



UvA-DARE (Digital Academic Repository)

Traversing the free-energy pathways of intricate biomolecular processes

Enhanced simulation development and applications

Pérez de Alba Ortíz, A.

Publication date

2021

[Link to publication](#)

Citation for published version (APA):

Pérez de Alba Ortíz, A. (2021). *Traversing the free-energy pathways of intricate biomolecular processes: Enhanced simulation development and applications*. [Thesis, fully internal, Universiteit van Amsterdam].

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

VI

Summary, publications,
contributions and
acknowledgements

Summary

Advances in the atomistic understanding of biomolecular structure and function come with wide applications in medicine, pharmacology and biomaterials, as well as with fundamental answers about the nature of living things. Such insight can be brought by molecular dynamics (MD) simulations, which nowadays can model a huge variety of biosystems, such as proteins, nucleic acids, polysaccharides, etc. However, even with nowadays vast computational resources, the timescales that are usually accessible in standard MD simulations cannot reliably sample many relevant biological processes, which occur either too slowly or too infrequently. We refer to this as the *rare event* problem. To overcome it, the MD community has devised an arsenal of enhanced sampling techniques. One specific family of techniques, known as free-energy methods, work by exerting a biasing potential on key descriptive molecular degrees of freedom—i.e. collective variables (CVs)—that drive the transition of interest. By correcting for the bias afterward, one can extract insight in the form of free-energy landscapes—projected onto the CVs—with stable states, transition channels and barriers that can be related to experimental results. However, the computational cost of converging a free-energy calculation scales exponentially with the number of CVs; meaning that complex transitions that require many descriptors are typically out of grasp. Path-based methods offer a solution to this challenge. By performing the biasing on the one-dimensional progress parameter along an adaptive path connecting two known states in CV-space, rather than on the CVs themselves, path-based methods circumvent the curse of dimensionality. In this thesis, we introduce novel strategies to path-based enhanced sampling, and also combine them with other powerful simulation techniques; thus, advancing the capabilities to simulate complex molecular transitions. We assess our strategies by studying intricate conformational and chemical changes in a variety of biomolecules; including oligopeptides, sensor proteins, DNA and polysaccharides.

Part I of this thesis presents a general introduction to biomolecular simulation and a methodology chapter. Part II deals with combining experiments and simulations of polysaccharides. In Chapter 3, we study the ion-responsive rigidity of hyaluronan, a natural polyelectrolyte that regulates the mechanical properties of the extracellular matrix. Our multi-method approach employs 2D infrared spectroscopy, single-chain force spectroscopy and molecular simulations. Infrared spectroscopy and MD simulations evidence a specific binding of Ca^{2+} ions to the amide and carboxylate groups of hyaluronan, which does not occur for Na^+ ions. To study the implications of this binding, we use free-energy calculations and force spectroscopy. Both methods find that the presence of Ca^{2+} ions triggers a drastic decrease, of $\sim 50\%$, in the persistence length of hyaluronan. We discover that the mechanistic origin for this responsive flexibility is the weakening of hyaluronan's intramolecular

hydrogen bonds, due to the distinctive binding of Ca^{2+} ions. The understanding of this finely tuned responsive mechanism can elucidate how hyaluronan regulates cell behavior in the extracellular matrix. Additionally, simulations could guide the design of biopolymers with specific monomer combinations, i.e. ion-binding side chains, to achieve tunable rigidity responses. This is the only chapter that uses standard, rather than path-based, free-energy calculations. Standard methods are more than suitable when a single CV—in this case, the end-to-end distance of a hyaluronan chain—can describe the process of interest—i.e. the bending of the chain—and accessible simulation timescales are sufficient to sample relevant fluctuations, as we demonstrate in an error analysis.

In Part III of this thesis we present new developments in path-based methods, mainly concerning parallelism. Our novel approaches are tested in oligopeptides. In Chapter 4, we analyze the performance of multiple-walker path-metadynamics (PMD), our in-house developed path-based method. We consider the right- to left-handedness change in polyproline helices. Polyproline chains, whose conformation can be described in terms of ω -dihedrals around the peptide bonds, present an ideal test system for CV-dimensionality. Adding one more proline residue to a chain is equivalent to adding one more CV to the description of the transition. We study polyproline chains described by three, four and five CVs, i.e. a tetramer, a pentamer and a hexamer. Multiple-walker PMD delivers pathways and free-energy profiles with sub-exponential, and even sub-linear, performance scaling; thus surpassing the limitations of standard biasing techniques. The uniform sampling performed by the multiple replicas along the path also helps in the path optimization, which raises the performance above trivial parallelism. Regarding the handedness transition in polyproline, we observe a relation between the direction and the order of the residue rotations. By reversing the direction of the residue rotations, the order of the mechanism can change from an end-to-end (zipper-like) mechanism, to a center-to-end mechanism. This observation reconciles previous experimental and computational results. Moreover, we also observe that higher-dimensional CV-spaces tend to contain more possible pathways, which motivates the next chapter.

In Chapter 5, we introduce a novel method to perform a parallel search for reaction channels. Multiple-path-metadynamics (multiPMD) uses multiple repulsive and adaptive paths to simultaneously find competing transition mechanisms, and their free energies. The crucial aspect of the technique is the use of special walkers that induce repulsion between the paths, i.e. *repellers*, or that pinpoint known distinct intermediates, i.e. *attractors*. We assess the performance of multiPMD by simulating conformational transitions in alanine dipeptide—for which we find two pathways—and tetrameric polyproline—for which we find six pathways. We also introduce additional schemes using special walkers to handle cyclic paths and switching paths, i.e. paths that lie somewhere between two previously found mechanisms. More importantly, we also present another novel concept: the PathMap. After the multiple paths between two states have been located, a PathMap provides an interpretable view of the free-energy along each pathway, as well as of the merging and forking of the channels and the free-energy ridges between them. MultiPMD and PathMaps offer promising ways to efficiently handle competing, or debated,

transition mechanisms. In particular, we exploit the versatility of MultiPMD in Chapter 7.

In Part IV of this thesis, we study DNA base-pairing transitions using path-based methods. Specifically, we study the transition between Watson-Crick-Franklin (WCF) and Hoogsteen (HG) base-pairing. HG is an alternative base-pairing motif, in which the purine base is rotated 180° around the glycosidic bond with respect to the WCF conformation. The rotation of the purine can occur either remaining *inside* of the double helix, or flipping *outside* and then returning into the helix. In recent years, HG base pairs have been discovered to exist in a dynamical equilibrium with WCF base pairs. Moreover, the HG motif is required in many relevant biological functions. In Chapter 6, we present a mechanistic study of the WCF-to-HG transition of a specific A·T pair in a previously studied DNA sequence (A₆-DNA). Showcasing the flexibility of the framework, we apply the path approach in combination with several well-established sampling methods: steered MD, metadynamics, umbrella sampling, constrained MD and transition path sampling. This robustness might be of special interest to the simulation community. Additionally, the path formalism also allows us to study many CVs for the transition. We consider the purine base rotation and flipping, but also key hydrogen-bond distances, distances between neighboring bases and distances between the backbone atoms of the transitioning A·T pair. The relevance of the CVs is evaluated in the appendix of the chapter. Using the most relevant CVs, we find a preference for the outside pathway. We identify key mechanistic differences between the inside and the outside pathways. For example, the inside pathway induces a significant opening between the two neighboring bases of the adenine, in order to accommodate the rotation. This possible influence of the neighbors motivates the following chapter.

In Chapter 7, we employ the multiPMD method presented in Chapter 5. The efficiency of multiPMD enables the first systematic investigation of the WCF-to-HG base-pairing transition in multiple DNA sequences. We study seven different DNA chains, which are variations of the original A₆-DNA sequence treated in Chapter 6. In all sequences, we find that the outside pathway has the lowest free-energy barrier. Regarding the sequence-dependence of the transition, we observe a weak influence of the direct neighbors on the transitioning base. Instead, we observe a strong influence of the direction in which the purine base rotates, and if it is either towards the long or the short section of our DNA chain. This finding suggests a key role of local changes in DNA flexibility to modulate the dynamical WCF/HG base-pairing equilibrium. Such local changes in flexibility might be induced by stacking interactions in specific nucleotide sequences, or by protein-DNA complexes.

Part V of this thesis deals with the combination of path-based free-energy techniques with hybrid quantum mechanics/molecular mechanics (QM/MM) modeling. In Chapter 8, we investigate the dark-state recovery mechanism of light-using flavin (BLUF) photoreceptor proteins, which are known for their non-covalently bonded flavin chromophore and for their widely-varying signaling times. For example, the recovery time of the BlrB BLUF protein is of 2 s, which is 750 times shorter than that of the AppA BLUF protein (~ 25 min). Via path-based calculations we sample, for the first time, the dynamics of Gln rotation and tautomerization that lead to

the dark-state recovery. We calculate free-energy profiles for the BlrB and AppA proteins that corroborate a previously proposed mechanism. Our free-energy barriers also evidence the crucial role a Trp and a Met near the flavin chromophore as modulators of the recovery rate, by tuning the hydrogen-bonding interactions of the rotating Gln. This result shows how slight mechanistic adjustments in BLUF proteins are able to control large variations in signaling times, and therefore tune the light-responsive behavior of organisms. Multiscale modeling and path-based simulations hold the potential to efficiently evaluate directed mutations of BLUF photoreceptors, and thus calibrate signaling times for optogenetic or biosensor applications.

In this thesis, we have expanded the capabilities of path-based free-energy methods and put them to the test in various biosystems. We have incorporated well-established techniques—such as multiple replicas—to the path-based framework. More importantly, we have also developed novel schemes, such as multiPMD and PathMaps. Additionally, we combined path-based enhanced sampling with multiscale modeling, i.e. QM/MM, to enable even more cost-efficient calculations. These various method extensions have allowed us to overcome different challenges encountered while navigating the free-energy landscapes of various types of biomolecules. We hope for the advancements here presented to contribute to the long-term goals of biomolecular simulation; enabling the atomistic understanding of ever larger, more complex and realistic living systems.

Samenvatting

Vooruitgang in het atomistische begrip van de biomoleculaire structuur en functie gaat gepaard met brede toepassingen in de geneeskunde, de farmacologie en de biomaterialen, alsmede fundamentele antwoorden over de aard van levende wezens. Dergelijk inzicht kan worden verkregen door simulaties in moleculaire dynamica (MD), die tegenwoordig een enorme verscheidenheid aan biosystemen kunnen modelleren, zoals proteïnen, nucleïnezuren, polysacchariden, enzovoorts. Echter, zelfs met de tegenwoordige enorme computationele middelen, kunnen de tijdschalen die met standard MD simulaties worden bereikt niet alle relevante biologische processen omschrijven, die ofwel te traag zijn ofwel zelden voorkomen. We noemen dit het *rare event* probleem. Om dit probleem op te lossen, heeft de MD gemeenschap een arsenaal van verbeterde bemonsteringstechnieken ontwikkeld. Een specifieke familie van technieken, bekend als vrije-energie methoden, werken door het uitvoeren van een bias potentiaal op de belangrijkste beschrijvende moleculaire vrijheidsgraden—d.w.z. collectieve variabelen (CV's)—die de interessante moleculaire transitie omschrijven. Door achteraf te corrigeren voor de bias kan inzicht worden verkregen in de vorm van vrije-energielandschappen, geprojecteerd op de CV's, waarin stabiele toestanden, overgangskanalen en barrières in verband kunnen worden gebracht met experimentele resultaten. De computationele kosten van het convergeren van een vrije-energieberekening schalen echter exponentieel met het aantal CV's; wat betekent dat complexe overgangen waarvoor veel beschrijvende elementen nodig zijn, meestal buiten het bereik van deze technieken liggen. Pad-gebaseerde methoden bieden een oplossing voor deze uitdaging. Door de bias uit te voeren op de één-dimensionale progressieparameter langs een adaptief pad dat twee bekende toestanden in de CV-ruimte verbindt in plaats van op de CV's zelf, omzeilen pad-gebaseerde methodes de “vloek van de dimensionaliteit”. In deze dissertatie introduceren we nieuwe strategieën voor pad-gebaseerde verbeterde sampling, en combineren we ze met andere krachtige simulatietechnieken; op deze manier bevorderen we de mogelijkheden om complexe moleculaire overgangen te simuleren. We testen onze strategieën door ingewikkelde conformatie- en chemische veranderingen te bestuderen in een verscheidenheid van biomoleculen, waaronder oligopeptiden, sensoreiwitten, DNA en polysacchariden.

Deel I van dit proefschrift geeft een algemene inleiding tot biomoleculaire simulatie en een methodologisch hoofdstuk. Deel II behandelt de combinatie van experimenten en simulaties van polysacchariden. In Hoofdstuk 3 bestuderen we de ion-ontvankelijke stijfheid van hyaluronzuur, een natuurlijk polyelectrolyt dat de mechanische eigenschappen van de extracellulaire matrix regelt. Onze multi-methodische benadering maakt gebruik van 2D infraroodspectroscopie, enkelvoudige-keten krachtspectroscopie en moleculaire simulaties. Infraroodspectroscopie en MD simulaties tonen een specifieke binding aan van Ca^{2+} ionen aan

de amide- en carboxylaatgroepen van hyaluronzuur, die niet optreedt voor Na^+ ionen. Om de implicaties van deze binding te bestuderen, gebruiken we vrije-energieberekeningen en krachtspectroscopie. Beide methoden vinden dat de aanwezigheid van Ca^{2+} ionen leidt tot een drastische afname, van $\sim 50\%$, in de persistentielengte van hyaluronzuur. Wij ontdekken dat de mechanistische oorsprong voor deze responsieve flexibiliteit de verzwakking is van de intra-moleculaire waterstofbruggen van hyaluronzuur, als gevolg van de kenmerkende binding van Ca^{2+} ionen. Het begrijpen van dit fijn afgestemd en responsieve mechanisme kan verklaren hoe hyaluron het gedrag van cellen in de extracellulaire matrix reguleert. Bovendien kunnen simulaties een belangrijke leidraad zijn tijdens de ontwikkeling van biopolymeren met specifieke monomeercombinaties, d.w.z. ion-bindende zijketens, om afstembare stijfheidsreacties te bereiken. Dit is het enige hoofdstuk dat gebruik maakt van standaard, in plaats van pad-gebaseerde, vrije-energieberekeningen. Standaardmethoden zijn meer dan geschikt wanneer een enkele CV—in dit geval de eind-tot-eind afstand van een hyaluronzuurketen—het proces kan beschrijven dat van belang is—d.w.z. de buiging van de keten—en de toegankelijke simulatietijdschalen voldoende zijn om relevante fluctuaties te omschrijven, zoals we aantonen in een foutenanalyse.

In Deel III van dit proefschrift presenteren wij nieuwe ontwikkelingen in pad-gebaseerde methoden, voornamelijk met betrekking tot paralleliteit. Onze nieuwe benaderingen worden getest in oligopeptiden. In Hoofdstuk 4 analyseren we de prestaties van multiple-walker path-metadynamics (PMD), onze eigen ontwikkelde pad-gebaseerde methode. We beschouwen de verandering van rechtshandigheid naar linkshandigheid in polyproline helices. Polyproline-ketens, waarvan de conformatie kan worden beschreven in termen van ω -dihedralen rond de peptidebindingen, vormen een ideaal testsysteem voor de CV-dimensionaliteit. Het toevoegen van één proline residu aan een keten komt overeen met het toevoegen van één extra CV aan de beschrijving van de overgang. Wij bestuderen polyproline-ketens beschreven door drie, vier en vijf CV's, d.w.z. een tetramer, een pentamer en een hexamer. Multiple-walker PMD levert trajecten en vrije-energieprofielen met sub-exponentiële, en zelfs sub-lineaire, prestatie schaling; daarmee overtreffen we de beperkingen van standaard bias technieken. De uniforme sampling voortkomend uit de meervoudige replica's langs het pad helpt ook bij de optimalisatie van het pad, waardoor de prestaties boven triviaal parallele berekeningen uitstijgen. Met betrekking tot de overgang van de links- of rechtshandigheid in polyproline, zien we een relatie tussen de richting en de volgorde van de residu-rotaties. Door de richting van de residu-rotaties om te keren, kan de volgorde van het mechanisme veranderen van een eind-tot-eind (rits-achtig) mechanisme, naar een midden-tot-eind mechanisme. Deze observatie brengt eerdere experimentele en computationele resultaten met elkaar in overeenstemming. Bovendien zien we ook dat de hoger-dimensionale CV-ruimten meer mogelijke routes bevatten, wat de motivatie vormt voor het volgende hoofdstuk.

In Hoofdstuk 5 introduceren we een nieuwe methode om in parallel te zoeken naar reactiekanalen. Multiple-path-metadynamics (multiPMD) gebruikt meerdere afstotende en adaptieve paden om gelijktijdig concurrerende overgangsmechanismen

te vinden, en hun vrije energieën. Het cruciale aspect van de techniek is het gebruik van speciale walkers die afstoting tussen de paden induceren, d.w.z. *afstoters*, of die bekende verschillende tussenliggende paden aanwijzen, d.w.z. *attractors*. We evalueren de prestaties van multiPMD door conformatie-overgangen te simuleren in alanine dipeptide—waarvoor we twee paden vinden—en tetramerische polyproline—waarvoor we zes paden vinden. We introduceren ook extra schema's die gebruik maken van speciale walkers om cyclische paden en schakelpaden te behandelen, d.w.z. paden die ergens tussen twee eerder gevonden mechanismen liggen. Belangrijker nog is dat we ook een ander nieuw concept presenteren: de PathMap. Nadat de meervoudige paden tussen twee toestanden zijn gelokaliseerd, geeft een PathMap een interpreteerbaar beeld van de vrije-energie langs elk pad, en ook van de samenvoegingen en vertakkingen van de kanalen en de vrije-energie-richtels tussen de paden in. MultiPMD en PathMaps bieden veelbelovende manieren om efficiënt om te gaan met concurrerende, of omstreden, overgangsmechanismen. In Hoofdstuk 7 maken wij specifiek gebruik van de veelzijdigheid van MultiPMD.

In Deel IV van dit proefschrift bestuderen we de overgangen tussen DNA basenparen met behulp van op pad-gebaseerde simulatiemethoden. In het bijzonder bestuderen we de overgang tussen Watson-Crick-Franklin (WCF) en Hoogsteen (HG) basenparing. HG is een alternatief base-paringsmotief, waarbij de purinebase 180° geroteerd is rond de glycoside binding ten opzichte van de WCF-conformatie. De rotatie van de purinebase kan binnen de dubbele helix blijven, of erbuiten vallen en dan terugkeren in de helix. De laatste jaren is ontdekt dat HG-basenparen in een dynamisch evenwicht bestaan met WCF-basenparen. Bovendien speelt het HG-motief een belangrijke rol in veel biologische processen. In Hoofdstuk 6 presenteren we een mechanistische studie van de WCF naar HG-overgang van een specifiek A·T paar in een eerder bestudeerde DNA-sequentie (A_6 -DNA). Om de flexibiliteit van het raamwerk te demonstreren, passen we de pad-benadering toe in combinatie met verschillende beproefde simulatietechnieken: gestuurde MD, metadynamica, overkoepelende sampling, beperkte MD en simulering van overgangspaden. Deze robuustheid kan van bijzonder belang zijn voor de simulatiegemeenschap. Bovendien stelt het pad-formalisme ons ook in staat om vele CV's tijdens de overgang te bestuderen. We beschouwen de purinebase rotatie en flipping, maar ook belangrijke waterstofbrug-afstanden, afstanden tussen naburige basen en afstanden tussen de ruggengraat atomen van het overgaande A·T paar. De relevantie van de CV's wordt verder besproken in de appendix van dit hoofdstuk. Met behulp van de meest relevante CV's vinden we een voorkeur voor de buitenroute. We identificeren belangrijke mechanistische verschillen tussen de binnen- en de buitenroute. Bijvoorbeeld, de route binnendoor induceert een significante opening tussen de twee naburige basen van de adenine om de rotatie te accommoderen. Deze mogelijke invloed van de burens motiveert het volgende hoofdstuk.

In Hoofdstuk 7 gebruiken we de multiPMD methode die in Hoofdstuk 5 is gepresenteerd. De efficiëntie van multiPMD maakt het mogelijk om voor het eerst systematisch onderzoek te doen naar de WCF-naar-HG-baseparing overgang in meerdere DNA-sequenties. We bestuderen zeven verschillende DNA-ketens, die variaties zijn van de originele A_6 -DNA-sequentie die in Hoofdstuk 6 is behandeld. In alle sequen-

ties vinden we dat de buitenste route de laagste vrije-energie barrière heeft. Wat betreft de sequentie-afhankelijkheid van de overgang, zien we een zwakke invloed van de directe burens op de overgaande base. In plaats daarvan zien we een sterke invloed van de richting waarin de purinebase roteert, en of dit in de richting van het lange of het korte gedeelte van onze DNA-keten is. Deze bevinding suggereert een sleutelrol van lokale veranderingen in de flexibiliteit van het DNA bij het moduleren van het dynamische WCF/HG basenpaarevenwicht. Dergelijke lokale veranderingen in flexibiliteit kunnen worden geïnduceerd door stapelinteracties aanwezig in specifieke nucleotidesequenties, of door de vorming van eiwit-DNA-complexen.

Deel V van dit proefschrift behandelt de combinatie van pad-gebaseerde vrije-energietechnieken met hybride kwantummechanica/moleculaire mechanica (KM/MM) modellering. In Hoofdstuk 8 onderzoeken we het herstelmechanisme in de donkere toestand van licht-gebruikende flavine (BLUF) fotoreceptoreiwitten, die bekend staan om hun niet-covalent gebonden flavine chromofoor en om hun sterk variërende signaleringstijden. De hersteltijd van het BlrB BLUF-eiwit is bijvoorbeeld 2 s, wat 750 keer korter is dan die van het AppA BLUF-eiwit (~25 min). Via pad-gebaseerde berekeningen simuleren we, voor het eerst, de dynamica van Gln-rotatie en tautomerisatie die leiden tot het herstel van de donkere toestand. We berekenen vrije-energieprofielen voor de BlrB en AppA eiwitten die beide een eerder voorgesteld mechanisme bevestigen. Onze vrije-energie barrières tonen ook de cruciale rol aan van een Trp en een Met in de buurt van de flavine chromofoor als modulators van de herstelsnelheid, door het afstemmen van de waterstof-bindinginteracties van de roterende Gln. Dit resultaat toont aan hoe kleine mechanistische aanpassingen in BLUF-eiwitten in staat zijn om grote variaties in signaleringstijden te controleren, en dus het licht-responsieve gedrag van organismen af te stemmen. Meerschakige modellering en pad-gebaseerde simulaties bieden de mogelijkheid om gerichte mutaties van BLUF-fotoreceptoren efficiënt te evalueren, en zo signaaltijden te ijken voor optogenetische of biosensor toepassingen.

In dit proefschrift hebben we de mogelijkheden van op paden gebaseerde vrije-energiemethoden uitgebreid en getest in verschillende biosystemen. We hebben gevestigde technieken, zoals meervoudige replica's, opgenomen in het pad-gebaseerde raamwerk. Belangrijker nog is dat we ook nieuwe methoden hebben ontwikkeld, zoals multiPMD en PathMaps. Bovendien hebben we pad-gebaseerde simulatie verbeteringstechnieken gecombineerd met multidimensionale modellering, d.w.z. KM/MM, om nog meer kostenefficiënte berekeningen mogelijk te maken. Deze verschillende uitbreidingen van de simulatie methodiek hebben ons in staat gesteld om de verschillende uitdagingen te overwinnen die zich voordoen bij het navigeren door de vrije-energielandschappen van verschillende soorten biomoleculen. Wij hopen dat de hier gepresenteerde vorderingen zullen bijdragen aan de lange-termijn doelstellingen van biomoleculaire simulatie; het mogelijk maken van atomistisch begrip van steeds grotere, meer complexe en realistische levende systemen.

Resumen

Los avances en la comprensión atomística de la estructura y la función biomolecular suelen venir acompañados de amplias aplicaciones en medicina, farmacología y biomateriales, así como de descubrimientos sobre la naturaleza de los seres vivos. Este conocimiento puede ser alcanzado por simulaciones de dinámica molecular (DM), que hoy en día pueden modelar una enorme variedad de biosistemas, tales como proteínas, ácidos nucleicos, polisacáridos, etc. Sin embargo, incluso con los vastos recursos computacionales actuales, las escalas de tiempo a las que suelen acceder las simulaciones de DM estándar no pueden muestrear de forma fiable muchos procesos biológicos relevantes, que ocurren con demasiada lentitud o con muy poca frecuencia. A esto le llamamos el problema de los *eventos raros*. Para superarlo, la comunidad de DM ha ideado un arsenal de técnicas de muestreo mejorado. Una categoría específica de técnicas, conocida como métodos de energía libre, funciona ejerciendo un potencial de sesgo en los grados de libertad clave que describen a la molécula—es decir, variables colectivas (VCs)—que impulsan la transición de interés. Corrigiendo el sesgo a posteriori, se puede extraer información en forma de superficies de energía libre—proyectadas sobre las VCs—con estados estables, canales de transición y barreras que pueden relacionarse con los resultados experimentales. Sin embargo, el coste computacional de la convergencia de un cálculo de energía libre aumenta exponencialmente con el número de VCs, lo que significa que transiciones complejas que requieren muchos descriptores suelen estar fuera de nuestro alcance. Los métodos basados en caminos de transición ofrecen una solución a este reto y evitan la maldición de la dimensión. Esto se logra al realizar un sesgo en el parámetro de progreso unidimensional a lo largo de un camino adaptativo que conecta dos estados conocidos en el espacio de las VCs, en lugar de en las propias VCs. En esta tesis, introducimos nuevas estrategias para el muestreo mejorado basado en caminos, y también las combinamos con otras potentes técnicas de simulación; de este modo, aumentamos las capacidades para simular transiciones moleculares complejas. Adicionalmente, evaluamos nuestras estrategias estudiando intrincados cambios conformacionales y químicos en una variedad de biomoléculas, incluyendo oligopéptidos, proteínas sensoras, ADN y polisacáridos.

La Parte I de esta tesis presenta una introducción general a la simulación biomolecular y un capítulo de metodología. La Parte II trata de la combinación de experimentos y simulaciones de polisacáridos. En el Capítulo 3, estudiamos el efecto de los iones en la rigidez del ácido hialurónico, un polielectrolito natural que regula las propiedades mecánicas de la matriz extracelular. Nuestro enfoque multimétodo emplea espectroscopía infrarroja 2D, espectroscopía de fuerza de cadena única y simulaciones moleculares. La espectroscopía infrarroja y las simulaciones de DM evidencian una unión específica de los iones Ca^{2+} a los grupos amida y carboxilato del ácido hialurónico, que no se produce para los iones Na^+ . Para estudiar

las implicaciones de esta unión, utilizamos cálculos de energía libre y espectroscopía de fuerza. Ambos métodos encuentran que la presencia de iones Ca^{2+} desencadena una disminución drástica, de $\sim 50\%$, en la longitud de persistencia del ácido hialurónico. Descubrimos que el origen mecanístico de esta flexibilidad responsiva es el debilitamiento de los enlaces de hidrógeno intramoleculares del ácido hialurónico, el cual ocurre debido a la unión específica de los iones Ca^{2+} . La comprensión de este mecanismo de respuesta finamente ajustado puede aclarar cómo el ácido hialurónico regula el comportamiento celular en la matriz extracelular. Esto significa que las simulaciones podrían guiar el diseño de biopolímeros con combinaciones específicas de monómeros, es decir, cadenas laterales para la unión de iones, para lograr respuestas de rigidez controlables. Este es el único capítulo que utiliza cálculos de energía libre estándar, en lugar de los basados en caminos. Los métodos estándar son perfectamente adecuados cuando: 1) una sola VC puede describir el proceso de interés (en este caso, la distancia de extremo a extremo de una cadena de ácido hialurónico describe su flexión) y, 2) las escalas de tiempo computacionalmente accesibles son suficientes para muestrear las fluctuaciones relevantes, como demostramos en un análisis de errores.

En la Parte III de esta tesis presentamos nuevos desarrollos en los métodos basados en caminos de transición, principalmente en lo que respecta a la paralelización. Nuestros nuevos enfoques se prueban en oligopéptidos. En el Capítulo 4, analizamos el rendimiento de la metadinámica de camino (MDC) con múltiples caminantes, nuestro propio método basado en caminos. Para evaluarlo, consideramos el cambio de mano derecha a izquierda en hélices de poliprolina. Las cadenas de poliprolina, cuya conformación puede describirse en términos de diedros ω alrededor de los enlaces peptídicos, presentan un sistema de prueba ideal para la dimensionalidad de VCs. Añadir un residuo más de prolina a una cadena equivale a añadir una VC más a la descripción de la transición. Estudiamos cadenas de poliprolina descritas por tres, cuatro y cinco VCs, es decir, un tetrámero, un pentámero y un hexámero. La MDC con múltiples caminantes proporciona rutas y perfiles de energía libre con un escalado de rendimiento subexponencial, e incluso sublineal; superando así las limitaciones de las técnicas de muestreo estándar. El muestreo uniforme realizado por las múltiples réplicas a lo largo del camino también ayuda a la optimización del camino, lo que eleva el rendimiento por encima del paralelismo trivial. En cuanto a la transición de las hélices de poliprolina, observamos una relación entre la dirección y el orden de las rotaciones de los residuos. Al invertir la dirección de las rotaciones de los residuos, el orden del mecanismo puede cambiar de un mecanismo de extremo a extremo (tipo cremallera) a un mecanismo de centro a extremo. Esta observación concilia los resultados experimentales y computacionales anteriores. Además, observamos que los espacios VCs de mayor dimensión tienden a contener más vías posibles, lo cual sirve de motivación para el siguiente capítulo.

En el Capítulo 5, introducimos un novedoso método para realizar una búsqueda paralela de canales de reacción. La metadinámica de caminos múltiples (MDCmulti) utiliza múltiples caminos repulsivos y adaptativos para encontrar simultáneamente mecanismos de transición, que compiten entre sí, y sus energías libres. El aspecto crucial de esta técnica es el uso de caminantes especiales que inducen repulsión entre

los caminos, es decir, *repulsores*, o que señalan distintos estados intermedios conocidos, es decir, *atractores*. Evaluamos el rendimiento de MDCmulti simulando las transiciones conformacionales en un dipéptido alanina—para el que encontramos dos caminos—y en la poliprolina tetramérica—para la que encontramos seis caminos. También introducimos esquemas adicionales utilizando caminantes especiales para manejar caminos cíclicos y cambios de camino, es decir, caminos que se encuentran en algún lugar entre dos rutas encontradas previamente. Lo que es más importante, también presentamos otro concepto novedoso: el Mapa de Caminos. Una vez localizados los múltiples caminos entre dos estados, un Mapa de Caminos proporciona una visión interpretable de la energía libre a lo largo de cada camino, así como de la fusión y bifurcación de los canales y las crestas de energía libre entre ellos. La MDCmulti y los Mapas de Caminos ofrecen formas prometedoras de manejar eficientemente los mecanismos de transición que compiten entre sí o que son objeto de debate. En particular, explotamos la versatilidad de MDCmulti en el Capítulo 7.

En la Parte IV de esta tesis, estudiamos las transiciones de apareamiento de bases en ADN utilizando métodos basados en caminos de transición. En concreto, estudiamos la transición entre los apareamientos de bases Watson-Crick-Franklin (WCF) y Hoogsteen (HG). El apareamiento HG es una estructura alternativa, en el que la purina gira 180° alrededor del enlace glucosídico con respecto a la conformación WCF. La rotación de la purina puede producirse ya sea permaneciendo dentro de la doble hélice, o bien girando fuera y volviendo a entrar en la hélice. En los últimos años, se ha descubierto que los pares de bases HG existen en un equilibrio dinámico con los pares de bases WCF, y que la configuración HG es necesaria en muchas funciones biológicas relevantes. En el Capítulo 6, presentamos un estudio mecanístico de la transición de WCF a HG de un par A·T en una secuencia de ADN previamente estudiada (A_6 -ADN). Para mostrar la flexibilidad de nuestra técnica, aplicamos el enfoque de caminos en combinación con varios métodos de muestreo bien establecidos: DM dirigida y restringida, metadinámica, paraguas de muestreo, y muestreo de caminos de transición. Esta flexibilidad podría ser de especial interés para la comunidad de simulación. Además, el formalismo del camino también nos permite estudiar muchas VCs para la transición. Consideramos la rotación de la purina y su salida de la hélice, pero también las distancias clave de los enlaces de hidrógeno, las distancias entre las bases vecinas y las distancias entre los átomos de la columna vertebral del par A·T en transición. La relevancia de las VCs se evalúa en el apéndice del capítulo. Utilizando las VCs más relevantes, encontramos una preferencia por la transición fuera de la doble hélice e identificamos diferencias mecánicas clave entre las vías interior y exterior. Por ejemplo, la vía interior induce una apertura significativa entre las dos bases vecinas de la adenina, para acomodar la rotación. Esta posible influencia de los vecinos motiva el siguiente capítulo.

En el Capítulo 7, empleamos el método MDCmulti presentado en el Capítulo 5. La eficiencia de MDCmulti permite la primera investigación sistemática de la transición de apareamiento de bases WCF a HG en múltiples secuencias de ADN. Estudiamos siete cadenas de ADN diferentes, que son variaciones de la secuencia original de A_6 -ADN tratada en el Capítulo 6. En todas las secuencias, encontramos que la vía exterior tiene la barrera de energía libre más baja. En cuanto a la relación entre

la secuencia y la transición, observamos una débil influencia de los vecinos directos en la rotación. En cambio, observamos una fuerte influencia de la dirección en la que gira la purina, y si es hacia la sección larga o corta de la cadena de ADN. Este hallazgo sugiere un papel clave de los cambios locales en la flexibilidad del ADN, los cuales modulan el equilibrio dinámico entre pares de bases WCF y HG. Estos cambios locales en la flexibilidad podrían ser inducidos por interacciones de apilamiento en secuencias de nucleótidos específicas, o por complejos proteína-ADN.

La Parte V de esta tesis combina técnicas de energía libre basadas en caminos de transición con modelado híbrido de mecánica cuántica y mecánica molecular (MC/MM). En el Capítulo 8, investigamos el mecanismo de recuperación del estado oscuro en proteínas sensoras de luz azul que usan flavina (BLUF, por sus siglas en inglés), conocidas por su cromóforo no unido covalentemente y por sus tiempos de señalización ampliamente variables. Por ejemplo, el tiempo de recuperación de la proteína BlrB BLUF es de 2 s, que es 750 veces más corto que el de la proteína AppA BLUF (~25 min). Mediante cálculos basados en caminos, muestreamos por primera vez la dinámica de rotación y tautomerización de la Gln que conduce a la recuperación del estado oscuro. Calculamos perfiles de energía libre para las proteínas BlrB y AppA que corroboran un mecanismo propuesto previamente. Nuestras barreras de energía libre también evidencian el papel crucial de un Trp y una Met cerca del cromóforo de flavina para modular la tasa de recuperación a través del ajuste de interacciones de enlace de hidrógeno de la Gln en rotación. Este resultado muestra cómo ligeros ajustes mecanísticos en las proteínas BLUF son capaces de controlar grandes variaciones en los tiempos de señalización y, por tanto, afinar el comportamiento de respuesta a la luz de los organismos. El modelado multiescala y las simulaciones basadas en trayectorias tienen el potencial de evaluar eficientemente mutaciones dirigidas en los fotorreceptores BLUF, y así calibrar los tiempos de señalización para aplicaciones optogenéticas o de biosensores.

En esta tesis, hemos ampliado las capacidades de los métodos de energía libre basados en caminos de transición y los hemos puesto a prueba en varios biosistemas. Hemos incorporado técnicas bien establecidas—como las réplicas múltiples—al marco basado en caminos. Lo que es más importante, también hemos desarrollado nuevos esquemas, como MDCmulti y Mapas de Caminos; y hemos combinado el muestreo mejorado basado en caminos con el modelado multiescala, es decir, MC/MM, para permitir cálculos aún más rentables. Estas diversas extensiones del método nos han permitido superar diferentes retos encontrados al navegar por las superficies de energía libre de varios tipos de biomoléculas. Esperamos que los avances aquí presentados contribuyan a los objetivos a largo plazo de la simulación biomolecular, permitiendo la comprensión atómica de sistemas vivos cada vez más grandes, complejos y realistas.

List of publications

Part of this thesis

- Chapter 3: G. Giubertoni[†], **A. Pérez de Alba Ortíz**[†], F. Bano, X. Zhang, R.J. Linhardt, D.E. Green, P.L. DeAngelis, G.H. Koenderink, R.P. Richter, B. Ensing, and H.J. Bakker. *Strong reduction of the chain rigidity of hyaluronan by selective binding of Ca²⁺ ions*. *Macromolecules* **54**, 1137-1146 (2021).
- Chapter 4: **A. Pérez de Alba Ortíz**, A. Tiwari, R.C. Puthenkalathil, and B. Ensing. *Advances in enhanced sampling along adaptive paths of collective variables*. *J. Chem. Phys.* **149**, 072320 (2018).
- Chapter 5: **A. Pérez de Alba Ortíz** and B. Ensing. *Simultaneous sampling of multiple molecular transition pathways*. To be submitted.
- Chapter 6 and Chapter 2, Section 2.3.2: **A. Pérez de Alba Ortíz**, J. Vreede, and B. Ensing. *The Adaptive Path Collective Variable: A Versatile Biasing Approach to Compute the Average Transition Path and Free Energy of Molecular Transitions*. In: Bonomi M., Camilloni C. (eds) *Biomolecular Simulations. Methods in Molecular Biology*, vol. **2022**, pp. 255-290. Humana, New York, NY, (2019).
- Chapter 7: **A. Pérez de Alba Ortíz**, J. Vreede, and B. Ensing. *Sequence-dependence of transient Hoogsteen base-pairing in DNA*. To be submitted.
- Chapter 8: **A. Pérez de Alba Ortíz**, C. Rovira, and B. Ensing. *The dark-state recovery mechanism of blue light-using flavin proteins*. In preparation.

Other publications

- F. Hooft, **A. Pérez de Alba Ortíz**, and B. Ensing. *Discovering collective variables of molecular transitions via genetic algorithms and neural networks*. *J. Chem. Theory Comput.* (2021)¹.
- M.A. Mroginski,... , **A. Pérez de Alba Ortíz**,... , I. Schapiro. *Frontiers in multiscale modeling of photoreceptor proteins*. *Photochem. Photobiol.* **97**, 243-269 (2021).
- O. O. Sofronov[†], G. Giubertoni[†], **A. Pérez de Alba Ortíz**, B. Ensing, and H.J. Bakker. *Peptide side-COOH groups have two distinct conformations under biorelevant conditions*. *J. Phys. Chem. Lett.* **11**, 3466-3472 (2020).

[†]These authors contributed equally.

¹Referenced in Appendix 6.A.

- J. Vreede, **A. Pérez de Alba Ortíz**, P.G. Bolhuis, and D.W.H. Swenson. *Atomistic insight into the kinetic pathways for Watson-Crick to Hoogsteen transitions in DNA*. *Nucleic acids research* **47**, 11069-11076 (2019)².
- T. van Heesch, **A. Pérez de Alba Ortíz**, and B. Ensing. *DiSCo: Discovery of State based Collective Variables*. In preparation³.
- E. Schuler[†], **A. Pérez de Alba Ortíz**[†], B. Ensing, and G.J. Gruter. *Mechanistic insights in hydroxide-catalyzed formate coupling and its reversibility in the presence of potassium*. In preparation.

²Closely related to Chapter 6.

³Referenced in Appendix 8.A.

Author contributions

- Chapter 3: G.G. and **A.P.A.O.** equally contributed to the work. G.H.K. conceived the idea. H.J.B., G.G., **A.P.A.O.**, R.P.R., G.H.K., and B.E. designed the project. G.G. performed the linear IR and 2DIR measurements and together with H.J.B., analyzed the results. **A.P.A.O.** performed the MD simulations, and together with B.E. analyzed the results. F.B. performed the single chain force spectroscopy experiments and together with R.P.R. analyzed the results. X.Z., R.J.L., D.E.G., and P.L.D. produced the thiodonor and the SH-HA construct for the single-chain force spectroscopy experiments. H.J.B., G.G., **A.P.A.O.**, F.B., R.P.R., G.H.K. P.L.D., and B.E. wrote the paper.
- Chapter 4: B.E. conceived the idea. **A.P.A.O.**, A.T., R.C.P., and B.E. designed the project. **A.P.A.O.** wrote the theoretical background, extended the path-metadynamics code to use multiple walkers, produced code to pre- and post-process the data, and performed and wrote the case study on the PPI-to-PPII transition in proline oligomers. A.T. produced code to pre- and post-process the data, performed and wrote the case study on the acidity constant calculation of acetic acid in aqueous solution (not included in this thesis). R.C.P. produced code to pre- and post-process the data, performed and wrote the case study on the H₂ formation catalyzed by a hydrogenase enzyme (not included in this thesis). B.E. analyzed the results together with **A.P.A.O.**, A.T. and R.C.P., and revised the manuscript.
- Chapter 5: **A.P.A.O.** conceived the idea of multiple-path-metadynamics. **A.P.A.O.** and B.E. conceived the idea of PathMaps and designed the project. **A.P.A.O.** extended the path-metadynamics code, produced code to pre- and post-process the data, performed the simulations, analyzed the results together with B.E., and wrote the manuscript. B.E. revised the manuscript.
- Chapter 6 and Chapter 2, Section 2.3.2: J.V. and B.E. conceived the idea. **A.P.A.O.**, J.V., and B.E. designed the project. **A.P.A.O.** produced code to pre- and post-process the data, performed the simulations, analyzed the results together with J.V. and B.E., and wrote the manuscript. J.V. prepared the initial DNA system. J.V. and B.E. revised the manuscript.
- Chapter 7: J.V. conceived the idea. **A.P.A.O.**, J.V., and B.E. designed the project. **A.P.A.O.** prepared the systems, produced code to pre- and post-process the data, performed the simulations, analyzed the results together with J.V. and B.E., and wrote the manuscript. J.V. and B.E. revised the manuscript.

- Chapter 8: **A.P.A.O.** and B.E. conceived the idea. **A.P.A.O.**, C.R. and B.E. designed the project. **A.P.A.O.** prepared the MM system, prepared the QM/MM system guided by C.R., produced code to pre- and post-process the data, and performed the simulations. C.R. and B.E. analyzed the results together with **A.P.A.O.**, and revised the manuscript.

Acknowledgements

A PhD is quite a pathway to traverse (pun intended). The last four years have been one of the most transformative periods of my life, full of hard work, breakthroughs, growth and fun. I am extremely proud of this thesis, and I am equally thankful to all the people who helped making it possible. Here, I will make an attempt to show my appreciation to all of you.

First of all, I would like to acknowledge my PhD advisor, **Bernd Ensing**. I am immensely thankful for your genuine support—career and otherwise—and for all the doors you have opened for me. Your passion for science is admirable. I fondly look back at my spontaneous visits to your office whenever I had either a big problem or a big solution, and the long conversations that we would have afterward. I have really missed these chats during the pandemic and I will continue to heartily remember them for years to come. It has been a great fortune to have you as a supervisor. Like I said before, it has been an honor to be a Bernd-o-naut (and I really hope one day this name will catch on).

I also want to thank my co-promotor, **Evert Jan Meijer**. I have always admired your talent to guide and teach via questions (in a very Socratic manner). I very much appreciate both your lessons in quantum chemistry and your encouragement to pursue a research career. Thank you for being so approachable and supportive.

I also wish to thank my committee members: **Huib Bakker**, **Peter Bolhuis**, **Herma Cuppen**, **Alessandro Laio**, **Carne Rovira** and **Jocelyne Vreede**. It is an honor to have such highly-regarded experts evaluating my thesis and as opponents in my public defense. Moreover, I also had the privilege of learning from many you in other contexts, for which I will also thank you in the following paragraphs.

I also want to thank my wonderful paranymphs. **Thor**, dankjewel! You are a very kind person and I very much appreciate the chance not only to work with you, but also to share this kind of big events. I wish all the best for your future and I look forward to watching you in your own defense. **Maria**, muchas gracias por siempre estar ahí para tomar un café, ir al borrel o al molino, o contestar innumerables mensajes. Estoy muy feliz de haberte conocido y de haber compartido todo este tiempo contigo. ¿Puente?

I thank all members of the Computational Chemistry group: **Jocelyne**, I am very thankful for all your teachings, ranging from how to model biomolecular systems to how to supervise my first students. I also appreciate your advice and recommendation when looking for postdoc positions. I enjoyed collaborating on the DNA base rolling projects, and I hope that we stay in touch for a long time. **Peter**, I think that on each Monday meeting I took home at least one important scientific lesson, and I consider myself very lucky for that. I also really enjoyed coinciding with you in a few conferences abroad, and having the chance to talk about life in academia. I wholeheartedly thank you for your support during my PhD, and for

your recommendations to jumpstart my next career step. **Herma**, I highly appreciate the chance to learn from you whenever you joined our meetings. I was always captivated by the systems and topics that you study. I am also thankful for your interest in my work, as well as for your career advice and for your help in revising my HPC-Europa3 proposal. **David Dubbeldam**, thank you for always having great feedback and questions during the group presentations, and also for bringing up fascinating topics in your own talks, as well as during breaks. Also, thank you very much for your help with Carbon and thanks also to **Mahdi** for carrying on with this responsibility. **Tristan**, I have always found your research very inspiring, and I am glad to keep hearing about it, now as part of the Soft Matter community.

I would also like to thank my PhD colleagues at the Computational Chemistry group. First, the other two Bernd-o-nauts who accompanied me for most of the way: **Ambuj**, I still remember the day I met you when I had my interview. Your enthusiasm was already a sign of the great time that was ahead of me in Amsterdam. I am very glad to have collaborated with you and to have shared so many borrels. **Rakesh**, it was amazing to have a football-fan buddy in the group, except for that time when Liverpool eliminated Barcelona. I am very happy to have worked with you and to have shared a lot of time together in the office. **Nitish**, I very much enjoyed our coffee breaks, cinema outings and chats in general. It was a true honor to be your paranymp and I am happy that we are still in touch. I look forward to visiting you soon in Denmark, and also to go take a look at Govindalab. **Arjun**, I am very glad that you gave such nice talks about nucleation, because now it is becoming part of my day to day. I am also very happy about all those borrels we shared, and to have a rockstar friend. **Yuliia**, it is great to have such a talented friend. Thank you for sharing that talent with me in the form of a couple of paintings. Also, thanks to you and **Vlad** for hosting so many gatherings at your former place, and for letting me stay that snowy night a couple of years ago. **Verena**, it was great to always have a friend willing to get some dessert after lunch and some beers after work. And recently, also willing to cycle to Utrecht, together with the other “polymer girl”. I appreciate a lot hanging out with you and **Toni Jurn**, from the bottom of my heart, thank you for all those times you said “Beer?”. You are the kind of mate every office needs. I will always be thankful to you and **Claire** for inviting me to my first Dutch wedding. **Hannah**, thank you for all your kind and supportive words after presentations. You are a great friend to have around, and I am excited about working in more similar fields now, colloid buddy. **Sander**, it was great to have you as an office mate, always up to chat about science and fun stuff, and to offer help whenever needed. **Ferry**, it was a pleasure to work with you on the FABULOUS project. I wish we had more fun in person at the beginning of your PhD. We should compensate when the pandemic is over. Thank you also for your help with translating my summary. **Faidon**, although we coincided only shortly in Amsterdam, it has been a pleasure to find you in conferences. I wish you all the best and hope to encounter each other again soon. **Becca**, it was great to have you a couple of months a year here in the Netherlands. How to forget those coffee breaks with U.S. treats? **Chiara**, I loved going to your defense in Ghent. I hope soon we can meet more often in Amsterdam. **Chunhua**, it was an honor to

be present at your marriage proposal in Schiphol and it was very nice to celebrate Lunar New Year with you. **Mojgan**, it was very nice to have you in the team. I also appreciate your help when preparing my HPC-Europa3 proposal. **David Swenson**, it was great to have you in the Path meeting. I have learned a lot from you about transition path sampling and about software development. **György**, I hope to see you in Vienna next year. Thank you for your advice for the grant proposal and for nice discussions about path sampling. I also want to thank **Yingying**, **Adithya**, **Tamika**, **Jim**, **Johan** and **Ivan** for nice scientific and non-scientific chats in the coffee corner. I would also like to thank some students who have been part of the group: **Phillip**, **Teun**, **Thijs**, **Noortje** and **Saarika**. I especially want to mention **Denice** and **Joshka**, who were the first students I supervised.

Por supuesto, gracias también a los latinoamericanos en HIMS: **Ariana**, ha sido un gusto tener a una compatriota en el grupo, especialmente alguien tan buena onda y exitosa como tú. Aprecio mucho todos tus consejos a lo largo del doctorado. **Mauricio**, fue un gusto conocerte en Erice y hacernos famosos en Lausanne. Me parece genial que ahora estés en Ámsterdam y espero que sigamos en contacto. **Roberto**, una de las primeras fiestas a las que fui en Ámsterdam fue tuya. Creo que marcó estándares y fue una señal de lo buenos que serían los siguientes cuatro años. **Alan**, muchas gracias por guiarme en todos los asuntos de CONACYT cuando llegué y por organizar las buenísimas comidas mexicanas. **Hernán**, aunque seas merengue, siempre ha sido un gusto verte para tomar café o cerveza.

Also within HIMS, I was lucky enough to find amazing collaborators. **Eric**, it is a pleasure to work with you in the formate coupling project. Who would say this started in Brainwave? Thank you also to **Gert-Jan Gruter** and **Shiju Raveendran** for supporting our collaboration. A block away from HIMS, I also found great collaborators in AMOLF. **Giulia**, I consider myself very lucky to have met you. You are an amazing scientist to collaborate with (in two projects!) and a very kind friend. **Oleksandr**, it was very nice to work with you. I wish you the best in the future. **Gijsje**, thank you very much for all your feedback and advice. **Huib**, it was a great experience to collaborate with you and your team. Thank you very much for the opportunity. **Fouzia**, **Ralf**, **Xing**, **Bob**, **Dixy** and **Paul**, even though we never met in person, it has been a pleasure to work together.

Carne, moltes gràcies. Disfruté mucho mi tiempo en Barcelona, tanto la parte científica, como aprender de ti y convivir con el grupo. Espero que sigamos en contacto y vernos de nuevo en persona pronto. **Oriol**, gracias por darme la bienvenida y por tu interés en nuestros métodos. Espero verte pronto en Ámsterdam. **Joan** y **Alba**, ha sido un gusto conocerlos en Lausanne y vernos de nuevo en Barcelona. También me ha dado mucho gusto poder ver sus defensas, aunque fuera en línea. **Bea** e **Irene**, gracias por aclararme varias dudas y por estar ahí para charlar en la oficina. **Qinghua**, thank you very much for your interest in my work. All the best!

I also want to thank people who helped me in preparing for the PhD. I thank my MSc thesis supervisors in Munich—**Thomas Stecher** and **Harald Oberhofer**—for valuable lessons on quantum chemistry, multiscale modeling and reaction coordinates. I also want to say “gracias” to **Karsten Reuter** for accepting me in his group, for valuable advice regarding looking for a PhD in the Netherlands, and for

his recommendation. I would also like to thank **Martin Zacharias**, for teaching me about enhanced sampling, for giving me the opportunity to work on the topic as a research assistant in his group, and for extending me a recommendation.

Also in Munich I met amazing friends with whom I am glad to stay in touch. **Helisa** and **Saumitra**, who would think that supeprogramming of supercomputers would be key for my PhD? I am very glad to have met you and I look forward to seeing you in person again. **Juan Carlos** y **Alejandra**, es un gusto seguir en contacto con ustedes, les deseo lo mejor siempre. **Adriana**, me alegra mucho habernos conocido. Nos vemos pronto en Sevilla. Estoy seguro que lo pasaremos genial. **Cindy** y **David**, me da mucho gusto que nuestra amistad continúe hasta Utrecht. Aprecio mucho haber contado con ustedes este último año, y en especial haber pasado las fiestas de diciembre juntos.

I also want to thank many other friends. **Simon** and **Lisanne**, thanks for great times in Utrecht and across Canada. **Sahar**, you were the first person with whom I went out for dinner after moving to the Netherlands. Thank you for that and for still being around. **Pete**, thanks for helping me with the cover and for being part of the Utrecht gang. **Laura**, me da mucho gusto haberte conocido. El primer año del doctorado no hubiera sido ni la mitad de divertido sin nuestras salidas al molino. **Gustavo**, he disfrutado mucho nuestro tiempo en Ternium y nuestras visitas en Europa. Espero pronto felicitarte por tu defensa también. **Gil**, muchas gracias por estar siempre en contacto y siempre pendiente de mi y de mis publicaciones. **Dulce** y **Horacio**, me alegra mucho de que sigamos en contacto después de tanto tiempo y de que hayamos tenido la oportunidad de visitarnos mutuamente y vivir aventuras extremas. **Carolina** y **Emanuel**, gracias por seguir en contacto, por esas noches en Nueva York y por haber venido a Ámsterdam. **Pepefer** y **Fox**, es un gusto seguir en contacto y habernos visitado. **Erika**, espero verte pronto en Múnich otra vez y que visites por aquí también. **Diana**, tenemos que vernos pronto para celebrar, ya sea en Bélgica o en Holanda. **Lalo**, gracias por seguir en contacto constante y por visitar. ¡Nos vemos pronto en Berlín! **Pepe**, gracias por siempre estar atento, incluso desde la distancia. **María Velasco**, me da mucho gusto que ahora vivas más cerca. Espero que nos veamos muy seguido.

Ma, muchas gracias por todo lo que me has enseñado, incluyendo que no hay límites para lo que uno puede alcanzar. Como te he dicho antes, el hecho de que yo haya llegado hasta aquí sólo ha sido posible gracias a la formación y al ejemplo que me has dado. Sé que no siempre ha sido fácil a la distancia y por eso quiero que sepas que valoro infinitamente contar contigo y que siempre te siento cerca. Este logro, y los que vendrán, también son tuyos.

Su, ya llevamos más de 14 años atravesando fronteras juntos. No sé en dónde estaremos dentro de un par de años, pero estoy seguro de que estaremos unidos y apoyándonos para tener éxito en lo que sea que nos propongamos. Me encanta que seamos un equipo (y ahora un equipo de doctores). Gracias por compartir todo esto conmigo.

¡Gracias totales!

About the author

Alberto was born in Mexico City in 1989. His mother was key to forge his character, fuel his interest in science and facilitate his education. At 17, supported by a scholarship, he moved to the north of Mexico to study a BSc in Engineering Physics at Tecnológico de Monterrey. During his studies he enjoyed presiding the Physics student society and completed a couple of internships in the cement industry at Holcim and CEMEX. In the latter, he had his first contact with computational research. In 2011, after graduating, he worked as a planning and forecasting analyst at Ternium, the largest steel producer in Latin America. In 2014, he won two scholarships and moved to Germany to study a MSc in Computational Science and Engineering at Technische Universität München. He focused on Physics and Statistics, particularly applied to molecular simulation. As a research assistant in the group of Prof. Martin Zacharias, he worked for the first time in enhanced sampling methods for oligopeptides. His MSc thesis was done at Prof. Karsten Reuter's group, under the supervision of Dr. Harald Oberhofer and Dr. Thomas Stecher, working on adaptive multiscale simulations and parametrization of reaction coordinates. In 2017, he obtained another scholarship to pursue his PhD at the Universiteit van Amsterdam in the group of Dr. Bernd Ensing. The results of this period are presented hereby. During these career-motivated relocations, Alberto has been fortunate enough to solve the *two-body problem* together with his wife, who is a Dr. in Seismology. Since 2021, Alberto has joined Prof. Marjolein Dijkstra's group at Universiteit Utrecht, where he is expanding his research expertise into colloidal scales, inverse-design methods and machine learning. Free-energy calculations are still part of his day to day.