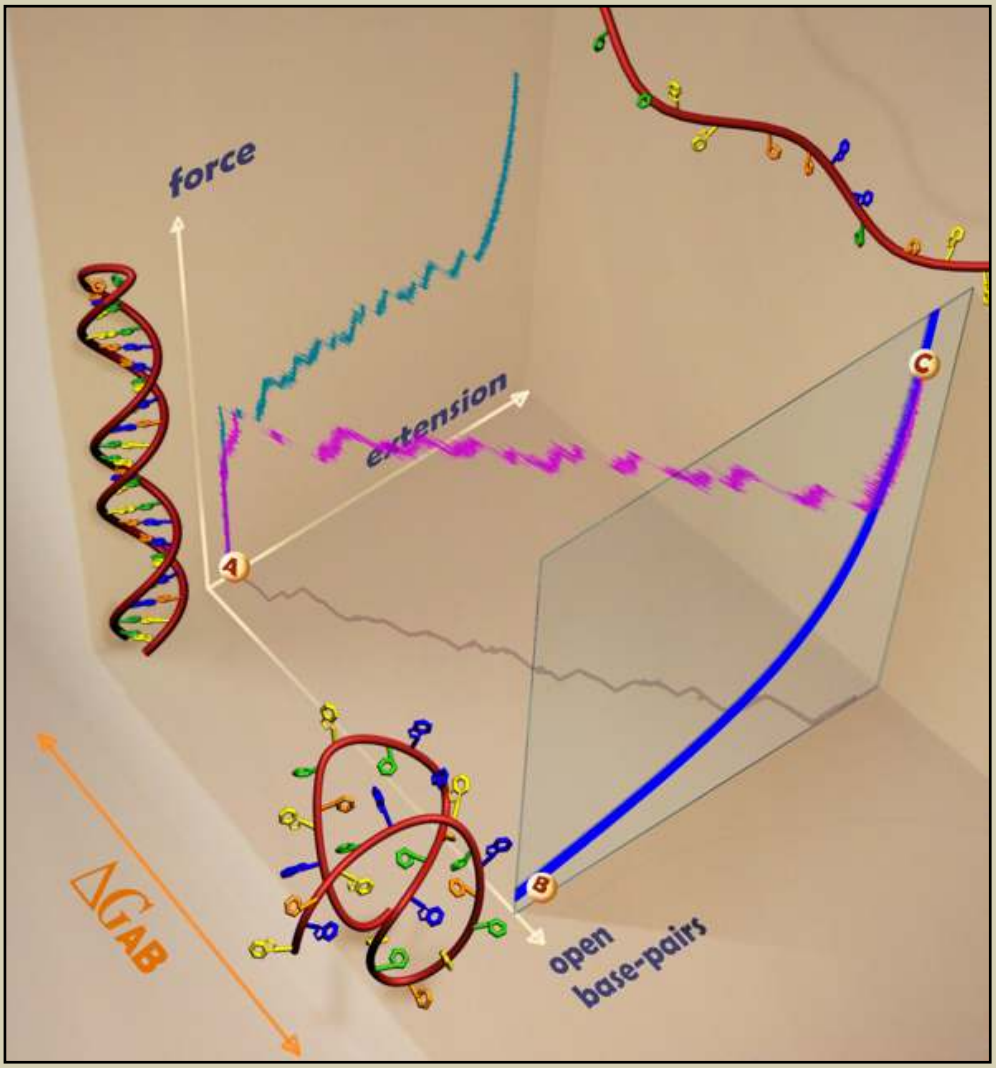
UNIVERSITAT DE BARCELONA

Statistical and thermodynamical properties of DNA unzipping experiments with optical tweezers

Josep Maria Huguet Casades

Statistical and thermodynamic properties of DNA unzipping experiments with optical tweezers

Josep Maria Huguet i Casades



UNIVERSITAT DE BARCELONA

PhD thesis

UNIVERSITAT DE BARCELONA

Departament de Física Fonamental
Programa de doctorat Física Avançada
Bienni 2004/2006

Tesi doctoral · PhD thesis

Statistical and thermodynamic properties of DNA unzipping experiments with optical tweezers

Josep Maria Huguet i Casades

Memòria presentada
el 26 d'octubre de 2010 per optar al
títol de Doctor en Física,
sota la direcció del
Dr. Fèlix Ritort i Farran

La realització d'aquesta tesi ha estat possible gràcies a la concessió d'una beca predoctoral del *Programa Nacional de Formación de Profesorado Universitario (FPU)* del *Ministerio de Educación y Ciencia*, i al suport del *Laboratori de Física de Biomolècules i sistemes petits* del *Departament de Física Fonamental* de la *Universitat de Barcelona*.

El que res no costa res no val.

Refrany popular català

Agraïments

Aquesta tesi que teniu entre les mans és el resultat de 6 anys d'esforç. Al moment d'iniciar aquest viatge no era conscient de la dedicació que em suposaria dur a terme un repte com aquest. Certament he d'admetre que hi ha hagut moments de dubtes i dificultats. Però veient en conjunt tot el que he pogut aprendre i viure en aquest temps em sento satisfet d'haver recorregut aquest camí. Tot plegat no hauria estat possible sense el suport i l'afecte de tota aquella gent que m'ha rodejat aquests darrers anys. En aquests moments sento una profunda i sincera gratitud envers tots ells i no puc deixar escapar l'ocasió de dedicar-los unes paraules.

En primer lloc vull agrair la confiança i el suport del meu director de tesi, el Fèlix Ritort. Ell ha estat un mestre de qui he après moltes lliçons, tant a nivell científic com personal. Li agraeixo especialment que m'hagi mostrat una manera de treballar basada en l'esforç, la tenacitat i la perseverança en l'assoliment dels objectius. El seu compromís amb la feina ben feta sempre serà un exemple per a mi. La seva actitud positiva i constructiva davant les dificultats i la seva imaginació per pensar nous reptes científics són dues qualitats que valoro notablement. Finalment, vull agrair-li que confiés en mi per ajudar-lo a muntar el seu laboratori. L'oportunitat que em va donar em va permetre viatjar als Estats Units i viure una experiència inoblidable.

La Maria va ser la meva companya de despatx els primers anys de tesi. Amb ella vaig compartir molts bons moments i vull agrair-li els ànims i els consells que em va donar per a afrontar tota aquesta etapa. La providència ha permès que també hàgim pogut compartir aquest últim any. Sens dubte, la seva amistat ha estat un regal. L'Ivan va aparèixer ben aviat. Un home encuriosit per la ciència que sempre m'ha complagut. Escoltador com és, he gaudit de molt bones converses amb ell. La Sara va venir poc després. Una dona plena de vitalitat amb un cor enorme que es fa estimar. Per a mi ha estat un puntal dins del laboratori. La seva comprensió m'ha donat forces en els moments de defalliment. La Núria em va mostrar que la paciència i la serenor són valors importantíssims per a fer la feina ben feta. La seva meticulositat i constància han estat un referent per a mi. Sento una immensa gratitud per la Sara i la Núria. Amb elles dues vam començar a fer els primers passos al laboratori. Plegats vam haver de superar moments de dificultat i conjurar-nos per a seguir lluitant per l'objectiu que el Fèlix ens havia encomanat: el muntatge del laboratori. La Kumiko va fer una estada memorable al laboratori.

Des de llavors, les seves anades i vingudes del Japó a intervals més o menys regulars m'han permès intercanviar vivències i, sobretot, riure molt. Més tard va aparèixer l'Alessandro, un home optimista i interessat per tot allò que afecta al ser humà. Un humanista de vasta cultura de qui cada dia s'aprèn alguna cosa i la presència del qual acoloreix la rutina. La Carol es va unir al grup poc després, somrient i amistosa. I tot seguit va arribar el Joan, que ha estat el meu company de despatx des de llavors. El Joan ha estat un molt bon col·lega, atent i comprensiu, i amb una capacitat de treball i d'aprenentatge dignes d'admiració. Amb ell he acabat establint una molt bona amistat. L'Anna es va incorporar just després i el seu caràcter jovial i desinhibit van fer augmentar l'estat d'ànim del laboratori. Ella és una física brillant, a qui vaig tenir ocasió de donar classe... dos dies. A continuació, el Cristiano va unir-se al laboratori. La seva irrupció va suposar un revulsiu per a mi. La seva actitud optimista i constructiva davant la ciència em va renovar la il·lusió per a seguir treballant. Li estic enormement agraït per l'ajuda que em va brindar. El Marco, un noi trempat amb un gran esperit crític, rebel i compromès, va incorporar-se posteriorment. El Javi va ser aire fresc per al laboratori, un tècnic ràpid, pragmàtic i capaç. I també un company amb un gran sentit de l'humor. El Lorenzo va venir més tard. Un home treballador, independent i eficient, amb les idees clares i els recursos per a dur-les a terme. Finalment, també vull agrair al Mario (gran coneixedor de la Barcelona nocturna), al Robert (enginyer robust), al Bing Feng (treballador incansable), a la Blanca (pencaire entusiasta), a la Sandra (sensible melòmana) i a l'Eckhard (joveníssima promesa de la ciència) el temps que hem compartit plegats al laboratori.

En la meva estada a Berkeley vaig tenir l'oportunitat de treballar al costat del Steve Smith. Amb ell vaig viure el període de la meva vida en què més coses vaig aprendre. La seva generositat i paciència em van permetre assolir tota una sèrie de coneixements que altrament hauria tardat anys a adquirir. La seva forma d'afrontar els reptes experimentals m'han inspirant al llarg de la realització de la tesi. També vaig compartir molt de temps amb el Shane Saxon, que em va introduir en les catacumbes de la informàtica i l'electrònica. Amb ell vaig acabar establint una relació de mestre-alumne força excèntrica, però molt cordial. A Berkeley vaig conèixer al Ricardo i al Borja, que van ser un referent per a mi i em van aconsellar en tot moment. Finalment, vull agrair l'hospitalitat i la generositat del Carlos Bustamante, per acceptar de bon grat la meva estada al seu laboratori.

D'entre els professors que m'han guiat aquests anys, m'agradaria destacar la influència del Jordi Ortín. Ell em va iniciar en la recerca i em va encoratjar a seguir per aquesta via. En els cursos de doctorat, vaig poder gaudir de les interessants reflexions del Jorge Wagensberg sobre biofísica. I també voldria mencionar els meus tutors dels treballs del DEA, l'Ignasi Pagonabarraga i l'Eduard Vives. Amb ells vaig aprendre altres maneres de fer ciència que m'han enriquit científicament i personal.

Dins mateix de la facultat, he passat molt bones estones rodejat dels companys de doctorat: el Javi, el Ramon, el Xavi, la Mireia, el Miquel, la Neus, el Rodrigo,

el Marc, el Fèlix, l'Isaac, el Xavifu i la Marta. Amb l'Albert i el Xavi he compartit bons dinars i intercanvis d'impressions els dijous. I també vull recordar-me de la Marta Dies i el Jordi Martínez, per tots els bons moments dins i fora del departament. A tots ells, gràcies per ser-hi.

Més enllà de l'àmbit purament acadèmic, vull agrair el suport que he rebut dels meus pares. Els seus ànims en els moments difícils m'han ajudat a no prendre decisions precipitades i renovar les forces per a seguir endavant. També vull agrair-li a la Gemma, la meva germana, els moments d'intercanvi d'impressions sobre la feina que estàvem fent. El fet de saber que compartíem situacions i recels similars davant els reptes que se'ns presentaven ha tingut un efecte balsàmic per a mi. Vull tenir un record molt especial per als meus quatre padrins. Crec que estarien contents de saber que la seva lluita no va ser en va. Les oportunitats que hem tingut la nostra generació són fruit del seu esforç. Al meus tiets, cosins i parents vull agrair-los l'interès que han demostrat per l'evolució de la tesi i que, insistentment, hagin preguntat quan tenia previst finalitzar-la. Gràcies al Gerard per ajudar-me en el disseny de la portada. I també vull agrair a aquells amics de la colla de Sant Martí la seva curiositat per entendre a què m'estava dedicant. Ha estat un plaer respondre a la pregunta de: "I això que fas per a què serveix?".

Finalment vull agrair a la Montserrat el seu suport incondicional i la seva immensa paciència. Ella és la persona que he tingut al meu costat tot aquest temps i que veritablement coneix l'esforç ingent que he hagut de fer. Aquesta tesi també ha suposat un esforç per a ella, que ha hagut de suportar el meu compromís amb el laboratori i els excessos de feina que ha comportat. La Montserrat ha viscut com a propis els meus desencisos. Compartir-los amb ella m'ha donat la fortalesa per a seguir endavant. M'ha aconsellat encertadament i m'ha ensenyat com afrontar les situacions difícils amb coratge, fermesa i decisió. La seva plena confiança en mi m'ha proporcionat la seguretat necessària per a persistir. I també ha viscut com a pròpies les meves alegries, demostrant-me que ha sigut al meu costat en tot moment. Per tot això i pel que ha de venir, moltes gràcies Montserrat.

Sant Martí de Maldà, 25 d'octubre de 2010

Contents

Part I

Resum de la tesi en català

Sumari

La biofísica molecular és una disciplina científica que estudia les biomolècules. Aquesta disciplina ha experimentat una revolució gràcies al desenvolupament de les tècniques de manipulació de molècules individuals. Aquestes tècniques permeten obtenir nous tipus de mesures que complementen les tècniques tradicionals realitzades en volum (és a dir, amb quantitats de molècules de l'ordre del mol). Les pinces òptiques són una tècnica experimental que utilitza la pressió de radiació de la llum per exercir forces en microesferes dielèctriques. Les biomolècules poden enllaçar-se amb aquestes microesferes per tal de realitzar experiments d'estirament. Les *Minipinces* són un instrument de pinces òptiques amb dos làsers contra-propagants que utilitza la conservació del moment lineal de la llum per tal de mesurar la força exercida sobre les microesferes. L'instrument té una gran estabilitat i resolució (0.1 pN en força i 0.5 nm en distància) en les mesures. Les propietats de l'ADN es poden estudiar amb les Minipinces a nivell de molècules individuals. L'ADN és una biomolècula formada per una doble hèlix que emmagatzema la informació genètica de la cèl·lula. Els experiments de ruptura mecànica de l'ADN consisteixen a separar les dues cadenes mitjançant l'aplicació de força als extrems de la molècula. En aquest procés, els parells de bases (pb) són romputs seqüencialment, mostrant una successió de regions cooperatives de ruptura (RCR) de diferents grandàries (entre 1–100 pb). En un experiment de ruptura d'ADN, es mesuren les corbes de força vs. distància (CFD) de la molècula. Aquestes corbes tenen una forma característica de dent de serra que depèn de la seqüència de la molècula. La CFD s'analitza d'acord amb una aproximació bayesiana per tal d'inferir la distribució de grandàries de les RCRs. La precisió experimental no permet observar RCRs de grandàries inferiors als 10 pb. A més, la ruptura dels parells de bases d'un en un només es pot assolir si es disposa d'una trampa òptica amb una rigidesa superior als 0.1 N/m. Aquest valor coincideix amb la rigidesa d'un únic nucleòtid d'ADN i s'ha deduït a partir d'un model de joguina específicament introduït per tal d'estudiar les distribucions de grandàries de les CFD. Per altra banda, les CFD es poden predir teòricament mitjançant el model de primers veïns (PV) adaptat als experiments de ruptura. El model de PV descriu la reacció d'hibridació de dues cadenes d'ADN. Ajustant les mesures experimentals de les CFD al model de PV, es poden obtenir les 10 energies de formació a PV amb una precisió de 0.1 kcal·mol^{-1} entre 10 mM–1 M de concentració de cations monovalents. Els resultats mostren que les CFD de ruptura i les temperatures de desnaturalització d'oligonucleòtids es poden descriure correctament amb una correcció específica de sal per a cadascuna de les 10 energies de

formació a PV. Per altra banda, la ruptura d'ADN també es pot realitzar a força controlada. Aquest tipus d'experiments presenten molta histèresi i són irreversibles. El paisatge d'energia lliure és una eina que ajuda a entendre aquest tipus d'experiments a força controlada. Finalment, el treball presentat en aquesta tesi es pot estendre per tal de trobar aplicacions pràctiques a la ruptura de l'ADN, com ara la seqüenciació de l'ADN per aplicació de força, i la mesura de les propietats termodinàmiques de les molècules en aquelles condicions en que els experiments de volum no són factibles.

Resum

0.1 Introducció

La ciència no és un catàleg de fenòmens, sinó un marc conceptual per comprendre la natura. El grans científics són capaços d'unificar les explicacions dels fenòmens, tal i com Newton va fer en el seu llibre *Principia*, en què va unificar la mecànica terrestre i la celeste. El 1928 el químic alemany Friederich Wöhler va unificar accidentalment la química i la biologia, en descobrir que es podia sintetitzar matèria orgànica (urea) a partir de matèria inorgànica. Es varen tardar uns quants anys més a unificar la física i la biologia en el que avui coneixem com a biofísica. El descobriment de l'estructura en doble hèlix de l'ADN l'any 1953 [1] es pot considerar com el punt de partida de la biofísica moderna, en què un físic i un biòleg van unir esforços per a respondre una pregunta científica comuna.

0.1.1 Què és la biofísica?

La biofísica és una disciplina científica que utilitza les eines i els mètodes propis de la física per estudiar sistemes biològics. La biofísica engloba un ampli rang de sistemes (des de biomolècules fins a ecosistemes, passant per cèl·lules i individus) i s'interessa per diversos aspectes d'aquests sistemes (estructura, cinètica, etc.).

Hi ha dues formes d'aproximació a la biofísica. La primera és des d'un punt de vista físic, en la que els físics utilitzen els sistemes biològics per a descobrir nous fenòmens i lleis físiques. La segona és des d'un punt de vista biològic, en la que els biòlegs utilitzen tècniques experimentals de la física i la matemàtica per a afrontar problemes biològics. Actualment, ambdós punts de vista estan convergint i els biofísics s'han convertit en un nou tipus de científic. Els biofísics no pretenen descriure els fenòmens amb el detall que ho fan els biòlegs ni amb la simplificació idealista del físics. Els biofísics tenen les seves pròpies preguntes científiques i busquen les respostes d'acord amb un mètode que és la fusió dels mètodes propis de la física i la biologia.

Les biofísica està íntimament relacionada amb altres disciplines com la bioquímica, la nanotecnologia i la biomedicina. Avui en dia trobem una gran interdisciplinarietat i, per tant, les fronteres entre elles no són clares. En qualsevol cas, la biofísica es caracteritza sobretot per l'ús de mesures i informació quantitativa.

0.1.2 Biofísica molecular

La biofísica molecular s'ocupa de l'estudi de l'estructura, la funció i la cinètica de les biomolècules. Les biomolècules són els maons que conformen la vida i n'hi ha de dos tipus: els àcids nucleics (ADN i ARN) i les proteïnes (que inclouen enzims i motors moleculars). Les energies dels enllaços de les biomolècules són del mateix ordre de magnitud que el bany tèrmic que les rodeja, és a dir, que les biomolècules estan governades per les fluctuacions tèrmiques. Una de les qüestions interessants en biofísica és entendre com les biomolècules poden treballar de manera tan eficient immerses en aquestes condicions.

Alguns dels temes interessants en biofísica molecular són el plegament d'ARN i proteïnes, la mecano-química dels motors moleculars, el transport en canals iònics i les interaccions ADN-proteïna. Tot plegat s'estudia amb combinacions de tècniques pròpies de la bioquímica, la química i la física. El desenvolupament de les tècniques de molècules individuals ha suposat una nova manera d'enfocar l'exploració de les biomolècules.

0.1.3 Tècniques de molècula individual

Tradicionalment, els experiments en química i biologia s'han realitzat en volum. Els experiments en volum involucren una gran quantitat de substància (mols, grams, mil·lilitres) quan es comparen amb una sola molècula de substància. Les tècniques de molècula individual consisteixen a realitzar experiments sobre una única molècula [2]. Tenint en compte que a la cèl·lula hi ha un gran nombre de molècules, es podria pensar que l'estudi d'una única molècula no pot aportar gran cosa. Això no obstant, les biomolècules tenen una estructura interna complexa i l'ús de tècniques de molècula individual aporta molta informació. A diferència de les tècniques de volum, les mesures no són la mitjana de totes les molècules i es poden fer mesures de desviacions o distribucions de probabilitat. Per exemple, a nivell de molècula individual s'ha pogut mesurar la velocitat instantània de motors moleculars, mentre que en els experiments en volum només es poden mesurar velocitats mitjanes.

El desenvolupament de les tècniques de cadena individual ha empès la instrumentació, que ha evolucionat extraordinàriament al llarg de la primera

dècada del segle XXI. Les tècniques de cadena individual més productives han estat la fluorescència, el microscopi de força atòmica, les pinces magnètiques i les pinces òptiques. La figura 1 mostra una breu descripció de cadascuna d'elles.

La recerca en biofísica molecular està definitivament lligada a les tècniques de molècula individual, que segueix evolucionant.

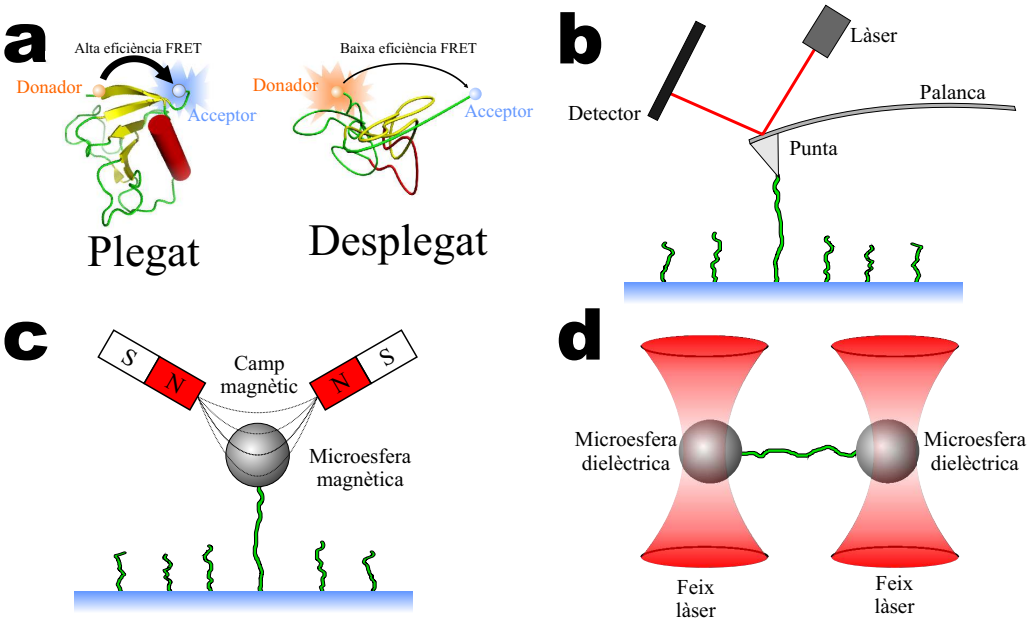


Figura 1: Tècniques de molècula individual. **(a)** Transmissió d'energia de ressonància de Förster (FRET). El FRET es basa en la transmissió d'energia entre dos fluoròfors (donador i acceptor). L'eficiència d'aquesta transmissió depèn de la distància entre aquests fluoròfors. Es pot utilitzar per detectar el plegament i desplegament d'una molècula. **(b)** Microscopi de força atòmica (AFM). L'AFM produeix i mesura les forces aplicades a una molècula. La força aplicada està directament relacionada amb la flexió d'una palanca sòlida, l'elasticitat de la qual és coneguda. La flexió de la palanca es mesura mitjançant la deflexió d'un feix làser reflectit a la palanca. **(c)** Pinces magnètiques (PM). La molècula s'enllaça entre una microesfera magnètica i la superfície d'un cobreobjectes. Les PM exerceixen força sobre la microesfera quan aquesta se sotmet a la influència d'un camp magnètic. La força es mesura a partir de la posició de la microesfera. El camp magnètic es genera mitjançant imants que es poden desplaçar. En alguns casos, els imants també poden rotar per tal d'exercir un parell de forces. **(d)** Pinces òptiques. Les pinces òptiques exerceixen força sobre microesferes dielèctriques utilitzant la pressió de radiació d'un feix làser. Una descripció més detallada es donarà a la secció següent.

0.1.4 Termodinàmica de sistemes petits

La termodinàmica de sistemes petits (també coneguda com a dinàmica mesoscòpica) és un subtema de la física que s'ocupa d'aquells sistemes que tenen una escala entremig dels sistemes microscòpics i els macroscòpics [3]. En aquesta escala d'entremig, el nombre de components del sistema és molt superior a 1 i molt inferior al nombre d'Avogadro. Aquests sistemes es caracteritzen per les elevades fluctuacions relatives de les seves magnituds observables. Tradicionalment, les fluctuacions de sistemes macroscòpics (gasos, imants) han estat molt difícils de mesurar. Amb el desenvolupament de les tècniques de molècula individual han aparegut tota una sèrie de sistemes que presenten fluctuacions rellevants i mesurables.

Els teoremes de fluctuació (TF) han estat i són un dels focus d'atenció de la termodinàmica de sistemes petits. Aquests teoremes relacionen les propietats d'equilibri d'un sistema amb el treball mecànic realitzat sobre aquest sistema al llarg de processos de no equilibri [4, 5]. Actualment, en biofísica s'utilitzen els TF per a obtenir les energies de formació de biomolècules. I, a la vegada, els experiments amb molècules individuals han contribuït a ampliar el nostre coneixement sobre els TF.

0.1.5 Sumari del treball presentat en aquesta tesi

El capítol 2 se centra en la descripció de les Minipinces. El capítol 3 mostra una descripció detallada dels experiments de ruptura d'ADN i els models que els descriuen. En el capítol 4 es realitza un estudi de les propietats estadístiques i dels estats metastables observats en els experiments de ruptura d'ADN. El capítol 5 mostra com extreure energies de formació d'hibridació a partir dels experiments. El capítol 6 descriu els experiments de ruptura a força constant. Finalment, el capítol 7 inclou les perspectives futures d'aquest treball i les conclusions.

0.2 Pinces òptiques

La llum porta moment lineal associat. Això és el que James Clerk Maxwell va deduir de les equacions del camp electromagnètic [6]. El principi de conservació del moment lineal comporta una transferència de moment quan la llum interacciona amb la matèria. És a dir, que la matèria experimenta una força quan interacciona amb la llum. Per descriure aquest fenomen, sovint es diu que la llum exerceix pressió de radiació sobre la matèria. A nivell macroscòpic, els efectes d'aquesta força són menyspreables. Això no obstant, a nivell microscòpic els objectes experimenten efectes observables.

Arthur Ashkin va ser el primer científic que va aconseguir controlar l'aplicació de la pressió de radiació d'un làser sobre microesferes [7]. Juntament amb els seus col·laboradors, ell mateix va perfeccionar la tècnica fins a aconseguir el confinament de les microesferes en una petita regió de l'espai [8]. Des de llavors, la tècnica es coneix amb el nom de 'pinces òptiques'. El següent pas va consistir en mesurar les forces exercides per les pinces òptiques. Entre el muntatges experimentals més reeixits per mesurar forces cal destacar l'interferometria [9], la detecció de microesferes per vídeo [10] i la deflexió de llum [11]. Aquest darrer sistema és el que s'utilitza en les Minipinces descrites en aquesta tesi.

Els avanços recents en pinces òptiques s'han dirigit per una banda cap a la millora de la resolució, la precisió i l'exactitud de les mesures, i per altra banda cap a la combinació d'altres tècniques experimentals (per exemple, pinces òptiques combinades amb fluorescència) [12]. La tendència actual en pinces òptiques consisteix a implementar múltiples trampes òptiques en un mateix muntatge experimental, ja que proporciona més versatilitat i millors mesures en els experiments [13]. Les pinces òptiques hologràfiques són una innovació prometedora, ja que l'experimentador pot controlar el nombre, la grandària, la posició i la rigidesa de les trampes òptiques de l'instrument [14].

Les pinces òptiques són una eina molt útil per als biofísics, ja que es tracta d'una tècnica no invasiva que la fa ideal per a treballar amb sistemes biològics.

0.2.1 Principi d'atrapament òptic

Hi ha diverses teories físiques que descriuen la interacció entre la radiació i la matèria. En el cas de les pinces òptiques, les equacions clàssiques de l'electromagnetisme de Maxwell descriuen correctament la interacció entre la llum del làser i les microesferes dielèctriques. Tot i que la pressió de radiació és l'única causa de les forces experimentades per la microesfera, aquestes forces generalment es divideixen entre forces de dispersió i forces

de gradient. Les forces de dispersió tendeixen a empènyer la microesfera en la direcció de propagació del làser, mentre que les forces de gradient ho fan en la direcció que apunta a la regió de màxima intensitat de llum. Si el feix làser es condiciona correctament (en termes de col·limació, distribució espacial d'intensitat, aberracions, etc.) és possible controlar la pressió de radiació que el làser exerceix sobre la microesfera. Una trampa òptica es pot aconseguir focalitzant un feix làser en una petita regió de l'espai. En aquesta situació, la microesfera experimenta una força restauradora que tendeix a empènyer-la a la regió de màxima intensitat de llum, és a dir, al focus del làser.

En funció de la relació entre la grandària de la microesfera (és a dir, el diàmetre d) i la longitud d'ona de la radiació làser (λ) es poden diferenciar tres aproximacions teòriques que descriuen l'atrapament de la microesfera. És important aclarir que les tres situacions corresponen a casos particulars de les equacions de Maxwell. En el règim de Mie, la partícula és molt més gran que la longitud d'ona del làser ($d \gg \lambda$) i la interacció es pot estudiar a partir de les lleis de l'òptica geomètrica i el traçat de rajos (vegeu la Fig. 2). En el règim de Rayleigh passa justament el contrari ($d \ll \lambda$) i la microesfera es tracta com si fos un dipol elèctric puntual que interactua amb un camp electromagnètic. Per estudiar la interacció en el règim d'entremig ($d \approx \lambda$), es fa ús de la teoria generalitzada de Lorentz-Mie [15], en la que cal resoldre un problema d'equacions diferencials amb condicions de contorn.

És suficient tenir en compte el règim de Mie per tal de dissenyar un sistema de mesura de la força. Tot i que els experiments rellevants en biofísica no pertanyen exactament a aquest règim sinó al règim de Lorentz-Mie, s'ha pogut comprovar empíricament que hi ha sistemes experimentals de mesura de força que pràcticament no depenen de la grandària relativa entre la microesfera i la longitud d'ona del làser.

0.2.2 Muntatge experimental de les Minipinces

Les minipinces són un instrument de pinces òptiques dissenyat per Steve Smith [16, 17, 18]. Es tracta d'un prototip compacte de dimensions reduïdes (uns 30 cm de diàmetre i 40 cm d'alçada) que es pot penjar al sostre amb una corda elàstica per tal de reduir les vibracions mecàniques de l'edifici que afecten a les mesures (vegeu la Fig. 3).

Òptica

La trampa òptica es forma mitjançant dos làsers contra-propagants focalitzats en el mateix punt amb objectius de microscopi (vegeu la Fig. 4). Els

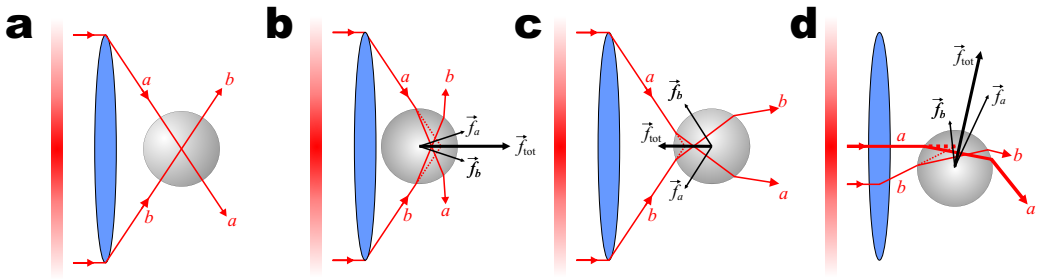


Figura 2: Atrapament òptic en el règim de Mie amb un feix làser gaussià focalitzat per una lent convergent. **(a)** Microesfera centrada al focus del feix gaussià. La microesfera no desvia els rajos marginals a and b i no s'exerceix força. **(b)** Microesfera localitzada en una posició anterior al focus del làser. Els rajos marginals a i b es desvien de tal manera que les forces radials es cancel·len i les forces axials se sumen. La llum empeny la microesfera cap al focus del feix. **(c)** Microesfera localitzada en una posició posterior al focus del làser. La força resultant és en sentit oposat a la del panell b. En aquest cas també està dirigida cap al focus del làser. **(d)** Microesfera localitzada fora d'eix. Les diferències d'intensitat entre els rajos centrals i els marginals indueixen una força de recuperació radial dirigida cap al focus del feix.

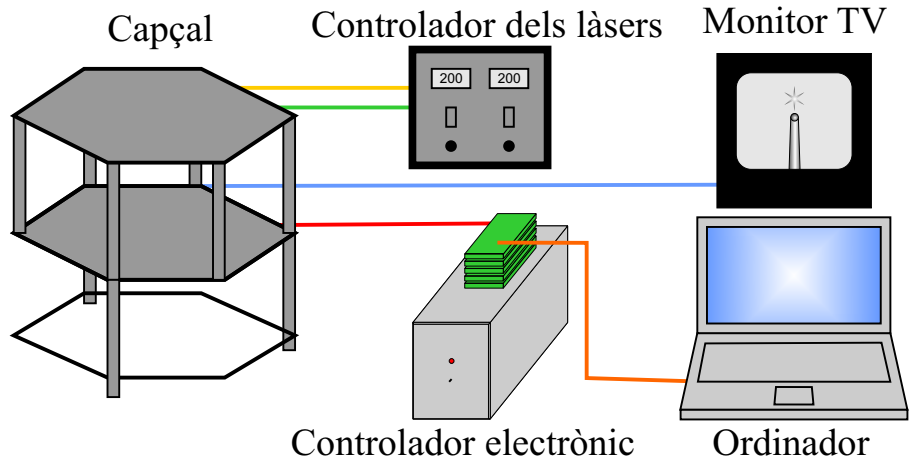


Figura 3: Esquema general de les minipinces. Els experiments es realitzen al capçal, que conté tota l'òptica i es pot penjar al sostre. Els làsers són alimentats pel controlador dels làsers. El controlador electrònic comunica el capçal i l'ordinador. L'usuari controla l'experiment a través de l'ordinador. El monitor de TV permet visualitzar l'experiment.

camins òptics d'ambdós làsers són simètrics. Les forces es mesuren a partir del principi de conservació del moment lineal de la llum. L'estratègia consisteix a recollir tota la llum dispersada per la microesfera i a analitzar la desviació que ha experimentat mitjançant fotodetectors sensibles a la posició (PSD). La diferència entre la intensitat de llum incident sobre la microesfera i la llum emergent dóna una mesura directa de la força que exerceix la trampa

òptica sobre la microesfera. La trampa òptica es pot posicionar en qualsevol lloc de l'espai en un rang de 10 μm sobre el pla focal mitjançant un dispositiu anomenat *redireccionador*[1]. El redireccionador consisteix en una fibra òptica que pot ser flexionada mecànicament mitjançant actuadors piezoelèctrics. La posició de la trampa es mesura gràcies a una palanca de llum, que reenvia part de la llum làser a un fotodetector sensible a la posició. La trampa òptica es forma dins d'una cambra fluídica construïda a partir de cobreobjectes de microscopi. És a l'interior d'aquesta cambra on es duen a terme els experiments d'estirament de biomolècules. La cambra conté una micropipeta que serveix de punt d'ancoratge per a fixar-hi les molècules que se sotmeten a tensió amb la trampa òptica. La imatge de l'interior de la cambra es projecta sobre una càmera CCD d'acord amb un sistema d'il·luminació Köhler, per tal que es pugui visualitzar l'experiment que s'està duent a terme.

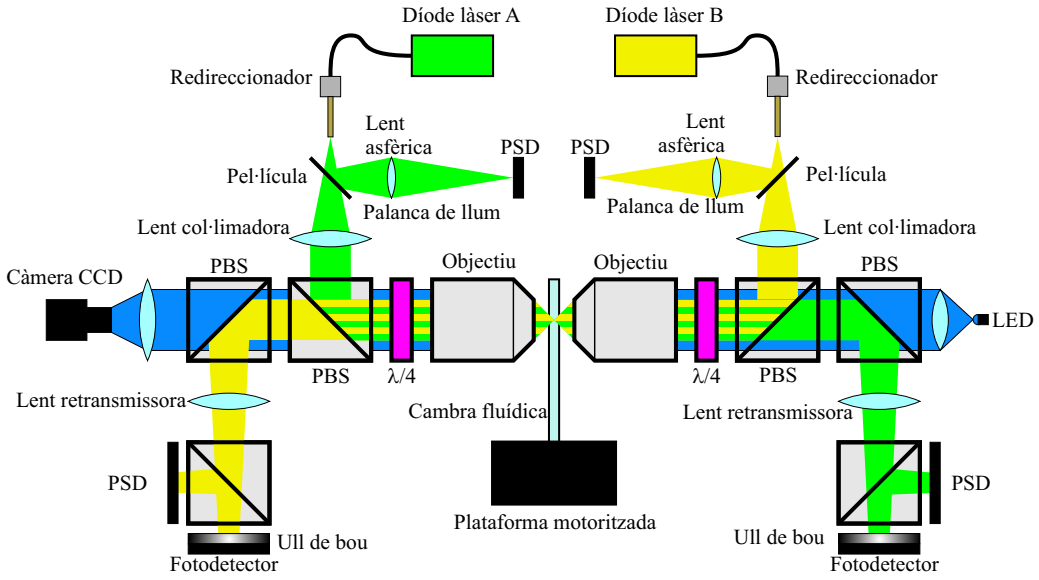


Figura 4: Sistema òptic de les minipinces (vegeu el text). Els camins òptics dels dos làsers estan dibuixats en verd i en groc. El camí òptic del sistema d'imatge està dibuixat en blau.

Un dels principals avantatges de les minipinces és que el calibratge de la força és independent del diàmetre de la microesfera, el seu índex de refracció, l'índex de refracció del medi que conté la cambra i la potència del làser. Això facilita molt els experiments, ja que només és necessari calibrar l'instrument una vegada. A més, el fet de disposar de dos làsers permet realitzar experiments amb doble trampa. Per altra banda, és necessari tenir un molt bon

[1] En anglès, *wiggler*.

alineament entre els dos làser per tal que el sistema funcioni correctament, la qual cosa dificulta la posada a punt de l'instrument.

Controlador electrònic

El controlador electrònic s'encarrega de l'adquisició de les mesures (força i localització de la trampa) i del processament de les ordres enviades per l'experimentador (posicionament de la trampa i de la cambra fluídica). El controlador està format per cinc plaques electròniques regulades per microcontroladors PIC (vegeu la Fig. 5). L'instrument té una resolució temporal d'1 kHz, 0.1 pN en força i 0.5 nm en distància. La placa principal s'encarrega de la transmissió de dades a l'ordinador (mitjançant un port USB) i a les altres plaques electròniques. Les dues plaques que converteixen la senyal analògica a digital (una per a cada làser) s'encarreguen de rebre les mesures de l'instrument i enviar-les a la placa principal. La placa convertidora de senyal digital a analògica s'encarrega de rebre les ordres de la placa principal i enviar-les als piezoelèctrics que controlen el redireccionador. La placa dels motors s'ocupa del posicionament de la cambra fluídica.

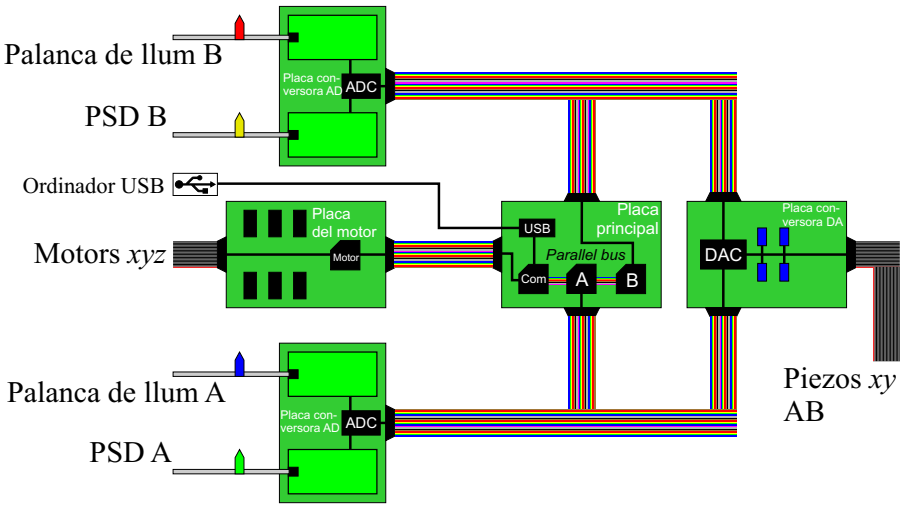


Figura 5: Esquema de les plaques electròniques (vegeu text).

Ordinador

L'ordinador té dues funcions. En primer lloc, s'encarrega de rebre, processar i emmagatzemar les dades adquirides per la placa principal del controlador electrònic. Per altra banda, l'ordinador és la interfície entre l'experimentador i l'experiment. L'ordinador rep les ordres de l'usuari i les envia a la placa del

controlador. En un experiment, s'executa l'aplicació 'lt', que són les inicials de *Laser Tweezers* (pinces òptiques en anglès). Aquesta aplicació va ser dissenyada per Shane Saxon [19]. La figura 6 mostra una captura de pantalla de la interfície gràfica.

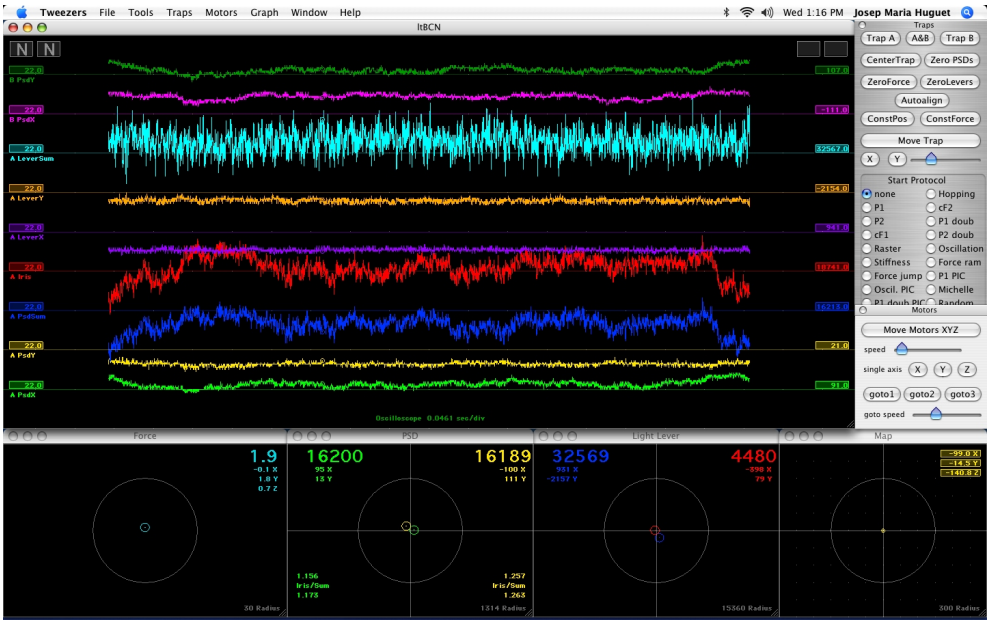


Figura 6: Interfície gràfica de l'aplicació lt. L'experimentador interacciona amb l'instrument mitjançant aquesta interfície.

0.2.3 Calibratge de l'instrument

Hi ha tres magnituds principals que cal calibrar en les minipinces: el temps, la força i la distància. La resta de magnituds es calibren a partir d'aquestes tres. El temps es calibra a partir del rellotge intern dels microcontroladors de les plaques electròniques. N'hi ha prou amb mesurar amb l'oscil·loscopi el temps que tarda el microcontrolador a completar un cicle del seu programa intern. Això dóna el factor de calibratge del temps. La força es pot calibrar per diferents mètodes. El primer mètode consisteix a utilitzar el principi de conservació del moment lineal de la llum. Sabent les característiques dels fotodetectors i mesurant la intensitat del feix làser es pot deduir el factor de calibratge. El segon mètode consisteix a utilitzar la llei de Stokes, que relaciona la força d'arrossegament d'una microesfera amb la velocitat del fluid que la rodeja. El tercer mètode es basa en les fluctuacions de força que experimenta una partícula browniana sotmesa a una força de recuperació. I el darrer mètode consisteix a utilitzar un patró. Per exemple, és sabut

que l'ADN presenta una transició estructural cooperativa a 67 pN de força. El calibratge de la distància de les palanques de llum es fa a partir de les especificacions de la plataforma motoritzada. El calibratge de l'instrument té error d'un 3% aproximadament.

0.2.4 Disseny de nous protocols

Les minipinces estan construïdes de tal manera que es poden dissenyar nous experiments o protocols d'estirament. D'aquesta manera es pot ampliar el ventall de possibilitats que ofereix l'instrument per tal d'estudiar aspectes més específics de les biomolècules. Al llarg de la realització d'aquesta tesi s'han dissenyat diversos nous protocols com el protocol d'oscil·lació (que aplica una força oscil·latòria) o la rampa de força (que augmenta progressivament la força aplicada).

0.2.5 Conclusions

D'ençà del descobriment de les pinces òptiques, aquesta tècnica ha evolucionat fins a esdevenir una eina clau en biofísica. La descripció de la interacció entre el làser i la microesfera es pot fer a partir de diverses aproximacions. Les minipinces són un aparell de pinces òptiques que mesura la força a partir del principi de conservació del moment lineal. L'instrument té un procés de calibratge ràpid i presenta una alta resolució en força i distància. A més, està dissenyat de tal manera que permet implementar nous tipus experiments.

0.3 Ruptura mecànica de l'ADN

El descobriment de l'estructura en doble hèlix de l'ADN l'any 1953 va suposar la culminació d'un segle de recerca [1]. Finalment es va poder establir la base físico-química de la genètica mendeliana. En síntesi, es va acceptar que la molècula d'ADN era la portadora de la informació genètica i que aquesta estava codificada en forma de seqüència de bases. L'any 1958, Francis Crick va formular el dogma central de la biologia molecular, que establia la forma com es replicava i es transferia la informació genètica de l'ADN a les proteïnes [20].

0.3.1 Estructura de l'ADN

L'ADN és un polímer format per la unió de dues cadenes complementàries de nucleòtids (vegeu la Fig. 7). Cadascun dels nucleòtids està format per una molècula d'àcid fosfòric, una molècula de desoxiribosa i una base nitrogenada (que pot ser adenina, citosina, guanina o timina). Una cadena aïllada presenta l'aspecte d'un esquelet (format per àcid fosfòric i desoxiribosa) en el qual hi ha enllaçades les bases nitrogenades. L'altra cadena de l'ADN té la mateixa estructura i està orientada en sentit contrari. Les bases d'una cadena s'enllacen amb les de la cadena complementària per ponts d'hidrogen. L'adenina sempre s'enllaça amb la timina i la citosina, amb la guanina. Tot i això, també s'han observat altres combinacions.

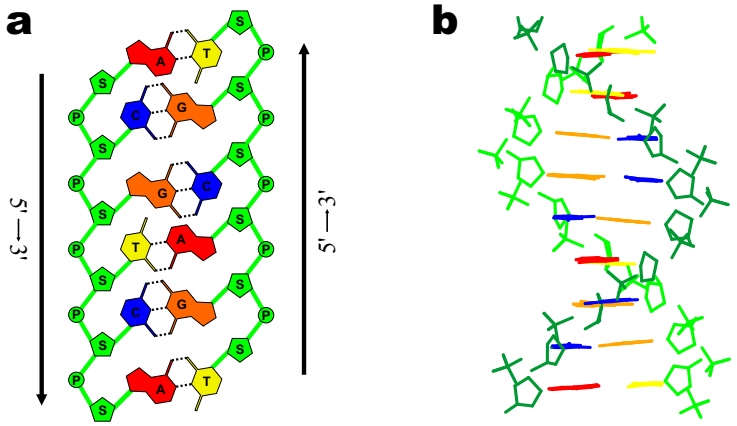


Figura 7: Estructura de l'ADN. **(a)** Representació esquemàtica de la composició de l'ADN. Hi ha dues cadenes antiparal·leles $5' \to 3'$ i $3' \to 5'$ enfrontades. Cada cadena està formada per un esquelet (dibuixat en verd) de sucres (S) i fosfats (P). Enmig hi ha les bases nitrogenades (A, C, G, T). Les línies contínues representen enllaços covalents, mentre que les discontínues, ponts d'hidrogen. **(b)** Estructura en doble hèlix de l'ADN. El dibuix té el mateix codi de colors que el panell a.

La informació genètica està codificada en la seqüència de bases. Cada tres bases forma un codó, que es tradueix en un aminoàcid. Els aminoàcids són els constituents de les proteïnes, que són les biomolècules encarregades de donar l'estructura i la funció a les cèl·lules. D'aquesta manera, llegint la seqüència de bases de l'ADN la cèl·lula coneix la seqüència d'aminoàcids que ha de sintetitzar per a produir una proteïna concreta.

0.3.2 Ruptura mecànica de l'ADN

Per tal d'accedir a la informació genètica emmagatzemada a l'ADN, la cèl·lula ha de separar les dues cadenes d'ADN, exposar la seqüència de bases i interaccionar amb elles per tal de llegir-les. El procés de separació de les dues cadenes implica una ruptura mecànica de les bases. Efectivament, les bases estan enllaçades per ponts d'hidrogen i cal aplicar força per trencar aquests enllaços. El procés de ruptura és reversible, de manera que si s'han separat les dues cadenes d'una molècula d'ADN es poden tornar a unir i formar de nou la doble hèlix. La cèl·lula disposa de diverses proteïnes (helicassa, ADN-polimerassa) encarregades d'obrir la doble cadena. Per tant, la separació de les dues cadenes d'ADN és un procés molt habitual dins de la cèl·lula.

La ruptura mecànica de l'ADN s'ha pogut aconseguir *in vitro* mitjançant diverses tècniques experimentals, com ara les microagulles [21], el microscopi de força atòmica [22], les pinces magnètiques [23] i les pinces òptiques [24, 25]. En aquestes tècniques s'aplica força al extrems de la cadena de l'ADN per tal de separa-les.

Els experiments de ruptura mecànica d'ADN exposats en aquesta tesi s'han realitzat mitjançant pinces òptiques. Per tal de dur a terme els experiments, s'han sintetitzat dues molècules de llargades diferents, una de 2252 parells de bases (pb) i una de 6828 pb. Les molècules tenen dues manetes als seus extrems que permeten ancorar-les a les microesferes amb què s'aplica força. Una microesfera està situada a la trampa òptica i l'altra es troba fixada a la micropipeta per succió d'aire (vegeu la Fig. 8).

Mesures experimentals de ruptura

Els experiments es poden dur a terme o bé controlant la posició de la trampa, o bé controlant la força aplicada sobre la molècula. En un experiment de ruptura, típicament es mesura la corba de força vs. distància (CFD). A distància controlada, es desplaça la trampa òptica, tot augmentant la separació entre les microesferes. A cada distància es mesura la força exercida (vegeu la Fig. 9). La CFD resultant té un aspecte de dent de serra, formada per una successió de pendents i caigudes de força. Els pendents corresponen a la

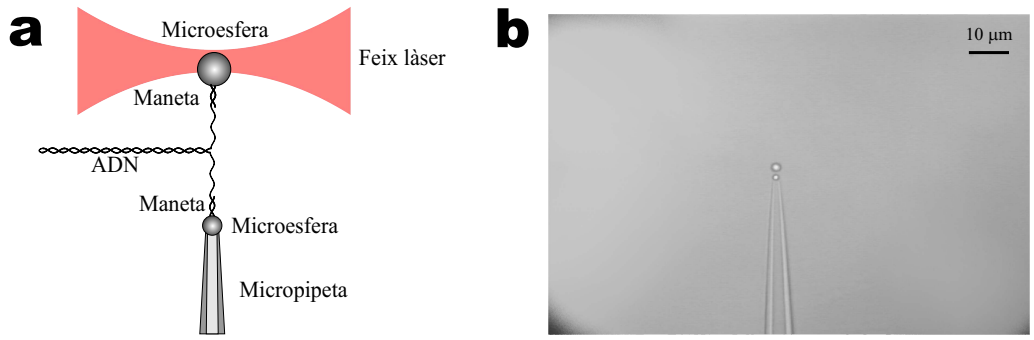


Figura 8: **(a)** Muntatge experimental per a produir la ruptura d'una molècula d'ADN. El dibuix no és a escala. **(b)** Aspecte que mostra l'experiment vist amb el monitor de TV.

resposta elàstica de la molècula, mentre que les caigudes de força apareixen quan s'obren (es trenquen) un grup de parells de bases. Si tot seguit es redueix de nou la distància entre microesferes, les bases es tornen a unir de nou i la molècula es torna a tancar. A força controlada, la CFD té un aspecte molt diferent. Quan la força aplicada s'acosta a la força de coexistència (al voltant d'uns 15 pN), la molècula d'ADN s'obre sobtadament i completa (vegeu la Fig. 10).

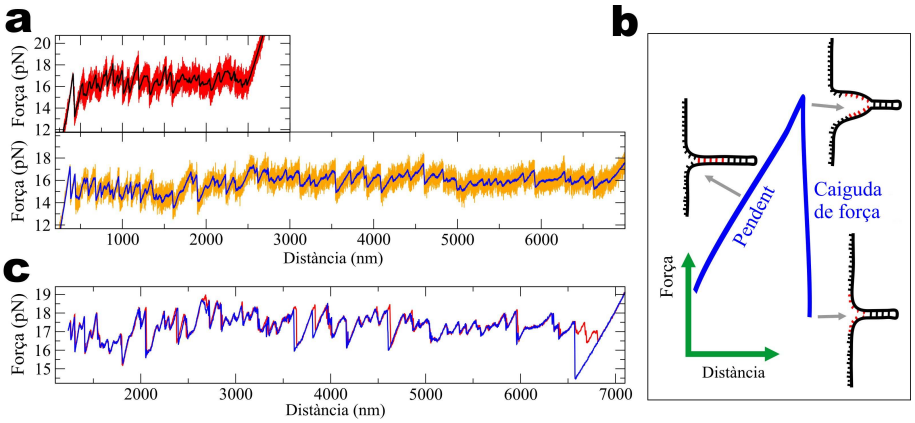


Figura 9: Ruptura d'una molècula d'ADN a distància controlada. **(a)** Patró en forma de dent de serra de la CFD. El panell superior (inferior) mostra la CFD per una molècula de 2.2 kpb (6.8 kpb). Les corbes vermella i taronja mostren les dades tal i com són mesurades per l'instrument. Les corbes negra i blava mostren les dades filtrades amb un ample de banda d'1 kHz. **(b)** Esquema que mostra la relació entre la dent de serra i la situació en què es troba la molècula d'ADN. El pendent es correspon amb la resposta elàstica i la caiguda de força amb la ruptura de parells de bases. El grup de bases que s'han obert estan dibuixades en vermell. **(c)** Cicle complet d'estirament i relaxació. La corba vermella (blava) mostra la ruptura (reenllaçament) dels parells de bases.

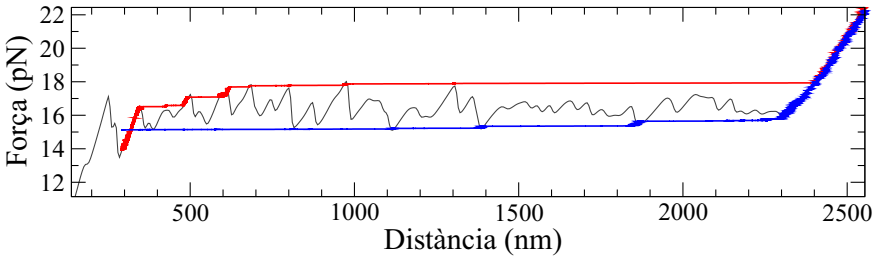


Figura 10: Ruptura d'una molècula d'ADN a força controlada. La corba vermella (blava) mostra la CFD de ruptura (reenllaçament). S'ha superposat en gris la CFD a distància controlada.

0.3.3 Modelització de l'experiment

Els experiments de ruptura mecànica de l'ADN realitzats amb pinces òptiques es poden modelitzar a partir del model de primers veïns (PV), que descriu la reacció d'hibridació de cadenes complementàries d'ADN [26, 27]. Els elements elàstics del sistema (trampa òptica, manetes, cadena senzilla) es modelitzen d'acord amb la llei de Hooke i els models sobre l'elasticitat dels polímers. La figura 11 mostra un esquema representatiu d'aquest model mesoscòpic. El model permet calcular una predicció de la CFD en equilibri termodinàmic. Aquest càlcul es realitza a partir de l'expressió de l'energia total del sistema i el càlcul de la funció de partició. La derivada de la funció de partició dóna l'expressió de l'equació d'estat. El model també permet calcular el paisatge d'energia lliure, que resulta molt útil per entendre el mecanisme de ruptura de l'ADN. El paisatge d'energia lliure és una expressió que dóna l'energia total del sistema en funció de la posició de la trampa òptica i del nombre de bases obertes.

0.3.4 Conclusions

L'ADN és probablement la biomolècula més rellevant que es troba en els sistemes biològics. La seva estructura és apropiada per a emmagatzemar la informació genètica de la cèl·lula, ja que els parells de bases que codifiquen les proteïnes estan aparellades i protegides a l'interior de l'esquelet de sucres i fosfats. Per tal d'accedir a les bases, la cèl·lula ha de separar la doble cadena. Aquest procés es pot produir artificialment mitjançant pinces òptiques. Els experiments típicament mesuren CFD que tenen una forma en dent de serra que depèn de la seqüència de bases de la molècula. El model mesoscòpic explicat en aquest capítol permet fer prediccions sobre la CFD mesurada experimentalment.

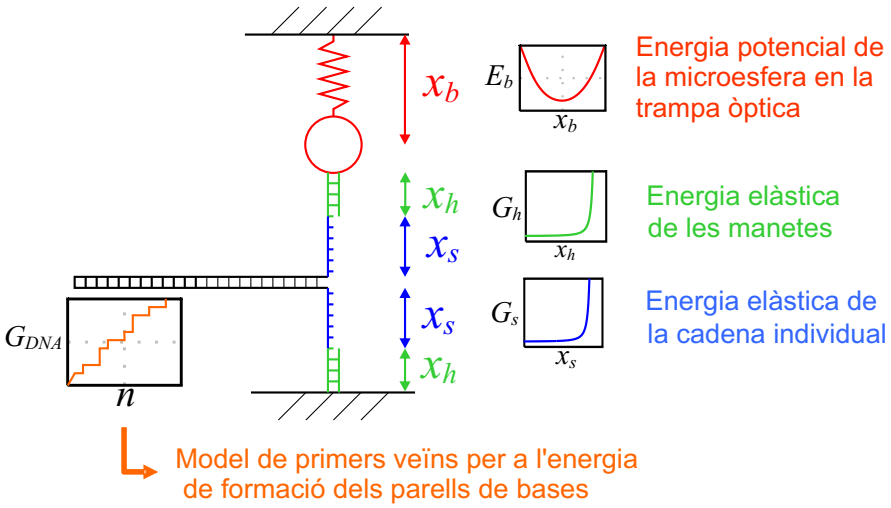


Figura 11: Model mesoscòpic. Cada element es representa amb un color diferent. Al costat de cadascun, es mostra un esbós de la seva contribució energètica

0.4 Intermediaris metastables en la ruptura mecànica de l'ADN

El procés de ruptura de l'ADN s'engloba dins la categoria de fenòmens de fractura, com per exemple els terratrèmols, la magnetització de dominis ferromagnètics o l'esquinç d'un full de paper. Aquests fenòmens presenten la característica dent de serra quan es trenquen, pròpia d'un procés d'acumulació d'energia elàstica que s'allibera sobtadament durant la ruptura. En l'ADN, cada dent de serra representa la ruptura d'un grup de parells de bases que són de diverses grandàries. Aquests trencaments s'anomenen regions cooperatives de ruptura (RCR). Després d'un trencament, la molècula es troba relaxada fins que es torna a augmentar de nou la força. Així, la ruptura de l'ADN amb pinces òptiques es pot entendre com una successió d'estats intermediaris que cada vegada tenen més parells de bases obertes, fins que la molècula té les dues cadenes completament separades.

0.4.1 Detecció d'estats intermediaris

A partir de les CFD i utilitzant el model mesoscòpic descrit en la secció 0.3.3, es pot inferir quantes bases obertes té cadascun dels estats intermediaris que apareixen en un procés de ruptura. Això es fa mitjançant una aproximació bayesiana, en què s'analitzen les dades experimentals d'acord amb un criteri probabilístic. L'estratègia consisteix a determinar per a cada punt experimental (donat per una distància i una força) el nombre de bases obertes més probable que pot tenir la molècula, d'acord amb la resposta elàstica que s'espera del sistema (vegeu la Fig. 12).

Aquest procés de classificació es repeteix per tots els punts de la CFD. D'aquesta manera es pot obtenir un histograma que indica el nombre de vegades que la molècula es troba en un estat intermediari concret al llarg de tot el procés de ruptura (vegeu la Fig. 13). L'histograma mostra una sèrie de pics que indiquen els estats intermediaris. Com més alt és el pic, més visible i estable és l'estat intermediari. Cada pic pot ajustar-se a una gaussiana o un conjunt de gaussianes que permeten establir el nombre de pb de l'estat intermediari. La distància entre els pics de les gaussianes dóna una mesura del nombre de pb que s'han obert entre un estat i l'altre, és a dir, la grandària d'una RCR. Després d'identificar totes les RCR de la molècula es pot construir un histograma de la distribució de grandàries (vegeu la Fig. 14).

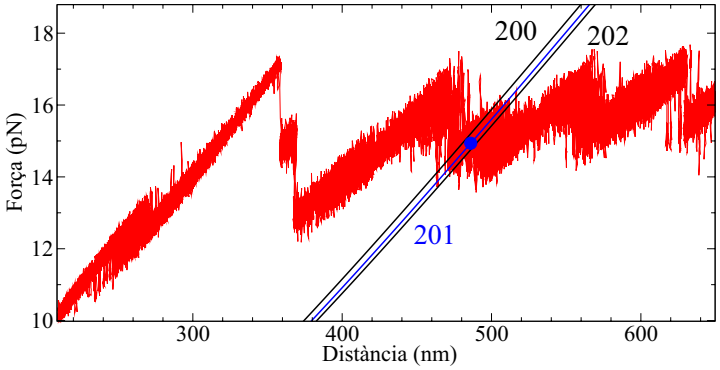


Figura 12: Classificació d'un punt experimental (dibuix no a escala). Per a punt experimental (en blau) de la CFD de ruptura (en vermell), triem la resposta elàstica que passa més a prop del punt experimental. En aquest cas particular, el més probable és que la molècula tingui 201 bases obertes a la força i distància indicada pel punt experimental, ja que és la resposta elàstica que passa més a prop del punt. Altres respostes elàstiques (200 i 202) ja estan més allunyades del punt blau.

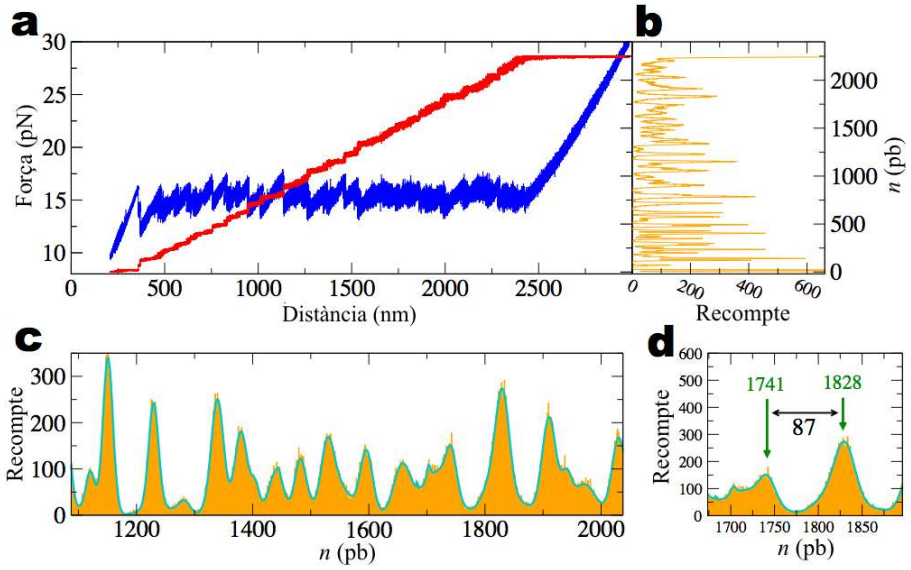


Figura 13: Histograma d'estats intermediaris. **(a)** Classificació de punts. La traça blava mostra la CFD experimental. La traça vermella mostra el nombre de parells de bases obertes (n) corresponent a cada punt experimental (l'eix de les y es llegeix al panell b). **(b)** Histograma dels valors de n^* mostrats al panell a. **(c)** Visualització detallada de l'histograma (corba taronja) sobreposada amb l'ajust a la suma de gaussianes (corba cian). **(d)** Detecció d'una RCR de 87 pb de grandària a partir de la distància entre els pics de dues gaussianes consecutives.

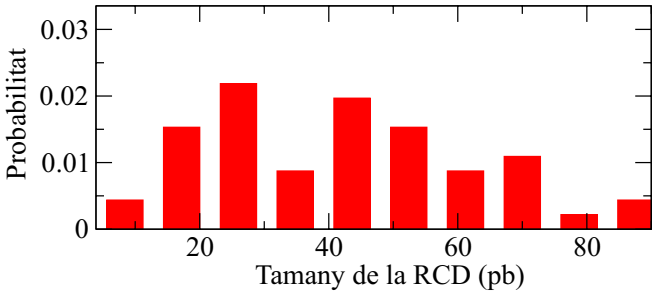


Figura 14: Distribució de grandàries de les RCR d'una molècula de 2.2 kpb. Les grandàries tenen un rang de 10-90 pb amb una gran quantitat de RCR que tenen entre 20 i 50 pb.

0.4.2 Model de joguina

Per tal d'interpretar els resultats de la distribució de grandàries de les RCR, hem desenvolupat un model simplificat al model mesoscòpic descrit en la secció 0.3.3. Aquest model conté els elements mínims per a reproduir qualitativament les propietats estadístiques de la ruptura de l'ADN (vegeu la Fig. 15). A més, els càlculs són més senzills i ràpids i permeten explorar moltes molècules amb seqüències diferents.

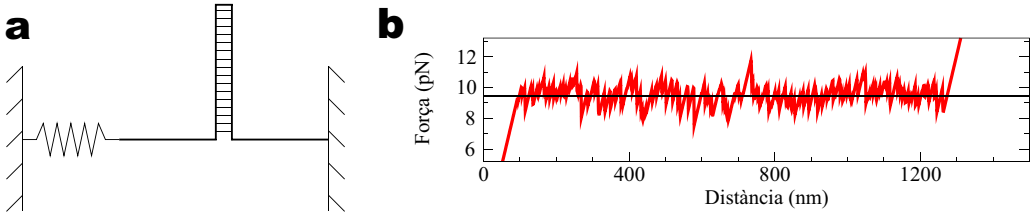


Figura 15: Model de joguina. **(a)** Esquema. El model només contempla la trampa òptica i l'energia d'hibridació de l'ADN. **(b)** CFD predita pel model, per a una seqüència de pb arbitrària. La força mitjana de ruptura està representada amb una línia negra.

0.4.3 Comparació de les distribucions de grandària de les RCR

La figura 16 mostra la comparació entre les mesures experimentals de les grandàries de les RCR, la predicció del model mesoscòpic de la secció 0.3.3 i la predicció del model de joguina. A la vista dels resultats, es pot comprovar com experimentalment es detecta menys quantitat de RCRs de grandària inferior als 10 pb que la quantitat predita pels models. Això indica que el muntatge experimental de les minipinces té una resolució de 10 pb. Per altra

banda, ambdós models descriuen bé la distribució de grandàries en el cas de RCR grans.

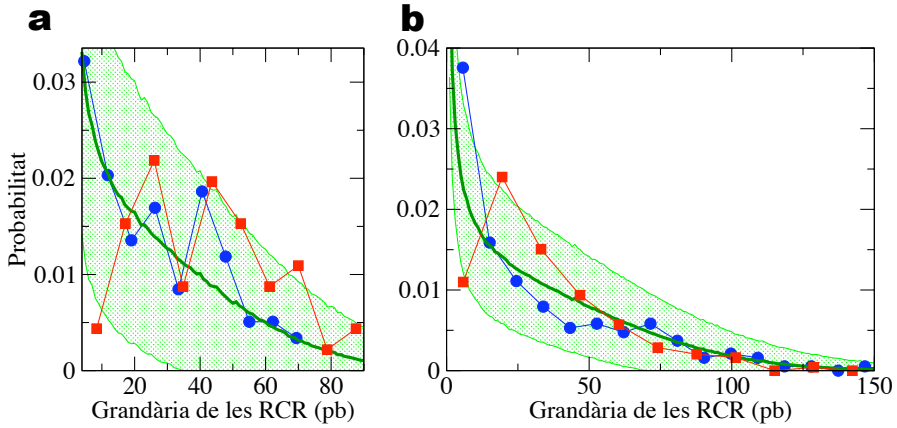


Figura 16: **(a)** Distribució de les grandàries de les RCR per a la seqüència de 2.2 kpb. La corba vermella mostra les mesures experimentals. La corba verda mostra la predicció del model de joguina (la zona ombrejada mostra la desviació estàndard per diferents realitzacions). La corba blava mostra la distribució predita pel model mesoscòpic. **(b)** Mateixa figura per a la seqüència de 6.8 kpb.

Una de les qüestions interessants que sorgeixen és saber quines condicions experimentals es poden canviar per tal de modificar la distribució de grandàries. El model de joguina prediu que quan s'augmenta la rigidesa de la trampa òptica disminueix la grandària mitjana de les RCR. A nivell quantitatiu, això indica que per tal d'observar RCR d'un sol pb seria necessari tenir una rigidesa de la trampa òptica de 100 pN/nm. Les minipinces tenen una rigidesa de 0.08 pN/nm. Per tant estem molt allunyats d'aquest règim i actualment no seria factible obtenir tal rigidesa amb pinces òptiques. Una rigidesa de 100 pN/nm és típica del microscopi de força atòmica. Curiosament, la mínima rigidesa de la trampa necessària per obrir els pb de l'ADN d'un en un es correspon amb el valor de la rigidesa que té una sola base d'ADN de cadena individual. Això suggereix que l'ADN té les propietats elàstiques adequades per tal que la maquinària cel·lular pugui produir la ruptura d'un únic pb de l'ADN si és necessari. Així s'evita obrir grans quantitats de bases que poden estar exposades a danys que alterin la informació genètica que conté l'ADN.

0.4.4 Conclusions

La ruptura de molècules d'ADN a distància controlada presenta una sèrie d'obertures cooperatives dels parells de bases. Mitjançant un mètode bayesià

es pot determinar el nombre de parells de bases obertes dels estats intermediaris i la grandària de les regions de ruptura. La resolució experimental de les minipinces situa el límit de detecció a 10 pb, per sota del qual no es poden detectar obertures de parells de bases. El model de joguina prediu aquestes obertures i fixa a 0.1 N/m la rigidesa mínima que ha de tenir la trampa òptica per tal que la molècula d'ADN s'obri d'una base en una. Aquest valor coincideix amb la rigidesa d'una sola base de cadena individual d'ADN.

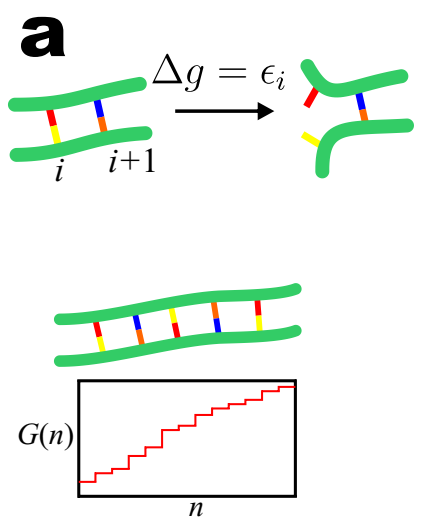
0.5 Dependència amb la sal de les energies lliures de formació dels parells de bases veïns

El model de primers veïns (PV) descriu acuradament les energies de formació de reaccions d'hibridació d'oligonucleòtids (vegeu la Fig. 17). El model estableix que l'energia de formació d'un parell de bases depèn del propi parell de bases en qüestió i del primer veí. Tenint en compte que hi ha 4 tipus de bases, en total s'esperaria que hi hagués $4 \times 4 = 16$ possibles combinacions de primers veïns. Això no obstant, degut a diverses simetries del problema, el nombre de combinacions de primers veïns es redueix a 10. L'any 1998, John SantaLucia Jr. va establir els valors d'aquestes 10 energies de formació a partir d'experiments de desnaturalització tèrmica de molècules d'ADN [28]. Des de llavors, aquests valors s'han conegut com els valors unificats d'oligonucleòtids (UO). Aquests valors es poden utilitzar en el model mesoscòpic descrit en la secció 0.3.3 per tal de predir la CFD d'una seqüència donada. Si bé la predicció de la forma de dent de serra de la CFD és qualitativament correcta, a nivell quantitatiu s'observen diferències notables quan es compara amb la mesura experimental de la CFD mesurada amb les minipinces (vegeu la Fig. 18). La predicció de la CFD amb els valors UO sobreestima la força mitjana de ruptura i això s'observa a diferents concentracions de sal. L'objectiu d'aquest capítol és ajustar els 10 valors de les energies de formació dels parells de bases (EFPB) d'acord amb el model de PV per tal de descriure correctament les CFD.

D'acord amb el model mesoscòpic, la CFD no només depèn de les 10 EFPB, sinó que la rigidesa de la trampa, les propietats elàstiques de les manetes i la cadena individual també afecten la CFD. La rigidesa de la trampa de l'instrument es pot mesurar directament. Les propietats elàstiques de les manetes són poc rellevants, perquè són molt curtes i gairebé es comporten com barres rígides. Ara bé, la rigidesa de la cadena individual és important, ja que afecta notablement la predicció de la CFD. Per tant, abans de continuar amb la descripció de l'ajust, és necessari determinar les propietats elàstiques de la cadena individual.

0.5.1 Resposta elàstica de la cadena individual

La resposta elàstica de la cadena individual s'obté prenent mesures de la CFD d'una molècula d'ADN de cadena individual de 3000 bases de llargada. Els experiments s'han realitzat en un ampli rang de concentració de sal (10 mM-1 M [NaCl]) i les corbes obtingudes s'ajusten a models de polímers



b

5'—3' AA TT 3'—5'	5'—3' AC TG 3'—5'	5'—3' AG TC 3'—5'	5'—3' AT TA 3'—5'
5'—3' CA GT 3'—5'	5'—3' CC GG 3'—5'	5'—3' CG GC 3'—5'	5'—3' CT GA 3'—5'
5'—3' GA CT 3'—5'	5'—3' GC CG 3'—5'	5'—3' GG CC 3'—5'	5'—3' GT CA 3'—5'
5'—3' TA AT 3'—5'	5'—3' TC AG 3'—5'	5'—3' TG AC 3'—5'	5'—3' TT AA 3'—5'

Figura 17: Model de PV. **(a)** Formació de la cadena doble d'ADN. L'energia de formació (Δg) del parell de bases i depèn d'ell mateix i del primer veí $i+1$. Cada parell de bases contribueix a l'energia total de formació de la cadena doble d'ADN ($G(n)$). **(b)** Les 16 combinacions de primers veïns. Les 12 combinacions ombrejades són simètriques respecte l'eix antidiagonal. En conclusió, només hi ha 10 energies diferents (6 de simètriques i 4 de l'antidiagonal).

ideals, com ara el model de cadena lliurement unida (CLU)[2] i el model de comportament de cuc (CDC)[3]. La figura 19 mostra els resultats. A partir de 100 mM de [NaCl] les respostes elàstiques experimentals mostren una desviació respecte la predicció dels models ideals. Això és degut a la formació d'estructura secundària en la molècula d'ADN, és a dir, hibridacions no desitjades entre bases. Es tracta d'un fenomen esperat a alta sal. De totes maneres, ens interessa descriure el comportament ideal del polímer i per tant, els models ideals automàticament exclouen l'efecte de formació d'estructura secundària.

0.5.2 Ajust de les CFD

Un cop disposem de la resposta elàstica de la cadena individual, podem procedir amb l'ajust de les 10 EFPB. És important mencionar que l'ajust de les dades té en compte els efectes de deriva que es produeixen en l'instrument. Efectivament, els fluxos d'aire i canvis de temperatura del laboratori poden produir dilatacions en l'instrument que alteren lleugerament les mesures de distància. En l'ajust, s'introdueix una correcció de la deriva que permet

[2]En anglès, *Freely-Jointed Chain*.
[3]En anglès, *Worm-Like Chain*.

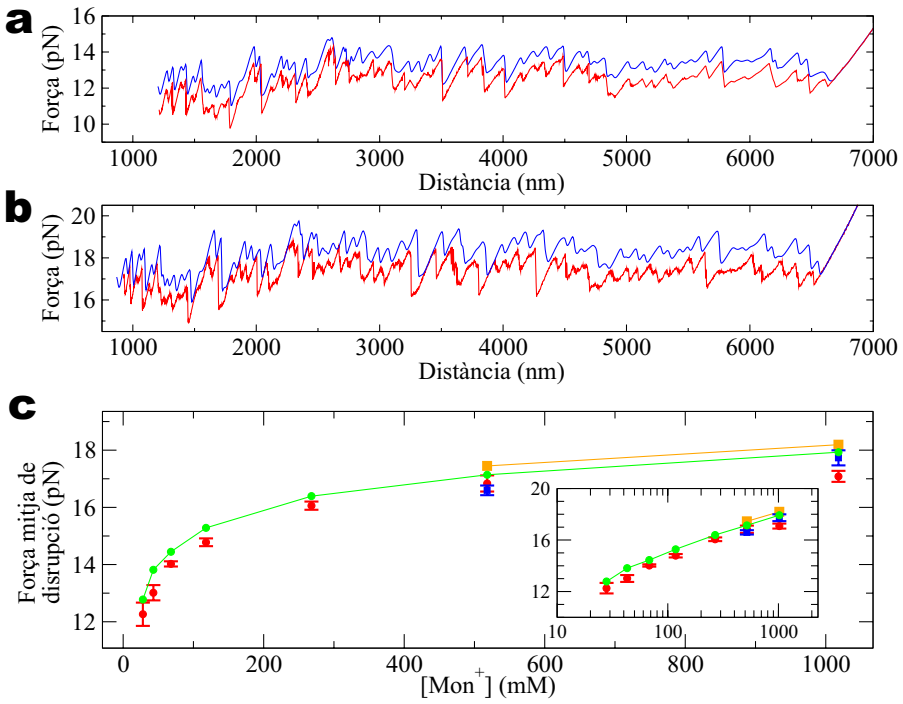


Figura 18: Discrepàncies entre els experiments de ruptura i la predicció del model de PV amb els valors UO. **(a)** CFD mesurada i filtrada amb un ample de banda d'1 Hz (corba vermella) i predicció amb els valors UO (corba blava) per a la molècula de 6.8 kpb a 10 mM [NaCl] de concentració de sal. **(b)** Seqüència de 6.8 kpb a 1 M [NaCl]. Mateix codi de colors que al panell a. **(c)** Dependència de la força mitjana de ruptura amb la concentració de sal. Punts vermells, mesures experimentals per a la molècula de 6.8 kpb; corba verda, predicció amb els valors UO; punts blaus, mesures per a la molècula de 2.2 kpb; corba taronja, predicció UO per a la molècula de 2.2 kpb. Els valors s'han obtingut a partir de la mitjana dels resultats per a sis molècules. Les barres d'error indiquen l'error en la desviació estàndard entre molècules.

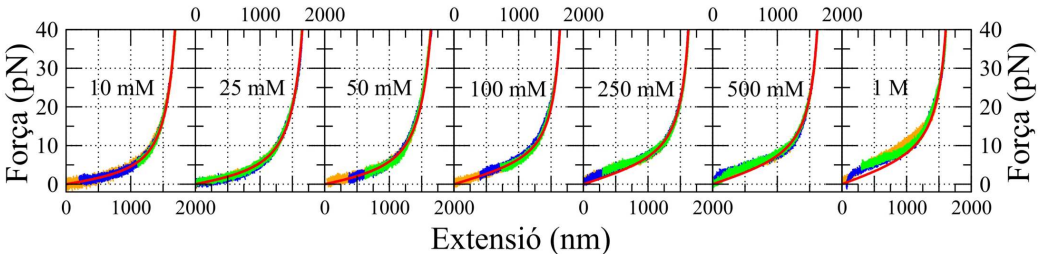


Figura 19: Ajust de la resposta elàstica de la cadena individual. Resposta elàstica d'una cadena de 3 kb a diferents concentracions de sal. Per a cada sal, es mostren les dades de tres molècules diferents (taronja, verd i blau). La corba vermella mostra el millor ajust. Els models són: CLU per [NaCl]$\leq$ 100 mM i CDC for [NaCl]$>$ 100 mM.

comparar millor les mesures experimentals i la predicció.

L'ajust es realitza minimitzant la diferència de força que hi ha entre les mesures experimentals i la predicció. Per a fer-ho, és necessari definir una funció error que es minimitza amb un algoritme Monte Carlo. Aquest algoritme consisteix a explorar aleatòriament diverses combinacions de les 10 EFPB fins a trobar la combinació que minimitza l'error comès. L'algoritme és robust, no depèn del valor inicial que es dóna a les 10 EFPB i dóna una solució que millora la predicció de les CFD. L'error de l'algoritme és d'aproximadament 0.05 kcal/mol per a cada energia de formació.

0.5.3 Dependència de les EFPB amb la sal

L'ajust es fa per a cadascuna de les CFD de totes les molècules i totes les condicions de sal. La figura 20 mostra els resultats obtinguts. D'acord amb els valors UO, la correcció de sal és homogènia. Això significa que s'aplica la mateixa correcció de sal per a totes les 10 combinacions de PV. Ara bé, si s'observen els resultats amb detall (vegeu la Fig. 20) es pot comprovar que una correcció homogènia no descriu correctament les nostres mesures. Per altra banda, una correcció heterogènia, en que cada una de les 10 combinacions de PV té la seva pròpia correcció de sal permet explicar correctament els resultats.

0.5.4 Predicció de les temperatures de desnaturalització d'oligonucleòtids

Per tal de comprovar els resultats que s'han trobat en la secció anterior, s'utilitza la correcció heterogènia per tal de predir les temperatures de desnaturalització d'oligonucleòtids. La temperatura de desnaturalització és aquella temperatura en la qual les dues cadenes d'un oligonucleòtid se separen. Aquesta temperatura depèn de la seqüència de l'oligonucleòtid i es pot predir a partir del model de PV. Per tant, predir temperatures de desnaturalització és una manera de comprovar que els nostres resultats són compatibles amb altres experiments.

Els resultats de la comparació mostren que amb la correcció heterogènia es prediuen correctament les temperatures de desnaturalització. A més, per a oligonucleòtids més llargs que 15 pb, la correcció heterogènia funciona millor que l'homogènia.

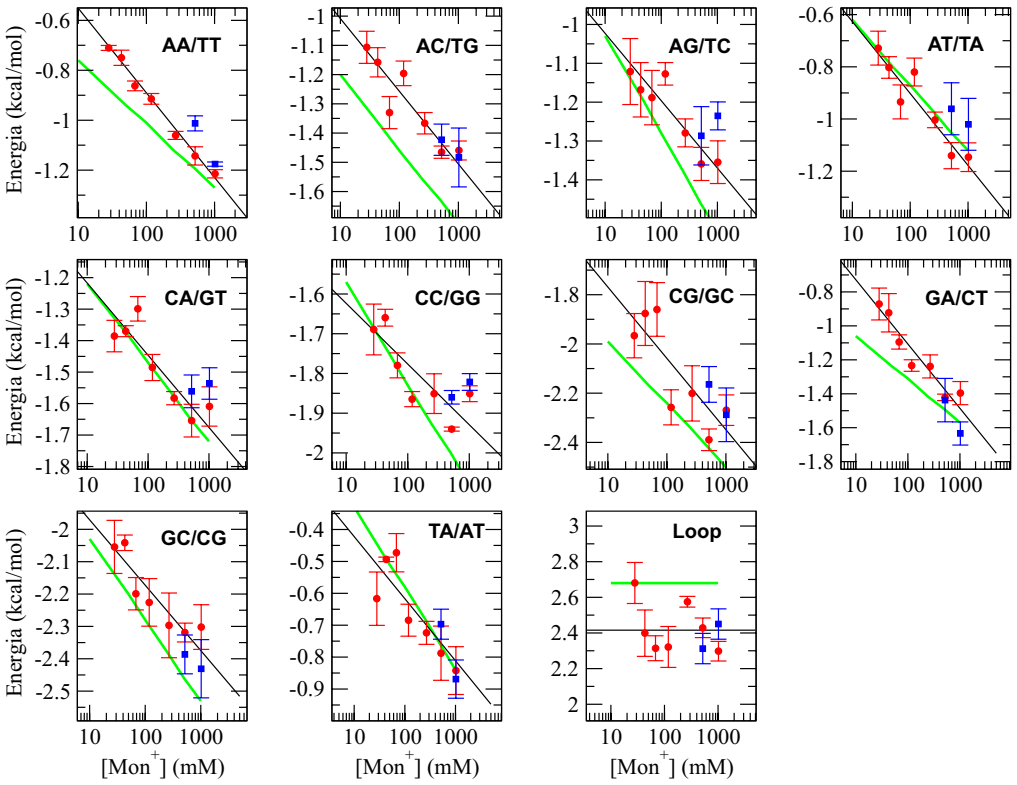


Figura 20: Correccions de sal de les EFPB. Cada panell mostra l'energia d'una combinació de PV. Els punts vermells (blaus) mostren els resultats per a la seqüència de 6.8 kpb (2.2 kpb); la corba verda mostra la correcció no específica dels valors de UO; la corba negra mostra l'ajust d'una correcció específica per a cada EFBP.

0.5.5 Conclusions

Els experiments de ruptura mecànica d'ADN proporcionen informació quantitativa sobre les propietats termodinàmiques de l'ADN. Les CFD experimentals poden ajustar-se al model mesoscòpic per tal d'obtenir els millors valors per a cadascuna de les 10 EFPB. L'ajust es realitza amb un algoritme Monte Carlo. Si aquest ajust es repeteix per a cada molècula i cada condició de sal es pot obtenir la dependència de les energies amb la concentració de sal. S'ha comprovat que una correcció heterogènia de sal descriu correctament els resultats i és compatible amb els resultats dels experiments de desnaturalització d'oligonucleòtids. La correcció heterogènia de sal pot ser deguda a la diferent solvatació dels parells de bases de l'ADN o bé a una dependència de la resposta elàstica de la cadena senzilla d'ADN amb la sal. Els resultats que es mostren aquí no permeten discernir entre una hipòtesi o l'altra.

0.6 Ruptura mecànica de l'ADN a força controlada

En la ruptura mecànica de l'ADN a força controlada s'utilitza un algoritme de realimentació, de tal manera que la posició de la trampa es corregeix contínuament per tal que la força exercida sobre la molècula tingui el valor desitjat. Les minipinces disposen d'un algoritme que corregeix la posició de la trampa a 4 kHz. Per tant, la força es pot mantenir constant dins d'un ample de banda de 4 kHz. Això no és suficient per anular les fluctuacions de força però, en mitjana, la força es manté constant.

Els experiments a força controlada consisteixen a augmentar progressivament la força aplicada sobre l'ADN fins que les dues cadenes d'ADN se separen. A continuació s'abaixa de nou la força aplicada fins que la molècula s'hibrida de nou. Els experiments es caracteritzen pel ritme de càrrega, és a dir, el ritme al que s'augmenta la força aplicada (en piconewtons per segon).

A diferència dels experiments a posició controlada, els experiments a força controlada presenten una gran irreversibilitat. Efectivament les corbes de ruptura i de reenllaçament no són iguals com en el cas de posició constant (vegeu la Fig. 9c). La figura 21a mostra diferents cicles d'estirament i relaxació a diferents ritmes de càrrega. A mesura que augmenta el ritme de càrrega la diferència entre l'estirament i la relaxació es fa més notable. L'àrea interior que deixen els dos processos equival al treball irreversible, és a dir, l'energia que es perd en forma de dissipació. La figura 21b mostra en escala logarítmica el treball dissipat en funció del ritme de càrrega. Resulta interessant comprovar com, a velocitats de càrrega molt baixes, la dissipació és encara significativa ($\sim 1000\ k_BT$). Això indica que el procés és molt irreversible a les escales de temps del laboratori. Per tal de tenir CFD més reversibles seria necessari reduir molt el ritme de càrrega de l'experiment.

Aquesta irreversibilitat té a veure amb les escales de temps. Efectivament, la força aplicada sobre la molècula d'ADN canvia molt més ràpidament que el temps que l'ADN necessita per acomodar-se a la nova força. Això fa que l'ADN no pugui assolir l'equilibri i presenti aquesta irreversibilitat. El paisatge d'energia lliure de la molècula permet comprendre el mecanisme que governa la ruptura. Efectivament, el paisatge d'energia lliure presenta barreres energètiques considerables, de manera que la molècula pot quedar atrapada en estats metastables. Les barreres disminueixen la seva alçada a mesura que augmenta la força fins al punt que la ruptura allibera gran quantitat d'energia de manera irreversible.

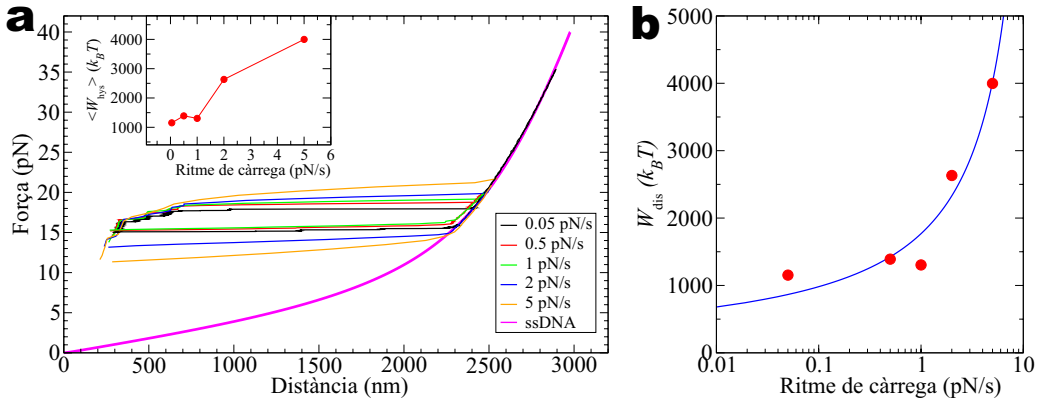


Figura 21: Experiments a força controlada. **(a)** Cicles d'estirament i relaxació a diferents ritmes de càrrega. **(Panell inserit)** Treball dissipat (area tancada pel cicle d'estirament i relaxació) vs. ritme de càrrega. **(b)** Histèresi en escala logarítmica. Mateix gràfic que el panell inserit en el panell a. La corba contínua tan sols és una guia visual.

0.6.1 Propietats d'escalament de la ruptura

L'any 2002, Lubensky i Nelson varen publicar un extens treball teòric sobre la ruptura de molècules d'ADN a força controlada [29]. En aquell article es feia una predicció del nombre de parells de bases que s'obrien a mesura que la força s'acostava al valor de la força de coexistència ($\sim$ 15 pN). Es tracta d'una relació en forma de llei de potències que està íntimament relacionada amb els fenòmens que no presenten una escala característica. En el cas concret que ens ocupa, aquesta relació indica que la ruptura de l'ADN no té una escala característica, ja el mecanisme d'obertura de la molècula no depèn de la seva llargada. És a dir, totes les seqüències, tinguin la llargada que tinguin, s'obren de la mateixa manera a mesura que ens acostem a la força de coexistència. A partir dels nostres experiments, es pot comprovar la validesa d'aquesta predicció (vegeu la Fig. 22). L'ajust és acurat i mostra que la predicció de Lubensky i Nelson és satisfactòria. En qualsevol cas, caldria comprovar els resultats per a la molècula de 6.8 kpb.

0.6.2 Conclusions

Encara que la força no és el paràmetre de control natural de les pinces òptiques, es poden realitzar experiments a força controlada mitjançant un algoritme de realimentació. Les CFD a força controlada tenen un aspecte molt diferent les CFD a distància controlada. Per altra banda, la ruptura a força controlada presenta una gran irreversibilitat i els cicles d'estirament i relaxació mostren una elevada histèresi. La molècula d'ADN no té temps suficient

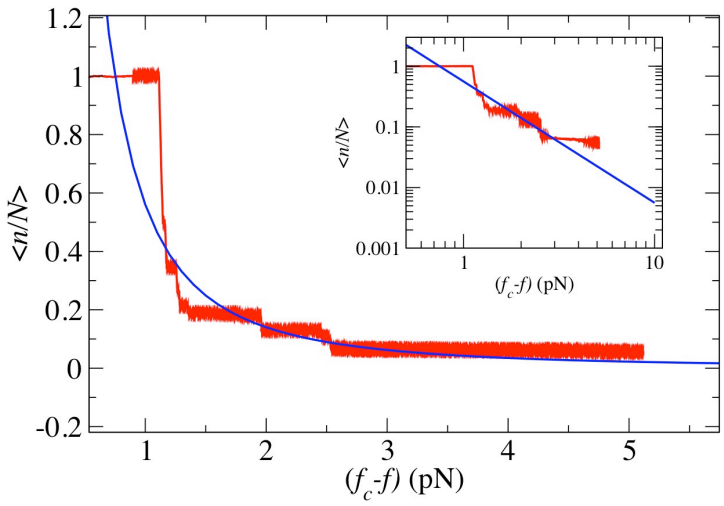


Figura 22: Propietats d'escalament a la transició de ruptura. La corba vermella mostra les dades experimentals obtingudes de la ruptura d'una cadena de 2.2 kpb a un ritme de càrrega de 0.05 pN/s. La corba blava és l'ajust de les dades a la funció predita per Lubensky i Nelson: $m = A/(f_c - f)^2$, on m $(0 < m < 1)$ és la proporció de bases obertes de la molècula, A és un paràmetre d'ajust i $f_c \simeq 19$ pN és la força crítica. El panell inserit mostra les mateixes dades en escala logarítmica.

per assolir l'estat d'equilibri en les escales de temps que dura l'experiment. El paisatge d'energia lliure permet entendre el mecanisme d'aquest fenomen. Finalment, s'ha comprovat que els experiments presenten les propietats d'escalament que foren predites en aquest tipus de sistema.

0.7 Perspectives futures i conclusions

La major part dels treballs científics s'inicien amb una pregunta. Per tal de respondre-la cal seguir el mètode científic, segons el qual cal recollir proves, avaluar críticament els resultats i formular conclusions. Sovint, en aquest procés sorgeixen qüestions i problemes no previstos que cal anar solucionant. Però en altres ocasions les qüestions queden obertes i cal posposar-les per a futures investigacions.

En aquesta tesi han sorgit noves preguntes. Algunes s'han hagut de resoldre per a prosseguir amb la investigació, com és el cas de la resposta elàstica de la cadena individual en el capítol 5. D'altres han resultat ser interessants però es desviaven del focus d'atenció. A continuació, s'enumeren algunes de les qüestions obertes més rellevants.

Com es poden obtenir les entalpies i entropies de formació dels parells de bases a partir dels experiments de ruptura?

Els experiments i l'anàlisi de dades del capítol 5 permeten obtenir les energies de formació dels parells de bases. Però els experiments de ruptura no donen informació sobre les entalpies i entropies d'hibridació. Per tal de mesurar aquestes quantitats, seria necessari repetir els experiments a una temperatura diferent. Això comporta seriosos reptes experimentals, ja que la temperatura produeix dilatacions en l'instrument que modifiquen el camí òptic i comprometen la qualitat de les mesures. Per tant, seria necessari dissenyar un sistema experimental que permeti controlar la temperatura. Mesurar l'entalpia i l'entropia pot resultar útil en aquelles situacions en què els experiments de volum (calorimetria, absorció ultraviolada) no són factibles perquè es produeixen interaccions entre molècules. Els experiments de molècula individual eviten aquests inconvenients.

Es pot resoldre un model analític que descrigui les RCR?

En aquesta tesi s'ha vist que el model mesoscòpic i el model de joguina donen bons resultats a l'hora de predir el comportament dels experiments de ruptura. Però aquests models no són analítics i cal resoldre'ls numèricament. Això passa perquè el model depèn fortament de la seqüència de bases de la molècula. Resoldre un model analític, que tingui en compte la mitjana sobre les seqüències, proporcionaria una millor comprensió del fenomen de ruptura. Aquest tipus de problemes resulten atractius per als físics teòrics que es dediquen a la física matemàtica.

La resposta elàstica de la cadena senzilla d'ADN depèn de la seqüència?

En el capítol 5 s'ha vist que una correcció de sal heterogènia descriu millor els resultats obtinguts. L'origen d'aquesta heterogeneïtat és incert. Alguns estudis suggereixen que la cadena senzilla d'ADN depèn de la seqüència, mentre que d'altres suggereixen que és la cadena doble. Respondre aquesta qüestió seria útil per tots aquells científics que treballen amb experiments que involucren cadenes senzilles d'ADN.

Com es pot seqüenciar l'ADN a partir dels experiments de ruptura?

Seqüenciar una molècula d'ADN consisteix a determinar la seqüència de bases que la formen. Per a seqüenciar, actualment s'utilitza un mètode bioquímic basat en el mètode de Sanger, que es va descobrir l'any 1975 [30]. Aquest mètode té l'inconvenient que perd eficiència per seqüències més llargues de 500-1000 pb. Els experiments de ruptura podrien, a priori, seqüenciar molècules d'ADN de grandària indefinida sense perdre eficiència. La raó és que el temps de ruptura d'una molècula és lineal amb la longitud d'aquesta. Però per tal d'aconseguir-ho, és necessari replegar la cadena senzilla d'ADN a mesura que es va alliberant i augmentar significativament la rigidesa de la trampa, tal i com s'ha establert en el capítol 4. La seqüenciació amb força és una aplicació interessant dels resultats establerts en aquesta tesi.

Conclusions

La irrupció de la nanociència i la nanotecnologia ha donat una nova empenta a la instrumentació. Això ha permès que les tècniques de molècula individual s'hagin desenvolupament de manera extraordinària durant la primera dècada del segle XXI. Aquest tipus d'experiments han beneficiat a la biofísica i han despertat l'interès de la comunitat científica pels sistemes biològics.

Les pinces òptiques és una tècnica que ha recollit gran fruits dins del camp de la biofísica. Aquest sistema experimental resulta molt atractiu per a realitzar experiments sobre biomolècules perquè es tracta d'una tècnica no invasiva que proporciona informació quantitativa. Les minipinces són un muntatge experimental compacte, precís i acurat. El seu calibratge és independent de la major part de paràmetres experimentals, cosa que facilita molt el treball dels experimentadors. A més, és un instrument versàtil de manera que es poden ampliar el tipus d'experiments que es poden realitzar.

La molècula d'ADN té un paper central en els éssers vius perquè s'encarrega d'emmagatzemar la informació genètica de les cèl·lules. Aquestes

necessiten separar les dues cadenes que formen la doble hèlix per tal d'accedir als parells de bases que codifiquen les proteïnes. La ruptura dels parells de bases es pot aconseguir *in vitro* mitjançant les minipinces. Les CFD obtingudes tenen forma de dent de serra i depenen de la seqüència de bases. El model mesoscòpic basat en el model de primers veïns descriu correctament aquestes CFD.

Des d'un punt de vista físic, la ruptura d'ADN és un procés de fractura. Les CFD mostren regions cooperatives de ruptura. Mitjançant tècniques d'anàlisi bayesià es poden inferir els estats intermediaris que s'observen durant la ruptura i la grandària de les regions cooperatives, que presenten un rang de 10-80 pb. El model de joguina és una versió simplificada del model mesoscòpic que recull les propietats estadístiques bàsiques dels experiments de ruptura. Els experiments no permeten discriminar RCR menors de 10 pb. Per altra banda, l'obertura de bases d'una en una es pot aconseguir augmentant la rigidesa de la trampa òptica fins a 100 pN/nm. Curiosament, aquest valor es correspon amb la rigidesa d'un sol nucleòtid de la cadena senzilla d'ADN. Això permet a la maquinària cel·lular accedir a la informació genètica d'una base en una base, sense necessitat de tenir un gran nombre de bases obertes. D'aquesta manera, les bases no s'exposen al solvent i s'eviten possibles d'anys.

Els valors de les energies lliures UO no descriuen quantitativament les CFD dels experiments de ruptura. Aquests valors es poden modificar per tal d'ajustar les mesures experimentals al model mesoscòpic. L'ajust es fa mitjançant un algoritme Monte Carlo que es mostra robust i dóna solucions satisfactòries. Els resultats mostren una dependència heterogènia de les 10 EFPB amb la sal. A més, els resultats prediuen correctament les temperatures de desnaturalització dels oligonucleòtids.

Les minipinces també permeten realitzar experiments a força controlada mitjançant un algoritme de realimentació. Les CFD a força controlada són irreversibles a l'escala de temps de l'experiment i sempre presenten histèresi. El nombre de bases obertes en funció de la força obeeix les lleis d'escalament predites.

Aquesta tesi s'ha centrat en les propietats estadístiques i termodinàmiques dels experiments de ruptura d'ADN realitzats amb pinces òptiques. Aquest estudi ha respost algunes preguntes i n'ha deixat altres d'obertes. El següent pas és estendre el treball per comprendre les extraordinàries propietats de l'ADN i trobar aplicacions pràctiques basades en els descobriments exposats aquí.

Part II

Thesis

Summary

Molecular biophysics is a scientific discipline that studies biomolecules. This discipline has experienced a revolution thanks to the development of single-molecule techniques. These techniques allows us to obtain new and valuable information that complements the traditional bulk assays. Optical tweezers is an experimental technique that uses the radiation pressure of light to exert forces on a tiny dielectric microsphere. Biomolecules can be bonded to such microspheres in order to perform pulling experiments at the single-molecule level. Minitweezers is a dual counter-propagating laser tweezers instrument that measures the force exerted on the microsphere by conservation of light momentum. The instrument has high stability and resolution (0.1 pN in force and 0.5 nm in distance) in the measurements. The properties of the DNA molecule can be studied with the Minitweezers at the single-molecule level. The DNA is a biomolecule that forms a double helix that stores the genetic information of the cells. DNA unzipping experiments consist in pulling apart the two strands of DNA by exerting mechanical forces on the extremities of the molecule. In such process, base-pairs (bp) are disrupted sequentially, showing a succession of cooperative unzipping regions (CUR) of different sizes (between 1–100 bp). In a DNA unzipping experiment, we measure the force vs. distance curve (FDC) of the molecule, which has a characteristic sawtooth-like shape that is sequence-dependent. The FDC is analyzed with a Bayesian approach to infer the size distribution of the CURs. The experimental accuracy does not allow to observe CURs of sizes below 10 bp. Furthermore, the unzipping of one bp at a time can only be achieved by having an optical trap stiffness value higher than 0.1 N/m, which corresponds to the stiffness of a single nucleotide of DNA. This has been deduced from a toy model specifically introduced to study the CUR size distributions. In addition, the FDCs are theoretically predicted by the nearest-neighbor (NN) model adapted to unzipping experiments. The NN model describes the hybridization reaction of two strands of DNA. By fitting the experimental FDCs to the model, the unique 10 NN bp free energies are obtained with 0.1 kcal·mol^{-1} precision between 10 mM–1 M of monovalent salt concentration. The results show that the unzipping FDCs and the melting temperatures of oligos are correctly described with a specific salt correction for each of the 10 NN bp free energies. Differently from the previous experiments, the unzipping of DNA can also be performed at controlled force. These last type of experiments exhibit large hysteresis and irreversibility. The free energy landscape is a tool that helps to understand the unzipping at controlled force. Finally, the work presented in this thesis

can be extended to find practical applications of DNA unzipping, such as sequencing of DNA by force, and measurement of thermodynamic properties of biomolecules in conditions not accessible by bulk methodologies.

List of abbreviations

AD	anti-digoxigenin
ADC	analog to digital converter
AFM	atomic force microscopy
API	application programming interface
CF	controlled force
CUR	cooperative unzipping region
DAC	digital to analog converter
DFC	distance versus force curve
DNA	deoxyribonucleic acid
dsDNA	double-stranded DNA
FBG	fiber Bragg grating
FDC	force versus distance curve
FDC$_f$	force versus distance curve at controlled force
FDC$_x$	force versus distance curve at controlled position
FEC	force versus extension curve
FJC	freely jointed chain
FRET	fluorescence resonance energy transfer
FFT	fast Fourier transform
FT	fluctuation theorem
GLMT	generalized Lorentz-Mie theory
HOT	holographic optical tweezers
LOT	laser optical tweezers
MC	Monte Carlo
MT	magnetic tweezers
NA	numerical aperture
NN	nearest neighbor
NNBP	nearest neighbor base pair
PBS	polarizing beam-splitter
PIC	programmable interface controller
PSD	position sensitive detector
PWM	pulse width modulation
RNA	ribonucleic acid
SA	streptavidin
SD	steepest descent
SLM	spatial light modulator
SME	single-molecule experiments
SPI	serial peripheral interface
ssDNA	single-stranded DNA

TE	transverse electric
TEM	transverse electromagnetic
TM	transverse magnetic
UO	unified oligonucleotide
WLC	worm-like chain

La inspiración existe, pero tiene que encontrarte trabajando.

Pablo Ruiz Picasso (1881-1973)

The scientist is not a person who gives the right answers, he is one who asks the right questions.

Claude Lévi-Strauss (1908-2009)

Chapter 1

Introduction

Most scientists agree that research requires imagination and tenacity. Once these two virtues are gathered, the discovery of a breakthrough is a matter of luck. Nevertheless, the history of science is not only made of breakthroughs. Instead, it is made of discoveries of different importance. Some discoveries just represent a little expansion of our knowledge. Nowadays, the frontiers of the science are expanded at the highest rate ever, thanks to millions of researchers from all over the world. However, science is not a collection of phenomena. It is also a frame to interpret the results. Therefore, all the little investigations need to be unified from time to time. Precisely, this is what great scientists do. For instance, Isaac Newton unified the celestial and the terrestrial mechanics in his book *Principia*. Indeed, a new paradigm usually provides a simplified way to look at phenomena and establishes a landmark to promote new investigations.

In 1828, the German chemist Friederich Wöhler accidentally unified the chemistry and the biology with a simple experiment. The experiment consisted in a chemical reaction in which two inorganic substances (silver cyanate and ammonium chloride) were mixed to produce an organic composite (urea). At that time, scientists believed that the inanimate (inorganic) matter was fundamentally different from the living (organic) matter. Wöhler's discovery was followed by other experiments supporting his observations. The scientists finally convinced themselves that both types of matter were made of atoms and there was no intrinsic difference between them. In 1903, the German scientist Carl Neuberg coined the word *Biochemistry*, to refer to a new

discipline that studied the processes of life (respiration, fermentation, etc.) from a chemical point of view.

The birth of biophysics was not that straightforward. Instead, it took decades to clarify whether physics could be used to answer biological questions [31]. In the mid 19th century, the *School of Berlin* was a group of scientists who firstly applied physics to the study of living systems. They strongly believed that physiology (i.e., the functioning of living systems) could be regarded as physical phenomena. For instance, Emil DuBois-Reymond modeled the nerves as an electrical system. Another example is Carl Ludwig, who firstly measured the blood pressure. This scientific discipline was known as *medical physics*. However, it remained in standby because the physics of the 19th century could not provide explanations to all phenomena observed. In 1944, Erwin Schrödinger, the Austrian physicist who co-formulated the dynamics of quantum systems, wrote an inspiring book entitled *What Is Life?* [32]. The book essentially suggested a physical approach to the questions concerning biology. In 1953, Watson and Crick discovered the double-helix structure of DNA (i.e., one of the most significant molecules in biology) by using X-ray diffraction (i.e., an experimental technique developed by physicists) [1]. This achievement represents the starting point of modern biophysics, which mainly focused on the structure and interaction of biomolecules and cells. Finally, the single-molecule techniques developed during the last decade of the 20th century represented a new revolution in the field of biophysics [33].

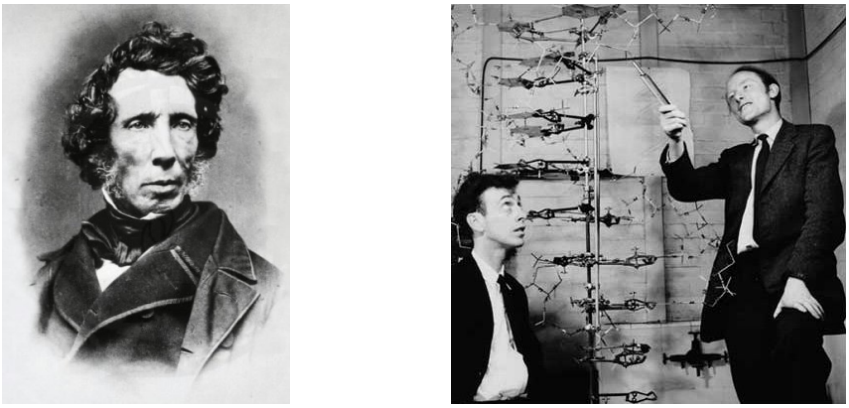


Figure 1.1: Portrait of Friederich Wöhler (1800-1882), who discovered that urea (organic matter) could be synthesized from inorganic matter (left picture). James Watson and Francis Crick showing the double-helix structure of DNA in 1953 (right picture).

The aim of this chapter is to focus the object of study of this thesis. We will overview the questions of interest in biophysics; we will state the scope of this work within the field of biophysics; and we will summarize the main

results presented here.

1.1 What is biophysics?

Biophysics is a scientific discipline that uses the tools and the methods of physics to study biological systems [34, 35]. In other words, biophysics is the study of physical processes governing the living cells. This general definition includes a variety of subtopics that focus on a wide range of levels, from molecular processes to ecological phenomena. Biophysicists also look at several aspects of biological systems such as structure, dynamics, mechanisms, complexity, applications, etc.

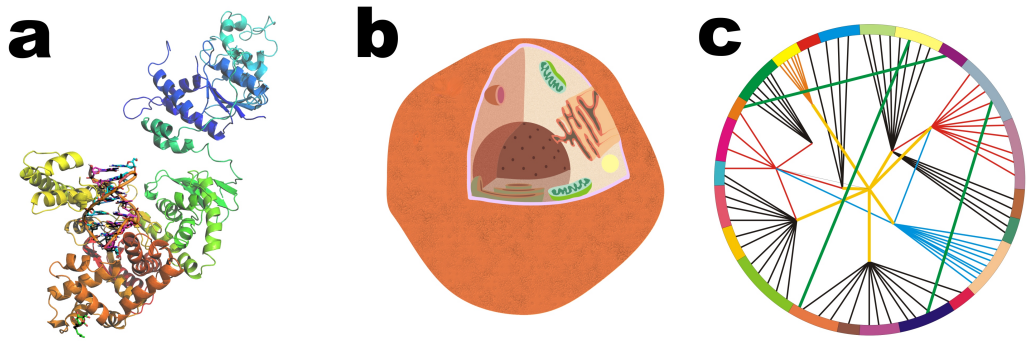


Figure 1.2: Some topics in biophysics. **(a)** Molecular biophysics. Structure of a DNA polymerase [36]. Picture of Protein Data Bank ID: 1TAU, created with PyMOL software (DeLano Scientific, 2002). **(b)** Cellular biophysics. Schematic representation of a cell. **(c)** Systems biology. A network of genes and their interaction.

The term biophysics is the result of mixing the words biology and physics. This suggests that both branches of knowledge are blended into a new discipline. However, the skills required in biology are very different from the ones required in physics. So the intersection between them is minimal. Indeed, biology is descriptive and complex, while physics is conceptual and simplistic. Biology is worried about the details, while physics tends to establish general laws. How can both ways reconcile?

A traditional answer to this question proposes either a biological or a physical approach to biophysics [37]. In the first case, physics is the servant to study a biological problem. Physicists contribute to biology by providing experimental techniques and mathematical tools. In the second case, the biologist provides a system that is studied by physicists like any other physical system. The physicist tries to find and understand new physical phenomena to establish new concepts and laws. Although the difference between both

approaches seems fuzzy, there is a central question that clarifies this. Who benefits from research in biophysics? If the answer is biology, we are facing a biological approach. Otherwise, this is a physical approach. Let us focus on two examples.

- The discovery of the double-helix of DNA [1] represented a great step for biology, that finally unveiled the carrier of genetic information in living beings. However, no new physics was applied because X-ray diffraction was known since the beginning of the 20th century. Besides, such discovery did not contribute to physics because no new phenomenon was discovered. This is an example of a biological approach in which the knowledge of biology is expanded by physics.

- An opposite example is the experimental verification of the Worm-Like Chain (WLC) model. The WLC model was developed by Kratky and Porod in 1949 to describe the elasticity of single semi-flexible polymers [38]. In 1992, Smith and co-workers pulled and measured the force vs. distance curve of a single DNA molecule using magnetic tweezers [39]. They showed that the WLC model predicts the elastic response of DNA [40]. In this case, a little contribution was made in biology. The synthesis of DNA was already known and no new biological insight was gained from the experiments. On the other hand, the experiment required the development of physical instrumentation (magnetic tweezers) and tested a purely physical model. Here, biology helped physics to expand its knowledge. This is a physical approach to biophysics.

Still, there are reluctances in the collaboration between physicists and biologists. Some biologists dislike the simplistic approach of physicists and prefer not to get rid of details. Some physicists consider that biosystems are too complicated for calculations. In fact, physicists feel much more attracted by biology. Indeed, they discover new phenomena where physical laws can be applied.

Nevertheless, this dual approach to biophysics is being left behind. Universities from all over the world devote degrees, masters, PhD programs and professors exclusively to biophysics. Besides, research institutes are created as a response to the demands of biophysics. All these efforts are changing the way we look at biophysics. Biophysics is no longer "physics for biologists" or "physicals methods applied to biology". On the contrary, biophysics is a discipline defined by its own scientific questions [41]. Biophysicists have become a new kind of researchers. They do not want to explain the biological processes in exact detail nor with an ideal model. They face scientific issues by extracting general laws of biological systems without excessive simplification.

Biophysics is closely related with other disciplines such as biochemistry, nanotechnology or biomedicine. The border among them is not sharp. Rather, there is an overlapping of topics, experimental techniques and theoretical frameworks. Though, the most significant feature of biophysics is the use of measurements and quantitative information.

1.2 Molecular biophysics

As mentioned before, biophysics covers a wide range of topics related to living systems. The object of study includes biomolecules, cells, organisms and ecosystems. Some studies focus on the structure of the elements such as proteins, membranes, tissues, etc. Some others focus on the dynamics such as protein folding, cell motility, self-assembly. Another topic is systems biology, that focuses on the dynamics of complex systems such as networks of proteins or population growth.

Molecular biophysics is a sub-topic of biophysics that studies the structure, function and kinetics of biomolecules. There are two main types of biomolecules: nucleic acids (DNA and RNA) and proteins, which include enzymes and molecular motors. Biomolecules are the building blocks of cells and they are also made of subunits (amino acids, nucleotides). So biomolecules are located in the middle level within the hierarchy of organization of the living systems (from atoms to ecosystems). The physical regime of biomolecules is governed by thermal fluctuations. The typical interactions of biomolecules involve weak bonds such as hydrogen bonds or Van der Waals forces. These bonds can be broken by thermal fluctuations. So the properties of biomolecules are determined by bonds that have similar energies to the ones of the surrounding thermal bath. Here emerges on of the most interesting questions in molecular biophysics. How can molecules work embedded in such thermal fluctuations?

Hot topics in molecular biophysics are RNA and protein folding, mechanochemistry of molecular motors, ionic channel transportation and DNA-protein interaction. The range of experimental techniques available is very wide. Apart from the well known biochemical techniques such as PCR and electrophoresis, biophysicists use X-ray diffraction, calorimetry, nuclear magnetic resonance, electronic and confocal microscopy among others. Finally, the development of single-molecule techniques such as AFM or optical tweezers have provided a new insight in molecular biophysics. The quantitative measurements obtained with these techniques allow biophysicists to understand the structure and the function of biomolecules and formulate models to characterize and predict their behavior. This will be discussed in the next section.

1.3 Single-molecule techniques

Traditionally, the experiments in chemistry and biology have been performed in bulk. Bulk experiments involve a large amount of substance (moles, grams, milliliters, etc.). Whatever the experiment is (reaction, measurement, detection), it is performed on a large collection of molecules. These experiments show the average behavior of the molecules and they have been used to set the basis of experimental sciences for centuries.

The detection of single particles has been pursuit by scientists since John Dalton proposed the modern atomic-molecular theory of matter in 1808. Physicists have been able to detect photons, electrons and sub-atomic particles using sensitive detectors (photomultipliers, particle accelerators, etc.). However, some of the most interesting phenomena in physics are cooperative and involve lots of particles: phase transitions, Bose-Einstein condensates. So single-molecule techniques might seem that have little to offer when studying elementary particles or molecules.

However, biomolecules have internal structure and they show complex behavior. So the study of biomolecules benefits from single-molecule techniques [42]. The key point here is that matter is viewed as the result of gathering complex individuals. Using single-molecule experiments (SME) the individual molecules can be manipulated an measured one at a time. The advantage is that the information is not averaged. Instead, one can measure deviations from the average bulk behavior and probability distributions. For instance, biophysicists have been able to measure the step size distribution of molecular motors; the energy consumption of enzymes; and the kinetics of biochemical reactions [33]. All these new amount of information complements the traditional bulk assays.

The first SME in biophysics were performed in the early 1990's [43]. Block *et al.* were able to characterize the movement of kinesin (i.e., a molecular motor in charge of transportation of substances within the cell) at the single molecule level using optical tweezers [44]. Later, Finer *et al.* described the working of myosin, another molecular motor that drives muscular contraction [45]. The SME of other molecular motors like F_1-ATPase [46] or RNA polymerase [47] were extensively studied. Among other properties of the motors, these works measured the step size, the stall force or the efficiency at different ATP concentrations. All these magnitudes are hardly measured in bulk. Other SME experiments [2] have focused on the elongational [48] and torsional [33] elasticity of single DNA molecules and the folding/unfolding kinetics of RNA [49] and proteins [50].

The development of SME have pushed the instrumentation forward. The technical devices related with positioning (piezoelectric crystals, motor driven

stages), detection (position sensitive detectors, high frequency cameras) and transduction (nanometric cantilevers, magnetic microspheres) have greatly expanded in the last years. Biochemistry has also standardized protocols and reactions commonly used in SME (labeling kits, fluorescent tags). The most productive single-molecule techniques are single-molecule fluorescence, Atomic Force Microscopy (AFM), Magnetic Tweezers (MT) and Laser Optical Tweezers (LOT). Figure 1.3 shows a brief description of them.

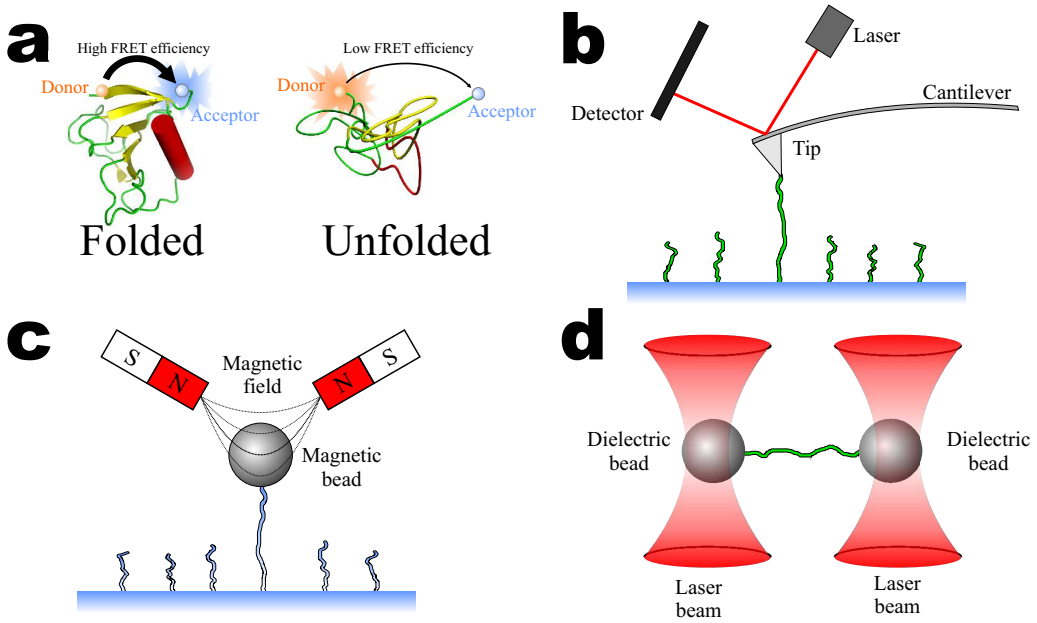


Figure 1.3: Single-molecule techniques. **(a)** Fluorescence resonance energy transfer (FRET). FRET is based on the transmission of energy between fluorescent dyes (donor and acceptor). The efficiency of such transmission depends on the distance between the dyes. It can be used to detect the folding/unfolding of a molecule. **(b)** AFM. AFM produces and measures the forces applied on a molecule (depicted in green). The force applied is directly related to the bending of a solid state cantilever, whose elasticity is known. The bending of the cantilever is measured from the deflection of a laser beam reflected on the cantilever. **(c)** MT. The molecule is attached between a magnetic bead and the surface of a cover glass. MT apply forces by immersing the bead in a magnetic field gradient. The force is measured by determining the position of the bead. The magnetic field is produced by magnets that can be moved and rotated to exert force and torque on the molecule. **(d)** LOT. LOT exerts forces on dielectric beads by using the pressure of radiation of a laser beam. A detailed description will be given in chapter 2.

SME are continuously evolving. The original pioneering experiments can be reproduced by most labs and such experiments have become starting points to study more complex systems. The current trend is to improve and combine techniques and perform several single-molecule experiments in

parallel. For instance, optical tweezers initially had one trap. Nowadays, most of the new constructed optical tweezers have two or more traps [51].

The research in molecular biophysics is definitely tied to single-molecule techniques.

1.4 Thermodynamics of small systems

The thermodynamics of small systems (also known as mesoscopic dynamics) is a subtopic of physics that deals with systems that have an intermediate scale between the microscopic and the macroscopic systems [3]. At this intermediate scale, the typical number of microscopic components of the system is much larger than 1 but much smaller than Avogadro's number. A characteristic property of these systems is that the energy exchanged between the system and the environment is of the same order of magnitude than the energy fluctuations. As a result, the energy fluctuations of the system contain valuable and meaningful information.

According to the classical statistical mechanics [52], the relative fluctuations of observable magnitudes decrease as $\sim 1/\sqrt{N}$, where N is the number of particles of the system. Traditionally, the fluctuations of macroscopic thermodynamic systems (gas, magnet) have been hardly observed (except for the critical points), due to the large number of particles involved in these systems ($N \sim 10^{23}$). However, the boom of single-molecule techniques have provided a new collection of experiments that exhibit significant fluctuations. Indeed, the biomolecules used in these experiments have few degrees of freedom (thousands of atoms). This has been very useful for the physicists interested in the equilibrium and non-equilibrium thermodynamics of small systems. In fact, some non-equilibrium theorems have been firstly tested on single-molecule experiments [53, 54].

Theorists devoted to the thermodynamics of small systems have focused their attention on the Fluctuation Theorems (FT). The FT relate the equilibrium properties of a system with the work performed on this system along irreversible processes. This was initially stated by Jarzynski [5] and generalized by Crooks [4]. The FT allowed to establish generic results (i.e., non-system dependent) in systems out of equilibrium. Nowadays, the application of FT in biophysics is quite frequent in order to obtain the free energies of formation of biomolecules. The experiments with biomolecules have also contributed to expand the FT. In particular, FT have been extended to include the partial equilibrium, which allows to infer the free energy of formation of different conformational states of biomolecules [55].

One of the more interesting topics in the thermodynamics of small sys-

tems are molecular motors of the cell. These systems are strongly affected by the environment in which they perform their functions (transportation, polymerization). In fact, thermal fluctuations dominate the behavior of such motors. In spite of the fluctuations, these motors are still extremely efficient. How can these systems work properly embedded in such wild conditions? The single-molecule micro-manipulation techniques shed light into this question [56].

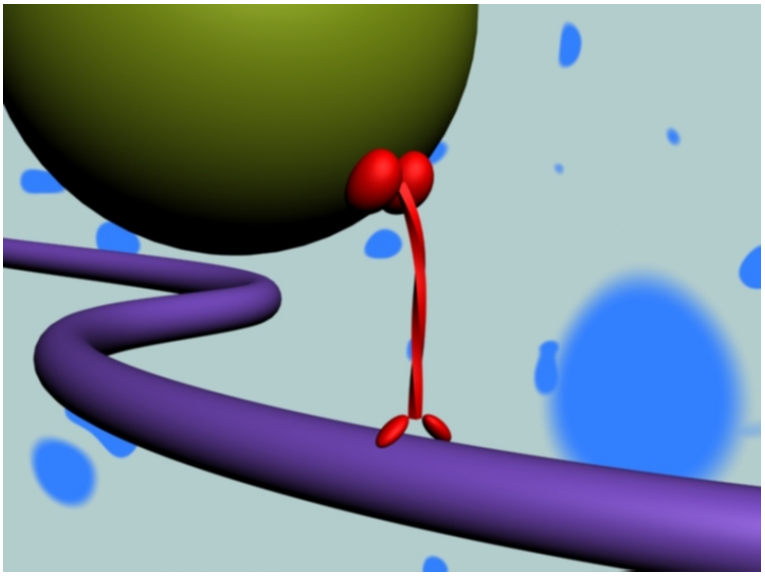


Figure 1.4: Artistic view of a kinesin at work. Kinesin (in red) is a molecular motor that carries vesicles of substances (yellow) by *walking* along the microtubules (violet) located within the cell.

1.5 Summary of work presented in this thesis

This PhD thesis is about the statistical and thermodynamic properties of DNA unzipping studied with optical tweezers.

Chapter 2 contains a thorough description of the optical trapping principle and the Minitweezers instrument used in the experiments reported in this thesis. The physics laying behind the optical tweezers is explained step by step. The technical details about the implementation of the instrument are also given. The maintenance (calibration, design of new protocols) of the instrument completes the chapter. All the information is complemented in the appendices.

Chapter 3 focuses on the molecule of DNA. There is an introductory description of the structure of the DNA. This description sets the basis to

understand the goal of the unzipping experiments performed with optical tweezers. Then the chapter describes the unzipping experiments and the models used to illustrate the phenomena observed.

Chapter 4 is a closer look to the statistical properties of the metastable states observed during DNA unzipping. The molecule of DNA is viewed as a system that exhibits cooperative transitions between its intermediate states. A toy model is developed to simplify and understand this phenomenon. This model also allows us to predict the experimental conditions under which sequencing of DNA by mechanical unzipping will be feasible.

Chapter 5 explains in detail how to extract the free energy of formation of duplexes of DNA from unzipping experiments. This is an original new approach in which the free energies of the building blocks of DNA (i.e., the base-pairs) are directly obtained. The chapter describes the analysis of the experimental data and the theoretical model that is used to reach this goal. The chapter concludes with a discussion about the salt dependence of free energies, one of the main conclusions of this thesis.

Chapter 6 is an exploratory study of DNA unzipping at controlled force. The chapter compares the controlled force experiments with those performed at controlled position and discusses the origin of the differences between them. There is also a brief investigation of the causes of the large hysteresis and irreversibility observed at controlled force.

Finally, chapter 7 sketches some future works and perspectives open by this thesis and summarizes the main conclusions derived from this PhD thesis.

And God said, "Let there be light". And there was light. And God saw the light, that it was good; and God divided the light from the darkness.

Genesis, 1:3-4

Chapter 2

Optical tweezers

Light carries momentum. This is what James Clerk Maxwell deduced from the equations of the electromagnetism [6]. The conservation of momentum and energy leads to a transfer of these two quantities from the electromagnetic field to the physical objects when they interact. Starting from Maxwell equations, in 1908 Gustav Mie computed the rigorous solution for the electromagnetic field of a plane wave diffracted by a homogeneous sphere [57]. Since then, the dispersion of light by small particles has been known as *Mie theory*. Following Mie's solution, Peter Debye calculated the mechanical force undergone by the spherical particle due to the radiation pressure of light [58]. The effect of the radiation pressure is hardly observed in the macroscopic world. However, in the microscopic level, the momentum of light might have the same order of magnitude than the momentum of the microscopic objects. In this situation, objects undergo forces that produce observable effects.

It was not until 1970 when Arthur Ashkin [7] firstly observed experimentally the interaction between a laser beam and a tiny particle in the laboratory. Using one single laser, he could accelerate particles along the direction of propagation of light. Moreover, combining two counter propagating lasers, he could trap particles. Ashkin understood that the scattering and gradient forces that pushed the particle were due to the interaction between the laser beam and the particle. Ashkin and Dziedzin [59] observed the optical levitation of a transparent particle. In that experiment, the scattering force of a vertically directed laser beam was canceled by the weight of the particle and the lateral restoring force was probed. Later, Ashkin [8] and collabora-

tors observed for the first time the trapping of particles using a single laser beam. They showed that the gradient force due to the focused laser beam was higher than the scattering force inducing the trapping of the particle. They also provided an explanation of trapping based on ray optics and they used Rayleigh dispersion (which assumes that the particles are much smaller than the wavelength of the laser) to estimate the trapping condition. Such device for capturing small dielectric particles has been known as *optical tweezers* and it has evolved rapidly to measure and control the applied forces.

The first attempt to measure the forces exerted by an optical trap was done by Block *et al.* [60]. They inferred the maximum trapping force at different laser powers from the viscous drag of a rotating bacterial cell. This method was initially used in other biophysical experiments [61, 44] but it could only provide information about the escape force, which limited the design of the experiment. The instrumentation and the applications evolved together in the early 90's. Denk and Webb [62] measured picometer displacements of cells and beads using a phase interferometer. This pioneering work set the bases of the *virtual spring* calibration technique widely used later to measure forces [9]. It provided high bandwidth measurements that allowed to observe the force fluctuations of a particle located in an optical trap. Kuo and Sheetz used video detection of beads to measure the forces applied by the optical tweezers [10]. Webb and co-workers used a photodetector to measure the light deflected by a particle to measure the roughness of a surface with nanometer precision [63]. This work inspired the detection of forces by light deflection [11]. Finally, Simmons *et al.* measured the forces by projecting an image of the bead to a quadrant detector [64]. A thorough description of back focal plane force detection was published in 2006 [65].

The recent advances in optical trapping have focused in two main directions: 1) improvement of the resolution and accuracy of the instrumentation and; 2) combination of new manipulation tools with optical tweezers [12]. Instruments with double optical traps are substituting the pioneering setups that tended to use coverslips and micropipettes as anchor points. The advantages of the double trap geometry are that they reduce the drift effects and give more information about the thermal noise [13]. Besides, back focal plane detection has become a standard setup to measure the position of the beads and the force applied by the optical trap. For instance, Dame and coworkers were able to create four optical traps by using acousto-optic deflectors [51]. This allowed them to simultaneously pull on two crossed strands of DNA and study how H-NS proteins link these two molecules of DNA. Another way to create double optical traps is the so-called time sharing technique [66]. It consists on switching the position of the laser beam much faster than the relaxation time of the bead, which effectively produces two optical traps.

Moreover, it can be used to produce novel optical potentials, i.e., different from the harmonic potential that induces a restoring force on the bead. Finally, the holographic optical tweezers (HOT) are a promising innovation in which the experimentalist can control the number, size, shape and position of the optical traps [14]. They are produced with spatial light modulators (SLM) that modify the wavefront of the laser beam and produce a desired pattern of traps. The main handicap of HOT is the measurement of forces. Indeed, the scattered light of the beads trapped in several traps cannot be split. Therefore, the force exerted on each bead cannot be measured independently. Besides, dynamical HOT require significant computational time to generate the holograms. Although successful algorithms have been implemented [67, 68], the repositioning of the optical traps is also limited by the refreshing time of the SLM ($\sim$ 60 Hz). Nevertheless, the HOT are a starting point to parallelize single-molecule experiments by producing arrays of optical traps and manipulate tridimensional biological systems at will.

The hybrid tools combine optical tweezers and other techniques. Bryant *et al.* used a rotating micropipette to induce torque on a DNA molecule [69]. The torque has also been exerted by using rotating magnets and paramagnetic beads embedded in the surroundings of an optical trap [70]. Studies of Laguerre-Gaussian laser modes (which are different from the axially symmetric TEM_{00} mode) have shown that they carry angular (and linear) momentum that can be transfered to the bead and induce torque [71, 72]. However, none of the previously mentioned experimental techniques were able to directly measure the torque exerted on the trapped particles. The simultaneous exertion and measurement of torque can be achieved by using cylindrical or birefringent beads (i.e., beads with two indexes of refraction) [73]. The fast axis of a birefringent particle tends to align with the polarization of the beam. The torque can be directly measured by analyzing the circular polarization of the outcoming light. Apart from the torque, optical tweezers have also been combined with other techniques such as nanopores [74] and fluorescence [75, 76]. Optical tweezers will eventually become a standard tool for the biophysicist. The noninvasive nature of optical trapping and its customizability makes it an ideal technique to carry out experiments on biological systems.

This chapter focuses on the description of the minitweezers, the optical tweezers instrument used to perform the DNA unzipping experiments detailed on the next chapters. After explaining the physical principles that lie behind the trapping of particles, we will deepen into the technical details of the instrument.

2.1 Optical trapping principle

2.1.1 Physics of optical trapping

There are several theories in physics to describe the interaction between light and matter: from the very fundamental Quantum Electrodynamics of electrons and photons to the Mie scattering of atmospheric particles. Thus, there is an appropriate theory for each scale of energy and mass involved in the interaction. The optical tweezers instruments devoted to experimental biophysics deal with interactions that are correctly described by classical theories (i.e., Maxwell's electromagnetism). In this classical regime, the forces exerted on matter by light are known as *radiation pressure*. The radiation pressure is the result of light-matter interaction and includes different phenomena such as absorption or scattering of light. The purpose of an optical tweezers setup is to control the radiation pressure of light in order to apply forces on tiny objects.

Although all optical forces exerted on a particle arise from the same physics, they are usually split into two main contributions: the scattering force and the gradient force. In both cases, the change in the light momentum is the ultimate responsible of the force exerted on the particle. The scattering force tends to push the particle along the beam in the direction of propagation, while the gradient force pushes the objects towards the regions of highest light intensity. If the light is correctly conditioned (in terms of collimation, intensity, aberration, etc.) the scattering and gradient forces can be combined to apply controlled forces on the particle. An optical trap is formed by focusing a laser beam in a small region so that a transparent particle feels a restoring force that tends to take it to the region of maximum intensity of light. An optical trap can only be formed if the gradient force along the optical axis is stronger than the scattering force that pushes the bead out of the focal point.

There are three main theoretical approaches to the physics of optical trapping, depending on the ratio between the wavelength λ of the light and the diameter of the particle d. In the ray optics (or Mie) regime, the particle is very large compared to the wavelength ($d \gg \lambda$), whereas in the Rayleigh regime the opposite is true ($d \ll \lambda$). The calculation of optical forces in the intermediate regime ($d \approx \lambda$) is quite complicated and requires the solution of the Maxwell's equations with the appropriate boundary conditions. It can be achieved by means of the Generalized Lorenz-Mie Theory (GLMT) [15]. Most of the work done with optical tweezers in biophysics falls in this last case, where infrared lasers have a wavelength of $\lambda \simeq$ 0.8-1.2 μm and the trapped particles are spherical beads of $d \simeq$ 2-4 μm.

The ray optics regime

In this regime, the wavelength of the light is much smaller than the typical dimensions of the object and the electromagnetic properties of radiation can be neglected (neither diffraction nor undulating effects are considered). This is the case of experiments in which cells are trapped using infrared lasers [77, 60]. In the ray optics description, a light beam is decomposed into individual rays characterized by their intensity and direction that follow the laws of geometrical optics. The ray tracing of each individual ray across the surfaces of the object (reflections and refractions) allows to determine the final direction of the ray and the net change in the linear momentum of light (see Fig. 2.1a). The overall rate of change of momentum of all rays produces a force on the object.

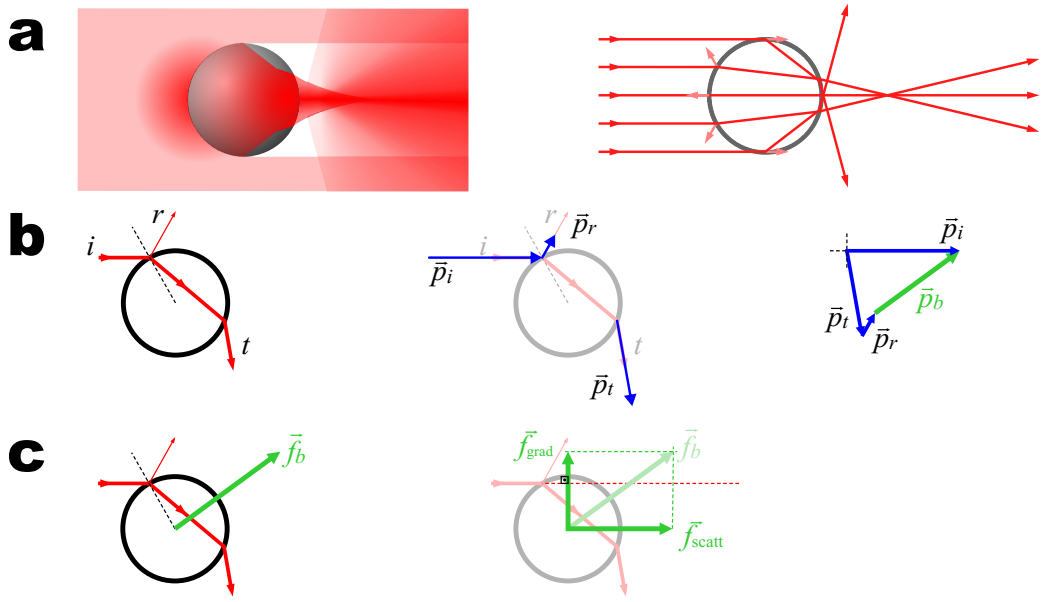


Figure 2.1: Ray optics regime. **(a)** Interaction of a collimated beam with a transparent particle. The left picture shows the dispersion of light and the distribution of intensities by reflections and refractions on the particle surface. The right picture shows a schematic representation of the ray tracing. **(b)** The left picture shows the optical path of one incident ray (i) that is reflected (r) and transmitted (t) on the particle. The central picture shows the directions and magnitudes of the linear momentum associated with the rays (blue arrows). The right picture shows how, due to the conservation of the linear momentum, the difference of linear momentum (green arrow) between the incident ($\vec{p}_i$) ray and the sum of the reflected ($\vec{p}_r$) and the transmitted ($\vec{p}_t$) rays induces a transfer of linear momentum to the particle ($\vec{p}_b$). **(c)** The linear momentum transferred to the bead and the resulting force have the same direction (left picture). The right picture shows the two components of the resulting force: the scattering ($\vec{f}_{\text{scat}}$) and the gradient ($\vec{f}_{\text{grad}}$) forces.

The resulting force exerted by a single ray of light on a transparent particle is the result of reflections and refractions of the incident ray (see Fig. 2.1b). A transparent particle is made of a material that does not absorb light (i.e., the index of refraction is a real -not complex- number) but the particle can reflect and refract light. The linear momentum carried by the incident light ($\vec{p_i}$) is split into the two outcoming rays: the reflected and the transmitted ray, each one carrying a different amount of linear momentum ($\vec{p_r}$ and $\vec{p_i}$ respectively). The balance between them is the resulting linear momentum transferred to the bead $\vec{p_b}$:

$$\begin{aligned} \vec{p_i} &= \vec{p_r} + \vec{p_t} + \vec{p_b} \\ \vec{p_b} &= \vec{p_i} - (\vec{p_r} + \vec{p_t}). \end{aligned} \tag{2.1}$$

The amount of linear momentum transferred per unit of time determines the force applied to the bead ($\vec{f_b}$):

$$\frac{d\vec{p_b}}{dt} = \vec{f_b}. \tag{2.2}$$

Usually, the ray reflected on a transparent particle is much weaker compared to the transmitted ray. So in most cases the reflected ray can be neglected. The resulting force is usually split into two perpendicular components (see Fig. 2.1c). The first one is the axial or scattering force ($\vec{f}_{\text{scat}}$) and it is parallel to the original direction of the beam. The second component one is called radial or gradient force ($\vec{f}_{\text{grad}}$).

The calculation of the force exerted by a beam of light is performed by repeating the previous computation for all the rays of the beam and summing their contributions (see Fig. 2.1a). This exerted force can be calculated numerically [78]. The summation of the radial forces for each ray gives the total gradient force exerted by the light beam. And so for the total scattering force. The calculation is strongly dependent of each particular case: position, collimation, intensity of the laser beam; and position, shape and index of refraction of the particle (see Figure 2.2). Besides, the optical forces depend on the polarization of the laser beam because the reflectivity and transmissivity on the surface of the particle are given by the Fresnel reflection and transmission coefficients respectively, which are polarization dependent. Thus, the calculation is usually performed on circularly polarized light, where the resulting trapping force is the average of the parallel and perpendicular polarizations. The forces calculated in the ray optics regime are independent of the particle size although we know that this is not true in a typical experimental setup. However, the ray optics regime provides a good qualitative description of the trapping phenomenon.

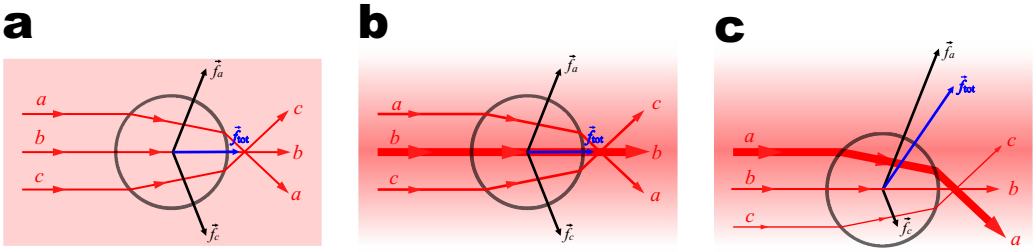


Figure 2.2: Interaction of a spherical particle with a light beam. Only a few rays of the beam are represented. Inclusion of more rays does not modify the qualitative description. No reflected rays are considered here. **(a)** Uniform light beam. The gradient forces of symmetric rays (a and c) cancel each other. Ray b does not contribute to the total force because it does not change its direction. Therefore, only the scattering forces contribute. The resulting force is directed along the propagation of the light beam. **(b)** Gaussian light beam with centered particle. Although the intensity of ray b is different from the ones of rays a and c, the gradient forces of these two rays still cancel each other. Only the scattering forces contribute to the total force. **(c)** Gaussian beam with off-centered particle. The gradient force of ray a is higher than the one of ray c. The resulting gradient force pushes the particle towards the region of maximum light intensity. The scattering forces are also present. Therefore the particle tends to align with the axis of the Gaussian beam (gradient force) and it is accelerated along the beam (scattering force).

In the particular case of a collimated Gaussian laser beam (i.e., a TEM_{00} mode) interacting with a spherical bead, the scattering force always pushes the particle along the direction of the beam (see Figs. 2.2b,c). Therefore, the particle cannot be confined (i.e., trapped) in a desired region of the space. However, a focused (not collimated) Gaussian beam can be used to trap the particle. The converging rays of the focused beam are deviated in such a way that the scattering force can be positive or negative, depending on the relative position between the center of the particle and the focus of the laser beam (see Figs. 2.3a,b,c). So an axial restoring force arises and the particle can be trapped. The more focused the laser beam, the higher the restoring force. The combination of this new axial restoring force with the radial one (already induced by the distribution of intensities of a Gaussian beam) produce a 3-dimensional optical trap (see Fig. 2.3d). So the particle is trapped by a central restoring force that tends to keep the particle in the focus of the laser beam. In most optical tweezers setups, the trapping is achieved by using Gaussian laser beams that overfill the back pupil of high numerical microscope objectives.

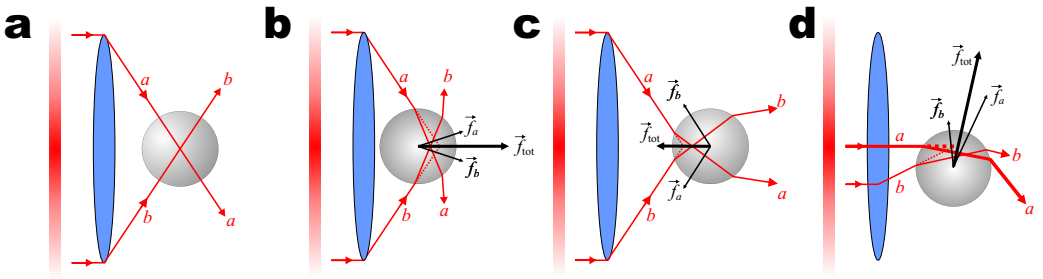


Figure 2.3: Optical trapping with a focused Gaussian beam. The reflection effects have been neglected. In all cases, the reflection would just push the particle a little bit along the direction of propagation of the beam light. **(a)** A particle centered at the focus of the Gaussian beam. The marginal rays a and b are not deviated by the particle because they cross the surface of the particle perpendicularly (this is a consequence of Snell's law). No force is exerted on the particle. **(b)** A particle located before the focus of the laser beam. The focus is located at the expected crossing point of the rays (where the dashed lines cross each other). The marginal rays a and b are deviated in such a way that the radial forces cancel each other and the axial ones push towards the focus of the laser beam. Other pairs of symmetric rays of the beam have different intensities and also contribute to the axial total force exerted on the bead. **(c)** A particle located after the focus of the laser beam. The force exerted has an opposite direction than in panel b and pulls the particle towards the focus. **(d)** A particle located off-axis. The different intensity between the centered and off-axis rays induce the radial restoring force. Therefore, the particle still tends to align with the axis of a focused (not collimated) Gaussian beam.

The Rayleigh regime

In this regime, the trapped particle is much smaller than the wavelength of the light beam. The electromagnetic field is uniform within the particle and it can be treated as a small spherical dipole. Although very few experiments in biophysics are done under this regime, it is useful because it provides simple and separate expressions for the scattering and gradient forces. The scattering force is the result of the change in light momentum after the interaction with the particle, which absorbs and reemits the light in all directions (according to Rayleigh dispersion). The scattering force of a spherical dipole in a medium of index of refraction n_m is given by [8]:

$$\vec{F}_{\text{scat}} = n_m \frac{\sigma I_0}{c} \hat{k} \tag{2.3}$$

where I_0 is the intensity of the laser beam, c is the speed of light, $\hat{k}$ is the unitary vector that points along the direction of propagation of the light beam, and σ is the scattering cross section of a Rayleigh particle which in

the case of a spherical particle of radius r is given by:

$$\sigma = \frac{128\pi^5 r^6}{3\lambda^4}\left(\frac{m^2-1}{m^2+2}\right)^2 \tag{2.4}$$

where m is the effective index of the particle (i.e., $m = n/n_m$, the index of the particle n divided by the index of the medium n_m) and λ is the wavelength of the light. Since the particle is much smaller than the wavelength, only the spatial distribution of the light intensity ($I_0(\vec{r})$) determines the spatial distribution of the scattering force ($\vec{F}_{\text{scat}}(\vec{r})$). Therefore, the scattering force is stronger in those regions where the intensity of light is higher.

The gradient force is due to the Lorentz force acting on the dipole, induced by the electromagnetic field. The time-averaged gradient force is given by [8, 79]:

$$\vec{F}_{\text{grad}} = \frac{2\pi\alpha}{cn_m^2}\vec{\nabla} I_0 \tag{2.5}$$

where α is the polarizability of the trapped particle. The polarizability of a dielectric particle is a magnitude that relates the applied external electric field ($\vec{E}$) and the electric dipole moment ($\vec{p}$) induced on the particle ($\vec{p} = \alpha\vec{E}$). In the case of a isotropic dielectric spherical particle of radius r, the polarizability is given by the Lorentz-Lorenz formula [80],

$$\alpha = n_m^2 r^3 \left(\frac{m^2-1}{m^2-2}\right) \ . \tag{2.6}$$

Equations 2.3 and 2.5 allow us to know the evolution of a Rayleigh particle illuminated by a light beam. The gradient force pushes the particle towards the highest light intensity (for $m > 1$) whereas the scattering force pushes it along the direction of light propagation. In order to trap particles, the gradient force along the optical axis must be higher than the scattering force. By increasing the numerical aperture, higher intensity gradients can be achieved and stronger gradient forces are exerted on the particle (see Fig. 2.4). This regime allows us to calculate the force exerted on a dielectric sphere under illumination of a paraxial Gaussian beam (i.e., a weakly focused beam which makes a small angle to the optical axis of the system). The strategy consists in writing the intensity of a Gaussian beam and introducing it in equations 2.3 and 2.5 [81]. Although the prediction of the optical forces fails when the size of the particle is of the order of the wavelength of the laser beam, the qualitative description of the optical trap is quite illustrative.

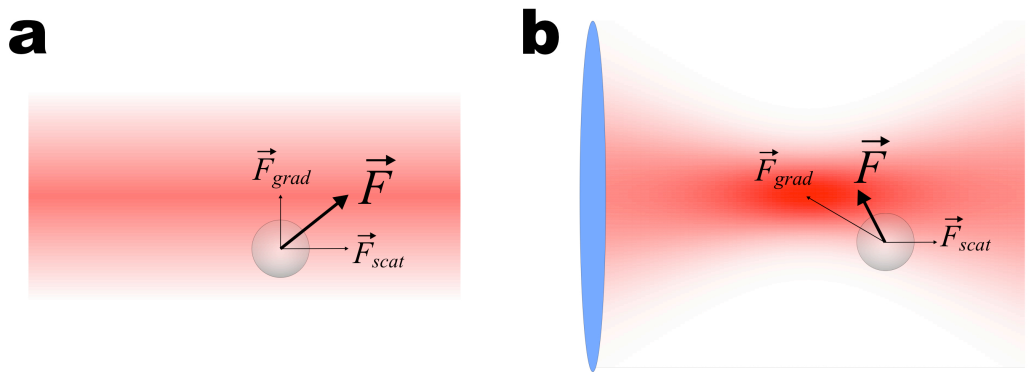


Figure 2.4: Forces arising in the Rayleigh regime. **(a)** A collimated Gaussian beam exerts a gradient (radial) force onto a particle that pushes the particle towards the beam axis. However, the scattering force pushes the particle in the direction of light. **(b)** A tightly focused Gaussian beam develops a strong intensity gradient along the propagation direction that prevents the particle from being pushed away by the scattering. The result is a restoring force and the optical trapping of the particle.

Generalized Lorenz-Mie Theory (GLMT)

The physics of the optical tweezers used in the common biological applications falls between the ray optics and Rayleigh regimes and cannot be described correctly with the previous approaches. In this intermediate regime, the computation of forces requires a whole characterization of the laser beam and the dielectric particle. Since the laser beam is highly focused, it can no more be described as a paraxial Gaussian beam. Under these conditions, the electromagnetic field of the laser beam shows its vector character and a scalar description is no longer valid. Moreover, the dielectric particle interacts with the incident laser beam and develops internal and scattered electromagnetic fields. In conclusion, the computation of forces requires the calculation of the resulting electromagnetic field after the particle has interacted with the laser beam. To do so, a boundary value problem for the electromagnetic field equations must be solved.

The solution to this problem was first calculated by Mie [57], who computed the resulting electromagnetic field of a plane monochromatic wave scattered by a spherical particle. The procedure is to obtain a general solution for the electromagnetic field equations and then impose the appropriate boundary conditions [80]. The calculation consists in writing the Maxwell equations in spherical coordinates and rewrite them in terms of two scalar fields ${}^{e}\Pi$ and ${}^{m}\Pi$ known as Debye's potentials. These potentials are two independent convenient solutions: the transverse electric (TE) field solution, which has zero radial electric field component; and the transverse magnetic

(TM) field, which has zero radial magnetic field component. Both potentials obey a wave differential equation and the solution for ${}^{i}\Pi$ ($i = e, m$) is expressed as a linear combination of solutions in spherical coordinates, where the boundary conditions (continuity of the tangential electric and magnetic fields over the spherical surface of the particle) have already been imposed. The final step is to find the coefficients of the general solution that describe the incident electromagnetic field and obtain the coefficients for the scattered and internal electromagnetic field. The general solution is expressed in terms of a dimensionless parameter $q = 2\pi r/\lambda$, where λ is the wavelength of the light and r is the radius of the trapped particle (see Fig. 2.5).

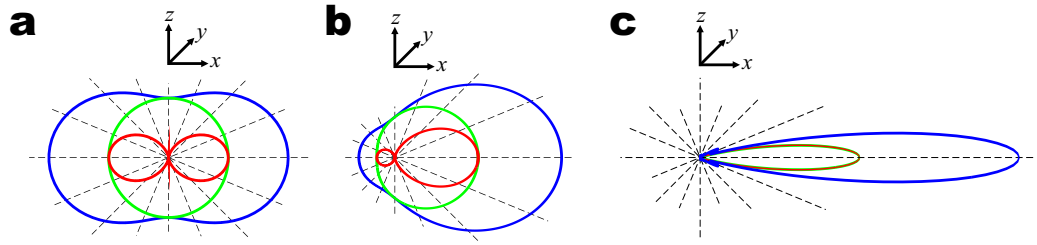


Figure 2.5: GLMT solution for different values of q obtained from Ref. [82]. A monochromatic planar wave polarized along the z axis propagates along the x axis and interacts with a particle (not depicted) that is located at the center of the diagram. The scattered light has different angular intensities and polarizations. The red curve shows the intensity of the scattered light that is parallel to the incident wave; the green curve shows the intensity of the scattered light that is perpendicular to the incident wave; and the blue line shows the combination of both (useful when considering circularly polarized light). **(a)** This diagram corresponds to $q = 0.1$, a situation close to the Rayleigh regime ($r \ll \lambda$) where the scattered light is quite isotropic. **(b)** $q = 1$. The scattered light shows an asymmetry between the forward and backward direction. **(c)** Diagram observed in the Mie regime ($q = 10$), where the backscattering almost vanishes.

The GLMT simply extends the previous calculation (i.e., Mie's calculation) to arbitrary incident waves [83] and to arbitrary positions of the particle [84]. The resulting force that acts on the particle can be determined from the calculated electromagnetic field. It has to be done by integration of the Maxwell's stress tensor $\overleftrightarrow{T}$ over a surface A enclosing the particle:

$$\langle \vec{F} \rangle = \left\langle \oint_A \overleftrightarrow{T}\, d\vec{a} \right\rangle \tag{2.7}$$

where $d\vec{a}$ is the surface element vector pointing outwards and $\langle ... \rangle$ represents a time average. A detailed derivation of equation 2.7 is shown in Appendix A. The Maxwell's stress tensor represents the flux of linear momentum of the

electromagnetic field and it is obtained from:

$$\overleftrightarrow{T}_{ij} = \epsilon \left(E_i E_j - \frac{1}{2}\delta_{ij} E^2 \right) + \frac{1}{\mu} \left(B_i B_j - \frac{1}{2}\delta_{ij} B^2 \right) \tag{2.8}$$

where E_i is the i-th component of the electric field, B_i is the i-th component of the magnetic field, ϵ is the electric permittivity and μ is the magnetic permeability of the surrounding medium. Following this methodology, the calculated net radiation force and torque exerted on a spherical particle by a highly focused Gaussian beam is qualitative consistent with experimental measurements [85]. However, the quantitative agreement between the theory and the experiments is unsatisfactory. The origin of such discrepancy is attributed to several causes: irregular shape or roughness of the particle, complex index of refraction, bad conditioning of the laser beam, etc. In the end, the achieved accuracy does not compensate the effort of going through the calculation. Eventually, the forces exerted by the laser beam must be determined empirically.

Fortunately, in optical tweezers the restoring force near the focus is proportional to the displacement of the particle from the center of the optical trap. This linearity allows to define a trap stiffness, which can be used to determine the applied force by measuring the position of the particle. In other words, the trap can be treated as a harmonic potential well or a *virtual spring*.

Still, there is another way to directly measure optical forces. The method consists in detecting the difference in the linear momentum of light before and after it has interacted with the particle. This is explained in the next section.

2.1.2 Conservation of linear momentum

The conservation of linear momentum is a fundamental principle in physics. It states that the total linear momentum of an isolated system remains constant. The momentum can be transferred between the different constituents of the system, provided that the momentum lost by one element is equal to the momentum gained by another. The principle can also be extended to open systems. In this case, the net change of momentum of the system equals the flux of momentum entering minus that exiting the system.

The principle is not only held by interacting particles but also by fields that carry momentum. Now, a balance equation for the conservation of the linear momentum can be written for every physical system. Under certain conditions, the conservation of linear momentum is useful to measure the forces involved in optical trapping. Starting from the Lorentz force and using

the Maxwell's equations, one can write an expression for the conservation of linear momentum of an object that interacts with an electromagnetic field within a volume V (see Appendix A for the derivation),

$$\frac{d}{dt}\left(\vec{P}_{\text{mech}} + \vec{P}_{\text{field}}\right) = \vec{\Phi}_{\text{field}} \tag{2.9}$$

where $\vec{P}_{\text{mech}}$ is the linear momentum of the object, $\vec{P}_{\text{field}}$ is the linear momentum of the electromagnetic field within the volume V and $\vec{\Phi}_{\text{field}}$ is the flux of linear momentum carried by an electromagnetic wave across the surface A that encloses the volume V. The left hand side of the equation accounts for the overall change of linear momentum inside the volume V and the right hand side accounts for the net flux of linear momentum than enters the volume V across the surface A. Figure 2.6a shows an schematic representation of the three terms of equation 2.9. The time variation of the linear momentum of the object is the force exerted on it

$$\vec{F} = \frac{d\vec{P}_{\text{mech}}}{dt} \ . \tag{2.10}$$

The total linear momentum of the electromagnetic field can be computed by summing the density momentum of the field $\vec{g}$ within the volume V (see Appendix A)

$$\vec{P}_{\text{field}} = \int_V \vec{g} \cdot dV = \int_V \frac{\vec{S}}{c^2} dV \tag{2.11}$$

where $\vec{S}$ is the Poynting vector calculated from the electromagnetic field $\vec{S} = \vec{E} \times \vec{H}$) and c is the speed of light. The total flux of linear momentum[1] $\vec{\Phi}_{\text{field}}$ is calculated by summing the Maxwell stress tensor $\overleftrightarrow{T}$ over the surface A

$$\vec{\Phi}_{\text{field}} = \oint_A \overleftrightarrow{T} \cdot d\vec{a} \ . \tag{2.12}$$

The Maxwell's stress tensor accounts for the flux of electromagnetic linear momentum per unit of area[2] and it is derived from the electromagnetic fields according to Eq. 2.8.

[1] Note that a flux of linear momentum has units of force:

$$[\text{Flux of momentum}] \equiv \frac{[\text{Momentum}]}{[\text{Time}]} = \frac{[\text{Mass}][\text{Velocity}]}{[\text{Time}]} = \frac{[\text{Force}][\text{Time}]}{[\text{Time}]} = [\text{Force}]$$

[2] It has units of pressure:

$$\frac{[\text{Flux of momentum}]}{[\text{Areal}]} = \frac{[\text{Force}]}{[\text{Area}]} = [\text{Pressure}]$$

By using equations 2.10, 2.11 and 2.12 we can rewrite equation 2.9 as

$$\vec{F} = \oint_A \overleftrightarrow{T} \cdot d\vec{a} - \frac{1}{c^2}\frac{d}{dt}\int_V \vec{S} dV \tag{2.13}$$

which provides a recipe to calculate the electromagnetic force exerted on an object from the resulting electromagnetic fields of the interaction.

Assuming that the electromagnetic field is stationary within the volume V (i.e., all the temporal dependence is due to the complex phase $\exp(i\omega t)$), the second term of the right side vanishes if a temporal average is performed. The resulting expression is equal to Eq. 2.7. Under these conditions, equation 2.9 can be rewritten as,

$$\vec{F} = \vec{\Phi}_{\text{field}} \tag{2.14}$$

which is a simple expression for the conservation of light momentum (we do not write the temporal average $\langle ... \rangle$). The force exerted on a particle is equal to the stationary flux of linear momentum carried by the interacting electromagnetic field. In the next section it is described how to measure the flux of linear momentum.

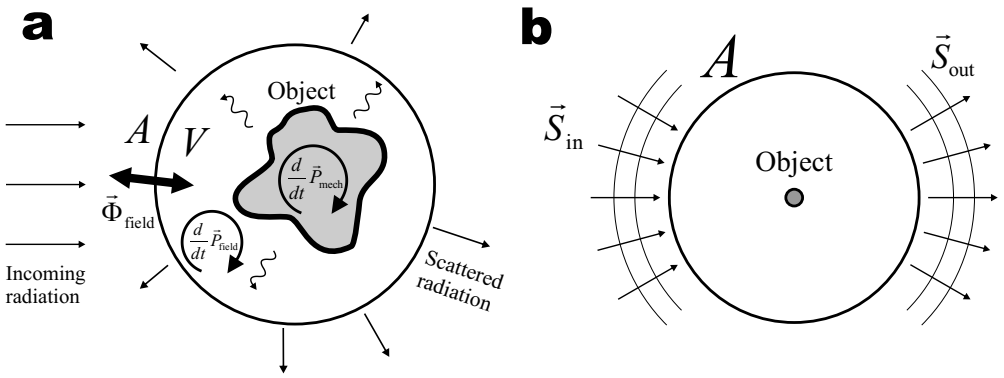


Figure 2.6: Conservation of light momentum. **(a)** The surface A encloses the volume V where the radiation interacts with the material object. The net flux of momentum $\vec{\Phi}_{\text{field}}$ that enters/exits the volume is equal to the change in the linear momentum of the internal radiation ($\vec{P}_{\text{field}}$) plus the change in the linear momentum (i.e., the force) of the objects ($\vec{P}_{\text{mech}}$). **(b)** In a stationary state, the total flux of linear momentum of light (i.e., $\vec{S}_{\text{in}} - \vec{S}_{\text{out}}$) that crosses the surface A is equal to the force applied to the object.

2.1.3 Measurement of linear momentum

In this section, we will assume that the diameter of the particle d is larger than the wavelength λ of the electromagnetic field (i.e., the laser light). Thus

the system falls into the ray optics regime (see Sec. 2.1.1) and the laser can be visualized as a bundle of rays that interact with the particle [86]. It allows us to calculate the Maxwell Stress Tensor by an alternative way. *A priori*, one might think that the results of this section are only valid when the Mie regime is valid. However, it has been check empirically that an optical tweezers instrument built under these assumptions still works properly when the approximation is not fulfilled [86, 16].

As deduced from the previous section, the force exerted on an object by an electromagnetic field is equal to the net flux of linear momentum that enters the object (Eq. 2.14). The net momentum flux that crosses the surface A can be written as an integral sum of the density of flux momentum[3] $\vec{\phi}_{\text{field}}$ of the electromagnetic wave, provided the elements of surface da are normal to $\vec{\phi}_{\text{field}}$,

$$\vec{\Phi}_{\text{field}} = \oint_A \vec{\phi}_{\text{field}}\, da\ . \tag{2.15}$$

We ended the previous section with the problem of measuring the total flux of momentum and now we are facing the problem of measuring the density of flux momentum. It seems we did not make much progress. However, in the ray optics regime, the laser light has a defined direction of propagation and the wave can be decomposed into rays. It allows us to write the density of flux momentum of a light wave as (see Appendix B)

$$\vec{\phi}_{\text{field}} = \frac{n_m}{c}\vec{S} \tag{2.16}$$

where n_m is the index of refraction of the surrounding medium, c is the speed of light and $\vec{S}$ is the Poynting vector. Equation 2.16 is quite relevant, since the temporal average value of the Poynting vector $\langle\vec{S}\rangle$ is the intensity[4] of the light I and it can be directly measured by a photodetector. Note that what we have done up to now is to write the Maxwell stress tensor for the ray optics regime (Eq. 2.16 is a restricted version of Eq. 2.8 and Eq. 2.15 is the analogous of Eq. 2.12).

Finally, combining equations 2.16, 2.15 and 2.14 we can write the total optical force exerted on an object as

$$\vec{F} = \frac{n_m}{c}\oint_A \vec{S}\, da \tag{2.17}$$

[3]It has units of pressure:

$$[\text{Density of flux momentum}] \equiv \frac{[\text{Flux momentum}]}{[\text{Area}]} = \frac{[\text{Force}]}{[\text{Area}]} = [\text{Pressure}]$$

[4]It has units of [Power]/[Area].

where the integral is extended on a surface A surrounding the object. $\vec{S}$ and the surface A must be perpendicular because Eq. 2.16 is a heuristic calculation of the Maxwell Stress Tensor that does not include the off-diagonal elements. The off-diagonal elements produce shear stresses that act parallel to the surface and they do not contribute to the total force if the surface of integration is perpendicular to the direction of the momentum flux. Finally, the Poynting vector has two contributions $\vec{S} = \vec{S}_{\text{in}} - \vec{S}_{\text{out}}$ that correspond to the light entering ($\vec{S}_{\text{in}}$) and the light leaving ($\vec{S}_{\text{out}}$) the volume enclosed in the surface A (see Fig. 2.6b).

2.1.4 A force transducer

A transducer is a device that transforms a physical magnitude into another. In this section, we describe a transducer that converts the intensity of light $\langle\vec{S}\rangle$ into a force $\vec{F}$ according to equation 2.17. We also consider that the transducer operates under the two following experimental conditions (see Fig. 2.7a):

(i) The incoming laser light $\langle\vec{S}_{\text{in}}\rangle$ is a focused spherical wave.

(ii) The outcoming light $\langle\vec{S}_{\text{out}}\rangle$ is spherical and emanates from the focus of the incoming light, where the trapped particle is also located.

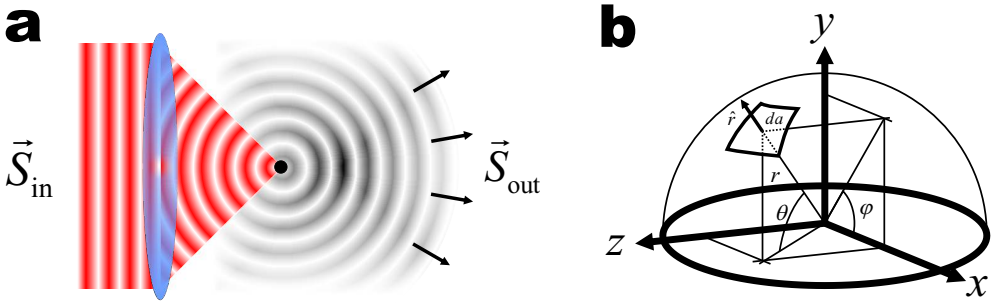


Figure 2.7: **(a)** The collimated light of the laser (in red) is focused by an objective (blue lens) producing a spherical converging beam. A particle is located at the focus and the light is scattered in all directions. The outcoming light is a diverging spherical wave. **(b)** Spherical coordinates and element of area.

The first condition can be easily achieved by focusing a collimated laser beam with a lens. The second condition is achieved when the outcoming light is observed far away from the source. It can be shown from the GLMT (see

Sec. 2.1.1) that the scattered wave has no radial components (therefore it is a spherical wave) in the so-called radiation zone, where the distance to the particle is much larger than the wavelength ($|\vec{R}| \gg \lambda$). Both conditions allow us to set the origin of coordinates at the focal point of the laser and write the intensity of light in terms of the spherical coordinates $\langle \vec{S}(r,\theta,\varphi)\rangle$. Since the incoming and outcoming lights are spherical, we can define the angular intensity distribution of light[5] (see Fig. 2.7b)

$$\frac{I(\theta,\varphi)\ \hat{r}}{r^2} = \langle \vec{S}(r,\theta,\varphi)\rangle \tag{2.18}$$

which is independent of the radius r of observation ($\hat{r}$ is a unit vector from the focus). Introducing Eq. 2.18 into Eq. 2.17 we can write

$$\vec{F} = \frac{n_m}{c} \oint_A I(\theta,\varphi) \left(\hat{\imath}\sin\theta\cos\varphi + \hat{\jmath}\sin\theta\sin\varphi + \hat{k}\cos\theta\right) d\Omega \tag{2.19}$$

where the differential of area ($da = r^2 d\Omega$) and the unit vector ($\hat{r} = \hat{\imath}\sin\theta\cos\varphi + \hat{\jmath}\sin\theta\sin\varphi + \hat{k}\cos\theta$) have been expressed in spherical coordinates. Again, $I(\theta,\varphi)$ is positive for rays entering the trap and negative for rays leaving it. The measurement of $I(\theta,\varphi)$ cannot be done directly from the emanating rays. It is necessary the project the angular distribution of intensities into a planar photodetector D, which can be achieved by using a condenser lens (see Fig. 2.8a). In a perfect condenser lens of focal length f, there is a univocal relation between the direction of a ray that emanates from the focus of the lens and the position of the ray at the image principal focal plane P (see Fig. 2.8a). This relationship is known as the Abbe sine condition and it is given by

$$\rho = f n_m \sin\theta \tag{2.20}$$

where ρ is the radial distance of the ray to the optical axis at the principal plane, n_m is the index of refraction of the medium, and θ is the angle of the emanating ray. On the other hand, if all the light is collected by the condenser lens, the conservation of the energy establishes a relationship between the total power of light that emanates from the focus and the power of light collected by the lens

$$\begin{aligned} \oint_A I(\theta,\varphi)\ d\Omega &= \int_P E(\rho,\varphi) da' \\ I(\theta,\varphi)\ d\Omega &= E(\rho,\varphi) da' \end{aligned} \tag{2.21}$$

[5]The angular intensity distribution of light has units of [Power]/[Solid angle]. As the spherical wave propagates, the total power of light is distributed on a spherical surface of area $A = 4\pi r^2$. Thus, the total amount of power P is independent of r and so it is constant: $P = \int \vec{S}(r,\theta,\varphi)\cdot d\vec{a} = \int \left(\frac{I(\theta,\varphi)\hat{r}}{r^2}\right)\cdot\left(r^2\sin\theta\, d\theta d\phi\ \hat{r}\right) = \int I(\theta,\varphi) d\Omega$, where $\hat{r}$ ($\hat{r}\cdot\hat{r} = 1$) is the radial unitary vector and $d\Omega = \sin\theta\, d\theta\, d\phi$ is the element of solid angle.

where $E(\rho,\varphi)$ and $da' = \rho\ d\rho\ d\varphi$ are the irradiance[6] and the element of area at the image focal plane, respectively. In Section 2.2.1 it is shown that the focusing lens is also useful to collect the backscattered light. Therefore, the integral sum over the image principal plane P of the condenser lens is extended to the image principal plane of the focusing lens, too. By introducing Eqs. 2.21 and 2.20 into Eq. 2.19 we can write the expression for the force as:

$$\vec{F} = \frac{1}{c}\int_P E(\rho,\varphi)\left(\hat{\imath}\frac{\rho}{f}\cos\varphi + \hat{\jmath}\frac{\rho}{f}\sin\varphi + \hat{k}\ n_m\sqrt{1-\left(\frac{\rho}{fn_m}\right)^2}\right)\rho\ d\rho\ d\varphi \tag{2.22}$$

where the integration is now taken over the surface of the image principal plane (P). Note that the resulting prefactor in the calculation of the z force is obtained from Eq. 2.20

$$\cos\theta = \sqrt{1-\sin^2\theta} = \sqrt{1-\left(\frac{\rho}{fn_m}\right)^2}\ . \tag{2.23}$$

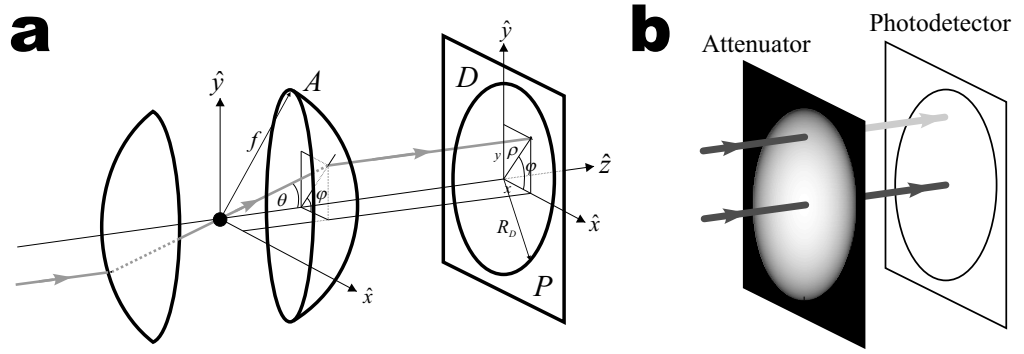


Figure 2.8: (a) Projection of the scattered light onto a photodetector. An incoming ray (in gray) is focused by the left lens and scattered by the particle. The outcoming ray is collected by the condenser lens (A) and redirected to a photodetector D. Actually, the condenser lens produces a change of coordinates, from spherical (θ,φ) to cylindrical (ρ,φ). **(b)** Measurement of axial (z) force. The attenuator has a transmission profile that looks like a bull's eye. An axial ray is fully transmitted to the photodetector. An off-axis ray, suffers an attenuation and its intensity is lower when it arrives at the photodetector.

Now, the forces exerted on the x and y directions (radial force) can be inferred from the measurement of a photodetector placed at the image principal plane P. A Position Sensitive Detector (PSD) is a device that produces two output signals (D_x,D_y) proportional to the irradiance of the light

[6]The irradiance has units of [Power]/[Area]. It is a magnitude equivalent to the intensity.

($E(x,y)$) weighted by the relative position from the center of the detector (x/R_D,y/R_D) according to

$$D_x = \Psi \int_D E(x,y)\frac{x}{R_D}\, da' = \Psi \int E(\rho,\varphi)\frac{\rho\cos\varphi}{R_D}\, da'$$
$$D_y = \Psi \int_D E(x,y)\frac{y}{R_D}\, da' = \Psi \int E(\rho,\varphi)\frac{\rho\sin\varphi}{R_D}\, da' \tag{2.24}$$

where Ψ is the sensitivity of the detector, R_D is the effective radius of the detector and the integrals are over the surface (D) of the detector (see Appendix C). Combining Eqs. 2.24 and 2.22 we can write the force in terms of the response of the detector

$$F_x = \frac{D_x R_D}{c\Psi f}$$
$$F_y = \frac{D_y R_D}{c\Psi f} \tag{2.25}$$

The force exerted on the z axis (axial force) can be measured provided that the irradiance of light is weighted by the factor of Eq. 2.23. It can be achieved by placing an attenuator in front of a photodetector with a transmission profile given by Eq. 2.23 (see Fig. 2.8b). The output signal of the photodetector is given by

$$D_z = \Psi \int_D E(\theta,\varphi)\sqrt{1-\left(\frac{\rho}{n_m f}\right)^2}\, da' \tag{2.26}$$

and the z force can be written as

$$F_z = \frac{D_z n_m}{c\Psi} \tag{2.27}$$

Eqs. 2.25 allow us to directly measure the transverse force (on the x-y plane) of a trapped particle, provided that all the scattered light can be collected. What is relevant from these expressions is that calibration does not depend on the refractive index of the medium (n_m), the radius of the particle nor the laser power. The factors involved in Eqs. 2.25 are constant in a typical experimental setup (R_D, f, Ψ). On the contrary, the measurement of the axial force in Eq. 2.27 depends on the refractive index of the medium and this dependency cannot be avoided. However, the error in the z force is smaller that 0.25% when the refractive index of the medium changes from $n_m = 1.334$ (pure water) to $n_m = 1.343$ (1 M NaCl buffer).

2.2 The Minitweezers experimental setup

Most of the optical tweezers setups are based on modified commercial microscopes to which a laser beam is introduced in the optical path to produce an optical trap [79]. Minitweezers is the name given by its inventor, Steve Smith, to the experimental setup used in this work [16, 17, 18]. This setup is a miniaturized evolution of previous instruments and achieves better stability and resolution in the measurements. Its compactness allows to hang it from the ceiling and isolate the instrument to prevent that vibrations and air flows affect the experiments. Minitweezers measure force by conservation of light momentum according to the ideas developed in Section 2.1.4. Therefore, the instrument was designed not only to produce an optical trap, but also to collect all the light emerging from the interaction between the laser light and the trapped particle. It requires the use of two microscope objectives: one to focus the laser beam and the other one to work as a condenser that collects the outcoming light. Thus, Minitweezers look like two microscopes facing each other with a common focal point. The most relevant innovation, though, is the steering of the laser beam, which is performed using a device called *wiggler*. The wiggler redirects the laser beam and repositions the optical trap by using piezoelectric crystals that provide high reliability and stability. The optical system is enclosed in the so-called head (see Figure 2.9). The control of the instrument and the measurement of signal is performed with customized electronic boards (the electronic controller) that use PIC microcontrollers. The experimentalist also interacts with the instrument with a Mac computer (also known as the host computer), which communicates with the electronic boards through a USB port. The following sections describe the parts of the experimental setup.

2.2.1 Optics

The formation of an optical trap using a single-beam laser requires focusing the laser with high Numerical Aperture (NA) so that the gradient force compensates the scattering force (see Section 2.1.1). Under this condition, it is very difficult to collect all the deflected light that emerges from the optical trap. Thus it is not possible to perform a measurement of force based on the change of the linear momentum of the light. The Minitweezers consist of two counter-propagating infrared laser beams that form a single optical trap. The advantage of using counter-propagating beams is that the scattering forces cancel out. This allows to reduce the NA of the lasers (i.e., the diameter of the laser beam) and all the exiting light can be collected. Now, the use of two low-NA laser beams with high-NA focusing lenses is the key to collect

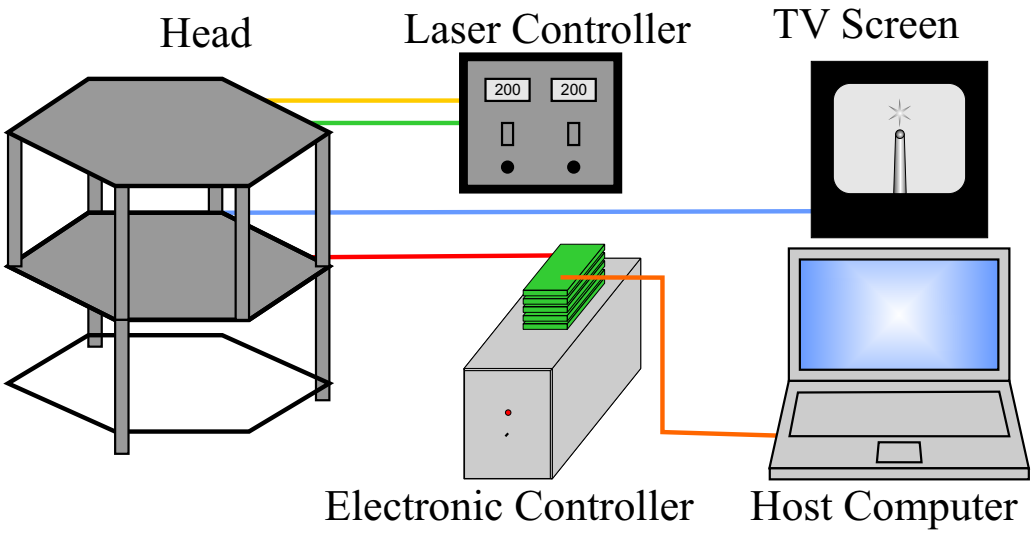


Figure 2.9: General scheme. The experiments are performed in the head, which contains all the optics. The lasers are powered by the laser controller. The electronic controller communicates the head and the host computer. The user interacts with the experiment by means of the host computer. The TV screen allows to visualize the experiment.

and measure the change in the light momentum.

Apart from having a calibration independent of several experimental conditions (bead size, index of refraction, etc.) there are other advantages in the double-beam optical trap. First, the laser beams do not need to be highly focused, which minimizes the effect of spherical aberration of the lenses. A lens with spherical aberration focuses the marginal rays of a laser beam more tightly than the rays near the optical axis, which produces a blurred focal point. Since the intensity of a low NA laser beam is concentrated near the optical axis, such beam is less affected by spherical aberration. Moreover, a low focused beam has a longer focal distance, which makes possible to focus the laser beam deeper inside the fluidics chamber. It allows to reduce the hydrodynamic effects of the boundaries of the fluidics chamber (i.e., the coverslips) on the particle trapped in the optical trap. Apart from the optical considerations, the use of low focused lasers reduces the heating of the medium by infrared absorption. Another advantage is that the instrument can operate in the two-traps mode. On the other hand, the double-beam is quite difficult to implement because it requires an accurate alignment of the laser beams. In fact, small misalignments produce optical traps that induce non-uniform forces on the trapped particles. An accurate alignment is achieved by electronically assisted feedbacks. Feedbacks such as *autoalign* read the PSD measurements and lightly reposition the optical traps in order

to cancel any misalignment between the focuses of the laser beams.

Optical schematics

The schematics of optics of the Minitweezers is symmetric for each laser. Both lasers share part of the optical path but are produced, controlled and measured independently. This section describes the optics of one laser and it is assumed that the other one has a symmetric schematics if it is not mentioned. Following, the different optical parts are described along with Fig. 2.10.

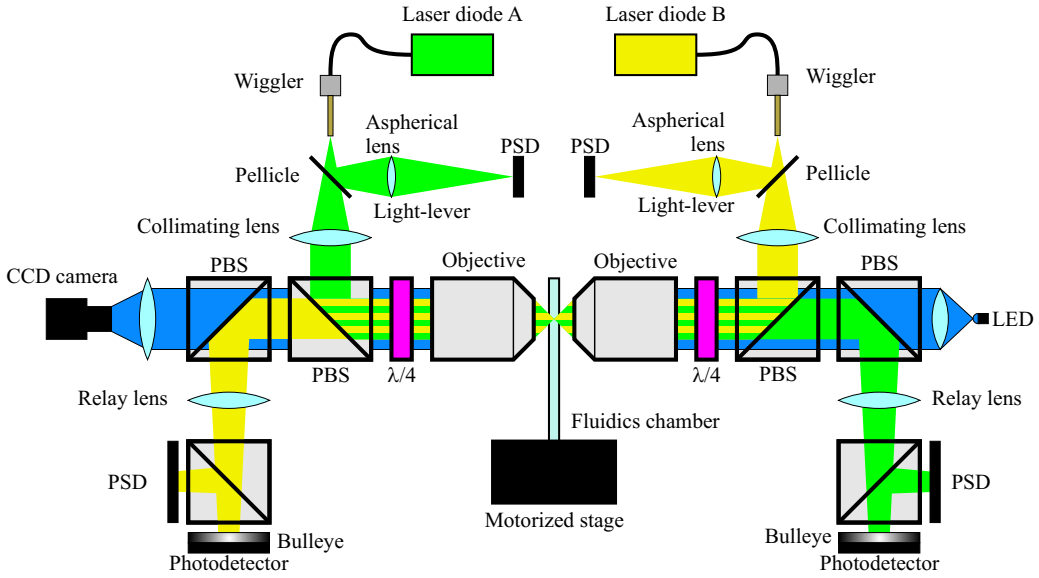


Figure 2.10: Optical schematics (see text). The optical paths of the lasers are depicted in green and yellow. The optical path of the imaging system is depicted in blue.

1. **Generation and conditioning of the laser beam.** A near-infrared laser diode of 845 nm wavelength is used to generate the optical trap. This is a suitable wavelength because the absorption of near-infrared radiation is quite low for water, which is the medium at which the experiments are performed. This prevents the heating of the sample. The laser diode (Lumix LU0845M200) is a single-mode fiber-coupled device of 200 mW of power that produces the fundamental transverse electromagnetic mode (TEM_{00}). The TEM_{00} mode has a Gaussian profile and it is linearly polarized. The intensity of the laser beams and the temperature of the laser diodes are controlled by an independent laser power supply (Appendix C.1). The laser diode is connected to an optical fiber that feed power to the wiggler.

2. **Beam steering and position detection.** The wiggler is a mechanical device that uses piezoelectric crystals to gently push the tip of the optical fiber and redirect the laser beam at will (Appendix C.2). The tip of the fiber can be repositioned throughout the perpendicular plane to the direction of propagation of the laser beam but not along this direction. This is useful to form the optical trap at different locations. About 8% of the the light emerging from the tip of the optical fiber is split by using a pellicle beam-splitter to form a *light-lever* that measures the position of the beam. The beam of the light-lever is projected to a Position Sensitive Detector (PSD) (see Appendix C.3) by means of a refocusing aspherical lens (aspherical lenses allow to focus highly diverging beams without introducing aberrations). The 92% of the remaining light is collimated by using a lens and it is introduced into the optical axis by using a Polarizing Beam-Splitter (PBS) that selects the horizontally polarized light to form the optical trap.

3. **Formation of the optical trap.** Once the light is into the optical path, a quarter-wave plate produces circular polarization before the beam is focused by a water-immersion microscope objective with NA=1.2 (Olympus UPLSAPO 60×W). The circular polarization ensures that the light scattered is an average between parallel and perpendicular polarized light (see Figure 2.5). By doing this, we ensure that the force exerted on the trapped particle will not depend on the polarization of the incident light. The exiting light is collected by the opposite objective and it is converted to vertically polarized light by another quarter-wave plate. The vertically polarized light can be extracted from the optical path using two PBSs and one relay lens that redirect the light to the PSD that measure the intensity of the beam. The first PBS is common to the other laser beam and performs a double task: 1) Redirects the vertically polarized light to the second PBS and 2) Introduces the other laser beam coming from the wiggler into the optical path. The second PBS projects the vertically polarized light to the photodetectors. So, the use of quarter-wave plates ensures that the light coming from the two laser beams do not interact with each other and they are correctly guided along their optical paths from the wiggler to the photodetectors. Moreover, the circularly polarized light guarantees that the light reflected from the particle is not returned to the wiggler but is reflected to the opposite photodetector.

4. **Detection and measurement of light.** The light redirected to the photodetectors is split into two parts using a (non-polarizing) beam-

splitter. One part is redirected to a PSD (OSI Optoelectronics, DL-10) (see Appendix C.3) that measures the deflection of the beam in the transverse direction, which is a measurement of the transverse force (i.e., the x and y forces). The other part is redirected to a bullseye filter and measured with a photo-diode (OSI Optoelectronics, PIN-10DI). The bullseye is a realization of the attenuator described in Figure 2.8b and it has the right profile to measure the force along the longitudinal axis, i.e., the z axis force.

5. **Imaging system.** A blue LED (wavelength 470 nm) and a CCD camera (Watec WAT-902H3 SUPREME EIA) are used to form a microscope to view the experiment (trapped particles and pipette). The light of the LED is expanded with a lens to produce a Köhler illumination system at the focal plane of the laser beams so that the field of view is uniformly illuminated. At the other side of the optical path, a lens projects the image of the focal plane to the CCD camera. The CCD camera is monochrome and it is visible and near-infrared sensitive, which permits to view both the experiment (i.e., the micropipette, the beads, etc.) and the laser beams. The intensity of the laser beam is so high that it saturates the CCD camera and the image on the TV monitor looks completely blank. Under these conditions, the elements of the fluidics chamber such as the micropipette cannot be observed and the experimentalist is not able to carry out the experiments. To solve this problem, an infrared filter is placed in the front camera. It prevents the laser spots from reaching the CCD camera but it lets the blue light of the LED pass. It allows to clearly see the micropipette, the beads and the dispenser tubes of the fluidics chamber. This is necessary to carry out the experiments.

6. **Fluidics chamber.** Optical trapping is done in a fludics chamber that is homemade with two microscope coverslips sealed with a nescofilm gasket in between. The gasket is cut to form channels in order to introduce and guide the particles into the chamber. The dimensions of the chamber are 24 mm$\times$60 mm$\times$200 μm and it also contains a micropipette to grab particles by suction (see Appendix C.4). The chamber is held by a motorized xyz stage with fine positioning control (< 0.5 μm) and it is placed between the two microscope objectives, where the optical trap is formed.

Ray tracing

Although in our case geometrical optics does not describe optical trapping correctly, it provides a qualitative understanding of the phenomenon. Ray tracing is useful to depict the manner in which the instrument measures distances (light-lever), radial forces (force PSD) and axial forces (Iris). Indeed, each combination of these 3 independent magnitudes produce a different ray tracing, which can be measured with the photodetectors. Figure 2.11 shows the ray tracing of six elementary situations and the corresponding reading of the detectors.

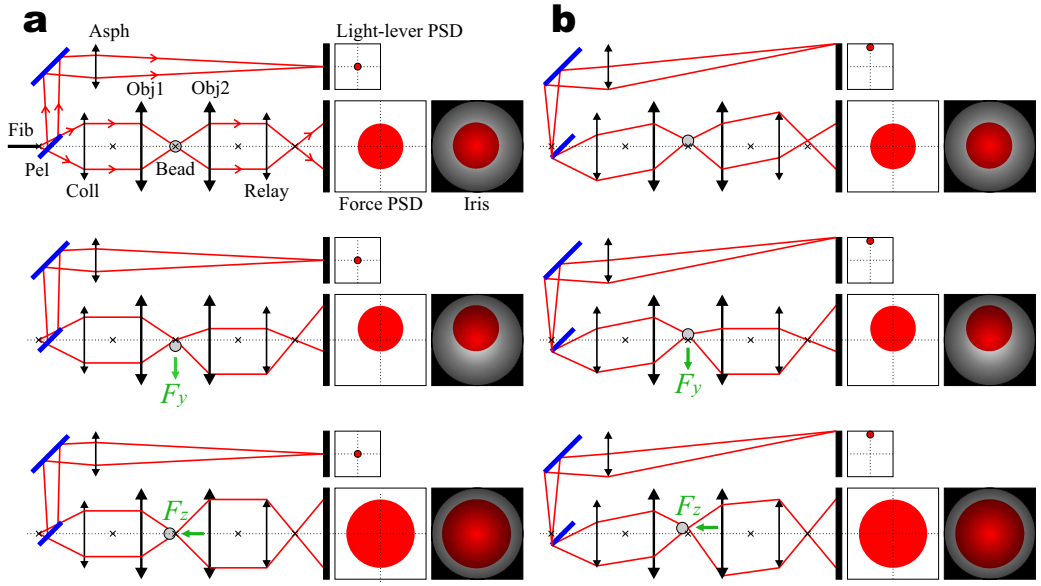


Figure 2.11: Ray tracing (see text). Only the marginal rays of one of the lasers are depicted. Some prisms and mirrors are not shown and all the lenses are assumed to be thin. The optical fiber is only shown in the top panel **a**. Pictures are not to scale. Legend: Asph=Aspherical lens; Obj1=objective 1 (focuser); Obj2=objective 2 (condenser); Coll=collimating lens; Pel=pellicle beamsplitter; Fib=optical fiber. The light travels from the optical fiber (left) to the detectors (right). The crosses in the optical path indicate the focal points of the lenses. The Coll and the Obj1 lenses form and infinite system (i.e., their focuses coincide and the light beam is parallel), as well as the Obj2 and the Relay lenses. The arrow next to the bead indicates the application of an external force. **(a)** Centered optical trap. **(Top)** Zero force. **(Middle)** Application of a radial external force (F_y). **(Bottom)** Application of an axial external force (F_z). **(b)** Optical trap at a different position. Note that the measurement of F_z is affected by that of F_y. Indeed, a y force induces a deflection on the outcoming light that hits the Iris detector off-axis. So the bullseye does not attenuate the laser beam radially and the reading of F_z is lightly biased.

When the optical trap is centered and no force is applied, the laser beam

hits the center of the detectors and their readings are zero (see top panel in Fig. 2.11a). By applying a radial external force (F_y), the outcoming light from the bead is still parallel with zero angle, but it is off-axis (see middle panel in Fig. 2.11a). In this case, the force PSD measures a deflection while the light-lever and the Iris measure the same as in the top panel. Note that the fact that the objective 1 is underfilled allows to collect all the deflected light in objective 2. Finally, when applying an axial external force (F_z), the outcoming light has a larger diameter, which is detected by the Iris (see bottom panel in Fig. 2.11a). In this situation, the beam hits the center of the force PSD and its reading is zero, as expected. If the force was applied in the opposite direction, the Iris would detect a smaller beam diameter. Generally speaking, the z force is detected from the change in the laser beam diameter. In figure 2.11b it is shown the same ray tracing than before, but with the optical trap located at a different position. When the optical trap is no longer centered (i.e., the wiggler redirects the light to another direction) the laser beam after the collimating lens is still parallel, but it has an angle with respect to the optical axis. The force readings in the upper panel of Fig. 2.11b are the same than in panel a but now the light-lever is off-axis, indicating the different position of the optical trap. The same criteria applies to the middle and bottom panels.

It is important to say that the distances between the optical elements (lenses and detectors) must be accurately adjusted. In fact, small misalignments induce artifacts in the measurement of position and force. For instance, if the force PSD was located too close (or too far) from the relay lens, the reading of force would depend on the position of the optical trap.

Summing up, the optical system is aligned so that the position of the optical trap is related to the angle between the collimated beam and the optical axis. In contrast, the offset of the outcoming beam is related to the radial force and the diameter of such beam is related to the axial force. Since each detector is only sensitive to one of these properties of the beam (angle, offset and diameter), the three magnitudes can be measured separately.

2.2.2 Electronic controller

The electronic boards are the elements that communicate the head of the instrument with the host computer. They perform two main tasks. The first one is the acquisition and conditioning of the signal coming from the photodetectors that measure the force and the position of the optical trap. The second one is the processing of orders given by the host computer to the wigglers and the motorized stage. Additionally, the feedback algorithms that perform the constant force or position protocols are also run in the

microprocessors of the electronic boards. The stack of boards is powered by power supplies that provide +5 V tension for the digital electronics, ±15 V for the analog electronics, ±12 V for the motors and +150 V for the piezoelectric crystals.

The boards combine analog and digital signal processing and each of them is devoted to a specific function. The electronic boards are set on a stack and they follow a hierarchical structure based on the head. There are 5 intercommunicated boards but 2 of them are identical (see Fig. 2.12). Following there is a description of each of them.

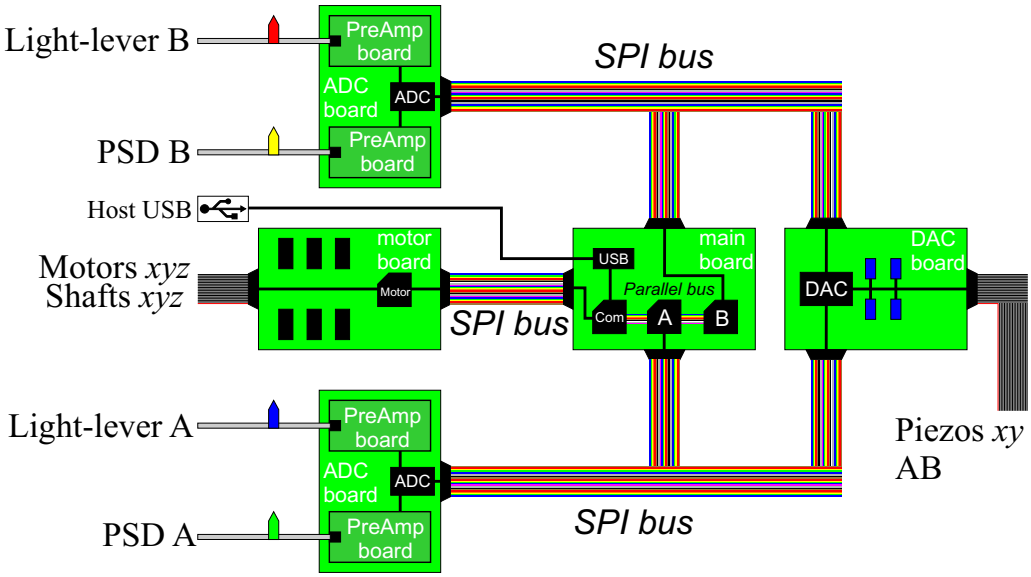


Figure 2.12: Electronics schematics (see text).

Main board. The main board is devoted to communication and data processing. It contains one USB transceiver and 3 microprocessors (Microchip PIC18F6520). The first microprocessor called ComPic controls the communication between the USB and the other pics. The other two microprocessors called TrapPics (A and B) control the two optical traps of the head. They read the data from the ADC board, send orders to the DCA board and perform the feedback algorithms. The three pics are intercommunicated with parallel buses and the ComPic is also connected with the MotorPic, located at the motor board.

Analog to Digital Converter (ADC) board. There are two identical and independent ADC boards, one for each optical trap (A and B). The ADC board receives the analog data from the PSDs (position and

force) and converts them into a digital signal. The ADC board has a sub-board called Pre-amplification (or Preamp) board that receives the 4 analog current signals (the 4 electrodes) from one PSD and converts them into 3 voltage signals: the x and y position and the total intensity (or PSDSum) of the beam. These 3 signals are converted into digital units by an AD converter. The ADC board is connected to its corresponding TrapPic in the Main board by a Serial Peripheral Interface (SPI) bus.

Digital to Analog Converter (DAC) board. This board receives the orders from the TrapPics in the main board and sends a voltage to the piezoelectric crystals of the wigglers to position the optical trap. It has a DA converter that communicates with the TrapPics via SPI. The DA converter outputs a voltage that is stabilized by 4 Operational Amplifiers (op-amps) that keep the 4 piezos (x and y for traps A and B) at the desired voltage.

Motor board. The motor board controls the motorized stage. It has a microprocessor (MotorPic) that communicates with the ComPic in the main board via SPI. The MotorPic has two main functions. The first one consists in moving the 3 motors of the xyz stage using 3 independent H-bridges. These are chips that allow to invert the direction of movement of the motor. The velocity of the motor is controlled by the Pulse Width Modulation (PWM) module of the MotorPic, which sends pulses of voltage of different frequencies to the motor. This allows to accurately control the amount of power given to the motor. The second function of the MotorPic is to measure the position of the motors. Each motor (Thorlabs Z606) has a shaft encoder that turns with the screw and sends pulses to a 24-bit counter. The counter reads the pulses and indicates the relative rotation of the motor, which is given in a 24 bit binary number. So the MotorPic reads this number to measure the position of the motor. It allows the MotorPic to perform feedback protocols such as Goto functions, in which the motor is running until a certain position is reached. Finally, the MotorPic also controls the blue LED that illuminates the optical system.

The digital processing of data is mainly performed by the 4 microcontrollers (ComPic, 2 TrapPics and MotorPic). A microcontroller (PIC) is a small processor capable of performing multiple tasks, including mathematical operations, time management, analog to digital conversions, parallel and serial communication and data storage. The list of operations that a PIC has

Parameter	Value
Analog data bandwidth	∼1 MHz
PIC internal clock	10 MHz
Sampling rate	4 kHz
Feedback running frequency	4 kHz
Digital data bandwidth	1 kHz
16-bit data channels	psdX, psdY, psdSUM, iris leverX, leverY, leverSUM
32-bit data channels	motorX, motorY, motorZ
ADC force resolution	> 0.01 pN
Maximum detectable force	< 400 pN
ADC distance resolution	> 0.1 nm
Maximum detectable distance	< 12 μm
Piezo update frequency	4 kHz
Wiggler relaxation time	<1 ms

Table 2.1: Specifications of the electronic controller.

to perform is known as firmware and it is written in standard C or assembly coding languages. The firmware is compiled and burn into the internal memory of the PIC. Appendix C.5 contains an expanded description of the firmware and connections of the PICs in each electronic board.

Table 2.1 summarizes the most relevant characteristics of the electronic controller.

2.2.3 Host

The host is the name given to the computer and to the software that allows the user to interact with the experiment. The main tasks of the host are to receive and save the data from the electronics controller and give orders to it. The host software is an application called "lt" (for "Laser Tweezers") and it was designed by Shane Saxon to provide a simple and efficient interface to the user [19]. It runs on a Mac computer and it is compiled with Xcode for Mac OSX 10.4 (Tiger). New versions of lt also run on updated versions of Mac OSX. The software was developed using Carbon, i.e., an Apple's Application Programming Interface (API). The code was written in C programming language and uses OpenGL libraries for the graphics display. The application is designed in layers of increasing level of programming so that the basic tasks are available to expand the features of the software. This way, new experimental protocols can be efficiently implemented. As mentioned,

the communication between the host computer and the controller is done by USB. Therefore, the instrument does not need to be linked to a specific computer. Instead, the host computer can be easily substituted by connecting the USB port to another computer, provided that the FTDI driver [87] is installed on it.

The lt software is an event-driven application with other internal threads that execute periodic tasks (communication, rendering). In other words, the application is continuously receiving and rendering data from the electronic controller until an event is triggered (e.g. a mouse click) and executed. All the variables of the application are grouped into a large data structure that is available to every subroutine of the code by passing the pointer to that structure. The mouse is mainly used to control the instrument (traps, motors, protocols) and the keyboard for the graphical display. A detailed description of the code run by the computer is explained in Appendix C.6. Figure 2.13 shows a screenshot of the user interface.

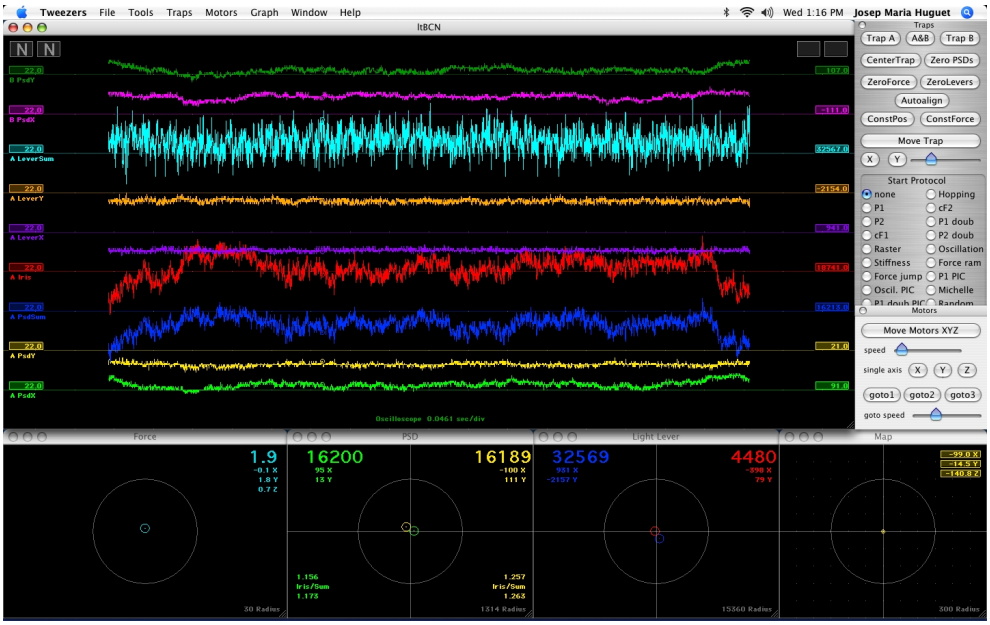


Figure 2.13: The user interface of the lt application. The experimentalist interacts with the instrument by means of this interface.

2.3 Calibration

In all instruments the raw data measured by the sensors (i.e., transducers or detectors) has to be converted into physical magnitudes. The process of

inference of such conversion is known as calibration. The calibration can be performed in either two ways: 1) by comparing a known magnitude of one physical process with the measurement of the instrument or 2) by establishing a conversion relation with another calibrated instrument.

The calibration of the Minitweezers involves three independent magnitudes: time, distance and force. The rest of magnitudes (e.g. pulling rate, motor velocity, etc.) measured by the instrument are obtained from combinations of those three ones. One interesting aspect of the optical tweezers is that the calibration of the force can be performed in different ways, allowing to cross-check the calibration procedures. During the following explanation, we use square brackets [...] to represent measured and converted magnitudes, and bare symbols for calibration factors and offsets.

2.3.1 Time

The measurement of time relies on an internal clock of the MainPic. The PIC sends a packet of data periodically to the USB transceiver at a frequency of 4 kHz (see Appendix C). One of the variables included in the packet is the CycleCount. It is a 32-bit number that increases one unit every time the MainPic completes a loop. Therefore, one unit of CycleCount represents a lapse of time of 1/4 kHz = 250 μs. The host computer performs the following operation to calculate the time,

$$[\text{Time}] = [\text{CycleCount}] \,/\, \text{Frequency} \tag{2.28}$$

where [Time] is the time in seconds and Frequency is the calibration factor (Frequency = 4000). Nevertheless, it is useful to perform a check with an oscilloscope. The frequency can be directly measured by connecting the probe of the oscilloscope to any wire of the parallel bus that communicates the three PICs in the main board (see Fig. 2.12). If the instrument works properly, the signal of the wire must be periodic with a frequency of 4 kHz. The error in the calibration of time directly depends on the error of the internal clock of the MainPic. The relative error of the oscillation frequency (i.e., $\Delta f/f_0$) is lower than 50 ppm, which is more than enough for the applications that we are interested in.

2.3.2 Distance

The distance has to be calibrated on two different devices: the motors and the light-levers. The calibration of the light-levers relies on the calibration of motors.

Motor

According to the specifications of the motor (Thorlabs Z606), the shaft encoder produces 48 pulses per revolution of the motor. Depending on the direction of the motor, each pulse increases or decreases by one unit the number stored in the 24-bit counter (see Appendix C). Moreover, there is a reduction of 256:1, which means that 256 revolutions of the motor produce 1 revolution of the screw that moves the xyz-stage. Finally, 1 revolution of the screw retracts or extends the screw by 0.05 μm. So, the conversion factor between distance and pulses can be calculated as follows,

$$\frac{48\,\text{pulses}}{1\,\text{motor rev.}} \times \frac{256\,\text{motor rev.}}{1\,\text{screw rev.}} \times \frac{1\,\text{screw rev.}}{0.05\,\mu\text{m}} = 245760\,\frac{\text{pulses}}{\mu m}$$

Now, the host computer performs the following operation to calculate the position of the motor,

$$[\text{MotorPosX}] = [\text{CounterX}] \,/\, 245760 \tag{2.29}$$

where [MotorPosX] is the position of the motor X in micrometers, [CounterX] is the number of pulses stored in the 24-bit counter. The same conversion is applied to the Y and Z motors. It is important to note that the calibration factor (245760) is only fixed by the specifications of the motor. However, the calibration factor can be verified by using an alternative method. The idea consists in making some marks to a fluidics chamber (by scratching it, or introducing fragments of micropipette) and measure the distance between them using a microscope and a micro-ruler. Then, the chamber is mounted on the optical tweezers and the distances measured by the motors are compared with the distances measured with the microscope. The discrepancy between both methods is lower than 0.5%.

Finally, it is important to mention that the shaft encoders show large errors in the measurement of short distances (< 0.5 μm) but they are very linear at large distances. This is due to the backlash of the motor, which is an irregular response of the motor to a voltage pulse. The origin of this phenomenon is the gap between the gear of the motor and the gear of the screw. In other words, the shaft encoder detects movement when the screw is not rotating yet. Therefore, the larger the distance measured by the motor, the lower the relative error.

Light-lever

The calibration of the light-lever position sensor is done using the motors that move the xyz-stage. A trapped microsphere is held fixed at the tip of

the micropipette and the optical trap follows the position of the microsphere by keeping the total force equal to zero using a force feedback mechanism (Fig. 2.14a). As the pipette is gently moved, the trap follows the center of the bead to maintain the preset zero force. Although the force is not calibrated yet, the light-lever calibration protocol is still valid because we only expect that the traps follow the movement of the microsphere, no matter what actual force is being applied. In order to make the reading of the light-lever independent of the total power of the optical trap, the deflection of the PSD [TrapALeverY] is normalized by the total power received by the PSD [TrapALeverSum]. It allows to establish a linear relation between the position of the motor (already calibrated) and the reading of the light-lever according to,

$$[\text{MotorPosY}] = \frac{[\text{TrapALeverY}]}{[\text{TrapALeverSum}]} \times D_y \tag{2.30}$$

where D_y is the calibration factor for trap A in the y direction. Figure 2.14b,c shows the linear fit of the measured data, whose slope is the calibration factor. There are also deviations from the linear regime at the ends of the range due to non-linearities of the PSDs at the borders. They are not relevant, though, since most of the experiments are performed at the central regions of the PSDs. The same procedure is repeated for the x axis and for trap B. This calibration protocol provides slightly different calibration factors (less than 3%) depending on the direction in which the motor is moving, due to the backslash of the motor. The average of them is used as the calibration factor.

In the end, the host software converts the raw data from ad units into nanometers according to the following relation,

$$[\text{A-DistY}] = \frac{[\text{TrapALeverY}]}{[\text{TrapALeverSum}]} \times D_y \tag{2.31}$$

2.3.3 Force

As mentioned before, the calibration of force can be performed by different methods that allow to check the process. According to the experimental setup (Fig. 2.10), the instrument detects the change in the light momentum of the laser beams that form the optical trap, which allows us to directly measure the force accurately. There is a linear relation (see Eqs. 2.25 and 2.27) between the PSD reading and the actual force exerted on the bead by the optical trap, which can be written as

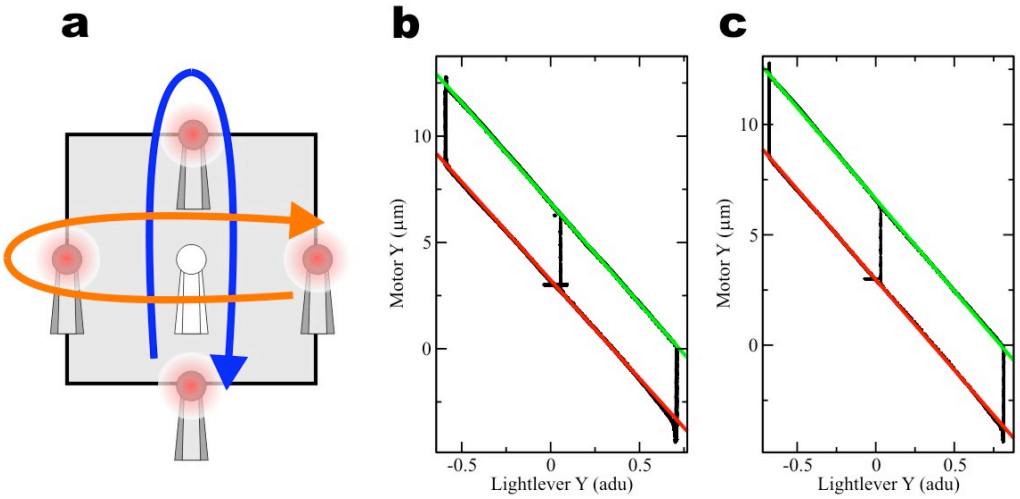


Figure 2.14: Calibration of light-lever (see text). **(a)** Calibration protocol. The micropipette is moved up and down (blue arrow) to calibrate the y distance and left and right (orange arrow) to calibrate the x distance. The gray square (side length of 11 μm) shows the range of the piezos. **(b)** Calibration of y distance for trap A. The horizontal axis shows the value of the normalized deflection of the light-lever in ad units given by [TrapALeverY]/[TrapALeverSum]. The black curve is obtained moving the trap up and down. Green line shows the linear fit when the bead is moved downwards and red line, when moved upwards. The vertical black trace is the backlash of the motor. **(c)** Calibration of y distance for trap B.

$$[\text{ForceY}] = [\text{PSDy}] \times M_y + O_y \tag{2.32}$$

where [ForceY] is the actual force on the y axis in pN, [PSDy] = [TrapAPsdY] + [TrapBPsdY] is the sum of the readings of the PSDs of both traps in the y direction, M_y is the calibration factor and O_y is a force offset, already corrected by the data acquisition board. The offset allows to make the reading of the [PSDy]=0 if the spot of the laser beam does not perfectly hit the center of the PSD detector when no force is applied. The value of M_y is independent of the trap power and the calibration protocols can be repeated at different laser powers as an extra test. It is important to mention that, unlike the measurement of distance, the force measurements are not normalized by the total power received. It is what makes the difference between a power-dependent magnitude (force) and a power-independent one (distance). Here we only show the method used in force calibration for the y axis. The same procedure applies to x and z axis. Three different methods were used to calibrate the PSD that measures the force.

Light-momentum

The calibration of the instrument can be performed from first principles, by measuring the optical parameters of the experimental setup. Indeed, Eqs. 2.25 and 2.27 require the knowledge of f (the objective focal length), R_D (the effective radius of the PSD chip) and Ψ (the power sensitivity of the PSD). The specification of the objective allows to obtain f from the quotient between the tube length of the microscope (usually equal to 160 mm) and the magnification of the objective (equal to $\times 60$). It gives a value of $f = 2.66$ mm which can also be checked experimentally [16]. The value of R_D is also given by the specifications of the detector but it has to be corrected due to magnification introduced by the relay lens ($R_D \simeq 2.25$ mm). This value can also be experimentally checked [16]. Ψ is a magnitude that relates the PSD detector output current [PsdSum] to the light intensity at the trap focus P according to $\Psi = [\text{PsdSUM}]/P$. This value has to include the attenuation effects of all the optical elements (objectives, beam splitters, prisms). Then the power at the focus P is estimated as the geometric mean of the light intensity of the laser beam at the entrance (P_{in}) and exit (P_{out}) of the objectives according to $P = \sqrt{P_{\text{in}} \cdot P_{\text{out}}}$. This estimation is correct if an exponential attenuation is assumed. The intensity is measured with an optical power meter (Thorlabs PM30-130).

Power spectrum of force

A microsphere suspended in water and located in an optical trap can be studied from the point of view of a Brownian particle submitted to a harmonic potential. The dynamics of this system is governed by a Langevin equation that allows to characterize the power spectrum of the thermal forces that act on the particle. In the case of an overdamped particle in a harmonic potential, the power density of force is expected to follow a Lorentzian distribution according to

$$S_{F_y}(\nu) = \langle F_y \cdot F_y^* \rangle = \frac{2kk_BT\omega_c}{\omega_c^2 + (2\pi\nu)^2} \tag{2.33}$$

where S_{F_y} is the force power density of the y force, ν is the frequency, F_y is the Fourier transform of the y force, k is the trap stiffness in the y direction, k_B is the Boltzmann constant, T is the temperature and ω_c is the corner frequency which is given by $\omega_c = k/\gamma$, where γ is the drag coefficient of the bead in distilled water. The drag coefficient for spherical particles at low Reynolds number regime can be calculated according to $\gamma = 6\pi\eta R$, where η is the viscosity of the surrounding fluid and R is the bead radius. Equation 2.33 can be used to obtain the calibration factor by comparing the theoretical

prediction and the experimental power spectrum.

The experimental power spectrum is obtained from a measurement of 30 seconds of Brownian PSDy signal of a trapped microsphere in distilled water at low laser power (see Fig. 2.15a) in order to have a low corner frequency. The data pairs (time vs. force) are split into 1.0-second windows and the Fast Fourier Transform (FFT) is applied to them. The square modulus of the FFT is computed and averaged over all windows, therefore obtaining an estimation of the experimental power spectrum S_{PSD_y} (see Fig. 2.15b,c). The spectrum is fit to a Lorentzian according to the following expression,

$$S_{\mathrm{PSD}_y}(\nu) = \langle \mathrm{PSD}_y \cdot \mathrm{PSD}_y^* \rangle = \frac{A}{B + (2\pi\nu)^2} \tag{2.34}$$

where A and B are fitting parameters. By using the relation given by Eq. 2.32, Eqs. 2.33 and 2.34 can be related according to

$$\begin{aligned} S_{F_y}(\nu) &= M_y^2 \cdot S_{\mathrm{PSD}_y}(\nu) \\ \frac{2kk_BT\omega_c}{\omega_c^2 + (2\pi\nu)^2} &= \frac{M_y^2 \cdot A}{B + (2\pi\nu)^2} \end{aligned} \tag{2.35}$$

which gives the calibration factor (M_y) and the trap stiffness (k) from the fitting parameters

$$\left.\begin{array}{l} 2kk_BT\omega_c = M_y^2 A \\ \omega_c^2 = B \end{array}\right\} \qquad \left.\begin{array}{l} M_y = \sqrt{2k_BT\gamma B/A} \\ k = \gamma\sqrt{B} \end{array}\right\} \tag{2.36}$$

where $k_BT = 4.11$ pN·nm at 25 °C, $\gamma = 2.67 \cdot 10^{-5}$ pN·s·nm^{-1} for a microsphere of diameter $2R = 3.00 \pm 0.05$ μm (polystyrene beads for calibration from Spherotech, Libertyville, IL) surrounded by distilled water of viscosity $\eta = 8.9 \cdot 10^{-4}$ Pa·s ($= 8.9 \cdot 10^{-10}$ pN·s·nm^{-2}).

The two force calibration factors extracted from the power spectra measured at two different laser powers differed less than 1%. The stiffness of the weak trap was $k = 3.22 \pm 0.05$ pN·μm^{-1}, while $k = 6.20 \pm 0.07$ pN·μm^{-1} for the stronger trap. The trap stiffness can also be deduced from the force fluctuations (i.e., the PSD_y reading) of the previous measured time series, without knowing the viscosity of water. The fluctuation-dissipation relation for a trapped particle can be written as:

$$\langle F_y^2 \rangle = k \cdot k_BT \tag{2.37}$$

where $\langle F_y^2 \rangle$ is the variance of the y force. Combining Eqs. 2.32 and 2.37 it is possible to obtain an expression for the trap stiffness given by

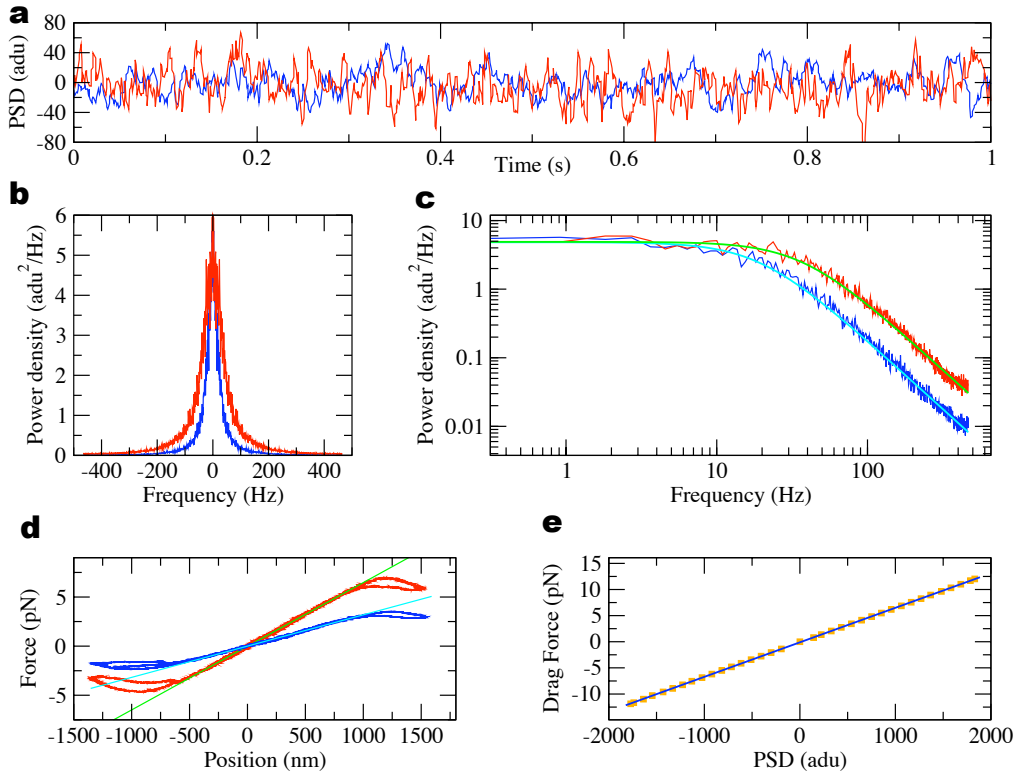


Figure 2.15: Calibration of force. **(a)** 1 second window of force reading (raw data from PSDs) versus time of two trapped particles. The red (blue) trace shows the force reading for a particle trapped in a stiff (soft) trap. **(b)** Noise power density for the previous recordings. **(c)** Noise power density in log-log scale and Lorentzian fits. Green curve shows the Lorentzian fit for the stiffer trap (red spectrum) and cyan curve shows the Lorentzian fit for the softer trap (blue spectrum). **(d)** Force vs. elongation of a particle stuck at the tip of the micropipette. The slope of the linear region of the red curve is the trap stiffness (slope of green curve). Analogous curves for the softer trap (blue and cyan curves). (e) Stokes law calibration. Orange dots show the experimental measurements and the blue curve depicts the linear fit. The drag force (y axis) is obtained from the velocity of the motor (v) and using the Stokes law ($F = 6\pi\eta R v$). The slope of the linear fit is the calibration factor.

$$k = \frac{M_y^2 \cdot \langle \mathrm{PSD}_y^2 \rangle}{k_B T} \tag{2.38}$$

where $\langle \mathrm{PSD}_y^2 \rangle$ is the variation of PSD in the y direction and M_y is the calibration factor obtained from Eq. 2.36. The values obtained were $k = 3.26$ pN·μnm^{-1} for the soft trap and $k = 6.15$ pN·μm^{-1} for the stiff trap, which represents an error of 1.5%.

A last check can be performed to measure the stiffness using another

independent method (see Fig. 2.15d) which does not use the force noise. A bead is stuck at the tip of the micropipette. The optical trap is aligned in the center of the bead and short displacements of the optical trap are performed while recording the force vs. trap displacement and keeping the micropipette at fixed position. The slope of the linear region of this curve is the trap stiffness. The values of the stiffness obtained were $k = 3.200 \pm 0.002$ pN$\cdot\mu$m^{-1} for the soft trap and $k = 6.28 \pm 0.02$ pN$\cdot\mu$m^{-1} for the stiff trap. These values differ by less than 1.5% with respect to the power spectrum method.

Stokes law

This is the most straightforward method to calibrate the force. The host software has a specific protocol to perform this calibration and it is convenient to periodically check the calibration factors. The analysis of data is also fast and efficient.

The Stokes law establishes a linear relationship between the drag force exerted on a particle F_y and the speed at which this particle is moving in a viscous fluid v. It is given by the following expression,

$$F_y = 6\pi\eta R v \tag{2.39}$$

where η and R have already been defined previously. In order to use the Stokes law to calibrate the force, a microsphere is trapped and the whole fluidics chamber is moved at a fixed speed using the motorized stage along the y direction. This process is repeated at different motor speeds and a collection of speed vs. force measurements are collected. The drag force can be deduced from the speed of the motor (which is given by the position of the shaft encoder along time) and Eq. 2.39. Figure 2.15e shows the linear relation between the drag force and the PSD$_y$ reading, whose slope is the calibration factor. It is obtained after averaging 40 different beads and it agreed within 2% with the previous method.

A final test to the calibration force is the reproducibility of some well-established experiments. The overstretching transition of half λ-DNA is commonly accepted to occur at 67 pN at 500 mM [NaCl] [48] (see Fig. 2.16a). This referential value of force is measured when performing the experiments with the calibrated instrument. Similarly, the unzipping of the cosL end of λ-DNA shows a characteristic FEC. Such FEC can be reproduced by different optical tweezers instruments provided that all them are well calibrated (see Fig. 2.16b).

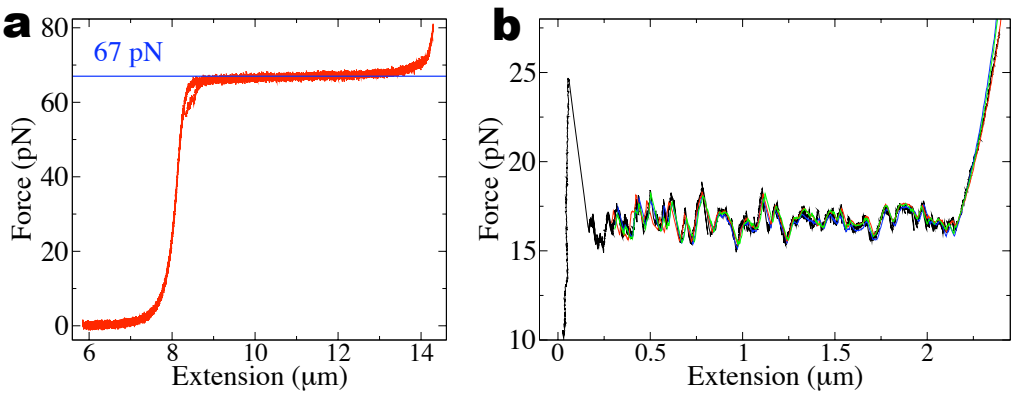


Figure 2.16: Calibration tests. **(a)** FEC of a half λ-DNA (red curve). The overstretching plateau occurs at 67 pN. **(b)** Measurements of unzipping in two optical tweezers instruments. Black curve shows the unzipping data measured with one instrument. Red, blue and green curves show the data of 3 different molecules obtained with another instrument.

2.4 Design of new protocols

A protocol is a type of experiment defined by the instructions that the instrument has to perform. The minitweezers can control the position and the force of the optical trap with a feedback. A protocol is defined by the steps that the instrument has to automatically perform on a molecule. For instance, one of the most common protocols is a pulling experiment. It consists on moving the optical trap up and down at a constant rate so that the molecule undergoes cycles of stretching and relaxing. Another protocol is the constant force protocol, in which the molecule is held at a constant force. This is achieved by means of a force feedback that corrects the position of the optical trap to keep constant the force exerted on the molecule.

The firmware and software of the instrument was designed so that new protocols can easily be implemented. Indeed, most of the essential routines and functions that perform the elementary steps are available to the user. Thus, a new protocol does not need to be designed from scratch. Instead, users can gather pieces of already coded subroutines to create their own protocol.

Here we have to distinguish between two types of new protocols. The first one only involves coding the software and it is easier to implement. The second one is more demanding and, apart from the software, it requires to code the firmware of the PICs. The advantage of coding the PICs is the update frequency. Indeed, the software of the Mac can update the instrument at 60 Hz, while the firmware of the PICs, at 4 kHz. Depending on the requirements of the experiment, one type of protocol will be preferred over

the other. Appendix D shows a detailed description of the steps to be followed to code a new protocol.

Figure 2.17 shows data obtained with five different protocols designed during the realization of this PhD thesis:

Double trap pulling protocol. This protocol assumes that each counter-propagating laser is used as an independent optical trap. Therefore, two beads located at two optical traps are used to pull on a molecule. The traps can move simultaneously or one respect to the other (see Fig. 2.17a).

Oscillation protocol. This protocol produces an oscillation of the force or the position of the optical trap. The frequency and the amplitude of the oscillation can be adjusted (see Fig. 2.17b). This is useful to study the stochastic resonance of single molecules [88].

Force jump protocol. This protocol is a combination of two already existing ones. It consists in a pulling experiment followed by a constant force protocol at a different arbitrary force (see Fig. 2.17c). The misfolding dynamics of DNA hairpins can be studied with this protocol.

Force ramp protocol. It is a pulling experiment in which the force applied to the molecule is increased at a constant loading rate. Instead of constantly changing the position of the optical trap, this protocol continuously increases the force by running a force feedback. The comparison between a force ramp protocol and a regular pulling protocol gives relevant information about the thermodynamics of small systems. This protocol was designed to carry out the work described in chapter 6.

Pulling protocol coded in the firmware. It is a standard pulling protocol in which the molecule is stretched and relaxed by moving the trap up and down. The advantage is that the update frequency of the PICs is 4 kHz and the position vs. time does not exhibit steps at high pulling speeds (see Fig. 2.17d).

2.5 Conclusions

Since the discovery of the optical tweezers, the accuracy of the instrumentation and the number of applications have multiplied. The research in molecular biophysics has gone hand in hand with the development of optical tweezers

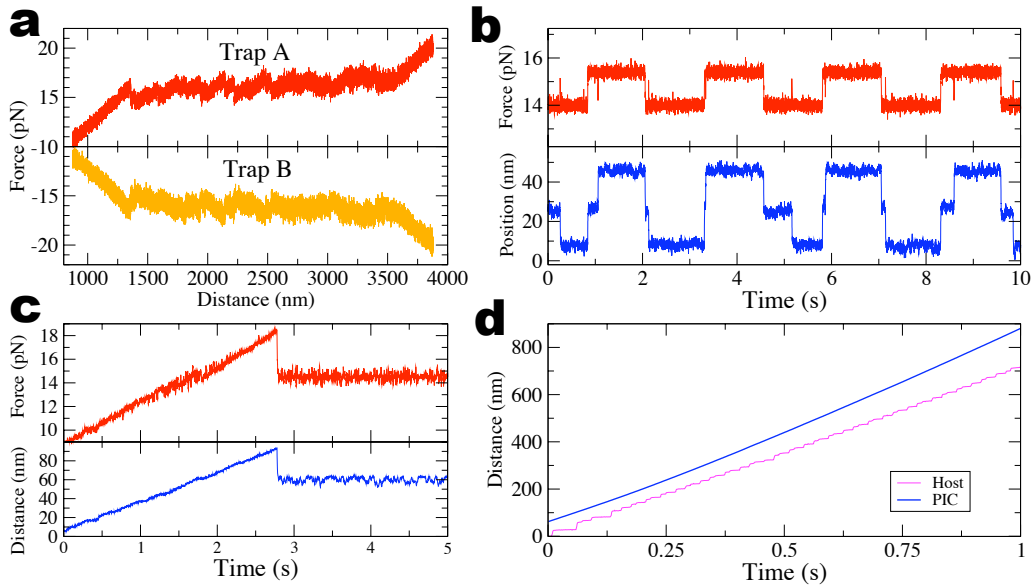


Figure 2.17: New designed protocols. **(a)** Unzipping of DNA with a double trap pulling protocol. The FDCs of the traps (red and orange) are measured simultaneously. Except for thermal fluctuations, the forces are equal and opposite. **(b)** Oscillation protocol applied to a DNA hairpin. The force oscillates between 14 and 15.4 pN (red curve) according to a square wave of frequency 0.4 Hz. The position of the optical trap (blue curve) is continuously adjusted by the force feedback algorithm. **(c)** Force jump protocol applied to a DNA hairpin. A pulling protocol of pulling rate 30 nm/s is followed by a constant force protocol of force 14.5 pN. **(d)** Position vs. time curves of two pulling protocols at a pulling rate of 950 nm/s. The blue (magenta) curve shows the output of the protocol performed by the PIC (host computer).

during the last 20 years. What is more, optical tweezers have become a standard tool for the biophysicists. The future will bring us the combination of optical tweezers with other techniques, the parallelization of experiments and the refinement of manipulation.

The optical trapping of particles can be described from different complementary theories. None of them is quantitatively accurate, but they provide satisfactory approaches to the problem. The minitweezers is an instrument designed to produce optical traps and measure the force using the conservation of light momentum. This is a stable instrument that has high resolution and accuracy thanks to its size and compactness.

The acquisition of data and the control of the instrument is performed with customized electronic boards based on PIC microcontrollers. The advantage of using PICs is that the feedback algorithms such as constant force can be implemented at high bandwidth compared to other optical tweezers instruments.

Besides, the user interface was designed for a Mac computer, which is intuitive and easy to use. The instrument was conceived to be used by several experimentalists working on different topics. So the calibration only needs to be done once to make the instrument available to all users. The host application also allows to design and implement new protocols easily.

The research described in the next chapters of this PhD thesis was based on the experiments performed with the minitweezers instrument. Understanding the working of the instrument has been a key stone to set up the experiments and analyze the data.

The information contained in this chapter and its appendixes is part of the know-how of the Small Biosystems Lab. It has been included in this thesis in order to make it available to everyone as a useful reference for future studies.

Why are we here? What's life all about? Is God really real, or is there some doubt? [...] Or are we just simply spiraling coils of self-replicating DNA?

Monty Python's
The Meaning of Life (1983)

Chapter 3

DNA unzipping

The discovery of DNA and its relation with the transmission of the genetic information was a process that took almost a century. The research involved many scientists with different backgrounds. All this prolonged investigation led to one of the most beautiful scientific achievements of the 20th century: the central dogma of molecular biology. Formulated by Francis Crick in 1958 [20], it eventually gave a physico-chemical basis to Mendelian genetics.

In 1865, Gregor Mendel set the foundations of genetics. He postulated the existence of genes as the basic units of heredity. So he established the rules of the heredity without knowing its physico-chemical basis. In 1869, Friedrich Miescher reported the existence of a new substance (nowadays known as nucleic acids) in the nucleus of the cells. Latter in 1882 Walther Flemming discovered the mitosis (i.e., the eukaryotic cell division) and the chromosomes. However, it was not until 1902 when two scientists (Theodor Boveri and Walter Sutton) independently identified the chromosomes as the carriers of the genetic information. This idea was experimentally verified by Thomas Morgan in 1910. Latter, the investigations focused on the composition of the chromosomes. In 1919, Phoebus Levene discovered that the nucleic acids were composed of nucleotides. However, he thought that the nucleotides were too simple to carry all the genetic information. At that time, most biologists believed that the proteins of the chromosomes were the carriers of the heredity. In 1944, the Avery-MacLeod-McCarty experiment –which was a revised version of Griffith experiment (1928)– concluded that DNA was the actual carrier of genes. The experiment consisted in mixing living avirulent

bacteria with a large inoculum of lethal heat-killed cells and injecting it into mice. The experimentalists understood that the DNA was the substance that induced the transformation of the avirulent bacteria into the lethal ones, which produced the death of the mice. In the early 1950's, Erwin Chargaff discovered that the amount of guanine in DNA is equal to cytosine and the amount of adenine is equal to thymine. He established the two so-called Chargaff's rules that helped to predict the base-pairing and the structure of DNA. In 1953, Watson and Crick [1] proposed the double helix structure of DNA, using a X-ray image taken by Rosalind Franklin. All this period of investigations concluded in 1968, when Khorana, Holley and Nirenberg were awarded with the Nobel prize for their discovery of the genetic code.

Since then, molecular biology has experienced a tremendous development. The processes of the cell have been explored in great detail and our knowledge about them is nowadays enormous. Nevertheless, the central dogma of molecular biology is still valid. It summarizes the very essential molecular mechanisms related to the flow of the genetic information.

Although some RNA (such as ribozymes) perform important cellular functions, generally speaking the proteins are the biomolecules that determine the structure and the function of each cell: different cells have different proteins. All proteins are made from a sequence of 20 different elementary components called aminoacids. So each cell is capable of producing its proteins by gathering and bonding the correct sequence of aminoacids. The sequence of aminoacids to produce one protein is coded in the DNA. Now, the DNA is a biomolecule composed of a sequence of 4 types of nucleotides. Each group of 3 nucleotides codes for one amino acid and this is the ultimate physical support of the genetic information. Figure 3.1 shows how the genetic information is transferred from the DNA to RNA in order to build the proteins.

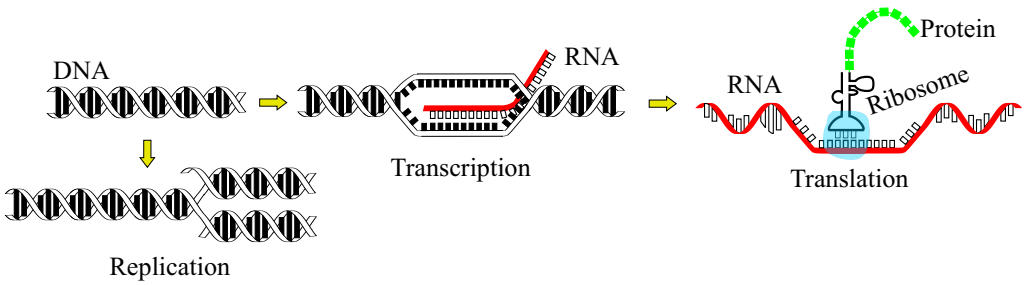


Figure 3.1: Central dogma of molecular biology [89]. The DNA stores the genetic information. Replication is the process by which this information is duplicated during the cell division. Transcription consists in the synthesis of a sequence of RNA from DNA. Finally, translation is the process by which proteins are synthesized according to the sequence of the RNA transcript. Translation is then carried out by ribosomes.

DNA forms a double helix that must be split for the cellular machinery to access the genetic information coded in the bases. Cells have specific molecular complexes to carry out such tasks. For instance DNA polymerases and their associated proteins are in charge of DNA replication. Similarly, RNA polymerases perform the transcription of DNA into RNA. During the last decades of the 20th century, scientists were able to reproduce these processes in vitro. Besides, the development of single-molecule techniques allowed them to directly manipulate and study these processes.

This chapter focuses on the process of DNA unzipping by force. DNA unzipping essentially consists in pulling apart the two strands of DNA by exerting mechanical forces on the extremities of the molecule. After describing the structure of DNA, this chapter focuses on the unzipping experiments and the models introduced to understand them.

3.1 Structure of DNA

The structure of biomolecules (including proteins and all nucleic acids) has been traditionally divided into four levels [90]. The primary structure indicates the chemical composition of the molecule, i.e., the atoms and the covalent bonds between them (see Fig. 3.2). DNA is a polymeric chain whose monomers are nucleotides [91]. A nucleotide (see Fig. 3.2a) is composed of one molecule of phosphoric acid, one molecule of 2′-deoxyribose (which is a cyclic pentose sugar) and a nitrogenous base (that can be adenine, cytosine, guanine or thymine). The phosphoric acid is bonded to the 5th carbon of the 2′-deoxyribose and the nitrogenous base, to the 1st one (see Fig. 3.2b). Two nucleotides are bonded each other by a phosphodiester bond between the phosphoric group of the first nucleotide and the 3rd carbon of the second nucleotide. The concatenation of nucleotides forms a phosphate-deoxyribose backbone from where the sequence of bases are linked (see Fig. 3.2c). The resulting structure has asymmetric ends (5′ and 3′) and the polynucleotide has a direction. The 5′ end has a terminal phosphate group and the 3′ end a terminal hydroxyl group. In general, the primary structure of DNA is usually given as a sequence of bases in the direction $5' \to 3'$.

The secondary structure of a biomolecule (see Fig. 3.3) is the result of non-covalent interactions (e.g., hydrogen bonds, hydrophobic interactions) between the atoms of the primary structure [92]. In the case of DNA, the secondary structure leads to the hybridization of two complementary and anti-parallel strands of DNA (i.e., a $5' \to 3'$ strand paired with a $3' \to 5'$ one). In the canonical base pairing, the bases of the two strands pair each other according to Watson-Crick rules: adenine is paired with thymine by two

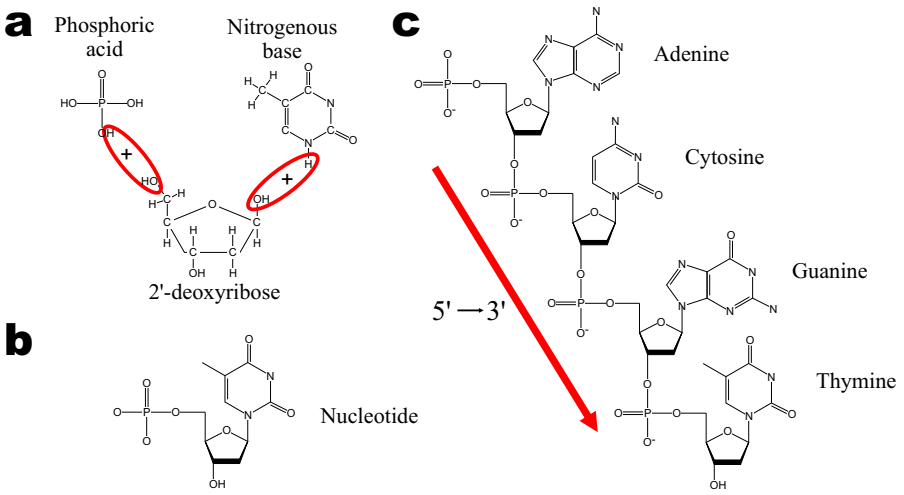


Figure 3.2: Primary structure of DNA. **(a)** Chemical components of a nucleotide. **(b)** Nucleotide. **(c)** Polynucleotide.

hydrogen bonds and cytosine is paired with guanine by three hydrogen bonds (see Fig. 3.3a). The hydrogen bonds only give specificity to the base pairing. The stacking interaction between consecutive base-pairs is what stabilizes the hybridized structure. Stacking is an intermolecular interaction observed in aromatic molecules that tend to arrange them in a pile (see Fig. 3.3b). There are two forces that stabilize base stacking: the hydrophobicity of the aromatic rings of the bases and the London dispersion of the dipoles (induced in the bases). Stacking forces are different for each combination of base-pair. In general, a stack of purines (adenine and guanine) is stronger than a stack of pyrimidines (cytosine and thymine). Apart from Watson-Crick base-pairs, there are also other motifs (such as loops or bulges) that can contribute to the thermodynamic stability of the secondary structure.

The tertiary structure of a biomolecule shows the spatial localization of atoms, i.e., the three-dimensional structure of the molecule. In the case of the DNA, the two strands form a double-helix [92](see Fig. 3.4). The backbones of the two anti-parallel strands face each other and twist themselves along the central axis of the molecule. The bases of one strand are paired with the complementary ones and they are localized in the cavity left between the two backbones. The outer envelope of the double helix is not cylindrically smooth. Instead, it exhibits two helical grooves: the major and the minor groove, which have different width and depth. As a result, the proteins that bind to DNA tend to interact with the major groove, since the base-pairs are more accessible. There are different kinds of double helices characterized by their geometric properties such as the tilt angle of the bases or the interphos-

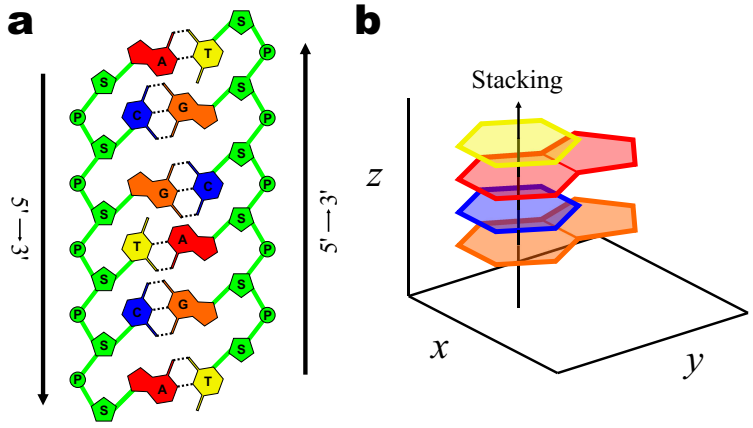


Figure 3.3: Secondary structure of DNA. **(a)** Hybridization of two antiparallel strands of DNA. The straight lines represent covalent bonds, while the discontinuous ones represent hydrogen bonds. **(b)** Stacking of four bases.

phate distance. The most common structures are the A-DNA (see Fig. 3.4a) and the B-DNA (see Fig. 3.4b), which both are right-handed double helices. In physiological conditions, DNA is found in the B form, while RNA is found in the A form. The A-DNA has a tilt angle of 20°, a base rise of 0.26 nm and a pitch of 11 bases per turn, while the values for the B-DNA are -3.6°, 0.34 nm and 10 bases per turn, respectively. The element that determines the difference between A-DNA and B-DNA is the sugar puckering (see Fig. 3.4c). Indeed, the 2′-deoxyribose is a cyclic molecule whose atoms do not lie in the same plane. In general, the 2nd ($C_{2'}$) and the 3rd ($C_{3'}$) carbons of the deoxyribose are out of the plane determined by the $C_{1'}$, $C_{4'}$ carbons and the oxygen. $C_{2'}$ and $C_{3'}$ can be found in two mutually exclusive conformations. In the $C_{2'}$-endo conformation, the 2nd carbon is above the plane of the sugar while the 3rd carbon is below. In the $C_{3'}$-endo conformation, the location of $C_{2'}$ and $C_{3'}$ carbons is inverted. Since the acid phosphoric is bonded to $C_{3'}$, the $C_{3'}$-endo conformation has a shorter interphosphate distance (i.e., the distance between the phosphates of two consecutive nucleotides) than the $C_{2'}$-endo conformation. Accordingly, the $C_{3'}$-endo conformation is observed in the A-DNA and the $C_{2'}$-endo conformation in the B-DNA. A different structure is Z-DNA, which is a left-handed double helix. There are some evidences that the Z form is a biologically relevant structure. There is another significant DNA structure: the S-DNA. It is a stretched double-helix with a large tilt angle and a low pitch. The S-DNA is postulated to have been observed after over-stretching the B-DNA and it was discovered using single-molecule techniques [48, 93]. However, this interpretation is now compromised [94].

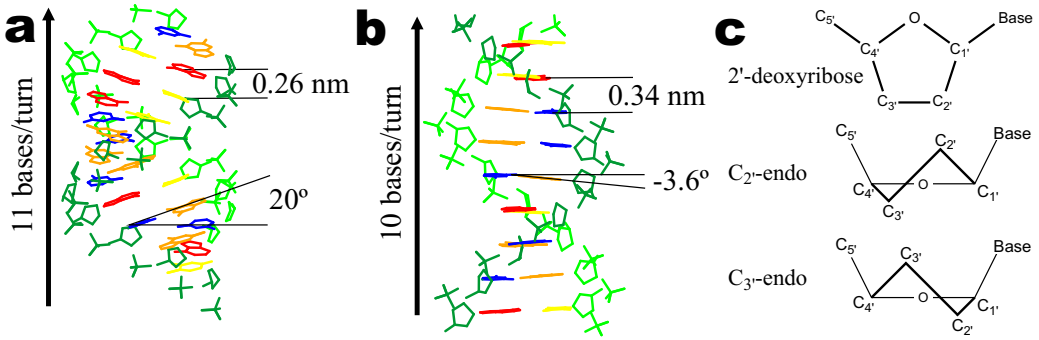


Figure 3.4: Tertiary structure of DNA (data obtained from [95]). **(a)** A-DNA. The phosphate-deoxyribose backbone is represented in green; adenine in red; cytosine in blue; guanine in orange; and thymine in yellow. **(b)** B-DNA. **(c)** Sugar puckering. The upper picture shows the chemical formula of the 2′-deoxyribose. The lower pictures show the two possible sugar pucker conformations.

Finally, the quaternary structure of a biomolecule is the assembly of different tertiary structures. This is quite relevant in proteins that form large complexes. In the case of eukaryotic DNA, the quaternary structure is called chromatin (see Fig. 3.5) [96]. The chromatin is a combination of DNA and essentially histones (i.e., a type of protein) that build complex structures that assemble to form the chromosomes. Chromatin is formed when DNA is wrapped around the histones and packed. A few single-molecule studies have studied nucleosome formation [97].

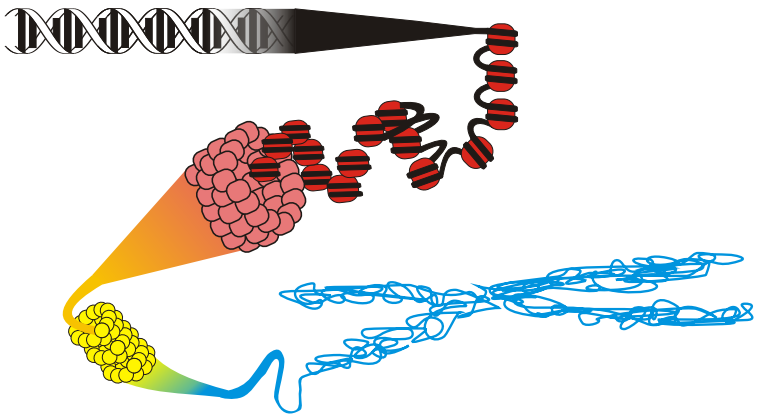


Figure 3.5: Quaternary structure of DNA (from top to bottom). The DNA is wrapped around the histones (proteins depicted in red). The histones are packed to form a helix (depicted in pink). Another super-helix is formed (depicted in yellow) which is the basic constituent of the chromosomes (depicted in blue).

What is the importance of the DNA structure? Charles Darwin sug-

gested in 1859 that natural selection is the key mechanism in the evolution of the species [98]. Similar ideas have been exported to other systems [99]. Biomolecules have been exposed to natural selection for millions of years. Therefore, the structure of nucleic acids and proteins has evolved to be capable of performing specific biological functions in an efficient way.

In the particular case of DNA, the double helix structure has several advantages. Here we enumerate some of the most relevant. First, genetic information is coded twice in the two complementary strands. This allows to safely store the information and check for errors during the replication process. Second, the external backbone protects the internal base-pairs (the ultimate carrier of the genetic information) from irreversible damage. Third, the lineal arrangement of the bases along the longitudinal axis of the DNA allows proteins to directly access to any fragment of the sequence. Fourth, the process of splitting and rejoining the two strands of DNA is reversible. This permits to carry out the replication and the transcription of DNA without damaging the original molecule. Finally, the double helix is a semi-flexible polymer that can be stretched, bended and twisted. As a result, the DNA can be compacted by forming super-structures (nucleosomes and chromosomes) that wrap the DNA around the histones.

3.2 Mechanical unzipping of DNA

In a very simplified view of life, all cells react to external impulses (food, heat) by performing some specific tasks (digest, crawl). All these specific tasks are carried out by proteins (enzymes, molecular motors). So cells need to continuously access to the genetic information stored in the sequence of DNA to synthesize proteins. Cell division is another example of a process that requires to access to the genetic information. In order to *read* the DNA sequence, the two strands of DNA must be split apart so that the cellular machinery can interact with the bases that code the instructions. For instance during replication, helicases are the proteins responsible of separating the two strands of DNA. The process of strand separation carried out by proteins was reproduced in vitro in the last decades of the 20th century [100]. Recently it has been studied at the single-molecule level [101, 102, 103].

The strand separation of DNA can also be produced in vitro without the proteins of the cell. This can be achieved by using an external agent such as temperature, chemical agents or force. Mechanical melting is a process that consists in pulling apart (i.e., unzipping) the two strands of a double-stranded DNA (dsDNA) molecule until the base pairs that hold the DNA duplex together are disrupted and two single-stranded DNA (ssDNA) molecules are

obtained. The term *unzipping* presumes that the DNA molecule is pulled from the $3'$ and the $5'$ extremities of one end of the molecule. The process of unzipping has its opposite: the rezipping. Indeed when the force is released, the DNA molecule tends to return to its native state, i.e., completely folded.

The mechanical separation of DNA was initially explored by pulling the DNA from its $5'$ extremities using an Atomic Force Microscopy (AFM) [104, 105]. The shear stress induced on the strands produced sudden disruptions of the base-pairs. The first unzipping of single molecules of DNA was carried out by Bockelmann and coworkers in 1997 using microneedles [106, 21, 107]. Inspired by a previous experiment [108], they tethered the DNA molecule between a coverslip and a movable glass microneedle, which allowed to measure the force by calibrating the deformation of the microneedle (see Fig. 3.6a). Mechanical unzipping was monitored by simultaneously measuring the displacement of the glass microneedle and the force applied on the molecule. The resulting force vs. extension curve (FEC) revealed a reproducible pattern which was correlated with the local content of Guanine and Cytosine (GC) along the DNA sequence. In particular, more force was required to unzip regions with high GC content as compared to regions with high AT content. Besides, the energies measured were compatible with the melting experiments performed in bulk [109]. Latter, Rief *et al.* were able to produce the overstretching and the unzipping of DNA in the same experiment using AFM (see Fig. 3.6b) [22, 110]. These experiments were a direct proof that the base-pairing forces in DNA were sequence specific. In 2001, Liphardt *et al.* [49] unzipped short RNA hairpins and observed coexistence and hopping between folded and unfolded states using optical tweezers. In a new improved experiment (see Fig. 3.6c), Bockelmann *et al.* repeated the DNA unzipping with optical tweezers [24]. Optical tweezers provided much more spatial and temporal resolution as well as a higher trap stiffness compared to glass microneedles. The unzipping experiments showed hopping between intermediate states and the experimental FEC was qualitatively reproduced by a mesoscopic model based on the Nearest-Neighbor model for nucleic acids [27, 26]. Still, the unzipping/rezipping curves showed some irreproducibility due to experimental drift, non-equilibrium effects (hysteresis) and statistical variation. A similar work done by the same group [111] showed that the unzipping experiments performed at high pulling speed deviated significantly from the quasistatic ones. They attributed such deviation to rotational drag effects. In 2003, Danilowicz *et al.* [23] unzipped DNA at constant force using magnetic tweezers (see Fig. 3.6d). The process showed a succession of breakages of base-pairs, typical of first order phase transitions. Since then, unzipping experiments have been the focus of attention of some studies [112, 113]. Unzipping has also been used as a tool to explore other

processes of the cell such as the protein-DNA interaction [114] or the translation of proteins observed in single ribosomes [115]. Besides, the theoretical studies have also focused their attention on different particular aspects of the unzipping experiments such as the scaling properties [29], the kinetics [116] or the problem of sequencing DNA by force [117].

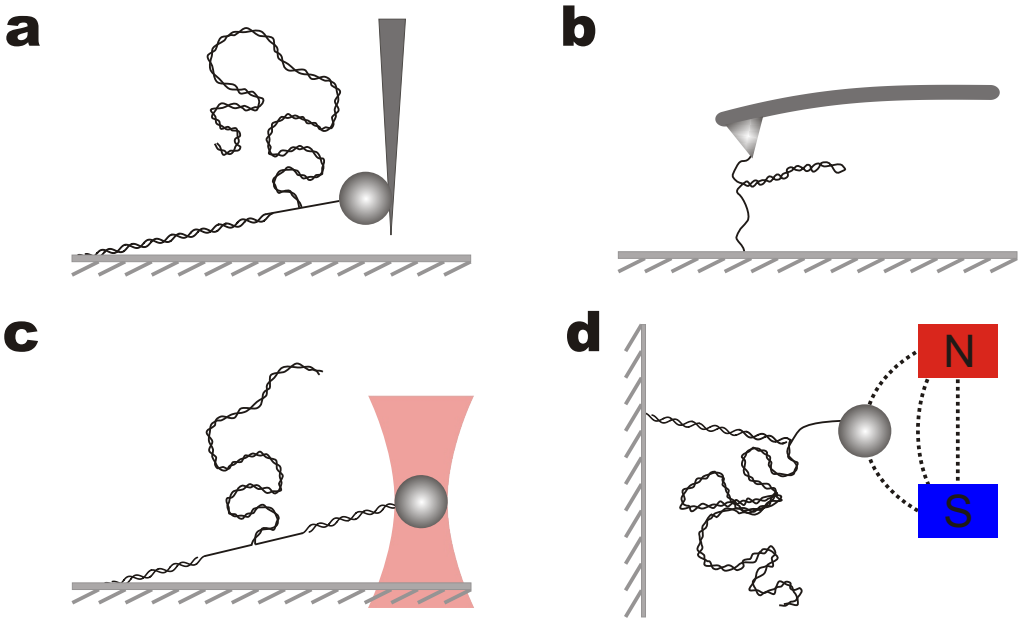


Figure 3.6: Unzipping of DNA with different techniques. **(a)** Microneedles [106]. **(b)** AFM [22]. **(c)** Optical tweezers [24]. **(d)** Magnetic tweezers [23].

3.2.1 Unzipping experiments

DNA unzipping can be performed in two different ways: 1) at controlled position or 2) at controlled force. The first one consists in separating both strands and applying whatever force is required to disrupt the bases. It is a sequential process and the disruption (i.e., opening) of bases occurs progressively. The second one consists in increasing the force applied to the strands of the molecule. In this protocol, all the base pairs of the DNA are suddenly disrupted when a critical force is reached. Depending on the experimental setup, the process of unzipping will be of one kind or the other. All these experiments can be performed at different speeds, which will determine whether the process is out of equilibrium or quasistatic.

One of the most important features of optical tweezers is that it allows to control and measure the extension of the DNA and the force applied to its ends. The possibility of measuring the force during the unzipping gives a lot

of information about the process. The typical forces exerted by an optical trap are of the order of magnitude of the critical force required to unzip the DNA molecule. Moreover, feedback algorithms are able to control the force at 4 kHz frequency. So both types of unzipping experiments (controlled position or controlled force) can be performed with optical tweezers.

In order to unzip DNA, the synthesized molecular constructs are attached to two homogeneous beads made of polystyrene that have an index of refraction of $n = 1.51$. One bead is a Streptavidin (SA) coated microsphere (2.0-2.9 μm; G. Kisker GbR, Products for Biotechnology). The other bead is a protein G microsphere (3.0-3.4 μm; Spherotech, Libertyville, IL) coated with anti-digoxigenin (AD) polyclonal antibodies (Roche Applied Science). The SA bead bonds to the biotin-labeled handle of the molecular construct, while the AD bead bonds to the other handle, which is labeled with digoxigenins. The rest of the elements of the molecular construct (ssDNA, dsDNA, loop) do not stick to polystyrene beads. This assures that unzipping experiments are not affected by unintended interactions. The SA bead is usually fixed at the tip of the micropipette by air suction, while the AD bead is captured in the optical trap (see Fig. 3.7).

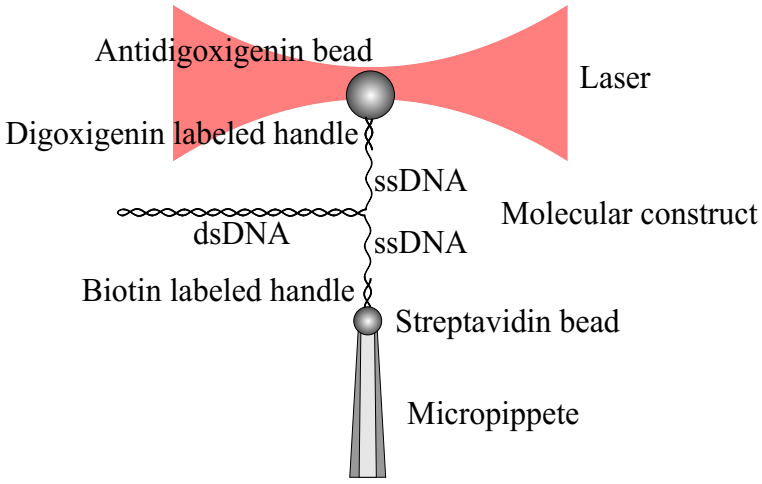


Figure 3.7: Experimental setup to unzip the DNA molecular constructs. Picture not to scale.

3.2.2 Experimental setup to unzip DNA molecules

Unzipping experiments using optical tweezers consist of two steps: 1) synthesis and preparation of the samples to be pulled and 2) attachment of the molecule between the two beads. The procedures to synthesize DNA, RNA

or proteins are very different. Thus, the synthesis of the molecule depends on the type of molecule to be pulled. Section 3.2.3 describes how the DNA molecules to be unzipped are synthesized in the lab.

1. The preparation of DNA molecules for pulling requires the incubation of the synthesized DNA molecules with the coated AD beads. This process consists in mixing the DNA molecules with beads in a ratio of approximately one bead per molecule. The incubation time is between 15-30 minutes at room temperature. During this time, the digoxigenin labeled handles of the DNA molecules are bonded to the antidigoxigenins that coat the beads. The sample is diluted and introduced in the upper channel of the fluidics chamber (see Fig. C.5 in Appendix C.4). Another sample is prepared by diluting a certain amount of SA beads in buffer. This second preparation is inserted in the lower channel of the fluidics chamber (Fig. C.5). The central channel is then loaded with the corresponding buffer at which the pulling experiment is performed.

2. The *tethering* or connection of the molecule between two beads is carried out within the fluidics chamber. First, a SA bead is trapped at the exit of the bearing tube of the lower channel. With the help of the optical trap, the bead is located and immobilized at the tip of the micropipette by air suction. Second, an AD bead (already incubated with DNA) is trapped at the exit of the bearing tube of the upper channel. Finally, the beads are brought close to each other until a tether between the SA bead an the biotin labeled handle of the DNA sample is achieved (see Fig. 3.8).

At this point, the optical trap measures a force when the beads are separated and the system is ready to start the DNA unzipping experiments.

3.2.3 Synthesis of DNA molecules

The unzipping of long molecules of DNA requires the design and synthesis of molecular constructs that can be pulled with optical tweezers. Optical tweezers use beads to apply forces on the molecules. Therefore, the DNA molecules have to be synthesized with labeled ends, so that the DNA can be bonded to the beads. In our case, the labels must be positioned in the $3'$ and $5'$ extremities of the handles.

The synthesis of DNA constructs involves a series of simple steps such as cleavage, electrophoresis or annealing among others. Protocols to carry out these steps are commercially available and are highly efficient. Our protocol

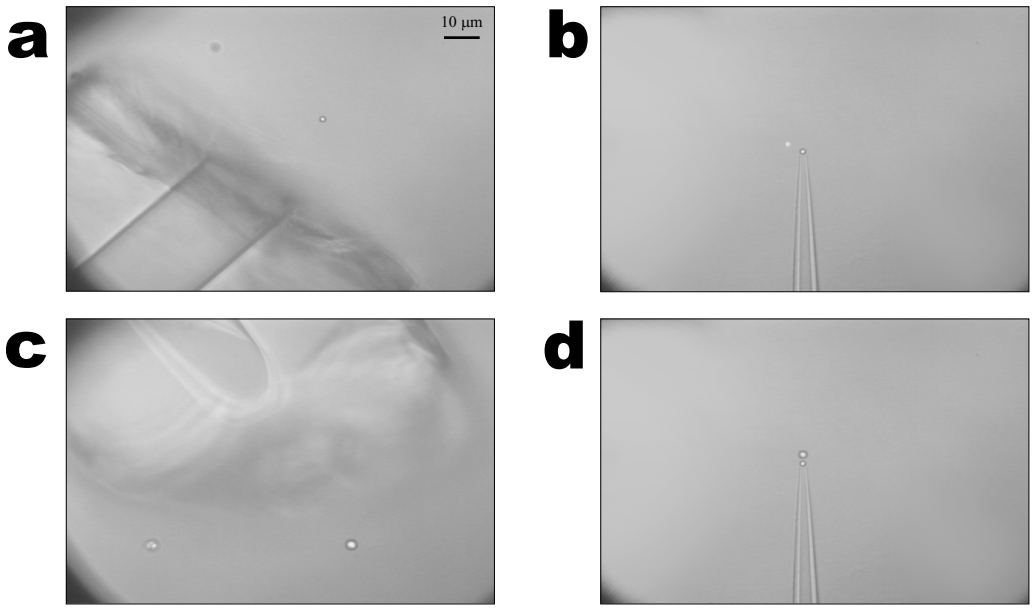


Figure 3.8: Screenshots that show the formation of an attachment between the molecular construct and the beads in the fluidics chamber. **(a)** Trapping of a SA bead. **(b)** Immobilization of the SA bead at the tip of the micropipette. **(c)** Trapping of a AD bead. **(d)** Tethering of the molecule between the two beads.

of synthesis starts from λ-DNA molecules which is then cut into several fragments. Latter, the desired fragments are selected and the labeled handles are added. Two different sequences of DNA have been used in our experiments. The first one has a length of 6.8 kbp and the second one has 2.2 kbp. In what follows, we describe the procedure to synthesize the 6.8 kbp sequence.

6.8 kbp sequence

A 6770 bp insert DNA is isolated by gel extraction of a BamHI digestion of λ phage DNA (see Fig. 3.9). Two short handles of 29 bps and one tetraloop (5′-ACTA-3′) are ligated to the insert that has the λ cosR end and a BamHI sticky end. To construct the DNA handles, an oligonucleotide (previously modified at its 3′ end with several digoxigenins using DIG Oligonucleotide Tailing Kit, 2nd Generation, Roche Applied Science) is hybridized with a second 5′ biotin-modified oligonucleotide giving a DNA construction with one cohesive end complementary to cosR and two 29 nucleotide long ssDNA at the other end. These two ssDNA have the same sequence and they are hybridized with a third oligonucleotide, which is complementary to them, resulting in two dsDNA handles. This construction is attached to the insert

DNA by a ligation reaction. A fourth self-complementary oligonucleotide, which forms a loop in one extreme and a cohesive BamHI end at the other, is ligated to the BamHI sticky end of the insert DNA. The DNA is kept in aqueous buffer containing 10 mM Tris-HCl (pH 7.5) and 1 mM EDTA. See Appendix E for the detailed sequence.

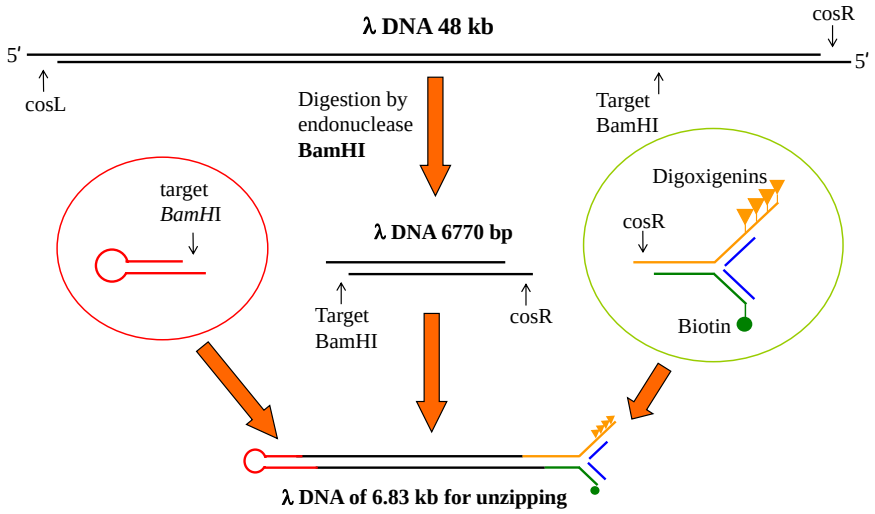


Figure 3.9: Synthesis of the 6.8 kbp sequence.

2.2 kbp sequence

The protocol of synthesis is similar to the previous one. The 2.2 kbp construct is obtained taking the 2215 bp fragment from a SphI digestion of λ-DNA. The same two short handles and tetraloop used for the 6.8 kbp DNA are used for the 2.2 kbp construct, with the following exceptions: the two handles are hybridized to the 2215 bp DNA through the cosL cohesive end of λ DNA, and the tetraloop is added to the insert DNA using the SphI sticky end (see Fig. 3.10). See Appendix E for the detailed sequence.

3.2.4 Experimental data of unzipping experiments

Once the molecule has been tethered between the two beads, the unzipping experiment can start. There are two types of unzipping experiments: controlled position and controlled force. The natural way to produce unzipping with optical tweezers is at controlled position. The reason is that the position of the optical trap is a parameter that we directly control by changing

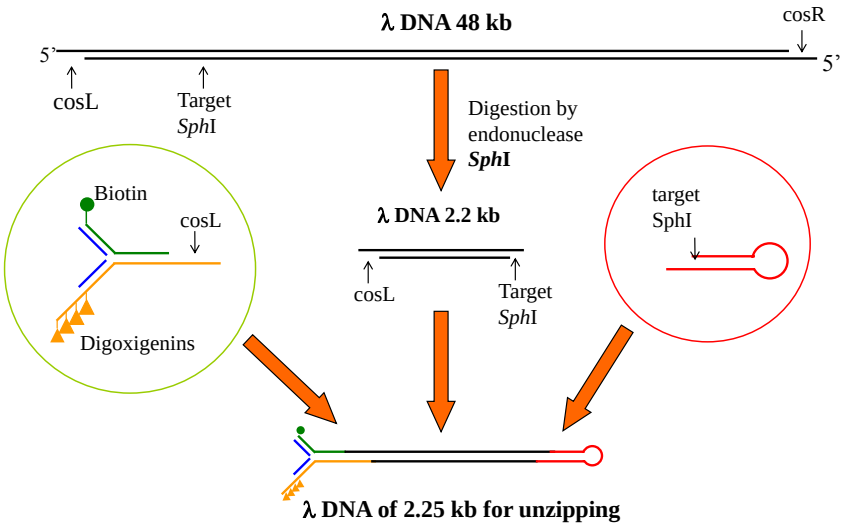


Figure 3.10: Synthesis of the 2.2 kbp sequence.

the voltage of the piezos. Instead, the controlled force experiments require the activation of a force feedback. The process of unzipping looks different if the experiment is performed at controlled position or force. Thus, the data measured with the instrument depends on the type of experiment.

Controlled position experiments

In the controlled position experiments, the optical trap is moved relative to the micropipette and the distance between the two beads increases. The separation of the beads induces a force on the handles of the DNA molecular construct, which is transmitted to the unzipping fork of the DNA. When the force is high enough ($\sim$ 15 pN), the base-pairs cannot withstand the force and suddenly break. The breakage of a group of bases release a certain amount of ssDNA, which increases the total length of the molecular construct. This reduces the force exerted on the ends of the molecule. When the separation between the traps is increased again, this very same process is repeated and more base-pairs are open until the two strands of the molecule are completely disrupted. Once this situation is reached, the loop keeps the two strands linked and the molecule exhibits the elastic response of the ssDNA (last part of the FDC). On the other hand, if the beads are brought close one to each other, the base-pairs of the DNA molecule tend to close and the molecule is rezipped.

During all this unzipping/rezipping process, the instrument simultaneously measures the position (i.e., the so-called distance) of the optical trap

and the force applied to the ends of the molecule. These two magnitudes are usually represented together in a Force vs. Distance Curve (FDC). The typical FDC of an unzipping experiment exhibits a sawtooth pattern, which is composed of a series of slopes and sudden decreases in force (see Fig. 3.11a). The slopes correspond to the elastic response of the molecule as it is being pulled. The sudden force decrements (also known as force rips) correspond to the breakage of base-pairs (see Fig. 3.11b). The slopes and force rips of the sawtooth have different shapes, widths and heights and they depend on the DNA sequence. Different sequences have different sawtooth patterns. The pulling speed is the rate (in nanometers per second) at which the beads are pulled apart. At slow pulling rate, the unzipping and rezipping processes are almost quasistatic and their FDC overlap, except for some unavoidable hysteresis when opening the loop (see Fig. 3.11c). At fast pulling rate, the force rips exhibit hysteresis, the process is no longer quasistatic and the unzipping and rezipping patterns differ significantly.

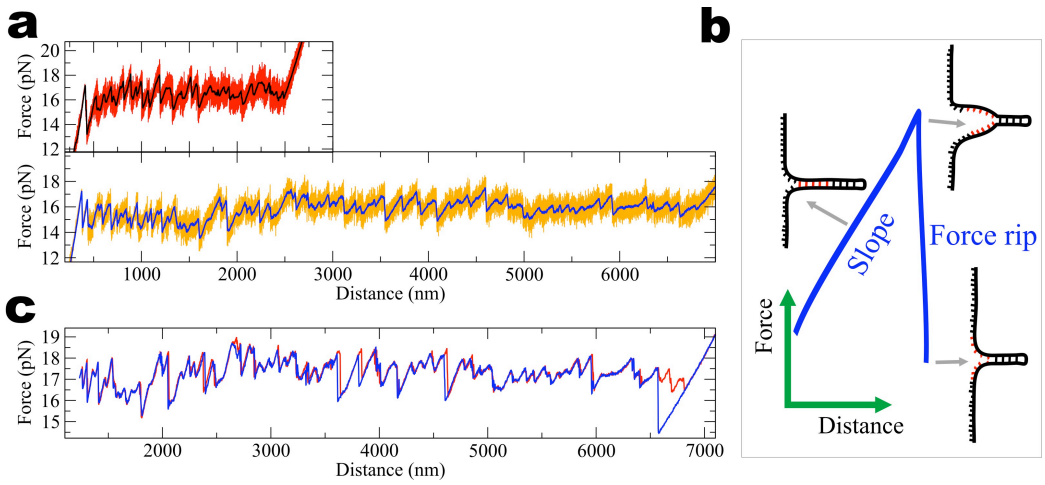


Figure 3.11: Unzipping of a DNA molecule at controlled position. **(a)** Sawtooth pattern of the FDC. Upper (lower) panel shows the FDC of a 2.2 kbp (6.8 kbp) sequence. Red and orange curves show the raw data. Black and blue curves show the filtered data at 1 Hz bandwidth. **(b)** The slope of the sawtooth corresponds to the elastic response whereas the force rip corresponds to the opening of base pairs. The group of bases that are disrupted in the force rip are depicted in red. **(c)** A quasistatic cycle of pulling/relaxing. Red (blue) curve shows the unzipping (rezipping).

Controlled force

The unzipping experiments at controlled force consist in continuously increasing the force applied at the ends of the DNA molecule until it disrupts at a

critical force. The disruption is much more abrupt than in the controlled position unzipping. In other words, the number of broken base-pairs increases dramatically as we reach the critical force. Indeed, within a range of 2 pN of force the molecule changes its state: from completely closed base-pairs to completely open.

The FDC measured in this kind of experiments is quite different from the ones obtained at controlled position. Its shape is different and it no longer exhibits a sawtooth pattern. Instead, it is a monotonically increasing step function, so the force and distance always increase together (see Fig. 3.12). The rezipping at controlled force is observed when the force applied to the molecule is released. The shape of the function is very similar but a large hysteresis is observed. Differently from the previous experiments, the hysteresis in the controlled force experiments cannot be reduced at will because the process is never quasistatic.

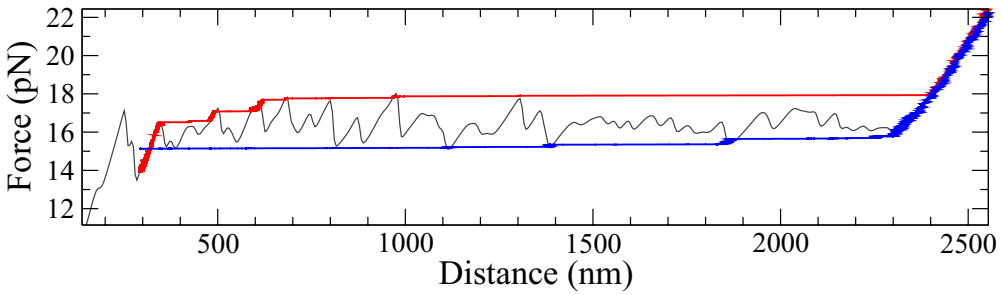


Figure 3.12: Unzipping of a DNA molecule at controlled force. Red (blue) curve shows unzipping (rezipping) FDC. The unzipping FDC at controlled position has been superimposed in gray.

3.3 The Nearest-Neighbor model

The Nearest-Neighbor (NN) model for DNA was developed to predict the free energy of formation (hereafter, simply referred as energy) of a double helix of DNA starting from two complementary strands of DNA [26, 27]. The model has been extended to include all possible secondary structures that can be formed with a single strand. The NN model assumes that the ssDNA is already formed (i.e., the phosphate backbone is already formed) and only base-pairing interactions are taken into account. The model predicts the formation energy of all the possible secondary structures we can form with the given sequence, but it does not predict the state of minimum energy (i.e.,

the native state). In order to find the native state, it is necessary to search the less energetic state among all possible secondary structures [118]. This requires the implementation of optimization algorithms such as *Mfold* that use the NN model to minimize the energy of a given sequence of ssDNA [119].

The basic idea of the NN model is that the base-pairing energy of two complementary bases only depends on the base itself and on the first neighbor located in the same strand. This allows to separate the computation of the global energy into many contributions called motifs (see Fig. 3.13a). Each motif represents a possible pairing of bases and the total energy of a state is given by the sum of the energies of different motifs. The energy of a given motif includes three main contributions: 1) the hydrogen bonds between bases, 2) the stacking interaction between consecutive bases and 3) the loss of entropy (see Fig. 3.13b).

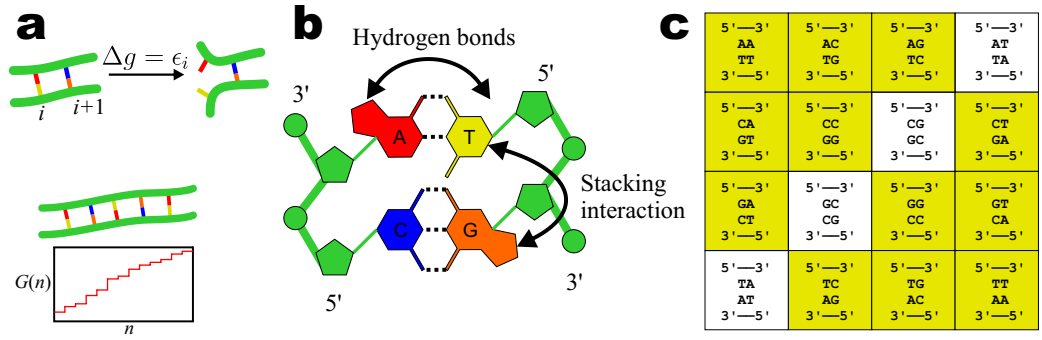


Figure 3.13: NN model. **(a)** Duplex formation. The formation energy (Δg) of the base pair i depends on itself and on the NN base pair $i+1$. Each base pair contributes to the total formation energy of the duplex ($G(n)$). **(b)** Interactions between bases. The hydrogen bonds between complementary bases, the stacking interactions and the loss of entropy (not depicted) constitute the free energy of one base pair. **(c)** The 16 NNBP motifs. The 12 highlighted motifs are symmetric with respect to the anti-diagonal. Each motif has the same energy than its symmetric. In the end, there are 10 different energies (6 symmetric + 4 anti-diagonal).

If the NN model is restricted to Watson-Crick base pairs, the number of possible motifs is dramatically reduced. Since DNA has four types of bases and the interaction involves the nearest neighbor, we end up with $4 \times 4 = 16$ different possible motifs. However, 6 pairs of them must have the same formation energy due to symmetry reasons (see Fig. 3.13c). Indeed, one given ssDNA determines the sequence of its complementary strand if we are restricted to Watson-Crick base pairs. This means that the energy of the duplex can be computed either knowing one strand or the other and the value of the energy must be the same in both cases, which imposes 6 restrictions. Therefore, only 10 different energy values define the free energy of formation

of Watson-Crick sequences of DNA.

The NN model has been largely investigated in calorimetry and UV absorbance experiments. Calorimetry measurements allows us to infer the enthalpy and entropy of formation of chemical substances by measuring the heat absorbed or emitted by a vessel that contains the reactive substances while the temperature is increased. In particular, it is possible to isolate the contribution of each motif by gathering calorimetric experiments performed on several sequences of DNA. In the end, a table of enthalpies and entropies for each motif can be obtained.

3.3.1 Thermodynamics of DNA duplex formation

As mentioned before, the NN model allows us to compute the thermodynamic properties of DNA duplex formation starting from the enthalpies and entropies of the elementary motifs that constitute all possible Watson-Crick sequences of DNA. This section describes how to calculate these magnitudes.

The sequence of an N-base ssDNA (σ_i) is given from the $5'$ end to the $3'$ one, where i=1 is the base located at the $5'$ end, i=N is the base located at the $3'$ one. σ_i can take four different values ($\sigma = \{A, C, G, T\}$). The standard enthalpy of hybridization of such sequence (ΔH^0) with its complementary strand can be calculated according to,

$$\Delta H^0 = \Delta H_{\text{init}} + \sum_{i=1}^{N-1} \Delta h^0(\sigma_i, \sigma_{i+1}) \tag{3.1}$$

where ΔH_{init} is the initiation term that depends on the ends of the sequence ($i = 1, N$); Δh^0 is the formation energy of Watson-Crick motifs and it depends on the nearest neighbor ($\sigma_i, \sigma_{i+1} \in \{AA, AC, \ldots, TT\}$); and the sum is extended over all NN base pairs. The notation is usually simplified by assuming that the $i+1$ base is already known from the sequence. Thus, the arguments of Δh^0 can be removed,

$$\Delta H^0 = \Delta H_{\text{init}} + \sum_{i=1}^{N-1} \Delta h_i^0 \tag{3.2}$$

where Δh_i^0 is the enthalpy of formation of the motif (σ_i, σ_{i+1}). Similarly, the entropy contribution (ΔS_i^0) in the hybridization reaction can be computed according to

$$\Delta S^0 = \Delta S_{\text{init}} + \sum_{i=1}^{N-1} \Delta s_i^0 \tag{3.3}$$

Term	Motif	ΔH°	ΔS°
Constant contribution	All	0.2	-5.6
A/T penalty	5′-A...-3′ 5′-T...-3′ 5′-...A-3′ 5′-...T-3′	2.2	6.9
TA penalty	5′-TA...-3′ 5′-...TA-3′	-0.4	0.5

Table 3.1: Initiation terms. The contributions are not mutually exclusive. It means that they are added to the total ΔH° and ΔS° for each motif of the oligo. For instance, the sequence 5′-TA...C-3′ will have the following initiation term for the enthalpy: $\Delta H_{\text{init}} = 0.2 + \Delta h^0_{\text{A/T}} + \Delta h^0_{\text{TA}}$. Enthalpies are given in kcal/mol and entropies, in cal/mol·K.

where ΔS_{init} is the initiation term and Δs^0_i is the entropy of formation of the motif $\sigma_i\sigma_{i+1}$.

The last bases of the sequence do not have a complete stacking interaction because they only have one neighbor. This is why the NN model has to introduce the initiation term. This term always adds a constant contribution, which is sequence independent. Moreover, a variable contribution has to be added depending on the type of bases located at the ends: an extra contribution of $\Delta h^0_{\text{A/T}}$ and $\Delta s^0_{\text{A/T}}$ must be added if the sequence starts with a 5′-A...-3′ or 5′-T...-3′; or ends with a 5′-...A-3′ or 5′-...T-3′. If the molecule has both motifs (start and end) the contribution has to be counted twice. Moreover, if the molecule specifically starts with 5′-TA...-3′ or ends with 5′-...TA-3′, there is another extra contribution of Δh^0_{TA} and Δs^0_{TA} to the initiation terms. Table 3.1 summarizes all these cases.

In general, ΔH^0_i and ΔS^0_i are considered to be temperature independent, meaning that the change in heat capacity can be neglected ($\Delta C_p = 0$). This allows to compute the free energy of formation (ΔG^0),

$$\Delta G^0 = \Delta H^0 - T\Delta S^0 \tag{3.4}$$

where T is the temperature at which the free energy is calculated. Following the same scheme, it is also possible to define the free energy of formation (Δg^0_i) of one NN motif according to

$$\Delta g^0_i = \Delta h^0_i - T\Delta s^0_i \ . \tag{3.5}$$

3.3.2 Melting temperatures

The melting temperature of a DNA hairpin is defined as the temperature at which half of the molecules are in the native state (i.e., the double helix state) and half are denaturated (i.e., the two strands are split). It is implicitly assumed that the hybridization of the molecule is well described by a two-state model, in which the base-pairs are either fully bonded or disrupted (no partial disruption of bases is considered). This model is convenient for oligos shorter than 100 bp and the predicted melting temperatures deviate from the experimental measurements if the DNA is longer (the so-called polymeric DNA).

The NN model is widely used to predict melting temperatures of DNA duplexes. In order to predict melting temperatures, the enthalpy (Eq. 3.2) and the entropy (Eq. 3.3) of formation of the DNA hairpin must be known. The melting temperature of a non-self-complementary duplex[1] is deduced from the the equation that relates the equilibrium constant with the free energy of formation of the hairpin,

$$\Delta G = \Delta G^0 + RT \ln K \tag{3.6}$$

where ΔG is the free energy of formation of the hairpin at the desired experimental conditions, ΔG^0 is the standard free energy calculated with Eq. 3.4, R is the gas constant, T is the temperature and K is the equilibrium constant. The hybridization reaction can be written as

$$\mathrm{ssDNA_1 + ssDNA_2} \rightleftharpoons \mathrm{dsDNA} \tag{3.7}$$

where $\mathrm{ssDNA_1}$ and $\mathrm{ssDNA_2}$ are the two complementary strands and dsDNA is the hybridized duplex. The equilibrium constant of the hybridization reaction is given by,

$$K = \frac{[\mathrm{dsDNA}]}{[\mathrm{ssDNA_1}] \cdot [\mathrm{ssDNA_2}]} \tag{3.8}$$

where $[\mathrm{ssDNA_1}]$, $[\mathrm{ssDNA_2}]$ and [dsDNA] are the concentrations (in molar units M=mols$\cdot$l^{-1}) of the single strands and the duplex, respectively. By definition, the melting temperature is the one in which [dsDNA] = $[\mathrm{ssDNA_1}]$ = $[\mathrm{ssDNA_2}]$. If the sample has a total concentration of strands equal to c_T, it is expected that at the melting temperature the strands will be distributed according to,

$$\begin{array}{rccccc} & \mathrm{ssDNA_1} & + & \mathrm{ssDNA_2} & \rightleftharpoons & \mathrm{dsDNA} \\ c_T = & x & + & x & + & 2x \end{array} \tag{3.9}$$

[1] A non-self-complementary duplex is an oligo which cannot hybridize with itself. For instance, 5′-ATCGCGAT-3′ is self-complementary.

where x is the concentration of strands at equilibrium (the concentration of strands is double for [dsDNA]). From this, we deduce that $c_T = 4x$ at the melting temperature and the concentrations are equal to $[\text{ssDNA}_1] = [\text{ssDNA}_2] = [\text{dsDNA}] = c_T/4$. Note that one molecule of dsDNA contributes twice to the total concentration of ssDNA, so we have

$$\begin{array}{rcccccc} c_T & = & [\text{ssDNA}_1] & + & [\text{ssDNA}_2] & + & 2 \times [\text{dsDNA}] \\ & = & c_T/4 & + & c_T/4 & + & 2 \times c_T/4 \ . \end{array} \tag{3.10}$$

By introducing these concentrations into Eq. 3.8 we find $K = (c_T/4)^{-1}$ at equilibrium. At the melting temperature (T_m) the hybridization reaction is in equilibrium, so $\Delta G = 0$. Considering this and introducing Eq. 3.4 into Eq. 3.6 it is possible to write,

$$0 \equiv \Delta G = \Delta H^0 - T_m \Delta S^0 + R T_m \ln (c_T/4)^{-1} \tag{3.11}$$

which leads to the following result after solving for T_m:

$$T_m = \frac{\Delta H^0}{\Delta S^0 + R \ln[c_T/4]} \tag{3.12}$$

3.3.3 Salt dependence

The prediction of melting temperatures is commonly done at standard conditions, which means that the concentration of salt is [NaCl]=1 M. However, this prediction can be extended when the hybridization reaction is performed at a different concentration of salt. According to the studies of calorimetry, the entropy is the only thermodynamic magnitude that depends on the salt due to rearrangements of the cloud of counterions along the DNA. This is deduced from the melting experiments performed on oligos, in which the enthalpy is directly obtained from the slope of the van't Hoff equation [120]. The results for several molecules show no significant enthalpic differences at different salt concentrations [121], meaning that the enthalpy is salt-independent. So, the salt correction only affects the entropy. The values of Δs_i^0 that enter in Eq. 3.3 must be substituted by Δs_i according to

$$\Delta s_i([\text{Na}^+]) = \Delta s_i^0 + \frac{m_i(T)}{T} \ln[\text{Na}^+] \tag{3.13}$$

where $\Delta s_i([\text{Na}^+])$ is the formation entropy of motif i at a salt concentration of $[\text{Na}^+]$, T is the temperature and $m_i(T)$ is a pre-factor that corrects for the salt and it is temperature-dependent. Although m_i depends on T, the whole prefactor $m_i(T)/T$ is temperature independent. Therefore, $m_i(T)$ is linear with T.

In general, the values of m_i do not need to be identical for all the NN motifs. However, in the studies of calorimetry they are taken to be all the same at first approach, and their value is $m_i = 0.114$ kcal·mol^{-1}·K^{-1} at $T = 310$ K [28]. On the other hand, the salt correction does not apply to the initiation term (ΔS_{init}) of Eq. 3.3. After introducing the salt correction, Eq. 3.12 can be rewritten as

$$T_m = \frac{\Delta H^0}{\Delta S^0 + \sum_{i=1}^{N-1} \frac{m_i(T)}{T} \ln[\text{Na}^+] + R \ln[c_T/4]} \tag{3.14}$$

where the new term in the denominator is a summation that rolls over all the bases of the sequence. Strictly speaking, the summation should be done for the number of phosphates of the DNA, divided by two [28]. In our case, it is identical to the number of base pairs of the molecular construct. Again, the prefactor $m_i(T)/T$ is temperature independent and the value of T at which it is measured is not important.

3.4 Modeling of experimental setup

In order to understand the mechanism of the unzipping process it is useful to have a model that reproduces the physics behind the experiments. A description of the experiment can be done at several degrees of accuracy: from an atomistic description to mesoscopic models. In general, each model focuses on a particular aspect of the experiments: thermodynamics, kinetics, scaling properties, etc. The aim of this section is to develop a mesoscopic model to study the thermodynamics of the DNA [112]. The model describes separately the different constituents of the experimental setup by using other submodels such as the NN model (described in the previous section) and the models for the elasticity of polymers. The standard techniques of statistical mechanics can be applied to the model in order to compute the partition function, the free energy landscape and the equation of state of the system.

In most macroscopic thermodynamic systems (e.g., ideal gas) the computation of the equation of state can be performed in different statistical ensembles (microcanonical, canonical, grand canonical). In principle, the result should not depend on the ensemble selected to perform the calculation if we take the thermodynamic limit. In the case of the DNA, the calculation of the thermodynamic properties of the model can be done in two ensembles: controlled position (also known as mixed [122, 112] or isometric [123]) ensemble and controlled force (also known as isotensional [123]) ensemble (see Fig. 3.14). In DNA unzipping, though, the resulting equation of state depends on the ensemble. The differences between them are due to the finite

size effects. Following, there are two separate descriptions of the model, one for each ensemble.

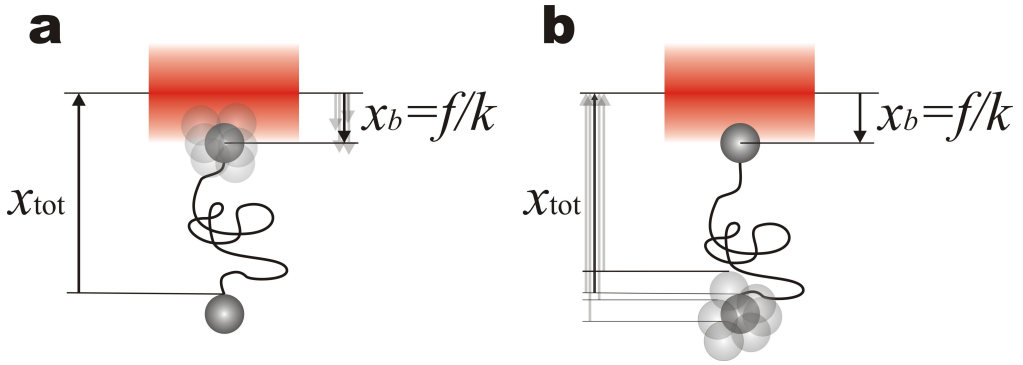


Figure 3.14: Ensembles. **(a)** Controlled position ensemble. The total distance (x_{tot}) between the anchor point and the center of the optical trap is held constant. The position of the bead (x_b) in the optical trap fluctuates and so does the applied force f (k is the trap stiffness). **(b)** Controlled force ensemble. The force is controlled by keeping constant the position of the bead in the optical trap (x_b). In order to do so, the total distance of the system (x_{tot}) must be corrected by a feedback. Therefore, x_{tot} fluctuates. This ensemble can also be implemented by applying a uniform field of force as in magnetic tweezers.

3.4.1 Controlled position

The position of the center of the optical trap is the control parameter that characterizes the controlled position ensemble. In this ensemble, the position of the trap is fixed and the force applied to the ends of the molecular construct fluctuates. The equation of state of the system in this ensemble can be experimentally reproduced by measuring the average force exerted on the molecular construct for each fixed position of the optical trap. The natural thermodynamic potential of this ensemble is the Helmholtz free energy, which is expressed in terms of the extensive variable (the distance) and the temperature. Therefore the energetic contributions of the model have to be expressed in terms of distances (or extensions).

The description of the molecule is split into three different parts: the handles, the open base pairs (which are in the form of single stranded DNA) and the closed base-pairs (see Fig. 3.15). We also have to take into account the effect of the optical trap. The bead in the optical trap is modeled by a Hookean spring,

$$f = kx_b \tag{3.15}$$

where f is the force applied, k is the measured stiffness of the optical trap and x_b is the elongation of the bead from the center of the trap. So, the potential

energy of the bead in the optical trap E_b is described by a harmonic potential

$$E_b = \frac{1}{2} k x_b^2 \ . \tag{3.16}$$

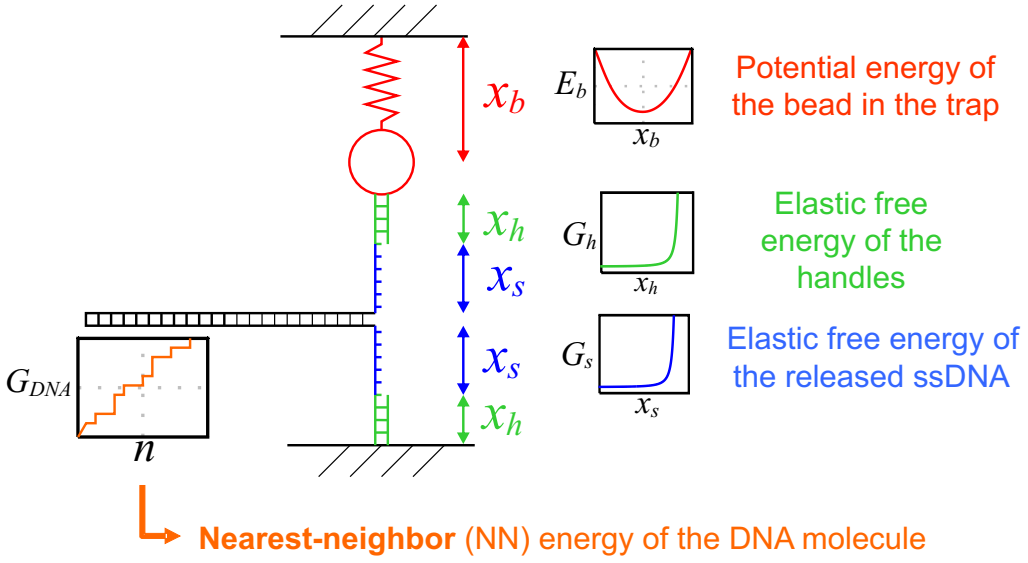


Figure 3.15: Mesoscopic model. Each element is represented using a different color. A sketch of the free energy contribution is also shown for each element.

We use the NN model to describe the free energy of formation of the DNA duplex. The free energy required to open n base pairs is given by the sum of the free energies required to open each consecutive nearest-neighbor pair,

$$G_{\mathrm{DNA}}(n) = -\sum_{i=0}^{n} \epsilon_i \tag{3.17}$$

where $G_{\mathrm{DNA}}(n)$ is the free energy of the hairpin when n bps are disrupted and ϵ_i is the free energy of formation of the bp i. The values of ϵ_i are negatively defined (because the base-pairing is a spontaneous reaction) and they are identical to Δg_i (see Eq. 3.5 for a definition of Δg_i^0 in standard conditions). The resulting function $G_{\mathrm{DNA}}(n)$ is positively defined and monotonically increasing with n (see Fig. 3.15). So, the energy is minimum and equal to 0 when all the base pairs are closed ($G_{\mathrm{DNA}}(0) = 0$) and it is maximum when all base pairs are open ($G_{\mathrm{DNA}}(N) = \sum_{i=1}^{N} \epsilon_i$). The free energy of the duplex depends on the sequence of base pairs. An extra free energy contribution is included in the model to account for the disruption of the end loop when all the base pairs are open (i.e., $n = N$).

Elastic models for polymers are used to describe the elasticity of the handles and the ssDNA released during the unzipping process. The handles are dsDNA and they are modeled using the force vs. extension curve of a Worm-Like Chain (WLC)

$$f_h(x_h) = \frac{k_B T}{4 l_p} \left(\left(1 - \frac{x_h}{L_0} \right)^{-2} - 1 + 4 \frac{x_h}{L_0} \right) \tag{3.18}$$

where k_B is the Boltzmann constant and T is the temperature, l_p is the persistence length and L_0 is the contour length. The elastic free energy of the handles ($G_h(x_h)$) is obtained by integrating the previous expression (see Appendix F for further details) according to

$$G_h(x_h) = \int_0^{x_h} f_h(x')\, dx' \tag{3.19}$$

where x' is a dummy variable. The ssDNA is modeled using either a WLC or a Freely-Jointed Chain (FJC) model. Depending on the salt concentration of the experiment, one model or the other describes better the elastic response of the ssDNA. In the case of the FJC model, the following equation gives the extension vs. force curve,

$$x_s(f, n) = nd \cdot \left(\coth \left(\frac{bf}{k_B T} \right) - \frac{k_B T}{bf} \right) \tag{3.20}$$

where k_B is the Boltzmann constant and T is the temperature, b is the Kuhn length, n is the number of bases and d is the interphosphate distance of the ssDNA. The product $L_0 = n \times d$ is the contour length of the molecule. It is convenient to write the explicit dependence on n in Eq. 3.20 because n is not a parameter but a variable that changes during unzipping. Again, the elastic free energy of the ssDNA is obtained by integrating the force versus molecular extension. However, in Eq. 3.20 f is the independent variable so the calculation of $G_s(x_s, n)$ requires the inversion of $x_s(f, n)$. It can be avoided by performing an integration by parts (see Appendix F) according to

$$G_s(x_s, n) = \int_0^{x_s} f_s(x', n)\, dx' = f \cdot x_s(f, n) - \int_0^{f} x_s(f', n)\, df' \tag{3.21}$$

where f' is a dummy variable and the limits of integration are related by $x_s = x_s(f, n)$. The total free energy of the global system (G_{tot}) is given by the sum of all free energy contributions,

$$G_{\text{tot}}(x_{\text{tot}}, n) = E_b(x_b) + 2G_h(x_h) + 2G_s(x_s, n) + G_{\text{DNA}}(n) \tag{3.22}$$

where x_{tot} is the total extension of the system. The energy of the handles and the ssDNA has to be counted twice because there are two handles and two fragments of released ssDNA. Eventually, the total free energy of the system is completely determined by the number of open base pairs and the total distance (x_{tot}), which is given the by sum of the extensions of all the elements

$$x_{\text{tot}}(f, n) = x_b(f) + 2x_h(f) + 2x_s(f, n) \tag{3.23}$$

where x_b and x_s are obtained directly from Eq. 3.15 and 3.20 respectively and x_h requires to invert Eq. 3.18 numerically. The extension of the elastic elements only depend on the force f applied to the ends of the molecular construct. By doing this we are implicitly assuming that x_b, x_h and x_s are in local equilibrium and they do not fluctuate. A more accurate calculation requires to let the forces applied to the ends of each element fluctuate. However, such calculation is more complex and it does not introduce any appreciable difference in the final result. Appendix G discusses this issue in more detail.

Besides, the extension of the ssDNA in Eq. 3.23 depends on the number of open base pairs n. Additionally, constant terms might also contribute to the total extension of the system (e.g., the bead diameters or the width of the DNA duplex). The thermodynamic properties of the model are not affected if these are neglected, though.

In the end, the calculation of the total energy at a given total extension x_{tot} and number of open base pairs n requires to solve f in Eq. 3.23 by using Eqs. 3.15, 3.18 and 3.20. Once the equilibrium force of the system is known, it can be used to recover the extension of each element and calculate the total energy of the system. Therefore, Eqs. 3.22 and 3.23 form a system of two coupled equations that determine the force and the total energy of the system.

Free energy landscape

The free energy landscape (also called potential of mean force) is the total energy of the system $G_{\text{tot}}(x_{\text{tot}}, n)$ (see Eq. 3.22) as a function of two variables: x_{tot} and n. These variables are defined between $0 < x_{\text{tot}} < +\infty$ and $0 < n < N$, where N is the total number of base pairs of the DNA molecule. The shape of this function can be explored by keeping one variable constant and letting the other vary. By fixing the value of n, we obtain a family of curves that give the free energy landscape of the system at different positions of the optical trap (see Fig. 3.16a). Each curve has an energy offset at $x_{\text{tot}} = 0$ nm as given by $G_{\text{DNA}}(n)$ and an elastic contribution that is monotonically increasing with x_{tot}. This specific shape makes that every curve crosses with all the others

and they alternate the minimum value of the energy as x_{tot} is increased. Therefore, there is not a value of n for which G_{tot} is minimum for all values of x_{tot}. Instead, there is an envelope curve that switches between all the functions that have minimum energy for each value of x_{tot}. Apart from them, there are several values of n for which their corresponding function is never a minimum, so they do not contribute to the envelope. They correspond to states with n open base pairs that are never observed during unzipping experiments because they have a high free energy. They might be detected in some cases if the difference between them and the states of minimum energy is a few k_BT, so thermal fluctuations make possible to explore those states.

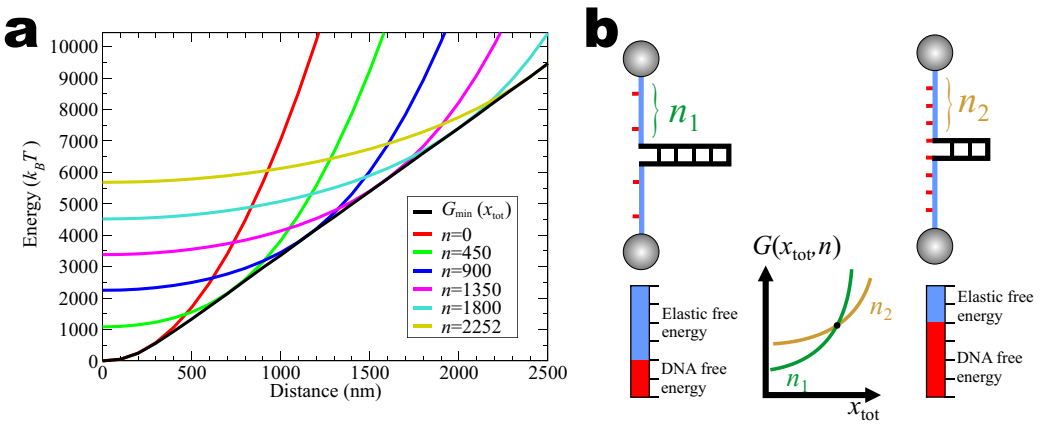


Figure 3.16: **(a)** Free energy landscape of a 2.2 kbp molecule at different values of fixed n. The black curve shows the minimum of energy for all values of x_{tot}. **(b)** Equilibrium of energies in a force rip. Left (right) picture shows the molecule with n_1 (n_2) open base-pairs. The total energy of the system in both states (n_1 and n_2) is equal but the internal energies (elastic and free energies) are distributed differently. The central graphic shows the total free energy of both states. The energies of these two states are identical when the force rip is produced (black dot).

Figure 3.16b shows the balance of the energetic terms when the DNA molecule is about to undergo a cooperative opening of base pairs. The system is originally at a state with n_1 open base pairs. When x_{tot} is increased, the elastic energy is also increased until the total energy of the system at n_1 is equal to the energy at n_2 ($n_1 < n_2$). This situation indicates the coexistence of the two states. Although the total energy of both states n_1 and n_2 is equal, the internal balance of the energetic terms (formation energy + elastic energy) is different. When x_{tot} is increased more, the state n_2 becomes more stable. Although the opening of base pairs increases the energy of the system, the released ssDNA relaxes the system and reduces the elastic energy. The global balance is that n_2 has a lower energy than n_1. So the unzipping at

constant position is governed by two energetic terms that are balanced by changing the number of open base pairs.

The free energy landscape can also be computed by fixing x_{tot} and exploring G_{tot} for different values of n (see Fig. 3.17a). At a given x_{tot}, the landscape has a parabolic shape when coarse grained. When the function is explored in deeper detail, the landscape is rough with a global minimum surrounded by other local minima that are separated by energy barriers (see Fig. 3.17b). Now, at a fixed x_{tot}, the system can jump between these minima by thermal fluctuations, provided that the energetic differences between them is of the order of the thermal bath energy. This phenomenon is known as hopping and it is observed in unzipping experiments.

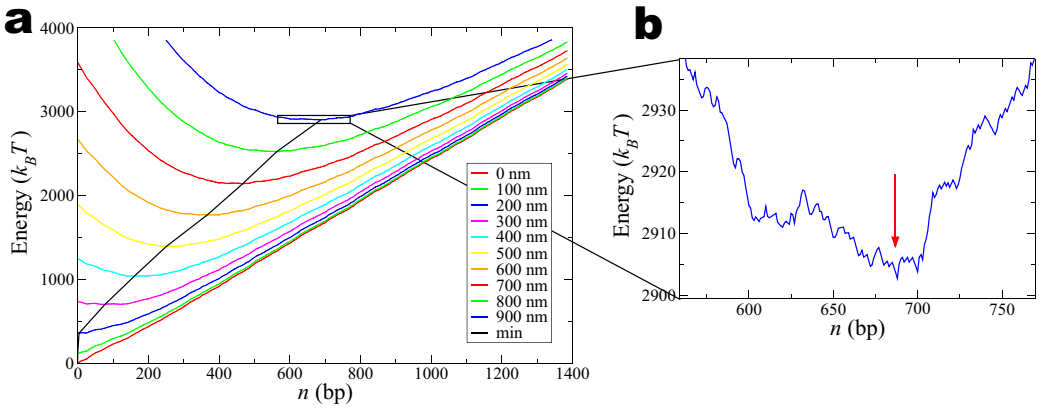


Figure 3.17: Free energy landscape of a 2.2 kbp molecule at fixed x_{tot}. **(a)** Coarse grained profile for different values of x_{tot}. At a fixed total distance, the energy landscape shows a minimum of energy, which is assumed to be the most probable states. **(b)** Detailed view where the roughness of the free energy landscape can be appreciated. The global minimum (pointed with an arrow) is surrounded by other local minima that may coexist with it.

Force vs. Distance Curve at $T = 0$

The FDC at $T = 0$ is the equation of state of the system when no thermal fluctuations are considered. The system is always assumed to be in the state of minimum energy (as in any other thermodynamic system studied at $T = 0$) and the dynamics of the system is deterministic.

For each given x_{tot} there is one (or two) values of n that minimize the total energy of the system $G_{\text{tot}}(x_{\text{tot}}, n)$. It defines a discontinuous function $n^*(x_{\text{tot}})$ that allows us to write an expression for the minimum energy of the system (G_{min}) at a given value of x_{tot} (see Fig. 3.18),

$$G_{\text{min}}(x_{\text{tot}}) = G_{\text{tot}}(x_{\text{tot}}, n^*(x_{\text{tot}})) \tag{3.24}$$

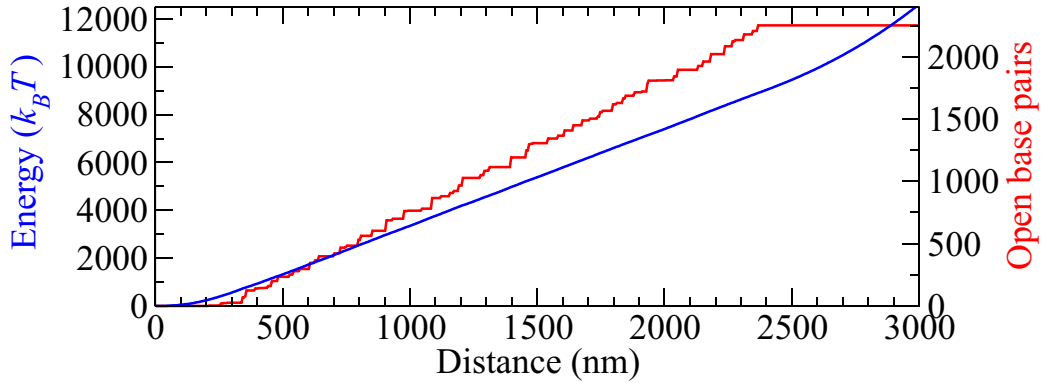


Figure 3.18: Blue curve shows the minimum free energy vs. x_{tot}, $G_{\text{min}}(x_{\text{tot}})$. Red curve shows the number of open base-pairs vs. x_{tot}, $n^*(x_{\text{tot}})$.

As mentioned before, G_{min} is the envelope function of the free energy landscape G_{tot} that minimizes its value. The FDC at $T = 0$ is obtained by computing the derivative of Eq. 3.24 according to,

$$f_m(x_{\text{tot}}) = \frac{\partial\, G_{\text{min}}(x_{\text{tot}})}{\partial x_{\text{tot}}}. \tag{3.25}$$

At each coexistence point, the derivative of the function G_{min} is not well defined because it switches between two functions from the energy landscape that are not smoothly connected. Therefore, the function $f_m(x_{\text{tot}})$ shows discontinuities at the coexistence points. Precisely, these points represent the force rips observed in the experimental FDC when an opening of base pairs is produced.

Partition function and equilibrium FDC

Starting from Eq. 3.22, the partition function of the system $Z(x_{\text{tot}})$ can be calculated by summing over all the possible states of the system ($0 < n < N$, where N is the total number of base pairs of the DNA) weighted with the Boltzmann factor

$$Z(x_{\text{tot}}) = \sum_{n=0}^{N} \exp\left(-\frac{G_{\text{tot}}(x_{\text{tot}}, n)}{k_B T}\right) . \tag{3.26}$$

From this expression, the equilibrium force vs. distance curve can be calculated according to

$$f(x_{\text{tot}}) = -k_B T \frac{\partial \ln Z(x_{\text{tot}})}{\partial x_{\text{tot}}} . \tag{3.27}$$

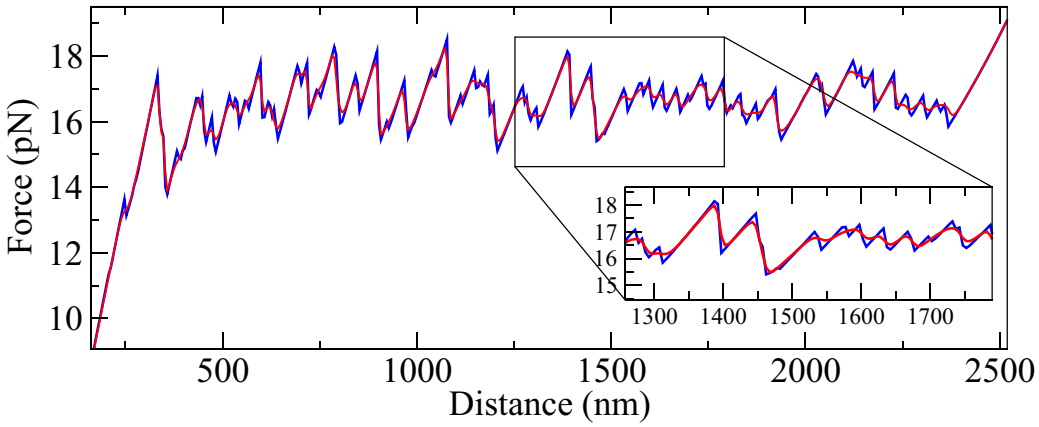


Figure 3.19: Blue curve, FDC at $T = 0$ according to Eq. 3.25. Red curve, equilibrium FDC according to Eq. 3.27. The inset shows a zoomed region where the smoothness of the equilibrium FDC contrasts with the abrupt FDC calculated at $T = 0$.

The resulting FDC calculated with Eq. 3.27 is very similar to the one calculated with Eq. 3.25. The main difference is that Eq. 3.27 is continuous and differentiable. The thermal fluctuations allow the system to explore states with more energy than the minimum so that the partition function $Z(x_{\text{tot}})$ is smooth.

The computation of the FDC for a given sequence is straightforward but it requires the use of numerical calculus. The exponential function in Eq. 3.26 has to be performed on values of G_{tot} that are larger and larger as x_{tot} increases. Appendix H shows how to deal with this issue.

3.4.2 Controlled force

In the controlled force ensemble, the force applied to the ends of the molecular construct is the control parameter and the total extension of the system is the magnitude that fluctuates. The natural thermodynamic potential of this ensemble is the Gibbs free energy, which is expressed in terms of the intensive variable (the force) and the temperature. The model is identical to the model described previously in the controlled position ensemble, but the expressions for the energies are different. Here we will follow the same description as in the previous section. The Lengendre transform converts one thermodynamic potential into the other. Using this transformation, the energetic contributions calculated at controlled position ensemble can be converted into the suitable expressions for the controlled force ensemble.

Let $G^X(x, T)$ be the free energy (i.e., the Helmholtz free energy) of an

elastic system defined by (x, f, T) in the control position ensemble (i.e., fixed position and temperature). The differential form of this energy is given by

$$dG^X(x,T) = f\,dx - S\,dT\ . \tag{3.28}$$

Now let $G^F(f,T)$ be the corresponding free energy (i.e., the Gibbs free energy) of the same system at the controlled force ensemble (i.e., fixed force and temperature). The relation between the two thermodynamic potentials is given by $G^F = G^X - fx$, which allows us to write

$$\begin{aligned} dG^F(f,T) &= dG^X(x,T) - d\,(fx) = f\,dx - S\,dT - f\,dx - x\,df \\ &= -S\,dT - x\,df\ . \end{aligned} \tag{3.29}$$

Now it is possible to calculate the free energy of an elastic element in the controlled force ensemble by evaluating the integral of Eq. 3.29 at constant temperature ($dT = 0$) according to,

$$G^F(f) = -\int_0^f x(f')\,df' \tag{3.30}$$

where G^F is the elastic free energy, f is the force at which the energy is evaluated, $x(f')$ is the extension vs. force curve and f' is the integration dummy variable. Eqs. 3.16, 3.19 and 3.21 are the (Helmholtz) elastic free energies (referred to as $G^X(x)$ in the new notation introduced in this section) of the bead, the handles and the ssDNA, respectively, in the controlled position ensemble. Now, we want to calculate the (Gibbs) free energy of each of them (i.e., $G^F(f)$) to proceed with the calculations in the controlled force ensemble, so we will apply Eq. 3.30 on all them. From now on, we will neglect the superscript F in G^F and we will understand that G must be read as the free energy in the controlled force ensemble.

The potential energy of the bead in the optical trap is given by

$$E_b(f) = -\frac{f^2}{2k}\ . \tag{3.31}$$

The calculation of the elastic energy of the handles involves the evaluation of the following integral $G_h(f) = -\int_0^f x_h(f')\,df'$ which has to be integrated by parts according to

$$G_h(f) = \int_0^{x_h} f_h(x')\,dx' - f x_h \tag{3.32}$$

where $f_h(x')$ is given by Eq. 3.18 and the integration limits are related by $f = f_h(x_h)$.

In the case of the ssDNA, the free energy is directly given by

$$G_s(f,n) = -\int_0^f x_s(f',n)\,df' \tag{3.33}$$

where $x_s(f',n)$ is given in Eq. 3.20. The total energy of the system at controlled force is given by the sum of the elastic contributions (Eqs. 3.31, 3.32, 3.33) and the formation energy of the DNA (Eq. 3.17) according to

$$G_{\rm tot}(f,n) = E_b(f) + 2G_h(f) + 2G_s(f,n) + G_{\rm DNA}(n) \tag{3.34}$$

and the total extension of the system is again given by Eq. 3.23

$$x_{\rm tot}(f,n) = x_b(f) + 2x_h(f) + 2x_s(f,n)\ . \tag{3.35}$$

From these two expressions it is possible to extract the properties of the DNA unzipping at controlled force as we did in the previous section with the controlled position ensemble.

Free energy landscape

Although Eqs. 3.22 and 3.34 have a similar functional form, the resulting expressions are different. The free energy landscape at controlled force can be calculated by fixing the value of f and giving values to n in Eq. 3.34 (see Fig. 3.20). Similar but slightly different aspects and details are emphasized by the free energy landscape of the system at controlled force. In this case, the force induces a net tilt in the free energy landscape. From this description, we can deduce that there is a coexistence force that induces no net tilt in the free energy landscape. Under this situation, there are a lot of minima along the molecule that coexist and may be observed. If the force is further increased, a net tilt appears, the states with higher number of broken bonds become more stable and the molecule starts to unzip. On the contrary, if the force is reduced respect to the coexistence value, the molecule is rezipped. Interestingly, the tilt can be adjusted by tunning the constant force applied on the molecule. So the evolution of the energy landscape as the force is increased is different from the evolution when the position of the trap is displaced. In the first case, the number of open of base pairs suddenly increases at the coexistence force because the stability of the minima changes abruptly. Precisely, this behavior is observed in the unzipping experiments at controlled force (see Chapter 6).

Another property of the free energy landscape calculated at fixed force is that at each given force there is always a global minimum separated by energetic barriers from other local minima. These energetic barriers can

be much higher than the thermal energy unit, k_BT. This implies that the system can be trapped in a metastable state for long periods of time before it is capable of finding the way to decay to the state of minimum energy.

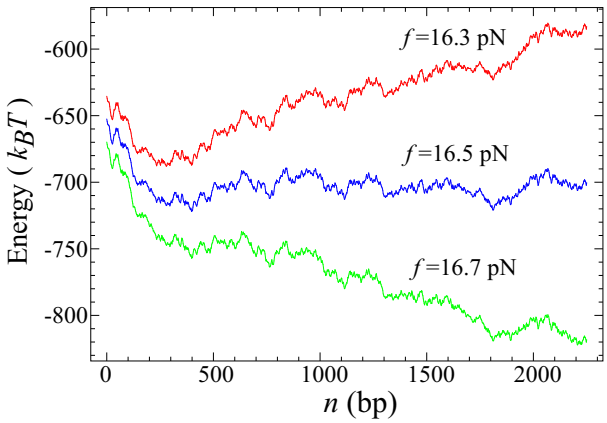


Figure 3.20: Free energy landscape of a 2.2 kbp molecule at controlled force for three different values of the force. There is a coexistence force ($f = 16.5$ pN) at which the net tilt of the landscape is zero. In this situation, many states with different number of open base-pairs coexist.

Distance vs. Force Curve (DFC) at $T = 0$

In the controlled force ensemble, the force is the independent variable and the position is the dependent one, so the equation of state is inverted as compared with the equation of state obtained in the controlled position ensemble. The calculation of the equation of state without considering the fluctuations requires to find the minimum of Eq. 3.34 with respect to n for each value of f according to

$$G_{\min}(f) = G_{\text{tot}}(f, n^*) = \min_n \left(G_{\text{tot}}(f, n)\right) \tag{3.36}$$

which allows to compute the DFC according to

$$x_{\text{tot}}(f) = -\frac{\partial G_{\min}(f)}{\partial f} \ . \tag{3.37}$$

This equation has discontinuities at the coexistence forces (see Fig. 3.21) but it does not show the force rips given by Eq. 3.25.

Partition function and equilibrium DFC

Again, assuming that all the elastic elements of the model are in mechanical and thermal equilibrium, the equilibrium force is the same for all them and

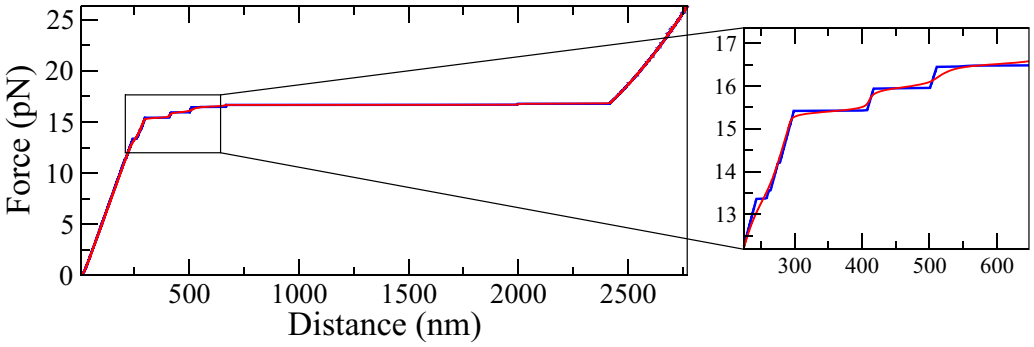


Figure 3.21: The axis in the DFCs have been inverted to look similar to the ones shown in Fig. 3.19. Blue curve, DFC at $T = 0$ according to Eq. 3.37. Red curve, equilibrium FDC according to Eq. 3.39. The inset shows a zoomed region where the smoothness of the equilibrium FDC contrasts with the abrupt FDC calculated at $T = 0$.

the extensions are univocally determined by the force. We can calculate the partition function $Z(f)$ by summing over all the intermediate states,

$$Z(f) = \sum_{n=0}^{N} \exp\left(-\frac{G_{\rm tot}(f,n)}{k_B T}\right) \tag{3.38}$$

and the resulting DFC is given by (see Fig. 3.21)

$$x_{\rm tot}(f) = k_B T \frac{\partial \ln Z(f)}{\partial f}. \tag{3.39}$$

3.5 Conclusions

The DNA is probably the most relevant biomolecule in living systems. Its structure allows it to carry out the tasks devoted to the storage, copy and transmission of the genetic information *hardcopied* into the base-pairs. In order to have access to the sequence, the two strands of DNA have to be split apart, in a process called *unzipping*. Cells have a specific machinery to produce the unzipping of DNA, whenever it is necessary. Nevertheless, the unzipping of DNA can be produced artificially by using single-molecule techniques such as optical tweezers. These experiments provide information about the thermodynamics and the kinetics of DNA duplex formation.

The unzipping of DNA using optical tweezers can be easily reproduced thanks to the great advance in molecular biology tools. DNA molecules can be synthesized with the required features in order to be pulled with optical tweezers (handles, loops, etc.).

The FDC is the characteristic measurement obtained with optical tweezers. Since the FDC is sequence-dependent, the FDC is a fingerprint of the DNA molecule. The FDC can be obtained with controlled position or controlled force protocols, which induce a different behavior in the opening of base-pairs. At controlled position, the NN model combined with a mesoscopic description of the other elements of the experiment (optical trap, handles, etc.) is able to accurately describe the quasistatic unzipping of DNA. Such mesoscopic model is quite versatile offering new insights on the physical phenomenon of unzipping. One interesting magnitude that can be calculated is the free energy landscape, which allows us to illustrate the unzipping mechanism.

Next chapters use the model and the methodology explained here to extract relevant information from DNA unzipping experiments. The results shown in those chapters reinforce the statements made here.

When individuals join in a cooperative venture, the power generated far exceeds what they could have accomplished acting individually.

Richard Buckminster Fuller (1895-1983)

Chapter 4

Metastable intermediates in DNA unzipping

From a pure physical point of view, the unzipping of DNA can be regarded as a process of fracture, in which the strength applied to the two strands produces the disruption of the base-pairs [124]. The FDC of unzipping has similar properties to the Barkhausen noise observed in the magnetization of ferromagnetic domains, paper tearing or earthquakes. The characteristic fingerprint in all of these cases is the sawtooth-like signal. The slopes of the sawtooth pattern correspond to the accumulation of energy (elasticity), while the drops in the signal are related to the release of energy (breakage). Bonnott and coworkers [125] have measured stress vs. strain curves on martensitic materials that look very similar to the FDC of DNA unzipping, showing that both phenomena exhibit the typical properties of fracture. Similarly, sawtooth-like FDCs have been observed in pulling experiments of gold nanowires with scanning tunneling microscopy [126]. In one of his articles about DNA unzipping with microneedles, Bockelmann *et al.* already identified the stick-slip motion (i.e., slopes and drops) of the unzipping fork, which is a characteristic signature of the cracking effects.

The unzipping of DNA can be studied globally, considering the succession of opening base-pairs as part of a whole process of fracture. The detailed information about the local effects is neglected and the relevant conclusions focus on the statistical properties of the intermediate states. In this approach, the DNA molecule is regarded as a disordered statistical system due to the

randomness of the base-pair sequence. A theoretical work by Lubensky and Nelson [29] explored the statistical properties of DNA unzipping at constant force, based on the experimental results of their collaborators [23]. One of the main results of that paper was the calculation of the size of the opening fork as the mean unzipping force is approached.

This chapter focuses on the statistical properties of the metastable intermediate states observed in DNA unzipping experiments at controlled position. The study allows us to deepen into the intrinsic mechanism that induces the opening of base pairs. Besides, the analysis of the experimental data is accompanied with the development of a toy model that contains the essential ingredients needed to reproduce the unzipping process. It is useful to establish what are the experimental conditions that permit to break up the metastable states that are observed during the unzipping into single base-pair events. Eventually, this might be useful to set the basis for sequencing DNA by force, a potentially interesting application in the field of biotechnology. A limiting factor in unzipping is the accuracy at which individual base pairs along the DNA can be resolved. Indeed, the unzipping process, even if carried out reversibly (i.e., infinitely slowly), shows a progression of cooperative unzipping/rezipping transitions that involve groups of base pairs of different sizes. These Cooperatively Unzipping Regions (CURs) of base pairs breath in an all-or-none fashion hindering details about the individual base pairs participating in such transitions. Here it is shown a Bayesian technique useful to extract as much information as possible from the noisy experimental data.

The treatment of the experimental data that exhibits CURs is complemented with a toy model. This model captures the very essential ingredients that reproduce the unzipping FDCs and the properties of the CURs. The advantage of such model with respect to the mesoscopic model introduced in the previous chapter is that with the toy model it is easier and faster to extract statistical properties of CURs. Indeed, it is computationally less demanding and the simulations of unzipping experiments can be extended to thousands of different sequences. This gives valuable information about the dependence of the CURs on the experimental conditions (trap stiffness, base-pair energies, etc.).

To sum up, here we will treat the DNA unzipping of long molecules as a whole phenomenon.

4.1 Detection of intermediate states

As described in the previous chapter, the quasistatic unzipping of DNA at controlled position is performed by moving the center of the optical trap at a very low speed (∼10 nm/s). Along these experiments, the dsDNA is progressively converted into ssDNA through a succession of intermediate states corresponding to the successive opening of the CURs. In these intermediate states, the DNA molecule is partially open. The molecule remains in this state if the distance between the two beads is not modified. If this distance is being increased, we can measure the FDC. The experimentally measured FDC shows a sawtooth-like pattern that alternates force rips and gentle slopes (see Fig. 4.1a). Slopes correspond to the elastic response of the molecule while the force rips correspond to the opening of groups of base-pairs. These openings are called CURs (see Fig. 4.1b).

The number of open base-pairs of a CUR is variable: It ranges from 5-10 base-pairs up to 80-100 and it depends on the sequence of the DNA molecule and the specific region that is being unzipped. The size of a CUR can be inferred from the difference of slopes that precede and follow a given force rip. However, the slopes can be hardly isolated because the experimental FDC exhibits noise (see Fig. 4.1b). The force fluctuations of the elements of the system (bead, handles) might be confused with the different forces at which the two intermediate states of the DNA duplex coexist. A classification by hand of the experimental points might lead to a biased interpretation of slopes and rips due to the human intervention. In order to avoid this, we adopt a Bayesian approach.

A Bayesian approach consists in using the probability inferred from the experimental data to accept or reject a hypothesis. In our particular case, the experimental data is first classified into intermediate states. Then, the most stable of them are identified as metastable states. Finally, the regions between each pair of consecutive metastable states are identified as CURs. Once this is achieved, the unzipping of DNA can be studied in terms of a new statistical object: the CUR size distribution.

Next section describes how to extract the experimental metastable states according to the Bayesian inference.

4.1.1 Bayesian analysis of FDCs

The classification of the experimental data into intermediate states is based on the elastic response of the system (the molecular construct plus the bead in the optical trap), which is given by the mesoscopic model described in Section 3.4. So each experimental data point (distance,force) is associated

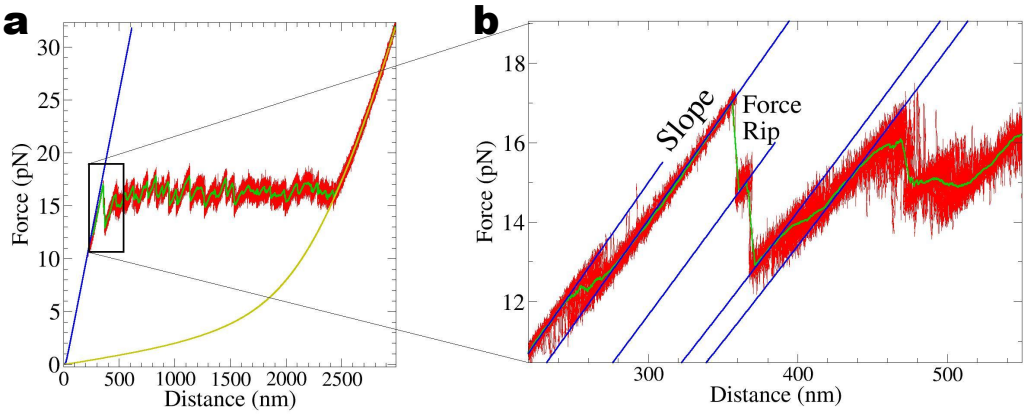


Figure 4.1: Intermediate states in experimental data. **(a)** Unzipping of a 2.2 kbp sequence. The raw experimental data is depicted in red and filtered data at 1 Hz bandwidth, in green. The yellow (blue) curve shows the elastic response of the system when all (no) base pairs are open. **(b)** Zoomed region. Same color code as in panel a. From left to right, the blue curves correspond to the elastic response of intermediate states with 0,19,70,123,142 open base pairs. Except for the force fluctuations, the slopes of the experimental data lie on the partial elastic responses of the system. Note that the number of open base pairs of the system is no longer clear for distances larger than 450 nm, where the force fluctuations are as large as the difference in force between the slopes of the coexisting intermediate states. Filtering data with a low-pass filter could be a solution. However, the force rips are smoothed and the hopping transitions are averaged out.

to the most probable intermediate state that is compatible with the elastic properties of the system. This way, each point belongs to a state.

For convenience, here we reproduce the expression (Eq. 3.23) that determines the elastic response of system when the DNA molecule is at the intermediate state I_n where n bases are open,

$$x_{\text{tot}}(f,n) = x_b(f) + 2x_h(f) + 2x_s(f,n) \tag{4.1}$$

where x_{tot} is the total distance of the system, x_b is the position of the bead, x_h is the extension of one handle and x_s is the extension of one strand of ssDNA. All these terms depend on the force applied f at the ends of the system. Note that this equation is a FDC defined for the intermediate state I_n.

Now, we will use Eq. 4.1 differently from Sec. 3.4. Here we consider the partial elastic response of a system with fixed number of open base pairs, so that the energetic contribution of the base pairs is irrelevant. Therefore, Eq. 4.1 can be understood as the expression of a family of curves passing through the coordinates origin and characterized by a parameter n. In unzipping experiments, the number of open base pairs varies while the rest of the properties of the system remain unchanged (trap stiffness, elasticity of

handles). So, as the molecule is being unzipped, we observe fragments of these curves (slopes) connected by force rips. In other words, an unzipping FDC is a piecewise-defined function of Eq. 4.1.

Since n is what changes the elastic properties of the system, it can be used to classify the experimental points into states. For each experimental data point along the FDC, (x, f), the intermediate state I_{n^*} that passes closest to that point for a fixed force f is determined by

$$|x - x_{\text{tot}}(n^*, f)| = \min_n \left(|x - x_{\text{tot}}(n, f)|\right) \tag{4.2}$$

where $x_{\text{tot}}(n, f)$ is given by Eq. 4.1. The function min ensures that n^* is the state that has the FDC that passes closest to the experimental point. In this way, each experimental data point (x, f) is associated to a unique value of n^* (see Fig. 4.2). Note that the experimental points have force fluctuations but Eq. 4.2 does not. It means that some points are incorrectly classified because force fluctuations are confused with different values of n. However, the overall outcome of the Bayesian approach shows its usefulness after all the experimental points are classified.

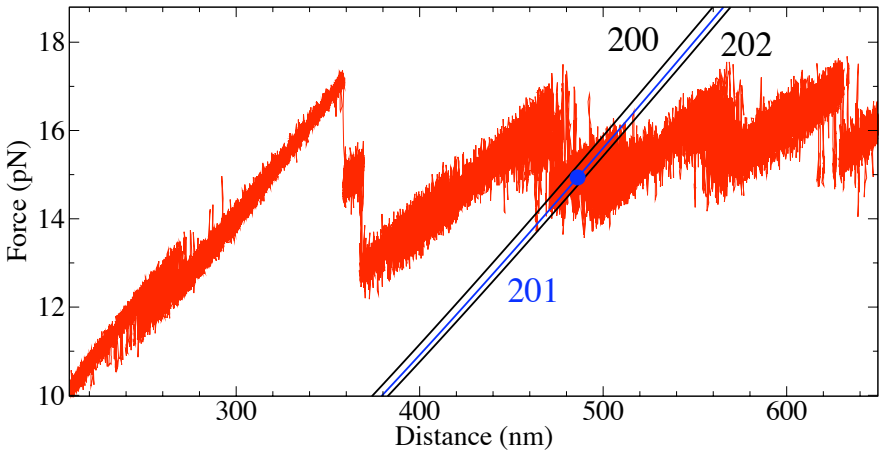


Figure 4.2: Classification of one experimental point (picture not to scale). For one experimental point (blue dot) of the unzipping data (red trace), we scan the elastic response of the system for different values of n (black curves) and select the one that passes close to it (blue curve). In this particular case, the experimental point is associated with an intermediate state of $n = 201$ open base pairs.

The use of Eqs. 4.1 and 4.2 to classify the experimental points requires the determination of some elastic parameters from the experimental data. Most of them are obtained by fitting the elastic response of the fully extended molecule (see yellow curve in Fig. 4.1a) to Eq. 4.1, because we know that the

number of open base pairs is $n = 2252$ in this part of the FDC. The stiffness of the trap $k \simeq$ 60-80 pN$\cdot\mu$m^{-1} is determined by performing measurements of the force vs. displacement when a bead is held fixed at the tip of the micropipette (see Fig. 2.15d). The parameters of the handles are taken from the established elastic properties of long dsDNA molecule [48]: $l_p = 50$ nm and $L_0 = m \times d = 9.86$ nm, where $m = 29$ is the number of base pairs of the handles and $d = 0.34$ nm is the interphosphate distance. Finally, the elastic parameters of the ssDNA are obtained by least-square non-linear fitting. The Kuhn length of the ssDNA is found to be $b \simeq 1.20$ nm (see Table 5.1). Among all the elements of the experimental setup, the dsDNA handles are those that do not modify appreciably the FDCs due to their large rigidity. Since the measurement of distances in the instrument is relative, it is also necessary to fit a global shift in Eq. 4.1 that determines the zero of distance.

4.1.2 Statistical inference of intermediate states

The association of each experimental point with its most probable intermediate state is repeated for all the points along the unzipping FDC (see Fig. 4.3a). The result is a list of numbers $\{n^*\}$ that indicate all the metastable states through which the molecule passes during the process of unzipping.

A histogram built from all values n^* results in a series of sharp peaks that can be identified with the many intermediate states I_n (see Fig. 4.3b). The histogram contains information about the stability of the intermediate states: the higher the peak, the higher the stability of that state. It has already been shown that the more stable states are related with higher GC content in the sequence [102].

The histogram can be fit to a sum of Gaussians each one characterized by its mean, variance and statistical weight (see Fig. 4.3c). The mean of each Gaussian indicates the number of open base pairs of that intermediate state. Other conditions such as the released ssDNA and the stability of the state determine the variance and the weight of the Gaussian. In general, a Gaussian distribution is sufficient to fit one peak. However, some peaks require the contribution of two or more Gaussians to be correctly fit. So this method allows to distinguish intermediate states that have similar number of open base pairs (about $\simeq$ 5-10 base pairs at the beginning of the unzipping).

Note that as the unzipping goes on, the peaks of the histogram look smoother and it is more difficult to differentiate intermediate states. This is due to the release of ssDNA as the DNA molecule is being unzipped. Indeed, the changes in the unzipping fork must be transmitted to the optical trap in order to be detected. So the opening of a CUR decreases the tension along the molecular construct and the optical trap detects a drop of force. A stiff tether

transmits the force towards the optical trap better than a compliant one. The reason is that a compliant connection can absorb the elastic energy released in the unzipping without hardly changing its extension. Therefore, as the ssDNA is released during unzipping, the amplitude of the sawtooth pattern decreases, the force signal is blurred and the histogram of intermediate states shows smoother peaks.

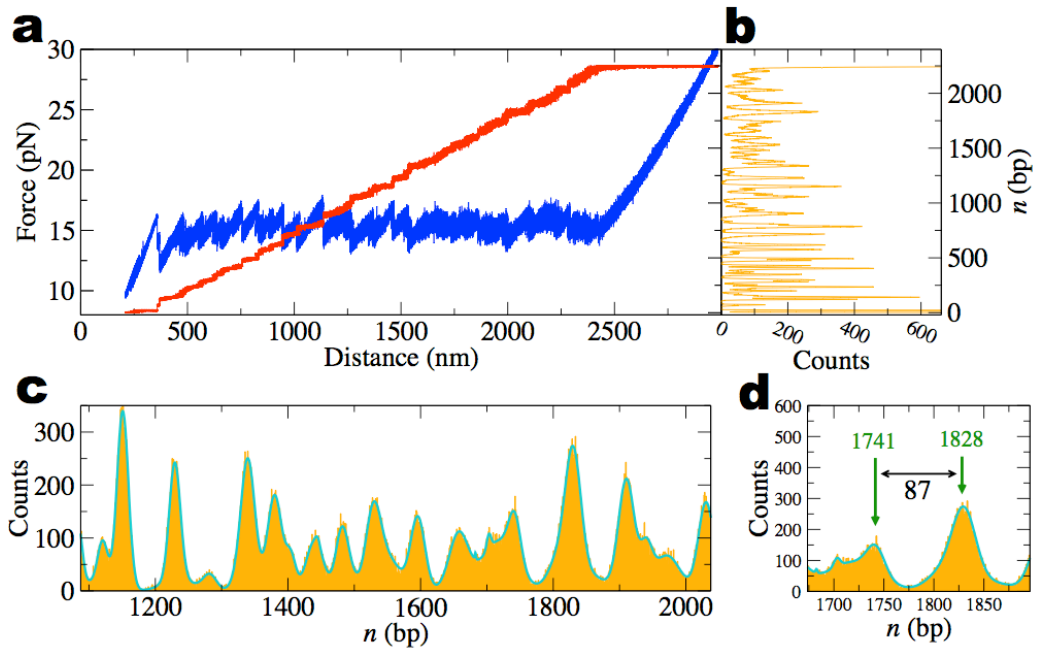


Figure 4.3: Histogram of intermediate states. **(a)** Classification of points. Blue trace shows the experimental FDC. Red trace shows the number of open base pairs n^* corresponding to each experimental data point (y-axis of this curve is shown in panel b). Although the curve is noisy, the plateaus indicate metastable states. **(b)** Histogram of the values for n^* shown in panel a. **(c)** Detailed view of the histogram (orange curve) overlapped with the fit to a sum of Gaussians (cyan curve). **(d)** Detection of a CUR of size 87 bp from the distance between two consecutive Gaussians centered at 1741 and 1828 open base-pairs.

4.1.3 CUR size distribution

The CUR size distribution is the function that relates the size of one CUR with the probability of being observed during the unzipping experiment. The size of all the CURs observed in the unzipping of one molecule is obtained by calculating the difference of the means (in base pairs) between consecutive Gaussians (see Fig. 4.3d). The histogram of all the CUR sizes gives the experimental distribution of CUR sizes (see Fig. 4.4).

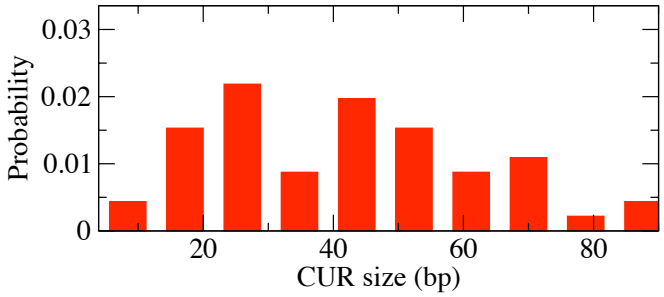


Figure 4.4: Experimental CUR size distribution for the 2.2 kbp sequence. Sizes range from a few base pairs up to 90 bp with a maximum number of detected CUR sizes between 20 and 50 bp.

4.1.4 Results

The previously method of analysis has been applied to several molecules of 6.8 and 2.2 kbp. In particular, the FDC of 6 molecules of each length have been measured and analyzed. Figure 4.5 shows complete FDC of some of them. Since the unzipping curves are obtained at very low pulling rates, the process is quasistatic and the FDCs resemble each other. This is quite relevant because it indicates that the position and the size of the CURs are reproducible and they only depend on the sequence, emphasizing the reliability of the approach.

The histograms of metastable states are also very similar among them because they are obtained from FDCs that resemble each other. Figure 4.6 shows some of the obtained histograms for the 6.8 kbp sequence. Finally, Fig. 4.7 shows the average histograms and the Gaussian fits for the 2.2 and the 6.8 kbp sequences.

The resulting distribution of CUR sizes will be discussed in Sec. 4.3 together with the theoretically predicted histograms.

4.2 Toy model

The cooperative dynamics of DNA unzipping at low pulling regimes has been described as a stick-slip motion elsewhere [24]. The opening of base pairs during the unzipping process is produced through a series of avalanches between stable intermediate states that have different number of open base pairs. We want to describe the most relevant properties of avalanche dynamics in terms of a simple model that reproduces the statistical properties of DNA unzipping. We have developed a toy model, simpler than the one described in section 3.4, that captures the essential behavior of DNA unzipping.

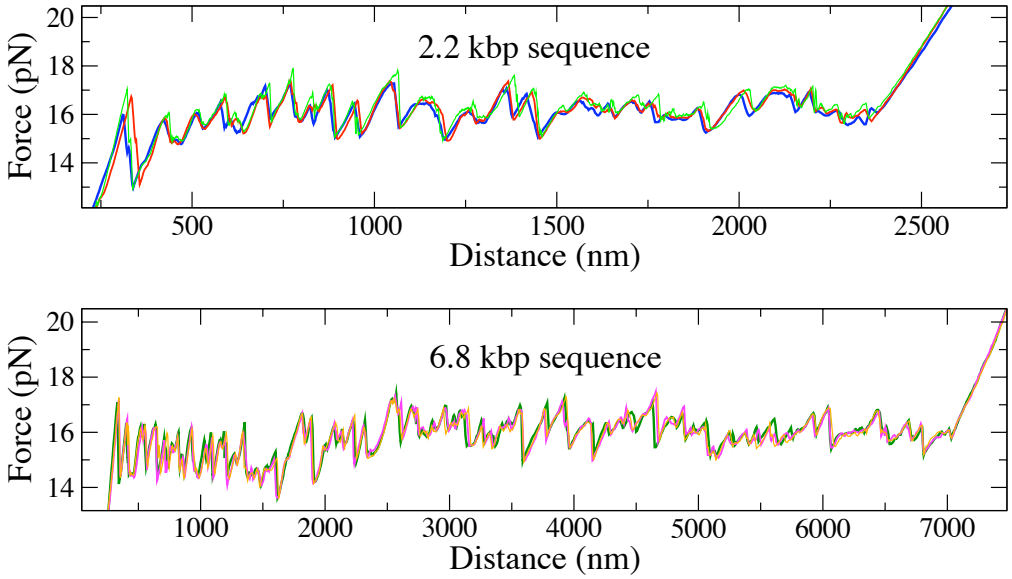


Figure 4.5: Experimental FDC of 3 molecules corresponding to the 2.2 kbp sequence (upper panel) and the 6.8 kbp sequence (lower panel). Since the raw data is too noisy, the data has been filtered with a low-pass running-average filter with a bandwidth of 1 Hz to clearly see the traces.

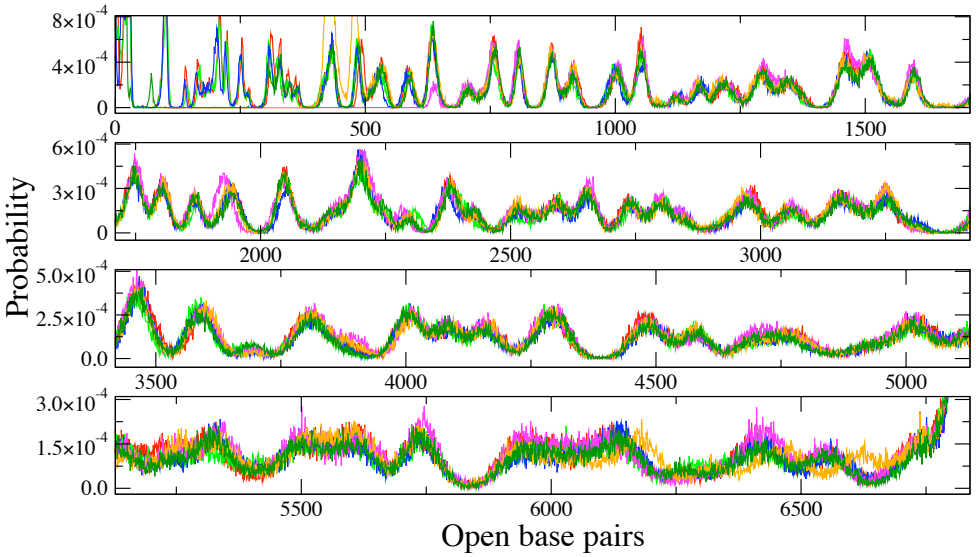


Figure 4.6: Histograms of intermediate states for six different molecules of 6.8 kbp. They are depicted in red, green, blue, magenta, orange and dark green. Although the height of each peak is different for the six histograms, the position of the peak is almost the same ($\pm$10 bp). The histograms for the 2.2 kbp molecules have similar reproducibility.

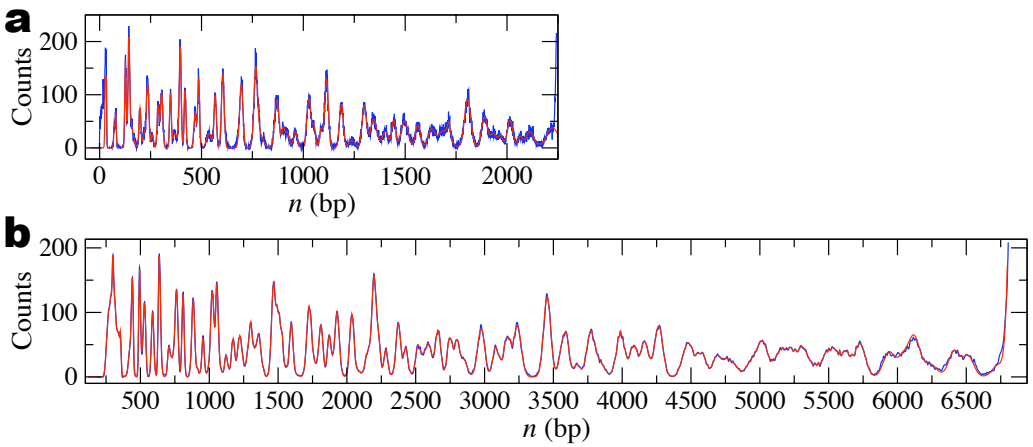


Figure 4.7: CUR size distributions. **(a)** Histogram of the intermediate states for the 2.2 kbp DNA sequence. Blue curve shows the experimentally measured histogram. Red curve shows the fit to a sum of Gaussians. **(b)** Histogram of intermediate states for the 6.8 kbp sequence. Same color code as in panel a.

In this toy model (see Fig. 4.8a), the energy of bead in the optical trap is also assumed to be quadratic (see Eq. 3.16). However, no elastic contributions of the molecular construct are taken into account and only the free energy of formation of the DNA duplex is considered (see Eq. 3.17). Summing the two energy contributions of the model we end up with:

$$E(x_b, n) = \frac{1}{2}kx_b^2 - \sum_{i=1}^{n} \epsilon_i \tag{4.3}$$

where $E(x_b, n)$ is the total energy of the system, k is the stiffness of the optical trap, x_b is the position of the bead in the optical trap and ϵ_i are the NN energies of the DNA hairpin. In the toy model, the energy of the DNA is no longer given by a specific sequence. Instead, we will consider that the NNBP energies are Gaussian distributed $\mathcal{N}(\mu, \sigma)$, with mean $\mu = \langle \epsilon_i \rangle$ and standard deviation $\sigma = \sigma_{\epsilon_i}$. Since the released ssDNA is taken as inextensible, its extension (x_s) is given by $x_s = 2dn$, where d is the interphosphate distance, n is the number of open bps and the factor 2 stands for the two strands of ssDNA. By using the relation $x_b = x_{\text{tot}} - 2dn$, the total energy of the system can be rewritten as:

$$E(x_{\text{tot}}, n) = \frac{1}{2}k(x_{\text{tot}} - 2dn)^2 - \sum_{i}^{n} \epsilon_i \ . \tag{4.4}$$

At fixed x_{tot}, the system will tend to occupy the state (n^*) that minimizes the total energy of the system, i.e., $E(x_{\text{tot}}, n^*) \leq E(x_{\text{tot}}, n), \forall n$. Therefore, it is

possible to define the function $n^*(x_{\rm tot})$ (see Fig. 4.8c). Combining this function with Eq. 4.4 we can obtain $E(x_{\rm tot}, n^*(x_{\rm tot})) = E_m(x_{\rm tot})$ (see Fig. 4.8b), which gives the FDC (Fig. 4.8d) according to,

$$f(x_{\rm tot}) = \frac{\partial E_m(x_{\rm tot})}{\partial x_{\rm tot}} \ . \tag{4.5}$$

The final shape of the FDC depends on the values given to ϵ_i, $(i = 0, N)$ in Eq. 4.4. Therefore, for each realization of $\{\epsilon_i\}$ there is a univocal FDC. The FDCs obtained from this model reproduces the formal sawtooth pattern that is experimentally observed. Moreover, the opening of base pairs is discontinuous and the size of the opening is not constant, but shows a distribution of sizes.

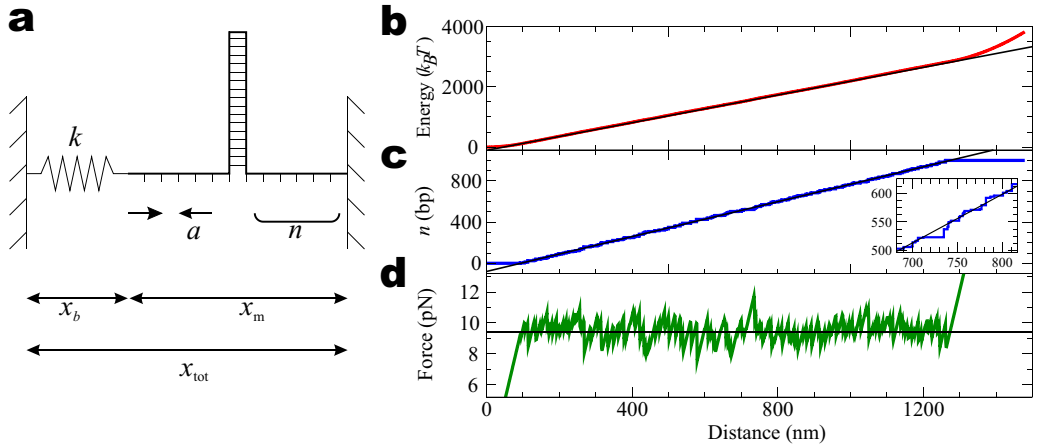


Figure 4.8: Toy model. **(a)** Scheme. **(b)** Minimum energy vs. distance, $E_m(x_{\rm tot})$, for one realization (red curve) of a 10^3 base-pair long sequence. The following parameters have been used: $k = 60$ pN·μm^{-1}, $d = 0.59$ nm, $\mu = -1.6$ kcal·mol^{-1}, $\sigma = 3.20$ kcal·mol^{-1}. **(c)** Open base pairs vs. distance, $n^*(x_{\rm tot})$, depicted in blue. Inset shows a detailed view of the step function. **(d)** FDC, $f(x_{\rm tot})$, depicted in green. The average behavior of the system is shown in black in panels b,c,d (see Sec. 4.2.1).

4.2.1 Approximate solution

If the disorder of the DNA energies (ϵ_i) is neglected, Eq. 4.4 can be rewritten as

$$E(x_{\rm tot}, n) \simeq \frac{1}{2}k(x_{\rm tot} - 2dn)^2 - \mu n \tag{4.6}$$

where $\mu = \langle \epsilon_i \rangle$. By minimizing this expression with respect to $x_{\rm tot}$, we immediately get the following result for the number of open base pairs:

$$n^*(x_{\rm tot}) = \frac{1}{2d}\left(x_{\rm tot} + \frac{\mu}{2dk}\right) \tag{4.7}$$

which allows to express the minimum energy of the system as

$$E_m(x_{\rm tot}) = -\frac{\mu}{2d}\left(x_{\rm tot} + \frac{\mu}{4kd}\right) \tag{4.8}$$

and obtain the FDC after calculating the derivative

$$f = -\frac{\mu}{2d} \tag{4.9}$$

The three previous expressions capture the dependence of the averaged number of open bps, energy and force on the external parameters. The solutions to this approximation are smooth expressions that collect the average behavior of the system over an ensemble of sequences (i.e., realizations of the disorder). Indeed, the bases open in a continuous fashion, the energy is linear with $x_{\rm tot}$ and the unzipping force is constant. The black curves in Figs. 4.8b,c,d show the approximated solution superimposed on one disorder realization. Interestingly enough, the toy model predicts a plateau of force whose value is only determined by formation energy of the base pairing (μ) and the properties of the ssDNA (d), and it is independent of the number of bases of the DNA (N).

4.2.2 Height of the force rips

As seen in Eq. 4.9, there is a mean unzipping force defined by the mean value of the NNBP energies. However, a single realization of the model shows slopes and force rips around the mean average force (see Fig. 4.8d), which reflect the discontinuous opening of base pairs. The height of the force rips (i.e., the amplitudes of the sawtooth pattern) depend on the standard deviation σ_{ϵ_i} of the energies and on the trap stiffness k (see Fig. 4.9). An interesting result is the fact that the stiffness of the trap directly affects the opening of base pairs. Therefore, the resolution in base pairs that can be observed during unzipping is partially determined by the stiffness. This feature of the model is quite relevant and it can be used to establish the experimental conditions that would allow to sequence the DNA through unzipping experiments.

4.2.3 CUR size distribution

Equation 4.7 is useful to describe the average number of open base pairs. However, for each realization of the energies, ϵ_i, the function $n^*(x_{tot})$ is a discontinuous function and has to be numerically calculated (see Fig. 4.10a black curve). Each discontinuity represents an opening of base pairs, i.e., a CUR. The size of a CUR is the difference of the number of open base pairs

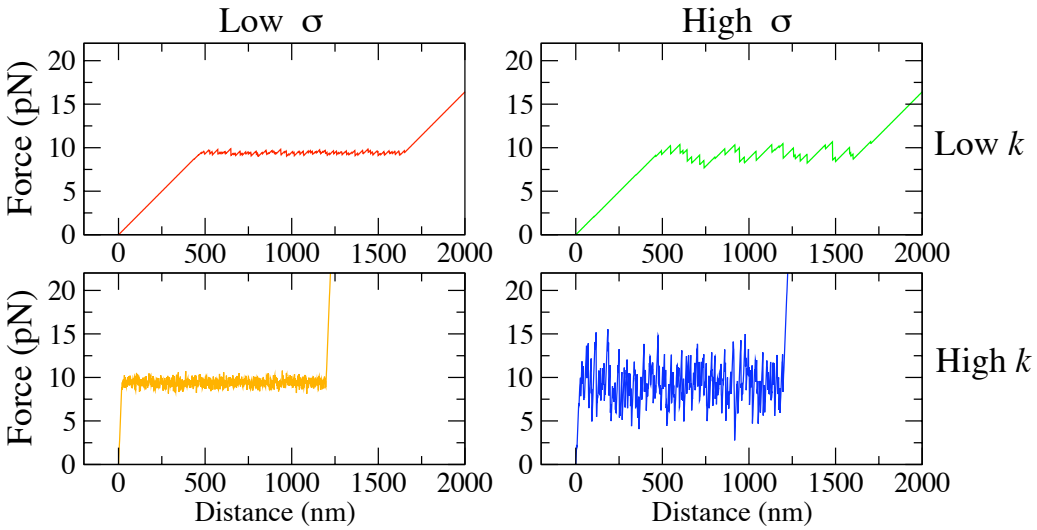


Figure 4.9: FDC for different values of the experimental parameters. The heights of the force rips depend on the standard deviation σ_{ϵ_i} and on the trap stiffness k. Low $\sigma_{\epsilon_i} = 0.1$ kcal·mol^{-1}, high $\sigma_{\epsilon_i} = 1.0$ kcal·mol^{-1}, low $k = 0.02$ pN·nm^{-1} and high $k = 0.5$ pN·nm^{1}. Note that the mean unzipping force is independent of σ_{ϵ_i} and k.

between two states. This way, we can calculate the size of the CUR from the $n^*(x_{tot})$ function by extracting the step size of the discontinuities (Fig. 4.10a, red curve). A distribution of CUR sizes can be obtained for one realization (Fig. 4.10b, blue curve, inset).

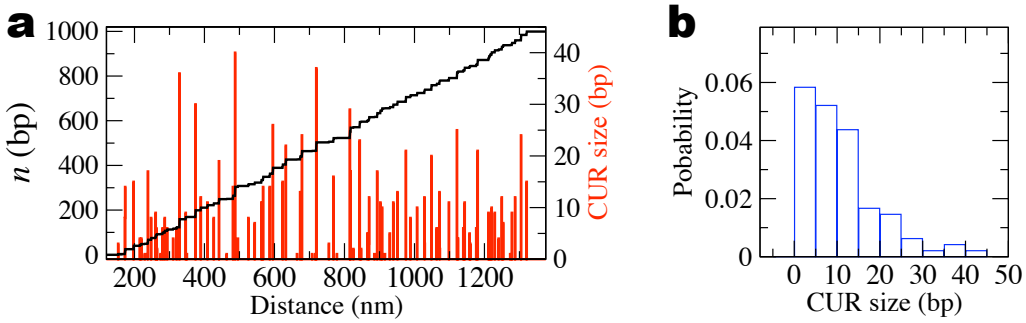


Figure 4.10: CUR size distribution. **(a)** Black curve shows the number of open base pairs for one sequence of 10^3 base pairs ($k = 0.06$ pN·nm^{-1}, $a = 0.59$ nm, $\mu = -1.6$ kcal·mol^{-1} and σ_{ϵ_i} =0.44 kcal·mol^{-1}). Red curve shows the size of the CURs. **(b)** CUR size distribution.

The average distribution of CUR sizes can be obtained by simulating random realizations, calculating the CUR size distribution for each realization of the disorder and averaging over them. By varying the parameters of the

model along a wide range we observe how the shape of the CUR size distribution depends on them. Generally speaking, we have observed that the shape of the CUR size distribution is independent of the mean NNBP energy (μ). It mostly depends on the standard deviation of the NNBP energies σ_{ϵ_i} and on the trap stiffness k. Figure 4.11 shows these dependencies for a model with the following parameters: $d = 0.59$ nm, $\mu = -1.6$ kcal·mol^{-1} simulated for a 10^4 bp long sequence over 10^4 realizations of the disorder.

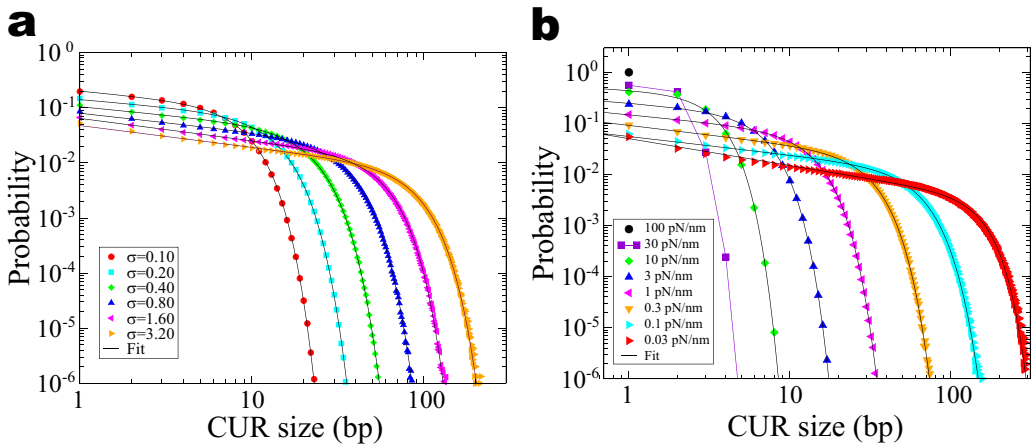


Figure 4.11: Average CUR size distributions plotted in log-log scale. **(a)** For different values of σ_{ϵ_i} at fixed $k = 60$ pN·μm^{-1}. **(b)** For different values of k at fixed $\sigma_{\epsilon_i} = 3.20$ kcal·mol^{-1}. The black curves show the fits of all the distributions to Eq. 4.10.

4.2.4 Fit of CUR size distributions

The analytical solution for the CUR size distributions is a problem that has not been solved yet. Although this calculation can be performed in the unzipping at controlled force [29], it is not that straightforward at controlled position. The computation requires extra efforts and it is not the primary goal of this section. Instead, an empirical expression for the CUR size distribution is provided in order to capture the qualitative dependence of the distribution with the tunable parameters of the model.

The CUR size distributions are well fit by an expression both in linear and log-log scale. It is a power law with a super-exponential cutoff:

$$P(n) = An^{-B} \exp\left(-(n/n_c)^C\right) \tag{4.10}$$

where $P(n)$ is the probability of observing a CUR of size n; A, B, C and n_c (cutoff size) are positive fitting parameters. All the distributions obtained

from the simulations with the toy model are fit to Eq. 4.10 (see black curves in Fig. 4.11).

Figure 4.12 shows the dependence of the fit parameters with σ_{ϵ_i} and k. The cutoff parameter n_c gives an idea of the maximum size of the CURs. It has a physical meaning and can be understood as the parameter that indicates the width of the CUR size distribution: the higher the value of n_c the wider the distribution. As expected, the value of n_c increases with the amount of disorder σ_{ϵ_i} (see Fig. 4.12a, panel n_c) because the NN energies are more dispersed and the intermediate states show more variety of open base pairs. However, what is interesting from the simulations is that the value of n_c can be reduced at will by increasing k. In fact, there is a limit in which n_c collapses down to 1 when $k \geq 100$ pN·nm^{-1} (see Fig. 4.12b, panel n_c), indicating that all CUR sizes are of size 1. Under this circumstances, the mechanical unzipping of DNA is done one base-pair at a time and establishes and interesting limit for the experimental conditions that would allow to sequence DNA by force.

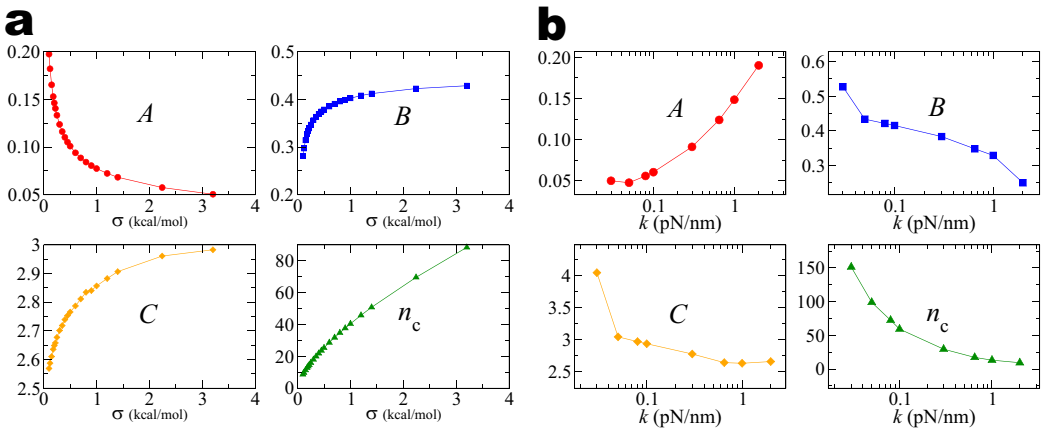


Figure 4.12: Fit parameters vs. toy model parameters. **(a)** Dependence on σ_{ϵ_i} at $k = 60$ pN·μm^{-1}. **(b)** Dependence on k at $\sigma_{\epsilon_i} = 3.20$ kcal·mol^{-1}.

4.2.5 Variability of CUR size distributions

Up to now, we have only explored the average CUR size distributions of the toy model and nothing has been said about their variability for different realizations. Although this variability should vanish for infinitely long sequences (CUR size distributions are self-averaging in the thermodynamic limit) there are large fluctuations for finite length molecules. This variability can be determined from the same simulations of the toy model that have been used in the previous sections. Now, for all the realizations we do not only calculate

the average CUR size distribution but the standard deviation of the distribution at each value of n. So if we have a collection of distributions obtained from n_r realizations, $\{P_i(n)\}$ i=1,...,n_r, we can compute the average distribution $\langle P(n)\rangle$ and the standard deviation of the distribution $\sigma_{P(n)}$ according to (see Fig. 4.13),

$$\begin{aligned}\langle P(n)\rangle &= \frac{1}{n_r}\sum_{i=1}^{n_r} P_i(n)\\ \langle P(n)^2\rangle &= \frac{1}{n_r}\sum_{i=1}^{n_r} P_i(n)^2\\ \sigma_{P(n)} &= \sqrt{\langle P(n)^2\rangle - \langle P(n)\rangle^2}\ .\end{aligned} \tag{4.11}$$

4.3 Comparison of CUR size distributions and discussion

At this point, we want to compare the CUR size distributions obtained from the unzipping experiments with the ones obtained with the toy model. The goal is to see if the toy model is capable of reproducing the statistical properties of the unzipping mechanism, search the causes of the differences and use the model to determine the best experimental conditions to extract information from the DNA unzipping.

How much can the toy model predict the experimental results? The experimentally obtained CUR size distributions for both molecular constructs are shown in red in Fig. 4.14. The fit of these distributions to Eq. 4.10 are also shown in green in Fig. 4.14. Considering $k = 60$ pN·μm^{-1} (equal to the stiffness of the trap that we can measure independently) and $d = 0.59$ nm (interphosphate distance for ssDNA), the parameters that best fit the experimental histograms for the 2.2 kbp sequence and their corresponding toy model parameters are

$$\left.\begin{array}{rcl} A &=& 0.058\\ B &=& 0.42\\ C &=& 2.95\\ n_c &=& 69 \end{array}\right\} \Longrightarrow \left\{\begin{array}{rcl} \mu &=& -2.80\ \text{kcal}\cdot\text{mol}^{-1}\\ \sigma &=& 2.2\ \text{kcal}\cdot\text{mol}^{-1}\end{array}\right. \tag{4.12}$$

The resulting CUR size distribution $P(n)$ is shown in green in Fig. 4.14a.

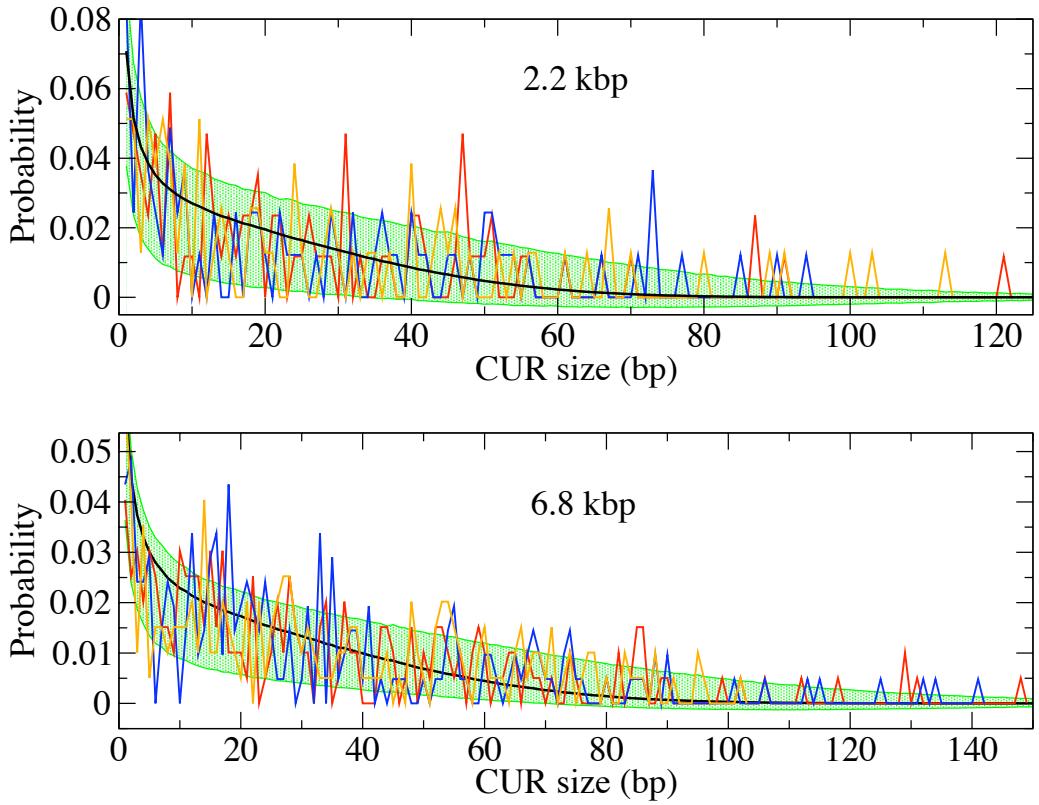


Figure 4.13: CUR size distributions calculated with the toy model. Upper (lower) panel shows the results for a 2252 (6838 bp) sequence. The black curve shows a CUR size distribution averaged over $n_r = 10^4$ realizations. The green region represents the upper and lower limits of the error bars that correspond to the standard deviation of those realizations. Red, blue and orange curves show 3 different realizations. Note the large deviations from the average histogram due to the finite length of the sequences. Simulations were performed with the following parameters: $k = 60$ pN·μm^{-1}, $\mu = -1.6$ kcal·mol^{-1} and $\sigma = 3.2$ kcal·mol^{-1}.

For the 6.8 kbp sequence we find

$$\left.\begin{array}{rcl} A &=& 0.050 \\ B &=& 0.43 \\ C &=& 3.0 \\ n_c &=& 91 \end{array}\right\} \Longrightarrow \left\{\begin{array}{rcl} \mu &=& -2.80 \text{ kcal} \cdot \text{mol}^{-1} \\ \sigma &=& 3.3 \text{ kcal} \cdot \text{mol}^{-1} \end{array}\right. \tag{4.13}$$

and the fit is shown in green in Fig. 4.14b. The values of μ and σ are not far from the actual mean and standard deviation of the energies of the nearest neighbor model for DNA,

$$\begin{array}{rcl} \mu &=& -1.60 \text{ kcal} \cdot \text{mol}^{-1} \\ \sigma &=& 0.44 \text{ kcal} \cdot \text{mol}^{-1} \ . \end{array} \tag{4.14}$$

Having not included the elastic effects of the ssDNA in the toy model we should not expect a good match between the fitting and the experimental values.

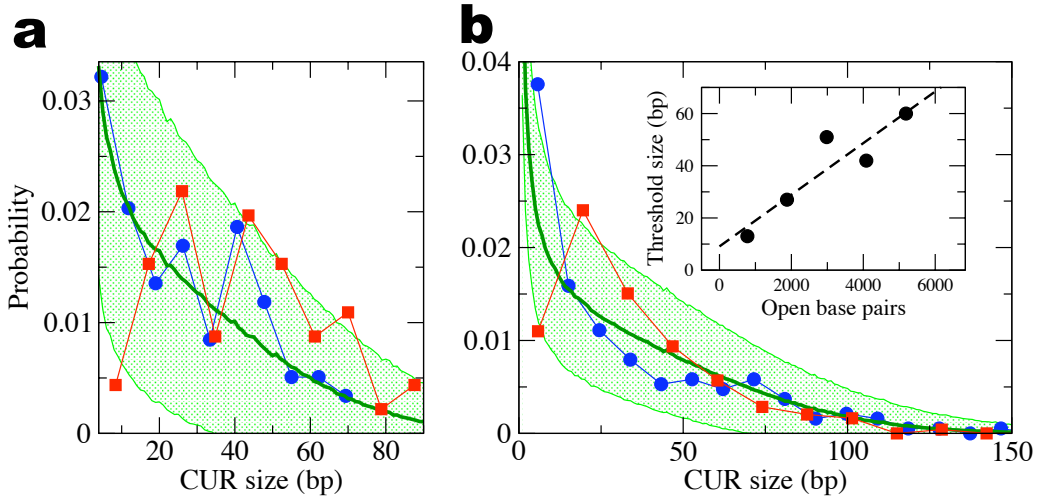


Figure 4.14: **(a)** Distribution of CUR sizes for the 2.2 kbp sequence. Red curve shows the experimentally measured distribution. Green curve shows the distribution predicted by the toy model and the shaded area shows the standard deviation from different sequence realizations of the same length. Blue curve shows the distribution predicted by the mesoscopic model for DNA. **(b)** Distribution of CUR sizes for the 6.8 kbp sequence. Same color code as in panel a. **(Inset of d)** Threshold size n^{thr} as a function of the number of open bps n. The dashed line is a linear fit, $n^{\mathrm{thr}} = 9.1 + 0.01n$.

However, there are two clear differences between the experimental and the predicted CUR size distributions. First, the experimental size distributions are not smooth but have a rough shape. We already know that this is a finite size effect described in Sec. 4.2.5. The distribution is smoother for the 6.8 kbp sequence because the sequence is longer, there is more statistics and the resulting CUR size distribution is better averaged. The second difference is that the toy model predicts a large fraction of CURs of size smaller than 10 bp that are not experimentally observed. There might be two explanations to this: 1) the toy model predicts small CURs that experimentally do not exist or 2) the method to detect metastable states is not capable of discriminating CURs smaller than 10 bps.

To better understand this, we can compute the CUR size distributions (depicted in blue in Fig. 4.14) with the mesoscopic model described in Sec. 3.4.1. Again, we find that the model predicts much more small CURs than we experimentally observe. Assuming that the model is correct, we conclude that the small CURs occur but the method of analysis has a limiting resolution of about $\sim$10 bp. In other words, for every large CUR detected experimen-

tally, the model predicts two (or more) small distributions. This limitation is due to the fact that the Bayesian analysis (Sec. 4.1) is not capable of distinguishing between force fluctuations and transitions between metastable states separated by less than 10 bp. A priori, it should be possible to do the pulling experiments at lower pulling rates and collect much more statistics. This would permit to have a better signal-to-noise ratio and discriminate the smaller metastable states. However, these experiments are much more difficult to carry out because the DNA molecule spends more time stretched and it breaks much more frequently before a whole pulling cycle can be completed.

Apart from these previous considerations, there is another issue that affects the discrimination of nearby metastable states. A quick look at Fig. 4.7 shows that histograms become smoother (i.e., the peaks are less sharp) as the molecule is progressively unzipped. The increased compliance of the molecular setup as ssDNA is released markedly decreases the resolution in discriminating intermediates (see the last paragraph of section 4.1.2 for a detailed explanation of this effect). In particular, for the 6.8 kbp construct we find that along the first 1500 bp of the hairpin only 30% of the total number of CURs smaller than 10 bp are detected, whereas beyond that limit no CUR smaller than that size is detected. If the threshold size $n^{\rm thr}$ is defined as the size of the CUR above which 50% of the predicted CUR are experimentally detected we find that $n^{\rm thr}$ increases linearly with the number of open bps, establishing a limit around 10 bp for the smallest CUR size that we can detect (Fig. 4.14b, inset).

Now let us focus on the other side of the distribution (large CUR sizes). The three CUR size distributions in Fig. 4.14 are long tailed distributions, which indicate that large CURs occur with finite probability. Unfortunately, large sized CUR hinder their internal DNA sequence, limiting the possibility of unzipping one base-pair at a time, which would permit to sequence the DNA. Under what experimental conditions is it possible to break up large sized CUR into individual bps? Only by applying local force on the opening fork (thereby avoiding the large compliance of the molecular setup) and by increasing the stiffness of the probe might be possible to shrink CUR size distributions down to a single base-pair [117]. Figs. 4.15a, 4.15b show how the CUR size distributions shrink and the largest CUR size decreases as the stiffness increases. Its value should be around 50-100 $\mathrm{pN{\cdot}nm^{-1}}$ for all CUR sizes to collapse into a single bp. Remarkably enough this number is close to the stiffness value expected for an individual DNA nucleotide stretched at the unzipping force. Any probe more rigid than that will not do better.

Similarly to the problem of atomic friction we can define a parameter η (defined as the ratio between the rigidities of substrate and cantilever) that

controls the transition from stick-slip to continuous motion [127]. For DNA unzipping we have

$$\eta = \frac{|\mu|}{kd^2} \tag{4.15}$$

where μ is the average free energy of formation of a single bp, k is the probe stiffness and d is the interphosphate distance. The value $\eta = 1$ determines the boundary where all CURs are of size equal to one bp ($\eta < 1$). In our experiments we have $\eta \simeq 500$ and to reach the boundary limit $\eta = 1$ we should have $k \sim 100$ pN·nm^{-1} consistently with what is shown in Figs 4.15a, 4.15d. Interestingly enough, the boundary limit of such stiffness is very similar to the expected stiffness of one ssDNA base. The stiffness of one nucleotide is the derivative of the Force vs. Extension Curve (Eq. 3.20) for the Freely Jointed Chain (FJC) model (see Appendix F), which is given by:

$$k_s(f) = \left[n \cdot d \left(-\frac{b}{k_B T} \mathrm{cosech}^2 \left(\frac{bf}{k_B T} \right) + \frac{k_B T}{bf^2} \right) \right]^{-1} \tag{4.16}$$

where $k_s(f)$ is the stiffness at force f, n is the number of open base-pairs, d is the interphosphate distance, k_B is the Boltzmann constant, T is the temperature and b is the Kuhn length. Using the following values for the parameters: $b = 1.2$ nm, $d = 0.59$ nm, k_BT=4.11 pN·nm and applying Eq. 4.16 to one single nucleotide ($n = 1$) we get a stiffness of $k_s = 113$ pN/nm at $f = 15$ pN and $k_s = 127$ pN/nm at $f = 16$ pN. It is remarkable that the elastic properties of ssDNA lie just at the boundary to allow for one bp discrimination. This suggests that the ssDNA has the correct elastic properties so that the DNA machinery can *read* the base-pairs one at a time, without having to expose hundreds of base-pairs to the solvent, thereby reducing the risk of damage.

4.4 Conclusions

The unzipping of DNA at controlled position exhibits cooperative openings of base-pairs (called CURs) as the distance between the two beads is increased. These CURs show characteristic statistical properties that depend on the sequence of the DNA and the parameters of the experiments (e.g., trap stiffness).

We have developed a Bayesian approach to extract the statistical properties of CURs directly from the experimental data. Such analysis can be extended to other systems (such as proteins or RNA) that show intermediate states in the unfolding process. The obtained histogram of intermediate

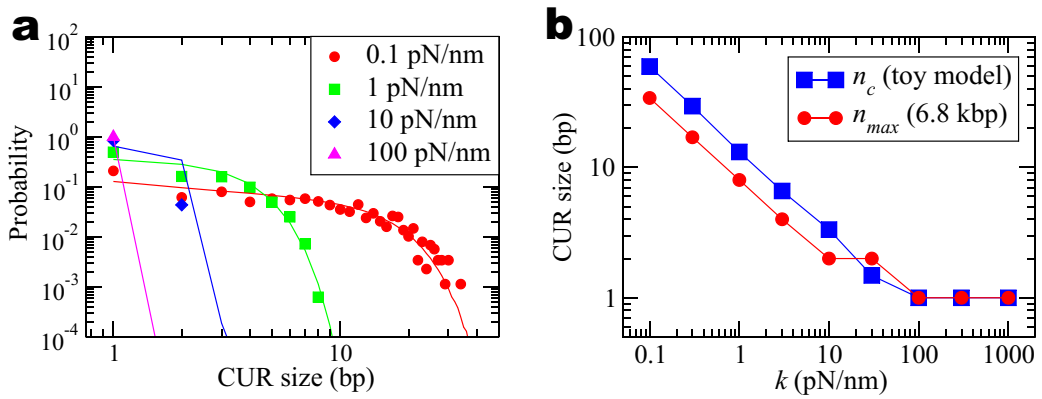


Figure 4.15: **(a)** CUR size distributions in log-log scale for some values of k using the toy model. Data plotted with points shows the CUR size distribution for the 6.8 kbp sequence. Data plotted with lines, shows the average CUR size distribution over 10^4 realizations (k=60 pN·μm^{-1}, d = 0.59 nm, μ=-1.6 kcal·mol^{-1}, σ=0.5 kcal·mol^{-1}). **(b)** The fit of the average CUR size distributions in panel a to Eq. 4.10 give the cutoff size n_c. It decreases like $n_c \simeq k^{-2/3}$. Blue curve shows n_c vs. k. Red curve shows the maximum CUR size (n_{max}) predicted by the toy model for the 6.8 kbp sequence. For $k > 100$ pN·nm^{-1}, both curves level off to CUR sizes of 1 bp.

states is a sign of the molecule and contains valuable information about the stability and the coexistence of these metastable states.

The properties of the CURs can be described by a simplified toy model that captures the essential features of the unzipping. The advantage of the toy model is that the calculations are easy to implement and the computational time required for the calculations is dramatically reduced. This is useful to intensively explore the predictions of the unzipping properties for different situations: trap stiffness, NNBP variances, sequences, etc. The results predicted by the toy model compare well with the experimental data obtained from unzipping.

The detection of intermediate states using the Bayesian analysis is blurred by the thermal noise of the experimental data. In our experimental conditions, the CURs smaller than ∼10 bp are hardly detected. Below this boundary, the fluctuations in force due to the thermal noise cannot be distinguished from the fluctuations due to the coexistence of states. A priori, this limit could be reduced down to 1 bp by collecting more accurate experimental data and finding the characteristic signature of each type of noise (correlation, spectrum, bandwidth).

The CUR size of the experimental unzipping is affected by the release of ssDNA. As the molecule is being unzipped, the amount of ssDNA between the unzipping fork and the optical trap increases. These reduces the compliance of the tether and the capability to distinguish between intermediate states is

less clear. The unzipping of a DNA molecule one base-pair at a time requires a minimum stiffness of 100 pN/nm (= 0.1 N/m). This could be achieved by applying local force on the unzipping fork, avoiding the accumulation of ssDNA that transmits the force to the optical trap.

Interestingly enough, the minimum stiffness required to unzip one base-pair at a time coincides with the expected stiffness of one single base of the ssDNA. On the other hand, the stiffness of the proteins that directly pull on DNA (e.g., helicases) or read the sequence (e.g., polymerases) can be assumed to be very large (proteins are very rigid objects) compared to the stiffness of a single base-pair. Therefore, the stiffness of the proteins are not a limiting factor to unzip DNA molecules one base-pair at a time. It is remarkable that ssDNA has the minimum required elastic properties so that the genetic information encoded in DNA can be accessed by the replication and transcription machinery.

To sum up, the unzipping of DNA one base-pair at a time could be used to infer the sequence of an unknown DNA fragment. This could be experimentally achieved by increasing the stiffness of the probe, applying local force to the unzipping fork and improving the Bayesian analysis to distinguish the thermal force fluctuations from the force fluctuations due to the coexistence of states.

Doutez toujours de vous-mêmes, jusqu'à ce que les données ne laissent aucun doute.

Louis Pasteur (1822-1895)

Chapter 5

Salt dependence of nearest-neighbor base-pair free energies

Nowadays, many biological applications require an accurate calculation of the free energy of formation of nucleic acids. The prediction of secondary structures is important in a wide range of fields such as self-assembled structures in DNA origami [128, 129]; achievement of high selectivity in the hybridization of synthetic DNAs [130]; antigene targeting and siRNA design [131]; characterization of translocating motion of enzymes that mechanically disrupt nucleic acids [102]; prediction of non-native states (e.g., RNA misfolding) [132]; and DNA guided crystallization of colloids [133].

The Nearest-Neighbor (NN) model for DNA thermodynamics described in Sec. 3.3 was developed in the early 1960's [26]. Since then, this model has been successfully applied to predict the free energy of secondary structures in nucleic acids. The main premise of the model is simple: the total free energy change to form a double helix can be written as the sum of the base-pair interactions (which depend on the base-pair itself and on the nearest-neighbor). This means that two fragments of DNA that have different sequences will generally have different free energies of formation. It is important to note that the model does not include the energetic values of the NN interactions. The values of these energies are called Nearest-Neighbor Base-Pair (NNBP) energies and they must be determined from experiments.

The estimates of the NNBP energies have been traditionally obtained from thermal denaturation experiments of DNA oligonucleotides [134, 135]. These experiments consist on heating a sample of a DNA oligo, measuring the melting temperature and inferring the thermodynamic properties of the molecules (enthalpy and entropy). In 1998, John SantaLucia Jr. gathered the experimental data from several labs (including his own) and provided a set of unified NNBP energies that best fit all the results [28]. Since then, these Unified Oligonucleotide (UO) energies have been considered the most reliable estimates and have been widely used as a reference. What is more, the validity of the NN model and the UO energies have exceeded the boundaries of melting experiments and they have been tested in single-molecule techniques. During the first decade of the 21st century, most of the works done on force denaturation experiments of nucleic acids used these values with satisfactory results [49, 24, 116, 136, 113]. Indeed, the UO values have become a referent to model the single-molecule experiments (see Sec. 3.4) because they reproduce quite well the force vs. extension curves of unfolding/refolding experiments of DNA and RNA molecules.

The UO energies are given at standard conditions (25°C and 1 M [NaCl]) and a few correction formulas are available to extend the values of the energies to other experimental conditions (temperature and salt concentration). However, some discrepancies appear when comparing the UO predictions with the single-molecule pulling experiments at different salt concentrations. For instance, at low salt concentration (10 mM [NaCl]), the measured mean unzipping force is ∼12 pN, while the force predicted by the UO energies is ∼13 pN. This is a significant discrepancy since the force resolution of the instrument is 0.1 pN.

The origin of this discrepancy is attributed to the differences in the assumptions and the treatment of data in both types of experiments. In melting experiments, the DNA duplexes are assumed to melt in a two-state fashion (all or none; hybridized or denaturated). This is acceptable for short oligos (less than ∼ 20 bp) but the two-state assumption fails as the length of the duplex increases. So, the longer oligos exhibit intermediate states where the DNA duplex is partially melted. This is the reason why there are two sets of NNBP energies inferred from melting experiments: one for the oligomers and one for the polymers [28]. The description of the unzipping experiments on DNA also has its own assumptions (elasticity of ssDNA, linearity of the optical trap, etc.). Nevertheless, unzipping experiments do not distinguish between the unfolding of oligonucleotides and polynucleotides. This is a useful feature because the NNBP energies are unique for all the DNAs of any length. The key point here is that single-molecule techniques allow one to control and monitor the denaturated state of the molecule along a full reac-

tion coordinate, without having to rely on a two-state model.

Up to now, we know that the NNBP energies given by the melting experiments are not capable of reproducing the unzipping experiments. Here emerges an interesting question. What are the values of the NNBP energies that quantitatively describe the DNA unzipping experiments? Can we extract them from our own experimental data? Since we have the data and we have the model, in order to extract the energies, we just need to tune the parameters to make the model fit the data.

This chapter describes how to infer the NNBP energies of DNA from unzipping experiments performed on optical tweezers. Here we develop the mathematical tool that allows us to reach this goal. The strategy consists in: 1) measuring the FDC of unzipping; 2) determining the elastic response of the ssDNA; 3) using the model described in Sec. 3.4 to compute the theoretical FDC; and 4) fit the experimental FDC to the theoretical one to obtain the NNBP parameters. The advantage of this technique is that it and can extended to a wide variety of conditions (salt concentration, pH, temperature) and molecules (RNA, proteins such as helical repeats like the leucine zipper [137]).

5.1 Discrepancies between melting and unzipping experiments

As previously said, the mesoscopic model (see Sec. 3.4) with the UO NNBP energies provides a prediction of the unzipping FDC that does not quantitatively reproduce the experimental FDCs. Figures 5.1a,b show the comparison. Note that the sawtooth pattern is well reproduced by the model, in the sense that the number of slopes and force rips coincide with the experimental measurements. However, there is an evident difference in the mean unzipping force: the UO prediction overestimates this value. Moreover, there is a misalignment between the positions of the slopes and the force rips. Such differences cannot be attributed to calibration errors.

Figure 5.1c shows the comparison of the mean unzipping force for several salt concentrations of the buffer (10 mM [NaCl] - 1 M [NaCl]). This discrepancy between the UO prediction and the unzipping experiments is systematic at all salts. Besides, the sign of such discrepancy is the same for both DNA sequences.

The fit of the NNBP energies can make the mesoscopic model quantitatively match the experimental FDC. In order to do so, the FDC of a few molecules is measured at different salt concentrations. Each FDC is fit to the

model and the resulting NNBP energies are averaged over all the molecules to give a final estimation of the energies at each salt. From these results, we can infer the rules to correct the NNBP energies at different salt concentrations.

In order to proceed with the fit of the FDCs, we must take into account several considerations. Next section discusses all them.

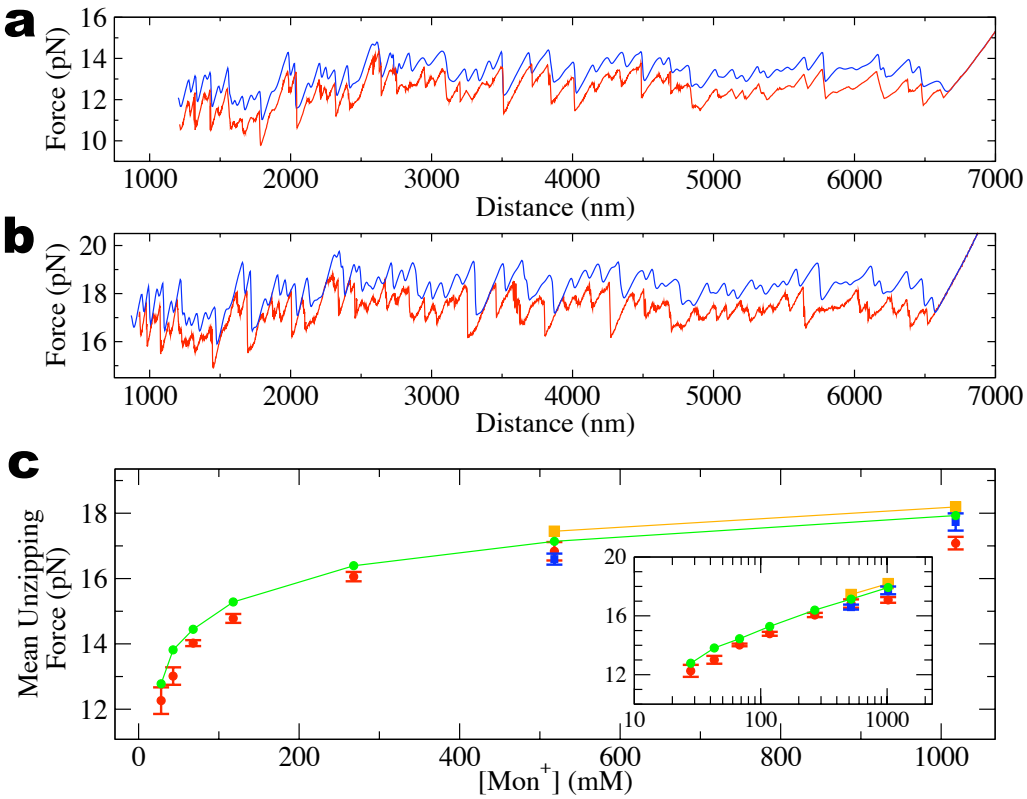


Figure 5.1: Discrepancies between the unzipping measurements and the UO theoretical prediction. **(a)** FDC measured and filtered at 1 Hz bandwidth (red curve) and UO prediction (blue curve) for a 6.8 kbp sequence at 10 mM [NaCl]. **(b)** 6.8 kbp sequence at 1 M [NaCl]. Same color code as in panel a. **(c)** Dependence of mean unzipping force with salt concentration. Red points, experimental measurements for the 6.8 kbp sequence; green curve, UO prediction for the 6.8 kbp sequence; blue points, experimental measurements for the 2.2 kbp sequence; orange curve, UO prediction for the 2.2 kbp sequence. The values for the 6.8 kbp and the 2.2 kbp molecules have been obtained after averaging over six molecules. Error bars are determined from the standard error among different molecules.

5.2 Parameters defining the theoretical FDC

The theoretical FDC computed in Sec. 3.4.1 requires the knowledge of several (theoretical and empirical) parameters. The equations that determine the total extension (Eq. 3.23) and the total energy (Eq. 3.22) implicitly depend on:

- **The stiffness of the optical trap**. A stiffer trap (see Sec. 4.2.2) entails a FDC with smaller and numerous force rips, while the rips of a softer trap are larger and and less frequent. This parameter appreciably changes the global shape of the FDC when its value is changed by orders of magnitude. In our particular case, the stiffness of the optical trap is restricted to $60 - 80$ pN/μm. Therefore, this parameter hardly affects the shape of the FDC (see Fig. 5.2a). The stiffness of the optical trap is obtained by immobilizing a bead at the tip of the micropipette and measuring the force vs. elongation (see Fig. 2.15d).

- **The persistence length and the contour length of the dsDNA handles**. The dsDNA handles are very short and short polymers are very rigid. In general, a polymer can be treated as a rigid rod when its persistence length is higher than its contour length. The accepted value of the persistence length for long molecules of dsDNA is ~ 50 nm [48], while the contour length of the handles (i.e., their length) is < 10 nm. Even if the actual persistence length of short molecules of dsDNA is lower that 50 nm, the handles will actually behave like rigid rods. Rigid rods fully transmit the forces along their ends. So the high rigidity of the dsDNA hardly affects the FDC (see Fig. 5.2b).

- **The persistence (or Kuhn) length and the contour length of the released ssDNA**. The elastic properties of the ssDNA strongly affect the mean unzipping force of the FDC. The parameters that define these elastic properties are not well defined and there are experimental evidences that they are sequence dependent in ssRNA [138]. In our unzipping experiments, the ssDNA is so important because it is an element that transmits the force from the unzipping fork to the handles (and the optical trap). Besides, it is not a static element. Instead, it increases its contour length as the molecule is being unzipped. Apart from that, the ssDNA is a soft polymer compared to dsDNA, so its persistence length is one order of magnitude lower (~ 1.5 nm). In ssDNA, the misprediction of the persistence length has dramatic consequences (see Fig. 5.2c). Indeed, a 10% difference in the persistence length of the ssDNA induces a difference of 1 pN in the mean unzipping force.

- **The NNBP energies**. The values of the energies affect the overall shape of the FDC. The mean unzipping force of the FDC can be increased (decreased) by globally increasing (decreasing) the absolute value of the NNBP energies. An average 10% correction in the NNBP energies introduces a difference of ∼1.3 pN in the mean unzipping force (see Fig. 5.2d). Besides, the local shape of the sawtooth pattern (heights of the rips and lengths of the slopes) change when the relative values among the NNBP energies are modified. The dependence of the FDC on the NNBP is non-linear. This means that the relative changes on the NNBP energies have unpredictable effects on the shape of the FDC. According to the NN model (see Sec. 3.3) there are 10 independent NNBP energies.

  An extra energy parameter is required to take into account the effect of the end loop. It basically increases or decreases the height of the last rip of the FDC.

In order to obtain an estimation of the NNBP, all the other parameters must be fixed. The stiffness of the optical trap and the elastic properties of the handles can be determined easily. However, some experiments must be carried out in order to determine the elastic properties of the ssDNA. Apart from this, there are two more issues that have to be addressed before proceeding with the fit of the NNBP. The following sections deal with them.

5.2.1 Elastic response of the ssDNA

The mean unzipping force of DNA depends on the elastic properties of the ssDNA according to the following explanation. Let us assume that the ssDNA is more rigid than usual. At the same force, a more rigid ssDNA is capable of storing less elastic energy ($E_b = f^2/2k$, see Eq. 3.16). Therefore, the force needed to acumulate enough elastic energy to break the base-pairs of the duplex increases. Thus, the elastic properties of ssDNA entering in Eq. 3.22 strongly determine the mean unzipping force of the FDC. In general, the elastic response of the ssDNA is basically determined by its stiffness (which is controlled by the persistence or Kuhn length) and the contour length (which is controlled by the interphosphate distance and the number of released bases).

A priori, one might think that the elastic parameters of the ssDNA are well-known and fixed. However, these values depend on the type of buffer at which the experiments are performed. It is crucial to know the elastic response of the ssDNA at forces below 20 pN because this is the range of forces at which unzipping is observed. Precisely at these forces it is more

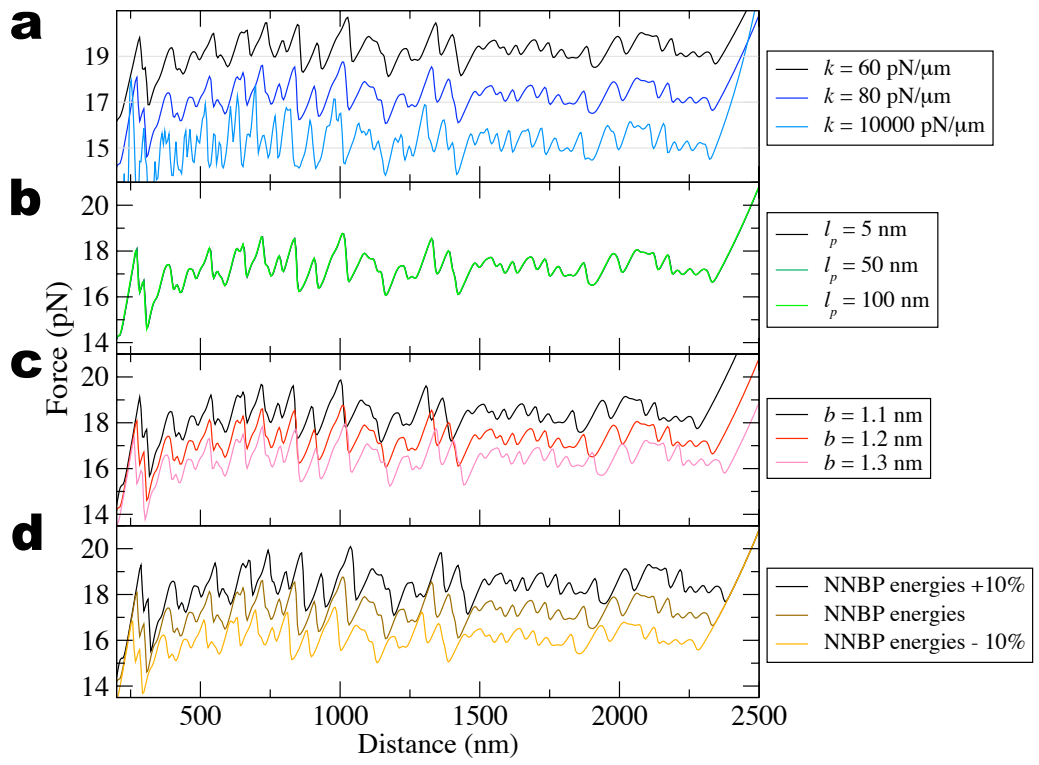


Figure 5.2: Dependence of the theoretical FDC on the experimental parameters for the 2.2 kbp sequence. **(a)** Trap stiffness. The three cases have the same mean unzipping force, so the curves have been displaced vertically ±2 pN for convenience. There are no significant differences between the 60 and the 80 pN/μm curves. However if the trap is orders of magnitude stiffer, the sawtooth pattern is appreciably different at short distances. As the ssDNA is released, the effective stiffness of the system decreases and the FDC looks like the other two cases. **(b)** Elasticity of the handles. The persistence length (l_p) of the dsDNA hardly changes the shape of the FDC (< 0.05 pN). **(c)** Elasticity of the ssDNA. The Kuhn length (b) of the ssDNA (which is directly related with the rigidity) changes the mean unzipping force. The higher the rigidity, the lower the unzipping force. **(d)** NNBP energies. A global increment or decrement of the NNBP energies modifies the mean unzipping force.

difficult to extract the ideal elastic response because the ssDNA tends to rezip and form secondary structures. So we need to measure a partial property of the experiment (ssDNA elasticity) that is hidden by the proper experiment (DNA unzipping).

The structure of the molecular construct used in our experiments helps to solve this problem. Indeed, the end loop added to the DNA duplex (see Sec. 3.2.3 and Appendix E) prevents the total separation of the two strands of DNA. When all the base pairs are disrupted, the molecular construct looks like a ssDNA molecule that can be pulled to obtain the elastic response.

However, it is only possible to measure the elasticity of the ssDNA after the molecule has been completely unzipped, i.e., for force values higher than 20 pN. Yet we face the problem that different models and different values for the elastic parameters are compatible with the same curve above 20 pN, but they show very different properties for forces below 20 pN. So the problem persists.

The easiest solution then is to synthesize a fragment of ssDNA and perform pulling experiments on it. Since this fragment has a random sequence, the formation of secondary structure is minimized and we can observe the elastic response of the ssDNA at low forces. We carried out pulling experiments with a 3 kbp piece of ssDNA from λ-DNA (see Appendix I for the details about the synthesis). In contrast to the molecular constructs for unzipping, this 3 kbp sequence is not self-complementary and it forms less secondary structure at low forces. The measured FDCs are converted to Force vs. Extension Curves (FECs) in which the force applied is depicted versus the extension of the ssDNA molecule x_m (rather than the relative position of the trap x_{tot}). The transformation from one to type of curve to another is performed point by point. So each point (x_{tot}, f) is converted into a point (x_m, f) by subtracting the elongation of the bead to the total distance of the system according to

$$x_m = x_{\mathrm{tot}} - \frac{f}{k} \tag{5.1}$$

where f is the force measured and k is the stiffness of the optical trap (see Fig. 5.3a,b). The resulting FEC is fit to a FJC or a WLC (Fig. 5.3a,b). The fitting parameters are the Kuhn length (b) or the persistence length (l_p) depending on the model, and the interphosphate distance (d). The FEC is forced to pass through the point $(x_m, f) = (0, 0)$, while the number of bases is fixed to $n = 3000$ (see Appendix I). It has also been checked that the obtained elastic properties of the ssDNA match the last part of the unzipping FDC, when the molecule is fully extended (see Fig. 5.3c,d).

Both the FJC and the WLC models are similar but have some differences at low forces. The WLC model correctly fits the ssDNA elastic response for salt concentrations below 100 mM [NaCl]. Above this value, FDCs develop a plateau at low forces that cannot be reproduced by the ideal models (FJC or WLC). Figure 5.3 shows the fits of the FEC of ssDNA at different salt concentrations. As the salt concentration increases, the matching between the experimental FEC and the model is worse. The problem is the appearance of a force plateau above 100 mM [NaCl]. Such force plateau is related to the formation of secondary structures (self-hybridization) in ssDNA [139]. How can we determine the elastic response of the ssDNA at high salt concentration? At 100 mM [NaCl] we find that FJC fits better than WLC. At salt

concentrations higher than 100 mM [NaCl], it is reasonable to assume that the ideal elastic response of the ssDNA is given by the FJC model, without considering the force plateau. So we fit the FEC to a WLC above 15 pN, because no secondary structure survives at this force. Indeed, above 15 pN, the FJC model fits data better.

Summing up, Fig. 5.3e shows the FEC at different salt conditions and the best fit to each of them and Table 5.1 shows the values of parameters for the best model at each salt.

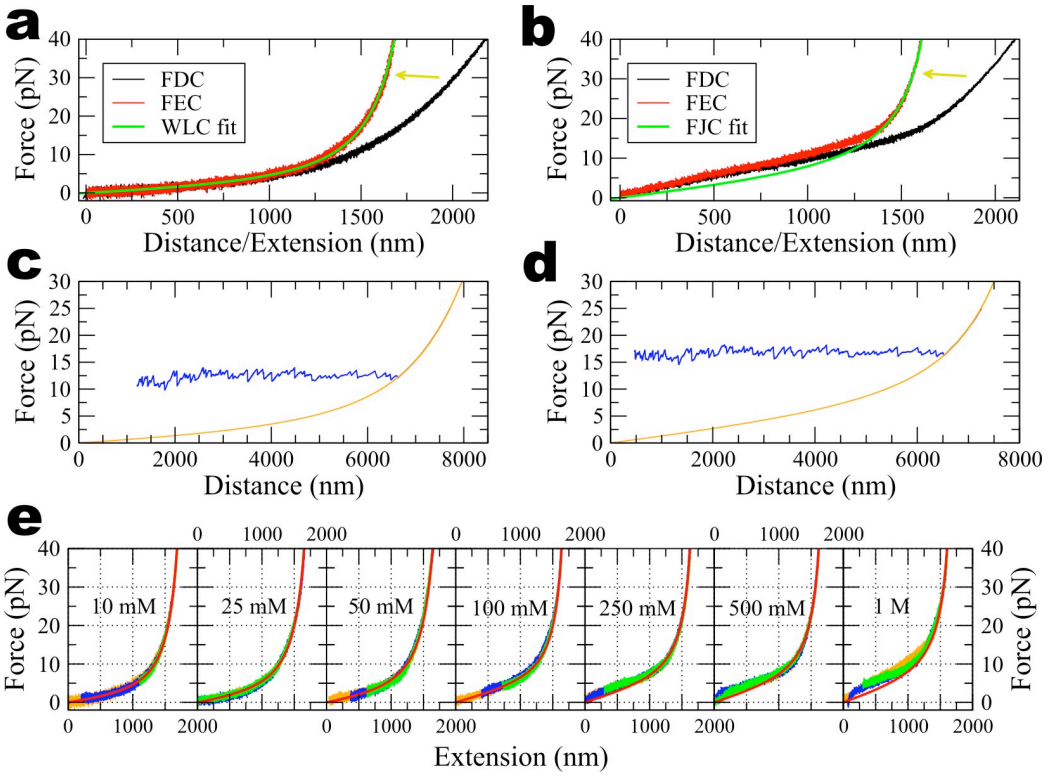


Figure 5.3: Fit of the elastic response of ssDNA. **(a,b)** Elastic response of the ssDNA at 10 mM [NaCl] and 1 M [NaCl], respectively. The panels show the conversion of FDCs (black curves) into FECs (red curves) for the 3 kb ssDNA molecule (yellow arrows indicate the direction of the conversion). The green curves show the fit of the FEC to the ideal models (WLC or FJC). **(c,d)** Predicted FDC for the fully unzipped molecule (orange curve) superimposed on the experimental unzipping FDC (blue curve) at 10 mM [NaCl] and 1 M [NaCl], respectively. **(e)** Elastic response of a 3 kb ssDNA molecule at various salt concentrations. For each salt, the raw data of three different molecules are shown (orange, green and blue curves). Red curve shows the best fit to the elastic model. The models are: FJC for [NaCl]$\leq$ 100 mM and WLC for [NaCl]$>$ 100 mM.

	WLC model $d = 0.665$ nm	**FJC model $d = 0.59$ nm**
Salt [NaCl]	**l_p (nm)**	**b (nm)**
10 mM	1.14 (0.1)	-
25 mM	0.93 (0.1)	-
50 mM	0.88 (0.1)	-
100 mM	-	1.37 (0.1)
250 mM	-	1.25 (0.1)
500 mM	-	1.20 (0.1)
1000 mM	-	1.10 (0.1)

Table 5.1: Elastic parameters of ssDNA at different salt concentration. d is the inter-phosphate distance for each model, l_p is the persistence length of the WLC and b is the Kuhn length of the FJC. The mean values were obtained after averaging over 5 molecules for each salt, except for 25 mM and 100 mM that were averaged over 4 molecules and for 50 mM that were averaged over 3.

5.2.2 Drift and shift function

The drift of the instrument is one of the major problems in single molecule experiments. The drift is a low frequency systematic deviation of measurements due to macroscopic effects. In the case of the minitweezers, the drift is mainly due to dilatations or contractions of the instrument produced by local changes of temperature or air flows in the room. The drift might be responsible of several effects such as the change in the relative position between the micropipette and the optical trap or the distortion of the optical path.

The importance of drift depends on the kind of the experiment and the protocol used. The unzipping experiments are performed at very low pulling speeds (typically around 10 nm/s), so measuring a whole unzipping/rezipping FDC may take 10 minutes or longer. Therefore it is useful to model the drift in order to remove its effects and extract accurate estimates for the NNBP energies. In order to do so, we introduce into the model a shift function that locally corrects the position of the trap along the FDC (see Appendix J). The final goal of the shift function is to make the slopes and force rips of the experimental and the theoretical FDC match each other. This leads to an improved match between the theoretical and the experimental FDCs.

The shift function is defined by equidistant control points that define a set of cubic splines. The shape of the shift function can be modified by tunning the location and the local shift of each of these control points. Appendix J describes in further detail the issues related with this function.

The shift function has a very important property: It is completely uncorrelated with the NNBP energies. Indeed, the shift function can be introduced into the mesoscopic model without affecting the NNBP energy values. Therefore, in the complete model, the shift function controls the horizontal matching between both curves (i.e., the distance) and the NNBP energies control the vertical matching (i.e., the force). The result is that an improved agreement between the experimental and the theoretical unzipping FDCs is achieved (see Fig. 5.4).

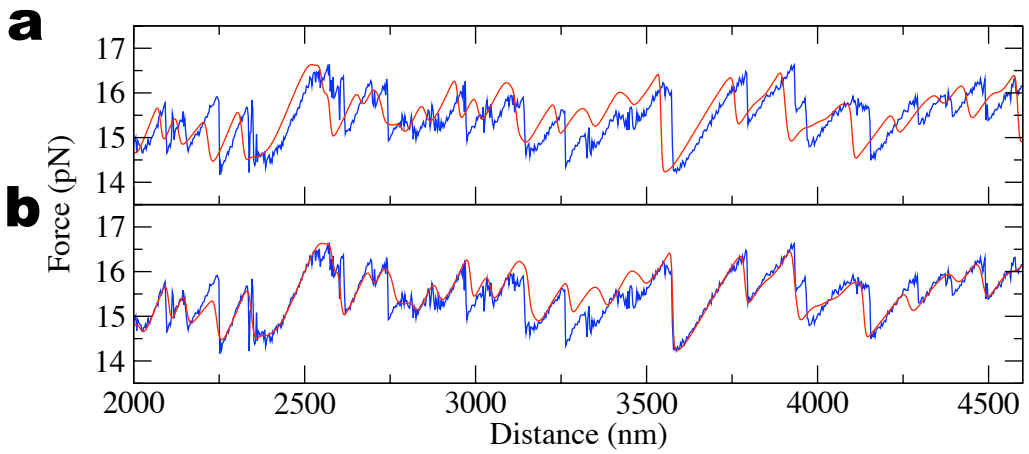


Figure 5.4: Effect of the shift function on the theoretical FDC. Blue curve shows the experimental FDC measured on a 6 kbp sequence. **(a)** Red curve shows the FDC predicted by the mesoscopic model without the shift function. **(b)** The inclusion of the shift function locally shifts each rip and the matching is better.

Although each molecule has its own shift function, these have similar shapes (see Fig. 5.5). We have checked that the undulations of the shift function are not an artifact of the spline interpolation (see Appendix J). The undulations remain when the number of control points of the shift function is increased. The net shift observed in some curves (around $\sim$100 nm) might be explained by improper calibration of the distance (around 4%). The undulations observed in the shift function might be due to non-linearities in the light-lever (i.e., trap position) measurements or interference fringes in the lenses or in the pellicle located along the optical path to the PSDs. The undulations in the shift function might also be correlated with the DNA sequence as emerges from the fact that undulations observed in different molecules of the same sequence appear at nearby positions. This might indicate new effects in the unzipping curves not accounted for in the NN model (e.g., the presence of next nearest-neighbor corrections).

It has also been checked whether the correction introduced by the shift

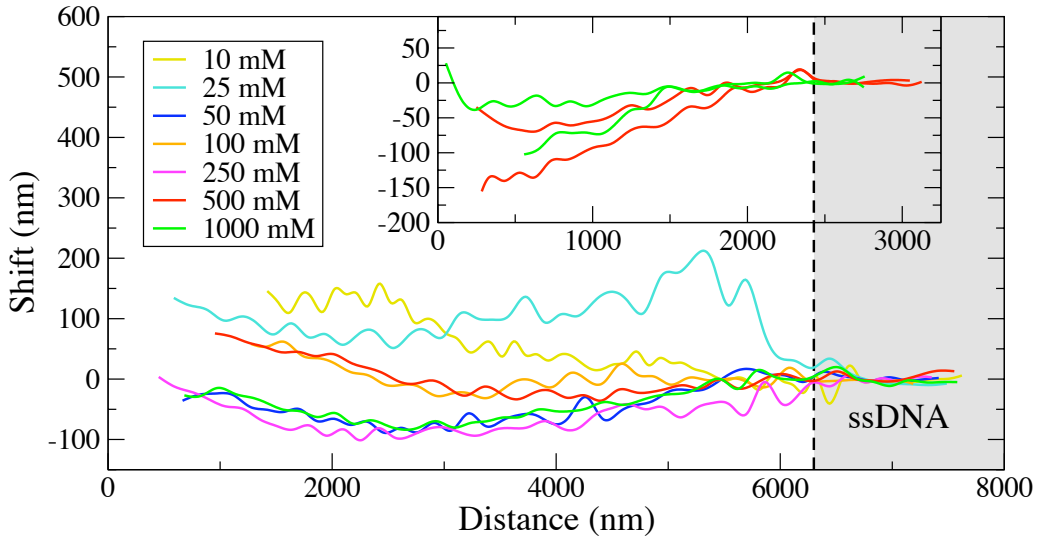


Figure 5.5: Shift function. The figure shows the shift function for a few 6.8 kbp molecules at different salt conditions. The inset shows the shift function for different molecules of 2.2 kbp at 500 mM [NaCl] (=red) and 1 M [NaCl] (=green). The gray shaded region corresponds to trap positions where DNA is fully unzipped. In this region, the local shift nearly vanishes.

function could be explained by a dependence of the Kuhn length on the contour length. Such dependence has not ever been reported. Yet it is interesting to evaluate the consequences of such hypothetic dependence. By letting the Kuhn length depend on the number of open base pairs, the position of the theoretical and experimental slopes and rips match each other if the Kuhn length increases as the contour length decreases. However this matching occurs at the price of an increasing average mean unzipping force as the molecule unzips and the ssDNA is released, an effect which is not experimentally observed. We conclude that the shift function is probably due to instrumental drift superimposed to imperfect calibration of the distance and non-linear optical effects.

5.2.3 Free energy of the end loop

The end loop of the molecule is a group of 4 bases that forms a different structure from the Watson-Crick base-pairs. The loop is a motif in which the backbone of the ssDNA molecule bends and twists itself. This induces a change in the orientation of the backbone that permits the formation of the DNA duplex (see Fig. 5.6a). The loop is quite useful in unzipping/rezipping experiments because the two strands of DNA are not completely split apart

after unzipping, which facilitates the rezipping. Moreover, the end loop allows us to measure the elastic response of a ssDNA above ~ 15 pN, which is an important piece of information to infer the NNBP energies from unzipping experiments (see Fig. 5.3c,d).

The free energy formation of the loop gets contributions from the bending energy, the stacking of the bases in the loop and the loss of entropy of the ssDNA. The energy formation of the loop is positive, meaning that the loop is an unstable structure at zero force. Upon decreasing the total extension, the formation of complementary base pairs along the sequence reduces the total energy of the molecule and the loop can be formed.

The effect of the loop is appreciated only in the last rip of the FDC. It introduces a correction to the free energy of the fully extended ssDNA molecule and modifies the force at which the last rip is observed (Fig. 5.6b).

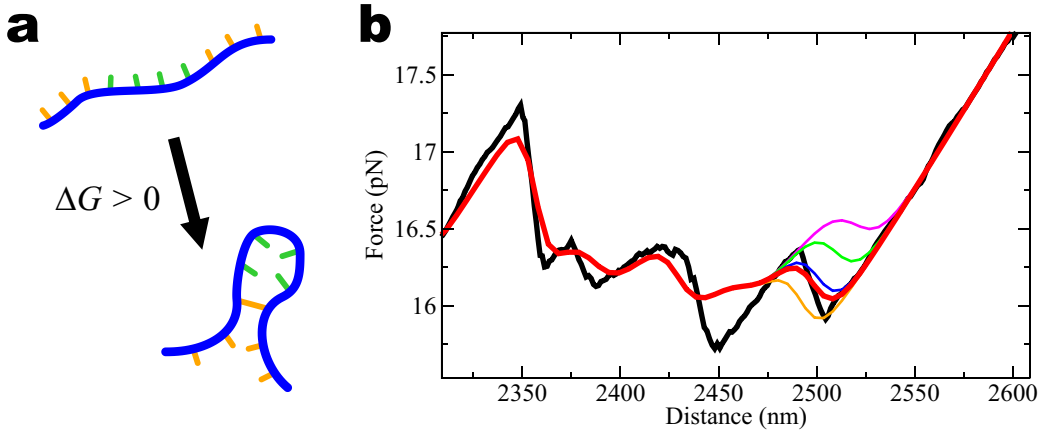


Figure 5.6: End loop. **(a)** The formation of the end loop involves the bending and twisting of the sugar-phosphate backbone of the DNA (depicted in blue). The four bases of the tetraloop (depicted in green) cannot form Watson-Crick base-pairs and the free energy of formation is positive (the loop is unstable). **(b)** Effect of the loop contribution on the FDC. The free energy of the loop modifies the shape of the theoretical FDC only at the last rip just before the elastic response of the full ssDNA is observed. The black curve is the experimental FDC. All other curves show theoretical FDCs with different values of ϵ_{loop}. Red curve, best fit with $\epsilon_{\text{loop}} = 2.27$ kcal·mol^{-1}; magenta curve, $\epsilon_{\text{loop}} = 0.0$ kcal·mol^{-1}; green curve, $\epsilon_{\text{loop}} = 1.00$ kcal·mol^{-1}; blue curve, $\epsilon_{\text{loop}} = 2.00$ kcal·mol^{-1} and orange curve, $\epsilon_{\text{loop}} = 3.00$ kcal·mol^{-1}.

5.3 Fit of FDCs

The values of the NNBP energies mostly define the location and the height of the force rips. If these values are modified, the shape of the FDC changes

dramatically. So if we consider that the other parameters (trap stiffness, elastic properties of ssDNA) are known and fixed, it is possible to state that the FDC only depends on the 10 NNBP energies ϵ_i (i=1,...,10) and on the free energy of the end loop ϵ_{loop}.

The question that we want to address here is how to modify the values of ϵ_i in order to obtain a theoretical FDC that is as close as possible to the experimentally measured one. In order to do so, it is necessary to perform a fit by using the Least-Square approach. This method consists in defining an error function $E(\epsilon_1, \ldots, \epsilon_{10}, \epsilon_{\text{loop}})$ that is the sum of the squared differences between a measured value and a theoretical value of the FDC according to

$$E(\epsilon_1, \ldots, \epsilon_{10}, \epsilon_{\text{loop}}) = \frac{1}{N} \sum_{i=0}^{N} \left(f_i^{\text{exp}} - f_i^{\text{the}}(\epsilon_1, \ldots, \epsilon_{10}, \epsilon_{\text{loop}}) \right)^2 \tag{5.2}$$

where N is the number of experimental points in the FDC, f_i^{exp} is the experimental equilibrium force and $f_i^{\text{the}}(\epsilon_1, \ldots, \epsilon_{10}, \epsilon_{\text{loop}})$ is the theoretically calculated FDC according to the model. Obviously, the values of f_i^{exp} and f_i^{the} are calculated at the same total extension of the system x_{tot}.

The theoretical FDC is calculated in equilibrium, which assumes that the bandwidth of our measurements is 0 Hz. The experimental data is filtered at a bandwidth of 1 Hz. If data is filtered at higher frequencies (>1 Hz), the hopping between intermediate states is observed and the experimental FDC does not compare well with the theoretical FDC at equilibrium. If the data is filtered at lower frequencies (<1 Hz), the force rips are smoothed and hopping transitions are averaged out. Appendix K explains why the (filtered) experimental FDC can be compared with the theoretical FDC.

The best values for the NNBP energies can be inferred by minimizing the mean squared error (i.e., the error function) between the experimental and theoretical FDCs. The complex dependence of the FDC with the NNBP energies and the large dimensional space spanned by the 10 NNBP energies makes the minimization a difficult optimization problem. In general, we have to minimize Eq. 5.2 which is a 11-dimensional function whose *error landscape* is unknown. There are several techniques to minimize the error function but not all of them are equally efficient.

We have tested two standard methods: 1) a steepest descent approach and 2) a Monte Carlo approach. The first one involves the numerical calculation of the first derivatives of the error function, which makes the method very straightforward but computationally inefficient. Indeed, each step of the steepest descend method leads to the solution that minimizes the error function. However, each step of the calculation requires 10 times more of time due to the calculation of the derivatives. On the contrary, the Monte Carlo

method involves a random search for optimal solutions in the space of parameters which turns out to be faster, although less accurate that the steepest descent method (see Fig. 5.7). So the minimization was finally performed with a Monte Carlo method.

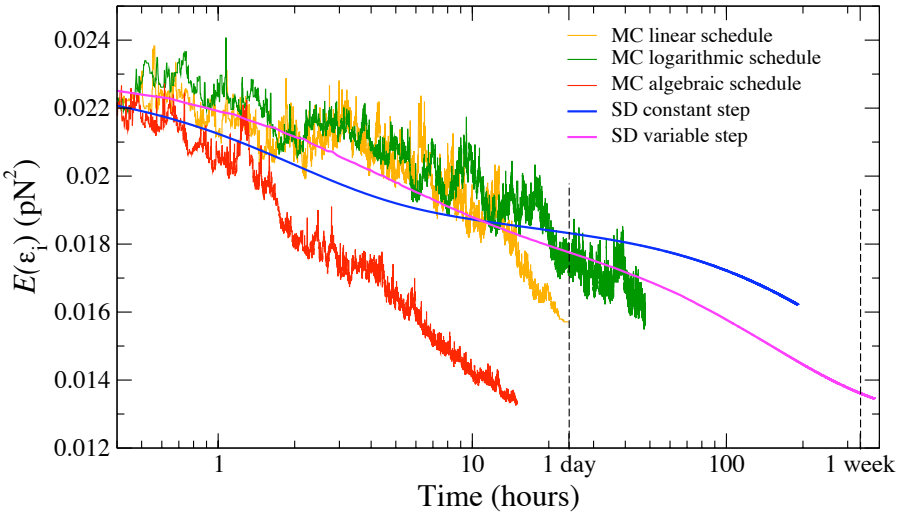


Figure 5.7: Optimization methods. Plot of the error function versus the computation time (MC=Monte Carlo; SD=Steepest Decent). Although the SD methods take less steps, the time spent in each step is much longer because the computation of the derivatives are much time consuming. The MC method allows to define different protocols depending of how the fictive temperature that controls the minimization is varied.

5.3.1 Monte Carlo optimization

In the Monte Carlo (MC) optimization we start from an initial guess for the energies ϵ_i (i=1,...,10,loop) and do a random walk in the space of parameters in order to minimize the error function defined in Eq. 5.2. All MC techniques require the introduction of a fictive temperature that controls how the space of parameters is explored. This method is just a standard Simulated Annealing optimization algorithm [140] adapted to our particular problem that speeds up considerably the time to find the minimum as compared to standard Steepest Descent algorithms. The MC optimization is composed of the following steps:

1. **Initial guess.** The error function is initially evaluated with the UO energies. We start from the unified values to enforce the system to explore the basin of attraction of these values.

2. **Quench algorithm.** The system is allowed to evolve until the total error reaches a minimum. This stage involves several steps. First, a random variation of the NNBP energies is proposed. The set of new energies is accepted if the new calculated value of the error function decreases and the process is repeated. If the error function increases, the set of energies is rejected and a new energy values are proposed. When the acceptance ratio of the proposed steps is lower than 3%, the system is considered to have newly reached a minimum. This method is essentially a steepest descent algorithm with the advantage that the free energy derivatives need not be calculated (see Fig. 5.8a).

3. **Heat-quench algorithm.** Once the system has found the first minimum, we start another MC search to explore other solutions in the vicinity of the minimum. Here we have to define a fictive temperature T to accept or reject the proposed NNBP energies. Each new proposal is accepted or rejected using a Metropolis algorithm. According to the Metropolis algorithm, a move is always accepted if $\Delta E < 0$ (where ΔE is the change in the error after the proposal). If $\Delta E > 0$, a random number r from a standard uniform distribution $r \in \mathcal{U}(0,1)$ is generated and the move accepted if $e^{-\Delta E/T} > r$ and rejected if $e^{-\Delta E/T} < r$.

   During the heat-quench algorithm, the system is heated up to a large fictive temperature until the error is 50% higher than the error of the first minimum. Afterwards, the system is quenched until the acceptance ratio of steps is lower than 3%. This procedure is repeated many times. The multiple solutions found after the initial minimum allow us to estimate the error in the NNBP energies (see Fig. 5.8b). The possible values for the NNBP energies are Gaussian distributed in a region of width approximately equal to 0.05 kcal·mol^{-1} (see Fig. 5.8c).

The error landscape defined by Eq. 5.2 is not rough and there is no necessity to use a MC algorithm. Nevertheless, we find that the MC optimization with the heat-quench scheme is computationally more efficient (i.e., faster) than other optimization algorithms and other fictive temperature schemes (see Figs. 5.7 and 5.9). The analysis revealed that the optimization algorithm is robust and leads to the same solution when the initial conditions are modified, as will be shown in the next section. Different molecules and different sequences converge to energy values that are clustered around the same value. Appendix L describes in further details the errors in the MC optimization.

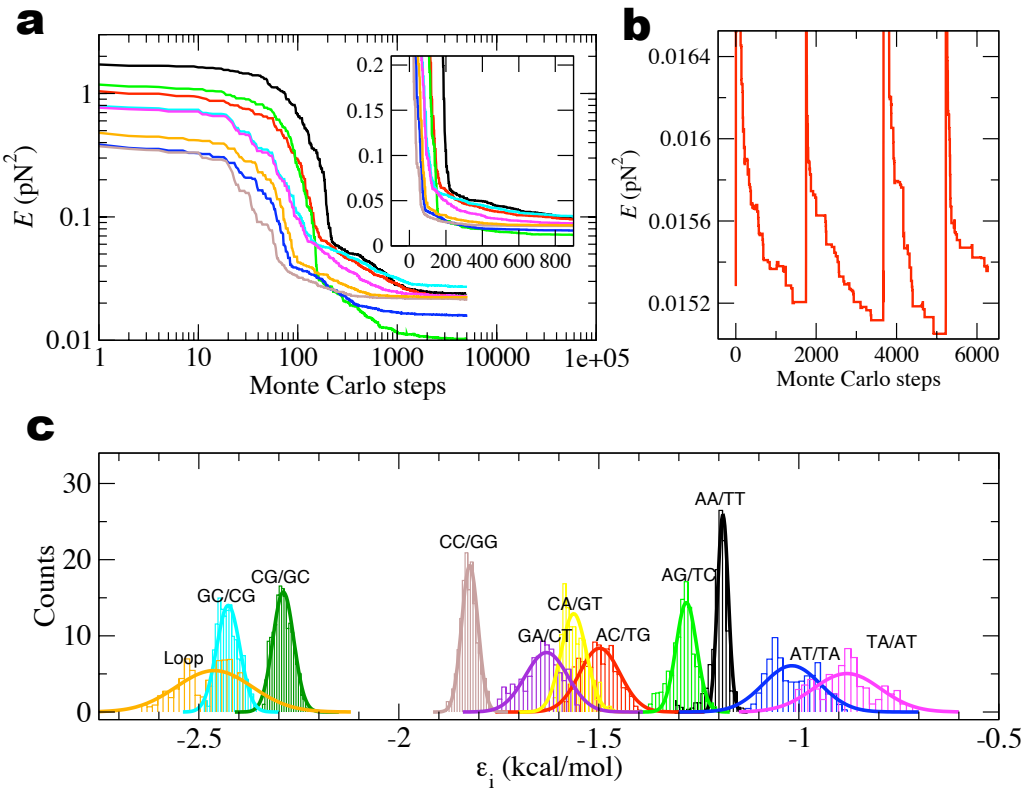


Figure 5.8: Monte Carlo optimization. **(a)** Quench algorithm. Evolution of the error function of different molecules during the quenching minimization. The main figure shows a log-log plot where the mean quadratic error decreases down close to 0.01 pN^2. The inset figure shows a linear plot of the same evolution. **(b)** Heat-quench algorithm. Evolution of the error during the heat-quench algorithm. **(c)** Histograms of solutions for one representative molecule obtained using the heat-quench algorithm. Each color represents one NNBP parameter and its Gaussian fit profile. Optimal solutions correspond to the most probable values of the distribution.

5.3.2 Independence of initial conditions

The optimization algorithm gives the same solution within 0.05 $kcal \cdot mol^{-1}$ when the initial conditions are modified. A detailed study of the optimization algorithm has been carried out in order to check that the final solution does not depend on the initial conditions given to the algorithm.

An ensemble of initial conditions have been selected from the values of the different labs that were unified by SantaLucia [28] (see Fig. 5.10a). The FDCs predicted by the different energy values are depicted in Fig. 5.10c. Here we can observe how an overestimation (underestimation) in the absolute value of the NNBP energies leads to an overestimation (underestimation) of the

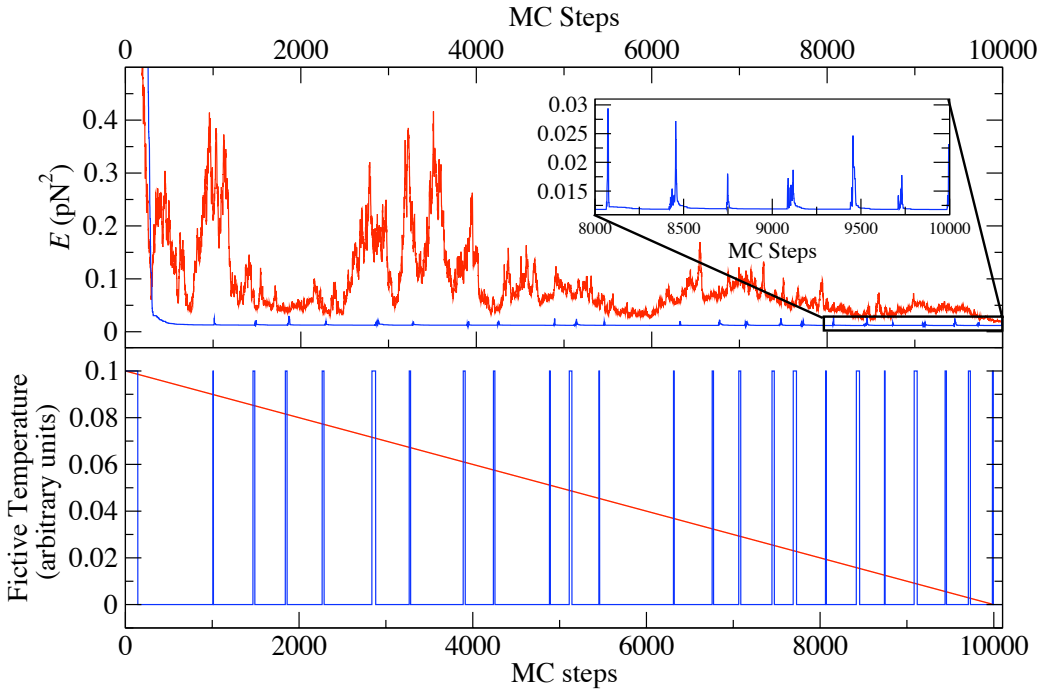


Figure 5.9: Temperature schedules. Upper panel shows the evolution of the error for a linear schedule (red curve) and a heat-quench algorithm (blue curve). The inset shows a zoomed region of the heat-quench algorithm. Note the rapid error decrease of the heat-quench algorithm. The lower panel shows the evolution of the temperature for each schedule.

mean unzipping force. The same elastic properties (ssDNA, handles and optical trap) have been used in all cases. Figure 5.10b shows the optimal values of the NNBP energies for the ensemble of initial conditions. All the NNBP energies have an error smaller than 0.1 kcal·mol^{-1}.

The optimization algorithm converges to essentially the same solution when starting from various initial conditions because the error bars of the different solutions obtained for each initial condition overlap with each other. Figure 5.10d shows the final theoretically predicted FDCs obtained when using the optimal values of the NNBP energies obtained for each initial condition. The different FDCs are indistinguishable and they reproduce quantitatively the experimental FDC.

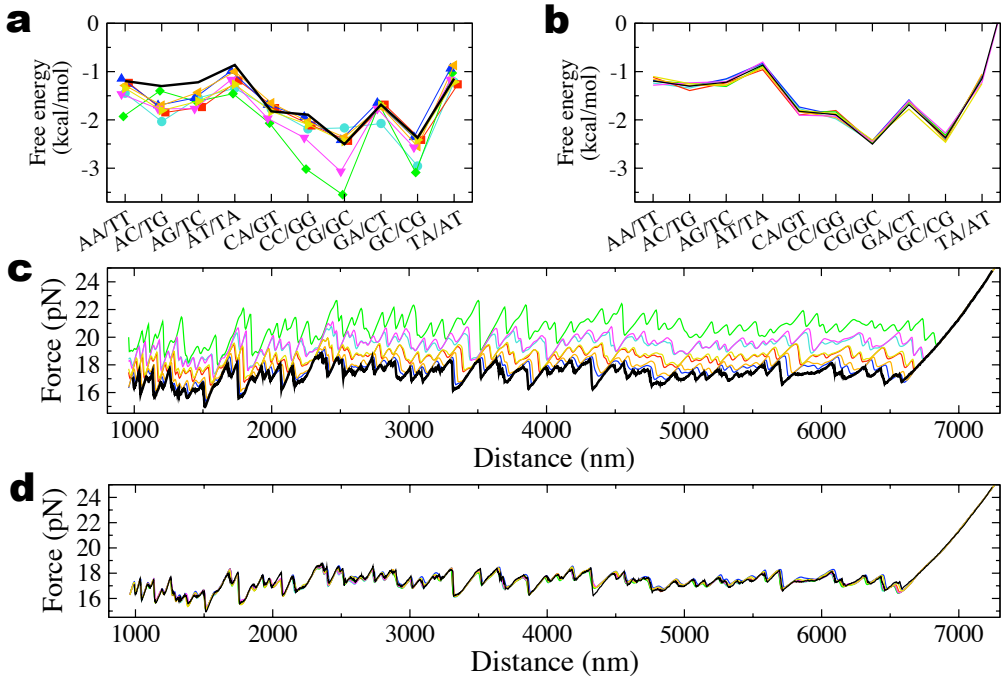


Figure 5.10: Dependence on initial conditions taken from bulk measurements [28]. **(a)** Ensemble of initial conditions that we tested. Data was obtained from SantaLucia [28] and it was corrected according to temperature (298 K) and salt conditions (1 M [NaCl]) of our experiments. Cyan points, values of Benight; red points, values of Blake; green points, values of Breslauer; blue points, values of Gotoh; orange points, values of SantaLucia; magenta points, values of Sugimoto; yellow points, values of Vologodskii; black curve, our fit values. **(b)** Solutions found after the fitting algorithm for the different initial conditions. Same color code as in panel a. The heterogeneous ensemble of initial conditions has converged to similar values for all the NNBP energies that differ by less than 0.1 kcal·mol^{-1}. **(c)** FDCs obtained using the NNBP energies from the different initial conditions. The color code is the same as in panel a. The black curve shows our fit FDC. Some of the initial conditions (Gotoh and SantaLucia) are compatible with the experimental FDC. **(d)** FDCs obtained using the optimal NNBP energies obtained for each initial condition. Experimental and optimal FDCs differ by less than 0.1 pN throughout the molecule.

5.4 Salt dependence and thermodynamic properties of NNBP interactions

Up to now, we have seen how to extract the NNBP energies from an unzipping FDC. This section describes how this process can be repeated for several molecules and several salt conditions in order to extract conclusions about the salt dependence of the NNBP free energies, enthalpies and entropies.

The unzipping experiments have been performed at 7 different salt conditions for the 6.8 kbp sequence and at two salt conditions for the 2.2 kbp

sequence. All them are quasistatic pulling experiments from which the equilibrium FDC can be obtained (see Fig. 5.11). At each salt condition, the experiment has been repeated for several molecules. The Monte Carlo optimization method has also been applied to fit all the collected experimental data to the model. The result is a collection of 10 NNBP energies (plus the loop contribution) for each salt condition and for each molecule. The error can be estimated from the averages between the individual molecules measured at each salt.

The best-fit energy parameters reduce the error between the measured FDC and the theoretical prediction (see Fig. 5.12a,b). That is, they give improved agreement between the experimental and the theoretical unzipping FDCs as compared with the UO values. Figures 5.12c,d show the average value and the standard error of the NNBP obtained for the 6.8 kbp sequence and the UO prediction for the NNBP energies at 10 mM and 1 M [NaCl]. It is interesting that some of the new values are in good agreement with the results given by SantaLucia [28] (e.g., CA/GT and AT/TA motifs) while others differ significantly (e.g., AA/TT and GA/CT at 10 mM NaCl and AC/TG and CC/GG at 1 M NaCl).

According to the UO salt correction, the NNBP energies are extrapolated homogeneously (i.e., the same salt correction is taken for all base-stack combinations) from standard salt conditions (1 M [NaCl]) down to lower salt concentrations (e.g., 50 mM) (see Sec. 3.3.3). However, such correction does not predict the observed unzipping force at low salt, especially for certain NNBP such as AA/TT or GA/CT.

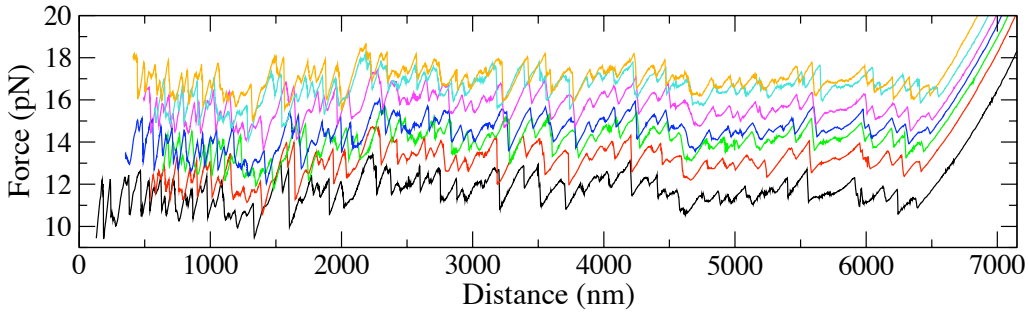


Figure 5.11: FDCs at various monovalent salt concentrations ([Mon$^+$]=[Na$^+$]+[Tris$^+$], [Tris$^+$]=18 mM). Black curve [Na$^+$]=10 mM, red curve [Na$^+$]=25 mM, green curve [Na$^+$]=50 mM, blue curve [Na$^+$]=100 mM, magenta curve [Na$^+$]=250 mM, cyan curve [Na$^+$]=500 mM, orange curve [Na$^+$]=1 M.

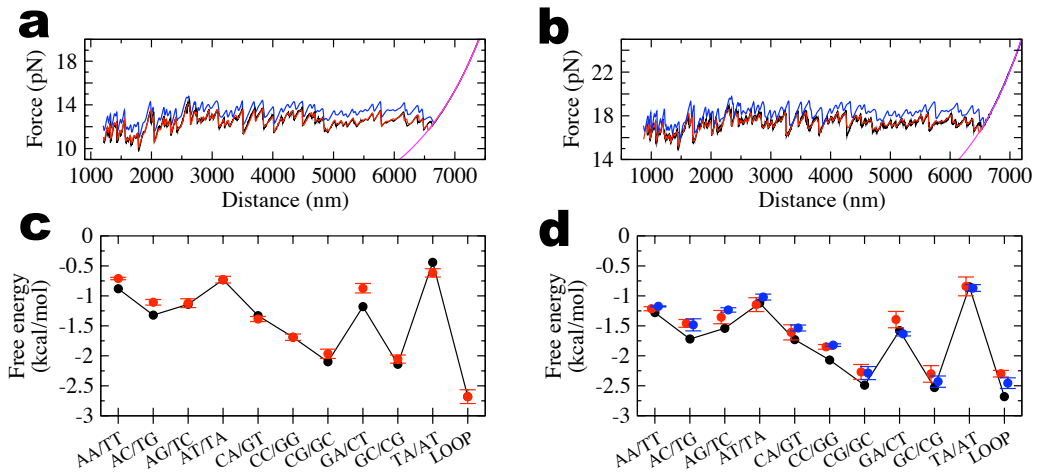


Figure 5.12: Salt dependencies. **(a,b)** FDCs for the 6.8 kbp sequence at 10 mM NaCl (panel a) and 1 M NaCl (panel b). Black curve, experimental measurements; blue curve, UO prediction; red curve, our fit; magenta curve, elastic response of the fully unzipped molecule. The theoretical FDC is calculated in equilibrium, which assumes that the bandwidth is 0 Hz and the experimental data is filtered at bandwidth 1 Hz. If data is filtered at higher frequencies (>1 Hz), hopping between states is observed and the experimental FDC does not compare well with the theoretical FDC at equilibrium. If data is filtered at lower frequencies (<1 Hz), the force rips are smoothed and hopping transitions are averaged out. **(c,d)** NNBP energies and comparison with UO values at 10 mM [NaCl] (panel c) and 1 M [NaCl] (panel d). The following notation is used for NNBP: AG/TC denotes 5′-AG-3′ paired with 5′-CT-3′. Black points, UO values; red points, values for the 6.8 kbp molecule; blue points, values for the 2.2 kbp molecule. The values for the 6.8 kbp and the 2.2 kbp molecules have been obtained after averaging over six molecules. Error bars are determined from the standard error among different molecules.

A heterogeneous (sequence specific) salt correction could provide consistent results with the experiments. Such deviations are not unexpected, given the differences in solvation between specific nucleotides and salt ions [141, 142]. However the effect has never been quantified. With this goal in mind, we apply the fitting algorithm to extract NNBP energies for data taken at many salt concentrations (see Fig. 5.13 red points). We find compatible NNBP energies between the two molecules (Fig. 5.13 blue dots).

As mentioned in Sec. 3.3.3 the UO prediction uses a non-specific salt correction for the different NNBP energies. Equation 3.13 gives the entropy correction of the NNBP for salt conditions different from 1 M [NaCl], which can be combined with Eq. 3.5 to write and expression for the free energy of formation of the NNBP at any salt concentration according to:

$$\epsilon_i([\mathrm{Mon}^+]) = \epsilon_i^0 - m \cdot \ln([\mathrm{Mon}^+]) \tag{5.3}$$

where $\epsilon_i([\mathrm{Mon}^+])$ is the energy of formation of the i th NNBP (i=1,...,10)

at a monovalent salt concentration of [Mon^+] (expressed in molar units), ϵ_i^0 is the NNBP energy at 298 K, 1 M monovalent salt and m is the non-specific prefactor equal to $m = 0.110$ kcal·mol^{-1} at 298 K (see Fig. 5.13, green lines) [28, 118]. The monovalent salt concentration [Mon^+] accounts for the total amount of ions with charge +1 that are in the buffer. In the case of the TE buffer at which the unzipping experiments are performed, the total concentration of monovalent ions is given by two contributions: the [Na^+] ions and the [$Tris^+$] ones, which is [$Tris^+$] = 18 mM at pH = 7.5.

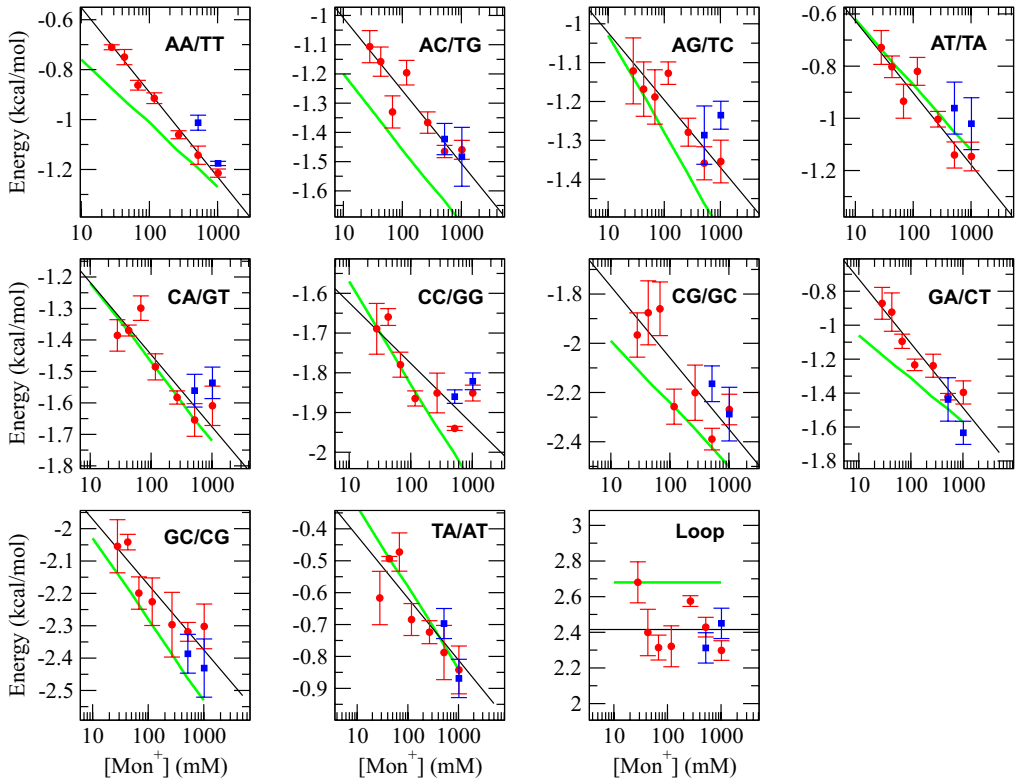


Figure 5.13: Salt corrections of the NNBP energies. Each panel shows the energy of a different NNBP parameter. Red (blue) points are the experimental results for the 6.8 kbp (2.2 kbp) sequence; green curve, UO non-specific salt correction; black curve, fit to Eq. 5.3 with adjustable parameters m_i (i=1,...,10,loop) and ϵ_i^0.

To define a heterogeneous salt correction within this scheme, it is only necessary to establish 10 sequence-specific prefactors m_i to be used with the same logarithmic dependence as shown in Eq. 5.3. Thus we fit all NNBP energies using NNBP-dependent parameters m_i (i=1,...,10,loop) and ϵ_i^0. The fits to each NNBP are shown in Fig. 5.13 (black lines) and the resulting fit parameters (ϵ_i^0 and m_i) are listed in Table 5.2. There we observe that the salt de-

NNBP	6.8 kbp	2.2 kbp	UO	ϵ_i^0	m_i
AA/TT	-1.21 (0.02)	-1.18 (0.01)	-1.27	-1.23 (0.01)	0.145 (0.006)
AC/TG	-1.46 (0.04)	-1.48 (0.10)	-1.71	-1.49 (0.05)	0.10 (0.02)
AG/TC	-1.35 (0.07)	-1.24 (0.04)	-1.53	-1.36 (0.03)	0.070 (0.014)
AT/TA	-1.15 (0.06)	-1.02 (0.05)	-1.12	-1.17 (0.04)	0.12 (0.02)
CA/GT	-1.61 (0.07)	-1.54 (0.05)	-1.72	-1.66 (0.05)	0.09 (0.02)
CC/GG	-1.85 (0.02)	-1.82 (0.02)	-2.08	-1.93 (0.04)	0.06 (0.02)
CG/GC	-2.27 (0.06)	-2.29 (0.10)	-2.50	-2.37 (0.09)	0.13 (0.04)
GA/CT	-1.40 (0.07)	-1.63 (0.04)	-1.57	-1.47 (0.05)	0.15 (0.02)
GC/CG	-2.30 (0.06)	-2.43 (0.10)	-2.53	-2.36 (0.04)	0.08 (0.02)
TA/AT	-0.84 (0.08)	-0.87 (0.06)	-0.84	-0.84 (0.05)	0.09 (0.02)
Loop	2.30 (0.06)	2.46 (0.09)	2.68	2.43 (0.05)	-

Table 5.2: 6.8 kbp (2.2 kbp) are the NNBP energies (in kcal·mol^{-1}) obtained from the averaged results for the two sequences (standard error in parenthesis) at 1 M [NaCl]. UO are the values extracted from ref. [28]. ϵ_i^0 and m_i are the standard energies and prefactors obtained from the fits shown in Fig. 5.13. These values are used in Eq. 5.3 to extrapolate the NNBP energies to other salt concentrations.

pendence of some NNBP parameters is well described by the UO non-specific correction (e.g., AT/TA and CA/GT) but most of them are better fit with some correction in parameters ϵ_i^0 and m_i (e.g., AA/TT,AC/TG,AG/TC).

Interestingly enough, a fit of the previous data with homogeneous salt correction (i.e., one single value m for all the NNBP energies) gives a value of $m = 0.104$ kcal·mol^{-1}, which is very similar to the value reported by SantaLucia [28]: $m = 0.114$ kcal·mol^{-1}. However, the root mean square error of a homogeneous fit is twice the error of a heterogeneous fit. Already with naked eye, it is possible to observe clear discrepancies in the slopes of some NNBP motifs (such as AA/TT, AG/TC, GA/CT) with a homogeneous salt correction (green curves in Fig. 5.13). Therefore, the use of a heterogeneous salt correction is worth, given the improvement in the fit. This is discussed in Sec. 5.5.

5.4.1 Inference of NNBP enthalpies and entropies

The unzipping experiments provide direct measurements of the free energies (ϵ_i) and the salt corrections (m_i) at $T = 298$ K for all the NNBP motifs ($i = 1, \ldots, 10$), but no information about the enthalpy (Δh_i) and the entropy (Δs_i) is provided. In order to infer these two magnitudes, we should perform more unzipping experiments at different temperatures an apply Eq. 3.5. At

present this cannot be achieved with the minitweezers experimental setup because the temperature cannot be controlled at will. The changes in temperature dramatically affect the optics of the instrument, which introduces undesirable drift effects that compromise the resolution of the experiment.

However, combining the results from unzipping experiments with the measurements of melting temperatures of several oligos obtained by optical melting experiments [143] we can infer the enthalpies and the entropies. In order to do so, we define an error function χ^2 that accounts for the mean squared error between the experimental melting temperatures [143] ($T_i^{\rm exp}$) and the predicted ($T_i^{\rm pred}$) ones for $N = 460$ different oligos and salt conditions,

$$\chi^2(\Delta h_1, \ldots, \Delta h_{10}) = \frac{1}{N} \sum_i^N \left(T_i^{\rm exp} - T_i^{\rm pred}(\Delta h_1, \ldots, \Delta h_{10}) \right)^2 \qquad (5.4)$$

where Δh_i ($i = 1, \ldots, 10$) are the NNBP enthalpies and $T_i^{\rm pred}$ are obtained according to Eq. 3.14 derived from the NN model. The NNBP entropies are fixed by 10 constraints that relate the free energies, the enthalpies and the entropies according to

$$\epsilon_i = \Delta h_i - T\Delta s_i \qquad \Longrightarrow \qquad \Delta s_i = \frac{\Delta h_i - \epsilon_i}{T} \qquad (5.5)$$

where $i = 1, \ldots, 10$; ϵ_i ($= \Delta g_i$) are the experimentally measured free energies with unzipping experiments and $T = 298$ K. Here, the enthalpies are fitting parameters that fix the entropies. Therefore the enthalpies and the entropies are fully correlated (their correlation coefficients are equal to 1).

The error function (Eq. 5.4) is minimized with respect to the 10 enthalpies using a steepest descent algorithm that rapidly converges to the same solution when starting from different initial conditions. Table 5.3 contains the best fit values for the enthalpies and the entropies which are compatible with the UO values. Appendix M describes how the error of the fitting parameters was estimated.

5.4.2 Prediction of melting temperatures

The aim of this section is to check how well the new thermodynamic values (see Table 5.3) work to predict the melting temperatures of oligonucleotides under various salt conditions. However, it has to be admitted that the confrontation of our results with the melting temperature of oligos is not the best way to show the reliability of the new NNBP energies. The reason is that most published sets of NNBP values obtained with different calorimetric or optical methods agree within 2°C (in average), revealing that the melting

Method	UO values				Force measurements			
NNBP	ϵ_i	Δh_i	Δs_i	m_i	ϵ_i	Δh_i	Δs_i	m_i
AA/TT	-1.28	-7.9	-22.2	0.114	-1.23	-7.28 (0.3)	-20.28 (1.2)	0.145
AC/TG	-1.72	-8.4	-22.4	0.114	-1.49	-5.80 (0.3)	-14.46 (1.3)	0.099
AG/TC	-1.54	-7.8	-21.0	0.114	-1.36	-5.21 (0.3)	-12.89 (1.2)	0.070
AT/TA	-1.12	-7.2	-20.4	0.114	-1.17	-4.63 (0.6)	-11.62 (2.1)	0.117
CA/GT	-1.73	-8.5	-22.7	0.114	-1.66	-8.96 (0.3)	-24.48 (1.2)	0.091
CC/GG	-2.07	-8.0	-19.9	0.114	-1.93	-8.57 (0.3)	-22.30 (1.2)	0.063
CG/GC	-2.49	-10.6	-27.2	0.114	-2.37	-9.66 (0.5)	-24.43 (2.1)	0.132
GA/CT	-1.58	-8.2	-22.2	0.114	-1.47	-8.16 (0.3)	-22.46 (1.3)	0.155
GC/CG	-2.53	-9.8	-24.4	0.114	-2.36	-10.10 (0.5)	-25.96 (1.8)	0.079
TA/AT	-0.85	-7.2	-21.3	0.114	-0.84	-8.31 (0.6)	-25.06 (2.1)	0.091

Table 5.3: Free energies and enthalpies given in kcal·mol^{-1}; entropies given in cal·mol^{-1}·K^{-1}. Left block of columns show the UO values at standard conditions (25°C and 1 M [NaCl]). Right block of columns show the thermodynamic values inferred from our unzipping experiments. The values in parenthesis indicate the estimation of the error (see Appendix M).

temperature is not necessarily a very robust indicator. In fact, the fit of the melting temperatures in Eq. 5.4 provides a better estimation for the melting temperatures than for enthalpies and entropies, which are the relevant magnitudes. The melting temperature is probably not the best estimator, but we do not have any other experimental observable with which we can infer and compare our results. In the end, we are inferring the important magnitudes (enthalpies and entropies) by means of a derived magnitude (melting temperature) that is compatible with large variations of such important magnitudes (enthalpies and entropies).

Moreover, there seems to be a circular reasoning here. If our thermodynamics values (enthalpies and entropies) are obtained after fitting the melting temperatures, our thermodynamic values will obviously predict the melting temperatures well. However, our fit does not necessarily lead to good melting temperature predictions (i.e., within 2°C). Indeed, our results might have led to a solution where the root mean squared error in Eq. 5.4 was much higher than the accepted experimental error in melting experiments (2°C). Fortunately, this is not the case and our thermodynamic values predict the melting temperatures of ref. [143] with an average error of 1.90°C.

What is more, we have checked that if our found NNBP energies (ϵ_i) increase only by 0.15 kcal·mol^{-1} (i.e., slightly beyond the standard error of the NNBP values) the standard deviation error for temperature melting prediction goes from 2°C up to $5-6$°C. So, if we had obtained the NNBP energies

with a systematic error of 0.15 kcal·mol^{-1} we could never have predicted the melting temperatures within the 2°C accuracy.

Once these preliminary issues have been addressed, let us focus on the prediction of the melting temperatures. Using our new thermodynamic values (Table 5.3 and Eq. 3.14), we calculate the melting temperatures of 92 oligos at 5 different salt conditions and compare our results with melting data taken from Owczarzy *et al.* [143]. Figure 5.14a illustrates how, in general, the UO values predict the melting temperatures better (average error of 1.53 °C) than our values (average error of 1.90 °C). The good news here is that our results are compatible and agree with the melting experiments. A deep look into the results reveals that our values work better for oligos longer than 15 bp (see Fig. 5.14b). Indeed, the UO prediction has an average error of 1.5 °C, while our values, 1.3 °C. All these results and their explanations will be discussed in the next section.

5.5 Discussion of results

5.5.1 Regarding the heterogeneous salt correction

The single molecule force unzipping experiments allowed us to extract the DNA base pair free energies at various salt concentrations. The heterogeneous salt correction found in our measurements is a remarkable result and it could be due to electrostatic effects. The NNBP formed as a combination of purine-purine or pyrimidine-pyrimidine (5′-YY-3′ or 5′-RR-3′, i.e., AA/TT, AG/TC, CC/GG, GA/CT) differ most from the UO homogeneous salt correction than mixed purine-pyrimidine combinations (5′-RY-3′ and 5′-YR-3′). A difference between these combinations can be observed in how charges (e.g., hydrogen bond acceptor and donor groups) are distributed along the major groove of the double helix. The latter have charged groups that tend to be uniformly distributed between the two strands along the major groove, whereas the former have donor and acceptor groups unevenly distributed between the two strands. The specific salt correction found in our measurements could be consequence of how monovalent cations bind the two strands along the major groove. There are precedents to such results. Sugimoto and collaborators [144] have reported that cation binding is correlated to duplex stability. Computer simulations have identified acceptor groups in guanine ($N_{7'}$,$O_{6'}$) and adenine ($N_{7'}$) as preferential cation binding sites [145]. Our experimentally determined specific salt corrections might be interpreted as a corroboration of such hypothesis.

An alternative explanation could be that the heterogeneity of the salt cor-

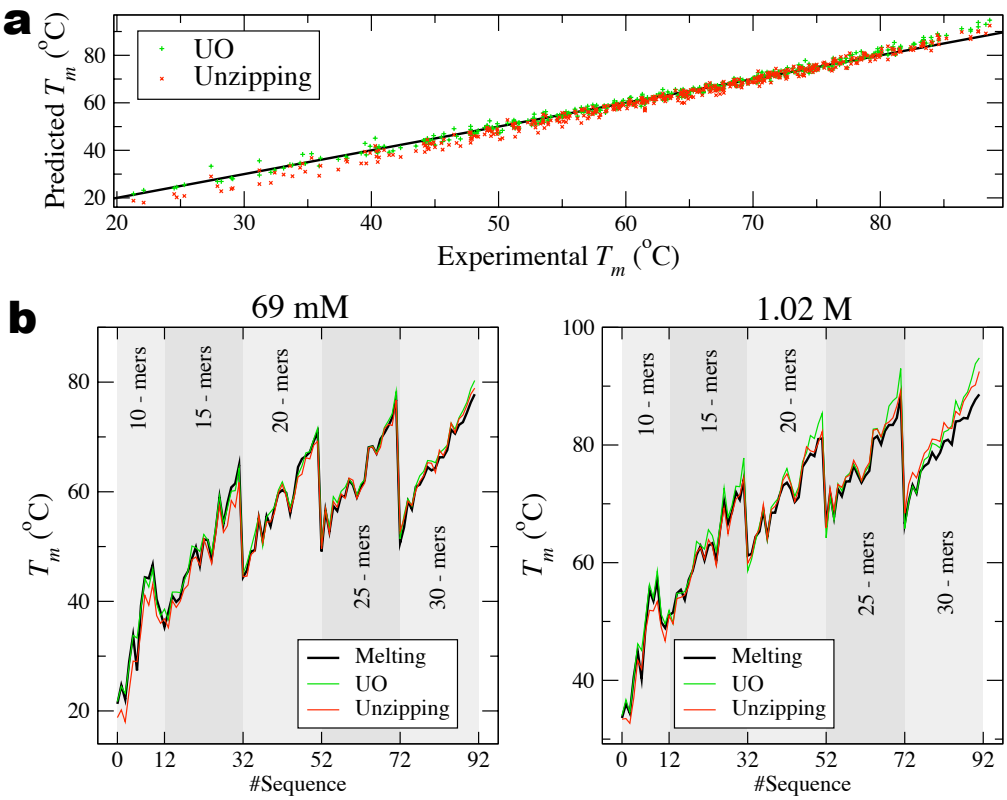


Figure 5.14: Prediction of melting temperatures. Comparison of melting temperatures for the 92 oligos ranging from 10 to 30 bp reported in ref. [143]. **(a)** Predicted versus experimentally measured melting temperatures at 5 salt conditions ($[Na^+]$ = 69, 119, 220, 621, 1020 mM). The values obtained from unzipping have less error at higher temperatures (corresponding to longer oligos). **(b)** Prediction at 69 mM [NaCl] (left panel) and 1.02 M [NaCl] (right panel). Black lines are the experimentally measured melting temperatures, green line is the UO prediction and red line our prediction from unzipping data. Note that unzipping experiments predict the melting temperatures better above 15 bp, while the UO prediction works fine below this length.

rection is consequence of sequence specific elastic properties of the ssDNA. Previous studies of ssDNA elasticity [48] have shown how the contour length of the ssDNA gradually increases from low to high forces suggesting a conformational transition of the sugar pucker that goes from the A-form ($C_{3'}$-endo) to the B-form ($C_{2'}$-endo) at high forces (see Sec. 3.1 and Fig. 3.4c for a detailed description of the sugar puckering). A similar phenomenon has been reported in recent studies of homopolymeric RNA sequences [138] that reveal conformational transitions under tension (in the form of force plateaus) in poly-C and poly-A sequences but not for poly-U sequences (studies for poly-G sequences were not available). For random ssDNA sequences one might

expect that such specific effects are averaged out and a gradual A→B transition is observed. Conformational transition effects are also observed in the elastic measurements of ssDNA when varying the salt. As shown in Table 5.1, the value of the interphosphate distance increases when decreasing the salt (it changes from 0.59 nm at high salts to 0.67 nm at low salts) suggesting a similar sugar pucker (B→A) conformational transition for the ssDNA. How this consideration could affect the interpretation of the heterogeneous salt correction? In other words, might it be possible that the contribution of each pair of stacked bases to the Kuhn length is different? This result has never been reported before in any experimental study of the ssDNA elasticity [48, 139]. It represents a completely different interpretation of the salt dependence and there is no reason why such hypothesis should be rejected. In order to determine the dependence of the Kuhn length with the sequence, it would be necessary to perform a detailed study of the elastic properties of periodic ssDNA sequences along the lines of what has been recently reported for homopolymeric RNA sequences [138]. A detailed investigation should be carried out for homopolymeric DNA sequences containing all possible combinations of stacking bases. Candidate sequences should be poly-dA, poly-dC, poly-dT and other periodic sequences such as poly-(dA-p-dT), poly-(dA-p-dT-p-dC). In the end, the Kuhn length associated to each combination of stacked bases (16 in total) could be written as a constant value plus a base (or sequence) dependent correction. Although an exhaustive research of the elastic response of different ssDNA sequences would shed light into this issue, this is beyond the scope of the present work.

5.5.2 About the prediction of melting temperatures

Why do our free energy numbers predict fairly well melting temperatures of oligos longer than 15 bp but do worse for shorter ones? Discrepancies between predicted and measured melting temperature for short oligos have been already reported in bulk measurements [146] and attributed to differences in analytical methods used to extract melting temperatures. For instance, there are a few definitions of the melting temperature that not always coincide: maximum heat capacity of the sample, maximum derivative of UV absorbance, etc. Another possible explanation is that short oligos (< 15 bp) might not have the double helix perfectly formed and the formation energies involved in the duplex are slightly different from the energies for longer sequences (i.e., typical boundary effects of small systems). Although there is no conclusive answer to this question, it is worth underlining that UO free energy values are obtained in order to correctly predict the melting

temperatures for all oligo lengths. This might lead to error compensation between the melting temperature data sets corresponding to short and long oligos. Let us stress that with increasing length, the prediction of melting temperatures is more tolerant of errors in the details of the NNBP energies, where sequence effects are averaged out. In addition, deviations from the bimolecular model (Eq. 3.14) arise for sequences with $n \geq 20$, as their melting process begins to shift toward pseudo-monomolecular behavior [120]. Still, our predicted melting temperatures for oligos with $n \geq 20$ agree well for the sequences reported in ref. [143].

5.6 Conclusions

The mesoscopic model (that uses the NN model) with the UO energies qualitatively describes the FDC of DNA unzipping experiments. At a quantitative level, the UO energies overestimate the mean unzipping force, especially at low salt concentration.

In order to reach a quantitative agreement between the experiments and the mesoscopic model, it has to be completed with: 1) an accurate model for the ssDNA elasticity, 2) an empiric model to correct the experimental drift and 3) a specific contribution for the end loop.

The NNBP energies can be fit to match the mesoscopic model with the experimental FDC. A Monte Carlo fit of the data gives rapid and accurate estimations of the NNBP energies. The fit is robust and provides acceptable confidence intervals for the energies.

The extension of the experiments and the Monte Carlo fit to several molecules and salt concentrations provides valuable information about the salt dependence of the NNBP energies. A homogeneous salt correction does not describe the results found. Instead, a sequence-specific salt correction offers a more trustworthy rule.

The heterogeneous salt correction might be justified by two different explanations: 1) the differences in solvation between the specific nucleotides and monovalent ions at different salt concentrations or; 2) the differences in the elastic response of the bases of the ssDNA at different salt concentrations. Specific experiments should be carried out to discern between the two possibilities.

The unzipping experiments do not provide the enthalpies and the entropies of formation of the NNBP elements. However, the unzipping and melting experiments can be combined to obtain an estimation of these thermodynamic magnitudes.

Although melting and unzipping experiments are based on disruption

processes triggered by different external agents (temperature and force respectively), the agreement between the thermodynamic magnitudes is remarkable. Both types of experiments are correctly described by the NN model.

The thermodynamic magnitudes obtained with the unzipping experiments not only predict the melting temperatures well but also improve the melting temperature prediction as compared to the UO prediction for oligos longer than 15 bp. The misprediction of melting temperatures for short oligos might be due to underestimated boundary effects.

The unzipping experiments provide an alternative determination of the NNBP parameters in which the folding/unfolding transition does not need to be two-state. Besides, instead of several short oligos of different sequence, one long molecule is sufficient to infer the NNBP energies.

The main limitations of the method exposed are the strong dependence of the results on the elastic response of the ssDNA and the impossibility to determine the bimolecular initiation factors.

The unzipping approach can be extended to extract free energies, entropies and enthalpies in DNA and RNA structures under different solvent and salt conditions. The method can be also applied to extract free energies of other structural motifs in DNAs (e.g., sequence dependent loops, bulges, mismatches, junctions). The enthalpies and the entropies can be directly extracted from unzipping experiments performed at different temperatures. The force methods make possible to extract free energies in conditions not accessible to bulk methods (such as the dsRNA that hydrolyzes in the presence of magnesium in melting experiments). Another extension of the method can be used to obtain the binding free energies of DNAs and RNAs bound to proteins, where the proteins denaturalize below the dissociation melting transition. Finally, this method could be also useful in cases where molecular aggregation and other collective effects in bulk preclude accurate free energy measurements.

This methodology can serve as a basis to search for long-range context effects (e.g., 2nd-nearest and 3rd-nearest interactions) in DNA. This might be especially interesting in those regions along the sequence where the NN model seems to fail.

Summing up, this chapter establishes a novel methodology to obtain thermodynamic information from single-molecule experiments.

When in doubt, use brute force.

Ken Thompson (1943)

Chapter 6

Unzipping of DNA at controlled force

A significant group of techniques in biophysics are capable of applying forces to individual molecules. Some of them directly control the force such as the magnetic tweezers. Some others apply forces by means of indirect control parameters such as the optical tweezers. Apparently, there should be no difference between both kinds of techniques. However, the control parameter determines what magnitude is fixed (force or extension) and what magnitude fluctuates (extension or force, respectively), while establishing the type of experiments that can be performed. Recent theoretical studies have shown the inequivalence between statistical ensembles for the elasticity of polymers [123]. The question that emerges here is whether this inequivalence can be observed experimentally.

Detailed studies of DNA unzipping at Controlled Force (CF) have been carried out with magnetic tweezers [23]. The results show how unzipping is always a non-equilibrium process with dissipation of energy that leads to large hysteresis cycles [147]. These experiments have been accompanied with theoretical studies of the process that shed light into the phenomenon [29]. On the other side, the unzipping experiments at Controlled Position (CP) [24, 148, 25] have been capable of obtaining more information from the internal structure of the DNA molecule. So, the published works on DNA unzipping use different experimental techniques, different molecular constructs and different instruments. How the unzipping processes compare

when the only difference is in the choice of the control parameter?

This chapter describes how optical tweezers combined with a high bandwidth force feedback are capable of measuring the FDCs of the same molecular construct at CP and CF. The experimental data measured at CF can be analyzed with the tools developed up to now. In the end, we are able to directly compare the differences between one type of experiments and the other and determine the origin of the inequivalence of ensembles.

The work presented in this chapter is exploratory and descriptive and the obtained results have to be considered preliminary. Nevertheless, they raise interesting questions about DNA unzipping, disordered systems and non-equilibrium thermodynamics. The open questions exposed here will be answered in the future with more experiments and calculations.

6.1 Controlled force experiments in optical tweezers

The control of force in optical tweezers is usually achieved by two different methods. The first one consists in setting up a feedback that corrects the position of the optical trap so that the position of the bead with respect to the center of the optical trap (and so the force) is always constant in average (see Fig. 3.14b). This method requires the use of analog or digital electronic feedbacks that always have a limited bandwidth. The bandwidth of a feedback is measured in Hertz and it indicates the maximum number of times the position of the trap can be corrected per unit of time. For instance, a force feedback of 100 Hz indicates that the position of the optical trap can be corrected 100 times per second. Ideally, the CF experiments should be done at infinite bandwidth.

The second method is called *force-clamp* and it was developed by Block and coworkers [149]. It consists in locating the bead in the region of the optical trap where it is about to escape. The bead undergoes a constant force in a local region with zero stiffness, which is an effective infinite bandwidth force feedback. The main disadvantage of this method is that the force is constant only on short ranges of extensions (about 100 nm). So the method is useful to exert forces on short DNA hairpins (< 100 bp) but it is not possible to perform unzipping experiments of long DNA molecules because the opening of base pairs moves the bead away from the region of zero stiffness.

There is a third technique not developed yet that consists in forming a uniform gradient of laser light in a confined region where the bead would always undergo the same force. Apart from the TEM_{00}, other laser modes such

as Laguerre-Gaussian [72] or Bessel [150] have been used to trap particles. However, a practical laser beam with a uniform gradient of light intensity has never been produced.

The experiments described in the present chapter are performed with the first method, i.e., with a finite bandwidth force feedback.

6.1.1 Force feedback

The force feedback of the instrument follows a proportional protocol (see Eq. C.2) that corrects the position of the optical trap to keep the force constant. The feedback runs at 4 kHz but the natural bandwidth of the wiggler is about $\sim$ 2 kHz [19]. It means that the electronic controller can give orders faster than the time spent by the wiggler to reposition the optical trap. Strictly speaking, the bandwidth of the wiggler limits the chances to perform unzipping at CF. In addition, the thermal force spectrum of a polystyrene bead surrounded by distilled water in a trap stiffness of $\sim$75 pN$\cdot\mu$m^{-1} has a corner frequency of about 1 kHz. Our force feedback has a similar bandwidth and it is not capable of canceling out the noise of the thermal forces. Nevertheless, we are not really interested in having the thermal forces constant. Instead, we want to control the force when the opening of base pairs is produced during unzipping, which is a phenomenon that occurs at a time scale of 1 s (=1 Hz). Because of the difference in time scales, the force feedback is useful to perform our experiments.

Figure 6.1 shows the typical measurements in a CF unzipping experiment performed on a 2.2 kbp sequence. These experiments are characterized by the loading rate, which is 0.05 pN$\cdot$s^{-1} in this particular example. The force is the control parameter that increases monotonically while the distance evolves according to the opening of base pairs of the DNA molecule. We compare the new measurements with a previous ones with controlled trap position.

6.1.2 Unzipping/Rezipping cycles at controlled force

The optical tweezers instrument is capable of producing DNA unzipping and rezipping at CF within an acceptable bandwidth. Figure 6.2 shows a full cycle of unzipping/rezipping of a 2.2 kbp sequence. The experiment is ready to start when the DNA molecular construct is stretched between two beads in the optical tweezers. Depending on how much the DNA is frayed (fraying is the melting of DNA in the extremities), the molecule might have some open base pairs at the beginning. In this particular case, we set the unzipping fork at a region in which the molecule has about $\sim$100 open base pairs. When the pulling protocol starts, the force increases at the indicated loading rate.

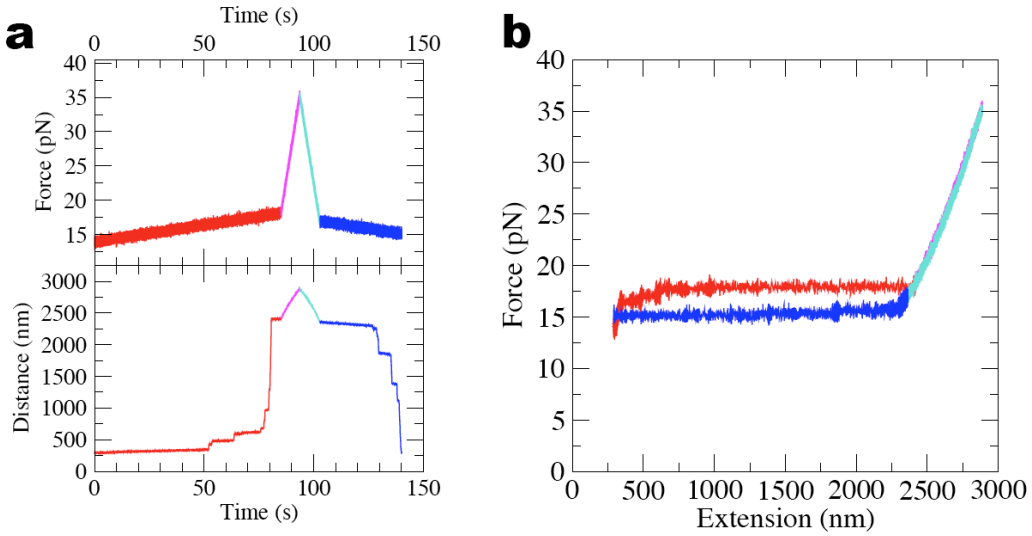


Figure 6.1: Data from a CF pulling experiment. **(a)** Force vs. time (upper panel) and Distance vs. time (lower panel). The force is increased monotonically at a loading rate of 0.05 pN/s (red curve) until the molecule is fully unzipped (i.e., converted into ssDNA). Then, the loading rate is increased up to 2 pN/s (magenta curve) to pull the released ssDNA. Even at high loading rates, the pulling of ssDNA is reversible. So the loading rate is increased in order to reduce the time spent in the pulling cycle. During the rezipping, the force is decreased at -2 pN/s (cyan curve) and at -0.05 pN/s (blue curve). The distance suddenly increases (decreases) when a group of bases is unzipped (rezipped). Note that the force feedback is not capable of reducing the fluctuations of the thermal noise. The force undergoes some bumps during the opening of base pairs due to the finite bandwidth of the force feedback. **(b)** Resulting FDC of previous measurement. This is a typical FDC at controlled force that exhibits an appreciable cycle of hysteresis.

As the force increases, the distance increases exhibiting sudden horizontal hops. The resulting FDC monotonically increases and the force never retracts (except for some thermal fluctuations). Once the DNA molecule is fully open, the elastic response of the ssDNA is measured just like in the CP protocol. When the maximum force is reached, the loading rate is inverted and the force starts to decrease (rezipping). The measured elastic response of the ssDNA during the rezipping is almost identical to the unzipping. However, during the rezipping the closing of base pairs starts at lower forces than the opening of base pairs during the unzipping. As the force decreases, the extension of the molecule is reduced (and so the distance) until returning to the original starting point of the unzipping.

The most prominent characteristic of these unzipping/rezipping curves measured in one single cycle of pulling is the fact that they do not overlap, so a whole pulling cycle shows hysteresis. The hysteresis is one of the main indicators of irreversibility in thermodynamical processes. The area enclosed

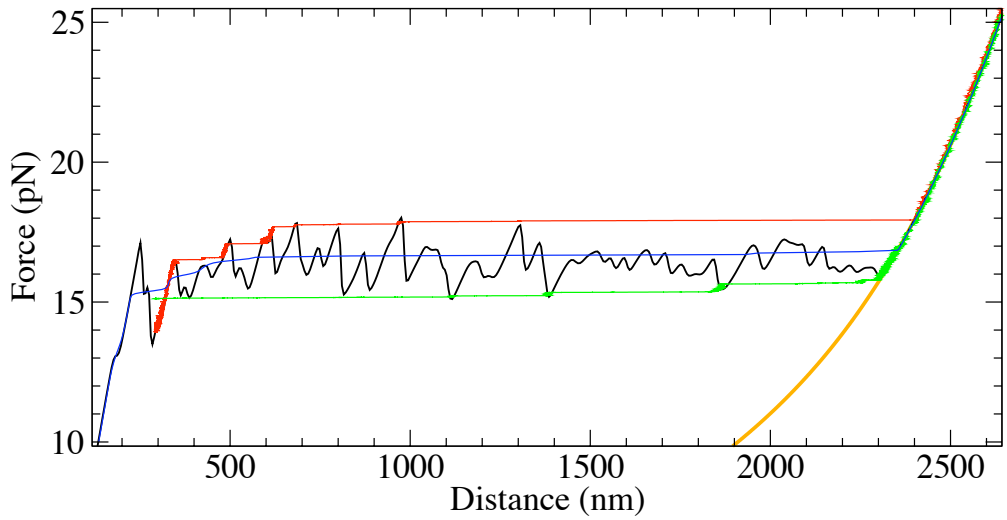


Figure 6.2: Full cycle of a CF unzipping/rezipping experiment of a 2.2 kbp sequence performed at a loading rate of 0.05 pN/s. Red (green) curve shows the unzipping (rezipping) at CF. The blue curve shows the expected equilibrium FDC at CF. We have superimposed the FDC at CP (in black) and the elastic response of the ssDNA (in orange).

between the two unzipping/rezipping curves is equivalent to the work dissipated by the system during the whole cycle. This means that although the process has been carried out at a very low loading rate it is not reversible.

The time spent in a pulling cycle is not enough to equilibrate the system and obtain the equilibrium FDC. According to the Kramers theory [151, 152], the time scale τ to overcome an energetic barrier of height ΔE is given by:

$$\tau \approx \tau_0 \exp\left(\Delta E / k_B T\right) \tag{6.1}$$

where $\tau_0 \approx 10^{-7}$ s is the microscopic time at which the base-pairs breath (i.e., open and close) [136, 23], k_B is the Boltzmann constant and T is the temperature. The highest barriers at the coexistence force are of the order of 25-30 $k_B T$, which gives a mean passage time between hours and weeks. In principle, if we perform a CF pulling experiment slow enough, we should be able to obtain the equilibrium FDC. However, this is not experimentally feasible, because the time spent in such experiment would be extremely large and the pulling experiment could be ruined by several issues such as the breakage of the tether or the unavoidable long term drift.

Figure 6.2 also shows the predicted theoretical Force vs. Distance Curve at Controlled Force (FDC$_f$) according to the calculations of Sec. 3.4.2. The difference between the experimental and theoretical FDC$_f$ is extremely significant. It is another evidence that the CF experiments cannot be performed

quasistatically at the laboratory time scale.

There is a major difference between the CF experiments and the CP experiments. For the latter, the equilibrium FDC can be obtained just by reducing the pulling speed down to reasonable values. At first sight, one can see that the FDC at CP (FDC_x) lies between the unzipping and rezipping FDCs at controlled force. Although both curves have radically different shapes, the mean unzipping force is around $\sim$16.5 pN in both cases. What is more, there seems to be a way to relate one type of curve with the other. Note that the FDC_f follows the slopes of the FDC_x until a force rip is achieved. At this point, the FDC_f increases its extension until it intercepts the next slope of the FDC_x. This process is repeated until no slope is intercepted and the molecule is fully extended showing the elastic response of the ssDNA. Similarly, the rezipping FDC_f has a symmetric relation with the FDC_x. Somehow, the experimental FDC_f that we can measure are the envelope curves of the FDC_x.

6.1.3 Hysteresis

In general, relaxation processes in non-equilibrium systems are characterized by the intrinsic relaxation time (τ_0) of the system and the external characteristic time at which the system is driven (τ_{ext}). In the case of the DNA unzipping at CF, there is an intrinsic relaxation time that indicates the average time the molecule needs to reach the thermal equilibrium and an external time related with the loading rate of the pulling experiment. The hysteresis is a phenomenon that appears when the driven rate is faster than the relaxation time of the system ($\tau_{\mathrm{ext}} \ll \tau_0$). In this situation, the system gets stuck in metastable states because it is not capable of reacting fast enough to the changes induced by the external control parameter. This phenomenon is specially observed in systems with rough free energy landscapes (e.g., disordered or glassy systems) that have energetic barriers much higher than the thermal noise. Under this circumstances, the barriers prevent the system to explore the other regions of the free energy landscape.

Here we are precisely facing this effect. In the CF unzipping experiments, the rate of change of the force, although very slow to our lab time scale, is much faster than the relaxation time of the DNA molecule. The slowest feasible experimental unzipping/rezipping cycles do not reproduce the equilibrium FDC. And even if we further reduce the pulling rate, there is no guarantee that we will be able to obtain it. In fact, the relaxation time of a system depends exponentially on the height of the typical energy barriers involved, which is about 20-30 k_BT at the coexistence force (see Fig. 3.20). So there are firm reasons to believe that the equilibrium FDC_f will be ob-

tained after performing pulling experiments that last orders of magnitude much longer than the ones described here (see Eq. 6.1 and the corresponding discussion for an estimation of the time required).

So far, we have focused on the experiments performed at low pulling rates. Considering that we will obtain similar results by decreasing the loading rate, we can start looking at the other side. By increasing the loading rate, the irreversibility of the process increases. The cycles of hysteresis become wider and the total amount of dissipated work increases. Figure 6.3a shows several unzipping/rezipping cycles at different loading rates and the corresponding area of the cycles of hysteresis. The unzipping of DNA at CF has already been studied for short hairpins [153, 154] and a linear relation between the dissipated work and the loading rate has been found in the limit of slow loading rates. Moreover, the dissipated work tends to zero when the time spent in the pulling tends to infinity (zero pulling rate). However, in our case this does not hold, at least for the range of loading rates explored (0.05–5 pN·s^{-1}). In fact, the dissipated work tends to saturate at a value of $\simeq 1000\ k_BT$.

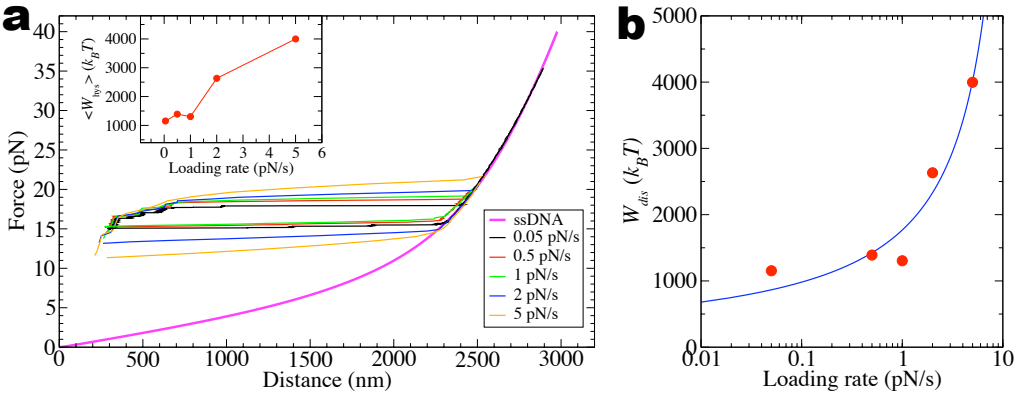


Figure 6.3: Hysteresis. **(a)** Pulling cycles at different loading rates. **(Inset)** Dissipated work (area enclosed by the cycle) vs. loading rate. **(b)** Hysteresis in logarithmic scale. Same graph as the inset in panel a. The continuous curve is just a visual guide.

Figure 6.3b shows a logarithmic representation of the loading rate, which is indicative of the trend of the dissipated work at very low loading rates. According to this, the pulling rate must be decreased orders of magnitude to let the system equilibrate and reduce the dissipated work down to zero. A detailed study of the free energy landscape of the system would give valuable information about the energetic barriers that cannot be surpassed during an unzipping process at CF.

6.2 Detection of metastable intermediate states

Using the same analysis technique developed in Sec. 4.1, here we identify and classify the experimental points measured in the FDC_f into the most probable state to which they might belong. This gives us information about the intermediate states observed during the unzipping at CF.

As a short reminder, the analysis consists in: 1) determine the origin of coordinates of the FDC by using the elastic response of the ssDNA when the molecule is fully open; 2) calculate the family of curves that give the elastic response of the partially unzipped molecule with n open base pairs and 3) associate each experimental point with the curve n that passes closest to this point. Finally, a histogram is computed with the list of associated states. The peaks of the histogram (see Fig. 6.4) indicate the metastable states in which the molecule is retained during the unzipping.

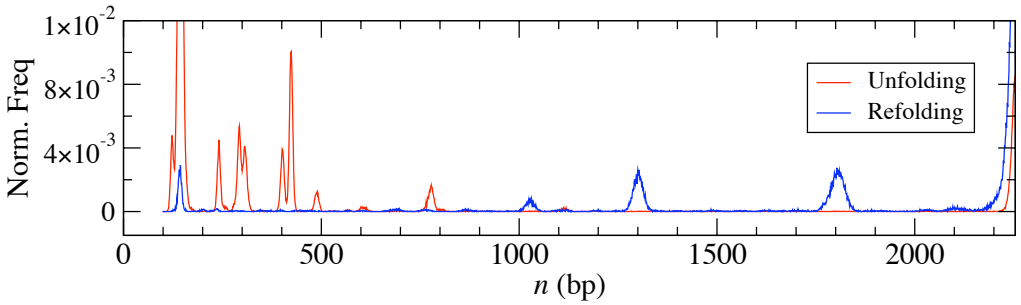


Figure 6.4: Histogram of intermediate states at CF of a 2.2 kbp sequence unzipped/rezipped at a loading rate of 0.05 $\mathrm{pN{\cdot}s^{-1}}$.

As expected from the clear differences between the FDC_f and the FDC_x, the histograms of intermediate states are also significantly different (compare Figs. 6.4 and 4.7a). Here, the number of intermediate states is much lower and they are much more localized at the beginning of the molecule during unzipping, and at the end of the molecule during rezipping. Indeed, the sizes of opening base pairs are smaller and they are concentrated at the beginning of the FDC_f. Summing up, we get less information from the sequence of DNA.

6.2.1 Free energy landscape

Once the intermediate states are identified, we can explore the localization of these states in the free energy landscape. The free energy landscape is calculated according to the tools exposed in Sec. 3.4.2. The force applied to the DNA molecule induces a tilt in the free energy landscape that tends to

progressively decrease the energies of the states with a large number of open base pairs. The result is that the system is more stable for larger values of n as the force is increased.

Figure 6.5a shows the free energy landscape computed at a fixed force, together with the metastable states detected during unzipping according to the analysis of the previous section (see also Fig. 6.4). These states are localized at the beginning of the molecule and the separation between them increases as the molecule is unzipped. What is the characteristic property of these states that we observe during the unzipping? Why do we observe these states and not the other ones? The answer is not straightforward because there is no general theory to study the kinetics of systems that exhibit rough landscapes.

Nevertheless, we can suggest a heuristic explanation. These metastable states are separated by energy barriers that decrease their height as the force increases. At the beginning of the pulling (i.e., low forces), the tilt of the landscape induced by the force strongly affects the energy of the states with a few open base-pairs. This induces transitions between the initial states, characterized by the opening of a few base-pairs. As the force increases, the tilt reduces the energy of the states with a larger number of open base-pairs. The transitions then become less frequent and the number of open base-pairs in each transition increases significantly.

This very same effect is observed during the rezipping but in the opposite sense (see Fig. 6.5b). Initially we observe fast and small transitions between the states with a large number of open base-pairs. As the force decreases, the transitions become larger and less frequent until the molecule is completely folded.

This mechanism prevents the system to explore all the free energy landscape at the theoretical equilibrium force. This is the reason why we always observe hysteresis in CF experiments performed at the lab time scale.

6.3 Scaling properties of unzipping

In 2002, Lubensky and Nelson published an extensive theoretical study of DNA unzipping at constant force [29]. Starting from a mesoscopic model, their work treated the unzipping of DNA as a phase transition. They provided the scaling properties of the system and the expected critical exponents at the coexistence (i.e., critical or mean) unzipping force.

One year latter, Lubensky and Nelson also contributed as co-authors in an experimental work carried out by Danilowicz *et al.* [23]. The unzipping of DNA was performed with magnetic tweezers and the results were compared

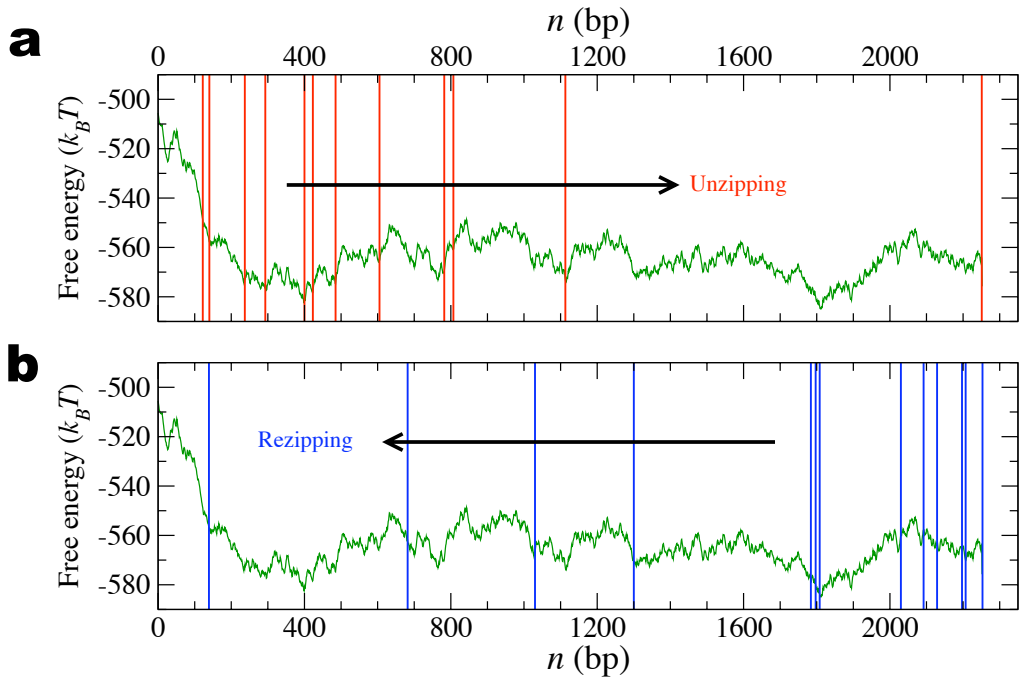


Figure 6.5: Location of intermediate states in the free energy landscape for the 2.2 kbp sequence. **(a)** Free energy landscape (green curve) and localization of intermediate states observed during unzipping (red marks) at CF. The landscape has been calculated at the coexistence force, $f = 16.5$ pN. Note that more metastable states are detected at the beginning of the molecule. The arrow indicates the succession of observed states. **(b)** Same graphic for rezipping. The observed intermediate states are depicted in blue.

with a coarse-grained model that qualitatively predicted the experimental observations. However, the resolution of the instrument was not good enough to compare the data with the theoretical results calculated one year earlier.

According to the calculations of Lubensky and Nelson, the number of open base-pairs as the exerted force approaches the critical force goes like,

$$\overline{\langle m \rangle} = \overline{\left\langle \frac{n}{N} \right\rangle} \sim \frac{1}{(f_c - f)^2} \tag{6.2}$$

where n is the number of open base-pairs, N is the total number of base-pairs of the molecule, $\langle \cdots \rangle$ indicates the average over the statistical ensemble, $\overline{\cdots}$ indicates the average over realizations of the sequence, f_c is the critical force and f is the force exerted on the molecule. This calculation is valid of a random heteropolymer. The 2.2 kbp sequence is a random heteropolymer, because the sequence of base-pairs can be considered random (with 4 different types of bases in the sequence). Lubensky and Nelson obtained a different

law for homopolymers (i.e., with only one type of base-pair) which is given by: $\overline{\langle m \rangle} \sim (f_c - f)^{-1}$

The detection of metastable states in Sec. 6.2 allows us to determine the number of open base-pairs of each experimental point of the FDC$_f$. Since we know the total number of base-pairs of the molecule ($N = 2252$) we can obtain an experimental estimation of the curve defined in Eq. 6.2, i.e., a curve of $\overline{\langle m \rangle}$ vs. f (see Fig. 6.6).

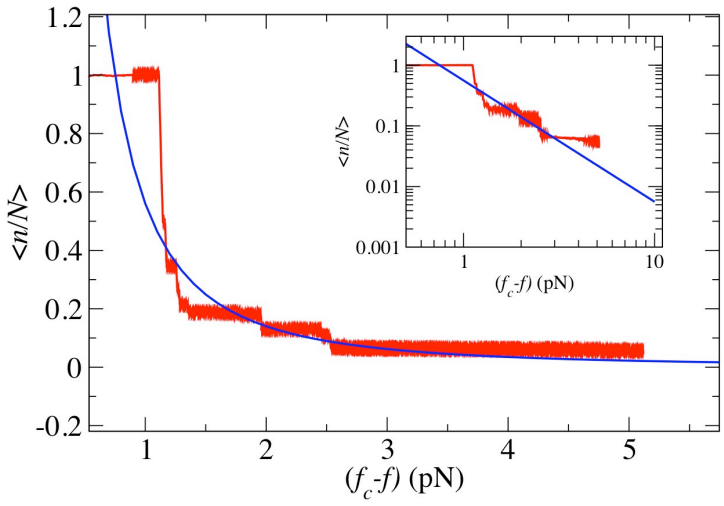


Figure 6.6: Scaling properties at the unzipping transition. The red curve shows the experimental data obtained from the unzipping of a 2.2 kbp sequence at a loading rate of 0.05 pN/s. The blue curve is the fit of the experimental data to the following function: $m = A/(f_c - f)^2$, with $f_c \simeq 19$ pN. The inset shows the same data in log-log scale.

The fit of the experimental data to Eq. 6.2 is satisfactory. A priori, we should not expect a perfect agreement. The reason is that Eq. 6.6 was predicted in equilibrium and averaged over realizations, while the experiments are not performed in equilibrium and we only have one sequence. Besides, the 2.2 kbp sequence is far from being an infinite molecule in which the thermodynamic prediction can be applied. So the fact that the parameter that fits the data ($f_c \simeq 19$ pN) is significantly higher than the actual critical force ($f_c = 16.5$ pN) might indicate that we are comparing non-equilibrium data with an equilibrium prediction. This result must be corroborated with more experiments performed on longer sequences.

6.4 Conclusions

Although force is not the natural control parameter in the minitweezers, we can perform unzipping experiments of DNA at controlled force by using a force feedback. The bandwidth of the force feedback is wide enough to keep the force constant while the DNA molecule undergoes transitions between metastable states. The force feedback, though, cannot follow the fluctuations of the force induced by the thermal motion.

The unzipping FDC_f differs significantly from the FDC_x. The FDC_f is a step-like monotonically increasing function and does not exhibit the typical sawtooth pattern of the FDC_x. Even at a very low loading rate, the unzipping and rezipping FDC_f do not overlap each other. The unzipping/rezipping cycles are not quasistatic at the timescale of the lab and they always exhibit hysteresis. The hysteresis has a minimal value and it increases with the loading rate. The minimal value corresponds to the slowest loading rate that is experimentally feasible in the instrument.

The irreversibility of the pulling experiments at controlled force is due to the height of the energetic barriers of the energy landscape. In a pulling experiment, the system does not have enough time to overpass such energetic barriers, which means that the entire distribution of energy states is not correctly sampled and full equilibrium cannot be reached. The number of metastable states detected in experiments is lower (and the barriers are larger) at controlled force as compared to controlled position. Besides, at controlled force the metastable states observed during unzipping are different from the ones observed during rezipping.

The unzipping at controlled force starts with small and frequent transitions between metastable states. As the force increases and the base-pairs are being disrupted, the transitions become larger and less frequent. This effect can be understood by looking at the free energy landscape. The progressive tilt of the free energy explains these transitions. The rezipping also shows this behavior.

Although the molecule is far from being an extensive thermodynamic system and the experiments are not quasistatic, the measurements verify the prediction of the number of open base-pairs vs. the force exerted.

Science is always wrong. It never solves a problem without creating ten more.

George Bernard Shaw (1856-1950)

What today started as a science fiction novel, tomorrow will end as a news report.

Arthur Charles Clarke (1917-2008)

Chapter 7

Future perspectives and Conclusions

Future perspectives

Most scientific works start with a main question. The strategy required to answer such question is well established by the scientific method. This involves the finding of evidences, a rigorous and critical evaluation of the results and the formulation of conclusions. However, this process usually opens new questions. Some of them cannot be avoided and must be addressed to reach the final goal. Some others are out of the scope of the research and they are left open for future investigations. This fact is specially marked in experimental sciences. Sometimes, the results of the experiments are unexpected and the investigations must focus on issues initially not predicted. This is what science is about: answering one question leads to another one.

A lot of new questions have emerged during the realization of this PhD thesis. Some of them were crucial and they had to be addressed. For instance, the measurement of the elastic response of the ssDNA in Chapter 5 was an unexpected experiment. Such experiment became essential to complete the data analysis. So the main question forced to answer secondary questions. Apart from that, new questions that have not been addressed have emerged. Following, there is a summary of the most relevant of them.

How to infer the NNBP enthalpies and entropies from unzipping experiments?

Chapter 5 has shown in detail how to perform the experiments and analyze the data to obtain the free energy of formation of the NNBP motifs. As pointed out there, there is no way to extract the enthalpies and the entropies from the unzipping experiments. The solution is to repeat the experiments at a different temperature. However, this is not an easy task. After having developed all the methodology to analyze the data, the challenge seems to be purely experimental.

In 2005 Mao *et al.* [155] showed that the fluidics chamber developed convective flows when it was heated locally. The convection is mainly due to differences in the temperature throughout the fluidics chamber. In SME, this is really inconvenient because the flows of fluid might exert significant forces on the molecules being pulled. From that work, it was clear that the chamber has to be heated (or cooled) uniformly. They used copper jackets to cover the microscope objectives and keep all the fluidics chamber at a constant temperature. The temperature could be controlled between 8.4 and 45.6°C and the DNA was overstretched at different temperatures.

In our case, the accuracy in the force measurements is an essential point. Indeed, the relative error committed in the overstretching of DNA is much lower than the one committed in unzipping. Besides, our aim is to quantitatively measure the enthalpies and entropies at the single-molecule level. So the temperature of the fluidics chamber must be very homogeneous and stable.

One solution is to increase the whole temperature of the room (provided that the experimentalist can hold such temperature). This would ensure that the room, the instrument and the fluidics chamber are at the same temperature and the convective flows are avoided. However, this is not feasible nor efficient. The best way is to control the temperature of the fluidics chamber and all the surrounding area, including the objectives and the air in between. This would require to enclose the optical path with an isolating box and flow air into it at the desired temperature. A priori, this seems a good strategy. Nevertheless, there might be some issues related with the optical alignment of the instrument such as changes in the index of refraction of the air or dilatation of the aluminum supports that hold the detectors.

Successful unzipping experiments at different temperatures might open the door to a new methodology to measure enthalpies and entropies of biomolecules. This could establish unzipping as a standard technique that would complement the traditional bulk methods, such as calorimetry or UV absorbance. What is more, unzipping experiments could provide information

at conditions not accessible by bulk experiments.

Is there an analytical model to describe the CURs?

The statistical properties of DNA unzipping are sequence dependent (opening of base-pairs, force rips, FDC, etc.). From a physical point of view, it is annoying to study a system that is sequence dependent. Indeed, understanding a physical phenomenon from a generic model is much more gratifying. The reason is that it provides information about the global behavior of the system without having to look into the details.

Lubensky and Nelson [29] did a great work in which they were able to describe the unzipping of DNA at controlled force. This work uses mathematical tools to proceed with analytical calculations. The idea of studying hard theoretical problems has always fascinated the physicists. These problems require a combination of imagination and mathematical capabilities. From a physical point of view, the conclusions of these works are usually bright and beautiful. Sometimes the predictions found cannot be experimentally tested. For instance, the relaxation time in glassy or frustrated systems is so large that few experiments can be designed to probe the longest timescales. Nevertheless, it is worth exploring these problems because the mathematical machinery developed is useful and can be applied to other fields.

One interesting question that emerges from Chapter 3 is whether the distribution of CUR sizes can be calculated analytically. In similar problems, this quantity is computed in equilibrium and averaged over realizations, i.e., over sequences. An analytical expression would be useful to compare unzipping with other physical phenomena.

Does the ssDNA have a sequence-dependent elastic response?

An unsolved question in Chapter 5 is the origin of the heterogeneous salt dependence of the NNBP energies. There we discussed two main causes: 1) the different solvation of the dsDNA base-pairs or 2) the sequence-dependent elastic response of ssDNA. Concerning its elasticity, the ssDNA is still an unknown polymer. The pioneering literature measured the elastic response of long ssDNA molecules and gave the parameters of the models that fit the data. However, the elasticity of ssDNA depends on the salt conditions of the buffer. Moreover, the persistence (or Kuhn) length of the ssDNA is of the same order of magnitude than the typical interphosphate distance (i.e., the size of the monomers). So there is a strong correlation between these two parameters and the fit of experimental data is not conclusive yet.

From these results, it is worth checking whether the salt dependence of

the NNBP energies is due to the elastic properties of ssDNA. This hypothesis implies that the elastic response of the ssDNA is sequence dependent. In order to test the validity of this idea and quantify the salt dependence, it would be necessary to perform pulling experiments on several sequences of ssDNA as we did in Chapter 5. By gathering the information from the experiments, averaging over molecules and considering the possible formation of secondary structure we could possibly answer this question.

Knowing the elastic properties of the dsDNA has been really useful since they were established. The reason is that dsDNA has been widely used as a molecular handle in many SME. Similarly, in case the elastic response of ssDNA is sequence-dependent, most biophysicists working on DNA or DNA-protein interaction would benefit from the determination of these parameters.

How can we sequence DNA from unzipping experiments?

Sequencing a given molecule of DNA consists in determining the primary structure (i.e., the sequence of base-pairs or nucleotides) of such molecule. This is an important procedure in biology and it has been the nucleus of the Human Genome Project, completed in 2003. Among other applications, sequencing permits to identify genes or compare two samples of genetic material.

The sequencing of DNA has been performed with Sanger method since its implementation in 1975 [30, 156]. This is a biochemical technique that uses a polymerase to elongate the DNA molecules of the sample and chain-terminator nucleotides, which are chemically modified and they terminate the elongation. In the end, the DNA molecules of the sample have different lengths, depending on the random incorporation of terminating nucleotides. Besides, these terminating nucleotides are fluorescent (they used to be radioactive in Sanger original method) and they have a different glowing color depending on the type of base (A,C,G,T). The fragments of DNA are separated by size using electrophoresis and the last nucleotide of each of them is identified according to its glowing color.

The Sanger method has been improved and nowadays the sequencing of DNA is almost automatic. However, the method has a limitation. The main handicap is that it is efficient for fragments of DNA smaller than 500-1000 nucleotides. Longer sequences must be split into fragments, sequenced and reassembled to establish the original order. Nowadays, the assembling of fragments is the most time consuming task. The assembly is also the limiting step in the so-called *next-generation* sequencing methods (such as MPSS, pyrosequencing) developed in the 1990's.

The single molecule unzipping experiments could be used as an alter-

native method to sequence DNA molecules. The strategy would consist in performing pulling experiments according to Chapter 3 and analyzing the FDC according to Chapter 5. The underlying principle of this idea is that each base-pair disrupts at a different force. So as the molecule is being unzipped, the force applied to break one base-pair can be used to infer the type of base-pair. However, as we have seen in Chapter 4, the unzipping of DNA exhibits a series of CURs that hide the sequence of bases. So the experimental setup has to focus on how to produce the unzipping of one base-pair at a time. This requires two conditions: 1) apply local force on the unzipping fork, avoiding the accumulation of released ssDNA and 2) have a probe stiffness larger than 100 pN/nm.

A priori a larger stiffness can be obtained by increasing the power of the laser that produces the optical trap. However, the effect of a high intensity laser on the fluidics chamber or the DNA should be tested. The stiffness of an optical trap can also be increased with the leverage technique [157], in which the bead leans and rolls over the coverslip. The bead works as a lever that amplifies the force applied by the optical trap. A higher stiffness can also be achieved with other single-molecule techniques that apply forces, such as the AFM. The application of local force is somehow more complicated to implement. The key point here is how to retract the ssDNA released during unzipping. While the unzipping of DNA with nanopores already has this advantage [158], the optical tweezers setup has not. Although some strategies (such as wrapping the ssDNA around the bead) could be designed to overcome this handicap, a good way is to use a helicase. In this experimental setup, the helicase would be bounded to be bead in a such a manner that the helicase slides along the released ssDNA an the force is always exerted at the unzipping fork.

The main advantage of sequencing by force is that it would be faster. Indeed, the time spent would be linear with the length of the DNA molecule. Besides, the amount of chemicals used is not length dependent, which might reduce the cost of sequencing. The sequencing by force would be included in the group of *third generation* sequencing techniques, based on single-molecule techniques.

Other interesting issues

Regarding instrumentation, it is worth to consolidate the double optical trap setup in the Minitweezers (see Fig. 2.17a). Besides, the research of better algorithms might be useful to produce more efficient experimental protocols (e.g., constant force). The current trends in SME demand higher bandwidth and sampling rates in data acquisition, which allows the experimentalist to

get much more information from the measured fluctuations.

Although the biochemistry in DNA unzipping is well established, it is worth to test other protocols in order to improve the strength of the bonds between the probes (i.e., beads) and the samples (i.e., DNA). Stronger bonds make the experiments last longer and allow to apply more complex protocols (e.g., force jumps, high pulling rates, etc.) without breaking the tether.

An interesting measurement not performed yet is the dependence of the CUR size distribution with the laser power (i.e., the trap stiffness). It would shed light into the unzipping of DNA seen as a fracture phenomenon. It is also good to extend the unzipping experiments to other sequences such as homopolymers or palindromes.

Currently, the analysis of a FDC to obtain the NNBP energies requires intensive human supervision. The analysis of data could be improved an automatized to find the values of the fitting parameters much more efficiently. Apart from that, it would be useful to design an application similar to *Mfold* to calculate experimental FDCs of given sequences of DNA at different salt conditions. This certainly would be a valuable tool for the experimentalists working on DNA unzipping.

Finally, the unzipping experiments at constant force have to be extended to the 6.8 kbp sequence in order to gain insight and complete the whole picture sketched in Chapter 6.

Conclusions

The boom of nanoscience and nanotechnology has pushed the instrumentation forward. The experimental setups have been developed so intensively that the single-molecule experiments have now become standard techniques. Such experiments represent a new way to observe, measure and get information from nature. The control and measuring devices are more accurate and precise every day. This allows the scientists to observe, explore and quantify processes that some years ago were considered too complex. Thanks to the single-molecule techniques, biophysics has experienced a revolution. The traditional experiments performed in bulk have been repeated at the single-molecule level, which has revealed the extraordinary capabilities of the cellular machinery. All these advances have established biophysics as a scientific discipline, different from biology and physics; with its own topics, concerns and issues.

Optical tweezers were initially developed by physicists who were fascinated by the radiation pressure of light. Biophysicists soon foresaw the useful applications of optical tweezers to manipulate tiny objects. The range of forces measured and exerted by optical tweezers is suitable to carry out experiments with biomolecules. Optical tweezers is a clean and non-invasive technique which also allows to do experiments *in vivo*. The Minitweezers is an optical tweezers experimental setup characterized by its compactness, stability, accuracy and user-friendliness. The optical trap is generally formed with two counter propagating lasers and a micropipette is used as an anchor point. The force is measured by conservation of light momentum and the distances are measured with light-levers. The calibration is independent of most experimental parameters (e.g., laser power, bead size) so that it only has to be done once. The Minitweezers were designed to be easily customized. Therefore, new experimental pulling protocols can be easily implemented and the already existing ones can be modified and expanded.

The molecule of DNA has a central role in life: It stores the genetic information. The central dogma of molecular biology states the general flow of genetic information in life (replication and transcription of DNA, and translation of RNA). The double-helix structure of DNA preserves the base-pairs located in between, which are the physical carriers of genetic information. The cellular machinery needs to disrupt the hydrogen-bonds that hold the base-pairs together in order to read the sequence of bases. Similarly, mechanical unzipping of DNA consists in pulling on the two strands of DNA from the same end in order to split them apart. Unzipping can be performed with several single-molecule manipulation techniques (microneedles, AFM, magnetic tweezers). Optical tweezers can produce unzipping at controlled

position or at controlled force, which give different experimental measurements. At controlled position the FDC exhibits a sawtooth pattern, whose slopes correspond to the stretching of ssDNA and the force rips correspond to the disruption of base-pairs. At controlled force, the base-pairs are disrupted in large groups and the FDC is a monotonically increasing function. In both cases, the FDC is sequence-dependent. The NN model describes the hybridization reaction of two strands that form a duplex of dsDNA. Such model can be completed with elastic models of polymers to describe the unzipping experiments. The so-called mesoscopic model provides a prediction of the FDC, an estimation of the free energy landscape and the number of open base-pairs as the DNA is unzipped.

From a physical point of view, the unzipping of DNA can be studied as a cracking phenomenon. Here, the details about the sequence of the molecule are less relevant. The attention is focused on the statistical properties of the metastable states observed during unzipping. A Bayesian approach has been developed in order to infer the number of open base-pairs of the intermediate states observed during unzipping. This allows us to calculate the size of the unzipping regions. The distribution of sizes depends on the experimental conditions (trap stiffness, NNBP free energies). The sizes range from 10-80 bp and the smaller ones are more frequent than larger ones. The experimental accuracy does not allow to observe all regions of sizes below 10 bp. On the other hand, the effective trap stiffness must be much higher to observe single bp openings of large unzipping regions. The toy model is capable of reproducing the statistical properties of unzipping with the minimal necessary elements. The model predicts and qualitatively reproduces the experimental distribution of sizes. The unzipping of one base-pair at a time can only be achieved by applying local force on the unzipping fork and by having a trap stiffness value higher than 100 pN/nm. The stiffness of a ssDNA nucleotide coincides with this value, which is a remarkable property of DNA that allows the cellular machinery to access the genetic information one base-pair at a time.

The NNBP free energies obtained from temperature melting experiments cannot quantitatively reproduce the FDC obtained from DNA unzipping experiments. The discrepancies are markedly significant at low salt concentration. The unzipping experiments performed on DNA can be used to extract the formation free energy of NNBP motifs. The experimental data can be fit to the mesoscopic model in order to determine the values of the NNBP energies. The model has to be completed with an accurate description of the elastic properties of the ssDNA, a shift function that corrects the instrumental drift and the specific free energy formation of the end loop. The fit is performed with a Monte Carlo optimization algorithm that provides

the values of the fitting parameters rapidly and robustly. Differently from the UO rules, the resulting NNBP energies at all salt concentrations are well described by a heterogeneous salt correction. Such dependency can be attributed to the different solvation of the base-pairs, or to the sequence-dependent elastic response of the ssDNA. The unzipping experiments cannot provide the enthalpies and entropies of the NNBP motifs nor the initiation factors. However, the enthalpies and entropies can be inferred by fitting the melting temperatures of oligonucleotides. The results show that the new values of the NNBP energies correctly describe both the melting and unzipping experiments. Furthermore, the new NNBP energies predict the melting temperatures of oligos longer than 15 bp better than the UO NNBP energies. This methodology can be extended to other experimental conditions in which the melting experiments cannot be applied. For instance, in a melting experiment RNA is hydrolyzed by magnesium when the temperature increases. The unzipping of RNA would circumvent such problem. In the end, the NN model is capable of describing the disruption and hybridization of nucleic acids, no matter what external agent (temperature or force) triggers the reaction.

The Minitweezers experimental setup can also be used to perform unzipping experiments at controlled force using a force feedback. The bandwidth of the force feedback is not high enough to keep the force constant in the presence of thermal fluctuations. Nevertheless, the average force is kept constant while the DNA molecule undergoes the opening of base-pairs. The experimental FDC_x is significantly different from the FDC_f and the latter always exhibits hysteresis at the timescale of the measurements. The hysteresis is not suppressed at the lowest feasible loading rate and increases with the loading rate. The mesoscopic model cannot predict the experimental FDC_f well, because this is not a quasistatic measurement. The intermediate states observed at controlled force are different from the ones observed at controlled position. At controlled force, there are small openings of base-pairs in the beginning of the unzipping and large openings in the end. This is a consequence of the tilt of the free energy landscape induced by the force. An analysis of the intermediate states shows that the number of open base-pairs vs. force follow the theoretically predicted scaling properties.

This thesis has focused on the statistical and thermodynamic properties of DNA unzipping measured with optical tweezers. This study has answered some questions and has open new ones. The next step is to extend the work to comprehend the remarkable properties of DNA, and to find practical applications based on the achievements exposed here.

Part III

Appendixes

Appendix A

The Maxwell Stress Tensor

The Maxwell equations of the electromagnetism provide a frame in which the interaction between radiation and matter can be studied from a classical point of view. It allow us to understand what are the mechanisms that make possible the optical trapping of particles. Here, we will develop the general theory of radiation-matter interaction starting from the principle of conservation of momentum combined with the Maxwell equations [159]. The derivation is done for radiation and matter that interact in the vacuum. The obtained expressions can be extended to other homogeneous and isotropic dielectric mediums by substituting the electric permittivity (ϵ_0) and the magnetic permeability (μ_0) of the vacuum by the actual values of the medium (ϵ, μ).

The total electromagnetic force $\vec{F}$ that acts on a charged particle is given by the Lorentz force [159]

$$\vec{F} = \frac{d\vec{p}}{dt} = q\left(\vec{E} + \vec{v} \times \vec{B}\right) \tag{A.1}$$

where q is the charge of the particle, $\vec{v}$ is the velocity of the particle and $\vec{E}$ and $\vec{B}$ are the electric and magnetic fields.

If we sum the contribution of all the particles contained in the volume V we can write the total force ($\Delta\vec{F}$) acting on the volume,

$$\Delta\vec{F} = \int_V \left(\rho\vec{E} + \rho\vec{v} \times \vec{B}\right) dV \tag{A.2}$$

where ρ is the charge density and $\vec{J} = \rho\vec{v}$ is the current density vector.

We can also define the force acting on a tiny volume,

$$d\vec{F} = \left(\rho\vec{E} + \vec{J} \times \vec{B}\right) dV \tag{A.3}$$

that allow us to define the force per unit volume as,

$$\vec{f} = \frac{d\vec{F}}{dV} = \left(\rho\vec{E} + \vec{J} \times \vec{B}\right) \tag{A.4}$$

At this point, we can use the two following Maxwell equations (Poisson and Maxwell-Ampère equations) to eliminate ρ and $\vec{J}$,

$$\begin{aligned} \rho &= \epsilon_0 \left(\vec{\nabla} \cdot \vec{E}\right) \\ \vec{J} &= \frac{1}{\mu_0}\vec{\nabla} \times \vec{B} - \epsilon_0 \frac{\partial \vec{E}}{\partial t} \end{aligned} \tag{A.5}$$

and write the density of force in terms of the electromagnetic fields,

$$\begin{aligned} \vec{f} &= \epsilon_0 \left(\vec{\nabla} \cdot \vec{E}\right) \vec{E} + \left(\frac{1}{\mu_0}\vec{\nabla} \times \vec{B} - \epsilon_0 \frac{\partial \vec{E}}{\partial t}\right) \times \vec{B} \\ &= \epsilon_0 \left(\vec{\nabla} \cdot \vec{E}\right) \vec{E} + \frac{1}{\mu_0}\left(\vec{\nabla} \times \vec{B}\right) \times \vec{B} - \epsilon_0 \frac{\partial \vec{E}}{\partial t} \times \vec{B} \ . \end{aligned} \tag{A.6}$$

We want to group the time dependent contributions. We note that using the product derivative rule and another Maxwell equation (Maxwell-Faraday equation) $\left(-\frac{\partial \vec{B}}{\partial t} = \vec{\nabla} \times \vec{E}\right)$ we can write the last term as,

$$\begin{aligned} \frac{\partial \vec{E}}{\partial t} \times \vec{B} &= \frac{\partial\left(\vec{E} \times \vec{B}\right)}{\partial t} - \vec{E} \times \frac{\partial \vec{B}}{\partial t} \\ &= \frac{\partial\left(\vec{E} \times \vec{B}\right)}{\partial t} + \vec{E} \times \left(\vec{\nabla} \times \vec{E}\right) \\ &= \mu_0 \frac{\partial \vec{S}}{\partial t} + \vec{E} \times \left(\vec{\nabla} \times \vec{E}\right) \\ &= \mu_0 \frac{\partial \vec{S}}{\partial t} - \left(\vec{\nabla} \times \vec{E}\right) \times \vec{E} \end{aligned} \tag{A.7}$$

where we have used the definition of the Poynting vector $\vec{S} = \frac{1}{\mu_0}\left(\vec{E} \times \vec{B}\right)$.

The second terms of equations A.6 and A.7 are the cross product of the curl of a vector with itself, which can be written as:

$$\begin{aligned} \left(\vec{\nabla} \times \vec{B}\right) \times \vec{B} &= \left(\vec{B} \cdot \vec{\nabla}\right) \vec{B} - \frac{1}{2}\vec{\nabla}\vec{B}^2 \\ \left(\vec{\nabla} \times \vec{E}\right) \times \vec{E} &= \left(\vec{E} \cdot \vec{\nabla}\right) \vec{E} - \frac{1}{2}\vec{\nabla}\vec{E}^2 \ . \end{aligned} \tag{A.8}$$

Substituting A.8 into A.6 and A.7, and A.7 into A.6 we have

$$\begin{aligned}\vec{f} &= \epsilon_0 \left[\vec{E}\left(\vec{\nabla}\cdot\vec{E}\right) + \left(\vec{E}\cdot\vec{\nabla}\right)\vec{E}\right] \\ &+\frac{1}{\mu_0}\left[\vec{B}\left(\vec{\nabla}\cdot\vec{B}\right) + \left(\vec{B}\cdot\vec{\nabla}\right)\vec{B}\right] \\ &-\vec{\nabla}\left(\frac{\epsilon_0}{2}\vec{E}^2 + \frac{1}{2\mu_0}\vec{B}^2\right) - \frac{1}{c^2}\frac{\partial\vec{S}}{\partial t}\ . \end{aligned} \tag{A.9}$$

An extra term proportional to $\vec{\nabla}\cdot\vec{B} = 0$ that vanishes has been added to make the expression look more symmetric. Here it is convenient to introduce the Maxwell Stress Tensor $\overleftrightarrow{T}$,

$$\overleftrightarrow{T}_{ij} = \epsilon_0\left(E_iE_j - \frac{1}{2}\delta_{ij}E^2\right) + \frac{1}{\mu_0}\left(B_iB_j - \frac{1}{2}\delta_{ij}B^2\right)\ . \tag{A.10}$$

The component T_{ij} of the tensor is the flux (i.e., the time rate of change) of the i component of the electromagnetic momentum across the plane j. The tensor has units of pressure. The diagonal elements $\overleftrightarrow{T}_{ii}$ represent a force acting in a direction that is perpendicular to the surface. The off-diagonal elements $\overleftrightarrow{T}_{ij}(i \neq j)$ represent a shear stress, acting in a direction that is parallel to the surface. All the terms in A.9 except the last one are exactly the divergence of the Maxwell Stress tensor. Using this notation, the force density can be written in a simpler manner,

$$\vec{f} = \vec{\nabla}\cdot\overleftrightarrow{T} - \frac{1}{c^2}\frac{\partial\vec{S}}{\partial t}\ . \tag{A.11}$$

This equation is the local version of the conservation law of momentum and it holds in every region of the space. If we now want to calculate the total force acting on a material system of volume V enclosed in a boundary surface A that interacts with the electromagnetic field we can integrate equation A.11 and write,

$$\begin{aligned}\vec{F} &= \int_V\left(\vec{\nabla}\cdot\overleftrightarrow{T} - \frac{1}{c^2}\frac{\partial\vec{S}}{\partial t}\right)dV \\ \vec{F} &= \oint_A \overleftrightarrow{T}\cdot d\vec{a} - \frac{d}{dt}\frac{1}{c^2}\int_V \vec{S}dV \end{aligned} \tag{A.12}$$

where we have used the Gauss' theorem to convert a volume integral into a surface one ($d\vec{a}$ is the normal element of area). Now we can identify the three terms of the previous equation and give physical meaning to them.

The first term is the total force acting on volume V and can be written as $\vec{F} = d\vec{P}_{\text{mech}}/dt$, where $\vec{P}_{\text{mech}}$ is the mechanical linear momentum of the system enclosed in V. On the other hand, the last term can be identified as the derivative of the total electromagnetic momentum $\vec{P}_{\text{field}}$ in the volume V,

$$\vec{P}_{\text{field}} = \frac{1}{c^2} \int_V \vec{S} dV \ . \tag{A.13}$$

Putting everything together we can write,

$$\frac{d\vec{P}_{\text{mech}}}{dt} = \oint_A \overleftrightarrow{T} \cdot d\vec{a} - \frac{d\vec{P}_{\text{field}}}{dt} \tag{A.14}$$

and grouping the time-dependent terms it can be expressed as a conservation law

$$\frac{d}{dt}\left(\vec{P}_{\text{mech}} + \vec{P}_{\text{field}}\right) = \oint_A \overleftrightarrow{T} \cdot d\vec{a} \ . \tag{A.15}$$

The right term of the equation accounts for the flux of momentum across the surface A into the volume V. In other words, it is the time rate of change of electromagnetic momentum inside the volume V. This flux acts on the combined system of particles and fields. The left hand side of the equation accounts for the changes in the linear momentum of the particles and the electromagnetic field. Thus it is a balance equation. It stands that *what* gets into volume V (i.e., right hand side) affects *whatever* was inside (i.e., left hand side).

The experimental setup described in section 2.2.1 was designed to measure forces using the conservation of linear momentum. The force exerted by a laser beam on a particle can be measured starting from equation A.15.

Appendix B

Electromagnetic density of flux momentum

The flux momentum is a physical magnitude that represents the amount of linear momentum that crosses a surface. The density of such magnitude can be defined locally in space by means of a vector field. The density of flux momentum of an electromagnetic field is defined in all the regions of the space where the electromagnetic field is present. The Maxwell's Stress tensor is the rigorous expression for the density of electromagnetic flux momentum and is given in Eq. A.10. This appendix shows a less formal way to derive the Maxwell's Stress tensor in order to provide a more intuitive meaning of such quantity.

The density of flux momentum of a plane wave

A plane electromagnetic wave exerts a Lorentz force on a charged particle. The direction of such force depends on the speed of such particle. In general, the movement of the particle will be a combination of circular motion and linear drift, i.e., similar to an helical trajectory. The linear drift suggests that there is a net flux of linear momentum from the plane wave to the particle. Therefore, a plane wave propagating along one direction carries a linear momentum directed along the same direction, which can be transfered to material objects. The Maxwell's Stress tensor quantifies the flux of such linear momentum and the calculation is quite straightforward in this case. A plane wave only has contributions in the direction of wave propagation. The electric field of a polarized plane wave in the y axis propagating in vacuum along the z direction can be written as

$$\vec{E}(x, y, z, t) = \hat{\jmath} E_y = \hat{\jmath} E_0 e^{i(kz - \omega t)} \tag{B.1}$$

where E_0 is the amplitude of the electric field, $k = 2\pi/\lambda$ is the wave number, $\omega = kc$ is the angular frequency and c is the speed of light. The corresponding magnetic field is given by

$$\vec{B}(x, y, z) = \hat{\imath} B_x = \hat{\imath} B_0 e^{i(kz - \omega t)} \tag{B.2}$$

where $B_0 = E_0/c$ according to Maxwell equations. The calculation of the Maxwell's Stress tensor can be done using Eq. A.10

$$\begin{aligned}
\overleftrightarrow{T}_{ij} &= \epsilon_0 \left(E_i E_j - \frac{1}{2}\delta_{ij} E^2 \right) + \frac{1}{\mu_0} \left(B_i B_j - \frac{1}{2}\delta_{ij} B^2 \right) \\
&= \epsilon_0 \left\{ \begin{pmatrix} 0 \\ E_y \\ 0 \end{pmatrix} \begin{pmatrix} 0 & E_y & 0 \end{pmatrix} - \frac{1}{2} E_y^2 \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right\} + \\
&\quad + \frac{1}{\mu_0} \left\{ \begin{pmatrix} B_x \\ 0 \\ 0 \end{pmatrix} \begin{pmatrix} B_x & 0 & 0 \end{pmatrix} - \frac{1}{2} B_x^2 \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right\} \\
&= \epsilon_0 \left\{ \begin{pmatrix} 0 & 0 & 0 \\ 0 & E_y^2 & 0 \\ 0 & 0 & 0 \end{pmatrix} - \begin{pmatrix} E_y^2/2 & 0 & 0 \\ 0 & E_y^2/2 & 0 \\ 0 & 0 & E_y^2/2 \end{pmatrix} \right\} + \\
&\quad + \frac{1}{\mu_0} \left\{ \begin{pmatrix} B_x^2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} - \begin{pmatrix} B_x^2/2 & 0 & 0 \\ 0 & B_x^2/2 & 0 \\ 0 & 0 & B_x^2/2 \end{pmatrix} \right\} \\
&= \epsilon_0 \begin{pmatrix} -E_y^2/2 & 0 & 0 \\ 0 & E_y^2/2 & 0 \\ 0 & 0 & -E_y^2/2 \end{pmatrix} + \\
&\quad + \frac{1}{\mu_0} \begin{pmatrix} B_x^2/2 & 0 & 0 \\ 0 & -B_x^2/2 & 0 \\ 0 & 0 & -B_x^2/2 \end{pmatrix} \\
&= \begin{pmatrix} -\frac{\epsilon_0}{2} E_y^2 + \frac{1}{2\mu_0} B_x^2 & 0 & 0 \\ 0 & \frac{\epsilon_0}{2} E_y^2 - \frac{1}{2\mu_0} B_x^2 & 0 \\ 0 & 0 & -\frac{\epsilon_0}{2} E_y^2 - \frac{1}{2\mu_0} B_x^2 \end{pmatrix} .
\end{aligned} \tag{B.3}$$

Now, by substituting E_x and B_y according to Eqs. B.1 and B.2 and knowing that $E_0 = B_0 c$ and $c^{-2} = \epsilon_0 \mu_0$ the Maxwell's Stress tensor can be written as

$$\overleftrightarrow{T}_{ij} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -\epsilon_0 E_x^2 \end{pmatrix} . \tag{B.4}$$

Note that all the components of the stress tensor vanish but one. It is a diagonal component that represents the density flux of momentum along the direction of propagation of the plane wave. If we perform the product of the stress tensor with a differential area element that is perpendicular to that tensor $d\vec{a} = da\,\hat{k}$, then we obtain the flux of electromagnetic moment (i.e., the force) across this surface

$$\begin{aligned} d\vec{F} = d\vec{\Phi}_{\text{field}} &= \overleftrightarrow{T}_{ij} \cdot d\vec{a} \\ &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -\epsilon_0 E_x^2 \end{pmatrix} \cdot \begin{pmatrix} 0 \\ 0 \\ da \end{pmatrix} \\ &= -\epsilon_0 E_x^2 \cdot da\,\hat{k}\ . \end{aligned} \tag{B.5}$$

If we divide both sides of the previous equation by the area, we can define a density of flux momentum

$$\phi_{\text{field}} = \frac{d\Phi_{\text{field}}}{da} = -\epsilon_0 E_x^2 \cdot \hat{k}\ . \tag{B.6}$$

Now we can recall the definition of the Poynting vector $\vec{S} = \vec{E} \times \vec{H}$ that in the case of a plane wave has the following expression:

$$\begin{aligned} \vec{S} &= \frac{1}{\mu_0}\vec{E} \times \vec{B} = \frac{1}{\mu_0} E_x\, B_y (\hat{\imath} \times \hat{\jmath}) \\ &= \frac{1}{\mu_0} E_x\, B_y\, \hat{k} = \frac{1}{\mu_0} E_x\, \frac{E_x}{c}\, \hat{k} \\ &= \epsilon_0 c E_x^2\, \hat{k} \end{aligned} \tag{B.7}$$

and rewrite Eq. B.6 as

$$\vec{\phi}_{\text{field}} = \frac{\vec{S}}{c} \tag{B.8}$$

which is equivalent to the definition of the density of flux momentum given in Eq. 2.16.

To sum up, the density of flux momentum across a surface perpendicular to the direction of propagation of light can be written in terms of the Poynting vector. This result can be used to calculate the Maxwell's Stress tensor of electromagnetic fields in the ray optics regime, in which the wave can be seen as a bundle of rays, characterized by their Poynting vector.

The density of flux momentum of a spherical wave

The previous calculation of the flux momentum can be extended to spherical waves. The electric and magnetic field of the most simple spherical wave that verifies Maxwell equations is given by

$$\begin{aligned} \vec{E}(r) &= -\frac{1}{r}E_0 e^{i(kr-\omega t)}\hat{\theta} = E_\theta \hat{\theta} \\ \vec{B}(r) &= -\frac{1}{r}B_0 e^{i(kr-\omega t)}\hat{\varphi} = E_\varphi \hat{\varphi} \end{aligned} \tag{B.9}$$

where r is the distance to the origin of coordinates and $\hat{\theta}$ and $\hat{\varphi}$ are the unitary vectors in spherical coordinates. The amplitudes of the electric and magnetic field are related according to $E_0 = cB_0$. Eqs. B.9 represent a diverging spherical wave that emanates from the origin of coordinates. At every time, the amplitude of the wave decreases with the inverse of the distance to the origin ($\sim r^{-1}$), but the energy of the wavefront is spread over a spherical surface of area $4\pi r^2$. This is consistent with the principle of conservation of energy, since the energy per unit of area (i.e., the intensity) of the wave is proportional to the square of the amplitude. Indeed, the calculation of the Poynting vector can be done in spherical coordinates

$$\begin{aligned} \vec{S} &= \frac{1}{\mu_0}\vec{E}\times\vec{B} \\ &= \frac{1}{\mu_0}E_\theta B_\varphi \cdot \hat{\theta}\times\hat{\varphi} \\ &= c\epsilon_0 E_\theta^2 \hat{r} \\ &= \frac{c\epsilon_0 E_0^2}{r^2}\hat{r} \end{aligned} \tag{B.10}$$

where the explicit temporal dependence has been omitted ($e^{2i(kr-\omega t)}$). Note that the intensity is radial and follows an inversely square law. The total power of the wave (L) is given by the sum of the intensity over an arbitrary surface, and L is independent of such surface. In the case of a spherical

surface A, the total power is independent of the radius:

$$\begin{aligned} L &= \int_A \vec{S} \cdot d\vec{a} \\ &= c\epsilon_0 E_0^2 \int_A \frac{1}{r^2} \hat{r} d\vec{a} \\ &= c\epsilon_0 E_0^2 \int_A \frac{1}{r^2} \hat{r} \cdot r^2 \hat{r} d\Omega \\ &= c\epsilon_0 E_0^2 \int_A d\Omega \\ &= 4\pi c\epsilon_0 E_0^2 \end{aligned} \tag{B.11}$$

where $d\vec{a} = r^2 \hat{r} d\Omega$ is the surface element of a spherical shell.

The calculation of the Maxwell Stress tensor can also be done in spherical coordinates. Here we use a more convenient notation to write the components of a tensor. The component $\widehat{e_i e_j}$ represents the tensor product between the two unitary vectors $e_i \otimes e_j$, with $e_i, e_j = \{\hat{r}, \hat{\theta}, \hat{\varphi}\}$.

$$\begin{aligned} \overleftrightarrow{T} &= \epsilon_0 \left(E_\theta \hat{\theta} \otimes E_\theta \hat{\theta} - \frac{1}{2} \left(\widehat{rr} + \widehat{\theta\theta} + \widehat{\varphi\varphi} \right) E_\theta^2 \right) + \\ &\quad + \frac{1}{\mu_0} \left(B_\varphi \hat{\varphi} \otimes B_\varphi \hat{\varphi} - \frac{1}{2} \left(\widehat{rr} + \widehat{\theta\theta} + \widehat{\varphi\varphi} \right) B_\varphi^2 \right) \\ &= \epsilon_0 \left(E_\theta^2 \widehat{\theta\theta} - \frac{1}{2} E_\theta^2 \left(\widehat{rr} + \widehat{\theta\theta} + \widehat{\varphi\varphi} \right) \right) + \\ &\quad + \frac{1}{\mu_0} \left(B_\varphi^2 \widehat{\varphi\varphi} - \frac{1}{2} B_\varphi^2 \left(\widehat{rr} + \widehat{\theta\theta} + \widehat{\varphi\varphi} \right) \right) \\ &= \left(-\frac{\epsilon_0}{2} E_\theta^2 - \frac{1}{2\mu_0} B_\varphi^2 \right) \widehat{rr} + \left(\frac{\epsilon_0}{2} E_\theta^2 - \frac{1}{2\mu_0} B_\varphi^2 \right) \widehat{\theta\theta} + \left(-\frac{\epsilon_0}{2} E_\theta^2 + \frac{1}{2\mu_0} B_\varphi^2 \right) \widehat{\varphi\varphi} \ . \end{aligned} \tag{B.12}$$

The components $\widehat{\theta\theta}$ and $\widehat{\varphi\varphi}$ vanish, because $B_\varphi = E_\theta / c$ and $\mu_0^{-1} = \epsilon_0 c^2$. The only surviving term is the pressure along the radial direction, which coincides with the direction of propagation of the radiation.

$$\overleftrightarrow{T} = -\epsilon_0 E_\theta^2 \widehat{rr} \ . \tag{B.13}$$

Physical interpretation of flux momentum

The Poynting vector $\vec{S}$ is a physical magnitude that represents the local flux of electromagnetic energy. Assuming that an electromagnetic field can be

described as a flux of photons, one can write

$$\vec{S} = E\rho\vec{c} \tag{B.14}$$

where $E = \hbar\omega$ is the energy of one photon and ρ and $\vec{c}$ are the density and the speed of photons respectively. The flux of momentum can be calculated similarly. The momentum of a photon (p) can be obtained from the relativistic expression for the kinetic energy of one particle $E = \sqrt{m^2c^4 + p^2c^2}$, where m is the mass of the particle. A photon has zero mass ($m = 0$) and the resulting expression is

$$E = pc \tag{B.15}$$

and the density of flux momentum can be expressed in terms of the previous relations according to

$$\begin{aligned} \vec{\phi}_{\text{field}} &= p\rho\vec{c} \\ &= \frac{E}{c}\rho\vec{c} = \frac{E\rho\vec{c}}{c} \\ &= \frac{\vec{S}}{c} \, . \end{aligned} \tag{B.16}$$

Thus we have informally justified Eq. 2.16

Appendix C

Experimental setup

This appendix describes in further detail some aspects of the experimental setup.

C.1 Laser diode

The Lumix LU045M200 is a laser diode mounted on a 14-pin *butterfly* chip and connected to a single-mode optical fiber (see Fig. C.1). It has a power of 200 mW and a wavelength of 845 nm. The optical fiber has a Fiber Bragg Grating (FBG) that filters the light to produce a better monochromatic wave. Apart from the laser diode, the chip also has a photodiode, a thermistor and a Thermoelectric Cooler (TEC). The photodiode measures the intensity of light that is produced into the optical resonator cavity of the laser. The thermistor measures the temperature of the laser and the TEC is a device that uses the Peltier effect to refrigerate the laser chip. The laser diode is driven at constant current by a laser diode controller (Thorlabs IP500), which also displays the instantaneous intensity of the laser radiation read by the photodiode. A TEC controller (Thorlabs TCM1000T) controls the temperature of the laser by reading it from the thermistor and performing a proportional/integral (PI) feedback on the TEC. There is one laser driver and one TEC controller for each laser and they are assembled in the so called laser controller box.

C.2 The wiggler

The wiggler (see Fig. C.2a) consists of two concentric brass tubes fixed to a solid block. An optical fiber is located and fixed along the inner tube and remains straight. One side of the fiber is feed by the laser diode and the other

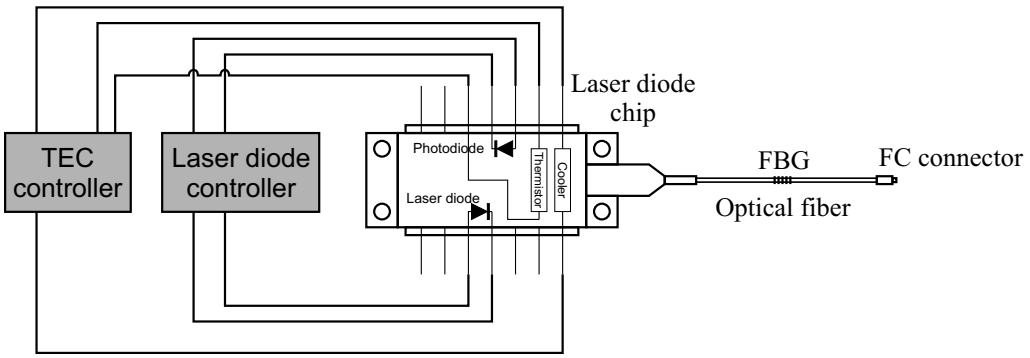


Figure C.1: Laser diode mount. Only 8 pins of the butterfly chip are needed to measure and control the intensity and the temperature of the laser diode. The controllers (intensity and temperature) read the signals and perform the feedbacks.

one is bare. This side protrudes from the inner and outer tubes and exposes its tip, which delivers the laser light to the optical path. The extreme of the outer tube is glued with the fiber and the inner tube fixes a pivot point. Two perpendicular piezoelectric actuators push the outer tube in a hard spherical enlargement located in the middle of the tube, inducing a tilt at the fiber that changes the direction at which the light emerges from the tip of the optical fiber. In the end, the wiggler has a high frequency response (>2 kHz) and a position range of about 11 μm after the displacement is amplified by the microscope objective.

The diverging beam that comes out from the tip of the fiber has a Numerical Aperture (NA) of 0.12 and it is located at the focal plane of a collimating lens of focal length $f = 20.0$ mm that projects the parallel light to the microscope objective, which refocuses the light to form the optical trap (see Fig. C.2b).

C.3 Position Sensitive Detector (PSD)

A PSD is an optoelectronic device that provides an analog output current proportional to the displacement of the centroid of a light spot that hits the sensitive area of the device. Essentially, a PSD is a PIN diode that is sandwiched between two conductive layers connected to 4 electrodes, two by two (see Fig. C.3a). The top layer electrodes measure the position of the spot in the x direction and the other two ones measure the y direction. The diode usually operates in reverse biased mode: the P layer is kept at a negative voltage and the N layer, at a positive. When a photon hits the intrinsic (intermediate) layer of the diode, it produces a pair of hole-electron (see

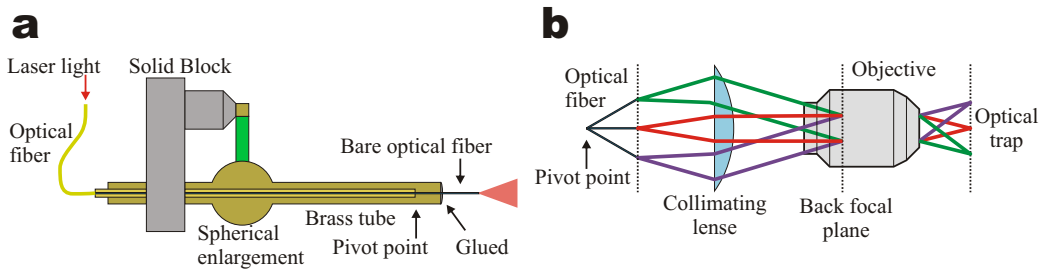


Figure C.2: The wiggler. **(a)** When the piezo (green prism) pushes the spherical enlargement, the outer tube also pushes the optical fiber at the glued point. Since the inner tube remains straight, the optical fiber is bended at the pivot point and the light can be redirected. **(b)** A different bending of the optical fiber (only represented from the pivot point) translates into a different position of the optical trap. According to the definition of Numerical Aperture (NA) NA $= n \sin\theta_m$ (where n is the index of refraction of the medium and θ_m is the marginal ray) the half-angle of the cone of light that emerges from an optical fiber with NA=0.12 is 6.9 °. The beam expands until it reaches a diameter of $\phi = 4.83$ mm ($\tan\theta_m = (\phi/2)/f$, where $f = 20$ mm is the focal length of the collimating lens) before it enters the collimating lens. The entrance pupil of the objective is about 11 mm, so the laser beam is underfilling the objective.

Fig. C.3b) that travel in opposite directions towards the conductive layers. Since there is a large amount of photons in a laser beam, a lot of pairs hole-electron are produced. The excess of charge in the conductors induce a current in the electrodes (see Fig. C.3c). The current of each electrode is proportional to the distance between the spot of light and the electrode. Therefore, the signals of the electrodes can be combined to infer the location of the light spot. If we consider a light beam that hits the sensitive area S of the PSD detector we can write the following relationships between the irradiance of the light ($E(x,y)$) and the output signal of the PSD,

$$\begin{aligned} I_{\text{left}} - I_{\text{right}} &\propto \int_S x\, E(x,y) \cdot dS \\ I_{\text{top}} - I_{\text{bottom}} &\propto \int_S y\, E(x,y) \cdot dS \\ I_{\text{left}} + I_{\text{right}} = I_{\text{top}} + I_{\text{bottom}} &\propto \int_S E(x,y) \cdot dS \end{aligned} \tag{C.1}$$

where $I_{\text{left}}, I_{\text{right}}, I_{\text{top}}, I_{\text{bottom}}$ are the currents of the electrodes and dS is the differential of surface. The differences of current between the electrodes provide the averaged position of the light spot, while the addition of currents provide the total power of light. The addition and subtraction of currents and the application of a constant voltage to the electrodes is performed by using analog electronics in the Preamp board.

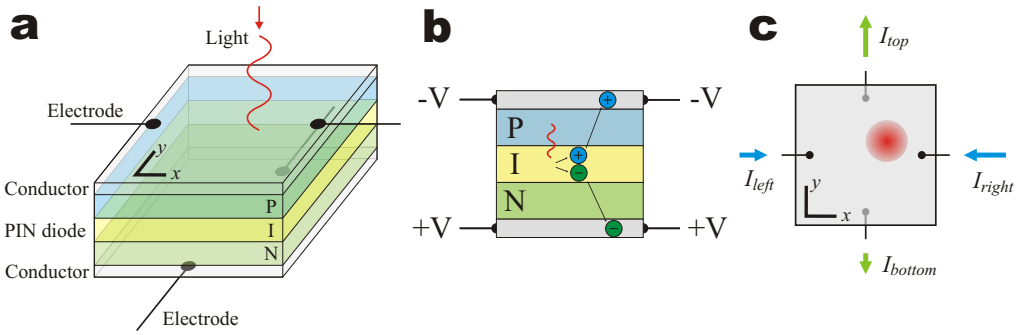


Figure C.3: Position Sensitive Detector. **(a)** Design of a PSD. **(b)** Side view of a PSD. Interaction with light, production of a pair hole-electron and traveling towards the electrodes. **(c)** Front view of a PSD. The currents produced in the electrodes depend on the position at which the light spot hits the detector.

C.4 Fluidics chamber

Figure C.4a shows the three layers of materials that have to be assembled in order to build a fluidics chamber. Using a laser engraver, 6 holes are made in the front coverslip, where the samples are introduced (and removed) into the chamber. The same laser engraver is used to make the channels in the nescofilm layer. A micropipette and two dispenser (or bearing) tubes are also sandwiched between the front coverslip and the nescofilm gasket.

Figure C.4b shows the sealed chamber after the nescofilm has been heated and melted at 119 °C. A zoomed view of the central channel (red frame) shows the disposition of the inner tubes. The micropipette is oriented vertically. The bearing tubes connect the upper and lower channels with the central ones and they are oriented diagonally. The samples introduced in these channels reach the central one by flowing along the bearing tubes.

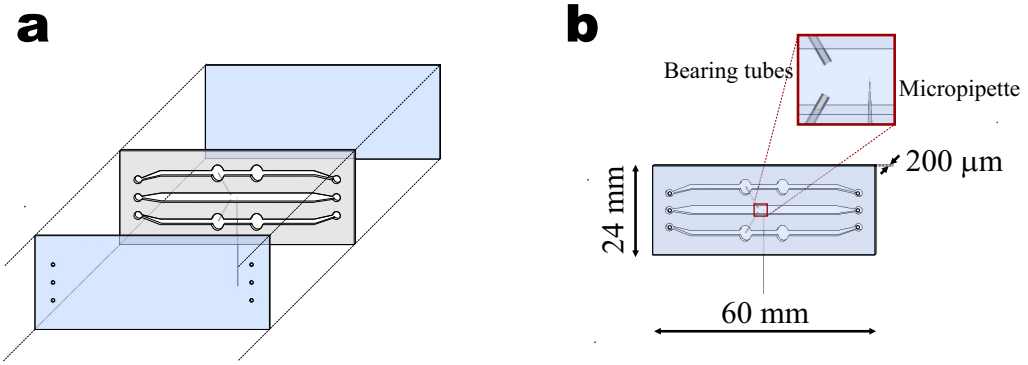


Figure C.4: Fluidics chamber. **(a)** Assembly. **(b)** Dimensions, final appearance of the chamber and zoomed region of the central channel.

Figure C.5 shows how the chamber is mounted in the black frame, which is moved by the xyz-stage. The buffer is introduced into the central channel by means of a large syringe that pushes the buffer stored into the bottle. Two independent small syringes (blue needles) filled with the samples (beads and molecules) are connected to the upper and lower channels. The 3 channels are drained with the opposite holes and the wasted buffer and samples are stored in the trash bottle. The micropipette is connected to another syringe (red needle) filled with air, which is used to immobilize a bead at the tip of the micropipette by suction of air.

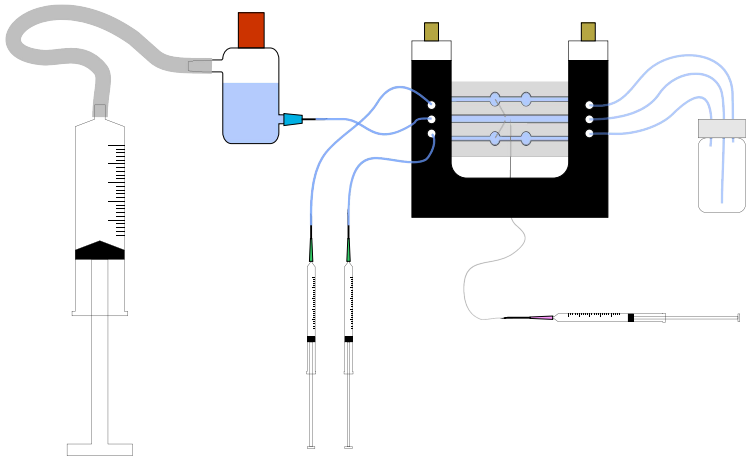


Figure C.5: Tubing of fluidics chamber. The samples flow from left to right.

C.5 Electronic controller

This section describes in further detail the schematics of the electronic controller and the algorithms run by the PICs (see [19] for the exact design). It is convenient to split the description into two parts. The first one only explains the connection between the boards and the devices and the communication between them, i.e., the hardware. The second one, focuses into the firmware, i.e., the software, which is only in the PICs. The main and motor boards use the PIC18F6520 microcontroller, which is a 64-pin surface mount chip. The pins are grouped in ports that allow the communication with other devices. Some of these ports have modules that perform specific tasks such as AD conversion or Pulse Width Modulation.

Hardware

In what follows we describe the internal structure of each board. See 2.12 for a global description of the electronic controller.

- **Main board** (see Fig. C.6). The ComPic is connected to the TrapPics by two ports. Port E controls the flow of data and the Port F is a custom parallel bus devoted to data transfer. Similarly, Port A is used to control the communication between the ComPic and the USB transceiver (DLP-USB245M-G) and the Port D is an 8-bit parallel bus used to write and read data. The 3 PICs use the port C to communicate with other chips of the electronic board via an SPI module. By using the SPI, the ComPic communicates with the MotorPic. The TrapPics communicate with the AD and DA converters. The Chip Select (CS) pin controls the communication. Finally, the Port B is used to burn and debug the firmware in the PICs with an In-Circuit Debugger (ICD).

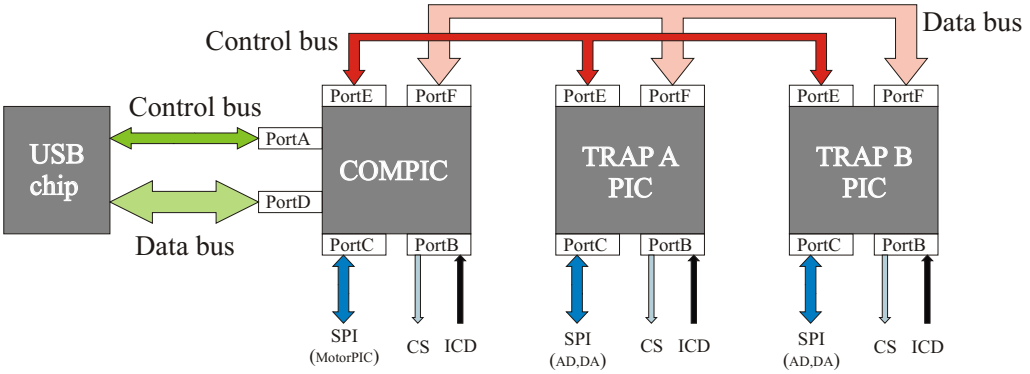


Figure C.6: Main board schematics.

- **Preamp board** (see Fig. C.7a). The current provided by the 4 electrodes of the PSD is converted into a voltage by using operational amplifiers that work as current-to-voltage converters. The current from the IRIS detector is also converted into voltage. The converted voltages from the PSD are combined to obtain the desired signal: subtraction gives the deflections (DX and DY) and addition gives the total power (SUM or PSDSUM). There are 2 Preamp boards per trap: one for position PSD and one for the force PSD.

- **ADC board** (see Fig. C.7b). The signals that come from the Preamp board are stabilized and filtered by using voltage followers and RC analog filters (1 kHz bandwidth) before their enter the AD converter

chip (Allegro MAX1168). It is a 16-bit 8-channel (although 7 channels are used only) converter that communicates with the TrapPic located at the main board via SPI bus.

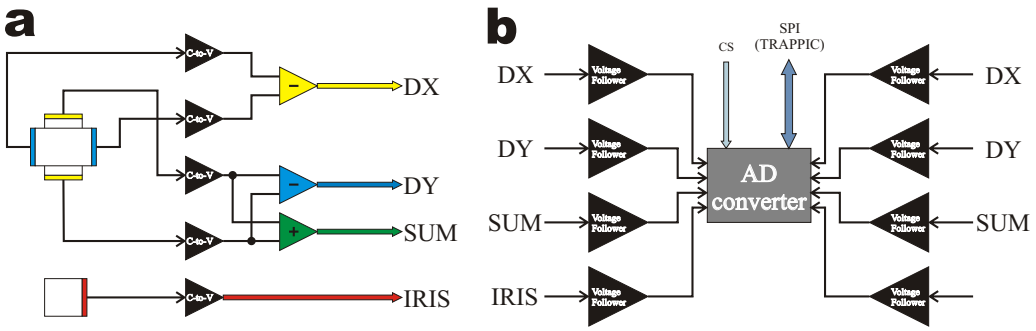


Figure C.7: **(a)** Preamp board. **(b)** ADC board. The bottom right channel of the ADC is not used.

- **Motor board** (see Fig. C.8). This board has a complex combination of analog driving of motors (H-bridges), digital control of motor limits using logical gates (TTL technology) and digital processing of data (MotorPic and counters). The 3 motors are driven independently by 3 H-bridges located in two chips (one chip is shared by the y and z motors). The H-bridges provide the power to the motors in the correct polarity in order to make them rotate in one direction or the other. The direction and the speed of the motors is controlled by the MotorPic by means of Ports A and G respectively. The output of Port G is a Pulse Width Modulation signal that drives the motors. Between the motors and the MotorPic there are logical circuits (TTL protection made with AND, NOR and NOT gates) that prevent the breakage of the motors by disconnecting the H-bridge when the limit switches are exceeded. The MotorPic also reads the status of the limit switches in Port F. The shaft encoders of the motors send their pulses to three 24-bit counters, one for each motor. The MotorPic reads and controls the data from the counters by means of the Ports E and D respectively. Port C is used to power the blue LED that illuminates the optical path and to communicate with the ComPic in the main board. Finally, Port B is used by the ICD to burn the firmware as in the PICs of the main board.

- **DAC board** (see Fig. C.9). This board is shared by the two traps. There are two independent digital to analog converters (Texas Inst DAC8534IP) that communicate with the TrapPics via SPI. The analog output voltage from the DAC is send to 4 op-amps that amplify the

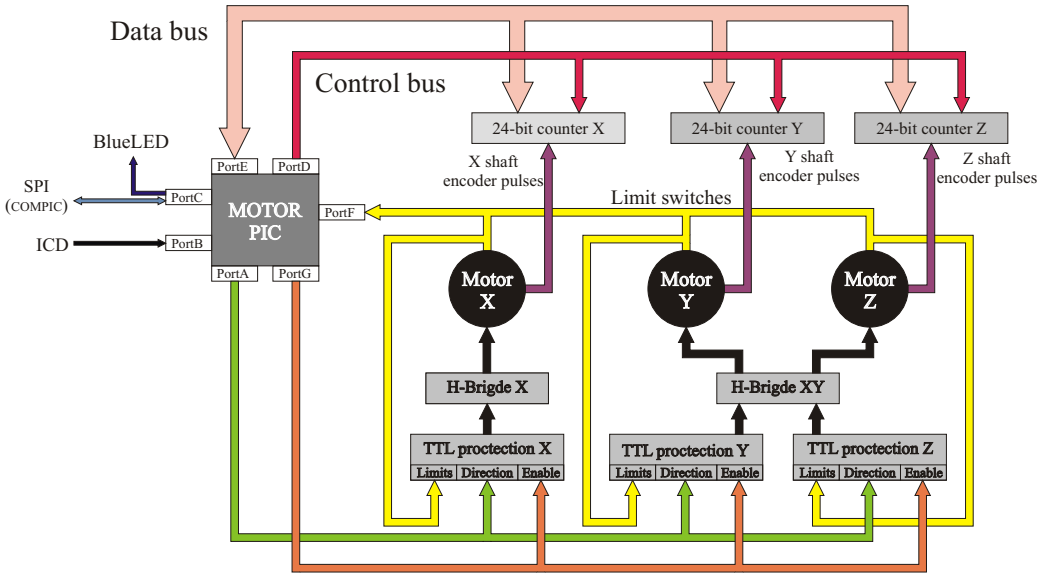


Figure C.8: Motor board.

voltage to a range between 0-150 V. The voltage is applied to the piezos of the wigglers that reposition the two optical traps in the x and y direction.

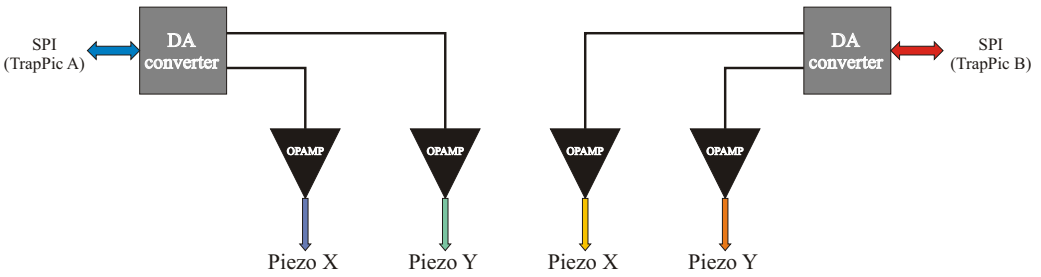


Figure C.9: DAC board.

Firmware

Following, there is a description of the algorithms run by the 4 PICs of the board. The two TrapPics run the same code.

- **ComPic algorithm** (see Fig. C.10). The initialization of the PIC consists in setting up the ports and the variables. The main loop of the firmware (left scheme) continuously communicates with the USB transceiver, sending data and receiving orders from the host. A 4 kHz

timer interrupt is enabled before entering the main loop. So, the interrupt is triggered every 250 μs and the ComPic runs the instructions shown on the right scheme. The ComPic works as the master PIC that synchronizes the communication between the TrapPics and the MotorPic.

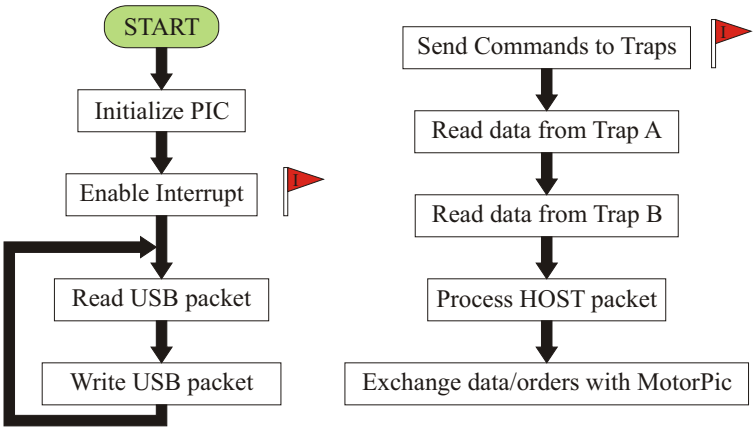


Figure C.10: ComPic algorithm.

- **TrapPic algorithm** (see Fig. C.11). After initializing the code, the TrapPic sends the order to start a new conversion to the AD converter. The watchdog timer is a routine that reboots the PIC if it is not called, which automatically prevents possible code crashes. At this point, the TrapPic waits until the ComPic enters the interrupt subroutine and starts the interchange of data and orders. After processing the orders, the TrapPic performs the selected feedback. Other feedbacks allow to implement custom designed protocols such as oscillation or fast movement of the traps. Once the new position of the trap is calculated, it is written to the DAC and a new loop starts.

- **MotorPic algorithm** (see Fig. C.12). After setting up the 24-bit counter chips, the MotorPic enters the main loop. The MotorPic sends the value of the counter to the ComPic and receives the commands that it has to execute. The most relevant commands involve the movement of the motors. The control of the speed of the motor is also done by the MotorPic. The Time Out subroutine stops the motors in case a Goto command is not able to reach the target and prevents the everlasting movement of the motors. Finally, in the Goto check the motor is stopped if it has reached the desired position and the MotorPic starts another loop.

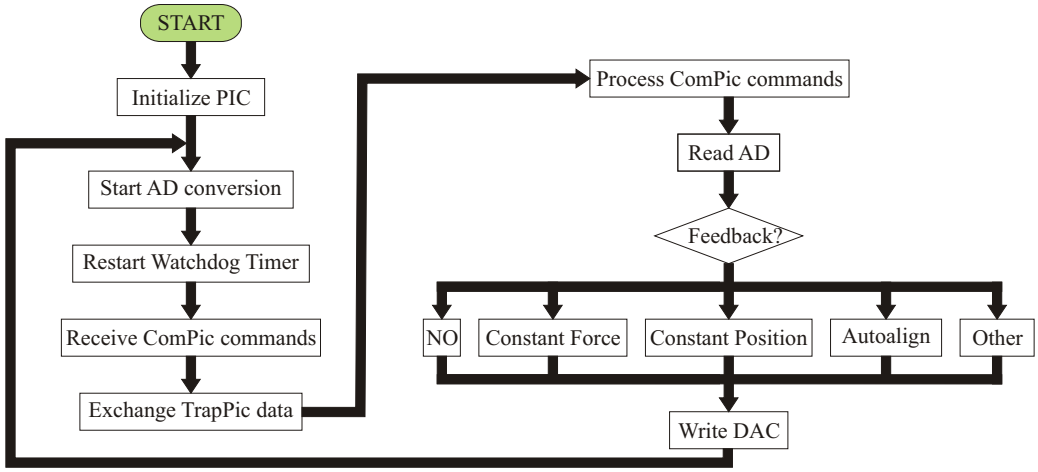


Figure C.11: TrapPic algorithm.

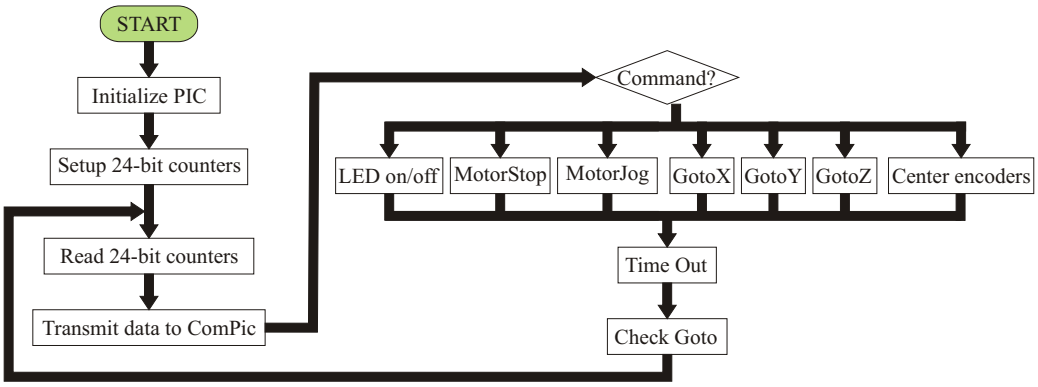


Figure C.12: MotorPic algorithm.

Data structures

The digital data processed by the PICs is organized into packets that are interchanged between them, and between the ComPic and the USB transceiver. There are three types of data depending on their length: the 8-bit, 16-bit and 32-bit numbers. The first ones are status variables, while the second ones are physical magnitudes (PSDs, piezo voltage). The third ones are used by the MotorPic. Following there is a list of the more relevant variables contained in the structures of data in the PICs.

- **mode**. It is a 8-bit variable that holds the working mode of the TrapPic. Depending on its value, the TrapPic works in manual mode or performs some type of feedback such as constant position, constant force or autoalign.

- **gain and rateDivider**. These are two 8-bit variables that are used to perform the feedback algorithms. A feedback is a routine that changes an output controlled variable to produce a desired value (target) to an uncontrolled input variable. For instance, the position of the optical trap (output) can be changed in order to produce a desired force (input variable). All the feedbacks perform the following loop:

$$O_{\text{new}} = O_{\text{old}} + (I_{\text{target}} - I_{\text{actual}}) \times G \tag{C.2}$$

  where O_{new} is the new value that the output variable will have after the calculation, O_{old} is the value that the output variable had before the calculation, I_{target} is the desired value for the input magnitude, I_{actual} is the actual value of the input magnitude and G is a tunable parameter called Gain. This algorithm is named proportional, since the correction in the output variable is proportional to the difference between the actual and the target values of the input magnitude. The Gain controls the response to this difference and it has to be tunned. Large gains produce large responses that might lead to oscillations. Low gains might lead to slow responses that never reach the target value. Finally, the rate divider is another control variable that allows to skip the algorithm every once in a while, in order to produce more stable feedbacks. It also has to be tunned.

- **psdX, psdY, psdSum, iris, leverX, leverY, leverSum**. These are 16-bit variables that hold the data from the photodetectors and there is one set of them for each TrapPic. They can take values between 0-65535 in adu. The variables that represent deflections (psdX, leverY, etc.) can be positive or negative. As a convention, the middle value 32768 is taken as 0, so that positive (negative) values are larger (smaller) that this value. All these variables are sent to the host that has to perform the conversion to physical magnitudes after the calibration procedure.

- **piezoX, piezoY**. There are two of these 16-bit variables for each TrapPic. They represent the voltage of the piezo crystals of the wiggler. There is a conversion between the 0-65535 adu units and the 0-150 V applied to the piezos.

- **positionX, positionY, positionZ**. These are three 32-bit variables used by the MotorPic. These variables hold the values of the 24-bit counters and permit to infer the position of the motor. They must be so long in order to cover the whole range of the motors.

C.6 Host software

The lt application allows the user to interact with the experiment and save the measured data. Following, there is a description of the software.

Files

The code of lt application is split into different C source and header files. The most relevant files are the following:

- **main.c** It is the main file present in all C programs. It calls the routines that set up the variables and the USB port, and it starts the main loop that catches the events of the application by running the function RunApplicationEventLoop().
- **ltdata.c** and **ltdata.h** These two files define the structure of data that contains the variables used by the code, as well as the routines that initiate and give the default values to the variables. The main structure of data is called Data which has several substructures such as,
  - Control, which holds the variables related to the event handling (e.g., keymap, mouse sensitivity, etc.);
  - Instrument, which contains all the status variables of the instrument (calibration factors, running protocol);
  - Quicktime, which stores all the experimental data (distance, force);
  - FileData, which holds the data file parameters (file name, writing speed);
- **ltcontrol.c** and **ltcontrol.h** These files contain the routines that allow the application to communicate with the electronic controller, run the experimental protocols and dispatch the events. The most important routine is ControlEvent().
- **ltosx.c** and **ltosx.h** These files contain the Mac OSX functions that set up the event-handling of the application. The application has two types of events: 1) triggered by the user or 2) triggered by a timer. The first kind of event is handled by routines such as AppEventHandler(), while IdleTimer() is an example of the second kind.
- **ltopengl.c** and **ltopengl.h** All the OpenGL windows that plot data and their rendering are defined here. Each window has its own data

structure called RecContext (for recreation context) and the rendering is performed by the routine DrawGL().

- **ltFileIO.c** and **ltFileIO.h** All the routines related to the input and output files are defined here. Output files are mainly data and comment files. Input files are the calibration file and preferences file (i.e., user defined). The routine ltFileIO() does all these tasks.

- **ltposix.c** and **ltposix.h** These files contain the routines that use the POSIX library to set up a thread that communicates with the USB port (i.e., the electronic controller). The most important routine is the InstrumentTask().

- **main.nib** It is a Mac specific file that contains the information to create the graphical interface of the applications (menu bar, windows, etc.).

Algorithm

The application starts by setting up the variables, functions and routines required for the correct acquisition of data and the dispatch of events (see Fig. C.13). The variables are initialized in InitData() and the menu bar and the graphical interface are set in InitNib(). Afterwards, the event handlers are initialized in InitEvents() which calls two different handlers. The first one is AppEventHandler() and it is responsible to catch the user events and send them to ControlEvent() where they are dispatched. The second one is IdleTime() which is triggered every 1/60 Hz and performs periodic tasks such as writing data files (ltFileIO()) and moving the piezos according to the different protocols (AutoMovePiezos()). The InitControl() routine creates the default windows (trap and motor windows, oscilloscope, PSDs, etc.) and starts the thread that opens the USB port to communicate with the electronic controller (InstrumentTask()). Once all the initialization has been completed, the algorithm enters the RunApplicationEventLoop() which keeps running the application until the user quits.

The ControlEvent() is one of the central routines and it performs two types of tasks: 1) give orders to the instrument (e.g., center traps, motor gotos) by calling CommandInstrument() (the orders will be send in the next call of InstrumentTask()) and 2) create and setup windows by calling MakeWindow(). The windows can be either of two types: dialog windows and OpenGL windows. The first ones allow to modify the parameters of the protocols and the routine UtilityWindow() sets up the event handlers of them. The second

ones are graphical windows that display the data measured by the instrument. The routine GraphWindow() creates and installs two different types of event handlers on these windows: 1) a timer event (timerContextCB()) that refreshes the graphics at a frequency of 60 Hz by using the DrawGL() routine and 2) a keyboard event (graphKeyEventHandler()) that allows to control the properties of the graph (zoom, shift, rotation, etc.).

To sum up, the application continuously runs a central loop that catches and dispatches all the possible events, which are managed by several event handlers that report to that loop. The events include user actions (keyboard, mouse) and periodic tasks (graphical rendering, pulling protocols). Besides, a parallel thread communicates with the instrument by using the USB port.

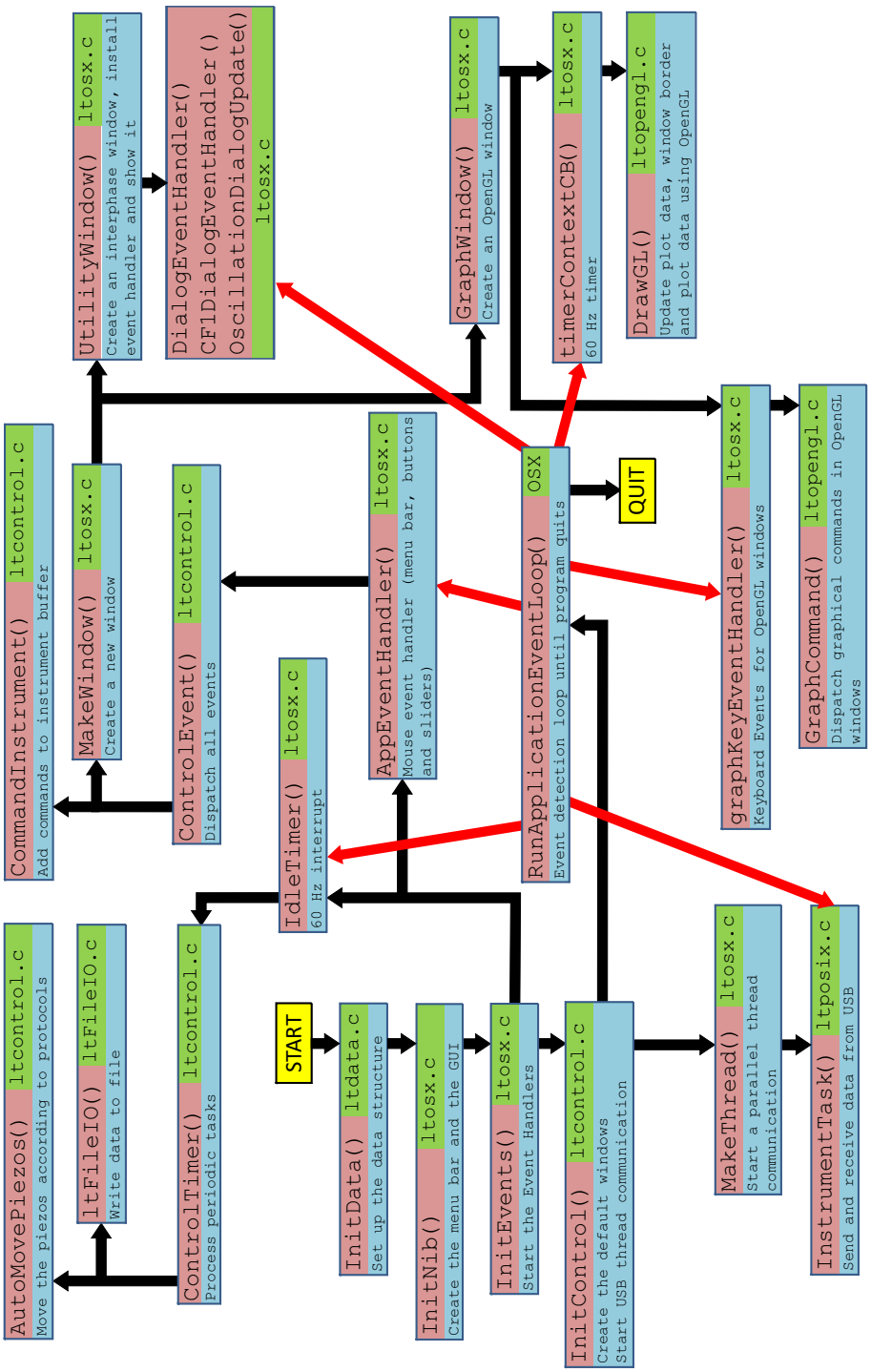


Figure C.13: Host algorithm (see text). The boxes represent the most important routines of the lt application. Each box indicates the name of the routine, a brief description of it and the file where it is defined. Horizontal and vertical arrows indicate the natural flow of the algorithm. Tilted arrows indicate the routines that report events to the main loop.

Appendix D

Design of new protocols

This appendix describes the recipe to be followed in order to implement a new protocol. A new protocol can be executed by the host computer or the electronic boards. In both cases, the host algorithm has to be coded and only in the last case, the PICs firmware has to be recoded. In the first section of this appendix we show how to modify the host algorithm. In the second one, we show how to code the PICs. It is not necessary to understand the whole host algorithm (see the section Host Software in Appendix C) to be able to code a new protocol.

Coding the host algorithm

Most of the tasks performed by the host algorithm (communication, graphical rendering, event handler, etc.) do not need to be modified when designing a new protocol. So the user only has to expand a little piece of code to write a new protocol. There are two different steps in the designing process: 1) the algorithm of the protocol itself (what the instrument has to do) and 2) the integration of the algorithm in the user interface of the lt application. The first one involves control theory and the algorithm has to be written in C. Thus, the algorithm has to include the orders to read the inputs (time, force, position, limits); take decisions (decrease force, switch direction, etc.); and send orders to the instrument (move trap, move motor, etc.). This first step of the design depends on the type of protocol to be implemented. The second step, i.e., the integration on the application, is common for all protocols. It is only necessary to prepare the graphical interface (buttons, sliders, windows of parameters) and the events (start/stop protocol).

1. Writing the protocol

The instructions that the protocol has to execute are written in a C function. So we have to create a function that contains all the orders for our new designed protocol. For instance, AutoMovePiezos(), Oscillate(), ForceRamp() are names of functions that perform a specific protocol. These functions are located in the ltcontrol.c file and predefined in the ltcontrol.h header file. The ControlTimer() function is responsible for calling the protocol algorithms 60 times per second. Therefore, a IF statement has to be added in ControlTimer() to call our new designed function when it is activated by the user.

The algorithm of the new protocol has to be designed considering that it will be called (and run) at 60 Hz. The following example shows the pseudocode to produce a pulling protocol at constant pulling rate.

```
IF <forceLimit> OR <positionLimit> TRUE
   <pullingRate> = - <pullingRate>;
<newPosition> = ( <previousTime> - <Time> ) * <pullingRate>;
<previousTime> = <Time>;
```

The protocol algorithm will have direct access to all the data structure that contains the information about the status of the instrument (time, force, motor and trap position, etc.). On the other hand, the CommandInstrument() function is the most important function to give orders to the instrument such as set the feedback algorithms (autoalign, constant force, passive mode), move the motors (gotos, motor speed), move the traps (set the piezos output voltage), etc.

So the protocol is written in a function that reads the status of the instrument and gives orders at 60 Hz.

2. Integration of the protocol into the algorithm

In general, the integration of the protocol needs: 1) a start/stop control and 2) a dialog window to introduce the parameters of the protocol. In order to do so, several modifications of the code have to be done.

1. Creation of new variables.
   - Define all the additional variables required to hold the parameters of the protocol (e.g., speed, force limit, waiting time, etc.) in the structure data->instrument located at the file ltdata.h.
   - In case these new defined variables need default values, initialize them in the function InitInstrument() located at the file ltdata.h.

- Define an event variable that starts the protocol in the enumeration of events located in the file ltcontrol.h.
- Define an event variable that opens the dialog window. This variable has to be included in the enumerated list of window events located in the file ltcontrol.h.

2. Changes in the nib file. The main.nib file contains the information about the user interface, which can be easily modified using the application Interface Builder. The nib file has to be modified in order to include the new graphical objects (e.g., windows, buttons, sliders, text boxes, etc.) required to control the protocol.

   - The graphical interface is organized so that all the start/stop controls of the protocols are located in a radio button. So a new option in the radio button that selects our new designed protocol has to be created.
   - Create the graphical design of the dialog window (labels, text boxes, enter button, etc.) where the user introduces the values of the parameters of the protocol.

3. Changes in already existing functions.

   - Define mouse clicks that activate the protocol as a new case in the AppEventHandler() function located in the ltosx.c.
   - Define mouse clicks that make the dialog window as a new case in the AppEventHandler() function located in the ltosx.c.
   - Include the instructions to make the dialog window in the UtilityWindow() function in the ltosx.c file. The instructions are: 1) read and create the window from the nib file, 2) install the event handler, 3) show the window, and 4) dispose other windows.
   - Include the calling of the protocol algorithm in the ControlTimer() function in the ltcontrol.c file.
   - Include the event that creates the dialog window in the ControlEvent() function in the ltcontrol.c file.
   - Include the event that starts the protocol in the ControlEvent() function in the ltcontrol.c file. Other automatic orders that have to be simultaneously started along with the protocol (e.g., write a comment, switch writing speed, etc.) are included here.

4. Definition of new functions.

- Define the *get values* dialog function in the ltosx.c file. This function reads the values of the parameters located in the text boxes of the dialog window and stores them in the variables defined in the data->instrument structure. This function is activated when the protocol starts or the enter button is pushed.
- Define the *update* dialog function in the ltosx.c file. This function reads the values stored in the data->instrument structure and updates the dialog box accordingly.
- Define the event handler (i.e., a C function) for the dialog window in the ltosx.c file. This function detects the events produced within the window and calls the two previous functions to get or update the values.

After all these changes have been made to the code, this has to be recompiled in order to make them functional.

Coding the firmware of the PICs

As mentioned before, the implementation of a protocol in the PICs requires to modify the host software too. Once the start/stop event and the dialog window are created we have to focus on the firmware. In other words, when we modify the firmware we are expanding the capabilities of the CommandInstrument(). This function has a few options but they are very complete to design any protocol. It might be possible that we need some options not included in this function. When this happens, it is time to modify the firmware.

In general, only the TrapPic firmware will have to be modified. Both TrapPics have information of the position of the optical trap and the force applied and they are capable of giving orders to the DA converters that finally move the piezos. So the TrapPics have the inputs, the outputs and the computational requirements to perform a protocol.

Figure C.11 shows a scheme of the TrapPic algorithm. A new protocol is introduced as a *Feedback* routine that the TrapPic runs when it is activated. It requires coding in C, compiling and burning the firmware into the PICs. The following example shows the pseudocode of a pulling protocol implemented in the TrapPic. It is much more elementary than the host algorithm. In this case, the voltage applied to the piezo (voltagePiezoY) is increased by a fixed amount (deltaV) one of every certain cycles (skipMove). The control of force limits and directions is performed by the host algorithm.

```
IF <skipMove> TRUE
   <voltagePiezoY> = <voltagePiezoY>;
ELSE
   <voltagePiezoY> = <voltagePiezoY> + <deltaV>;
```

Appendix E

Sequences of DNA constructs

6.8 kbp sequence

The target of the BamHI restriction enzyme is 5′-ggatccc-3′ and cleaves the λ-DNA at several points. The target we are interested in is located between the bases 41732-41738 of λ-DNA (see Fig. E.1a). The loop is a single oligonucleotide that folds in such a manner that exposes its BamHI end (see Fig. E.1b). The handles are synthesized from 3 different oligonucleotides. Two of them hybridize leaving a sticky end that is complementary to the cosR end of λ-DNA (orange and magenta oligos in Fig. E.1c) and another one completes the handles (violet oligo in Fig. E.1c).

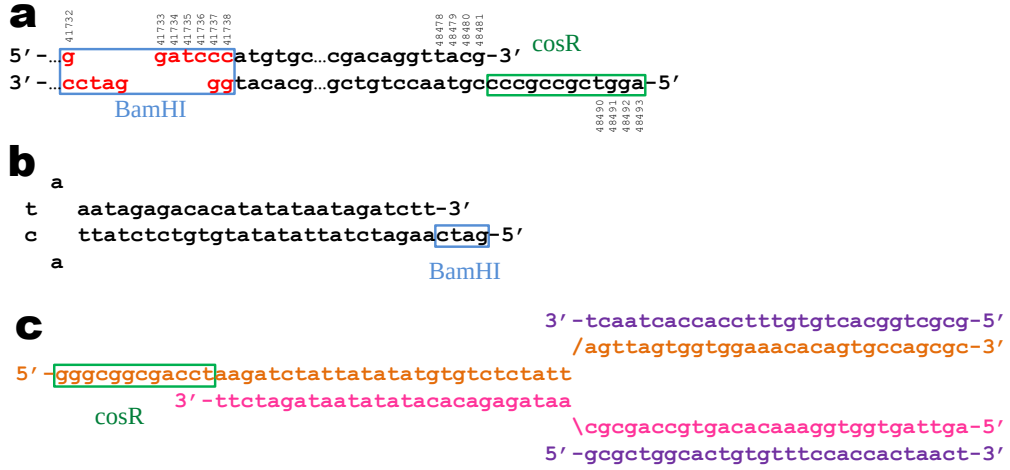


Figure E.1: 6.8 kb sequence. **(a)** Fragment of λ-DNA. The position of the base in λ-DNA is indicated with a number next to it. Red highlighted sequence is the BamHI target. The blue square shows how the enzyme cleaves the target. The green square shows the cosR end sequence. **(b)** Loop. **(c)** Handles.

2.2 kbp sequence

Similarly, the target of the SphI restriction enzyme is 5′-gcatgc-3′. The digestion of the target located between the bases 2212-2217 of λ-DNA (see Fig. E.2a) produces a fragment of DNA of ≈ 2.2 kbp. As in the 6.8 kbp sequence, the loop is a single oligonucleotide that exposes its SphI end (see Fig. E.2b). The combination of three different oligonucleotides produce the handles, which form the cosL end (see in Fig. E.2c).

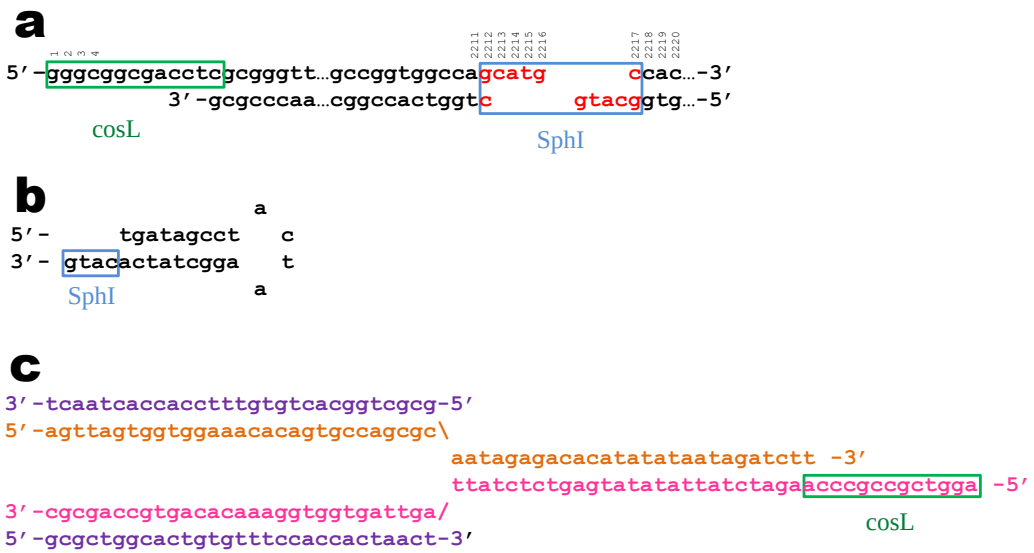


Figure E.2: 2.2 kbp sequence. **(a)** Fragment of λ-DNA. **(b)** Loop. **(c)** Handles.

Appendix F

Elastic energy of polymers

Polymers are chemical substances formed by the union of single subunits (called monomers) in a chain. Here, we will focus on the study of the elasticity of linear polymers, that can be treated as 1-dimensional chains. The model is treated from the point of view of statistical physics and the goal of this section is to characterize the elastic energy and the equation of state of the polymer.

Worm Like Chain (WLC) model

The WLC model is used to describe the elastic response of semi-flexible polymers. The main property of these polymers is that some energy is required to bend them. The model assumes that the polymer is a continuous rod with a characteristic bending energy. The persistence length (l_p) is the only parameter that controls the flexibility of the polymer. The other important parameter is the contour length (L_0) that determines the length of the polymer chain.

The exact solution of the model requires to solve a transcendental equation numerically. However, Marko & Siggia [40] proposed the following approximated solution for the Force vs. Extension Curve in the WLC model

$$f(x) = \frac{k_B T}{4 l_p} \left(\left(1 - \frac{x}{L_0} \right)^{-2} - 1 + 4 \frac{x}{L_0} \right) \tag{F.1}$$

where f is the force applied and the ends of the polymer, x is the molecular extension, k_B is the Boltzmann constant and T is the temperature. This equation was derived after calculating the partition function of the model, computing its free energy and approximating the equation of state. However, this equation of state does not come from the actual free energy. So if we

want to know the free energy $G(x)$ of this expression, we need to calculate its integral which is

$$\begin{aligned} G(x) &= \int_0^x f(x')\,dx = \\ &= \frac{k_B T L_0}{4 l_p}\left(\left(1-\frac{x}{L_0}\right)^{-1} - 1 - \frac{x}{L_0} + 2\left(\frac{x}{L_0}\right)^2\right) . \end{aligned} \tag{F.2}$$

This expression is consistent with Eq. F.1 because the elastic free energy is $G = 0$ when the extension of the polymer is $x = 0$, i.e., when no force is applied at the ends of the polymer.

There are two main extensions of this model [160]. The first one involves polynomial corrections to Eq. F.1 up to order $\mathcal{O}(x/L_0)^7$ so that the equation of state is as close as possible to the exact solution. In the case of dsDNA, this correction permits to fit the measured FEC with the same relative error at all range of forces. The second one involves an enthalpic correction that allows the contour length of the polymer to increase at high stretching forces. It enters in Eq. F.1 as a stretch modulus and it converts the vertical asymptote located at $x = L_0$ into an oblique one.

In our unzipping problem, though, these two corrections are not necessary. We are interested in the elastic energy of the handles, so the total area under the FDC is hardly affected by the polynomial correction. Moreover, the unzipping of DNA occurs below 20 pN, where the effect of the enthalpic correction is small. Indeed, using the enthalpic correction to fit the FEC at low forces (<20-25 pN) induces some instability in the values of the fitting parameters. In particular, the persistence length increases whereas the stretch modulus decreases [161].

Freely Jointed Chain (FJC) model

This is the simplest model for the elasticity of polymers. It assumes that the polymer is made of rigid monomers of length b (also known as the Kuhn length) connected by junctions that are free to rotate and they do not interact with each other. The contour length of the polymer (L_0) is determined by the number of monomers N so that $L_0 = Nb$. It is a suitable model for flexible polymers and the elastic response is purely entropic.

The problem can be mapped to a paramagnetic system, where the magnetization is the analogous to the extension and the external magnetic field is the analogous to the force. Differently from the WLC model, the equation

of state can be calculated exactly and it is given by,

$$x(f) = L_0 \cdot \left(\coth\left(\frac{bf}{k_B T}\right) - \frac{k_B T}{bf} \right) \tag{F.3}$$

were x is the extension, f is the force applied, k_B is the Boltzmann constant and T is the temperature. Strictly speaking, $L_0 = Nb$ where b is the Kuhn length. However, the resulting FEC does not describe correctly the experimental data. In order to fit the data, this model (Eq. F.3) is used by taking $L_0 = Nd$ where d is the distance between monomers and b is the Kuhn length (different from d).

The free energy of the model can also be calculated by integration of Eq. F.3. Note that the equation of state is given in terms of the extension vs. the force, so the calculation of the free energy requires an integration by parts (see Sec. 3.4.2 for further details). Here we focus on the integral of Eq. F.3 which is given by,

$$\begin{aligned} G(f) &= \int_0^f x(f')\, df = \\ &= \frac{k_B T L_0}{b} \left[\ln\left(\sinh\frac{bf}{k_B T}\right) - \ln\frac{bf}{k_B T} \right] . \end{aligned} \tag{F.4}$$

For completeness, here we write the free energy of the FJC at controlled position

$$\begin{aligned} G(x) &= xf - \int_0^f x(f')\, df = \\ &= xf - \frac{k_B T L_0}{b} \left[\ln\left(\sinh\frac{bf}{k_B T}\right) - \ln\frac{bf}{k_B T} \right] \end{aligned} \tag{F.5}$$

where $x = x(f)$ according to Eq. F.3.

Appendix G

Elastic fluctuations in the mesoscopic model

In order to obtain a full description of the model that includes the effects of the elastic fluctuations of the polymers, we have to split the contribution of the handles and the ssDNA, so that they can have independent lengths (see Fig. G.1).

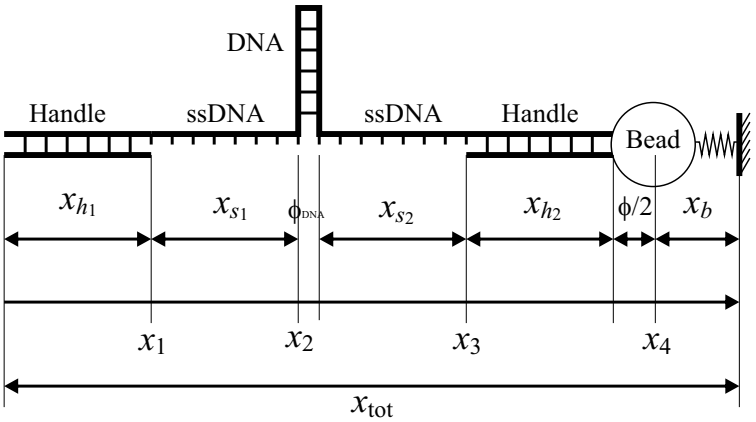


Figure G.1: Full model. Each element is treated indepentdently.

The total energy of the system is then written as,

$$\begin{aligned} G_{\rm tot}(x_{h_1}, x_{s_1}, x_{s_2}, x_{h_2}, x_b, n) &= G_{h_1}(x_{h_1}) + G_{s_1}(x_{s_1}, n) + G_{\rm DNA}(n) \\ &+ G_{s_2}(x_{s_2}, n) + G_{h_2}(x_{h_2}) + E_b(x_b) \end{aligned} \tag{G.1}$$

where the explicit dependencies of the handles are given by Eq. F.2; either Eq. F.2 or Eq. F.5 give the dependency of the ssDNA; Eq. 3.17 gives the free

energy of the duplex and Eq. 3.16 gives the energy of the bead in the optical trap. The total extension of the system is the control parameter and it can be written as

$$x_{\text{tot}} = x_{h_1} + x_{s_1} + \phi_{\text{DNA}} + x_{s_2} + x_{h_2} + \phi/2 + x_b \tag{G.2}$$

where ϕ_{DNA} is the diameter of the DNA duplex and ϕ is the diameter of the bead. The partition function can be calculated by summing over all the possible states of the system according to

$$Z(x_{\text{tot}}) = \sum_{n=0}^{N} \int_{\mathbb{R}^5 \geq 0} e^{-\beta\, G_{\text{tot}}(x_{h_1}, x_{s_1}, x_{s_2}, x_{h_2}, x_b, n)|_{x_{\text{tot}}}} \cdot dx_{h_1} dx_{s_1} dx_{s_2} dx_{h_2} dx_b \tag{G.3}$$

where the integral is extended over all positive values of the extensions $x_{h_1}, x_{s_1}, x_{s_2}, x_{h_2}$ and x_b; provided that Eq. G.2 is fulfilled. Now, the calculation of this integral is quite complex because the extensions in Eq. G.3 are restricted by Eq. G.2, so they are not independent variables. Our goal is to write the total free energy of the system in terms of independent variables that can be integrated separately. So we can define a new set of positively defined variables x_1, x_2, x_3 and x_4 (see Fig. G.1) that indicate the position of the ends of the elastic elements, which are related to the extensions according to the following relations:

$$\begin{cases} x_{h_1} &= x_1 \\ x_{s_1} &= x_2 - x_1 \\ x_{s_2} &= x_3 - x_2 - \phi_{\text{DNA}} \\ x_{h_2} &= x_4 - \phi/2 - x_3 \\ x_b &= x_{\text{tot}} - x_4 \end{cases} \tag{G.4}$$

and allow us to rewrite Eq. G.1 as

$$\begin{aligned} G_{\text{tot}}(x_1, x_2, x_3, x_4, x_{\text{tot}}, n) &= G_{h_1}(x_1) + G_{s_1}(x_2 - x_1, n) + \\ &+ G_{\text{DNA}}(n) + G_{s_2}(x_3 - x_2 - \phi_{\text{DNA}}, n) + \\ &+ G_{h_2}(x_4 - \phi/2 - x_3) + E_b(x_{\text{tot}} - x_4) \end{aligned} \tag{G.5}$$

Now the total energy of the system can be calculated for different extensions of the elements without affecting the total extension of the system. For instance, we can give different values to x_1 that modify the extensions of x_{h_1} and x_{s_1} without changing the value of x_{tot}. Now, the partition function can be calculated by integrating over the new variables of the elements according to

$$Z(x_{\text{tot}}) = \sum_{n=0}^{N} \int_{\mathbb{R}^4} e^{-\beta G_{\text{tot}}(x_1, x_2, x_3, x_4, x_{\text{tot}}, n)} \cdot dx_1\, dx_2\, dx_3\, dx_4 \tag{G.6}$$

where the integral is extended over all values of $-\infty < x_i < +\infty$, $i = 1, \ldots, 4$ while keeping the value of $x_{\rm tot}$ constant. Note that Eqs. G.4 define a set of limits of integration for the new variables that do not coincide with the ones used in Eq. G.6. Nevertheless, we extend the limits of integration of the newly defined variables over all the space. We can do this because the contribution to the integral vanishes for large extensions, due to the exponential dependence of the energy. The advantage of such approximation is that we will have to perform Gaussian integrals, which are easy to calculate when the limits of integration expand over all the space.

For each value of $x_{\rm tot}$ and n, there is always a combination of (x_1, x_2, x_3, x_4) that minimizes the total energy in Eq. G.5. It is convenient to define this global minimum as $G_{\rm min}(x_{\rm tot}, n)$ because it allows to perform the integral by using a saddle-point approximation. If we define $\vec{x} = (x_1, x_2, x_3, x_4)$ as the vector that contains all the variables, we can perform a Taylor expansion of Eq. G.5 around the minimum $\vec{x}_{\rm m}$ of $G_{\rm tot}(\vec{x}, x_{\rm tot}, n)$ up to order $\mathcal{O}(\vec{x} - \vec{x}_{\rm min})^2$ according to,

$$\begin{aligned} G_{\rm tot}(\vec{x}, x_{\rm tot}, n) \;\simeq\; & G_{\rm min}(x_{\rm tot}, n) + \vec{\nabla}_{\vec{x}} G_{\rm tot}(\vec{x}, x_{\rm tot}, n)\Big|_{\vec{x}_{\rm min}} \cdot (\vec{x} - \vec{x}_{\rm min}) \\ & + \frac{1}{2} (\vec{x} - \vec{x}_{\rm min})^T \cdot \mathbf{H}\Big(G_{\rm tot}(\vec{x}, x_{\rm tot}, n)\Big)_{\vec{x}_{\rm min}} \cdot (\vec{x} - \vec{x}_{\rm min}) \end{aligned} \tag{G.7}$$

where $(\vec{x} - \vec{x}_{\rm min})^T$ stands for the transposed vector and $\mathbf{H}$ is the Hessian matrix calculated at fixed $x_{\rm tot}$ and n. Formally, the Hessian matrix is a quadratic form that acts over the vector $(\vec{x} - \vec{x}_{\rm min})$ and its transposed $(\vec{x} - \vec{x}_{\rm min})^T$ and returns the change in $G_{\rm tot}$. Note that the variables of the function are $\vec{x}$, while $x_{\rm tot}$ and n are fixed parameters. Since the Taylor expansion has been performed around the minimum, the gradient (i.e., first derivative) of the function vanishes and the second term in the right side of Eq. G.7 does not contribute.

In order to obtain the Hessian matrix, we have to calculate all the second order derivatives of Eq. G.1. The first derivatives of elastic energy terms give forces and the second derivatives give stiffnesses. So for the first derivatives of Eq. G.5 we can write

$$\begin{aligned} \left.\frac{\partial\, G_{\rm tot}(\vec{x})}{\partial x_1}\right|_{x_{\rm tot}, n} &= \frac{\partial\left(G_{h_1}(x_1) + G_{s_1}(x_2 - x_1)\right)}{\partial x_1} \\ &= f_{h_1}(x_1) - f_{s_1}(x_2 - x_1) \end{aligned}$$

$$\left.\frac{\partial\, G_{\rm tot}(\vec{x})}{\partial x_2}\right|_{x_{\rm tot}, n} = \frac{\partial\left(G_{s_1}(x_2 - x_1) + G_{s_2}(x_3 - x_2 - \phi_{\rm DNA})\right)}{\partial x_2}$$

$$
\begin{aligned}
&= f_{s_1}(x_2 - x_1) - f_{s_2}(x_3 - x_2 - \phi_{\mathrm{DNA}}) \\
\left.\frac{\partial\, G_{\mathrm{tot}}(\vec{x})}{\partial x_3}\right|_{x_{\mathrm{tot}},n} &= \frac{\partial\left(G_{s_2}(x_3 - x_2 - \phi_{\mathrm{DNA}}) + G_{h_2}(x_4 - \phi/2 - x_3)\right)}{\partial x_3} \\
&= f_{s_2}(x_3 - x_2 - \phi_{\mathrm{DNA}}) - f_{h_2}(x_4 - \phi/2 - x_3) \\
\left.\frac{\partial\, G_{\mathrm{tot}}(\vec{x})}{\partial x_4}\right|_{x_{\mathrm{tot}},n} &= \frac{\partial\left(G_{h_2}(x_4 - \phi/2 - x_3) + E_b(x_{\mathrm{tot}} - x_4)\right)}{\partial x_4} \\
&= f_{h_2}(x_4 - \phi/2 - x_3) - f_b(x_{\mathrm{tot}} - x_4) \qquad \text{(G.8)}
\end{aligned}
$$

where $f_i(x_i)$ $i = h_1, s_1, h_2, s_2$ are the FEC of the elastic elements and the explicit dependencies of G_{tot} on x_{tot} and n have not been written. The components of the Hessian matrix will be given by

$$
\mathbf{H}\left(G_{\mathrm{tot}}(\vec{x})\right) = \begin{pmatrix}
\dfrac{\partial^2 G_{\mathrm{tot}}}{\partial x_1^2} & \dfrac{\partial^2 G_{\mathrm{tot}}}{\partial x_1 \partial x_2} & \dfrac{\partial^2 G_{\mathrm{tot}}}{\partial x_1 \partial x_3} & \dfrac{\partial^2 G_{\mathrm{tot}}}{\partial x_1 \partial x_4} \\
\dfrac{\partial^2 G_{\mathrm{tot}}}{\partial x_2 \partial x_1} & \dfrac{\partial^2 G_{\mathrm{tot}}}{\partial x_2^2} & \dfrac{\partial^2 G_{\mathrm{tot}}}{\partial x_2 \partial x_3} & \dfrac{\partial^2 G_{\mathrm{tot}}}{\partial x_2 \partial x_4} \\
\dfrac{\partial^2 G_{\mathrm{tot}}}{\partial x_3 \partial x_1} & \dfrac{\partial^2 G_{\mathrm{tot}}}{\partial x_3 \partial x_2} & \dfrac{\partial^2 G_{\mathrm{tot}}}{\partial x_3^2} & \dfrac{\partial^2 G_{\mathrm{tot}}}{\partial x_3 \partial x_4} \\
\dfrac{\partial^2 G_{\mathrm{tot}}}{\partial x_4 \partial x_1} & \dfrac{\partial^2 G_{\mathrm{tot}}}{\partial x_4 \partial x_2} & \dfrac{\partial^2 G_{\mathrm{tot}}}{\partial x_4 \partial x_3} & \dfrac{\partial^2 G_{\mathrm{tot}}}{\partial x_4^2}
\end{pmatrix} \qquad \text{(G.9)}
$$

which can be computed from Eqs. G.8 according to

$$
\begin{aligned}
h_{11} = \left.\frac{\partial^2 G_{\mathrm{tot}}(\vec{x})}{\partial x_1^2}\right|_{x_{\mathrm{tot}},n} &= \frac{\partial}{\partial x_1}\left(\frac{\partial G_{\mathrm{tot}}}{\partial x_1}\right) \\
&= \frac{\partial}{\partial x_1}\Big(f_{h_1}(x_1) - f_{s_1}(x_2 - x_1)\Big) \\
&= k_{h_1}(x_1) + k_{s_1}(x_2 - x_1) \\
&= k_{h_1}(x_{h_1}) + k_{s_1}(x_{s_1}) \\
h_{21} = \left.\frac{\partial^2 G_{\mathrm{tot}}(\vec{x})}{\partial x_2 \partial x_1}\right|_{x_{\mathrm{tot}},n} &= \frac{\partial}{\partial x_2}\left(\frac{\partial G_{\mathrm{tot}}}{\partial x_1}\right) \\
&= \frac{\partial}{\partial x_2}\Big(f_{h_1}(x_1) - f_{s_1}(x_2 - x_1)\Big) \\
&= -k_{s_1}(x_2 - x_1) \\
&= -k_{s_1}(x_{s_1})
\end{aligned}
$$

$$
\begin{aligned}
h_{22} = \left.\frac{\partial^2 G_{\text{tot}}(\vec{x})}{\partial x_2^2}\right|_{x_{\text{tot}},n} &= \frac{\partial}{\partial x_2}\left(\frac{\partial G_{\text{tot}}}{\partial x_2}\right) \\
&= \frac{\partial}{\partial x_2}\Big(f_{s_1}(x_2 - x_1) - f_{s_2}(x_3 - x_2 - \phi_{\text{DNA}})\Big) \\
&= k_{s_1}(x_2 - x_1) + k_{s_2}(x_3 - x_2 - \phi_{\text{DNA}}) \\
&= k_{s_1}(x_{s_1}) + k_{s_2}(x_{s_2})
\end{aligned}
$$

$$
\begin{aligned}
h_{12} = \left.\frac{\partial^2 G_{\text{tot}}(\vec{x})}{\partial x_1 \partial x_2}\right|_{x_{\text{tot}},n} &= \frac{\partial}{\partial x_1}\left(\frac{\partial G_{\text{tot}}}{\partial x_2}\right) \\
&= \frac{\partial}{\partial x_1}\Big(f_{s_1}(x_2 - x_1) - f_{s_2}(x_3 - x_2 - \phi_{\text{DNA}})\Big) \\
&= -k_{s_1}(x_2 - x_1) \\
&= -k_{s_1}(x_{s_1})
\end{aligned}
$$

$$
\begin{aligned}
h_{32} = \left.\frac{\partial^2 G_{\text{tot}}(\vec{x})}{\partial x_3 \partial x_2}\right|_{x_{\text{tot}},n} &= \frac{\partial}{\partial x_3}\left(\frac{\partial G_{\text{tot}}}{\partial x_2}\right) \\
&= \frac{\partial}{\partial x_3}\Big(f_{s_1}(x_2 - x_1) - f_{s_2}(x_3 - x_2 - \phi_{\text{DNA}})\Big) \\
&= -k_{s_2}(x_3 - x_2 - \phi_{\text{DNA}}) \\
&= -k_{s_2}(x_{s_2})
\end{aligned}
$$

$$
\begin{aligned}
h_{33} = \left.\frac{\partial^2 G_{\text{tot}}(\vec{x})}{\partial x_3^2}\right|_{x_{\text{tot}},n} &= \frac{\partial}{\partial x_3}\left(\frac{\partial G_{\text{tot}}}{\partial x_3}\right) \\
&= \frac{\partial}{\partial x_3}\Big(f_{s_2}(x_3 - x_2 - \phi_{\text{DNA}}) - f_{h_2}(x_4 - \phi/2 - x_3)\Big) \\
&= k_{s_2}(x_3 - x_2 - \phi_{\text{DNA}}) + k_{h_2}(x_4 - \phi/2 - x_3) \\
&= k_{s_2}(x_{s_2}) + k_{h_2}(x_{h_2})
\end{aligned}
$$

$$
\begin{aligned}
h_{23} = \left.\frac{\partial^2 G_{\text{tot}}(\vec{x})}{\partial x_2 \partial x_3}\right|_{x_{\text{tot}},n} &= \frac{\partial}{\partial x_2}\left(\frac{\partial G_{\text{tot}}}{\partial x_3}\right) \\
&= \frac{\partial}{\partial x_2}\Big(f_{s_2}(x_3 - x_2 - \phi_{\text{DNA}}) - f_{h_2}(x_4 - \phi/2 - x_3)\Big) \\
&= -k_{s_2}(x_3 - x_2 - \phi_{\text{DNA}}) \\
&= -k_{s_2}(x_{s_2})
\end{aligned}
$$

$$
\begin{aligned}
h_{43} = \left.\frac{\partial^2 G_{\rm tot}(\vec{x})}{\partial x_4 \partial x_3}\right|_{x_{\rm tot},n} &= \frac{\partial}{\partial x_4}\left(\frac{\partial G_{\rm tot}}{\partial x_3}\right) \\
&= \frac{\partial}{\partial x_4}\Big(f_{s_2}(x_3 - x_2 - \phi_{\rm DNA}) - f_{h_2}(x_4 - \phi/2 - x_3)\Big) \\
&= -k_{h_2}(x_4 - \phi/2 - x_3) \\
&= -k_{h_2}(x_{h_2}) \\
\\
h_{44} = \left.\frac{\partial^2 G_{\rm tot}(\vec{x})}{\partial x_4^2}\right|_{x_{\rm tot},n} &= \frac{\partial}{\partial x_4}\left(\frac{\partial G_{\rm tot}}{\partial x_4}\right) \\
&= \frac{\partial}{\partial x_4}\Big(f_{h_2}(x_4 - \phi/2 - x_3) - f_b(x_{\rm tot} - x_4)\Big) \\
&= k_{h_2}(x_4 - \phi/2 - x_3) + k_b(x_{\rm tot} - x_4) \\
&= k_{h_2}(x_{h_2}) + k_b(x_b) \\
\\
h_{34} = \left.\frac{\partial^2 G_{\rm tot}(\vec{x})}{\partial x_3 \partial x_4}\right|_{x_{\rm tot},n} &= \frac{\partial}{\partial x_3}\left(\frac{\partial G_{\rm tot}}{\partial x_3}\right) \\
&= \frac{\partial}{\partial x_3}\Big(f_{h_2}(x_4 - \phi/2 - x_3) - f_b(x_{\rm tot} - x_4)\Big) \\
&= -k_{h_2}(x_4 - \phi/2 - x_3) \\
&= -k_{h_2}(x_{h_2})
\end{aligned}
\tag{G.10}
$$

where $k_i(x_i)$ ($i = h_1, s_1, s_2, h_2, b$) are the stiffnesses of the elastic components (handles, ssDNA and optical trap) and the variables x_i ($i = 1, \ldots, 4$) have been written in terms of the extensions again. All the other components of the Hessian matrix ($h_{13}, h_{14}, h_{24}, h_{31}, h_{41}, h_{42}$) vanish because the first derivative does not depend on the the variable that is being derivated in second order. For instance, $h_{14} = 0$ because $\frac{\partial G_{\rm tot}(\vec{x})}{\partial x_4}$ does not depend on x_1. Note that all second derivatives are symmetric ($h_{ij} = h_{ji}$), as expected in a function that has continuous second partial derivatives. Now the Hessian matrix can be written as

$$
\mathbf{H}\left(G_{\rm tot}(\vec{x})\right) = \begin{pmatrix} h_{11} & h_{12} & 0 & 0 \\ h_{21} & h_{22} & h_{23} & 0 \\ 0 & h_{32} & h_{33} & h_{34} \\ 0 & 0 & h_{43} & h_{44} \end{pmatrix} \tag{G.11}
$$

and its determinant is equal to

$$
\begin{aligned}
\det \mathbf{H}\left(G_{\rm tot}(\vec{x})\right) &= h_{11}h_{22}h_{33}h_{44} - h_{11}h_{22}h_{34}h_{43} - h_{11}h_{23}h_{32}h_{44} \\
&\quad - h_{12}h_{21}h_{33}h_{44} + h_{12}h_{21}h_{34}h_{43}
\end{aligned}
\tag{G.12}
$$

which will be necessary to compute the integral. At this point it is important to remember that the components of the Hessian matrix (and so the determinant) depend on x_{tot} and n and the calculation of the stiffnesses (i.e., the second order derivatives) is performed at fixed values of x_{tot} and n.

Now we recall the Taylor expansion written in Eq. G.7 and we introduce it into the calculation of the partition function in Eq. G.6 to obtain the following expression,

$$\begin{aligned} Z(x_{\text{tot}}) &= \sum_{n=0}^{N} \int_{\mathbb{R}^4} e^{-\beta G_{\text{tot}}(x_1, x_2, x_3, x_4, x_{\text{tot}}, n)} \cdot dx_1\, dx_2\, dx_3\, dx_4 \\ &\simeq \sum_{n=0}^{N} \int_{\mathbb{R}^4} e^{-\beta\Big(G_{\min}(x_{\text{tot}}, n) + \frac{1}{2}(\vec{x} - \vec{x}_{\min})^T \cdot \mathbf{H} \cdot (\vec{x} - \vec{x}_{\min})\Big)} \cdot d\vec{x} \end{aligned}$$

where the gradient of $G_{\text{tot}}(\vec{x})$ vanishes because we are expanding around the minimum. Now we can perform the integral of the quadratic contribution according to

$$\begin{aligned} Z(x_{\text{tot}}) &= \sum_{n=0}^{N} e^{-\beta G_{\min}(x_{\text{tot}}, n)} \int_{\mathbb{R}^4} e^{-\frac{\beta}{2}(\vec{x} - \vec{x}_{\min})^T \cdot \mathbf{H} \cdot (\vec{x} - \vec{x}_{\min})} \cdot d\vec{x} \\ &= \sum_{n=0}^{N} e^{-\beta G_{\min}(x_{\text{tot}}, n)} \int_{\mathbb{R}^4} e^{-\frac{\beta}{2}\, \vec{y}^T \mathbf{H} \vec{y}} \cdot d\vec{y} \\ &= \sum_{n=0}^{N} e^{-\beta G_{\min}(x_{\text{tot}}, n)} \sqrt{\left(\frac{2\pi}{\beta}\right)^4 \frac{1}{\det \mathbf{H}(x_{\text{tot}}, n)}} \\ &= \sum_{n=0}^{N} \left(\frac{2\pi}{\beta}\right)^2 \left[\det \mathbf{H}(x_{\text{tot}}, n)\right]^{-1/2} e^{-\beta G_{\min}(x_{\text{tot}}, n)} \end{aligned} \tag{G.13}$$

where we have applied a change of variable $(\vec{x} - \vec{x}_{\min}) \to \vec{y}$ that keeps the same integration limits and allows to perform the Gaussian integration by computing the determinant of the Hessian matrix.

So, in order to calculate the partition function of the whole model we have proceed as follows:

1. Fix a value of x_{tot} at which $Z(x_{\text{tot}})$ will be calculated.

2. For this fixed value of x_{tot}, we have to find the total energy for all the values of n. It requires to solve the transcendental equation G.2 for fixed values of x_{tot} and n, find the extensions of all the elastic elements and introduce them into Eq. G.1.

3. For each value of n (at the same fixed value of x_{tot}), we have to calculate the Hessian matrix according to Eqs. G.10 and compute its determinant, given by Eq. G.12.

4. Once we have all the values of $G_{\mathrm{min}}(x_{\mathrm{tot}}, n)$ and $\det \mathbf{H}(x_{\mathrm{tot}}, n)$ from steps 2 and 3 respectively for all values of n, we have to multiply them and sum over n according to Eq. G.13.

5. Go back to step 1 to calculate $Z(x_{\mathrm{tot}})$ at another value of x_{tot}.

Once the partition function is calculated at all desired values of x_{tot}, the equation of state is obtained after derivating it according to Eq. 3.27,

$$f(x_{\mathrm{tot}}) = -k_B T \frac{\partial \ln Z(x_{\mathrm{tot}})}{\partial x_{\mathrm{tot}}} \tag{G.14}$$

Figure G.2 shows the correction introduced in the calculation of the FDC when considering the elasticity of the elements that constitute the whole model. Note that the differences are less than 0.05 pN. Therefore the effort to compute all the second derivatives of the Hessian matrix does not compensate the improvement in the computation of the FDC, which is lower than the experimental resolution of the experiments. However, this effect is important when the length of the handles is longer and they no longer behave like rigid rods.

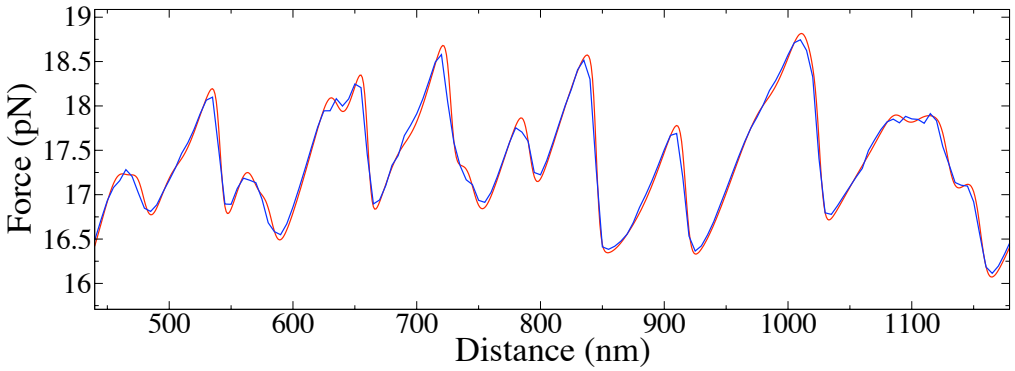


Figure G.2: FDCs for the simplified and full models. Red curve shows the simplified model described in Sec. 3.4.1. Blue curve shows the full model described here in Appendix G.

Appendix H

Numerical calculation of the FDC

The calculation of the FDC has to be performed numerically due to the length of the DNA sequence. The computation of the DNA energy (Eq. 3.17) is a simple sum of terms that extends over all the base pairs of the molecular construct. So this large summation must be done numerically. Apart from that, the exponential terms that enter the partition function have a wide range of orders of magnitude and they must be treated correctly.

The partition function is a function of the distance (i.e., the total extension) x_{tot}, which is a variable that has to be discretized in order to calculate the value of Z at each position. The distance is divided into m equidistant points x_i, $i = 0, \ldots, m$ separated by a distance Δx (see Fig. H.1). A value of $\Delta x = 5$ nm is enough for our calculations. The details of the calculated FDC are missed for higher values of Δx and lower values of it do not improve the calculation.

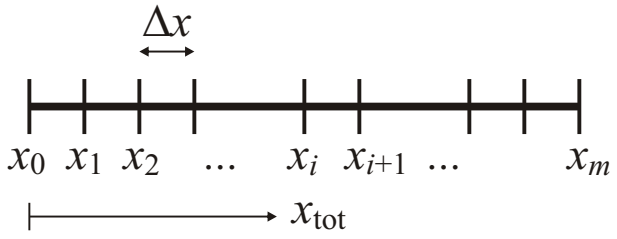


Figure H.1: Discretization of the distance.

Now, for each value of x_i we have to find the minimum of Eq. 3.22 with respect to n. In order to do this, we have to solve Eq. 3.23 for all values of n according to $x_i = x_{\text{tot}}(f, n)$, compute the energies and get the value n_i^* that minimizes the energy at fixed x_i. Equation 3.23 is a trascendental

equation and it can be solved numerically by using the Newton's method. This method gives the equilibrium force f of the system after few interations (no more than 10 to obtain a numerical estimation of f with a relative error smaller than 10^{-5}) of the following expression

$$f_{\text{new}} = f_{\text{old}} - \frac{x_{\text{tot}}(f,n) - x_i}{x'_{\text{tot}}(f,n)} \tag{H.1}$$

where f_{new} and f_{old} are the previous and the posterior values obtained from the iteration, respectively; $x_{\text{tot}}(f,n)$ is Eq. 3.23 evaluated at f and n; and $x'_{\text{tot}}(f,n) = \frac{\partial}{\partial f} x_{\text{tot}}(f,n)$ is its derivative with respect to f. An initial guess of $f = 0$ ensures the convergence of the solution because the first and second derivatives of function x_{tot} have positive values for all values of $f > 0$, which fulfills the Newton's method conditions (see Fig. H.2).

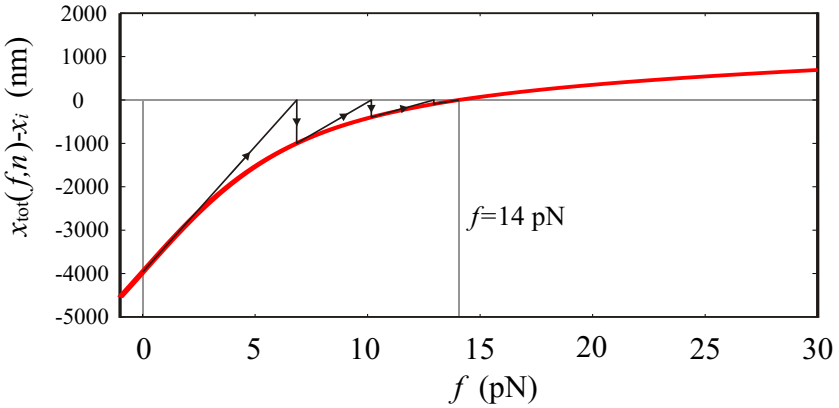


Figure H.2: Newton's method. Red curve shows the approximate elastic response of the system for $n = 4000$ and trap stiffness $k = 80$ pN·μm^{-1} according to $x_{\text{tot}}(f,n) - x_i$. The value of $x_i = 4000$ determines the origin ordinate. The first iteration of the Newton's method starts at $(f, x_{\text{tot}} - x_i)$=$(0, -4000)$ and after some steps it converges to the solution $(f, x_{\text{tot}} - x_i)$=$(14, 0)$. Since the first and second derivatives of $x_{\text{tot}}(f,n) - x_i$ are monotonically decreasing for any value of $x_i > 0$, $f > 0$ and n, the shape of the function is very similar to the one depicted here. Therefore, the evolution of the iterations has a similar pattern and the Newton's method always converges when starting from $f = 0$.

Once we have the combination of values (x_i, n_i^*) we can now calculate the minimum energy $G_{\text{min}}(x_i) = G_{\text{tot}}(x_i, n_i^*)$ according to Eq. 3.22. The

discretized Eq. 3.27 can be written as

$$\begin{aligned} f(x_i) &= -k_B T \frac{\ln Z(x_{i+1}) - \ln Z(x_i)}{x_{i+1} - x_i} \\ &= -k_B T \frac{\ln Z(x_{i+1}) - \ln Z(x_i)}{\Delta x} \\ &= -\frac{k_B T}{\Delta x} \ln \frac{Z(x_{i+1})}{Z(x_i)} \end{aligned} \tag{H.2}$$

where the partition functions have to be calculated according to

$$Z(x_i) = \sum_{n=0}^{N} \exp\left(-\frac{G_{\text{tot}}(x_i, n)}{k_B T}\right) \tag{H.3}$$

The problem with this equation is that the values of $G_{\text{tot}}(x_i, n)$ become larger as the value of x_i increases. When these energies are introduced in the exponential, the value of $Z(x_i)$ can be orders of magnitude smaller than 1. Therefore the quotient $\frac{Z(x_{i+1})}{Z(x_i)}$ of Eq. H.2 is numerically less accurate for large values of x_i. The result is a loss of accuracy in the calculation of the FDC as x_i increases.

This problem can be fixed by taking advantage of the calculated value of $G_{\min}(x_i)$. The idea consists in rewriting the partition function with all the energies referred to this state. So we have,

$$\begin{aligned} Z(x_i) &= \sum_{n=0}^{N} \exp\left(-\frac{G_{\text{tot}}(x_i, n)}{k_B T}\right) \\ &= \sum_{n=0}^{N} \exp\left(-\frac{G_{\text{tot}}(x_i, n) - G_{\min}(x_i) + G_{\min}(x_i)}{k_B T}\right) \\ &= \exp\left(-\frac{G_{\min}(x_i)}{k_B T}\right) \sum_{n=0}^{N} \exp\left(-\frac{G_{\text{tot}}(x_i, n) - G_{\min}(x_i)}{k_B T}\right) \\ Z(x_i) &= e^{-\beta G_{\min}(x_i)} \cdot \tilde{Z}(x_i) \end{aligned} \tag{H.4}$$

where

$$\tilde{Z}(x_i) = \sum_{n=0}^{N} \exp\left[-\beta\big(G_{\text{tot}}(x_i, n) - G_{\min}(x_i)\big)\right] \tag{H.5}$$

is the partition function calculated with an energy offset given by the state of minimum energy at fixed x_i. Equation H.4 can be introduced into Eq. H.2 to obtain

$$f(x_i) = -\frac{k_B T}{\Delta x} \ln \frac{Z(x_{i+1})}{Z(x_i)}$$

$$
\begin{aligned}
&= -\frac{k_B T}{\Delta x} \ln \frac{e^{-\beta G_{\min}(x_{i+1})} \tilde{Z}(x_{i+1})}{e^{-\beta G_{\min}(x_i)} \tilde{Z}(x_i)} \\
&= -\frac{k_B T}{\Delta x} \left[\ln \frac{e^{-\beta G_{\min}(x_{i+1})}}{e^{-\beta G_{\min}(x_i)}} + \ln \frac{\tilde{Z}(x_{i+1})}{\tilde{Z}(x_i)} \right] \\
f(x_i) &= -\frac{k_B T}{\Delta x} \left[\beta \Big(G_{\min}(x_i) - G_{\min}(x_{i+1}) \Big) + \ln \frac{\tilde{Z}(x_{i+1})}{\tilde{Z}(x_i)} \right] \qquad \text{(H.6)}
\end{aligned}
$$

where $G_{\min}(x_i)$, $G_{\min}(x_{i+1})$, $\tilde{Z}(x_{i+1})$ and $\tilde{Z}(x_i)$ can be easily calculated numerically.

To sum up, the numerical calculation of the FDC requires the following steps,

1. Discretize the distance with equidistant values x_i.
2. For each value of x_i, calculate the energies $G_{\text{tot}}(x_i, n)$ for all the values of n by solving Eq. H.1.
3. Among the previously calculated energies, find the value of n_i^* that minimizes the energy for each value of x_i. This energy is called $G_{\min}(x_i)$.
4. For each value of x_i, calculate $\tilde{Z}(x_i)$ according to Eq. H.5
5. For each value of x_i, calculate $f(x_i)$ according to Eq. H.6.

Appendix I

Synthesis of ssDNA

A 3 kb ssDNA molecular construct was obtained by pH denaturation (strand separation) of a 3 kb dsDNA (see Fig. I.1). The dsDNA was obtained from PCR amplification of a $\sim$3 kbp fragment of λ-DNA. One of the primers used in the process was already labeled with Biotin. The resulting product was cleaved with the endonuclease XbaI producing a cohesive end. Another 24 base oligonucleotide (previously labeled with several digoxigenins at its $3'$ end by using terminal transferase) was hybridized with a second 20-base long oligonucleotide giving a DNA construction with one cohesive end complementary to XbaI. Both products were annealed and ligated resulting in one 3 kbp dsDNA molecule.

To produce ssDNA, the molecular construct was incubated with Strepavidin coated beads for 30 min at room temperature in a volume of 15 μl of 10 mM TE buffer. Afterwards, 35 μl of 0.1 M [NaOH] were added in order to cause the separation (i.e., denaturation) of the strands. After 30 min, the sample was centrifuged. The white precipitate of beads and ssDNA was re-suspended in TE buffer. The second attachment with the antidigoxigenin beads was achieved in the fluidics chamber with the help of the micropipette (see Sec. 3.2.2).

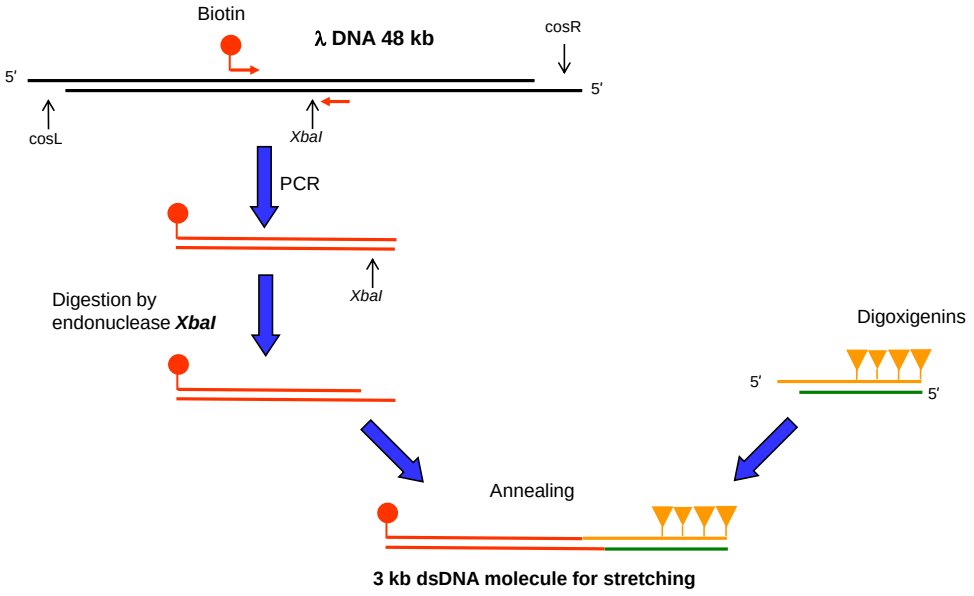


Figure I.1: Synthesis of the 3 kb dsDNA. The resulting dsDNA is obtained after pH denaturation. The ssDNA molecule can be stretched between two coated beads, since the Biotin and Digoxigenins labels are located on the same strand.

Appendix J

Shift function

The experimental FDC measurements are force vs. trap position as measured by the light-levers. The unzipping/zipping curves contain reproducible and recognizable landmarks (i.e., slopes and rips) which indicate the true position of the trap. Therefore it is possible to to take advantage of these landmarks to correct for the instrumental drift. Correction for drift is introduced in terms of a shift function $s(x_{\rm tot})$ which is built in several steps. Due to its relevance for data analysis we describe the steps in some detail:

- **Step 1**. Since the distance $x_{\rm tot}$ of the FDC is a relative (not absolute) magnitude, we firstly determine the origin of coordinates of the experimental FDC (see blue arrow in Fig. J.1a). In order to do so, we fit the last part of the experimental FDC (black and yellow curves in Fig. J.1a) that correspond to the stretching of the ssDNA when the hairpin is fully unzipped.

- **Step 2**. Having fixed the origin of coordinates we calculate the predicted FDC by using the Unified Oligonucleotide (UO) NNBP energies. It is shown in red (Fig. J.1a,b). The qualitative behavior is acceptable (all force rips are reproduced). However, the predicted mean unzipping force is higher than the value found experimentally and the force rips are not located at the correct position.

- **Step 3**. Next we generate a FDC with NNBP energies lower than the UO values until the mean unzipping forces of the predicted and the experimental FDC coincide. What we typically do is multiplying all the 10 UO NNBP energies by a factor (e.g., 0.95). The new NNBP energies have an absolute value 8-10% lower than the UO NNBP energies. The resulting FDC with these new energies is shown in green in Fig. J.1b (in this particular case we took $\epsilon_i^{\rm New} = 0.92 \cdot \epsilon_i^{\rm UO}$). Although the mean

unzipping force of the green and black curves is nearly the same, there is a misalignment between the rips along the distance axis. Moreover, there are discrepancies between the predicted and measured heights of the force rips. As we will see below, the shift function will correct the horizontal misalignments and the NNBP energies will correct the discrepancies along the force axis.

- **Step 4**. We now introduce a shift function that uses the slopes of the sawtooth pattern of the FDC as *landmark points* to locally correct the distance to align the experimental data with the theoretical prediction. First, we want to know the approximate shape of the shift function and latter we will refine it. This is done by looking for some characteristic slopes of the sawtooth pattern along the FDC and measuring the local shift that would make the two slopes (theoretical and experimental) superimpose. Figure J.1c shows zoomed regions of the FDC and the blue arrows indicate the local shift that should be introduced in each slope to correct the FDC. The orange dots shown in the upper panel of Fig. J.1d depict the local shifts vs. the relative distance that have been obtained for the landmark points. These orange dots represent a discrete sampled version of an ideal shift function that would superimpose the predicted and the experimental FDC. Because these dots are not equidistant, we use cubic splines to interpolate a continuous curve every three landmark points. The resulting interpolated function that describes the local shift for any relative distance is shown in violet. Note that the violet curve passes through all the orange dots.

- **Step 5**. Starting from the cubic splines interpolation of the shift function that we have found (violet curve in upper panel Fig. J.1d) we can define new equidistant points (yellow dots in lower panel Fig. J.1)d) that define the same shift function. The yellow equidistant points are separated 100 nm. We call these yellow points *Control points*.

- **Step 6**. We now introduce the shift function into the calculation of the theoretical FDC. The results are shown in Fig. J.1e. Again, the black curve is the experimental FDC, the green curve is the predicted FDC without the local shift correction and the magenta curve is the predicted FDC with the local shift function obtained previously. Note that the magenta and green curves are identical, except for local contractions and dilatations of the magenta curve. The slopes of the magenta and the black curves now coincide. Still, the NNBP energies must be fit to make the height of the rips between the theoretical and experimental curve coincident.

- **Step 7**. At this point we start the Monte Carlo fitting algorithm. At each Monte Carlo step we propose new values for the 10 NNBP energies and we also adjust the shift function in order to superimpose the theoretical and experimentally measured FDCs. The shift function is adjusted by modifying the values of the control points (yellow dots in lower panel of Fig. J.1)d). The horizontal position of the yellow points is always the same (e.g., the yellow dots located at Relative Distance = -1000 nm will always be located there). What we change when we adjust the shift function is the value of each point (e.g., the yellow dot located at Relative Distance = -1000 nm may change its shift value from -54 to -20 nm). During the Monte Carlo optimizing procedure the shift function modifies its shape as the NNBP energies are modified. The results are shown in Fig. J.1f. Black curves show snapshots of the evolution of the shift function during the optimization process from the initial shape (red curve). The green curve shows the final shift function (the control points are not depicted in Fig. J.1f, only the interpolated shift function). When we finally calculate the theoretical FDC using the optimal shift function and the optimal NNBP energies we get the maximum overlap between the theoretical prediction and the experimental FDC. Black curve in upper panel in Fig. J.1g is the experimental FDC and red curve is the predicted FDC after having fit the NNBP energies and the shift function. The correction for drift has now finished. The optimal shift function is also shown in the right panel in Fig. J.1g.

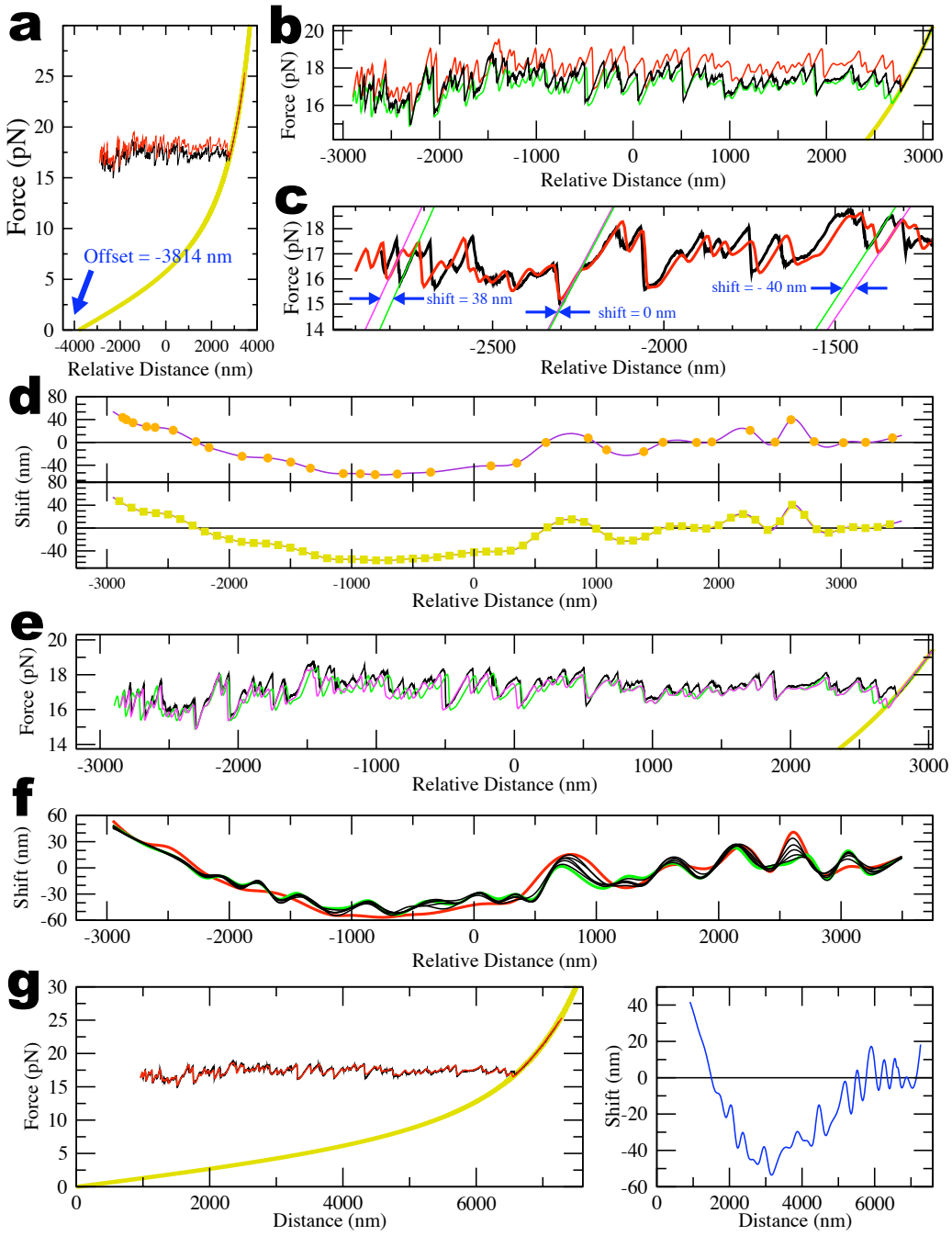


Figure J.1: Fit of the shift function. **(a)** Steps 1 and 2. **(b)** Steps 2 and 3. **(c)** Step 4. **(d)** Steps 4 and 5. **(e)** Step 6. **(f)** Step 7. **(g)** Step 7.

Appendix K

Sampling of energy states distribution

By definition, the magnitudes required to calculate a mean squarred error (between a theoretical prediction and an experimental measurement) must be the same. Therefore, we must be sure that the experimental FDC (f_i^{exp}) that enters Eq. 5.2 is the same magnitude than the theoretical FDC (f_i^{the}). According to our calculations, the theoretical FDC is in equilibrium. How do we know that the experimental (i.e., measured) FDC is an estimation of the equilibrium FDC too? This appendix describes how the unzipping experiments at low pulling rate are capable of sampling the entire distribution (or at least the most significant part of it) of energy states of the DNA molecule, which is a prerequisite for calculating ensemble averages, partition functions and thermodynamic parameters such as the equilibrium FDC.

Let us discuss the states of DNA that have higher energy than the minimum. In unzipping experiments, the position of the unzipping fork (see Fig. K.1a) exhibits thermally induced fluctuations in such a way that the system can explore higher free energy states. Such fluctuations represent the first kind of excitations in the system and will be discussed in the next paragraphs. However there is a second kind of excitation: breathing fluctuations. The breathing is the spontaneous opening and closing of base pairs produced in the dsDNA, far away from the unzipping fork (see Fig. K.1b). During this process, the DNA explores states of higher free energy while the unzipping fork is kept at the same position. So the breathing does not induce any change in the position of the unzipping fork. Consequently, we are not able to distinguish the breathing in our unzipping experiments because breathing fluctuations are not coupled to the reaction coordinate that we measure, i.e., the molecular extension, and should have a small effect on the measured FDC. Note that breathing fluctuations are expected to be relevant

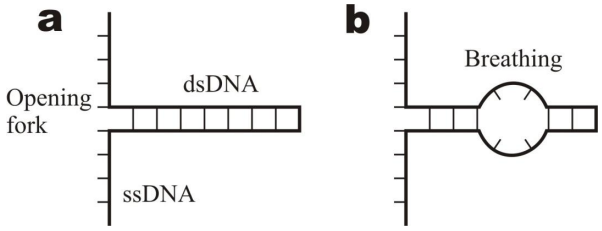


Figure K.1: **(a)** Opening fork. **(b)** Breathing.

only at high enough temperatures. While the inclusion of breathing fluctuations should be considered at high enough temperatures their contribution at 25°C is expected to be minimal. The fact that our model reproduces very well the experimental FDC supports this conclusion.

Now let us focus on the fluctuations of the unzipping fork. The unzipping of DNA is performed so slow that the system has enough time to reach the equilibrium at every fixed distance along the pulling protocol. Figure K.2a shows a fragment of the FDC in a region where 3 states (each one with a different number of open base pairs) of the 2.2 kbp molecule coexist ($n_1 = 1193$, $n_2 = 1248$ and $n_3 = 1300$). Figure K.2b shows the hopping in force due to the transitions that occur between these 3 states. The slow pulling rate guarantees that the hopping transitions are measured during unzipping (i.e., many hopping events take place while the molecule is slowly unzipped). The filtering of the raw FDC data at 1 Hz (black curve in Fig. K.2a) produces a reasonably good estimation of the equilibrium FDC. It is also important to remark that the unzipping and rezipping curves are reversible (see Fig. 3.11c). This supports the idea that the unzipping process is quasistatic and correctly samples the energy states.

In general, the hopping frequency between coexistent states is around $\sim 10 - 50$ Hz and the area of coexistence extends over 40 nm of distance (see Fig. K.2b). At a pulling rate of 10 nm/s, we can measure around $10 - 40$ transitions, which in most cases is sufficient to obtain a good estimation of the FDC after averaging out the raw data at a bandwidth of 1 Hz.

Figure K.3a shows the free energy landscape of one molecule at different fixed distances, which is given by $G(x_{\rm tot}, n)$ in Eq. 3.22. A detailed view of the free energy landscape (see Fig. K.3b) shows that it is a rough function and its coarse grained shape is parabolic. It means that for each value of $x_{\rm tot}$ there is always a state of minimum global energy surrounded by other states of higher free energies (see Fig. K.3b). Although there are lots of states in the phase space, in the experiments we only observe those states that differ in free energy by less than $\sim 5\ k_BT$ with respect to the state of minimum

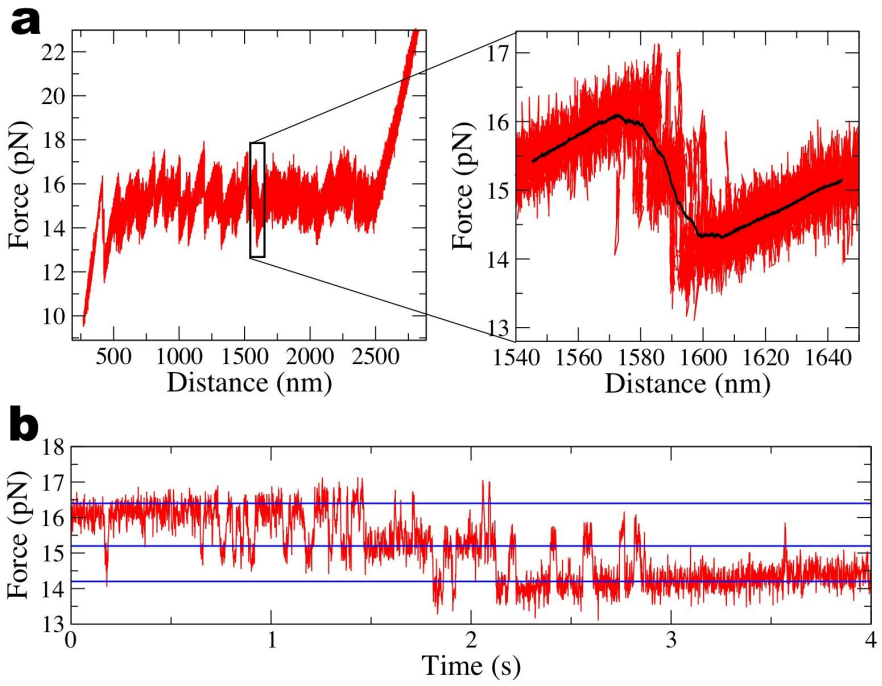


Figure K.2: Coexistence of states. **(a)** Left panel shows the measured FDC for the 2.2 kbp sequence. Right panel shows the fragment of the FDC (framed in the left panel) where 3 states coexist. Red curve shows the raw data and black curve shows the data filtered at 1 Hz. **(b)** Red curve shows the force vs. time of the previous fragment where the transitions between these 3 states can be observed. The blue lines indicate the average forces corresponding to each of these 3 states.

free energy. So the hopping transitions described in Fig. K.2b are between states that have similar free energies. Outside this range of free energies, the higher energetic states are rarely observed and their contribution to the equilibrium FDC is negligible.

Summing up, the temperature of the experiment (so low that the breathing is neglegible), the slow pulling rate and the shape of the free energy landscape ensure us that we explore higher energetic states (within a range of $\sim 5\ k_BT$ with respect to the global minimum) during the unzipping process. Therefore, the experimental FDC filtered at 1 Hz is a good estimation of the equilibrium FDC.

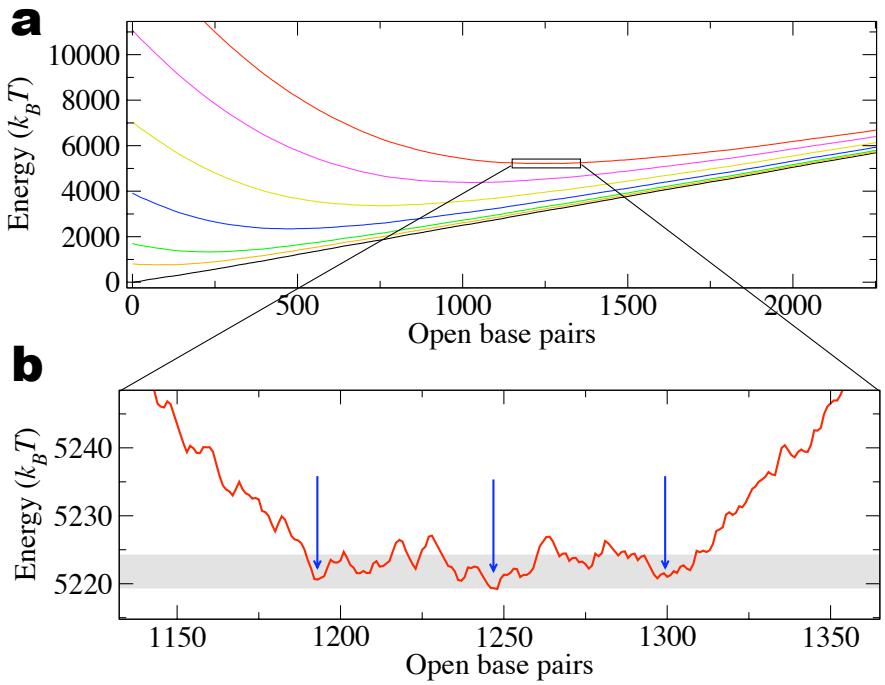


Figure K.3: Free energy landscape for the 2.2 kbp sequence at fixed distance. **(a)** The parabolic-like shape of the free energy landscape around the minima can be identified in a coarse grained view of what in truth is a rough landscape (see zoomed part of the landscape). Black, orange, green, blue, yellow, magenta and red curves show the free energy landscape at $x_{\text{tot}} = 0$, 350, 500, 750, 1000, 1250 and 1455 nm, respectively. **(b)** Zoomed region of the free energy landscape at the distance in which the 3 states of Fig. K.2 coexist. The blue arrows indicate the minima that correspond to these states. The highlighted gray area shows an energy range of 5 k_BT.

Appendix L

Error estimation in the Monte Carlo optimization

This aim of this appendix is to clarify which are the sources of errors in the estimates of the NNBP values when applying the MC optimization. There are three kinds of errors at different levels that we will denote as σ_1, σ_2 and σ_3:

1. The first error σ_1 comes from the fitting algorithm. The uncertainties of the estimated NNBP energies (σ_{ϵ_i}) indicate how much the error function ($E(\epsilon_1, \ldots, \epsilon_{10}, \epsilon_{\text{loop}})$ see Eq. 5.2) changes when the fitting parameters ϵ_i are varied around the minimum. For instance, a variation of the AA/TT motif ($\delta\epsilon_1$) around the minimum (see Fig. L.1) produces a larger change in the error function than a variation of the TA/AT motif ($\delta\epsilon_{10}$). This indicates that the uncertainty of AA/TT is lower than that of TA/AT. The curvature of the minimum in each direction ϵ_i gives the uncertainty. There is a different set of σ_{ϵ_i} uncertainties for each fit (i.e., each molecule). A quantitative evaluation of the uncertainty of the NNBP parameters requires the evaluation of the χ^2 function for each FDC (i.e., each fit), which is given by:

$$\chi^2(\vec{\epsilon}) = \sum_{i=1}^{N} \left(\frac{f_i - f(x_i; \vec{\epsilon})}{\sigma_y} \right)^2 \tag{L.1}$$

   where N is the number of experimental points of the FDC; x_i and f_i are the position and the force measurements, respectively; $\vec{\epsilon}$ is the vector of fitting parameters $\{\epsilon_i\}$, $i = 1, \ldots, 10$,loop; $f(x_i; \vec{\epsilon})$ is the theoretically predicted FDC according to the model (see Sec. 3.4.1); and σ_y is the experimental error of the force measurements performed with the optical

tweezers. The resolution of the instrument is taken as $\sigma_y = 0.1$ pN. The uncertainty of the fit parameters is given by the following expression [162]:

$$\sigma_{\epsilon_i} = \sqrt{C_{ii}} \tag{L.2}$$

where C_{ii} are the diagonal elements of the variance-covariance matrix C_{ij}. In a non-linear least square fit, this matrix can be obtained from $C_{ij} = 2 \cdot H_{ij}^{-1}$, where H_{ij}^{-1} is the inverse of the Hessian matrix

$$H_{ij} = \frac{\partial^2 \chi^2(\vec{\epsilon}_m)}{\partial\epsilon_i \, \partial\epsilon_j} \tag{L.3}$$

of $\chi^2(\vec{\epsilon})$ evaluated at the point $\vec{\epsilon}_m$ that minimizes the error. Note that the error function and the χ^2 function are related by a constant factor, $\chi^2(\vec{\epsilon}) = (N/\sigma_y^2) \cdot E(\vec{\epsilon})$, so their Hessians are related by one constant factor, as well. The calculation of σ_{ϵ_i} is quite straightforward and it gives values between $0.003 - 0.015$ kcal·mol^{-1}. These values represent the first type of error that we call σ_1. Note that the Hessian matrix evaluated at the minima found with the heat-quench algorithm is very similar to the Hessian matrix evaluated at the minimum, which means that the curvature is almost the same in all heat-quench minima. Therefore the error of the fit (σ_1) takes the same value within a region of ± 0.1 kcal·mol^{-1}.

2. The second error comes from the dispersion of the heat-quench minima. As we saw previously, there are several minima corresponding to different possible solutions (each solution being a set of 10 NNBP energies) for the same molecule. The values of the NNBP energies corresponding to the different solutions are Gaussian distributed (see Fig. 5.8c) and the average standard deviation is about 0.05 kcal·mol^{-1}. All these considerations result in a second typical error $\sigma_2 = 0.05$ kcal·mol^{-1}.

3. Finally, the third error corresponds to the molecular heterogeneity intrinsic to single molecule experiments. Such heterogeneity results in a variability of solutions among different molecules. Indeed, the FDCs of the molecules are never identical and this variability leads to differences in the values of the NNBP energies. This variability is the major source of error in the estimation of our results. The error bars in Figs. 5.12c,d and 5.13 indicate the standard error of the mean, which is around 0.1 kcal·mol^{-1} on average. This is what finally determines the statistical error of our analysis, $\sigma_3 = 0.1$ kcal·mol^{-1}.

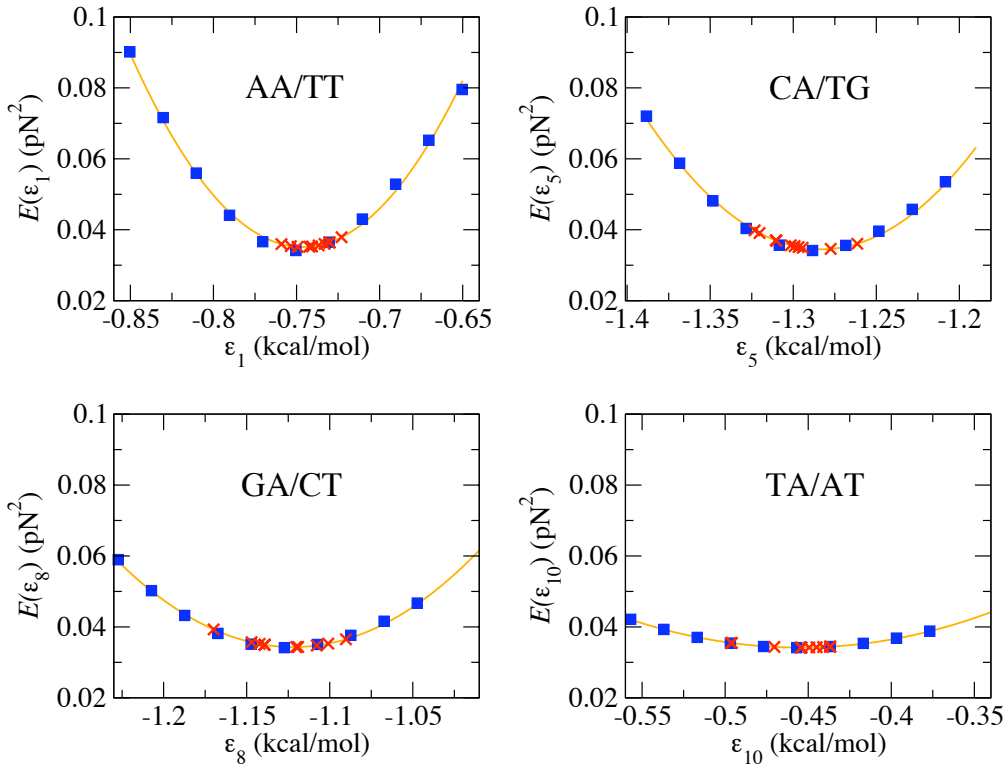


Figure L.1: Error function (see Eq. 5.2) around the minimum for small variations of some NNBP energies. Blue dots show the error function evaluated at different values of ϵ_i. Orange curves show the quadratic fits around the minimum according to $E(\epsilon) = c/2 \cdot (\epsilon - \epsilon_0)^2 + E_0$, where c, ϵ_0 and E_0 are fitting parameters. Red crosses show the solutions found with the MC algorithm, which differ by less than 0.05 kcal·mol^{-1} with respect to the minimum. Note that the values of the curvatures (c) of the error function are different for each NNBP parameter. The curvature allows us to estimate the error of the fitting parameters. We have checked that the curvature of the quadratic fit (i.e., the value of the parameter c) for each NNBP parameter coincides with the diagonal elements of the Hessian matrix (Eq. L.3), which give the curvature of the error function in the 11-dimensional space.

Since the major source of errors is the variability of the results from molecule to molecule, we simply report this last error in the manuscript. Because $\sigma_3 > \sigma_2 > \sigma_1$ we can safely conclude that the propagation of the errors of the heat-quench algorithm will not increase the final value of the error bar.

Appendix M

Error estimation in the enthalpy and entropy inference

Here we provide and estimation of the error ($\sigma_{\Delta h_i}$) of the 10 parameters Δh_i^0 ($i = 1, \ldots, 10$) involved in the fit of Eq. 5.4. We simplify the notation by writing the Δh_i^0 ($i = 1, \ldots, 10$) values given in Table 5.3 in vectorial form according to $\Delta\vec{h}_m$, where the subscript m stands for minimum. Note that $\Delta\vec{h}_m$ minimizes the $\chi^2(\Delta\vec{h})$ error function (Eq. 5.4). By definition, the first derivatives of $\chi^2(\Delta\vec{h})$ with respect to $\Delta\vec{h}$ vanish at the minimum ($\vec{\nabla} \cdot \chi^2(\Delta\vec{h}_m) = 0$). So we can write a Taylor expansion of $\chi^2(\Delta\vec{h})$ around $\Delta\vec{h}_m$ up to second order:

$$\chi^2(\Delta\vec{h}_m + \delta\Delta\vec{h}) \simeq \chi^2(\Delta\vec{h}_m) + \frac{1}{2}\delta\Delta\vec{h}^T \cdot \mathbf{H}\left(\chi^2(\Delta\vec{h}_m)\right) \cdot \delta\Delta\vec{h} \qquad \text{(M.1)}$$

where $\delta\Delta\vec{h}$ is a variation of the $\Delta\vec{h}_m$ vector and $\mathbf{H}\left(\chi^2(\Delta\vec{h}_m)\right)$ is the Hessian matrix of second derivaties

$$H_{ij} = \frac{\partial^2 \chi^2(\Delta h_1^0, \ldots, \Delta h_{10}^0)}{\partial \Delta h_i^0 \, \partial \Delta h_j^0} \qquad \text{(M.2)}$$

evaluated at the minimum $\Delta\vec{h}_m$. After fitting Eq. 5.4, the estimation of $\chi^2(\Delta\vec{h}_m) = 1.74$ (in units of squared Celsius degrees $^\circ\text{C}^2$) is lower than the typical experimental error in melting experiments, which is 2°C (i.e., $\chi^2 = 4$). So there is a range of $\Delta\vec{h}$ values around the minimum $\Delta\vec{h}_m$ that still predict the melting energies within an average error of 2°C. This range of values is what determines the error in the estimation of $\Delta\vec{h}_m$.

Following this criterion, we look for the variations around the minimum ($\delta\Delta\vec{h}$) that produce a quadratic error of 4°C^2. We divide this quadratic

error into the 10 fitting parameters (Δh_i^0 $i = 1, \ldots, 10$) and the 10 related ones (Δs_i^0 $i = 1, \ldots, 10$). So we look for each $\delta\Delta h_i$ that induces an error of $4°C^2/20 = 0.2°C^2$. Now, introducing $\chi^2(\vec{\Delta h}_m + \delta\vec{\Delta h}) = 4$ and $\chi^2(\vec{\Delta h}_m) = 1.74$ (divided by 20 parameters) into Eq. M.1 and isolating $\delta\vec{\Delta h}$, we get one expression to estimate the errors of the 10 fitting parameters:

$$\delta\Delta h_i = \sigma_{\Delta h_i} = \sqrt{\frac{2 \cdot (4 - 1.74)/20}{H_{ii}\left(\chi^2(\vec{\Delta h}_m)\right)}} = \sqrt{\frac{0.226}{H_{ii}\left(\chi^2(\vec{\Delta h}_m)\right)}}, \qquad i = 1, \ldots, 10 \tag{M.3}$$

which gives values between $0.3 - 0.6$ kcal·mol^{-1} (see Table 5.3). Now, the error in the estimation of the entropies ($\sigma_{\Delta s_i}$) can be obtained from error propagation of Eq. 5.5:

$$\sigma_{\Delta s_i} = \frac{1}{T}\left(\sigma_{\Delta h_i} + \sigma_{\Delta \epsilon_i}\right) \qquad i = 1, \ldots, 10 \tag{M.4}$$

where $T = 298.15$ K is the temperature and $\sigma_{\Delta\epsilon_i}$ are the experimental errors of our estimated NNBP energies from the unzipping measurements (Table 5.2). The errors range between $1.2 - 2.2$ cal/mol·K (Table 5.3).

Bibliography

[1] Watson J.D. and Crick F.H.C. A structure for deoxyribose nucleic acid. *Nature*, 171(4356):737–738, 1953.

[2] Ritort F. Single-molecule experiments in biological physics: methods and applications. *J. Phys. (Cond. Matt.)*, 18:R531–R583, 2006.

[3] Mossa A., Huguet J.M., and Ritort F. Investigating the thermodynamics of small biosystems with optical tweezers. *Physica E: Low-dimensional Systems and Nanostructures*, 42:666–671, 2010.

[4] Crooks G.E. Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences. *Phys. Rev. E*, 60:2721–2726, 1999.

[5] Jarzynski C. Nonequilibrium equality for free energy differences. *Phys. Rev. Lett.*, 78:2690, 1997.

[6] Maxwell J.C. *A Treatise on Electricity and Magnetism, Vol. II.* Clarendon Press, Oxford, 1873.

[7] Ashkin A. Acceleration and trapping of particles by radiation pressure. *Phys. Rev. Lett.*, 24(4):156–159, 1970.

[8] Ashkin A., Dziedzic J.M., Bjorkholm J.E., and Chu S. Observation of single-beam gradient force optical trap for dielectric particles. *Opt. Lett.*, 11(5):288, 290 1986.

[9] Svoboda K., Schmidt C.F., Schnapp B.J., and Block S.M. Direct observation of kinesin stepping by optical trapping interferometry. *Nature*, 365:721–727, 1993.

[10] Kuo S.C. and Sheetz M.P. Force of single kinesin molecules measured with optical tweezers. *Science*, 260:232–234, 1993.

[11] Ghislain L.P., Switz N.A., and Webb W.W. Measurement of small forces using an optical trap. *Rev. Sci. Instrum.*, 65(9):2762–2768, 1994.

[12] Moffitt J.R., Chemla Y.R., Smith S.B., and Bustamante C. Recent advances in optical tweezers. *Annu. Rev. Biochem.*, 77:205–228, 2008.

[13] Moffitt J.R., Chemla Y.R., Izhaky D., and Bustamante C. Differential detection of dual traps improves the spatial resolution of optical tweezers. *Proc. Natl. Acad. Sci. USA*, 103:9006–11, 2006.

[14] Grier D.G. A revolution in optical manipulation. *Nature*, 424:810–16, 2003.

[15] Gouesbet G., Grehan G., and Maheu B. Scattering of a Gaussian beam by a Mie scatter center using a Bromwich formalism. *J. Optics (Paris)*, 16(2):83–93, 1985.

[16] Smith S.B., Cui Y., and Bustamante C. Optical-trap force transducer that operates by direct measurement of light momentum. *Methods Enzymol.*, 361:134–160, 2003.

[17] Smith S.B. and Bustamante C. Light-force sensor and method for measuring axial optical-trap forces from changes in light momentum along an optical axis. *U.S. Patent*, 7,133,132 B2, 2006.

[18] Smith S.B. and Bustamante C. Optical beam translation device and method using a pivoting optical fiber. *U.S. Patent*, 7,274,451 B2, 2006.

[19] Smith S.B. and Rivetti C. http://tweezerslab.unipr.it. 2010.

[20] Crick F.H.C. The biological replication of macromolecules. *Symp. Soc. Exp. Biol.*, XII:138–163, 1958.

[21] Bockelmann U., Essevaz-Roulet B., and Heslot F. Molecular Stick-Slip Motion Revealed by Opening DNA with Piconewton Forces. *Phys. Rev. Lett.*, 79(22):4489–4492, 1997.

[22] Rief M., Clausen-Schaumann H., and Gaub H.E. Sequence-dependent mechanics of single DNA molecules. *Nat. Struct. Biol.*, 6(4):346–349, 1999.

[23] Danilowicz C., Coljee V.W., Bouzigues C., Lubensky D.K., Nelson D.R., and Prentiss M. DNA unzipped under a constant force exhibits multiple metastable intermediates. *Proc. Natl. Acad. Sci. USA*, 100(4):1694–1699, 2003.

[24] Bockelmann U., Ph. Thomen, Essevaz-Roulet B., Viasnoff V., and Heslot F. Unzipping DNA with optical tweezers: high sequence sensitivity and force flips. *Biophys. J.*, 82(3):1537–1553, 2002.

[25] Huguet J.M., Bizarro C.V., Forns N., Smith S.B., Bustamante C., and Ritort F. Single-molecule derivation of salt dependent base-pair free energies in DNA. *Proc. Natl. Acad. Sci. USA*, 107:15431–15436, 2010.

[26] Devoe H. and Tinoco Jr. I. The stability of helical polynucleotides: base contributions. *J. Mol. Biol.*, 4:500–517, 1962.

[27] Crothers D.M. and Zimm B.H. Theory of the melting transition of synthetic polynucleotides: evaluation of the stacking free energy. *J. Mol. Biol.*, 9:1–9, 1964.

[28] SantaLucia Jr. J. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl. Acad. Sci. USA*, 95(4):1460–1465, 1998.

[29] Lubensky D.K. and Nelson D.R. Single molecule statistics and the polynucleotide unzipping transition. *Phys. Rev. E*, 65:031917, 2002.

[30] Sanger F. and Coulson A.R. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J. Mol. Biol.*, 94:441–446, 1975.

[31] Bischof M. Some remarks on the history of biophysics and its future. *Proceedings of the 1st Hombroich Symposium of Biophysics, Oct.3-6, 1995, Neuss, Germany*, pages 10–21, 1995.

[32] Schrödinger E. *What is life?* Cambridge University Press., 1992.

[33] Strick T., Allemand J.-F., Croquette V., and Bensimon D. The manipulation of single biomolecules. *Phys. Today*, 54:46–51, 2001.

[34] Biophysical Society. http://www.biophysics.org/.

[35] Daune M. *Molecular Biophysics.* Oxford University Press, Oxford, 1999.

[36] Eom S.H. and Wang J. Steitz T.A. Structure of Taq polymerase with DNA at the polymerase active site. *Nature*, 382:278–81, 1996.

[37] Several authors. Edited by French A.P. *Physics in a Technological World.* AIP, New York, 1988.

[38] Kratky O. and Porod G. Röntgenuntersuchung gelöster Fadenmoleküle. *Rec. Trav. Chim. Pays-Bas*, 68:1106–1123, 1949.

[39] Smith S.B., Finzi L., and Bustamante C. Direct mechanical measurements of the elasticity of single DNA molecules by using magnetic beads. *Science*, 258(5085):1122–1126, 1992.

[40] Bustamante C., Marko J.F., Siggia E.D., and Smith S.B. Entropic elasticity of lambda-phage DNA. *Science*, 265(5178):1599–1600, 1994.

[41] Glaser R. *Biophysics.* Springer-Verlag, Berlin, 5th ed., 2001.

[42] Bai C., Wang C. Xie X.S., and Wolynes P.G. Single molecule phyisics and chemistry. *Proc. Natl. Acad. Sci. USA*, 96:11075, 1999.

[43] Mehta A.D., Rief M., Spudich J.A., Smith D.A., and Simmons R.M. Single-Molecule Biomechanics with Optical Methods. *Science*, 283:1689–1695, 1999.

[44] Block S.M., Goldstein L.S.B., and Schnapp B.J. Bead movement by single kinesin molecules studied with optical tweezers. *Nature*, 348(348-352), 1990.

[45] Finer J.T., Simmons R.M., and Spudich J.A. Single myosin molecule mechanics: piconewton forces and nanometre steps. *Nature*, 368:113–119, 1994.

[46] Noji H., Yasuda R., Yoshida M., and Kinosita Jr K. Direct observation of the rotation of F_1-ATPase. *Nature*, 386:299–302, 1997.

[47] Yin H., Wang M.D., Svoboda K., Landick R., Block S.M., and Gelles J. Transcription against an applied force. *Science*, 270:1653 – 1657, 1995.

[48] Smith S.B., Cui Y., and Bustamante C. Overstretching B-DNA: the elastic response of individual double-stranded and single-stranded DNA molecules. *Science*, 271:795–799, 1996.

[49] Liphardt J., Onoa B., Smith S.B., Tinoco Jr. I., and Bustamante C. Reversible unfolding of single rna molecules by mechanical force. *Science*, 292:733–737, 2001.

[50] Carrion-Vazquez M., Oberhauser A.F., Fowler S.B., Marszalek P.E., Broedel S.E., Clarke J., and Fernandez J.M. Mechanical and chemical unfolding of a single protein: a comparison. *Proc. Natl. Acad. Sci. USA*, 96:3694–9, 1999.

[51] Dame R.T., Noom M.C., and Wuite G.J.L. Bacterial chromatin organization by H-NS protein. *Nature*, 444:387–390, 2006.

[52] Pathria R.K. *Statistical Mechanics*. Butterworth-Heinemann, Oxford, 1996.

[53] Liphardt J., Dumont S., Smith S.B., Tinoco Jr I., and Bustamante C. Equilibrium information from nonequilibrium measurements in an experimental test of Jarzynski's equality. *Science*, 296:1832–5, 2002.

[54] Collin D., Ritort F., Jarzynsk C., Smith S.B., Tinoco Jr. I., and Bustamante C. Verification of the Crooks fluctuation theorem and recovery of RNA folding free energies. *Nature*, 437:231–234, 2005.

[55] Junier I., Mossa A., and Manosas M. Ritort F. Recovery of free energy branches in single molecule experiments. *Phys. Rev. Let.*, 102:070602, 2009.

[56] Ritort F. Nonequilibrium fluctuations in small systems: from physics to biology. *Advances in Chemical Physics (Ed. Stuart. A. Rice, Wiley publications)*, 137:31–123, 2007.

[57] Mie G. Beiträge zur Optik trüber Medien, speziell kolloidaler Metallösungen. *Ann. der Phys.*, 25:377–445, 1908.

[58] Debye P. Der Lichtdruck auf Kugeln von beliebigem Material. *Ann. der Phys.*, 335:57–136, 1909.

[59] Ashkin A. Optical levitation by radiation pressure. *Appl. Phys. Lett.*, 19(8):283–285, 1971.

[60] Block S.M., Blair D.F., and Berg H.C. Compliance of bacterial flagella measured with optical tweezers. *Nature*, 338:514–518, 1989.

[61] Ashkin A., Schütze K., Dziedzic J.M., Euteneuer U., and Schliwa M. Force generation of organelle transport measured in vivo by an infrared laser trap. *Nature*, 348:346–348, 1990.

[62] Denk W. and Webb W.W. Optical measurement of picometer displacements of transparent microscopic objects. *Appl. Opt.*, 29(16):2382–2391, 1990.

[63] Ghislain L.P. and Webb W.W. Scanning-force microscope based on an optical trap. *Opt. Lett.*, 18(19):1678–1680, 1993.

[64] Simmons R.M., Finer J.T., Chu S., and Spudich J.A. Quantitative measurements of force and displacement using an optical trap. *Biophys. J.*, 70(4):1813–1822, 1996.

[65] Tolic-Nørrelykke S.F., Schäffer E., Howard J., Pavone F.S., Jülicher F., and Flyvbjerg H. Calibration of optical tweezers with positional detection in the back focal plane. *Rev. Sci. Instrum.*, 77:103101, 2006.

[66] Visscher K., Gross S.P., and Block S.M. Construction of multiple-beam optical traps with nanometer-level position sensing. *IEEE J. Sel. Top. Quantum Electron. 2*, pages 1066–1076, 1996.

[67] Montes-Usategui M., Pleguezuelos E., Andilla J., and Martín-Badosa E. Fast generation of holographic optical tweezers by random mask encoding of Fourier components. *Optics Express*, 14:2101–2107, 2006.

[68] Pleguezuelos E., Carnicer A., Andilla J., Martín-Badosa E., and Montes-Usategui M. Holotrap: Interactive hologram design for multiple dynamic optical trapping. *Comp. Phys. Commun.*, 176:701–709, 2007.

[69] Bryant Z., Stone M.D., Gore J., Smith S.B., Cozzarelli N.R., and Bustamante C. Structural transitions and elasticity from torque measurements on DNA. *Nature*, 424:338–341, 2003.

[70] Sacconi L., Romano G., Ballerini R., Capitanio M., De Pas M., Giuntini M, Dunlap D., Finzi L., and Pavone F.S. Three-dimensional magneto-optic trap for micro-object manipulation. *Opt. Lett.*, 26:1359–61, 2001.

[71] Barnett S.M. and Allen L. Orbital angular momentum and nonparaxial light beams. *Opt. Commun.*, 110:670, 1994.

[72] He H., Friese M.E.J., Heckenberg N.R., and Rubinsztein-Dunlop H. Direct observation of transfer of angular momentum to absorptive particles from a laser beam with a phase singularity. *Phys. Rev. Lett.*, 75:826–829, 1995.

[73] La Porta A. and Wang M.D. Optical torque wrench: Angular trapping, rotation, and torque detection of quartz microparticles. *Phys. Rev. Lett.*, 92:190801, 2004.

[74] Keyser U.F., van der Does J., Dekker C., and Dekker N.H. Optical tweezers for force measurements on DNA in nanopores. *Rev. Sci. Instrum.*, 77:105105, 2006.

[75] Ishijima A., Kojima H., Funatsu T., Tokunaga M., Higuchi H., Tanaka H., and Yanagida T. Simultaneous observation of individual ATPase and mechanical events by a single myosin molecule during interaction with actin. *Cell*, 92:161–171, 1998.

[76] Mameren J., Modesti M., Kanaar R., Wyman C., Wuite G.J., and Peterman E.J. Dissecting elastic heterogeneity along DNA molecules coated partly with Rad51 using concurrent fluorescence microscopy and optical tweezers. *Biophys. J.*, 91:L78–80, 2006.

[77] Ashkin A. and Dziedzic J. M. Optical trapping and manipulation of viruses and bacteria. *Science*, 235:1517–1520, 1987.

[78] Ashkin A. Forces of a single-beam gradient laser trap on a dielectric sphere in the ray optics regime. *Biophys. J.*, 61:569–582, 1992.

[79] Neuman K.C. and Block S.M. Optical trapping. *Rev. Sci. Instrum.*, 75(9):2787–2809, 2004.

[80] Born M. and Wolf E. *Principles of optics: electromagnetic theory of propagation, interference and diffraction of light.* Cambridge University Press., 7th expanded ed. edition, 1999.

[81] Harada Y. and Asakura T. Radiation forces on a dielectric sphere in the Rayleigh scattering regime. *Opt. Commun.*, 124:529–541, 1996.

[82] Bazhan V. http://www.scatlab.org. 2003.

[83] Barton J.P., Alexander D.R., and Schaub S.A. Internal and near-surface elecromagnetic fields for a spherical particle irradiated by a focused laser beam. *J. Appl. Phys.*, 64:1632–1639, 1988.

[84] Maheu B., Gouesbet G., and Grehan G. A concise presentation of the generalized Lorenz-Mie theory for arbitrary location of the scatterer in an arbitrary incident profile. *J. Optics (Paris)*, 19:59–67, 1988.

[85] Barton J.P., Alexander D.R., and Schaub S.A. Theoretical determination of net radiation force and torque for a spherical particle illuminated by a focused laser beam. *J. Appl. Phys.*, 66(10):4594–4602, 1989.

[86] Grange W., Husale S., GÃ$\frac{1}{4}$ntherodt H.-J., and Hegner M. Optical tweezers system measuring the change in light momentum flux. *Rev. Sci. Instrum.*, 73(6):2308–2316, 2002.

[87] Future Technology Devices International Ltd. http://www.ftdichip.com/drivers/vcp.htm. 2010.

[88] Hayashi K., Forns N., Manosas M., Huguet J.M., and Ritort F. Stochastic resonance in the mechanical folding/unfolding of single DNA hairpins. *Submitted for publication*, 2010.

[89] Crick F. Central dogma of molecular biology. *Nature*, 227:561–563, 1970.

[90] Cantor C. and Schimmel P.R. *Biophysical chemistry*. Freeman, cop, New York, 1980.

[91] Calladine C.R. and Drew H.R. *Understanding DNA: the molecule and how it works*. Academic Press, San Diego, 1997.

[92] Saenger W. *Principles of Nucleic Acid Structure*. Springer-Verlag, New York, 1988.

[93] Cluzel P., Lebrun A., Heller C., Lavery R., Viovy J.-L., Chatenay D., and Caron F. DNA: An Extensible Molecule. *Science*, 271:792–794, 1996.

[94] van Mameren J., Gross P., Farge G., Hooijman P., Modesti M., Falkenberg M., Wuite G.J., and Peterman E.J. Unraveling the structure of DNA during overstretching by using multicolor, single-molecule fluorescence imaging. *Proc. Natl. Acad. Sci. USA*, 106:18231–6, 2009.

[95] Premilat S. and Albiser G. Conformations of A-DNA ad B-DNA in agreemet with fiber X-ray and infrared dichroism. *Nucl. Acids Res.*, 11(6):1897–1908, 1983.

[96] Watson J.D., Baker T.A., Bell S.P., Gann A., Levine M., and Losick R. *Molecular Biology of the Gene*. Benjamin Cummings, 5th. ed., San Francisco., 2004.

[97] Cui Y. and Bustamante C. Pulling a single chromatin fiber reveals the forces that maintain its higher-order structure. *Proc. Nat. Acad. Sci. USA*, 97:127–132, 2000.

[98] Darwin C. *On the origin of species*. Oxford University Press, London, 1951.

[99] Wagensberg J. *La rebelión de las formas*. Ed. Tusquets, Barcelona, 2004.

[100] Modrich P. and Richardson C.C. Bacteriophage T7 deoxyribonucleic acid replication *in vitro*. A protein of *Escherichia coli* required for bacteriophage T7 DNA polymerase activity. *J. Biol. Chem.*, 250:5508–5514, 1975.

[101] Dumont S., Cheng W., Serebrov V., Beran R.K., Tinoco Jr. I., Pyle A.M., and Bustamante C. RNA translocation and unwinding mechanism of HCV NS3 helicase and its coordination by ATP. *Nature*, 439(7072):105–108, 2006.

[102] Johnson D.S., Bai L., Smith B.Y., Patel S.S., and Wang M.D. Single molecule studies reveal dynamics of DNA unwinding by the ring-shaped T7 helicase. *Cell*, 29:1299–1309, 2007.

[103] Lionnet T., Spiering M.M., Benkovic S.J., Bensimon D., and Croquette V. Real-time observation of bacteriophage t4 gp41 helicase reveals an unwinding mechanism. *Proc. Natl. Acad. Sci. USA*, 104(50):19790–19795, 2007.

[104] Lee G.U., Chrisey L.A., and Colton R.J. Direct measurement of the forces between complementary strands of DNA. *Science*, 266:771 – 773, 1994.

[105] Boland T. and Ratner D. Direct measurement of hydrogen bonding in DNA nucleotide bases by atomic force microscopy. *Proc. Natl. Acad. Sci. USA*, 92:5297–5301, 1995.

[106] Essevaz-Roulet B., Bockelmann U., and Heslot F. Mechanical separation of the complementary strands of DNA. *Proc. Natl. Acad. Sci. USA*, 94:11935–11940, 1997.

[107] Bockelmann U., Essevaz-Roulet B., and Heslot F. DNA strand separation studied by single molecule force measurements. *Phys. Rev. E*, 58(2):2386–2394, 1998.

[108] Kishino A. and Yanagida T. Force measurements by micromanipulation of a single actin filament by glass needles. *Nature.*, 334:74–76, 1988.

[109] Breslauer K.J., Frank R., Blöcker H., and Marky L.A. Predicting DNA duplex stability from the base sequence. *Proc. Natl. Acad. Sci. USA*, 83:3746–3750, 1986.

[110] Clausen-Schaumann H., Rief M., Tolksdorf C., and Gaub H.E. Mechanical stability of single DNA molecules. *Biophys. J.*, 78:1997–2007, 2000.

[111] Thomen P., Bockelmann U., and Heslot F. Rotational drag on DNA: A single molecule experiment. *Phys. Rev. Lett.*, 88(24):248102, 2002.

[112] Manosas M. and Ritort F. Thermodynamic and kinetic aspects of RNA pulling experiments. *Biophys. J.*, 88(5):3224–3242, 2005.

[113] Woodside M.T., Anthony P.C., Behnke-Parks W.M., Larizadeh K., Herschlag D., and Block S.M. Direct measurement of the full, sequence-dependent folding landscape of a nucleic acid. *Science*, 314(5801):1001–1004, 2006.

[114] Koch S.J., Shundrovsky A., Jantzen B.C., and Wang M.D. Probing protein-DNA interactions by unzipping a single DNA double helix. *Biophys. J.*, 83:1098, 2002.

[115] Wen J.-D., Lancaster L., Hodges C., Zeri A.-C., Yoshimura S.H., Noller H.F., Bustamante C., and Tinoco Jr. I. Following translation by single ribosomes one codon at a time. *Nature*, 452:598–603, 2008.

[116] Cocco S., Marko J.F., and Monasson R. Theoretical models for single-molecule DNA and RNA experiments: from elasticity to unzipping. *arXiv:cond-mat/0206238v1*, 2002.

[117] Voulgarakis N.K, Redondo A., Bishop A.R., and Rasmussen K.Ø. Sequencing DNA by dynamic force spectroscopy: Limitations and prospects. *Nano Lett.*, 6:1483–1486, 2006.

[118] Peyret N. *Prediction of nucleic acid hybridization: Parameters and algorithms.* PhD thesis, Wayne State University, Department of Chemistry, Detroit, MI., 2000.

[119] Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucl. Acids Res.*, 31:3406–3415, 2003.

[120] Marky L.A. and Breslauer K.J. Calculating thermodynamic data for transitions of any molecularity from equilibrium melting curves. *Biopolymers*, 26:1601–1620, 1987.

[121] Erie D., Sinha N., Olson W., Jones R., and Breslauer K. A dumbbell-shaped, double-hairpin structure of DNA: a thermodynamic investigation. *Biochemistry*, 26:7150–7159, 1987.

[122] Gerland U., Bundschuh R., and Hwa T. Force-induced denaturation of RNA. *Biophys. J.*, 81:1324–1332, 2001.

[123] Süzen M., Sega M., and Holm C. Ensemble inequivalence in single-molecule experiments. *Phys. Rev. E*, 79:051118, 2009.

[124] Garcimartín A., Guarino A., Bellon L., and Ciliberto S. Statistical properties of fracture precursors. *Phys. Rev. Lett.*, 79:3202, 1997.

[125] Bonnot E. Romero R., Illa X. Mañosa L., Planes A., and Vives E. Hysteresis in a system driven by either generalized force or displacement variables: Martensitic phase transition in single-crystalline Cu-Zn-Al. *Phys. Rev. B*, 76:064105, 2007.

[126] Rubio-Bollinger G., Bahn S.R., Agraït N., Jacobsen K.W., and Vieira S. Mechanical properties of formation mechanisms of a wire of single gold atoms. *Phys. Rev. Lett.*, 87:026101, 2001.

[127] Socoliuc A., Bennewitz R., Gnecco E., and Meyer E. Transition from stick-slip to continuous sliding in atomic friction: Entering a new regime of ultralow friction. *Phys. Rev. Lett.*, 92:134301, 2004.

[128] Rothemund P.W.K. Folding dna to create nanoscale shapes and patterns. *Nature*, 440:297–302, 2006.

[129] Douglas S.M., Dietz H., Liedl T., Hogberg B., Graf F., and Shih W.M. Self-assembly of dna into nanoscale three-dimensional shapes. *Nature*, 459:414–418, 2009.

[130] Shoemaker D.D., Schadt E.E., Armour C.D., He Y.D., Garrett-Engele P., McDonagh P.D., Loerch P.M., Leonardson A., Lum P.Y., Cavet G., Wu L.F., Altschuler S.J., Edwards S., King J., Tsang J.S., Schimmack G., Schelter J.M., Koch J., Ziman M., Marton M.J., Li B., Cundiff P., Ward T., Castle J., Krolewski M., Meyer M.R., Mao M., Burchard J., Kidd M.J., Dai H., Phillips J.W., Linsley P.S., Stoughton R., Scherer S., and Boguski M.S. Experimental annotation of the human genome using microarray technology. *Nature*, 409:922–927, 2010.

[131] Dorsett Y. and Tuschl T. siRNAs: Applications in functional genomics and potential as therapeutics. *Nat. Rev. Drug Discovery*, 3:318–329, 2004.

[132] Russel R. RNA misfolding and the action of chaperones. *Front. Biosci.*, 13:1–20, 2008.

[133] Nykypanchuk D., Maye M.M., van der Lelie D., and Gang O. DNA-guided crystallization of colloidal nanoparticles. *Nature*, 451:549–552, 2008.

[134] Gotoh O. and Tagashira Y. Stabilities of nearest-neighbor doublets in double-helical DNA determined by fitting calculated melting profiles to observed profiles. *Biopolymers*, 20:1033–1042, 1981.

[135] Vologodskii A.V., Amirikyan B.R., Lyubchenko Y.L., and Frank-Kamenetskii M.D. Allowance for heterogeneous stacking in the DNA helix-coil transition theory. *J. Biomol. Struct. Dyn.*, 2:131–48, 1984.

[136] Manosas M., Collin D., and Ritort F. Force dependent fragility in RNA hairpins. *Phys. Rev. Lett.*, 96:218301, 2006.

[137] Bornschlögl T. and Rief M. Single molecule unzipping of coiled coils: sequence resolved stability profiles. *Phys. Rev. Lett.*, 96:118102, 2006.

[138] Seol Y., Skinner G.M., and Visscher K. Stretching of homopolymeric RNA reveals single-stranded helices and base-stacking. *Phys. Rev. Lett.*, 98:158103, 2007.

[139] Dessinges M.-N., Maier B., Zhang Y., Peliti M., Bensimon D., and Croquette V. Stretching single stranded DNA, a model polyelectrolyte. *Phys. Rev. Lett.*, 89(24):248102, 2002.

[140] Kirkpatrick S., Gelatt Jr. C.D., and Vecchi M.P. Optimization by simulated annealing. *Science*, 220:671–680, 1983.

[141] Tan Z.-J. and Chen S.-J. Nucleic acid helix stability: Effects of salt concentration, cation valence and size, and chain length. *Biophys. J.*, 90:1175–1190, 2006.

[142] Alemayehu S., Fish D.J., Brewood G.P., Horne M.T., Manyanga F., Dickman R., Yates I., and Benight A.S. Influence of buffer species on the thermodynamics of short DNA duplex melting: Sodium phosphate versus sodium cacodylate. *J. Phys. Chem. B*, 113:2578–2586, 2009.

[143] Owczarzy R., You Y., Moreira B.G., Manthey J.A., Huang L., Behlke M.A., and Walder J.A. Effects of sodium ions on DNA duplex oligomers: improved predictions of melting temperatures. *Biochemistry*, 43:3537–3554, 2004.

[144] Nakano S., Fujimoto M., Hara H., and Sugimoto N. Nucleic acid duplex stability: influence of base composition on cation effects. *Nucleic Acids Res.*, 27:2957–2965, 1999.

[145] Cheatham III T.E. and Kollman P.A. Md simulations highlight the structural differences among DNA:DNA, RNA:RNA and DNA:RNA hybrid duplexes. *J. Am. Chem. Soc.*, 119:4805–4825, 1997.

[146] Owczarzy R. Melting temperatures of nucleic acids: Discrepancies in analysis. *Biophys. Chem.*, 117:207–215., 2005.

[147] Hatch K., Danilowicz C., Coljee V., and Prentiss M. Measurements of the hysteresis in unzipping and rezipping double-stranded DNA. *Phys. Rev. E*, 75:051908, 2007.

[148] Huguet J.M., Forns N., and Ritort F. Statistical properties of metastable intermediates in DNA unzipping. *Phys. Rev. Lett.*, 103:248106, 2009.

[149] Greenleaf W.J., Woodside M.T, Abbondanzieri E.A, and Block S.M. Passive all-optical force clamp for high-resolution laser trapping. *Phys. Rev. Lett.*, 95:208102, 2005.

[150] Meresman H., Wills J.B., Summers M., McGloin D., and Reid J.P. Manipulation and characterisation of accumulation and coarse mode aerosol particles using a Bessel beam trap. *Phys. Chem. Chem. Phys.*, 11:11333–11339, 2009.

[151] Evans E. and Williams P. *Physics of Biomolecules and Cells.* Proceedings of the Les Houches Summer School, Session LXXV, edited by H. Flyvberg and F. Julicher. Springer-Verlag, Berlin, 2002.

[152] Zwanzig R. *Nonequilibrium Statistical Mechanics.* Oxford University Press, New York, 1st ed., 2001.

[153] Ritort F., Bustamante C., and Tinoco Jr. I. A two-state kinetic model for the unfolding of single molecules by mechanical force. *Proc. Nat. Acad. Sci.*, 99:13544, 2002.

[154] Manosas M., Mossa A., Forns N., Huguet J.M, and Ritort F. Dynamic force spectroscopy of DNA hairpins (II): Irreversibility and dissipation. *J. Stat. Mech (Theor. and Exp.)*, P02061, 2009.

[155] Mao H., Arias-Gonzalez J.R., Smith S.B., Tinoco I., and Bustamante C. Temperature control methods in a laser tweezers system. *Biophys J.*, 89:1308–1316, 2005.

[156] Sanger F., Nicklen S., and Coulson A.R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA*, 74:5463–7, 1977.

[157] Stout A.L. Detection and characterization of individual intermolecular bonds using optical tweezers. *Biophys. J.*, 80:2976–2986, 2001.

[158] Clarke J., Wu H.C., Jayasinghe L., Patel A., Reid A., and Bayley H. Continuous base identification for single-molecule nanopore DNA sequencing. *Nature Nanotech.*, 4:265–270, 2009.

[159] J.D. Jackson. *Classical electrodynamics*. Wiley, New York, 2d ed edition, 1975.

[160] Bouchiat C., Wang M.D., Allemand J., Strick T., Block S.M., and Croquette V. Estimating the persistence length of a worm-like chain molecule from force-extension measurements. *Biophys. J.*, 76:409–413, 1999.

[161] Wang M.D., Yin H., Landick R., Gelles J., and Block S.M. Stretching DNA with optical tweezers. *Biophys. J.*, 72:1335–1346, 1997.

[162] Press W.H., Teukolsky S.A., Vetterling W.T., and Flannery B.P. *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, New York, 2nd Ed., 1992.

List of publications

- Huguet J.M., Bizarro C.V., Forns N., Smith S.B., Bustamante C., and Ritort F. Single-molecule derivation of salt dependent base-pair free energies in DNA. *Proc. Nat. Acad. Sci. USA*, 107:15341-15346 (2010)
- Mossa A., Huguet J.M., and Ritort F. Investigating the thermodynamics of small biosystems with optical tweezers. *Physica E: Low-dimensional Systems and Nanostructures*, 42:666-671 (2010)
- Huguet J.M., Forns N., and Ritort F. Statistical Properties of Metastable Intermediates in DNA Unzipping. *Phys. Rev. Lett.*, 103:248106 (2009)
- Illa X., Huguet J.M., and Vives E. Work distributions in the T=0 random field Ising model. *Phys. Rev. E*, 79:021123 (2009)
- Mossa A., de Lorenzo S., Huguet J.M., and Ritort F. Measurement of work in single-molecule experiments. *Journal of Chemical Physics*, 130:234116 (2009)
- Mossa A., Manosas M., Forns N., Huguet J.M., and Ritort F. Dynamics force spectroscopy of DNA hairpins (I): force kinetics and free energy landscapes. *J. Stat. Mech. (Theor. and Exp.)* P02060 (2009)
- Manosas M., Mossa A., Forns N., Huguet J.M., and Ritort F. Dynamic force spectroscopy of DNA hairpins (II): Irreversibility and dissipation. *J. Stat. Mech. (Theor. and Exp.)* P02061 (2009)

- Hayashi K., Forns N., Manosas M., Huguet J.M., and Ritort F. Stochastic resonance in single DNA hairpins. *Submitted for publication* (2010)
- Forns N., de Lorenzo S., Manosas M., Hayashi K., Huguet J.M., and Ritort F. Improving signal-to-noise resolution in single molecule experiments using molecular constructs with short handles. *Submitted for publication* (2010)

Small Biosystems Lab
Departament de Física Fonamental, UB
Director • Dr. Fèlix Ritort Farran
(2010)