed by the photolysis of the molecule (T_2/S_0), while intersystem crossing to the $^3\pi\pi^*$ state (S_1/T_1) does not lead to dissociation, and the molecule is expected to finally decay to the ground state by intersystem crossing (T_1/S_0). Overall, 17% of the trajectories lead to photolysis from S_1, 31% to photolysis from T_2 and 52% to intersystem crossing to T_1.

Figure 6.7 shows the temporal evolution of the excited state populations of the trajectories that do not lead to photolyisis. The populations were fitted to a sequential kinetic model, which allowed the calculation of a time constant for intersystem crossing of $450 \pm 100\ fs$, which is in good agreement with the experimental value of 0.4 ps reported in ref[226].

6.7 Final remarks and graphical summary

The surface hopping dynamics simulations presented in this Chapter help uncover the structural and energetic modifications that accompany the deactivation of isolated 5-bromouracil after excitation with UV light. The computed absorption spectrum results from excitation to a bright $^1\pi\pi^*$ state. After an ultrafast internal conversion from the spectroscopically active $^1\pi\pi^*$ to the dark $^1n_O\pi^*$ state (S_1), an important decay pathway involves dissociation of the $C5-$Br bond (48% of the trajectories), that can take place from both singlet and triplet $n_O\pi^*$ states, T_2 being populated by intersystem crossing from S_1. This process is accompanied by a com-

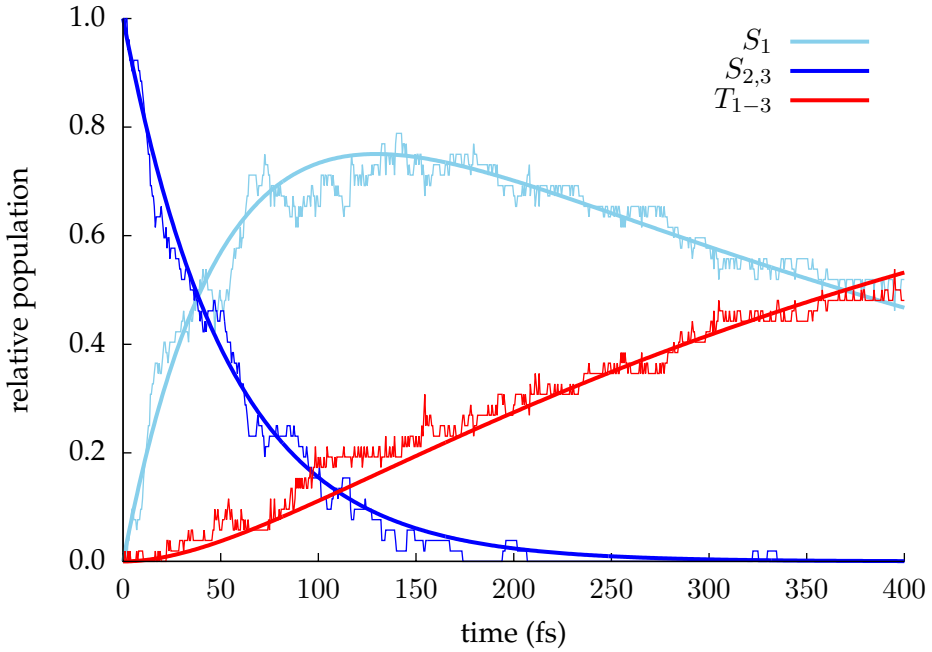


Figure 6.7: Temporal evolution of the excited state populations of non-dissociating trajectories.

bination of elongation and out-of-plane bending of the $C5-$Br bond and is barrierless, resulting in a smooth decrease in the energy of the S_1 state. Competing with photolysis is an intersystem crossing from S_1 to T_1 ($^3\pi\pi^*$), which does not lead to dissociation and accounts for 52% of the trajectories.

Deactivation through a conical intersection from S_1 to S_0, which has been suggested by static calculations, has not been observed. This process is accompanied by the out-of plane bending of the $C5-$Br bond up to a 90°, with negligible variations of the bond length with respect to the ground state structure. Trajectories evolving towards this kind of geometries clearly show that the energy of S_1 is not affected by the bending of the $C5-$Br bond, meaning that there is no energy gain in performing this deformation, which explains the marked preference of 5-bromouracil for photolysis.

The simulation presented in this Chapter cannot be compared with experimental values nor can they provide insight into the deactivation processes undergone by 5-bromouracil in the biological environment. This is due to two main limitations of these simulations. First, the lack of description of the environment. Since the system at hand undergoes photolysis, the solvent is expected to severely affect the rate of this process, and must be included in the description. Secondly, since photolysis is initiated by an electron transfer from an adjacent nucleobase, also part of the DNA chain must be described to some extent. A further problem of these simulations lies in the choice of the computational method: the MRCI(10,8) employed here overestimates the energy of the spectroscopically active $\pi\pi^*$ state, which is also expected to severely affect the results.

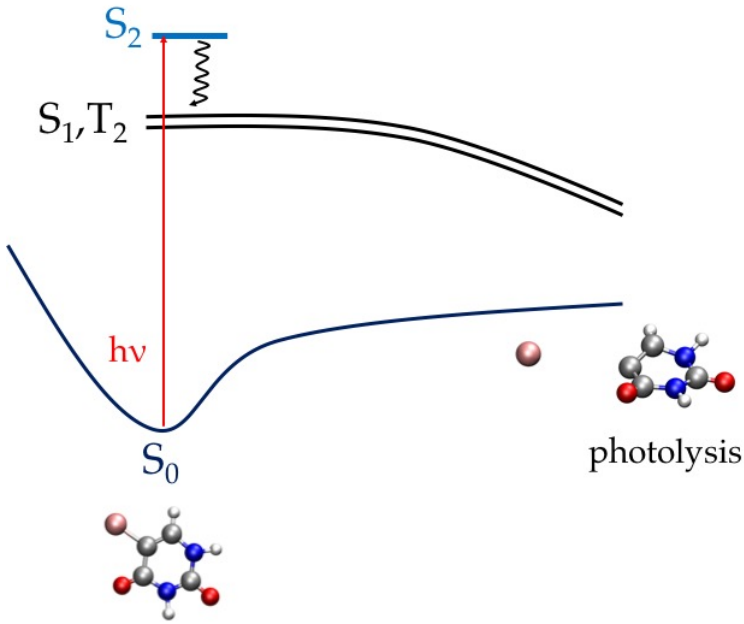


Figure 6.8: Graphical summary

Chapter 7

Final remarks

The simulation of photophysical processes of biological interest is a complex problem. In this thesis, TDDFT has been employed for the excited state calculations conducted on fluorescent markers ($\sim$ 35 atoms), while post-Hartree-Fock calculations, such as CASPT2, have been used for calibration. The protein energy landscape exploration method has been used to study the fibril-marker interaction and additionally DFT and force field methods have been used to run molecular dynamics simulations aimed at exploring the configuration space of the systems. The combination of these techniques is effective because it allows a full understanding of the properties of the fibril/marker complexes.

Surface hopping dynamics including arbitrary couplings has also been performed. When post-Hartree-Fock methods are employed, these simulations are computationally demanding. For this reason, it is necessary

to reach a compromise between cost and accuracy, sometimes at the expense of the quality of the calculation. Indeed, the simulations presented here on the photochemistry of 5-bromouracil are not satisfactory and offer a chance to reflect on the importance of the inclusion of environmental effects in the study of processes of biological interest.

Based on the results obtained on two classes of markers, NIADs and DANIRs, some general considerations may be drawn on the relationship between structural, photophysical and binding properties of amyloid markers. Concerning binding to amyloid fibrils, our results uncovered how this interaction is driven by purely geometric and polarity consideration, rather by the specific interactions of the markers with certain aminoacidic residues.

About the photophysical properties, it can be said that despite the fact that both families of markers share a common donor-π-acceptor structure, resulting in a charge transfer transition, they undergo quite different deactivation processes in aqueous solution. For example, the presence of heavy atoms such as sulfur opens the possibility of nonradiative deactivation through intersystem crossing, as is the case of NIAD markers. Furthermore, the presence of double bonds, observed in both classes of markers, results in effective internal conversion through conical intersections. For markers of the DANIR family, the fluorescence enhancement is determined by the sterical hindrance resulting from the inclusion of the fluorescent molecule in the hydrophobic channels of amyloid fibrils, which prevents deactivation via internal conversion. This internal conversion involves distorted geometries, corresponding to conical intersections, which are no longer accessible to the marker enclosed in the fibril. This,

however, only results in a small fluorescence enhancement observed upon interaction with amyloid fibrils.

These processes alone do not provide a satisfactory description of the photophysics of the makers in the biological environment because the activity of these molecules is strongly affected by external factors, such as solubility. For NIAD-4, the solubility is sufficiently low to promote dimerization in water, resulting in fluorescence quenching. While disaggregation of these dimers upon interaction to amyloid-β deposits provides an optimal fluorescence enhancement, solubility is a delicate mechanism to rely upon for an amyloid marker. Since the distribution of the molecule among the monomeric and aggregated forms in the biological environment is impossible to determine in advance, nor is the permeability of these species through the blood-brain barrier, it is desirable to find alternative ways to tune the different fluorescence intensity in water and in presence of amyloid fibrils.

Moreover, the rational design of amyloid markers should strive towards the maximization of the fluorescence enhancement observed when the molecule is bound to amyloid-β deposits. Unfortunately, unless solubility intervenes, only a small fluorescence enhancement is observed for markers, such as those presented, that spontaneously planarize upon excitation. This result highlights an intrinsic limitations of these compounds, resulting in an upper bound of $\sim$ 10-fold for the fluorescence enhancement observed on interaction with amyloid-β deposits.

For instance, it would much increase the fluorescence enhancement to design molecular rotors, built in analogy to, for instance, Thioflavin-T. These markers are expected to spontaneously evolve to a twisted internal

charge transfer nonemitting state when excited in aqueous solution, resulting in an utter lack of fluorescence emission. When bound to amyloid-β fibrils, these markers planarize, favoring radiative deactivation through fluorescence emission. In this way, the molecules that are completely dark when free in solution, would be "switched on" upon binding. This transition from a fully dark to a fully emitting state greatly enhances fluorescence ($\sim$ 1000-fold). However, such a rotor for *in vivo* applications still remains to be proposed. In this regard, relying on the insights provided by these studies on the complex relationship between structure, fluorescence properties and amyloid-β affinity, it may be possible to design *in silico* a rotor-based amyloid marker satisfying all the requirements for *in vivo* amyloid staining, and at the same time offering a good imaging contrast. This achievement could eventually lead to a further development of fluorescence imaging, which may become a valid tool for clinical practice.

Bibliography

[1] Alzheimer's Association. *Alzheimer Dement.*, 12(4):459–509, 2016.

[2] R. Katzman. *Arch. Neurol.*, 33(4):217–218, 1976.

[3] Alzheimer's Disease International. Policy brief for heads of government: the global impact of dementia 2013–2050. 2013.

[4] Y. Wu, L. Fratiglioni, F. E. Matthews, A. Lobo, M. M. B. Breteler, I. Skoog, and C. Brayne. *Lancet Neurol.*, 15(1):116–124, 2016.

[5] P. J. Nestor, P. Scheltens, and J. R. Hodges. *Nature Rev. Neurosci.*, 5: S34–S41, 2004.

[6] H. Braak and E. Braak. *Acta Neuropathol.*, 82(4):239–259, 1991.

[7] G. Benzi and A. Moretti. *Neurobiol. Aging*, 16(4):661–674, 1995.

[8] J. Mayes, C. Tinker-Mill, O. Kolosov, H. Zhang, B. J. Tabner, and D. Allsop. *J. Biol. Chem.*, 289(17):12052–12062, 2014.

[9] J. Alí-Torres, J. D. Marechal, A. Mirats, C. Rodríguez-Rodríguez, L. Rodríguez-Santiago, and M. Sodupe. *AIP Conf. Proc.*, 1618(1):106–108, 2014.

[10] S. S. Sisodia. *Proc. Natl. Acad. Sci. USA*, 89(13):6075–6097, 1992.

[11] C. Thinakaran and E. Koo. *J. Biol. Chem.*, 283(44):29615–29619, 2008.

[12] A. Mahiuddin, J. Davies, D. Aucoin, T. Sato, S. Ahuja, S. Aimoto, J. I. Elliott, W. E. Van Nostrand, and S. O. Amith. *Nat. Struct. Mol. Biol.*, 17(5):561–567, 2010.

[13] C. L. Masters, G. Simms, N. A. Weinman, G. Multhaupt, B. L. McDonald, and K. Beyreuther. *Proc. Natl. Acad. Sci. USA*, 88(44):4245–4249, 1985.

[14] D. J. Selkoe and M. S. Wolfe. *Cell*, 131(2):215–221, 2007.

[15] R. Tycko. *Quarterly Rev. Biophys.*, 39(1):1–55, 2006.

[16] A. K. Paravastu, R. D. Leapman, W. M. Yau, and R. Tycko. *Proc. Natl. Acad. Sci. USA*, 105(47):18349–18354, 2008.

[17] T. Lührs, C. Ritter, M. Adrian, D. Riek-Loher, B. Bohrmann, H. Döbeli, D. Schubert, and R. Riek. *Proc. Natl. Acad. Sci. USA*, 102 (48):17342–17347, 2005.

[18] A. E. Roher, J. D. Lowenson, S. Clarke, A. S. Woods, R. J. Cotter, E. Gowing, and M. J. Ball. *Proc. Natl. Acad. Sci. USA*, 90(22): 10836–10840, 1993.

[19] Y. Masuda, S. Uemura, R. Nakanishi, A. Ohashi, K. Takegoshi, T. Shimizu, Shirasawa T., and Irie K. *Bioorg. Med. Chem. Lett.*, 18 (11):3206–3210, 2008.

[20] L. Gu and Z. Guo. *J. Neurochem.*, 126(3):305–311, 2013.

[21] J. Alí-Torres, J. D. Maréchal, L. Rodríguez-Santiago, and M. Sodupe. *J. Am. Chem. Soc.*, 133(38):15008–15014, 2011.

[22] Y. Miller, B. Ma, and R. Nussinov. *Proc. Nat. Acad. Sci. USA*, 107(21): 9490–9495, 2010.

[23] A. I. Bush, W. H. Pettingell, G. Multhaup, M. Paradis, J. P. Vonsattel, J. F. Gusella, K. Beyreuther, C. L. Masters, and R. E. Tanzi. *Science*,

265(5177):1464, 1994.

[24] K. Hensley, J. M. Carney, M. P. Mattson, M. Aksenova, M. Harris, J. F. Wu, R. A. Floyd, and D. A. Butterfield. *Proc. Natl. Acad. Sci. USA*, 91 (8):3270–3274, 1994.

[25] M. M. Pallitto, J. Ghanta, P. Heinzelman, L. L. Kiessling, and R. M. Murphy. *Biochem.*, 38(12):3570–3578, 1999.

[26] Y. Miller, B. Ma, and R. Nussinov. *Biophys. J.*, 97(4):1168–1177, 2009.

[27] B. Caughey and P. T. Jr. Lansbury. *Annu. Rev. Neurosci.*, 26(1):267–298, 2003.

[28] I. A. Mastrangelo, M. Ahmed, T. Sato, W. Liu, C. Wang, P. Hough, and S. O. Smith. *J. Mol. Biol.*, 358(1):106–119, 2006.

[29] A. T. Petkova, W. M. Yau, and R. Tycko. *Biochemistry*, 45(2):498–512, 2006.

[30] R. Tycko. *Annu. Rev. Phys. Chem.*, 62:279–299, 2011.

[31] J. X. Lu, W. Qiang, W. M. Yau, C. D. Schwieters, S. C. Meredith, and R. Tycko. *Cell*, 154(6):1257–1268, 2014.

[32] B. H. Toyama and J. S. Weissman. *Annu. Rev. Biochem.*, 80(1):557–585, 2011.

[33] M. T. Colvin, R. Silvers, Q. Z. Ni, T. V. Can, I. Sergeyev, M. Rosay, K. J Donovan, B. Michael, J. Wall, S. Linse, and R. G. Griffin. *J. Am. Chem. Soc.*, 138(30):9663–9674, 2016.

[34] M. A. Wälti, F. Ravotti, H. Arai, C. G. Glabe, J. S. Wall, A. Böckmann, and Güntert, P. and Meier, B. H. and Riek, R. *Proc. Natl. Acad. Sci. USA*, 113(34):E4976–E4984, 2016.

[35] A. G. Vlassenko, T. L. S. Benzinger, and J. C. Morris. *BBA Mol. Basis Dis.*, 1822(3):370–379, 2012.

[36] M. D. Ikonomovic, W. E. Klunk, E. E. Abrahamson, C. A. Mathis, J. C. Price, N. D. Tsopelas, B. J. Lopresti, S. Ziolko, W. Bi, W. R. Paljug, M. L. Debnath, C. E. Hope, B. A. Isanski, R. L. Hamilton, and S. T. DeKosky. *Brain*, 131(6):1630–1645, 2008.

[37] W. E. Klunk, H. Engler, A. Nordberg, Y. Wang, G. Blomqvist, D. P. Holt, M. Bergström, I. Savitcheva, G. F. Huang, S. Estrada, B. Ausén, M. L. Debnath, J. Barletta, J. C. Price, J. Sandell, B. J. Lopresti, A. Wall, P. Koivisto, G. Antoni, C. A. Mathis, and B. Långström. *Ann. Neurol.*, 55(3):306–319, 2004.

[38] W. E. Klunk and C. A. Mathis. *Curr. Opin. Neurol.*, 21(6):683–687, 2008.

[39] J.M. Ollinger and J.A. Fessler. *IEEE Signal Processing Mag.*, 14(1):43–55, 1997.

[40] W. Jagust, R. Thisted, M.D. Devous, R. Van Heertum, H. Mayberg, K. Jobst, A.D. Smith, and N. Borys. *Neurology*, 56(7):950–956, 2001.

[41] C. A. Mathis, Y. Wang, D. P. Holt, G. F. Huang, Manik L. D., and W. E. Klunk. *J. Med. Chem.*, 46(13):2740–2754, 2003.

[42] C. Wu, M. T. Bowers, and J. E. Shea. *Biophys. J.*, 100(5):1316–1324, 2011.

[43] G. W. Small, S. Y. Bookheimer, P. M. Thompson, G. M. Cole, S.C. Huang, V. Kepe, and J. R. Barrio. *Lancet Neurol.*, 7(2):161–172, 2008.

[44] M. J. de Leon, A. Convit, O. T. Wolf, C. Y. Tarshish, S. DeSanti, H. Rusinek, W. Tsui, E. Kandil, A. J. Scherer, A. Roche, A. Imossi, E. Thorn, M. Bobinski, C. Caraos, P. Lesbre, D. Schlyer, J. Poirier, B. Reisberg, and J. Fowler. *Proc. Natl. Acad. Sci. USA*, 98(19):10966–10971, 2001.

[45] M. Staderini, M. A. Martin, M. L. Bolognesi, and J. C. Menendez. *Chem. Soc. Rev.*, 44(7):1807–1819, 2015.

[46] E. E. Nesterov, J. Skoch, B. T. Hyman, W. E. Klunk, B. J. Bacskai, and T. M. Swager. *Angew. Chem. Int. Ed.*, 44(34):5452–5456, 2005.

[47] V. Ntziachristos. *Annu. Rev.*, 8:1–33, 2006.

[48] A. Becker, C. Hessenius, K. Licha, B. Ebert, U. Sukowski, W. Semmler, B. Wiedenmann, and C. Grotzinger. *Nat. Biotech.*, 19(4):327–331, 2001.

[49] G. S. Kaminski Schierle, S. van de Linde, M. Erdelyi, E. K. Esbjörner, T. Klein, E. Rees, C. W. Bertoncini, C. M. Dobson, M. S auer, and C. F. Kaminski. *J. Am. Chem. Soc.*, 133(33):12902–12905, 2011.

[50] M. Heilemann, E. Margeat, R. Kasper, M. Sauer, and P. Tinnefel d. *J. Am. Chem. Soc.*, 127(11):3801–3806, 2005.

[51] M. Heilemann, S. van de Linde, M. Schüttpelz, B. Kasper, R. a nd Seefeldt, A. Mukherjee, P. Tinnefeld, and M. Sauer. *Angew. Chem. Int. Ed.*, 47(33):6172–6176, 2008.

[52] M. Monici. *Biotechnol. Annu. Rev.*, 11:227–256, 2005.

[53] H. E. de Vries, J. Kuiper, A. G. de Boer, T. J. C. Van Berkel, and D. D. Breimer. *Pharmacol. Rev.*, 49(2):143–156, 1997.

[54] J. Raymond, S. B.and Skoch, I. D. Hills, E. E. Nesterov, T. M. Swager, and B. J. Bacskai. *Eur. J. Nucl. Med. Mol. Imaging*, 35(1):93–98, 2008.

[55] C. Ran, X. Xu, S. B. Raymond, B. J. Ferrara, K. Neal, B. J. Bacskai, Z. Medarova, and A. Moore. *J. Am. Chem. Soc.*, 131(42):15257–15261, 2009.

[56] X. Zhang, Y. Tian, Z. Li, X. Tian, H. Sun, H. Liu, A. Moore, and C. Ran. *J. Am. Chem. Soc.*, 135(44):16397–16409, 2013.

[57] X. Zhang, Y. Tian, P. Yuan, Y. Li, M. A. Yaseen, J. Grutzendler, A. Moore, and C. Ran. *Chem. Comm.*, 50(78):11550–11553, 2014.

[58] M. Cui, M. Ono, H. Watanabe, H. Kimura, B. Liu, and H. Saji. *J. Am. Chem. Soc.*, 136(9):3388–3394, 2014.

[59] V. I. Stsiapura, A. A. Maskevich, V. A. Kuzmitsky, V. N. Uversky, I. M. Kuznetsova, and K. K. Turoverov. *J. Phys. Chem. B*, 112(49): 15893–15902, 2008.

[60] N. Amdursky, Y. Erez, and D. Huppert. *Acc. Chem. Res.*, 45(9):1548–1557, 2012.

[61] R. Sabate, L. Rodriguez-Santiago, M. Sodupe, S. J. Saupe, and S. Ventura. *Chem. Comm.*, 49(51):5745–7, 2013.

[62] M. Biancalana and S. Koide. *BBA-Proteins. Proteom.*, 1804(7):1405–1412, 2010.

[63] H. Puchtler, F. Sweat, and M. Levine. *J. Histochem. Cytochem.*, 10(3): 355–364, 1962.

[64] C. Wu, J. Scott, and J. E. Shea. *Biophys. J.*, 103(3):550–557, 2012.

[65] R Khurana, V. N. Uversky, L. Nielsen, and A. L. Fink. *J. Biol. Chem.*, 276(25):22715–22721, 2001.

[66] P. Frid, S. V. Anisimov, and N. Popovic. *Brain Res. Rev.*, 53(1):135–160, 2007.

[67] M. L. Schmidt, T. Schuck, S. Sheridan, M. P. Kung, H. Kung, Z. P Zhuang, C. Bergeron, J. S. Lama rche, D. Skovronsky, B. I. Giasson, V. M. Y. Lee, and J. Q. Trojanowski. *Am. J. Pathol.*, 159(3):937–943, 2001.

[68] D. M. Skovronsky, B. Zhang, M. P. Kung, H. F. Kung, J. Q. Trojanowski, and Vs. M. Y. Lee. *Proc. Natl. Acad. Sci. USA*, 97(13):7609–

7614, 2000.

[69] M. Grundman and P. Delaney. *Proc. Nutr. Soc.*, 61:191–202, 2002.

[70] G. M. Cole, B. Teter, and S. A. Frautschy. Neuroprotective effects of curcumin. In *The molecular targets and therapeutic uses of curcumin in health and disease*. Springer US, 2007.

[71] K. Ono, K. Hasegawa, H. Naiki, and M. Yamada. *J. Neurosci. Res.*, 75 (6):742–750, 2004.

[72] C. F. Chignell, P. Bilskj, K. J. Reszka, A. G. Motten, R. H. Sik, and T. A. Dahl. *Photochem. Photobiol.*, 59(3):295–302, 1994.

[73] X. Zhang, Y. Tian, C. Zhang, X. Tian, A. W. Ross, R. D Moir, H. Sun, R. E. Tanzi, A. Moore, and C. Ran. *Proc. Natl. Acad. Sci. USA*, 112(31): 9734–9739, 2015.

[74] M. Hintersteiner, A. Enz, P. Frey, A. L Jaton, W. Kinzy, R. Kneuer, U. Neumann, M. Rudin, M. Staufenbiel, M. Stoeckli, K. H. Wiederhold, and H. U. Gremlich. *Nat. Biotech.*, 23(5):577–583, 2005.

[75] H. Chen, S. S. Ahsan, M. B. Santiago-Berrios, H. D. Abruña, and W. W. Webb. *J. Am. Chem. Soc.*, 132(21):7244–7245, 2010.

[76] Z. F. Mainen, M. Maletic-Savatic, S. H. Shi, Y. Hayashi, R. Malinow, and K. Svoboda. *Methods*, 18(2):231–239, 1999.

[77] F. Helmchen. Two-photon functional imaging of neuronal activity. In *In vivo optical imaging of brain function, 2nd edition*. CRC Press/Taylor & Francis, 2009.

[78] R. Gómez-Balderas, D. F. Raffa, G. A. Rickard, P. Brunelleand, and A. Rauk. *J. Phys. Chem. A*, 109(24):5498–5508, 2005.

[79] I. A. Karpenko, Y. Niko, V. P. Yakubovskyi, A. O. Gerasov, D. Bonnet, Y. P. Kovtun, and A. S. Klymchenko. *J. Mater. Chem. C*, 4(14):3002–

3009, 2016.

[80] A. Loudet and K. Burgess. *Chem. Rev.*, 107(11):4891–4932, 2007.

[81] M. Ono, H. Watanabe, H. Kimura, and H. Saji. *ACS Chem. Neurosci.*, 3(4):319–324, 2012.

[82] R. Adhikary, P. Mukherjee, T. W. Kee, and J. W. Petrich. *J. Phys. Chem. B*, 113(15):5255–5261, 2009.

[83] A. Treibs and F. H. Kreuzer. *Justus Liebigs Ann Chem.*, 718(1):208–223, 1968.

[84] S. Erten-Ela, M. D. Yilmaz, B. Icli, Y. Dede, S. Icli, and E. U. Akkaya. *Org. Lett.*, 10(15):3299–3302, 2008.

[85] Y. Zhong, L. Si, H. He, and A. G. Sykes. *Dalton Trans.*, 40(43):11389–11395, 2011.

[86] Y. Ni and J. Wu. *Org. Biomol. Chem.*, 12(23):3774–3791, 2014.

[87] H. L. Kee, C. Kirmaier, L. Yu, P. Thamyongkit, W. J. Youngblood, M. E. Calder, L. Ramos, B. C. Noll, D. F. Bocian, W. Robert Scheidt, Robert R. Birge, J. S. Lindsey, and D. Holten. *J. Phys Chem. B*, 109(43): 20433–20443, 2005.

[88] S. Chibani, B. Le Guennic, A. Charaf-Eddin, A. D. Laurent, and D. Jacquemin. *Chem. Sci.*, 4(5):1950–1963, 2013.

[89] S. T. Meek, E. E. Nesterov, and T. M. Swager. *Org. Lett.*, 10(14):2991–2993, 2008.

[90] S. Bae, E. Lim, D. Hwang, H. Huh, and S. K. Kim. *Chem. Phys. Lett.*, 633:109–113, 2015.

[91] F. Peccati, J. Hernando, L. Blancafort, X. Solans-Monfort, and M. Sodupe. *Phys. Chem. Chem. Phys.*, 17(30):19718–19725, 2015.

[92] F. Peccati, X. Solans-Monfort, and M. Sodupe. *Theor. Chem. Acc.*, 135 (8):184, 2016.

[93] H. Fu, M. Cui, L. Zhao, P. Tu, K. Zhou, J. Dai, and B. Liu. *J. Am. Chem. Soc.*, 58(17):6972–6983, 2015.

[94] F. Jensen. *Introduction to Computational Chemistry*. Dover, 2013.

[95] F. L. Pilar. *Elementary Quantum Chemistry*. Wiley, 1990.

[96] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox. Gaussian09 Revision D.01. Gaussian Inc. Wallingford CT 2013.

[97] R. Ditchfield, W. J. Hehre, and J.A. Pople. *J. Chem. Phys.*, 54:724–728, 1971.

[98] J. VandeVondele, M. Krack, F. Mohamed, M. Parrinello, T. Chassaing, and J. Hutter. *Comput. Phys. Commun.*, 167(2):103–128, 2005.

[99] J. VandeVondele and J. Hutter. *J. Chem. Phys.*, 127(11):114105, 2007.

[100] S. Goedecker, M. Teter, and J. Hutter. *Phys. Rev. B*, 54(3):1703–1710, 1996.

[101] M. Krack. *Theor. Chem. Acc.*, 114(1):145–152, 2005.

[102] F. Aquilante, J. Autschbach, R. K. Carlson, L. F. Chibotaru, M. G. Delcey, L. De Vico, I. Fdez. Galván, N. Ferré, L. M. Frutos, L. Gagliardi, M. Garavelli, A. Giussani, C. E. Hoyer, G. Li Manni, H. Lischka, D. Ma, P. A. Malmqvist, T. Müller, A. Nenov, M. Olivucci, T. B. Pedersen, D. Peng, F. Plasser, B. Pritchard, M. Reiher, I. Rivalta, I. Schapiro, J. Segarra-Martí, M. Stenrup, D. G. Truhlar, L. Ungur, A. Valentini, S. Vancoillie, V. Veryazov, V. P. Vysotskiy, O. Weingart, F. Zapata, and R. Lindh. *J. Comput. Chem.*, 37(5):506–541, 2016.

[103] P. O. Widmark, P. A. Malmqvist, and B. O. Roos. *Theor. Chim. Acta*, 77(5):291–306, 1990.

[104] P. O. Widmark, B. J. Persson, and B. O. Roos. *Theor. Chim. Acta*, 79 (6):419–432, 1991.

[105] B. O. Roos, R. Lindh, P. A. Malmqvist, V. Veryazov, and P.O. Widmark. *J. Phys. Chem. A*, 108(15):2851–2858, 2004.

[106] K. Andersson, P. A. Malmqvist, B. O. Roos, A. J. Sadlej, and K. Wolinski. *J. Phys. Chem.*, 94(14):5483–5488, 1990.

[107] K. Andersson, P. A. Malmqvist, and B. O. Roos. *J. Phys. Chem.*, 96(2): 1218–1226, 1992.

[108] G. Ghigo, B. Roos, and P. A. Malmqvist. *Chem. Phys. Lett.*, 396(1-3): 142–149, 2004.

[109] N. Forsberg and P. A. Malmqvist. *Chem. Phys. Lett.*, 274(1-3):196–204, 1997.

[110] J. Finley, P. A. Malmqvist, B. O. Roos, and L. Serrano-Andrés. *Chem. Phys. Lett.*, 288:299–306, 1998.

[111] P. Hohenberg and W. Kohn. *Phys. Rev. B*, 136:B864–B871, 1964.

[112] W. Kohn and L. J. Sham. *Phys. Rev. A*, 140:1133–1138, 1965.

[113] I. Ciofini, C. Adamo, and H. Chermette. *Chem. Phys.*, 309(1):67–76, 2005.

[114] J. P. Perdew and K. Schmidt. *AIP Conference Proceedings*, 577(1):1–20, 2001.

[115] A. D. Becke. *J. Chem. Phys.*, 98(7):5648–5652, 1993.

[116] Ch. Lee, W. Yang, and R. G. Parr. *Phys. Rev. B*, 37:785–789, 1988.

[117] S. Grimme. *J. Comput. Chem.*, 27(15):1787–1799, 2006.

[118] S. Grimme, J. Antony, S. Ehrlich, and H. Krieg. *J. Chem. Phys.*, 132 (15):154104, 2010.

[119] E. Runge and E. K. U. Gross. *Phys. Rev. Lett.*, 52:997–1000, 1984.

[120] M. A. A. Marques, Maitra N. T., Nogueira F. M. S., Gross E. K. U., and A. Rubio. *Fundamentals of Time-Dependent Density Functional Theory*. Springer, 2008.

[121] M. E. Casida. Time-dependent density functional response theory of molecular systems: Theory, computational methods, and functionals. In *Recent Developments and Applications of Modern Density Functional Theory*. Elsevier, 1996.

[122] R. E. Stratmann, G. E. Scuseria, and M. J. Frisch. *J. Chem. Phys.*, 109 (19):8218–8224, 1998.

[123] T. Yanai, D. P. Tew, and N. C. Handy. *Chem. Phys. Lett.*, 393(1–3): 51–57, 2004.

[124] M. J. G. Peach, M. J. Williamson, and D. J. Tozer. *J. Chem. Theory Comput.*, 7(11):3578–3585, 2011.

[125] J. Tomasi, B. Mennucci, and R. Cammi. *Chem. Rev.*, 105(8):2999–3093, 2005.

[126] S. Miertuš, E. Scrocco, and J. Tomasi. *Chem. Phys.*, 55(1):117–129, 1981.

[127] G. Scalmani, M. J. Frisch, B. Mennucci, J. Tomasi, R. Cammi, and V. Barone. *J. Chem. Phys.*, 124:094107, 2006.

[128] A. V. Marenich, C. J. Cramer, and D. G. Truhlar. *J. Phys. Chem. B*, 113 (18):6378–6396, 2009.

[129] D. Mark and J. Hutter. Ab initio molecular dynamics: theory and tmplementation. In *Modern Methods and Algorithms of Quantum Chemistry,*. NIC series, John von Neumann Institute for Computing, 2000.

[130] T.L. Hill. *An introduction to statistical tehrmodynamics*. Dover, 1986.

[131] M. Iannuzzi. Molecular simulations: lecture notes. 2014.

[132] Amber 14 reference manual. 2014.

[133] M. Richter, P. Marquetand, J. González-Vázquez, I. Sola, and L. González. *J. Chem. Theory Comput.*, 7(5):1253–1258, 2011.

[134] S. Mai, P. Marquetand, and L. González. *Int. J. Quantum Chem.*, 115 (18):1195–1272, 2015.

[135] H. Lischka, R. Shepard, R. M. Pitzer, I. Shavitt, M. Dallos, Müller., P. G. Szalay, M. Seth, G. S. Kedziora, S. Yabushita, and Z. Zhang. *Phys. Chem. Chem. Phys.*, 3(5):664–673, 2001.

[136] A. C. T. van Duin, S. Dasgupta, F. Lorant, and W. A. Goddard III. *J. Phys. Chem. A*, 105(41):9396–9409, 2001.

[137] T. P. Senftle, S. Hong, M. M. Islam, S. B. Kylasa, Y. Zheng, Y. K. Shin, C. Junkermeier, R. Engel-Herbert, M. J. Janik, H. M. Aktulga, T. Verstraelen, A. Grama, and A. C. T. van Duin. *npj Comput. Mater.*, 2: 15011, 2016.

[138] S. Genheden and U. Ryde. *Expert Opin. Drug Discov.*, 10(5):449–461, 2015.

[139] J. Srinivasan, T. E. Cheatham III, P. Cieplak, P. A. Kollman, and D. A. Case. *J. Am. Chem. Soc.*, 120(37):9401–9409, 1998.

[140] D. A. Case, V. Babin, J. T. Berryman, R. M. Betz, Q. Cai, D. S. Cerutti, T. E. Cheatham III, T. A. Darden, R. E. Duke, H. Gohlke, A. W. Goetz, S. Gusarov, N. Homeyer, P. Janowski, J. Kaus, I. Kolossváry, A. Kovalenko, T. S. Lee, S. LeGrand, T. Luchko, R. Luo, B. Madej, K. M. Merz, F. Paesani, D. R. Roe, A. Roitberg, C. Sagui, R. Salomon-Ferrer, G. Seabra, C. L. Simmerling, W. Smith, J. Swails, R. C. Walker, J. Wang, R. M. Wolf, X. Wu, and P. A. Kollman. Amber 14. 2014.

[141] R. Salomon-Ferrer, A.W. Goetz, D. Poole, S. Le Grand, and R.C. Walker. *J. Chem. Theory Comput.*, 9(9):3878–3888, 2013.

[142] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling. *Proteins: Struct., Funct., Bioinf.*, 65(3):712–725, 2006.

[143] K. W. Borrelli, A. Vitalis, R. Alcantara, and V. Guallar. *J. Chem. Theory Comput.*, 1(6):1304–1311, 2005.

[144] D. J. Wales. *Energy landscapes: applications to clusters, biomolecules and glasses*. Cambridge University Press, 2003.

[145] B. Brooks and M. Karplus. *Proc. Natl. Acad. Sci. USA*, 80(21): 6571–6575, 1983.

[146] I. Bahar, T. R. Lezon, A. Bakan, and I. H. Shrivastava. *Chem. Rev.*, 110

(3):1463–1497, 2010.

[147] M. M. Tirion. *Phys. Rev. Lett.*, 77:1905–1908, 1996.

[148] K. Hinsen. *Proteins: Struct. Funct. and Bioinf.*, 33(3):417–429, 1998.

[149] T. R. Lezon, J. R. Banavar, A. M. Lesk, and A. Maritan. *Proteins: Struct. Funct. and Bioinf.*, 63(2):273–277, 2006.

[150] J. W. Ponder and F. M. Richards. Tertiary templates for proteins. *J. Mol. Biol.*, 193(4):775–791, 1987.

[151] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne. *Nucleic Acids Res.*, 28(1):235–242, 2000.

[152] S. Liang and N. V. Grishin. *Protein Sci.*, 11(2):322–331, 2002.

[153] S. C. Lovell, J. M. Word, J. S Richardson, and D. C. Richardson. *Proteins: Struct. Funct. and Bioinf.*, 40(3):389–408, 2000.

[154] A. Madadkar-Sobhani and V. Guallar. Pele web server: atomistic study of biomolecular systems at your fingertips. 2013.

[155] K. W. Borrelli, B. Cossins, and V. Guallar. *J. Comput. Chem.*, 31(6): 1224–1235, 2010.

[156] J. L. Banks, H. S. Beard, Y. Cao, A. E. Cho, W. Damm, R. Farid, A. K. Felts, T. A. Halgren, D. T. Mainz, J. R. Maple, R. Murphy, D. M. Philipp, M. P. Repasky, L. Y. Zhang, B. J. Berne, R. A. Friesner, E. Gallicchio, and R. M. Levy. *J. Comput. Chem.*, 26(16):1752–1780, 2005.

[157] D. Qiu, P. S. Shenkin, F. P. Hollinger, and W. C. Still. *J. Phys. Chem. A*, 101(16):3005–3014, 1997.

[158] D. E. Clark and S. D. Pickett. *Drug Discov. Today*, 5(2):49–58, 2000.

[159] P. Ertl, B. Rohde, and P. Selzer. *J. Med. Chem.*, 43(20):3714–3717, 2000.

[160] F. Peccati, S. Pantaleone, X. Solans-Monfort, and M. Sodupe. submitted. *Isr. J. Chem*, 2016.

[161] E. Sanchez-Garcia, M. Doerr, and W. Thiel. *J. Comput. Chem.*, 31(8): 1603–1612, 2010.

[162] P. Imhof. *J. Chem. Theory Comput.*, 8(11):4828–4836, 2012.

[163] C. M. Isborn, A. W. Götz, M. A. Clark, R. C. Walker, and T. J. Martínez. *J. Chem. Theory Comput.*, 8(12):5092–5106, 2012.

[164] J. P. Perdew, K. Burke, and M. Ernzerhof. *Phys. Rev. Lett.*, 77(18): 3865–3868, 1996.

[165] T. Le Bahers, C. Adamo, and I. Ciofini. *J. Chem. Theory Comput.*, 7(8): 2498–2506, 2011.

[166] I. Ciofini, T. Le Bahers, C. Adamo, F. Odobel, and D. Jacquemin. *J. Phys. Chem. C*, 116(22):11946–11955, 2012.

[167] D. Beljonne, Z. Shuai, G. Pourtois, and J. L. Bredas. *J. Phys. Chem. A*, 105(15):3899–3907, 2001.

[168] O. Valsson and C. Filippi. *J. Chem. Theory Comput.*, 6(4):1275–1292, 2010.

[169] M. Huix-Rotllant, M. Filatov, S. Gozem, I. Schapiro, M. Olivucci, and N. Ferré. *J. Chem. Theory Comput.*, 9(9):3917–3932, 2013.

[170] R. S. Sánchez, R. Gras-Charles, J. L. Bourdelande, G. Guirado, and J. Hernando. *J. Phys. Chem. C*, 116(12):7164–7172, 2012.

[171] J. A. Levitt, M. K. Kuimova, G. Yahioglu, P. H. Chung, K. Suhling, and D. Phillips. *J. Phys. Chem. C*, 113(27):11634–11642, 2009.

[172] M. Kasha, H. R. Rawls, and M. A. El-Bayoumi. *Pure Appl. Chem.*, 11 (3-4):371–392, 1965.

[173] F. Würthner, S. Yao, T. Debaerdemaeker, and R. Wortmann. *J. Am. Chem. Soc.*, 124(32):9431–9447, 2002.

[174] J. Hernando, M. van der Schaaf, E. M. H. P. van Dijk, M. Sauer, M. F. García-Parajó, and N. F. van Hulst. *J. Phys. Chem. A*, 107(1):43–52, 2003.

[175] F. Peccati, M. Wiśniewska, X. Solans-Monfort, and M. Sodupe. *Phys. Chem. Chem. Phys.*, 18(17):11634–11643, 2016.

[176] J. Kruszewski and T.M. Krygowski. *Tetrahedron Lett.*, 13(36):3839–3842, 1972.

[177] J. C. Dobrowolski and S. Ostrowski. *RSC Adv.*, 5(13):9467–9471, 2015.

[178] M. Rosenberg, C. Dahlstrand, K. Kilså, and H. Ottosson. *Chem. Rev.*, 114(10):5379–5425, 2014.

[179] www.molinspiration.com.

[180] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case. *J. Comput. Chem.*, 25(9):1157–1174, 2004.

[181] S. Genheden and U. Ryde. *J. Comput. Chem.*, 31(4):837–846, 2010.

[182] M. Anderson, O. V. Bocharova, N. Makarava, L. Breydo, V. V. Salnikov, and I. V. Baskakov. *J. Mol. Biol.*, 358(2):580–596, 2006.

[183] V. Ostapchenko, M. Gasset, and Ilia V. Baskakov. pages 157–167. Humana Press, 2012.

[184] S. Genheden, P. Mikulskis, L. Hu, J. Kongsted, P. Söderhje lm, and U. Ryde. *J. Am. Chem. Soc.*, 133(33):13081–13092, 2011.

[185] M. Richter, S. Mai, P. Marquetand, and L. González. *Phys. Chem. Chem. Phys.*, 16(44):24423–36, 2014.

[186] C. E. Crespo-Hernández, L. Martínez-Fernández, C. Rauer, C. Re-

ichardt, S. Mai, M. Pollum, P. Marquetand, L. González, and Inés Corral. *J. Am. Chem. Soc.*, 137(13):4368–4381, 2015.

[187] S. Mai, P. Marquetand, and L. González. *J. Phys. Chem. A*, 119(36): 9524–9533, 2015.

[188] S. Mai, P. Marquetand, M. Richter, J. González-Vázquez, and L. González. *Chem. Phys. Chem.*, 14(13):2920–2931, 2013.

[189] A Malyshev and F.E. Romsberg. *Angew. Chem. Int. Ed.*, 54(41):11930–11944, 2015.

[190] J. M. L. Pecourt, J. Peon, and B. Kohler. *J. Am. Chem. Soc.*, 123(21): 5166–5166, 2001.

[191] C. T. Middleton, K. de La Harpe, C. Su, Y. K. Law, C. E. Crespo-Hernández, and B. Kohler. *Annu. Rev. Phys. Chem.*, 60(1):217–239, 2009.

[192] N. Ismail, L. Blancafort, M. Olivucci, B. Kohler, and M. Robb. *J. Am. Chem. Soc.*, 124(24):6818–6819, 2002.

[193] S. Matsika. *J. Phys. Chem. A*, 108(37):7584–7590, 2004.

[194] P. M. Hare, C. E. Crespo-Hernández, and B. Kohler. *J. Phys. Chem. B*, 110(37):18641–18650, 2006.

[195] P. M. Hare, C. E Crespo-Hernández, and B. Kohler. *Proc. Natl. Acad. Sci. USA*, 104(2):435–440, 2007.

[196] P. Karran and N. Attard. *Nat. Rev. Cancer*, 8(1):24–36, 2008.

[197] W. Kwok, C. Ma, and D. L. Phillips. *J. Am. Chem. Soc.*, (15):5131–5139, 2008.

[198] M. Pollum, S. Jockusch, and C. E. Crespo-Hernández. *J. Am. Chem. Soc.*, 136(52):17930–3, 2014.

[199] L. Martínez-Fernández, L. González, and I. Corral. *Chem. Comm.*, 48 (15):2134, 2012.

[200] L. Martínez-Fernández, I. Corral, G. Granucci, and M. Persico. *Chem. Sci.*, 5(4):1336, 2014.

[201] A. Favre, C. Saintomé, J. L. Fourrey, P. Clivio, and P. Laugâa. *J. Photochem. Photobiol. B*, 42(2):109–124, 1998.

[202] Y. Matsumoto, V. Rodriguez, T. A Whitford, N. Beeharry, H. Ide, and A. E. Tomkinson. *Oncoscience*, 2(3), 2015.

[203] G. Miolo, C. Marzano, V. Gandin, A. C. Palozzo, D. Dalzoppo, A. Salvador, and S. Caffieri. *Chem. Res. Toxicol.*, 24(8):1319–1326, 2011.

[204] S. Yamazaki and T. Taketsugu. *J. Phys. Chem. A*, 116(1):491–503, 2012.

[205] F. Santoro, V. Barone, T. Gustavsson, and R. Improta. *J. Am. Chem. Soc.*, 128(50):16312–16322, 2006.

[206] S. Benzer and E. Freeze. *Proc. Natl. Acad. Sci. USA*, 44(2):112–119, 1958.

[207] M. Orozco. *J. Phys. Chem. B*, 5647(98):5228–5233, 1998.

[208] V. I. Danilov, T. van Mourik, N. Kurita, H. Wakabayashi, T. Tsukamoto, and D. M. Hovorun. *J. Phys. Chem. A*, 113(11):2233–2235, 2009.

[209] J.M. Campbell and S. F. Von Sonntag. *Z. Naturforsch.*, 29b:750–757, 1974.

[210] H. Sugiyama, Y. Tsutsumi, and I. Saito. *J. Am. Chem. Soc.*, 112(18): 6720–6721, 1990.

[211] Y. Zeng and Y. Wang. *J. Am. Chem. Soc.*, 126(21):6552–6553, 2004.

[212] T. Chen, G. P. Cook, A. T. Koppisch, and M. M. Greenberg. *J. Am. Chem. Soc.*, 122(16):3861–3866, 2000.

[213] T. M. Dietz, R. J. Von Trebra, B. J. Swanson, and T. H. Koch. *J. Am. Chem. Soc.*, 109(6):1793–1797, 1987.

[214] M. Dextraze, J. R. Wagner, and D. J Hunting. *Biochemistry*, 46:9089–9097, 2007.

[215] J. Tully. *J. Chem. Phys.*, 55(2):562, 1971.

[216] M. Kobyłecka, A. Migani, D. Asturiol, J. Rak, and L. Blancafort. *J. Phys. Chem. A*, 113(19):5489–5495, 2009.

[217] R. Abouaf, J. Pommier, and H. Dunet. *Chem. Phys. Lett.*, 381(3–4): 486–494, 2003.

[218] T. H. Dunning Jr. *J. Chem. Phys.*, 90:1007, 1989.

[219] A. K. Wilson, D. E. Woon, K. A. Peterson, and T. H. Dunning Jr. *J. Chem. Phys.*, 110:7667, 1999.

[220] H. Lischka, Müller, P. G. Szalay, I. Shavitt, R. M. Pitzer, and R. Shepard. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 1(2):191–199, 2011.

[221] J. P. Dahl and M. Springborg. *J. Chem. Phys.*, 88(7):4335, 1988.

[222] M. Barbatti, G. Granucci, M. Persico, M. Ruckenbauer, M. Vazdar, M. Eckert-Maksić, and H. Lischka. *J. Photochem. Photobiol. A*, 190 (2-3):228–240, 2007.

[223] M. Kasha. *J. Chem. Phys.*, 20(1):71–74, 1952.

[224] B. J. Swanson, J. C. Kutzer, and T. H. Koch. *J. Am. Chem. Soc.*, 103(5): 1274–1276, 1981.

[225] F. Peccati, S. Mai, and L.González. submitted. *Phil. Trans. R. Soc. A*, 2016.

[226] C. R. Wang, A. Hu, and Q. B. Lu. *J. Chem. Phys.*, 124(24):241102, 2006.

Appendices

Appendix A

Computational methods

Table A.1: TDDFT computed excitation and emission wavelengths (nm) of NIAD-4 in water, methanol, dichloromethane and diethylether computed with the PCM model. Linear response (LR), state specific (SS), equilibrium (EQ) and non-equilibrium (NEQ) approaches.

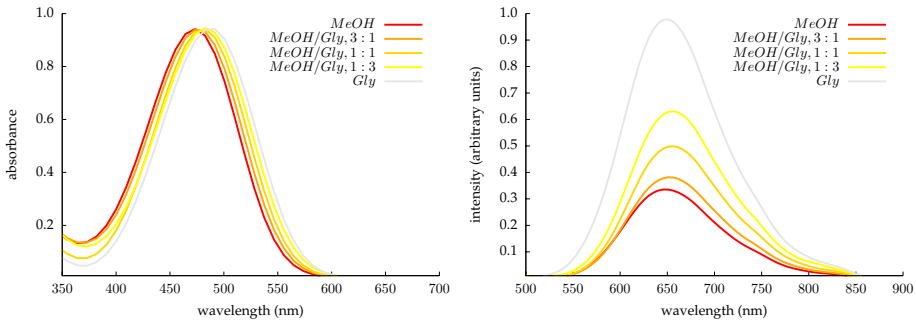
excitation wavelength	H_2O	$MeOH$	CH_2Cl_2	Et_2O
LR EQ	470	468	461	451
LR NEQ	439	438	440	435
SS NEQ	447	446	449	443
SS EQ	507	503	487	467
emission wavelength	H_2O	$MeOH$	CH_2Cl_2	Et_2O
LR EQ	625	620	598	572
LR NEQ	564	562	561	547
SS NEQ	554	553	534	535
SS EQ	549	548	542	532

Appendix B

Bithiophene derivatives

Table B.1: Gas phase dihedral angles (°) and rotational barriers $kJ\,mol^{-1}$ of three conformational isomers of NIAD-4 at the DFT(CAM-B3LYP) optimized geometry. Labels refer to Figure 3.1. Values in parenthesis correspond to geometries optimized using the PCM solvation model.

Isomer	**A**	**B**	**C**
α	33 (28)	33 (29)	33 (28)
β	-14 (-11)	-18 (-13)	33 (28)
γ	1 (0)	-179 (-179)	1 (0)
Rotational barriers		**A** $\rightarrow$ **B**	A $\rightarrow$ **C**
$\Delta E^{\ddagger}$		37.5 (37.6)	14.0 (15.4)
$\Delta G^{\ddagger}$		38.6 (39.5)	19.2 (21.2)



(a) Absorption spectra.

(b) Fluorescence spectra.

Figure B.1: Absorption and fluorescence emission spectra of NIAD-4 in mixtures of methanol and glycerol at increasing viscosity.

Table B.2: Relaxed S_1 energy scan around dihedral β of NIAD-4; excitation energies (eV) to S_1 and T_2 computed in gas phase and several solvents (PCM model). Values are corrected for the difference between MSCAPT2 and TDFFT energies at the S_1 optimized geometry. Labels refer to Figure 3.1.

β (°)	$S_1(H_2O)$	$T_2(H_2O)$	$S_1(MeOH)$	$T_2(MeOH)$	$S_1(CH_2Cl_2)$
192	2.14	3.08	2.15	3.08	2.22
172	2.13	3.08	2.15	3.08	2.22
152	2.18	3.06	2.19	3.06	2.26
132	2.27	3.01	2.29	3.00	2.35
112	2.41	2.94	2.43	2.94	2.49
92	2.60	2.90	2.61	2.90	2.65
72	2.46	2.93	2.48	2.93	2.54
52	2.31	2.99	2.32	2.99	2.39
32	2.19	3.06	2.21	3.06	2.28
12	2.13	3.10	2.15	3.10	2.22
β (°)	$T_2(CH_2Cl_2)$	$S_1(Et_2O)$	$T_2(Et_2O)$	$S_1(gp)$	$T_2(gp)$
192	3.07	2.32	3.06	2.70	3.04
172	3.07	2.31	3.07	2.69	3.04
152	3.05	2.36	3.05	2.73	3.03
132	3.00	2.44	3.00	2.81	2.99
112	2.93	2.57	2.93	2.89	2.94
92	2.90	2.71	2.90	2.92	2.89
72	2.93	2.63	2.93	2.96	2.96
52	2.99	2.49	2.99	2.87	2.99
32	3.05	2.38	3.05	2.77	3.03
12	3.10	2.31	3.09	2.70	3.06

Table B.3: Gas phase CAS(12,12) and MSCASPT excitation energies (eV) at the TDDFT optimized S_1 geometry of NIAD-4, NIAD-11 and NIAD-16.

root	S_1(CASSCF)	S_1(MSCASPT2)	T_2(CASSCF)	T_2(MSCASPT2)
NIAD-4				
1	0.00	0.00	1.24	1.81
2	2.91	2.57	2.57	2.99
3	3.66	3.41	3.79	4.12
4	4.25	3.67	3.97	4.50
5	4.91	4.55	4.37	4.72
6	5.27	4.79	4.88	5.27
7	5.43	5.16	-	-
8	5.66	5.42	-	-
9	5.85	5.85	-	-
10	6.28	6.03	-	-
11	6.48	6.31	-	-
12	6.54	6.65	-	-
NIAD-11				
1	0.00	0.00	0.64	1.72
2	2.77	2.17	2.04	3.00
3	3.48	3.41	3.26	4.15
4	4.24	3.59	3.43	4.41
5	4.60	4.03	3.57	4.61
6	5.16	4.39	4.42	5.32
7	5.20	4.78	-	-
8	5.45	4.91	-	-
9	5.56	5.25	-	-
10	5.97	5.72	-	-
11	6.08	6.09	-	-
12	6.31	6.41	-	-
NIAD-16				
1	0.00	0.00	1.32	1.51
2	2.64	2.50	2.59	2.79
3	3.22	3.05	3.80	3.75
4	3.86	3.14	3.90	3.96
5	4.77	4.32	4.46	4.24
6	5.14	4.41	4.81	4.76
7	5.20	5.02	-	-
8	5.36	5.37	-	-
9	5.56	5.58	-	-
10	5.92	5.92	-	-
11	6.02	6.12	-	-
12	6.18	6.73	-	-

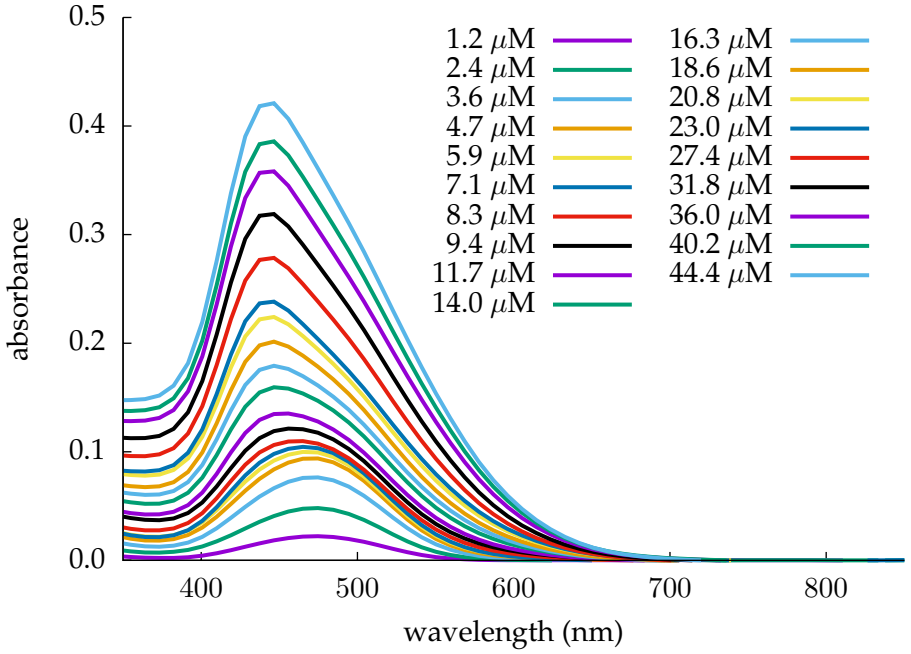


Figure B.3: Concentration dependent absorption spectra of NIAD-4 in a 2:1 mixture of water and methanol.

Table B.4: S_1 and S_2 vertical excitation energies of parallel (non-centrosymmetric), P, and antiparallel (centrosymmetric), AP, π stacking dimers at interchromophoric distances d ranging from 3.6 to 4.6 Å in water (PCM model).

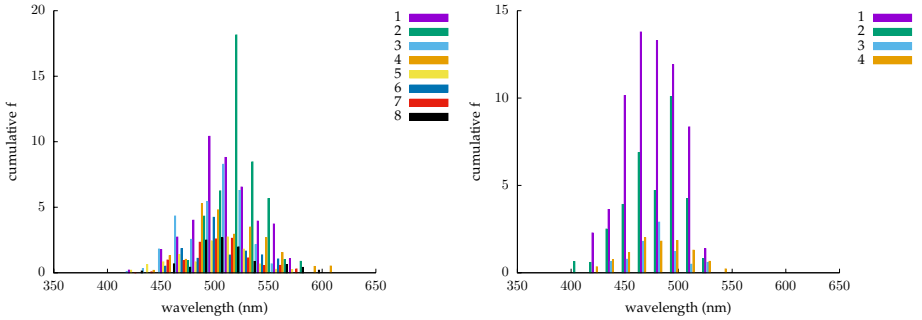
d	3.6	3.8	4.0	4.2	4.4	4.6
AP E(S_1)	505	477	461	451	446	442
AP f(S_1)	0.0	0.0	0.0	0.0	0.0	0.0
AP E(S_1)	449	434	426	423	422	421
AP f(S_1)	1.2	1.6	2.0	2.2	2.3	2.4
P E(S_1)	528	494	471	458	450	442
P f(S_1)	0.0	0.0	0.0	0.0	0.0	0.0
P E(S_1)	417	417	417	417	417	417
P f(S_1)	2.4	2.4	2.4	2.4	2.4	2.4

Table B.5: Dihedral angles α, β and γ, relative electronic (E, $kJ\,mol^{-1}$) energies and vertical excitation to S_1 with corresponding oscillator strength of the 8 low-lying isomers of NIAD-11 and of the 4 low-lying isomers of NIAD-16 in gas phase and water (SMD model) in brackets.

Isomer	α (°)	β (°)	γ (°)	E ($kJ\,mol^{-1}$)
N-11 is1	-52 (-52)	-21 (-21)	-178 (-179)	0.08 (1.33)
N-11 is2	122 (-121)	22 (23)	178 (179)	0.00 (0.72)
N-11 is3	53 (55)	-26 (-19)	2 (4)	0.53 (6.65)
N-11 is4	123 (121)	26 (-21)	-2 (2)	1.07 (5.14)
N-11 is5	52 (-54)	-144 (-147)	179 (179)	2.05 (0.94)
N-11 is6	-51 (-52)	143 (-146)	2 (-3)	2.11 (5.81)
N-11 is7	123 (-121)	-143 (144)	-2 (3)	2.13 (4.73)
N-11 is8	123 (121)	145 (-146)	180 (179)	2.61 (0.00)
N-16 is1	30 (23)	-12 (0)	179 (180)	0.00 (0.00)
N-16 is2	30 (23)	-16 (-1.83)	1 (1)	1.14 (5.41)
N-16 is3	30 (23)	157 (169)	180 (180)	3.89 (1.29)
N-16 is4	31 (25)	156 (170)	1 (1)	3.89 (6.62)
Isomer	λ_{abs} (nm)	f		
N-11 is1	469 (509)	0.95 (1.16)		
N-11 is2	464 (490)	0.98 (1.10)		
N-11 is3	461 (505)	1.07 (1.28)		
N-11 is4	457 (491)	1.07 (1.24)		
N-11 is5	455 (490)	0.84 (1.03)		
N-11 is6	452 (489)	0.96 (1.17)		
N-11 is7	448 (477)	1.00 (1.18)		
N-11 is8	447 (480)	0.85 (1.04)		
N-16 is1	432 (469)	1.29 (1.47)		
N-16 is2	430 (471)	1.34 (1.53)		
N-16 is3	428 (471)	1.09 (1.30)		
N-16 is4	424 (470)	1.30 (1.53)		

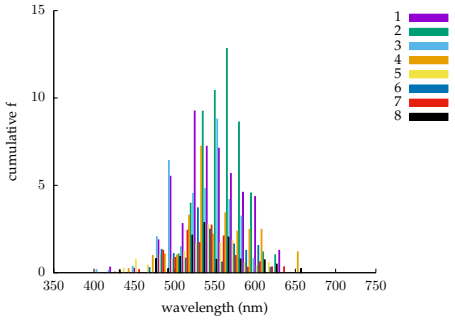
Table B.6: Emission wavelength from S_1 with corresponding oscillator strength of the 8 low-lying isomers of NIAD-11 and of the 4 low-lying isomers of NIAD-16 in gas phase and water (in brackets).

Isomer	λ_{em} (nm)	f
N-11 is1	573 (775)	1.02 (1.58)
N-11 is2	561 (746)	1.02 (1.53)
N-11 is3	570 (775)	1.10 (1.68)
N-11 is4	558 (745)	1.11 (1.66)
N-11 is5	578 (799)	0.87 (1.37)
N-11 is6	571 (803)	1.03 (1.57)
N-11 is7	561 (773)	1.05 (1.56)
N-11 is8	563 (636)	0.89 (1.33)
N-16 is1	513 (751)	1.47 (2.04)
N-16 is2	514 (756)	1.54 (2.11)
N-16 is3	516 (756)	1.28 (1.83)
N-16 is4	514 (761)	1.54 (2.11)

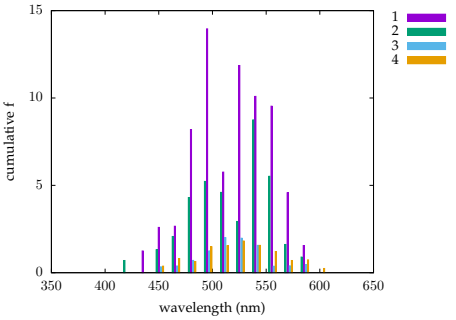


(a) NIAD-11 gas phase.

(b) NIAD-16 gas phase.



(c) NIAD-11 water.



(d) NIAD-16 water.

Figure B.4: Simulated absorption spectra from *ab initio* molecular dynamics simulations: isomer contributions weighted by their Boltzmann factor.

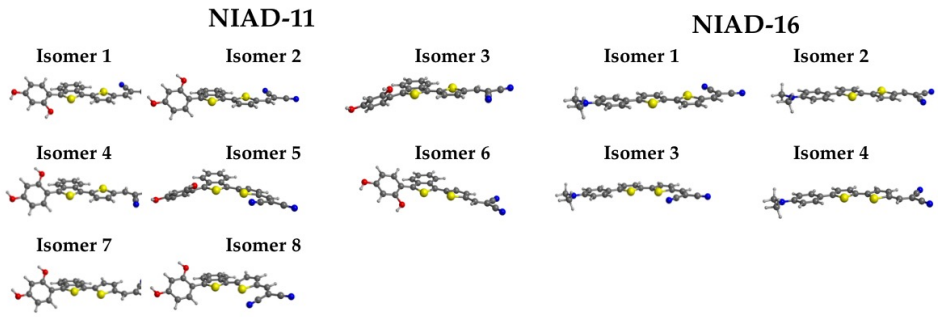


Figure B.6: Conformational isomers of NIAD-11 and NIAD-16 at the S_1 optimized geometry.

Table B.7: Relaxed S_1 energy scan around dihedral β of NIAD-11 and NIAD-16; excitation energies (eV) to S_1 and T_2 computed in gas phase and water. Values are corrected for the difference between MSCAPT2 and TDDFT energies at the S_1 optimized geometry. Labels refer to Figure 3.1.

β (°)	$S_1(gp)$	$T_2(gp)$	$S_1(H_2O)$	$T_2(H_2O)$
NIAD-11				
-135	2.26	2.89	1.62	3.00
-115	2.33	2.73	1.70	2.81
-95	2.33	2.53	1.84	2.58
-75	2.37	2.60	1.91	2.66
-55	2.32	2.78	1.75	2.86
-35	2.23	2.91	1.75	2.86
-15	2.15	2.99	1.64	3.01
5	2.11	3.02	1.56	3.11
25	2.17	2.97	1.54	3.14
NIAD-16				
-172	2.68	2.94	1.91	2.92
-152	2.72	2.92	1.95	2.89
-132	2.78	2.86	2.02	2.82
-112	2.83	2.78	2.13	2.68
-92	2.84	2.76	2.39	2.59
-72	2.86	2.78	2.17	2.69
-52	2.81	2.85	2.02	2.81
-32	2.73	2.92	1.95	2.89
-12	2.68	2.96	1.91	2.93

Appendix C

DANIRs: conjugated π system markers

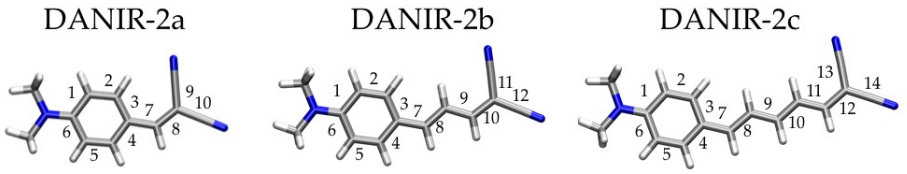


Figure C.1: Equilibrium bond lengths at the S_0 and S_1 optimized geometry are reported in Table C.1

Table C.1: Bond lengths (Å) of DANIR-2a, -2b and -2c at the S_0 and S_1 optimized geometries in gas phase. Bond labels refer to Figure C.1.

Bond leng. S_0	2a	2b	2c	Bond leng. S_0	2a	2b	2c
1	1.42	1.42	1.41	8	1.36	1.35	1.35
2	1.38	1.38	1.38	9	1.43	1.43	1.43
3	1.41	1.41	1.40	10	1.43	1.36	1.36
4	1.41	1.41	1.40	11	-	1.43	1.42
5	1.38	1.38	1.38	12	-	1.43	1.36
6	1.41	1.41	1.41	13	-	-	1.43
7	1.46	1.45	1.45	14	-	-	1.43
Bond leng. S_1	2a	2b	2c	Bond leng. S_1	2a	2b	2c
1	1.42	1.42	1.42	8	1.42	1.39	1.39
2	1.37	1.37	1.37	9	1.41	1.40	1.40
3	1.43	1.43	1.42	10	1.42	1.40	1.39
4	1.43	1.42	1.42	11	-	1.42	1.40
5	1.37	1.37	1.37	12	-	1.42	1.40
6	1.43	1.43	1.42	13	-	-	1.42
7	1.42	1.42	1.42	14	-	-	1.42

Table C.2: Relative potential E and free energies G ($kJ\,mol^{-1}$) of the conformational isomers of DANIR-2b and DANIR-2c in gas phase and water (SMD model).

	$\Delta E(gp)$	$\Delta G(gp)$	$\Delta E(H_2O)$	$\Delta G(H_2O)$
DANIR-2b isomer 1	0.0	0.0	0.0	0.0
DANIR-2b isomer 2	16.8	17.4	21.3	17.0
DANIR-2c isomer 1	0.0	0.0	0.0	0.0
DANIR-2c isomer 2	32.4	30.0	36.1	36.9
DANIR-2c isomer 3	14.0	11.8	14.0	12.2
DANIR-2c isomer 4	14.9	16.5	19.9	22.4

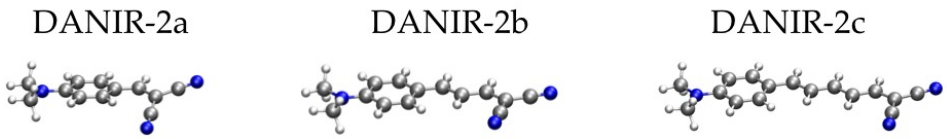


Figure C.2: S_1 optimized geometries of DANIR-2a, -b and -2c.

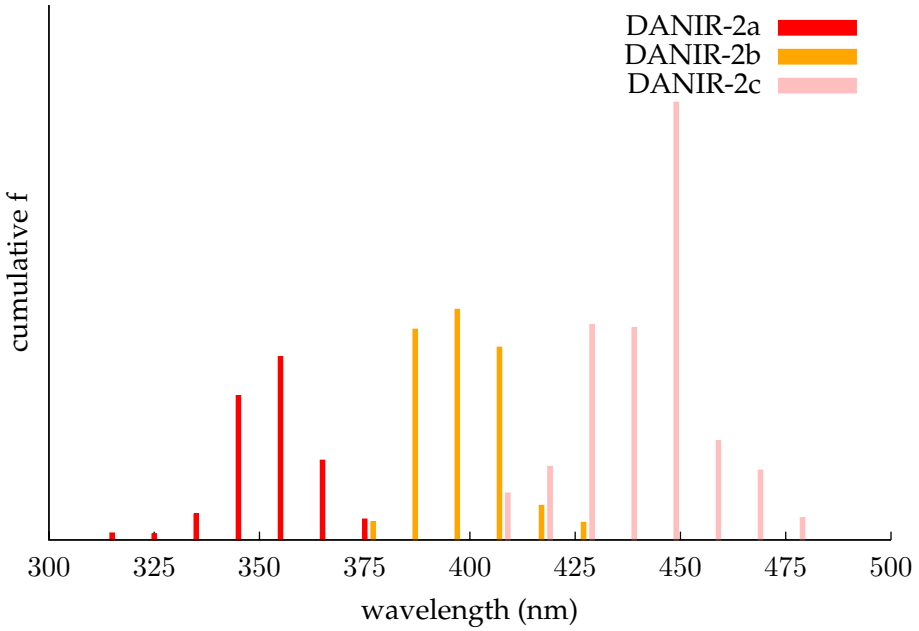
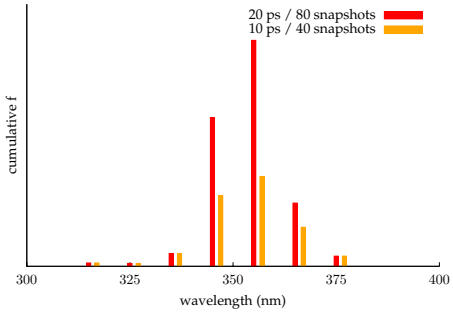
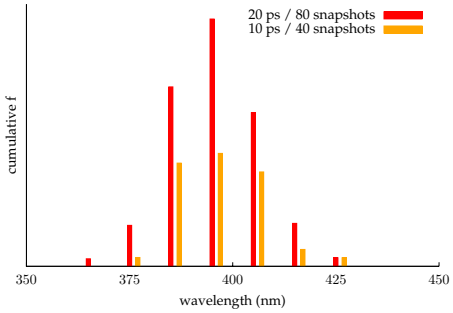


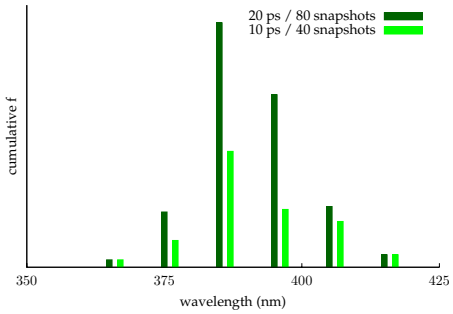
Figure C.3: Calculated absorption spectra from *ab initio* molecular dynamics simulations in gas phase of DANIR-2a, DANIR-2b and DANIR-2c.



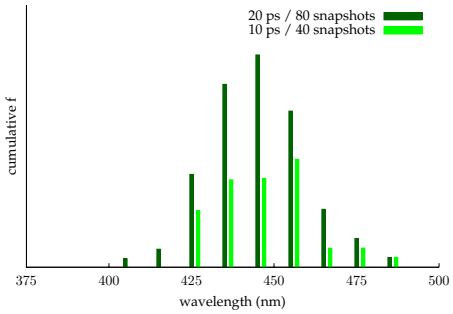
(a) DANIR-2a gas phase.



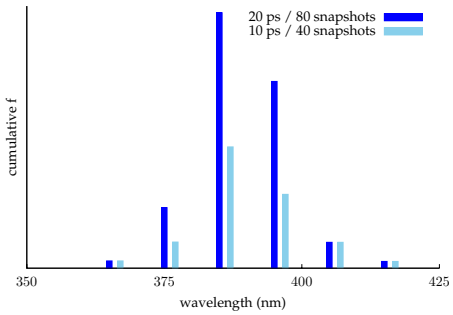
(b) DANIR-2b gas phase.



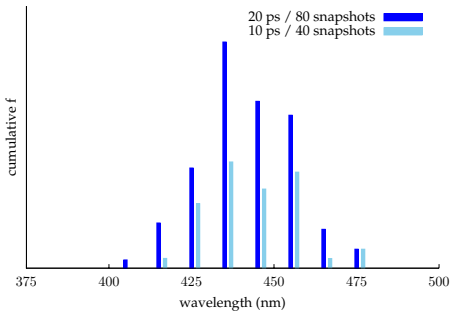
(c) DANIR-2a dichloromethane.



(d) DANIR-2b dichloromethane.

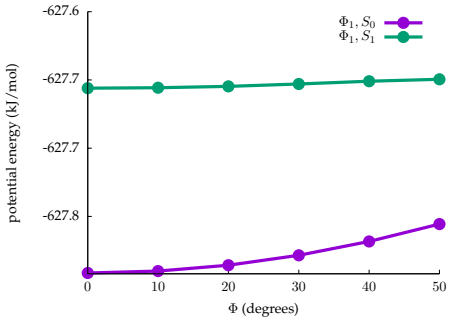


(e) DANIR-2a water.

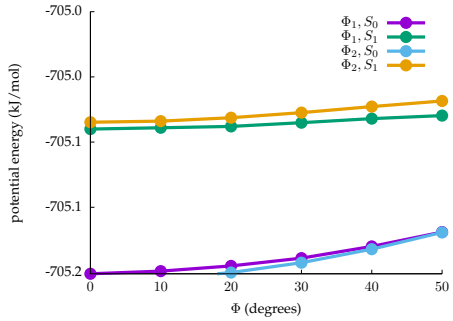


(f) DANIR-2b water.

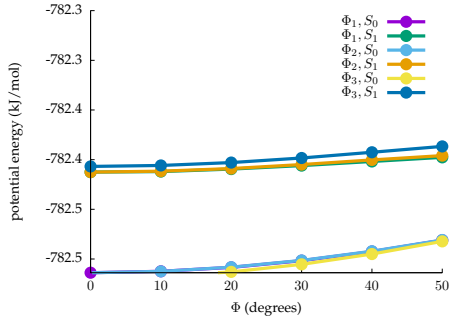
Figure C.4: Calculated absorption spectra from *ab initio* molecular dynamics simulations: effect of simulation time.



(a) DANIR-2a rigid scan.



(b) DANIR-2b rigid scan.



(c) DANIR-2c rigid scan.

Figure C.5: S_0 and S_1 energies for the rigid scans around double bonds Φ_1, Φ_2 and Φ_3 ($kJ\,mol^{-1}$) of DANIR-2a, -2b and -2c.

Appendix D

Binding to amyloid-β fibrils

Table D.1: Approximate dimensions of the binding channels of NIAD-4, NIAD-11, NIAD-16 and DANIR-2c on fibril models I, II and III (Figure 5.3). This description is not applicable to NIAD-11/I, in which the is accommodated externally, and no channel can be defined.

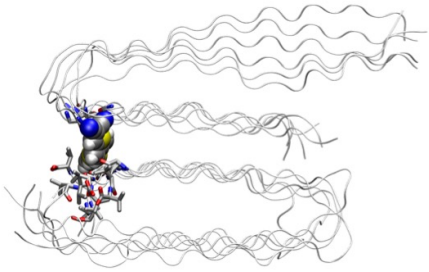
Binding pose	dimensions (Å)
NIAD-4/I-ji	9×11
NIAD-4/II-pe	10×12
NIAD-4/III-pi	13×8
NIAD-11/II-pi	8×9
NIAD-11/III-ji	10×9
NIAD-16/I-pe	8×11
NIAD-16/II-c	14×10
NIAD-16/III-pi	16×9
DANIR-2c/I-ji	9×6
DANIR-2c/II-c	7×9
DANIR-2c/III-pi	13×7

Table D.2: Binding energies ($kCal\,mol^{-1}$) of selected binding poses of NIAD-4, NIAD-11, NIAD-16 and DANIR-2c on fibril models I, II, III.

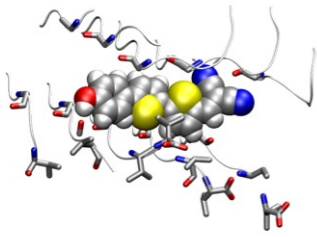
Binding pose	PELE	MM/GBSA	MM/PBSA
NIAD-4/I-je	-66.7	-42.1	-6.7
NIAD-4/I-pe	-73.8	-29.1	6.0
NIAD-4/II-c	-60.5	-38.2	-2.9
NIAD-4/III-ji	-58.3	-35.1	12.2
NIAD-11/I-pe	-38.1	-27.7	9.3
NIAD-11/II-c	-38.3	-52.7	-4.1
NIAD-11/III-pi	-43.0	-57.7	1.5
NIAD-16/I-je	-56.8	-50.4	-6.3
NIAD-16/II-je	-46.5	-20.6	2.6
DANIR-2c/I-je	-46.3	-46.6	-11.9
DANIR-2c/II-pi	-51.6	-36.7	-3.9
DANIR-2c/III-pi2	-42.6	-38.9	-4.0

Table D.3: Binding energies ($kcal\,mol^{-1}$) of selected binding poses of NIAD-4, NIAD-11, NIAD-16 and DANIR-2c on fibril models I, II, III (Figure 5.3). MM/PBSA values computed with a solute-solvent interaction model in which the nonpolar solvation free energy is modeled as a single term linearly proportional to the solvent accessible surface area.

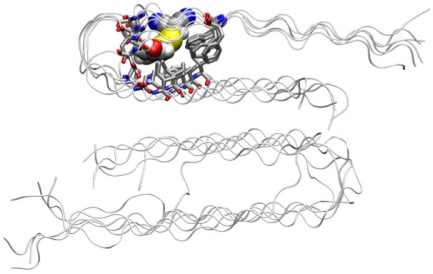
Binding pose	PELE	MM/GBSA	MM/PBSA
NIAD-4/I-ji	-80.6	-59.3	-40.5
NIAD-4/II-pe	-83.0	-41.9	-22.5
NIAD-4/III-pi	-60.4	-51.0	-39.3
NIAD-11/I-je	-53.2	-45.4	-32.6
NIAD-11/II-pi	-48.3	-38.0	-28.5
NIAD-11/III-ji	-45.7	-40.9	-20.8
NIAD-16/I-pe	-58.8	-44.5	-29.5
NIAD-16/II-c	-56.8	-43.9	-29.8
NIAD-16/III-pi	-44.9	-52.9	-38.9
DANIR-2c/I-ji	-60.4	-63.3	-36.4
DANIR-2c/II-c	-53.2	-45.7	-29.1
DANIR-2c/III-pi	-51.9	-57.7	-10.8



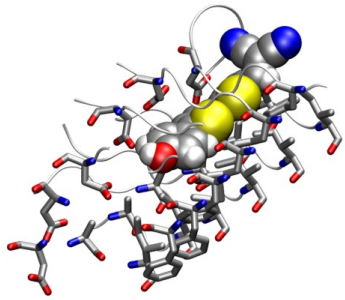
(a) NIAD-4/I-je.



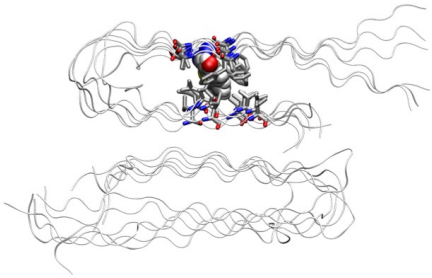
(b) GLY29, GLY37, VAL39, VAL40.



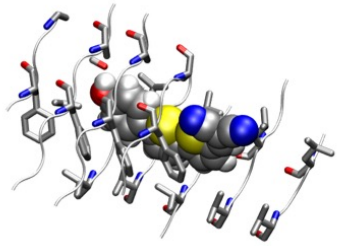
(c) NIAD-4/I-pi.



(d) PHE19, ASP23, ALA30, LEU32.

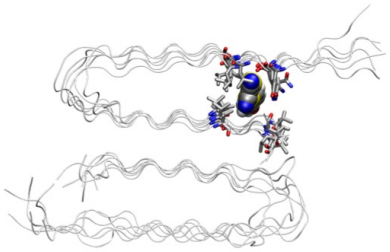


(e) NIAD-4/I-pcfects.

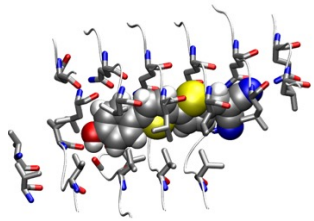


(f) PHE19, ALA21, VAL34, VAL36.

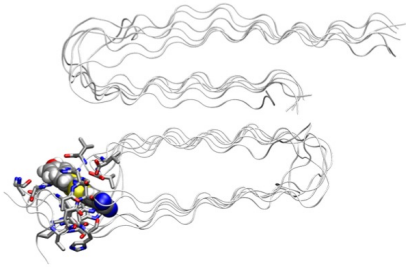
Figure D.1: Binding poses of NIAD-4 on fibril model I with indication of the aminoacidic residues involved in the interaction.



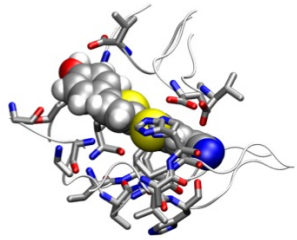
(a) NIAD-4/I-pe2.



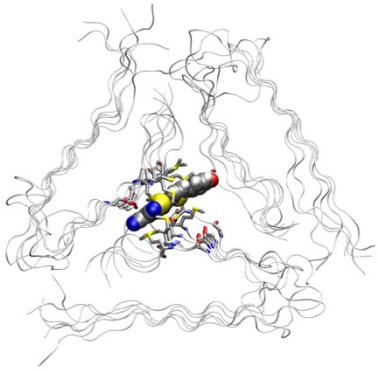
(b) GLU16, LEU17, VAL37, VAL39.



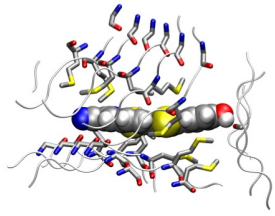
(c) NIAD-4/I-pe.



(d) GLY9, VAL12, HS13, VAL40.

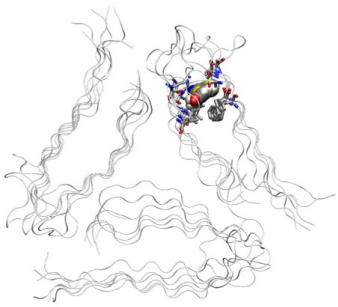


(e) NIAD-4/II-c.

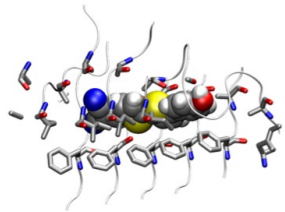


(f) GLY33, MET35.

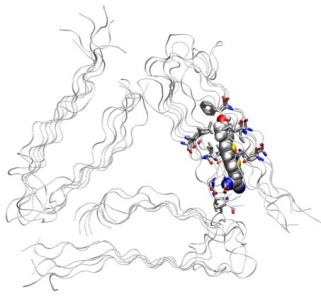
Figure D.2: Binding poses of NIAD-4 on fibril models I and II with indication of the aminoacidic residues involved in the interaction.



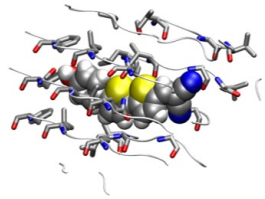
(a) NIAD-4/II-pi.



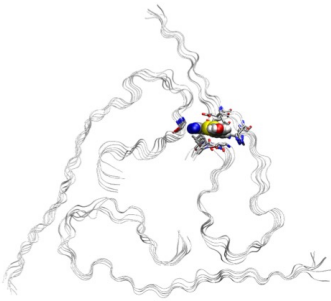
(b) PHE19, ALA21, ILE32, ALA33.



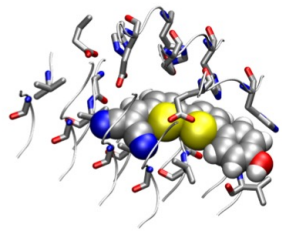
(c) NIAD-4/II-pc.



(d) GLN15, LEU17, PHE19, LEU34, VAl36, VAL39.

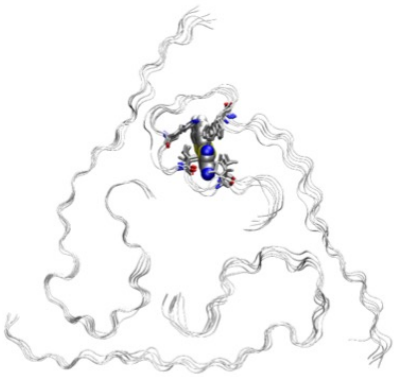


(e) NIAD-4/III-ji.

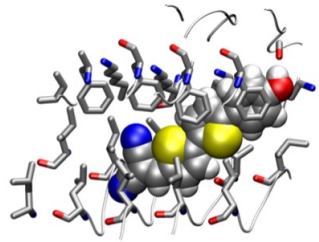


(f) GLU11, HIS13, GLY29, GLY38, VAL39.

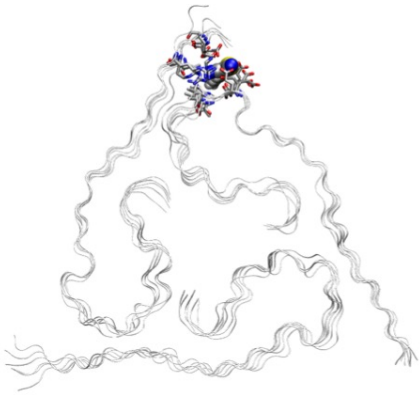
Figure D.3: Binding poses of NIAD-4 on fibril models II and III with indication of the aminoacidic residues involved in the interaction.



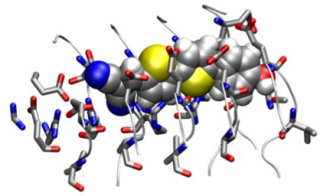
(a) NIAD-4/III-pi2.



(b) PHE20, LYS28, ISLE 31, LEU34.

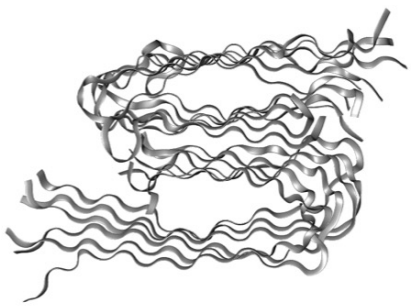


(c) NIAD-4/III-je.

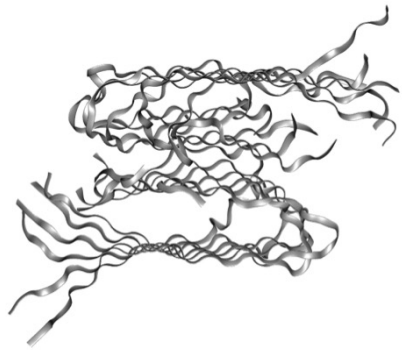


(d) GLU3, ARG5, GLU22, VAL24.

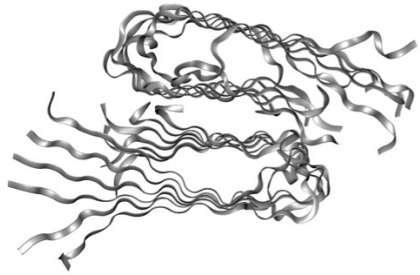
Figure D.4: Binding poses of NIAD-4 on fibril model III with indication of the aminoacidic residues involved in the interaction.



(a) Beginning of 10 ns simulation.

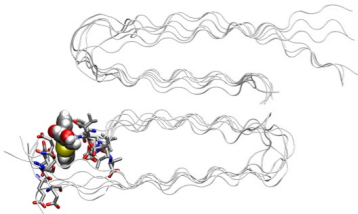


(b) After 5 ns of simulation.

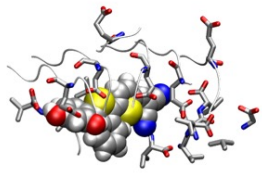


(c) After 10 ns of simulation.

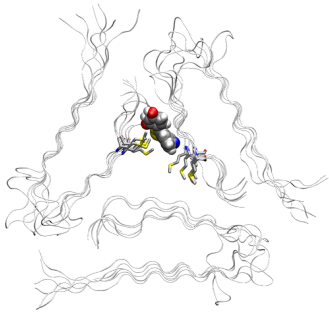
Figure D.5: Structural modifications undergone by fiber I (binding pose NIAD-4/I-ji) along a 10 ns molecular dynamics simulation. The corresponding RMSD is plotted in Figure 5.5.



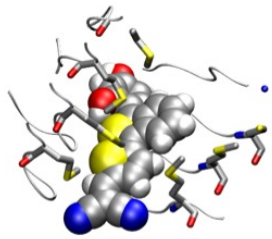
(a) NIAD-11/I-pe.



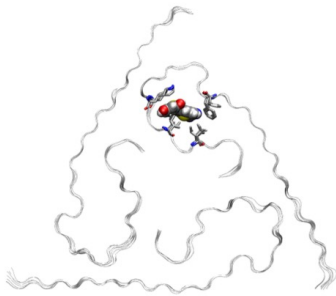
(b) GLU11, VAL39, VAL40.



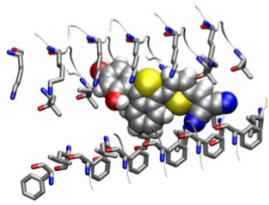
(c) NIAD-11/II-c.



(d) GLY33, MET35.

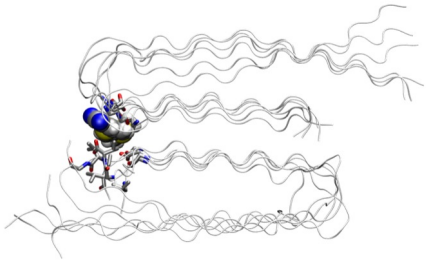


(e) NIAD-11/III-pi.

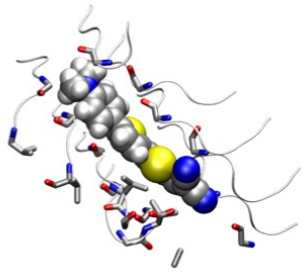


(f) PHE19, LYS28, ILE31, LEU34.

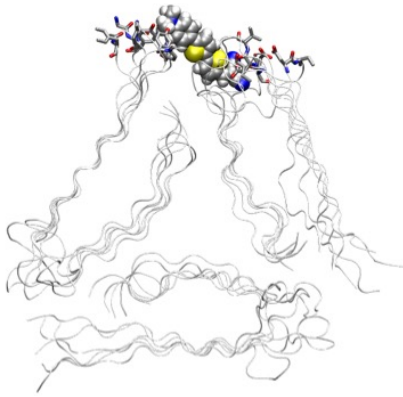
Figure D.6: Binding poses of NIAD-11 on fibril models I, II and III with indication of the aminoacidic residues involved in the interaction.



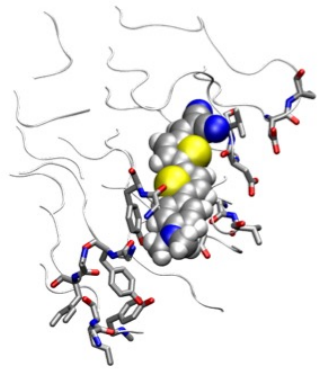
(a) NIAD-16/I-je.



(b) GLY29, GLY37, VAL39, VAL40.

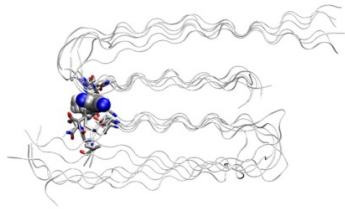


(c) NIAD-16/II-je.

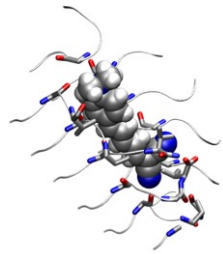


(d) GLY9, TYR10, ASP23, VAL24.

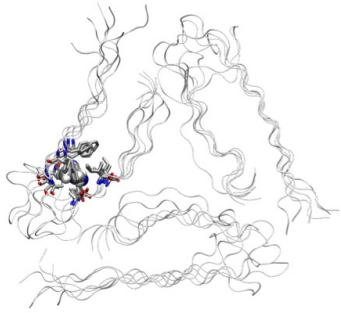
Figure D.7: Binding poses of NIAD-16 on fibril models I, II and III with indication of the aminoacidic residues involved in the interaction.



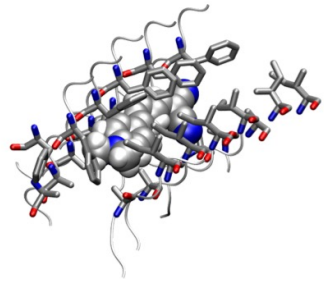
(a) DANIR-2c/I-je.



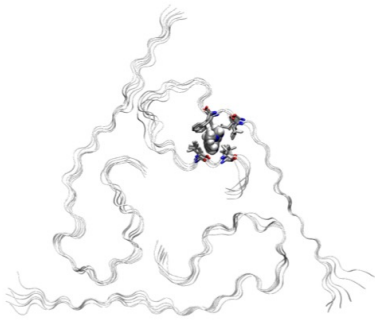
(b) GLU11, VAL39, VAL40.



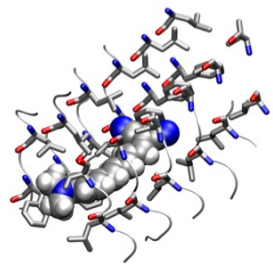
(c) DANIR-2c/II-pi.



(d) PHE19, ALA21, ALA30, ILE32.



(e) DANIR-2c/III-2.



(f) LEU17, PHE19, LEU34, VAL36.

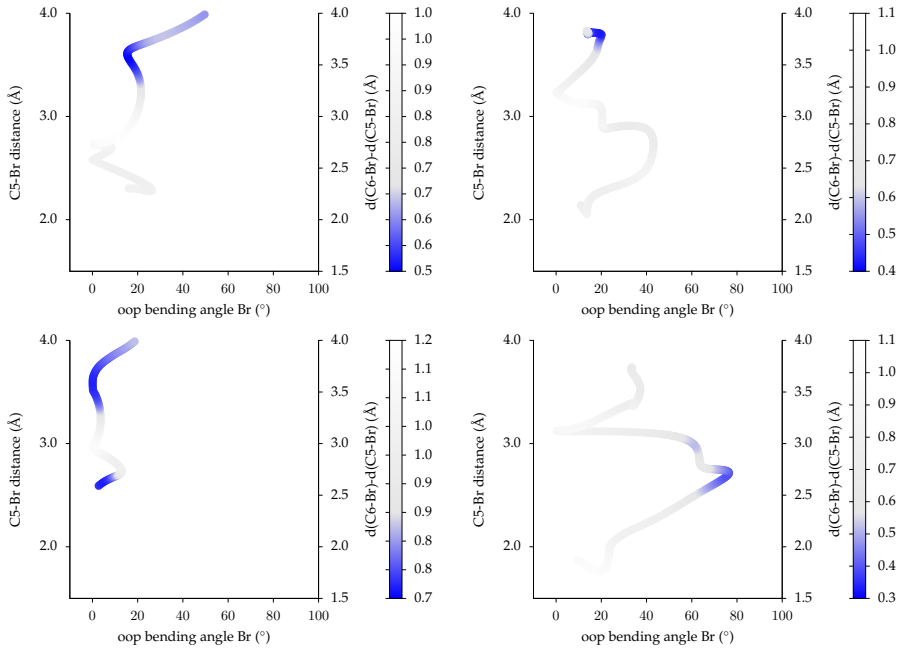
Figure D.8: Binding poses of DANIR-2c on fibril models I, II and III with indication of the aminoacidic residues involved in the interaction.

Appendix E

Photochemistry of 5-bromouracil

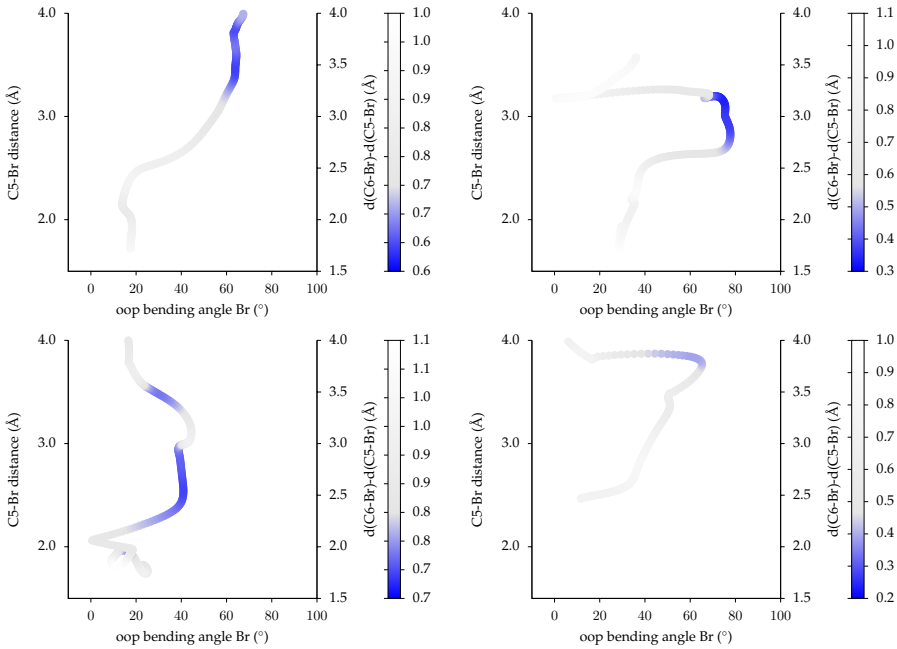
Table E.1: Computed MSCASPT2(16,12)/ANO-RCC-VDZP excitation energies at the Franck-Condon geometry of 5-bromouracil with the two different orbital spaces reported in Figure 6.1.

	E (eV)	
S_1	5.0	$\pi_4 \rightarrow \pi_1^*$
S_2	5.7	$\pi_4 \rightarrow \sigma CBr^*$
S_3	6.4	$\pi_3 \rightarrow \pi_1^*$
S_4	6.8	$n_Br \rightarrow \sigma CBr^*$
S_5	6.8	$n_Br \rightarrow \pi_1^*$
S_1	5.1	$\pi_4 \rightarrow \pi_1^*$
S_2	5.1	$n_{O1} \rightarrow \pi_1^*, n_{O2} \rightarrow \pi_1^*$
S_3	6.3	$\pi_4 \rightarrow \sigma CBr^*$
S_4	6.3	$\pi_3 \rightarrow \pi_1^*$
S_5	6.5	$n_{O1} \rightarrow \pi_2^*, n_{O2} \rightarrow \pi_2^*$
T_1	3.8	$\pi_4 \rightarrow \pi_1^*$
T_2	5.1	$n_{O1} \rightarrow \pi_1^*, n_{O2} \rightarrow \pi_1^*$
T_3	5.4	$\pi_4 \rightarrow \sigma CBr^*$
T_4	5.7	$\pi_3 \rightarrow \pi_1^*$
T_5	6.6	$n_{O1} \rightarrow \pi_2^*, n_{O2} \rightarrow \pi_2^*$



.

Figure E.1: Representation of the difference between bond lengths d($C6-$Br) and d($C5-$Br) along eight representative trajectories. The difference is always positive, indicating that no Br transfer is produced from C5 to C6, in contrast with what suggested in ref[216]



.

Figure E.2: Representation of the difference between bond lengths d($C6-$Br) and d($C5-$Br) along eight representative trajectories. The difference is always positive, indicating that no Br transfer is produced from C5 to C6, in contrast with what suggested in ref[216]