

Dipartimento di / Department of

Biotecnologie e Bioscienze

Dottorato di Ricerca in / PhD program Tecnologie convergenti per i sistemi biomolecolari (TeCSBi)

> Ciclo / Cycle XXXVII

Thermodynamic Behavior of Complex Molecular Systems: From Redox Potential Estimation to the Functionality of Ion Channels and Molecular Machines

Cognome / Surname: Silvestri Nome / Name: <u>Giuseppe</u> Matricola / Registration number: <u>861791</u>

Tutore / Tutor: Prof. De Gioia Luca

Coordinatore / Coordinator: Prof. Galimberti Andrea

ANNO ACCADEMICO / ACADEMIC YEAR 2024/2025

i

Summary

A molecular biologist seeking to understand how a protein or other biomolecules works, faces a challenge similar to observing molecules in action. Having an atomic-level structure is extremely useful and provides valuable insights into how the biomolecule works; however, since atoms in a biomolecule are in constant motion, molecular dynamics and intermolecular interactions are also critical. Ideally, one would not only want a static "snapshot" of the molecule, but also the ability to observe how it moves and responds to atomic-level perturbations. However, it is difficult to directly observe the atomic motion and manipulate it in a targeted manner. An attractive alternative is the use of computer simulations to model biomolecules at an atomic level. Molecular dynamics (MD) simulations predict the movement of every atom in a molecular system based on physical models of interatomic interactions. These simulations allow us to observe key biomolecular processes, such as conformational changes, ligand binding, and protein folding, with temporal resolution on the femtosecond scale. Furthermore, they enable us to predict how biomolecules respond to perturbations, such as mutations, phosphorylation, or the addition/removal of ligands. These simulations are often combined with experimental structural biology techniques, including X-ray crystallography, cryo-electron microscopy (cryo-EM), nuclear magnetic resonance (NMR), and Förster resonance energy transfer (FRET).

Although MD simulations are not new, with the first models of simple gases emerging in the 1950s, their use has increased significantly in recent years. Two main factors have contributed to the growing prominence of MD simulations. First, there has been an expansion of experimental structures for certain classes of molecules, especially membrane proteins. Recent advances in crystallography and cryo-EM have made it possible to determine the structures of many of these proteins, providing ideal starting points for MD simulations. Second, MD simulations have become much more powerful and accessible, thanks to advances in hardware, such as the use of GPUs, which allow for complex simulations at reduced costs, and improvements in software, which have made these tools easier to use even for non-experts.

The increasing accessibility and accuracy of MD simulations have opened new avenues for studying complex molecular systems, providing deeper insights into their dynamic behavior and interactions. These advancements have made it possible to apply computational techniques to a wide range of biomolecular processes, from fundamental research to practical applications, such as drug design and biomolecular engineering. By integrating MD simulations with experimental data, researchers can achieve a more comprehensive understanding of molecular function, predict responses to perturbations, and even design molecules with tailored properties. Building on these developments, this thesis will explore how advanced simulation techniques, such as free energy estimation through well-tempered metadynamics (WT-MTD) and non-equilibrium thermodynamic integration (NE-TI), can be applied to study complex molecular systems. Three main projects will guide our investigation: the estimation of redox potential in flavodoxins, the exploration of rotaxanes as molecular machines, and the structural and functional analysis of the Fluc-Ec2 ion channel. Each project demonstrates the power of computational approaches in addressing key questions about molecular behavior, stability, and function. In the first project, we focused on estimating the redox potential of flavodoxins, a class of enzymes containing the flavin mononucleotide (FMN) cofactor. Flavodoxins play a crucial role in electron transfer and energy conversion processes, and their redox properties are highly dependent on the surrounding molecular environment. We used the NE-TI method to accurately estimate the redox potential of wild-type flavodoxin from *Clostridium Beijerinckii* and eight single-point of its mutants. This approach has shown remarkable accuracy, with most calculated values deviating less than 1 kcal/mol from experimental data. These results suggest that the NE-TI method is a promising tool for the rational design of flavoproteins with optimized redox properties.

In the second project we explored the dynamics and stability of rotaxanes, molecular machines capable of harnessing chemical reactions to produce unidirectional motion. Rotaxanes consist of a ring (macrocycle) that shuttles along an axle, stabilized by bulky terminal groups, and this shuttling mechanism is highly influenced by the surrounding

solvent environment. Through the study of rotaxane shuttling in solvents such as DMSO, ACN, and CHCl₃, we revealed how solvent polarity and hydrogen bonding capacity play critical roles in modulating the energy landscape and dynamic behavior of these molecular machines. In highly polar solvents like DMSO and ACN, solvent interactions act as lubricants, lowering the free-energy barrier for the macrocycle's translocation between binding stations, leading to faster shuttling due to greater flexibility in both the axle and the macrocycle. Conversely, in low-polarity solvents like CHCl₃, the distal binding station becomes more stabilized, resulting in a higher free-energy barrier and slower shuttling, primarily due to stronger hydrogen bonding at the carboxylate group. Given the complexity of the rotaxane system and the limitations of unbiased molecular dynamics in sampling the full phase space, we employed well-tempered metadynamics to perform free energy calculations. This advanced technique allows for more efficient exploration of the energy landscape and provides a deeper understanding of the molecular mechanics involved. The ability of rotaxanes to function as "information ratchets", using chemical energy to drive directional movement, offers potential for designing advanced molecular devices. By leveraging solvent effects and WT-MTD, this project aims to deepen the behavior of rotaxanes as molecular machines, providing insight for performance opimization and future application in nanotechnology and molecular-scale devices.

Finally, the third project focused on the *Fluc-Ec2* ion channel, specialized in fluoride export and critical for microbial resistance to environmental fluorides. The aim is to understand how the protonation of key residues, such as His106 and Glu86, affects the channel's conformation and fluoride binding. Through unbiased and restrained MD simulations, we examined how different protonation states alter fluoride permeation, hypothesizing that His106 protonation favors a functional open state of the channel.

This thesis demonstrated how integrating various MD simulation techniques, including free energy estimation, can provide deeper insight into the energetic stability and dynamics of complex molecular systems, in order to potentially optimize proteins and molecular machines for their application in biotechnology and biomolecular device development.

Riassunto

Un biologo molecolare che desidera comprendere il funzionamento di una proteina o di un'altra biomolecola si trova di fronte a una sfida simile all'osservazione diretta delle molecole in azione. Avere una struttura a livello atomico è estremamente utile e fornisce informazioni preziose su come la biomolecola funzioni; tuttavia, poiché gli atomi in una biomolecola sono in costante movimento, anche le dinamiche molecolari e le interazioni intermolecolari risultano critiche. Idealmente, non si vorrebbe solo una "istantanea" statica della molecola, ma anche la possibilità di osservare come essa si muova e risponda a perturbazioni. Tuttavia, è difficile osservare direttamente il movimento atomico e manipolarlo in modo mirato. Un'alternativa interessante è l'uso di simulazioni al computer per modellare le biomolecole a livello atomico. Le simulazioni di dinamica molecolare (MD) predicono il movimento di ogni atomo in un sistema molecolare basandosi su modelli fisici delle interazioni interatomiche. Queste simulazioni ci permettono di osservare processi biomolecolari chiave, come i cambiamenti conformazionali, il legame con ligandi e il ripiegamento delle proteine, con una risoluzione temporale dell'ordine dei femtosecondi. Inoltre, ci consentono di prevedere come le biomolecole rispondano a perturbazioni, come mutazioni, fosforilazione o l'aggiunta/rimozione di ligandi. Spesso, queste simulazioni sono combinate con tecniche sperimentali di biologia strutturale, tra cui la cristallografia a raggi X, la microscopia crioelettronica (crio-EM), la risonanza magnetica nucleare (NMR) e il trasferimento di energia per risonanza di Förster (FRET). Sebbene le simulazioni MD non siano nuove, con i primi modelli di gas semplici risalenti agli anni '50, il loro uso è aumentato significativamente negli ultimi anni. Due fattori principali hanno contribuito alla crescente diffusione delle simulazioni MD. Primo, c'è stata un'espansione delle strutture sperimentali per alcune classi di molecole, in particolare le proteine di membrana. I recenti progressi nella cristallografia e nella crio-EM hanno reso possibile determinare le strutture di molte di queste proteine, fornendo punti di partenza ideali per le simulazioni MD. Secondo, le simulazioni MD sono

diventate molto più potenti e accessibili grazie ai progressi nell'hardware, come l'uso delle GPU, che permettono simulazioni complesse a costi ridotti, e ai miglioramenti nel software, che hanno reso questi strumenti più facili da usare anche per i non esperti.

L'aumentata accessibilità e accuratezza delle simulazioni MD ha aperto nuove strade per lo studio di sistemi molecolari complessi, fornendo una comprensione più profonda del loro comportamento dinamico e delle loro interazioni. Questi progressi hanno reso possibile l'applicazione di tecniche computazionali a una vasta gamma di processi biomolecolari, dalla ricerca fondamentale alle applicazioni pratiche, come la progettazione di farmaci e l'ingegneria biomolecolare. Integrando le simulazioni MD con dati sperimentali, i ricercatori possono ottenere una comprensione più completa del funzionamento molecolare, prevedere le risposte a perturbazioni e persino progettare molecole con proprietà specifiche. Sulla base di questi sviluppi, questa tesi esplorerà come tecniche di simulazione avanzate, come la stima dell'energia libera tramite Well-Tempered Metadynamics (WT-MTD) e l'Integrazione Termodinamica di Non-Equilibrio (NE-TI), possano essere applicate per studiare sistemi molecolari complessi. Tre progetti principali guideranno la nostra indagine: la stima del potenziale redox in flavodossine, l'esplorazione dei rotaxani come macchine molecolari, e l'analisi strutturale e funzionale del canale ionico Fluc-Ec2. Ogni progetto dimostra la potenza degli approcci computazionali nel rispondere a domande chiave sul comportamento molecolare, la stabilità e la funzione.

Nel primo progetto, ci siamo concentrati sulla stima del potenziale redox in flavodossine, una classe di enzimi che contiene il cofattore flavin mononucleotide (FMN). Le flavodossine svolgono un ruolo cruciale nei processi di trasferimento elettronico e conversione energetica, e le loro proprietà redox dipendono fortemente dall'ambiente molecolare circostante. Abbiamo utilizzato il metodo NE-TI per stimare accuratamente il potenziale redox della flavodossina wild-type di *Clostridium Beijerinckii* e di otto dei suoi mutanti. Questo approccio ha dimostrato una notevole accuratezza, con la maggior parte dei valori calcolati che deviano meno di 1 kcal/mol dai dati sperimentali. Questi risultati suggeriscono che il metodo NE-TI è uno strumento promettente per la progettazione razionale di flavoproteine con proprietà redox ottimizzate.

Nel secondo progetto, abbiamo esplorato la dinamica e la stabilità dei rotaxani, macchine molecolari capaci di sfruttare reazioni chimiche per produrre un movimento unidirezionale. I rotaxani consistono in un anello (macrociclo) che scorre lungo un asse, stabilizzato da gruppi terminali ingombranti, e questo meccanismo di scorrimento è fortemente influenzato dall'ambiente circostante. Attraverso lo studio del movimento del rotaxano in solventi come DMSO, ACN e CHCl₃, abbiamo rivelato come la polarità del solvente e la capacità di formare legami a idrogeno giochino ruoli critici nel modulare il paesaggio energetico e il comportamento dinamico di queste macchine molecolari. In solventi altamente polari come DMSO e ACN, le interazioni del solvente agiscono come lubrificanti, abbassando la barriera di energia libera per la traslocazione del macrociclo tra le stazioni di legame, portando a uno scorrimento più veloce grazie alla maggiore flessibilità sia dell'asse che dell'anello stesso. Al contrario, in solventi a bassa polarità come CHCl₃, la stazione di legame distale lega più stabilmente il macrociclo, risultando in una barriera energetica più alta e uno scorrimento più lento, principalmente a causa di una geometria di legame più favorevole con il gruppo carbossilato catalitico dell'asse. Data la complessità del sistema e le limitazioni della dinamica molecolare in termini di campionamento dello spazio delle fasi, abbiamo impiegato la WT-MTD per eseguire calcoli di energia libera. Questa tecnica avanzata consente un'esplorazione più efficiente del paesaggio energetico e fornisce una comprensione più approfondita della meccanica molecolare coinvolta. La capacità dei rotaxani di funzionare come ratchet informativi, utilizzando l'energia chimica per guidare il movimento direzionale, apre le porte per la progettazione di dispositivi molecolari avanzati, e lavori come quello appena descritto fornisce spunti per la loro ottimizzazione.

Infine, il terzo progetto si è concentrato sul canale ionico *Fluc-Ec2*, specializzato nell'esportazione del fluoro e critico per la resistenza microbica ai fluoruri ambientali. L'obiettivo è comprendere come la protonazione di residui chiave, come His106 ed Glu86, influenzi la conformazione del canale e il legame con il fluoro. Attraverso simulazioni di dinamica molecolare unbiased e non, abbiamo esaminato come diversi stati di protonazione influenzino la solvatazione dei pori del canale e il legame del fluoro nel

sito F2, suggerendo che la protonazione di His106 favorisca uno stato funzionale aperto del canale.

Questa tesi ha dimostrato come l'integrazione di varie tecniche di simulazione MD, inclusa la stima dell'energia libera, possa fornire una comprensione più profonda della stabilità energetica e delle dinamiche di sistemi molecolari complessi, al fine di ottimizzare potenzialmente proteine e macchine molecolari per la loro applicazione in biotecnologia e nello sviluppo di dispositivi biomolecolari.

Contents

Summary	iii
Riassunto	v
Introduction to Computational Approaches	1
1.1 Basic principles	2
1.2 Force Field	6
1.3 Free Energy Estimation	10
1.3.1 Definition of Free Energy	10
1.3.2 Non-Equilibrium Thermodynamic Integration	14
1.3.3 Well-Tempered Metadynamics	19
Bibliography	23
Assessing the Performance of NE-TI Method in	1
Flavodoxin's Redox Potential Estimation	28
2.1 Abstract	29
2.2 Introduction	30
2.3 Materials and methods	33

2.3.1 Simulated systems	33
2.3.2 Equilibrium MD setup	35
2.3.3 NE-TI setup	36
2.3 Results and discussion	37
2.4 Conclusions	49
Supplementary figures and tables	50
Bibliography	56

Computational Characterization of the Ring Shuttling Process in Amide-based [2]Rotaxanes: the Impact of the Solvent 66

3.1 Abstract	67
3.2 Introduction	68
3.3 Materials and methods	75
3.3.1 Simulation setup	75
3.3.2 WT-MTD parameters	76
3.4 Results and discussion	79
3.4.1 Dimethylsulphoxide	80
3.4.2 Acetonitrile	82
3.4.3 Trichloromethane	84
3.4.4 Effects on the shuttling mechanism	87
3.5 Conclusions	91

Exploring the Effect of Protonation Events at the SoutSite in the Fluc Ec2 Channel: Structural Insights fromMD Simulations102

4.1 Abstract	103
4.2 Introduction	105
4.3 Materials and methods	109
4.4 Results and discussion	112
4.4.1 Structural features by each system	113
4.4.2 Loop 1 dynamics are driven by electrostatic effects	
4.4.3 His106 protonation enhances F ⁻ retention in S _{out} site	120
4.5 Conclusions	
Supplementary figures and tables	125
Bibliography	137
Final Remarks	142
List of Publications	145

97

Chapter 1

Introduction to Computational Approaches

1.1 Basic principles

Biomolecular dynamics involves complex interactions occurring over various time and spatial scales, making the approach to their study highly dependent on the research question posed. Generally, experimental techniques like spectroscopy, which analyzes bond vibrations, or electrophysiology, which observes ion channel behavior, are preferred for studying these dynamics. Nevertheless, advancements in theoretical methods have notably enhanced our understanding of biomolecular systems. Today, computational modeling and simulations have become invaluable tools, often providing deeper insights that would be difficult to achieve through experimental work alone.

One of the most well-known methods in theoretical studies is Molecular Dynamics. These simulations are particularly favored when examining dynamic properties of molecular systems.

Historically, simulations have played a crucial role in validating theoretical models against experimental observations. For instance, ion channel simulations have clarified why certain ions penetrate membranes while others are blocked, complementing existing conductivity data from experiments. Furthermore, simulations can reveal information that may not be accessible through experimental means, such as the pressure distributions within cellular membranes. As computational techniques advance, the role of simulations has evolved from merely confirming experimental results to making predictions about molecular properties, including binding affinities and folding dynamics. The expectation is that this trend will continue to grow as computational resources expand.

From the perspective of theoretical physics, the time-dependent Schrödinger equation theoretically possesses the capability to accurately predict all molecular properties. However, when dealing with systems featuring multiple particles, approximations become necessary. In the realm of quantum chemistry, a widely used approximation is to assume static atomic nuclei and deploy an implicit solvent model. While these methodologies are effective, QM methods are computationally demanding, therefore their applicability is limited both in term of accessible spatial scales (i.e. size of the system usually limited to hundreds of atoms) and time scales (tenths of ps).

For most biomolecular systems, empirical, parameterized models are favored. These often involve classical Coulomb interactions between point-like atomic charges. The primary distinction between quantum chemistry and classical molecular dynamics lies in their focus areas; quantum chemistry excels in defining electronic structures and potential energy, while classical molecular dynamics is superior for sampling the extensive array of states a macromolecule might adopt. The latter captures the entropic contributions to free energy, making it faster and more aligned with empirical observations on the microsecond timescale compared to quantum models (Fig. 1.1).

The pioneering work of molecular dynamics simulations began in 1957 [2] but it was not until the 1970s that biomolecule simulations became feasible [23, 26]. It is important to note that macroscopic properties derived from experiments are averages across billions of molecules, representing a statistical mechanics ensemble rather than direct observations. This understanding introduces both theoretical challenges and practical considerations that researchers must navigate to obtain meaningful insights [3, 13].

One of the primary challenges lies in ensemble generation. Researchers cannot depend solely on the examination of individual molecular structures since these may not adequately represent the broader behavior of a system under specific conditions. Instead, they must develop a comprehensive analysis that generates a representative ensemble of structures. This ensemble must be reflective of key state variables, such as temperature and pressure, which play significant roles in determining the macroscopic properties of the system.

In addition to ensemble generation, researchers face challenges regarding the calculation of thermodynamic properties, particularly those related to free energy. Important equilibrium properties like binding constants and solubilities cannot be derived from unbiased MD simulations alone when the phenomena involves high energy barriers. Rather, they require more advanced techniques that account for entropic factors and the complexities of molecular interactions. These methods allow for a more accurate illustration of how molecules behave under varying conditions, ultimately leading to insights into their equilibria and behaviors.

Furthermore, there is a notable shift in focus from studying individual molecular trajectories to analyzing structural ensembles when it comes to equilibrium properties. By emphasizing ensemble analysis, researchers can capture the behavior of molecules, which is essential for a comprehensive understanding of the system as a whole. This approach facilitates the identification of trends and relationships that might be obscured when examining only single molecular paths.

By prioritizing ensemble generation, employing advanced methodologies for calculating thermodynamic properties, and focusing on the collective dynamics of molecular structures, researchers can attain a more nuanced grasp of the intricate behaviors that govern ensemble systems. This holistic approach is essential for advancing the field and fostering a deeper understanding of molecular interactions in various conditions.



Figure 1.1: Experimental and simulation techniques used to define protein structure and dynamics at different time and length-scales. ("MD" – molecular dynamics; "AFM" – atomic force microscopy; "SAXS" – small angle X-ray scattering; "SANS" – small angle neutron scattering; "EM" – electron microscopy; "NMR" – nuclear magnetic resonance spectroscopy; "smFRET" – single molecule fluorescence resonance energy transfer) [15].

1.2 Force Field

All classical simulation methods rely on empirical parameter sets known as *Force Fields* (FF) to calculate interactions and evaluate the potential energy of a system based on atomic coordinates [17, 21, 25, 29]. The force field refers to the functional form and parameter sets used to calculate the potential energy of a system as a function of the particles coordinates. Knowing the FF, the forces acting on each particle due to the rest of the system can be calculated (derivative of the FF).

Although these approximations are highly effective for most applications, they cannot accurately capture quantum phenomena such as bond formation or breaking. Common force fields divide potential energy functions into two categories: *bonded interactions* and *non-bonded interactions*. Bonded interactions encompass the stretching of covalent bonds, bending of angles, torsion potentials around bonds, and improper torsion potentials, all of which typically fluctuate around equilibrium values. Non-bonded interactions, on the other hand, involve not directly bonded atoms. These include *Lennard-Jones forces* (repulsion and dispersion) and *Coulomb electrostatic interactions*, which are usually calculated using neighbor lists that are updated periodically (Fig. 1.2).



Figure 1.2: Illustration of the fundamental force field energy terms.

Upon determination of the forces on all atoms, their coordinates are updated for progression to the subsequent step in the simulation, according with classical motion eqations. Before running the dynamics, the system is subjected to an energy minimization step: the steepest descent algorithm is employed to slightly adjust each atom's position towards lower energy, given that force equates to the negative gradient of energy. In molecular dynamics, the coordinates are updated by integrating Newton's equations of motion and so the system's potential energy too, as illustrated in the flowchart of Fig. 1.3.



Figure 1.3: A typical molecular dynamics (MD) simulation follows a simplified flow where structures are generated from a natural ensemble by calculating potential energy functions and integrating Newton's equations of motion. These structures are then analyzed to evaluate the system's equilibrium properties. The time step in an MD simulation is usually around 1 or 2 femtoseconds, unless specialized techniques are applied to extend it further [20].

However, the integration of this differential equation is challenging due to the many-body nature of the system. Therefore, it is discretized and solved numerically. The goal of this numerical integration is to determine the positions $\vec{r_i}(t + \Delta t)$ at time $t + \Delta t$, based on the known positions at time t. This is achieved, for instance, using the Verlet algorithm:

$$\vec{r}_i(t + \Delta t) \approx 2\vec{r}_i(t) - \vec{r}_i(t - \Delta t) + \frac{F_i(t)}{m_i}\Delta t^2$$

For the integration to be stable, Δt must be smaller than the timescale of the system's fastest motions. In biological macromolecules, these fast motions typically involve the vibrations of light atoms (e.g., the O-H bond), which occur over a few femtoseconds. As a result, Δt is typically set to a sub-femtosecond scale. Enhancing performance may involve extending the time step; however, bond vibrations can introduce inaccuracies even at intervals as short as 1 femtosecond. In the majority of simulations, the emphasis is not on bond vibrations, which can be removed using bonding constraint algorithms [14, 27]. These constraints facilitate extending the time step to 2 femtoseconds and fixed-length bonds often yield superior approximations of quantum mechanical oscillators compared to harmonic springs.

The initial velocities of the particles are generally chosen randomly from a Maxwell-Boltzmann distribution at a given temperature:

$$f(v) = \sqrt{\left(\frac{m}{2\pi k_{BT}}\right)^3} \cdot e^{-\frac{mv^2}{2k_BT}}$$

where k_B is the Boltzmann constant, *T* is the temperature, *m* is the particle mass, and *N* is the number of particles. The particle velocity *v* is selected accordingly to the system's thermal conditions.

Given that even the smallest chemical entities are excessively large for complete simulation, biomolecular simulations frequently utilize *periodic boundary conditions* to

mitigate artifacts at the simulation box's periphery. This methodology ensures that, for instance, a water molecule exiting the system on one side reemerges on the opposing side. If the simulation box possesses adequate dimensions, molecular interactions with their periodic images are negligible. This is particularly crucial for non-bonded interactions, ideally summed over all neighboring atoms in an infinite periodic system. Simple cutoffs are feasible for Lennard-Jones interactions, which decline swiftly, while a cutoff applied to Coulomb interactions may incur substantial errors. At present, the *Particle-Mesh-Ewald* (PME) summation is the preferred technique for calculating infinite electrostatic interactions, where calculations are partitioned into short- and long-range components. As the structure's potential energy diminishes during simulation, an increase in kinetic energy (and hence temperature) is observed should the system's total energy remain fixed. To mimic experimental conditions (i.e. constant T), the system is often linked to a thermostat, calibrating velocities during the simulation to preserve a stable temperature, generally around room temperature. Likewise, the system's pressure can be modulated by

adjusting the simulation box's dimensions, either uniformly in all directions or independently along the x, y, and z axes.

1.3 Free Energy Estimation

1.3.1 Definition of Free Energy

The spontaneity of a process is dictated by its free energy, as systems naturally progress toward their lowest free energy state when no external work is applied. The speed at which this state is achieved is influenced by free energy barriers along the pathways leading to that minimum. Consequently, free energies are crucial as they determine phenomena such as binding affinities (whether binding occurs spontaneously) and protein folding (where the folded state typically represents the free energy minimum). Furthermore, as barriers are tied to rates through rate theory, free energy barriers impact the kinetics of binding, folding, permeation, and reactions. Thus, accurately deriving free energies through computational techniques is vital, not only due to their fundamental importance but also because they can be directly compared quantitatively with experimental data.

Thermodynamically, a distinction exists between the Helmholtz and Gibbs free energies. The Helmholtz free energy, denoted as F, is given by:

$$F = U - TS$$

where U is the internal energy, T is the temperature, and S is the entropy of the system. The Gibbs free energy, denoted as G, is expressed as:

$$G = H - TS$$

where H is the enthalpy of the system, defined as H = U + pV, with p being the pressure and V the volume of the system. Both definitions include a term related to the system's internal (or potential) energy and another related to entropy. This implies that free energies, and thus affinities or stabilities, are influenced by both changes in interatomic interactions and entropic variations. Changes in solvent entropy can be particularly significant, as seen in the hydrophobic effect, where solvent molecules are released during the formation of a hydrophobic protein or membrane core.

The difference between Helmholtz and Gibbs free energy lies in the pV term. Helmholtz free energy is applicable at constant volume, whereas Gibbs free energy is used at constant pressure, with the pV term quantifying the work related to volume changes. Under physiological and typical experimental conditions, isobaric conditions are common, making the Gibbs free energy gradient a driving force for a system.

In equilibrium, the free energy represents the maximum amount of energy obtainable from a spontaneous process at constant temperature, or conversely, the minimum energy required to drive an uphill process. Free energy is the energy acquired from (or needed for) a process slow enough to maintain equilibrium with its surroundings. If the process is forced to take place faster, friction causes non-equilibrium work values to exceed the associated free energy change, with the excess energy dissipated as heat. Traditional free energy methods thus require slow transitions to consider systems in equilibrium. However, non-equilibrium work distributions can also be used to derive free energies using the Jarzynski equality and the Crooks Fluctuation Theorem (CFT).

In statistical mechanics, free energy plays a role in determining the probability of being in state x, which is directly related to its free energy:

$$p(x) \propto e^{-G(x)/k_BT}$$

where k_B is the Boltzmann constant. The term $e^{-G(x)/k_BT}$ is called the Boltzmann factor. This relation is useful for estimating free energy differences between states A and B:

$$\frac{p(A)}{p(B)} = e^{-(G(A) - G(B))/k_BT} = e^{-\Delta G/k_BT}$$

This provides a straightforward approach to estimate free energies from simulations, assuming a sufficiently converged ensemble with several reversible transitions between A and B. The free energy difference ΔG can be directly estimated by evaluating the populations in states A and B.

Alternative simulation methods for deriving free energies often focus on overcoming or avoiding free energy barriers related to binding, unbinding, conformational changes, or (re)folding. Techniques include umbrella sampling (US) [26] and thermodynamic integration (TI) [17]. In umbrella sampling, a biasing potential enforces transitions across high-energy states along a predefined reaction coordinate (a chemical process descriptor), with the biasing potential's effect corrected afterward to yield the free energy profile or landscape. Thermodynamic integration and other alchemical methods provides the overall free energy difference for the process and not the chemical mechanism itself, by transforming the Hamiltonian from system A to B. Since free energy is a state variable, the change in relative free energy can be approximated by the difference in transition free energy, such as that occurring in a protein binding pocket compared to the same transformation in solution, which is useful for calculating relative binding free energy. A notable characteristic of free energy is that its estimate often converges faster than the individual enthalpy and entropy contributions. This occurs because the largest contributions to both arise from a solvent-solvent term present in both but exactly canceling out in the free energy estimate [6].

The free energy landscape of a system fully encapsulates its thermodynamic and kinetic characteristics. An intuitive understanding of free energy is provided by its relationship to the probability p_A of the system being located in a particular phase space volume A:

$$p_A = e^{-\beta(F_A - F)}$$

where F_A is the free energy associated with phase space volume A, F is the Helmholtz free energy over the entire phase space, and $\beta = 1/k_BT$, with k_B as the Boltzmann constant and T as the temperature. In practice, the interest often lies in the differences between state populations rather than the absolute free energy values. For two states, *A* and *B*, the free energy difference ΔF_{AB} is expressed as:

$$\Delta F_{AB} = F_B - F_A = -\frac{1}{\beta} ln \left(\frac{p_B}{p_A}\right) = \frac{1}{\beta} ln \left(\frac{Q_B}{Q_A}\right)$$

Q represents the canonical partition function: Q = Q(N, V, T), with *N* as the number of particles and *V* as the volume. The free energy is linked to the partition function via $F = -\frac{1}{\beta} ln \ Q(N, V, T)$. From this expression, various methods for determining ΔF can be derived. A straightforward approach involves directly counting events that sample phase space volumes *A* and *B*, thereby obtaining probabilities p_A and p_B . For instance, in molecular binding scenarios, this method can be applied by simulating two molecules of interest (e.g., ligands *A* and *B*) and counting their respective probabilities of being in bound/unbound states. Although simple, this method is computationally intensive for large biologically relevant systems, exceeding the capabilities of current state-of-the-art atomistic simulations.

An alternative approach involves evaluating the partition functions. While the partition function has an analytical form for an ideal gas, for particles with complex interactions, numerical integration over coordinates and momenta is necessary. A canonical partition function is defined as:

$$Q(N,P,T) = \frac{1}{h^{3N}N!} \int \dots \int e^{-\beta \left(H\left(p_1 \dots p_{N}, q_1 \dots q_N\right) + PV\right)} dV dp_1 \dots dp_N dq_1 \dots dq_N$$

where H(p,q) is the Hamiltonian of the system, q and p are coordinates and momenta, respectively, and h is Planck's constant. For multi-particle systems, integrating over all degrees of freedom is computationally prohibitive, so simulations often sample the accessible phase space volume. In simulations, inaccessible phase space regions are omitted for both states *A* and *B*, allowing an accurate assessment of ΔF due to error cancellation.

Considering that a system's Hamiltonian comprises kinetic and potential energy components H(p,q) = K(p) + U(q), the partition function can be split into kinetic and configurational parts. Since K(p) depends solely on particle momenta, the kinetic partition function remains unchanged unless there's a mass change between states. Thus, often only the configurational partition function is considered.

1.3.2 Non-Equilibrium Thermodynamic Integration

Thermodynamic integration is a method used to calculate the free energy difference between two states by integrating the average force applied to the system as a function of a coupling parameter, known as the λ variable, during a transition:

$$\Delta G_{AB} = \int_0^1 \left\langle \frac{\partial H}{\partial \lambda} \right\rangle_{\lambda} d\lambda$$

This λ parameter controls the progression of the system from one state to another in an *alchemical* transition, which refers to a process where interactions between particles are gradually turned on or off, or where particles are morphed into others. Alchemical methods calculate the work required to transition a system from one state to another via non-physical pathways. Since the free energy of a system depends solely on its state, the free energy difference between two states is independent of the specific path taken during the transition. Adequate sampling is crucial for all free energy methods, and achieving convergence becomes significantly more challenging as the size of the perturbation increases. These transitions are not physical processes but hypothetical transformations

that allow the calculation of free energy differences between states (e.g., different molecular species or ligand-bound vs. unbound states).

In *slow growth* thermodynamic integration, the transition between states occurs very slowly as the λ variable is changed continuously over time, keeping the system near equilibrium. In this scenario, the free energy change is equated to the work done during the transition. However, because the system is only maintained in a quasi-equilibrium state due to the limitations in sampling, some work dissipates, causing inaccuracies in the free energy estimates and leading to known convergence issues.

Discrete thermodynamic integration offers a solution to these issues by dividing the λ path into a series of discrete steps, similar to methods like Free Energy Perturbation (FEP) and Bennett Acceptance Ratio (BAR). At each discrete state λ_i , an equilibrium simulation is performed to calculate the average force $\left\langle \frac{\partial H}{\partial \lambda} \right\rangle_{\lambda_i}$ at that point. The free energy

difference, ΔG , is then obtained by numerically integrating these average forces along the entire λ path.

In *fast growth* thermodynamic integration, unlike slow growth, the transition between states occurs rapidly, driving the system far from equilibrium. In this approach, the work done during the transition is no longer equal to the free energy difference due to the significant dissipation of energy, as the system does not have sufficient time to equilibrate between successive changes in the λ variable. However, multiple fast growth simulations can be averaged using Jarzynski's equality [11, 14] to estimate the free energy difference:

$$e^{-\beta \Delta G_{AB}} = \langle e^{-\beta W} \rangle$$

This method can provide accurate results but requires a large number of fast growth simulations to compensate for the lack of equilibrium, as each individual run may exhibit large deviations. The accuracy of free energy differences calculated using Jarzynski's formula depends on rare events in which little work is dissipated. Fast transitions that

drive a system far from equilibrium tend to yield large work values, contributing minimally to the exponential average, and thus slowing the convergence of the ΔG estimation.

Several free energy estimators have been developed based on the Jarzynski equality. Among them, a more general relation was derived by Crooks, allowing the combination of work value distributions from forward and backward transitions to obtain the Gibbs free energy difference, known as the Crooks Fluctuation Theorem (CFT) [8, 10]:

$$\frac{P_f(W)}{P_r(-W)} = e^{\beta(W - \Delta G)}$$

Where $P_f(W)$ and $P_r(-W)$ correspond to the work distributions during forward and reverse processes, respectively. Work values are calculated by integrating the Hamiltonians (H_A and H_B) of the system with respect to a coupling parameter (λ) along the transition path:

$$W = \int_0^1 \frac{\partial H}{\partial \lambda} \partial \lambda$$

The coupling parameter λ changes from 0 to 1 continuously and quickly to transform the system from state A to state B in order to obtain the forward work (*W*). The same happens to calculate the reverse work (–*W*), changing λ from 1 to 0. Both W and –*W* are calculated as an average over an ensemble of N trajectories that are started from an equilibrated canonical ensemble for both states A and B. In other words, a lot of very fast transitions (hundreds of transitions of a few tens of ns) are performed from A to B and vice versa, with a corresponding great saving of computational time.

This method is straightforward, but requires sufficient overlap between work histograms, typically only achieved in near-equilibrium transitions. Only work values from the overlap region contribute to the free energy estimate. To address these challenges, work histograms can be approximated by an analytical distribution through the Crooks Gaussian Intersection (CGI) estimator. The intersection point of two distributions corresponding to forward and reverse transitions indicates a work value equal to ΔG (Fig. 1.4), where the statistical error can be calculated by means of bootstrapping. As the estimator depends on a Gaussian approximation, the validity of this assumption can be assessed using a statistical test, e.g. Kolmogorov–Smirnov (KS) [22]. The BAR estimator can also be applied to non-equilibrium simulations, providing a maximum likelihood estimator for the free energy difference.



Figure 1.4: Non-equilibrium free energy calculations involve generating two equilibrium ensembles for the system in state A and state B, typically through MD simulations. From these ensembles, snapshots are selected as starting points for short non-equilibrium transition simulations, where the system is progressively transformed from state A to state B, and vice versa [20].

1.3.3 Well-Tempered Metadynamics

Although MD has become invaluable for studying complex systems and making predictions, they often face a significant timescale problem: atomistic simulations require a time step on the order of femtoseconds (10^{-15} seconds) to accurately capture fast motions like bond stretching and bending, but many important biological processes, such as protein folding, occur over much longer timescales. This mismatch between the timescale of the simulation and the actual biological process limits MD's ability to provide complete insights, as the system may not explore all relevant configurations. High energy barriers or slow diffusion in configuration space can further complicate the situation, making it difficult for the system to transition between metastable states and achieve ergodicity within practical simulation times. To address this issue, enhanced sampling techniques have been developed in recent years [1, 24, 30].

One such method is metadynamics (MTD), which accelerates sampling by applying an additional history-dependent bias potential to a few selected degrees of freedom, known as collective variables (CVs). Metadynamics allows the system to overcome free energy barriers more easily and reconstructs the free-energy surface (FES) as a function of these CVs. The bias potential is constructed by adding Gaussians along the system's trajectory in CV space, preventing the system from revisiting previously sampled configurations. The bias potential gradually grows, enabling the system to move between different energy minima, and the FES is reconstructed as the simulation progresses.

Despite its benefits, metadynamics has some limitations, including the fact that the bias potential does not converge precisely to the free energy in a single run and may overfill the energy landscape. Additionally, choosing appropriate CVs to describe complex processes can be challenging.

To address overfilling issues, well-tempered metadynamics was introduced, where the bias deposition rate decreases over time, allowing the simulation to approach thermodynamic equilibrium more smoothly. In MTD, CVs are functions S of the microscopic coordinates R of the system:

$$S(R) = \left(S_1(R), \dots, S_d(R)\right)$$

which are able to provide a coarse-grained description of the process under study. In particular, CVs must distinguish the relevant states of the system and include all the kinetically relevant degrees of freedom. The MTD bias potential (V_G) can be written as a sum of Gaussians deposited along the system trajectory in the CVs space. In the *well-tempered* approach [5], V_G has the following functional form at time *t*:

$$V_G(S,t) = \int_0^t dt' \omega(t') \cdot exp\left(-\sum_{i=1}^d \frac{\left(S_i(R) - S_i(R(t'))\right)^2}{2\sigma_i^2}\right)$$

where σ_i is the width of the Gaussian for the *i*th CV. The time-dependent energy rate $\omega(t)$ is defined as:

$$\omega(t) = \omega_0 \cdot exp\left(-\frac{V_G(S,t)}{k_B \Delta T}\right)$$

where ω_0 is an initial deposition rate, k_B is the Boltzmann constant, and ΔT is an input parameter with the dimension of a temperature.

It has been proven that, in the long-time limit, V_G converges to:

$$V_G(S, t \to \infty) = -\frac{\Delta T}{T + \Delta T} \cdot F(S) + C$$

where *T* is the temperature of the system, *C* is an irrelevant additive constant and *F*(*S*) is the free energy as a function of the CVs. The last equation is often expressed in terms of the so-called *bias-factor* $\gamma = (T + \Delta T) / T$ as:

$$V_G(S, t \to \infty) \approx -(1 - \gamma^{-1}) \cdot F(S)$$

In the long-time limit, the CVs probability density P(S,t) can be written as:

$$P(S, t \to \infty) \propto exp\left(-\frac{F(S)}{k_B(T + \Delta T)}\right)$$

which corresponds to sampling the FES at the fictitious temperature of $T + \Delta T$. The extent of FES exploration can thus be regulated by tuning ΔT .

To understand the effect of V_G , let us consider a system whose dynamics can be captured by a one-dimensional free energy, with local minima separated by barriers much higher than thermal fluctuations. In a standard MD simulation, the system would remain trapped in one of the local minima and would not be able to sample all the relevant regions in a reasonable simulation time. In a WT-MTD simulation, Gaussians are progressively added, increasing of the bias potential which ultimately facilitates barrier crossing. This leads to exploration of previously inaccessible regions of the FES and V_G will converge. At that point, the free-energy profile can be estimated.

In WT-MTD, the bias deposition rate decreases with time as 1/t. The dynamics of all the microscopic variables thus become progressively closer to thermodynamic equilibrium as the simulation proceeds. This feature allows to easily recover the equilibrium Boltzmann distribution of degrees of freedom other than the CVs, which are typically altered by the introduction of V_G [4, 7].

In practice, WT-MTD simulations can be reweighted to recover the unbiased probability distribution for all system variables, ensuring that quantitative comparisons with experimental data can still be made, even when the bias potential is applied indirectly to the system's degrees of freedom.

To conclude, it's worth noticing that proper selection of CVs is crucial for the effectiveness of the simulation. To ensure successful application, CVs must follow these guidelines:

- 1. Distinguish between states: CVs should differentiate the initial, final, and intermediate states of the system.
- 2. Include slow modes: CVs must capture the slow dynamics of the system, as neglecting them can lead to poor sampling and inaccurate results.
- 3. Limit the number of CVs: Keeping the number of CVs low is important, as increasing the number adds complexity and makes the exploration of the system's energy landscape computationally expensive.

The choice of CVs involves reducing the system's dimensions, projecting its potential energy on a FES. For an accurate study of a process, the selected CVs must effectively describe the transition between key states. Additionally, slow variables that cannot be fully sampled in the timescale of the simulation must be included, as omitting them can prevent proper convergence to the FES.

Bibliography

- Abrams, Cameron, and Giovanni Bussi. "Enhanced Sampling in Molecular Dynamics Using Metadynamics, Replica-Exchange, and Temperature-Acceleration." *Entropy* 16, no. 1 (December 27, 2013): 163–99. https://doi.org/10.3390/e16010163.
- Alder, B. J., and T. E. Wainwright. "Phase Transition for a Hard Sphere System." *The Journal of Chemical Physics* 27, no. 5 (November 1, 1957): 1208–9. https://doi.org/10.1063/1.1743957.
- Allen, M. P., and D. J. Tildesley. *Computer Simulation of Liquids*. Oxford University Press, 1989.
- Barducci, Alessandro, Massimiliano Bonomi, and Michele Parrinello.
 "Linking Well-Tempered Metadynamics Simulations with Experiments." *Biophysical Journal* 98, no. 9 (May 2010): L44–46. https://doi.org/10.1016/j.bpj.2010.01.033.
- Barducci, Alessandro, Giovanni Bussi, and Michele Parrinello. "Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method." *Physical Review Letters* 100, no. 2 (January 18, 2008). https://doi.org/10.1103/physrevlett.100.020603.
- Ben-Naim, Arieh Y. Statistical Thermodynamics for Chemists and Biochemists. Springer Science & Business Media, 2013.
- Bonomi, M., A. Barducci, and M. Parrinello. "Reconstructing the Equilibrium Boltzmann Distribution from Well-tempered Metadynamics." *Journal of Computational Chemistry* 30, no. 11 (May 6, 2009): 1615–21. <u>https://doi.org/10.1002/jcc.21305</u>.

- Chelli, Riccardo, Simone Marsili, Alessandro Barducci, and Piero Procacci. "Recovering the Crooks Equation for Dynamical Systems in the Isothermal-Isobaric Ensemble: A Strategy Based on the Equations of Motion." *The Journal of Chemical Physics* 126, no. 4 (January 23, 2007). https://doi.org/10.1063/1.2424940.
- Chipot, Christophe, and Andrew Pohorille. Free Energy Calculations: Theory and Applications in Chemistry and Biology. Springer Science & Business Media, 2007.
- Crooks, Gavin E. "Entropy Production Fluctuation Theorem and the Nonequilibrium Work Relation for Free Energy Differences." *Physical Review E* 60, no. 3 (September 1, 1999): 2721–26. https://doi.org/10.1103/physreve.60.2721.
- Cuendet, Michel A. "The Jarzynski Identity Derived from General Hamiltonian or Non-Hamiltonian Dynamics Reproducing NVT or NPT Ensembles." *The Journal of Chemical Physics* 125, no. 14 (October 12, 2006). https://doi.org/10.1063/1.2338535.
- Dellago, Christoph, and Peter G. Bolhuis. "Transition Path Sampling and Other Advanced Simulation Techniques for Rare Events." In *Advanced Computer Simulation Approaches for Soft Matter Sciences III*, 167–233. Berlin, Heidelberg: Springer Berlin Heidelberg. Accessed October 12, 2024. http://dx.doi.org/10.1007/978-3-540-87706-6_3.
- 13. Frenkel, Daan, and Berend Smit. *Understanding Molecular Simulation: From Algorithms to Applications*. Elsevier, 2023.
- Hess, Berk, Henk Bekker, Herman J. C. Berendsen, and Johannes G. E. M. Fraaije. "LINCS: A Linear Constraint Solver for Molecular Simulations." *Journal of Computational Chemistry* 18, no. 12 (September 1997): 1463–
72. https://doi.org/10.1002/(sici)1096-987x(199709)18:12<1463::aid-jcc4>3.0.co;2-h.

- 15. Hsu, Claire C., Markus J. Buehler, and Anna Tarakanova. "The Order-Disorder Continuum: Linking Predictions of Protein Structure and Disorder through Molecular Simulation." *Scientific Reports* 10, no. 1 (February 7, 2020). https://doi.org/10.1038/s41598-020-58868-w.
- Jarzynski, C. "Nonequilibrium Equality for Free Energy Differences." *Physical Review Letters* 78, no. 14 (April 7, 1997): 2690–93. https://doi.org/10.1103/physrevlett.78.2690.
- Kaminski, George A., Richard A. Friesner, Julian Tirado-Rives, and William L. Jorgensen. "Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides." *The Journal of Physical Chemistry B* 105, no. 28 (April 19, 2001): 6474–87. https://doi.org/10.1021/jp003919d.
- Karplus, Martin, and J. Andrew McCammon. "Molecular Dynamics Simulations of Biomolecules." *Nature Structural Biology* 9, no. 9 (September 2002): 646–52. https://doi.org/10.1038/nsb0902-646.
- Kirkwood, John G. "Statistical Mechanics of Fluid Mixtures." *The Journal* of Chemical Physics 3, no. 5 (May 1, 1935): 300–313. https://doi.org/10.1063/1.1749657.
- 20. Kukol, Andreas. Molecular Modeling of Proteins. Humana Press, 2017.
- MacKerell, A. D., Jr., D. Bashford, M. Bellott, R. L. Dunbrack Jr., J. D. Evanseck, M. J. Field, S. Fischer, et al. "All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins." *The Journal of*

Physical Chemistry B 102, no. 18 (April 1, 1998): 3586–3616. https://doi.org/10.1021/jp973084f.

- Massey, Frank J., Jr. "The Kolmogorov-Smirnov Test for Goodness of Fit." Journal of the American Statistical Association 46, no. 253 (March 1951): 68–78. https://doi.org/10.1080/01621459.1951.10500769.
- McCammon, J. Andrew, Bruce R. Gelin, and Martin Karplus. "Dynamics of Folded Proteins." *Nature* 267, no. 5612 (June 1977): 585–90. https://doi.org/10.1038/267585a0.
- 24. Miao, Yinglong, and J. Andrew McCammon. "Unconstrained Enhanced Sampling for Free Energy Calculations of Biomolecules: A Review." *Molecular Simulation* 42, no. 13 (July 5, 2016): 1046–55. https://doi.org/10.1080/08927022.2015.1121541.
- 25. Oostenbrink, Chris, Alessandra Villa, Alan E. Mark, and Wilfred F. Van Gunsteren. "A Biomolecular Force Field Based on the Free Enthalpy of Hydration and Solvation: The GROMOS Force-field Parameter Sets 53A5 and 53A6." *Journal of Computational Chemistry* 25, no. 13 (July 14, 2004): 1656–76. https://doi.org/10.1002/jcc.20090.
- Rahman, Aneesur, and Frank H. Stillinger. "Molecular Dynamics Study of Liquid Water." *The Journal of Chemical Physics* 55, no. 7 (October 1, 1971): 3336–59. https://doi.org/10.1063/1.1676585.
- 27. Ryckaert, Jean-Paul, Giovanni Ciccotti, and Herman J.C Berendsen.
 "Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes." *Journal of Computational Physics* 23, no. 3 (March 1977): 327–41. https://doi.org/10.1016/0021-9991(77)90098-5.

- Torrie, G.M., and J.P. Valleau. "Nonphysical Sampling Distributions in Monte Carlo Free-Energy Estimation: Umbrella Sampling." *Journal of Computational Physics* 23, no. 2 (February 1977): 187–99. https://doi.org/10.1016/0021-9991(77)90121-8.
- 29. Wang, Junmei, Piotr Cieplak, and Peter A. Kollman. 2000. "How Well Does a Restrained Electrostatic Potential (RESP) Model Perform in Calculating Conformational Energies of Organic and Biological Molecules?" *Journal of Computational Chemistry* 21 (12): 1049–74. <u>https://doi.org/10.1002/1096-987x(200009)21:12</u>.
- Yang, Yi Isaac, Qiang Shao, Jun Zhang, Lijiang Yang, and Yi Qin Gao.
 "Enhanced Sampling in Molecular Dynamics." *The Journal of Chemical Physics* 151, no. 7 (August 16, 2019). https://doi.org/10.1063/1.5109531.

Chapter 2

Assessing the Performance of NE-TI Method in Flavodoxin's Redox Potential Estimation

2.1 Abstract

Flavodoxins are enzymes that contain the redox-active flavin mononucleotide (FMN) cofactor and play a crucial role in numerous biological processes, including energy conversion and electron transfer. Since the redox characteristics of flavodoxins are significantly impacted by the molecular environment of the FMN cofactor, the evaluation of the interplay between the redox properties of the flavin cofactor and its molecular surroundings in flavoproteins is a critical area of investigation for both fundamental research and technological advancements, as the electrochemical tuning of flavoproteins is necessary for optimal interaction with redox acceptor or donor molecules. In order to facilitate the rational design of biomolecular devices, it is imperative to have access to computational tools that can accurately predict the redox potential of both natural and artificial flavoproteins. In this study, we have investigated the feasibility of using nonequilibrium thermodynamic integration protocols to reliably predict the redox potential of flavodoxins. Using as a test set the wild-type flavodoxin from Clostridium Beijerinckii and eight experimentally characterized single-point mutants, we have computed their redox potential. Our results show that 75% (6 out of 8) of the calculated reaction free energies are within 1 kcal/mol of the experimental values, and none exceed an error of 2 kcal/mol, confirming that non-equilibrium thermodynamic integration is a trustworthy tool for the quantitative estimation of the redox potential of this biologically and technologically significant class of enzymes.

2.2 Introduction

Due to the extraordinarily large number of known flavoproteins and the wide range of redox reactions that this family of enzymes may catalyze, flavoproteins rank among the most significant protein families engaged in redox processes [31, 38]. According to the particular protein environment, the redox potential of flavoproteins can vary by hundreds of mV from the redox potential of the free flavin cofactor in water and span a very wide range [20, 29]. The flavin cofactor is also peculiar because it can catalyze both one- and two-electron transfer reactions. In fact, flavins can exist in three redox forms (Fig. 2.1): quinone (OX), semiquinone (either as anionic, ASQ, or neutral, NSQ, species), and hydroquinone (HQ). Since electrochemical tuning of flavoproteins is essential for optimal reactivity with acceptor or donor redox molecules, evaluating the relationships between the redox properties of the flavin cofactor and their molecular environment in flavoproteins is an important area of study for both basic research and technological applications [4, 18, 37, 44]. In fact, the development of biosensors [16], biocatalysis [6, 16], bioremediation [1], and bioelectronics [1] are just a few biotechnological applications that flavoproteins represent a potential molecular system for. Therefore, the availability of computational methods able to reliably predict the redox potential of natural and de novo designed flavoproteins is crucial for the rational design of tailored molecular devices. Flavodoxins are a diverse group of flavoproteins that play a crucial role in a variety of biological processes, including electron transfer and energy conversion. These proteins contain a FMN cofactor, which serves as a redox-active molecule and provides flavodoxin with its unique redox properties. In recent years, significant progress has been made in the study of flavodoxins and their redox properties. Structural and functional studies have established that electrostatic interactions are a dominant factor affecting the redox properties of the FMN cofactor. In particular, mutations in flavodoxins have revealed a strong correlation between the redox potential of the cofactor and the number of negatively charged groups in the vicinity of the flavin [60]. Structural and functional studies on flavodoxins have also established that in

flavodoxin, the first reduction involves the transition from OX to NSO, i.e., it can be described as a proton-coupled redox reaction [11, 40, 52, 58], and electrostatic interactions are a dominant factor affecting the NSQ/HQ equilibrium. In particular, since the flavin hydroquinone in flavodoxins is not protonated at N1 [19], the isoalloxazine moiety is anionic, and it is expected to generate substantial repulsions in the negatively charged protein environment commonly observed in flavodoxins [41, 59]. Indeed, mutations in Desulfovibrio Vulgaris flavodoxin have revealed a strong correlation of the NSQ/HQ potential with the number of negatively charged groups in the neighborhood of the flavin [59], confirming that the flavin mononucleotide cofactor bound to flavodoxins is more difficult to convert to the fully reduced form compared to free FMN. According studies flavodoxins on wild-type and mutant from *Desulfovibrio* to vulgaris and Clostridium beijerinckii [41, 54, 55, 59], unfavorable aromatic stacking interactions can also be important in tuning the redox potential. Other investigations have also emphasized and shown the impact on the flavin reduction potential of particular hydrogen bonds, electrostatic, hydrophobic, and stacking interactions, as well as conformational alterations of the tricyclic ring or its surroundings [8, 9, 17, 30, 59, 60]. However, it is still difficult to quantitatively evaluate the correlations between structural factors and redox properties, which limits the ability to rationally design synthetic flavoproteins with specific redox properties. Indeed, in order to estimate the redox potential of flavoproteins, a number of computational techniques have been evaluated and employed [24, 56]. As a benchmark for subsequent QM/MM experiments aiming at examining the redox characteristics of tiny flavoproteins, Truhlar and associates presented a series of seminal DFT examinations of flavins in various solvents [45]. However, the systematic implementation of QM and QM/MM approaches is still hampered by the high computing cost of such calculations, despite the fact that QM and QM/MM studies enable predicting flavin reduction potentials with an average error of only 10-20 mV [35, 43]. In addition to QM and QM/MM investigations, studies based on a description of flavoproteins in terms of molecular mechanics have also been published. A thermodynamic integration investigation was conducted by Sattelle and Sutcliffe [48] on a variety of naturally occurring and artificially created flavodoxins, each of which differed for one amino acid in the cofactor environment, obtaining an average difference between estimated and actual redox potential of 20–100 mV.

To further assess the possibility of using thermodynamic integration protocols to reliably predict the redox potential of flavodoxins, we have used a fast growth non-equilibrium approach to compute the redox potential of the flavodoxin from *Clostridium Beijerinckii*. In particular, we have conducted a systematic analysis of a number of single-point mutants of the flavodoxin from *Clostridium Beijerinckii*, for which redox potentials were already experimentally measured, with the aim of evaluating and validating this approach in the calculation of relative redox potential in flavoproteins, comparing results with similar investigations carried out using different methodological approaches. NE-TI protocols have already been used to investigate other processes, such as docking of ligands to proteins and redox processes [2, 23, 36, 42]. However, to the best of our knowledge, this is the first report about the evaluation of a non-equilibrium thermodynamic integration study for the prediction of the relative redox potential of biological macromolecules.



Figure 2.1: Redox and protonation states for the isoalloxazine ring in flavoproteins: quinone (OX), semiquinone (either as anionic, ASQ, or neutral, NSQ, species), and hydroquinone (HQ)[51].

2.3 Materials and methods

2.3.1 Simulated systems

In this work, we explored the first reduction event in nine variants of Clostridium Beijerinckii flavodoxin (WT, G57T, D58P, E59A, M56A, M56G, M56L, M56I, and M56V). The chosen mutants are all electrochemically characterized, and the redox potentials associated with both flavin reductions are known [8, 34, 40]. The starting structures of WT, G57T, and D58P flavodoxins correspond to the X-ray oxidized (pdb id: 5NLL, 1FLD, 4NUL) and semiquinone (pdb id: 2FOX, 5NUL, 1FLN) structures. The crystallographic structure of WT_{OX} (5NLL) presents the three known alternative conformations for the 50's loop [40]; only the structure prevailing in solution (cis-Odown) was selected for the simulations. The G57T_{SO} crystal (5NUL) is also characterized by the presence of two conformers, of which only the prevalent one [40], featuring trans-O-up conformation, is used for the simulations. The pdb structure 1FLN corresponds to the fully reduced state of flavodoxin; however, due to the strong structural similarity between the SQ and HQ states found in the other crystallized structures, such a structure was also used for the semiquinone one. The starting structures of other systems are obtained by in silico mutagenesis of oxidized, semiquinone, and fully reduced X-ray structures of the WT system using the Pymol mutagenesis tool (Tab. 2.1). All structures are processed with the protein preparation wizard of Maestro 2021-4 software to remove the co-crystallized water molecules, add hydrogen atoms assuming pH = 7, and fill in missing side chain atoms and missing loops.

G (OX		NSQ			
System	Structure ^a	50's Loop ^b	Structure ^a	50's loop ^b		
WT	5NLL	cis-O-down	2FOX	trans-O-up		
G57T	1FLD	trans-O-down	5NUL	trans-O-down		
D58P	4NUL	cis-O-down	5FLN	trans-O-down		
E59A	5NLL in silico mutagenesis	cis-O-down	2FOX in silico mutagenesis	trans-O-up		
M56A	5NLL in silico mutagenesis	cis-O-down	2FOX in silico mutagenesis	trans-O-up		
M56G	5NLL in silico mutagenesis	cis-O-down	2FOX in silico mutagenesis	trans-O-up		
M56L	5NLL in silico mutagenesis	cis-O-down	2FOX in silico mutagenesis	trans-O-up		
M56I	5NLL in silico mutagenesis	cis-O-down	2FOX in silico mutagenesis	trans-O-up		
M56V	5NLL in silico mutagenesis	cis-O-down	2FOX in silico mutagenesis	trans-O-up		

Table 2 1.	Summary	of simulation	evetom	details of th	he OY an	A NSO	flavin_flav	adarin	complay
<i>Table 2.1</i> .	Summary e) simulation	system	aeiaiis oj ii	не Ол ини	a NSQ	jiavin-jiav	оаохіп	сотриех.

(*a*) PDB code. (*b*) Conformation of the reverse turn involving residues 56–59 in the Clostridium Beijerinckii flavodoxin.

2.3.2 Equilibrium MD setup

All equilibrium molecular dynamics simulations were produced using the GROMACS 2020.3 [57] software and the CHARMM36 [28] force field integrated with the parameters for flavins developed by Alexey Aleksandrov [3]. The investigated flavoproteins were solvated with TIP3P water molecules [33] in a cubic box of 6.22 nm³. Counter ions K⁺ and Cl⁻ were added to obtain a neutrally charged box and to reach a salt concentration of 0.1 M. All models were minimized using a steepest descent algorithm and then equilibrated at 300 K and 1 bar through a 4 ns NVT simulation and 1ns NPT simulation, respectively. Coupling to the external bath is achieved using two different algorithms: the modified Berendsen thermostat is used for temperature [5, 10] and the Parrinello-Rahman algorithm for pressure [47, 46]. The LINCS algorithm [27] was used to constrain all covalent bonds in which hydrogen atoms are involved, while the equations of motion are integrated using the leap-frog algorithm. The non-covalent interactions were computed according to the Verlet cutoff scheme [25], while van der Waals interactions and shortrange electrostatic interactions were evaluated at each time step within 1.2 nm of each particle. Long-range electrostatic interactions were evaluated using the PME method with an interpolation order of 4 and the remaining parameters set to default values [15]. All simulations were performed using periodic boundary conditions in all directions. The production simulations were carried out using an integration time step of 1 fs. We initially tested a first protocol of three MD simulations of 10 ns followed by three TI simulations (similarly to what has been performed by Aldeghi et al. [2]); then, we tried another approach in which a single MD simulation of 500 ns was followed by one TI calculation in order to evaluate the effect of extended sampling on the free energy estimation.

2.3.3 NE-TI setup

Hybrid topologies were generated using the *PMX* package developed by Daniel Seelinger and Bert L. de Groot [49]. A total of 500 equidistant snapshots were extracted from each MD simulation for each system (in the case of 3×10 ns simulations, 500 snapshots were extracted from each replica), and they were used as a starting point for 100 ps nonequilibrium transitions. These simulations were carried out using a $\Delta\lambda$ of 10^{-5} and setting the sc-alpha (soft core alpha) and sc-sigma (soft core sigma) to 1.2 and 0.3 [39], respectively. Finally, the ΔG corresponding to the redox reactions was calculated using the BAR approach [50]. When considering the 3×10 ns simulations, the BAR method was also applied to the three forward and three reverse concatenated trajectories. Uncertainties were estimated as standard errors ($\sigma_{\Delta G}$) by separately considering each equilibrium simulation and its related non-equilibrium trajectories as independent calculations (i.e., when 3 equilibrium simulations were used, 3 independent ΔG values were obtained, and these were used to estimate $\sigma_{\Delta G}$). These uncertainties were then propagated to the final $\Delta\Delta G$ estimate so as to obtain the standard error $\sigma_{\Delta\Delta G}$. The total amount of time simulated in this work is about 21 µs.

2.3 Results and discussion

The accurate calculation of absolute protein redox potentials, i.e., within ~1 kcal/mol (~50 mV) of the experimental values, remains extremely challenging. As initially underlined by Van Gunsteren, Canters, and collaborators [7, 53], and more recently by D'Abramo and collaborators [12], the direct computation of protein redox potential as a function of pH or protein composition using a comprehensive quantum-chemical treatment of the system is still essentially unattainable. In fact, such a task would necessitate a comprehensive quantum-chemical treatment of the reaction site, encompassing its immediate surroundings and accounting for entropic effects arising from alternative conformational states of the protein and the solvent. Consequently, researchers are compelled to adopt one or more simplifying assumptions to carry out such investigations. One viable approach involves employing high-level quantum-chemical methodology to characterize the redox center and its proximal environment while disregarding entropic contributions, the rest of the protein environment, and the solvent (or describing them using an MM forcefield). Alternatively, one may consider treating the protein in an explicit solvent using an empirical force field and classical dynamics. This approach incorporates entropic effects through extensive conformational sampling, treating oxidation and reduction as purely electrostatic phenomena and implicitly assuming that quantum-mechanical contributions are consistent across all mutations. A third possible approach involves the omission of both the quantum-mechanical nature of the transition and entropic contributions from the environment, treating oxidation and reduction solely as electrostatic phenomena. For example, conducting Poisson–Boltzmann calculations based on a given (X-ray) structure in the oxidized and reduced states.

With the aim of contributing to the evaluation of computational approaches for the calculation of protein thermodynamic properties, we have used the NE-TI method for the computation of flavodoxin's redox potential, studying a dataset comprising the wild-type flavodoxin from *Clostridium beijerinkii* and eight single-amino acid mutants for which experimentally derived redox potentials are available.

It should be noted that, due to approximations to the energy function and possible incomplete sampling [13], the relative values of calculated molecular properties are generally more reliable than the corresponding absolute values. Therefore, with the aim of achieving better accuracy, the redox properties of single-point mutants have been evaluated using the wild-type value as a reference [48]. Specifically, $\Delta\Delta G$ values have been calculated by subtracting from the ΔG value of the WT the one relative to each mutant, according to the following formula:

$$\Delta \Delta G_{MUT} = \Delta G_{MUT} - \Delta G_{WT}$$

In addition, we have focused our attention on the OX/NSQ redox process, in which the reduction/protonation event is well characterized and directly involves the flavin cofactor (Fig. 2.2). Of course, the computational investigation of the SQ/HQ redox process would also be relevant to the study. However, the protonation event associated with the SQ/HQ redox transition does not directly involve the flavin cofactor but has been suggested to imply the protonation of some aminoacids in the cofactor environment [30], making the protonation form of the protein to be used in the TI simulations more uncertain.

All three redox states for the wild-type flavodoxin from *Clostridium beijerinckii* were previously characterized by X-ray diffraction [40]. It is noteworthy that the conformational properties of the so called "50s loop", which includes the residues Met56-Gly57-Asp58-Glu59, have been suggested to be affected by the redox state of the isoalloxazine ring of the FMN cofactor [34]. In particular, the Gly57-Asp58 peptide bond in the OX redox state is found predominantly in the cis-O-down and trans-O-down conformations, which are characterized by the carbonyl group oriented opposite to the flavin ring (Fig. 2.2).

After one-electron reduction, the peptide bond rearranges in a trans-O-up conformation where the carbonyl group, which plays the role of a hydrogen-bond acceptor, interacts with N5H, stabilizing the reduced FMN. In fact, according to X-ray data, the wild-type flavodoxin in the OX state is a mixture of cis-O-down (50%), trans-O-down (20%), and



Figure 2.2: (A) structure of C. Beijerinckii flavodoxin (PDBID 5NLL). (B) Conformation of 50s loop [51].

trans-O-up (30%) conformations. In light of these observations, we have initially evaluated the effect of the residues 57–58 conformation upon the OX/SQ reduction in the wild-type flavodoxin, taking into account these transitions:

 $FMN_{OX/cis-O-down} \rightarrow FMN_{NSQ/cis-O-down}$ $FMN_{OX/cis-O-down} \rightarrow FMN_{NSQ/trans-O-down}$ $FMN_{OX/cis-O-down} \rightarrow FMN_{NSQ/trans-O-up}$

Three replicas of 10 ns were conducted on all four structures to obtain a representative ensemble for each state, then three replicas of the TI calculation were performed for each reaction, and the $\Delta G_{OX \rightarrow NSO}$ has been evaluated as the average of the replicas. The use of relatively brief MD replicas instead of longer MD trajectories as the source of conformational ensembles for the oxidized (OX) and reduced (SQ) states is based on previous findings by De Groot and colleagues, who reported that for obtaining reliable TI results, it is generally advantageous to sample conformations that are proximal to the experimentally determined structure [2]. Indeed, for the sake of completeness, we also tested the possibility of using 500 ns MD trajectories to generate starting points for the non-equilibrium TI simulations, but it turned out that the results were poorer (see below). The starting structures were retrieved from the pdb files 5NLL and 2FOX for OXflavodoxin and SQ-flavodoxin, respectively. Results show very similar reaction free energy values in the three cases: $\Delta G_{OX \rightarrow NSO}$ of -42.92 kcal/mol, -42.91 kcal/mol, and -43.76 kcal/mol for the OX_{(cis-O-down}/NSQ_(trans-O-up), OX_{(trans-O-down}/NSQ_(trans-O-up), and OX_(trans-O-up)/NSQ_(trans-O-up) redox transitions, respectively (Tab. 2.2). All values are in a range of 0.85 kcal/mol, suggesting that the OX/NSQ reaction energy is not strongly affected by the conformation of residues 56–57.

WT	Exp. Ratio	Replica ^a	ΔG_{cal} 10 ns ^b	Average ΔG_{cal} ^{b,c}	Std. Error ^b	MAE ^b	RMSE ^b
	50%	1	-43.31 ± 0.05	-42.92	0.31	1.37	1.54
Cis-O- down		2	-43.44 ± 0.04				
		3	-42.48 ± 0.04				
Trans.	20%	1	-42.92 ± 0.05	-42.91	0.19	0.64	0.87
0-		2	-43.97 ± 0.04				
down		3	-43.73 ± 0.05				
		1	-43.17 ± 0.05	-43.76	0.32	1.37	
Trans- O-up	30%	2	-42.70 ± 0.04				1.54
		3	-43.33 ± 0.04				
			Overall	-43.17 ^d	0.29 ^d	1.06 ^e	1.25 ^e

Table 2.2: Free energy values of the first reduction in with	<i>d-type flavodoxin considering all the 50's loop conformation</i>
for the OX.	

(a) 10 ns simulations. (b) Values are reported in kcal/mol. (c) ΔG is calculated using the BAR method on the three concatenated trajectories. (d) All values are calculated as the weighted arithmetic mean using the experimental relative abundance as weight. (e) MAE and RMSE are calculated using the overall ΔG as reference.

In light of this observation, to simplify and reduce the number of simulations, we have used only the cis-O-down conformation as the OX state for all the flavodoxin variants for which the crystallographic structure is presently not available. The same applies to the NSQ state, for which we adopted the trans-O-up conformation.

Computed root-mean-square deviation (RMSD) and root-mean-square fluctuation (RMSF) values (see Supplementary figures section) highlight the great structural stability of these proteins. In fact, the analysis of RMSF values allowed us to observe that the only protein portion undergoing conformational transitions along the MD simulations is the loop in front of FMN. This observation is relevant because the conformation of this loop is known to be important for the modulation of the flavin redox potential [34]. The calculated $\Delta\Delta G$ values for the OX to NSQ redox transition for the wild-type flavodoxin as well as for the single-point mutant investigated are collected in tables 2.3 and 2.4, while the corresponding distributions of forward and reverse work values (W) obtained during the TI simulations are shown in the Supplementary figures section. Computed $\Delta\Delta G$ values obtained from 10ns replicas can be generally considered satisfactory, with the minimum, maximum, and mean absolute error equal to 0.60, 3.00, and 1.37 kcal/mol, respectively. The method therefore fails at most by about 3 kcal/mol (G57T mutant), while the calculated RMSE is less than 0.93 kcal/mol. It is also very encouraging that the mean deviation of the calculated values from the experimental ones is only 31 mV. In fact, our results nicely compare to those obtained in a MD-FEP study carried out on Anabaena's Flavodoxins [48] and also with the results reported by Grater and coworkers, who used a free-energy calculation scheme based on the Crooks Gaussian intersection method to estimate the redox potential of thiol/disulfide pairs in 12 proteins belonging to the thioredoxin superfamily [39]. Our deviations from calculated values from experimental data are also in line with recently reported TI simulations carried out to estimate ligand binding affinity [2, 21, 22].

3×10 ns Simulations								
	ΔG_{exp} a	$\Delta\Delta G_{exp}^{a,b}$	Rep1 _{calc} ^a	Rep2 _{calc} ^a	Rep3 _{calc} ^a	ΔG_{calc} a,c,d	$\Delta\Delta G_{calc}{}^{a,b}$	Abs. Error ^a
WT	2.12 ± 0.11	-	${}^{-43.33\pm}_{0.05}$	$\begin{matrix}-43.46\pm\\0.04\end{matrix}$	$\begin{matrix}-42.50\pm\\0.04\end{matrix}$	-42.94 ± 0.31	-	-
G57T	6.23 ± 0.11	4.11 ± 0.23	$\begin{matrix}-41.17\pm\\0.04\end{matrix}$	-42.12 ± 0.05	${}^{-42.05\pm}_{0.04}$	-41.83 ± 0.30	1.11 ± 0.61	3.00
D58P	$\begin{array}{c} 3.58 \pm \\ 0.11 \end{array}$	1.46 ± 0.23	$\begin{matrix}-41.98\pm\\0.05\end{matrix}$	-42.81 ± 0.06	${}^{-43.31\pm}_{0.03}$	-42.81 ± 0.40	0.13 ± 0.71	1.33
E59A	4.29 ± 0.12	2.17 ± 0.23	$\begin{array}{c}-41.64\pm\\0.04\end{array}$	-42.53 ± 0.05	${}^{-43.22\pm}_{0.05}$	-42.18 ± 0.36	0.76 ± 0.67	1.41
M56A	1.66 ± 0.18	$\begin{array}{c}-0.46\pm\\0.30\end{array}$	$\begin{matrix}-43.98\pm\\0.04\end{matrix}$	$\begin{matrix}-43.66\pm\\0.06\end{matrix}$	$\begin{matrix}-44.56 \\ 0.05\end{matrix}$	-44.01 ± 0.32	-1.06 ± 0.63	0.60
M56G	1.94 ± 0.18	$\begin{array}{c}-0.18\pm\\0.30\end{array}$	$\begin{array}{c}-44.14\pm\\0.06\end{array}$	$\begin{array}{c}-43.31\pm\\0.05\end{array}$	$\begin{matrix}-44.45 \\ 0.07\end{matrix}$	$\begin{array}{c}-43.76\pm\\0.33\end{array}$	$\begin{array}{c} -0.82 \pm \\ 0.64 \end{array}$	1.24
M56L	$\begin{array}{c} 2.95 \pm \\ 0.18 \end{array}$	$\begin{array}{c} 0.83 \pm \\ 0.30 \end{array}$	-42.93 ± 0.06	$\begin{array}{c}-43.14\pm\\0.05\end{array}$	$\begin{matrix}-44.20\pm\\0.05\end{matrix}$	-43.21 ± 0.42	$\begin{array}{c} -0.27 \pm \\ 0.73 \end{array}$	1.10
M56I	$\begin{array}{c} 2.81 \pm \\ 0.18 \end{array}$	0.69 ± 0.30	$\begin{array}{c}-43.18\pm\\0.07\end{array}$	$\begin{matrix}-43.93\pm\\0.06\end{matrix}$	$\begin{matrix}-43.68\pm\\0.08\end{matrix}$	$\begin{array}{c}-43.65\pm\\0.20\end{array}$	$\begin{array}{c}-0.71\pm\\0.51\end{array}$	1.40
M56V	$\begin{array}{c} 2.93 \pm \\ 0.18 \end{array}$	$\begin{array}{c} 0.81 \pm \\ 0.30 \end{array}$	$\begin{matrix}-43.95\pm\\0.05\end{matrix}$	$\begin{matrix}-43.61\pm\\0.05\end{matrix}$	$\begin{matrix}-43.49\pm\\0.05\end{matrix}$	$\begin{matrix}-43.66\pm\\0.17\end{matrix}$	$\begin{array}{c}-0.72\pm\\0.48\end{array}$	1.53
						MAE	1.37	
						RMSσ	0.06	
						RMSE	1.54	

Table 2.3: ΔG and $\Delta \Delta G$ one-electron reduction potentials estimate both experimentally (exp) and in silico (calc) for 3 \times 10 ns.

(*a*) Values are reported in kcal/mol. (*b*) $\Delta\Delta G$ are calculated respect to WT. (*c*) Average of Replica1, Replica2 and Replica3. (*d*) ΔG is calculated using the BAR method on the three concatenated trajectories.

1 × 500 ns Simulations									
	ΔG_{exp} ^a	$\Delta\Delta G_{exp}^{\ a,b}$	ΔG_{calc} ^a	$\Delta\Delta G_{calc}$ ^{a,b}	Abs. Error ^a				
WT	2.12 ± 0.11	-	-43.60 ± 0.08	-	-				
G57T	6.23 ± 0.11	4.11 ± 0.23	-41.45 ± 0.04	2.15 ± 0.12	1.96				
D58P	3.58 ± 0.11	1.46 ± 0.23	-42.21 ± 0.04	1.39 ± 0.12	0.07				
E59A	4.29 ± 0.12	2.17 ± 0.23	-42.14 ± 0.04	1.46 ± 0.12	0.71				
M56A	1.66 ± 0.18	-0.46 ± 0.30	-43.49 ± 0.04	0.11 ± 0.12	0.57				
M56G	1.94 ± 0.18	-0.18 ± 0.30	-43.80 ± 0.05	-0.20 ± 0.13	0.02				
M56L	2.95 ± 0.18	0.83 ± 0.30	-41.99 ± 0.03	1.61 ± 0.11	0.78				
M561	2.81 ± 0.18	0.69 ± 0.30	-42.86 ± 0.05	0.74 ± 0.13	0.05				
M56V	2.93 ± 0.18	0.81 ± 0.30	-44.30 ± 0.04	-0.70 ± 0.12	1.51				
			MAE	0.70					
			RMS	0.95					

Table 2.4: ΔG and $\Delta \Delta G$ one-electron reduction potentials estimate both experimentally (exp) and in silico (calc) for 1 \times 500 ns simulations

(a) Values are reported in kcal/mol. (b) $\Delta\Delta G$ are calculated respect to WT.

The 0.89 \mathbb{R}^2 value also indicates that the method satisfactorily estimates the differences in redox potential among the tested mutants (fig. 2.3). On the other hand, computed $\Delta\Delta G$ values obtained from a 500 ns trajectory are characterized by poorer agreement with experimental values; in fact, the \mathbb{R}^2 value is 0.58, although the minimum, maximum, and mean absolute error are equal to 0.07 kcal/mol, 1.96 kcal/mol, and 0.70 kcal/mol, respectively. These data highlight that the best predictions are obtained by reproducing multiple short-time replicas rather than a single longer simulation. While the average absolute error is smaller in the 500 ns simulation, the relative behavior of the data shows improvement in the case of three 10 ns replicas, as shown by the \mathbb{R}^2 value being closer to 1. This implies that, by applying a correction factor, the method is more effective in predicting the relative redox potential trends of different mutations when multiple shorttime replicas are used.

The relative error calculated for $\Delta\Delta G$ appears to be smaller for the 500 ns simulation, but such an observation is biased by the different number of replicas carried out in the 500 and 10 ns protocols.



Figure 2.3: Scatter plots $\Delta\Delta G_{cal}$ Vs $\Delta\Delta G_{exp}$ for 3× replicas of 10ns (A) and for 1× replica of 500 ns (B). Trendlines are reported in red, which have a slope of 1.72 and 1.08 for 3 × 10 ns (A) and for 1 × 500 ns (B), respectively.

In fact, if each 10 ns simulation is individually compared to the 500 ns simulation, it is observed that the error calculated by the BAR method is nearly identical in both setups (values are in a range from 0.04 to 0.08). The error increases when $\Delta\Delta G$ is calculated on the three concatenated 10 ns trajectories, suggesting that the BAR method is more affected by the number of independent trajectories than their duration.

For selected systems, we have also produced replicas of 500 ns trajectories in order to analyze the consistency of the methodology, reproducibility, and influence of the number of fast transitions on the final ΔG value. In addition, to test the consistency of the results, we also evaluated whether the chosen thermodynamic cycle model (shown in fig. 2.4) closed properly. The chosen systems were the WT flavodoxin and the M56V mutant, as the latter is characterized by an unsigned error (UE) for the $\Delta\Delta G$ above 1 kcal/mol, i.e., the second highest calculated among all systems. The cycle was constructed using two protein mutagenesis reactions as vertical branches, where flavodoxin WT was chemically converted into its M56V variant, both in the oxidized form and in the semiguinone form of the flavin cofactor. The values of the horizontal branches ($WT_{OX} \rightarrow WT_{NSO}$ and $M56V_{OX} \rightarrow M56V_{NSQ}$) were obtained with 3 MD runs of 500 ns each, while the vertical branches (WT_{OX} \rightarrow M56V_{OX} and WT_{NSQ} \rightarrow M56V_{NSQ}) were obtained with a single 500 ns MD each. The same protocol is repeated using 10 ns MD simulations. Even if a statistical analysis on multiple mutants would be necessary to exclude the possibility that the discrepancy obtained is mutant-dependent, the results are very encouraging. In fact, the difference obtained between the horizontal and vertical branches is below 0.55 kcal/mol (using 500 ns MD simulations) and 0.65 kcal/mol (using 10 ns MD simulations), showing that the calculations are sound for this type of system and suggesting that the $\Delta\Delta G$ error associated with the redox reaction may mainly derive from an inadequate force field. Tests with other force fields will be carried out in the future to further test this hypothesis.



Figure 2.4: Thermodynamic cycle showing the free energy differences of reduction and the free energy differences of WT to M56V transformation for 3×500 ns (**A**) and 3×10 ns (**B**).

With the aim of providing a rationale for the difference in redox potential among wildtype and single-point mutant flavodoxins, we have also examined in detail the structural and dynamic properties of the flavin cofactor and of its environment. In particular, for such a purpose, four structural features were taken into account, both for OX and SQ states. (i) The distance between the centers of geometry (COGs) of the protein and the isoalloxazine ring; (ii) the distance between the COGs of the isoalloxazine ring and TRP90 that plays an important role in the affinity of FMN to the pocket [34]; (iii) the distance between N5 (HN5 in the case of SQ flavin) and carbonyl oxygen of residue 57; and (iv) the value of the ω -torsion angle of residues 57–58 (see Supplementary figures section). When considering 3×10 ns trajectories, the investigated parameters do not undergo any significant change, possibly due to the short timescale. Nevertheless, the calculated redox potentials well correlate with the experimental data. This observation suggests that the chemical properties of the environment surrounding the flavin cofactor have a significant impact on the redox potential, whereas structural rearrangement does not play an important role in redox potential modulation. When considering the 1×500 ns trajectory, the most relevant difference among the enzyme variants is the distance between the hydrogen H5 (H_{flavin}) and the carbonyl oxygen of residue 57 (O_{carbonyl}), which quantifies the extent of the formation of a hydrogen bond between the flavin moiety and residue 57. Single-point mutants G57T, E59A, and M56L are characterized by an increase in the O_{carbonyl}-H_{flavin} distance along the MD simulations, and a corresponding decrease in the redox potential is computed, in nice agreement with previous proposals indicating that such a bond stabilizes the SQ state [40].

2.4 Conclusions

In this study, we have shown that NE-TI calculations can reliably be used to predict even subtle thermodynamic properties, such as the difference in redox potentials for WT and single-point mutant flavodoxins. Therefore, NE-TI calculations have the potential to be a robust computational tool available to research laboratories for the design of flavoproteins with tailored redox properties. It remains to be evaluated how this computational approach remains effective for other redox enzyme families and/or proton-coupled redox reactions where not only the cofactor but also proximal aminoacids are directly involved in the process. Studies aimed at investigating such issues are currently underway in our laboratory.

Supplementary figures and tables



Figure S2.1: RMSD values of Wild-type and mutant variants for 3x10ns replicas (A) and 1x500ns replica (B). Red vertical lines in panel A divide the total RMSD showing the three replicas for each flavodoxin.

Time (ns)

M56V



Figure S2.2: RMSF values of Wild-Type and mutant variants for 3x10ns replicas (A) and 1x500ns replica (B)



Figure S2.3: ΔG values calculated with the BAR estimator for the nine flavodoxin variants in the 3x10ns simulations. For each system, ΔG value is calculated taking together all the transitions collected in the three replicas; as a result 1500 transitions are used for both forward and backward reaction. On the left, the work values for the forward (green) and backward (blue) transitions are reported for every starting structure. On the right, the distributions of these work values are shown as histograms and are used to draw the gaussians (green and blue dashed lines), the intersection (black dashed line) of which allows obtaining the free energy difference.



Figure S2.4: ΔG values calculated with the BAR estimator for the nine flavodoxin variants in the 1x500ns simulations. For each system ΔG value is calculated using 500 transitions for both forward and backward reaction. On the left, the work values for the forward (green) and backward (blue) transitions are reported for every starting structure. On the right, the distributions of these work values are shown as histograms and are used to draw the gaussians (green and blue dashed lines), the intersection (black dashed line) of which allows obtaining the free energy difference.



Figure S2.5: Structural data for wild-type and mutant flavodoxins in the 3x10ns simulations.



Figure S2.6: Structural data for wild-type and mutant flavodoxins in the 1x500ns simulations.

Bibliography

- Ackerley, D. F., C. F. Gonzalez, C. H. Park, R. Blake II, M. Keyhan, and A. Matin. "Chromate-Reducing Properties of Soluble Flavoproteins from Pseudomonas Putida and Escherichia Coli." *Applied and Environmental Microbiology* 70, no. 2 (February 2004): 873–82. https://doi.org/10.1128/aem.70.2.873-882.2004.
- Aldeghi, Matteo, Vytautas Gapsys, and Bert L. de Groot. "Accurate Estimation of Ligand Binding Affinity Changes upon Protein Mutation." ACS Central Science 4, no. 12 (December 13, 2018): 1708–18. https://doi.org/10.1021/acscentsci.8b00717.
- Aleksandrov, Alexey. "A Molecular Mechanics Model for Flavins." *Journal of Computational Chemistry* 40, no. 32 (August 31, 2019): 2834–42. https://doi.org/10.1002/jcc.26061.
- Batinic-Haberle, Ines, Artak Tovmasyan, Emily R. H. Roberts, Zeljko Vujaskovic, Kam W. Leong, and Ivan Spasojevic. "SOD Therapeutics: Latest Insights into Their Structure-Activity Relationships and Impact on the Cellular Redox-Based Signaling Pathways." *Antioxidants & amp; Redox Signaling* 20, no. 15 (May 20, 2014): 2372–2415. https://doi.org/10.1089/ars.2012.5147.
- Berendsen, H. J. C., J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak. "Molecular Dynamics with Coupling to an External Bath." *The Journal of Chemical Physics* 81, no. 8 (October 15, 1984): 3684–90. https://doi.org/10.1063/1.448118.
- 6. Berkel, W.J.H. van, N.M. Kamerbeek, and M.W. Fraaije. "Flavoprotein Monooxygenases, a Diverse Class of Oxidative Biocatalysts." *Journal of*

Biotechnology 124, no. 4 (August 2006): 670–89. https://doi.org/10.1016/j.jbiotec.2006.03.044.

- Bosch, Marieke van den, Marcel Swart, Jaap G. Snijders[†], Herman J. C. Berendsen, Alan E. Mark, Chris Oostenbrink, Wilfred F. van Gunsteren, and Gerard W. Canters. "Calculation of the Redox Potential of the Protein Azurin and Some Mutants." *ChemBioChem* 6, no. 4 (April 6, 2005): 738–46. https://doi.org/10.1002/cbic.200400244.
- Bradley, Luke H., and Richard P. Swenson. "Role of Glutamate-59 Hydrogen Bonded to N(3)H of the Flavin Mononucleotide Cofactor in the Modulation of the Redox Potentials of theClostridium beijerinckiiFlavodoxin. Glutamate-59 Is Not Responsible for the pH Dependency but Contributes to the Stabilization of the Flavin Semiquinone." *Biochemistry* 38, no. 38 (August 26, 1999): 12377–86. https://doi.org/10.1021/bi991172f.
- Bradley, Luke H., and Richard P. Swenson. "Role of Hydrogen Bonding Interactions to N(3)H of the Flavin Mononucleotide Cofactor in the Modulation of the Redox Potentials of the Clostridium Beijerinckii Flavodoxin." *Biochemistry* 40, no. 30 (June 30, 2001): 8686–95. https://doi.org/10.1021/bi010571j.
- Bussi, Giovanni, Davide Donadio, and Michele Parrinello. "Canonical Sampling through Velocity Rescaling." *The Journal of Chemical Physics* 126, no. 1 (January 3, 2007). https://doi.org/10.1063/1.2408420.
- 11. Chang, Fu-Chung, and Richard P. Swenson. "The Midpoint Potentials for the Oxidized–Semiquinone Couple for Gly57 Mutants of the Clostridium Beijerinckii Flavodoxin Correlate with Changes in the Hydrogen-Bonding Interaction with the Proton on N(5) of the Reduced Flavin Mononucleotide Cofactor As Measured by NMR Chemical Shift Temperature Dependencies." *Biochemistry* 38, no. 22 (May 8, 1999): 7168–76. https://doi.org/10.1021/bi982203u.

- Chen, Cheng Giuseppe, Alessandro Nicola Nardi, Andrea Amadei, and Marco D'Abramo. "Theoretical Modeling of Redox Potentials of Biomolecules." *Molecules* 27, no. 3 (February 5, 2022): 1077. https://doi.org/10.3390/molecules27031077.
- Cournia, Zoe, Bryce Allen, and Woody Sherman. "Relative Binding Free Energy Calculations in Drug Discovery: Recent Advances and Practical Considerations." *Journal of Chemical Information and Modeling* 57, no. 12 (December 15, 2017): 2911–37. https://doi.org/10.1021/acs.jcim.7b00564.
- Crooks, Gavin E. "Nonequilibrium Measurements of Free Energy Differences for Microscopically Reversible Markovian Systems." *Journal of Statistical Physics* 90, no. 5/6 (1998): 1481–87. https://doi.org/10.1023/a:1023208217925.
- Darden, Tom, Darrin York, and Lee Pedersen. "Particle Mesh Ewald: An N·log(N) Method for Ewald Sums in Large Systems." *The Journal of Chemical Physics* 98, no. 12 (June 15, 1993): 10089–92. https://doi.org/10.1063/1.464397.
- 16. Dijkman, Willem P., Gonzalo de Gonzalo, Andrea Mattevi, and Marco W. Fraaije.
 "Flavoprotein Oxidases: Classification and Applications." *Applied Microbiology and Biotechnology* 97, no. 12 (May 3, 2013): 5177–88. https://doi.org/10.1007/s00253-013-4925-7.
- 17. Druhan, Lawrence J., and Richard P. Swenson. "Role of Methionine 56 in the of of Control the Oxidation-Reduction Potentials theClostridiumbeijerinckiiFlavodoxin: Effects of Substitutions by Aliphatic Amino Acids and Evidence for a Role of Sulfur-Flavin Interactions." 37, 19, 1998): *Biochemistry* no. 27 (June 9668-78. https://doi.org/10.1021/bi980770b.
- 18. Ferreira, Patricia, and Milagros Medina. "Anaerobic Stopped-Flow Spectrophotometry with Photodiode Array Detection in the Presteady State: An

Application to Elucidate Oxidoreduction Mechanisms in Flavoproteins." In *Methods in Molecular Biology*, 135–55. New York, NY: Springer US, 2021. http://dx.doi.org/10.1007/978-1-0716-1286-6_9.

- Franken, Hans-Dieter, Heinz Rüterjans, and Franz Müller. "Nuclear-magneticresonance Investigation of 15N-labeled Flavins, Free and Bound to Megasphaera Elsdenii Apoflavodoxin." *European Journal of Biochemistry* 138, no. 3 (February 1984): 481–89. https://doi.org/10.1111/j.1432-1033.1984.tb07942.x.
- 20. Galuzzi, Bruno Giovanni, Antonio Mirarchi, Edoardo Luca Viganò, Luca De Gioia, Chiara Damiani, and Federica Arrigoni. "Machine Learning for Efficient Prediction of Protein Redox Potential: The Flavoproteins Case." *Journal of Chemical Information and Modeling* 62, no. 19 (September 20, 2022): 4748–59. https://doi.org/10.1021/acs.jcim.2c00858.
- 21. Gapsys, Vytautas, Laura Pérez-Benito, Matteo Aldeghi, Daniel Seeliger, Herman van Vlijmen, Gary Tresadern, and Bert L. de Groot. "Large Scale Relative Protein Ligand Binding Affinities Using Non-Equilibrium Alchemy." *Chemical Science* 11, no. 4 (2020): 1140–52. https://doi.org/10.1039/c9sc03754c.
- 22. Gapsys, Vytautas, Ahmet Yildirim, Matteo Aldeghi, Yuriy Khalak, David van der Spoel, and Bert L. de Groot. "Accurate Absolute Free Energies for Ligand– Protein Binding Based on Non-Equilibrium Approaches." *Communications Chemistry* 4, no. 1 (May 11, 2021). https://doi.org/10.1038/s42004-021-00498-y.
- Genheden, Samuel, Ingemar Nilsson, and Ulf Ryde. "Binding Affinities of Factor Xa Inhibitors Estimated by Thermodynamic Integration and MM/GBSA." *Journal of Chemical Information and Modeling* 51, no. 4 (March 18, 2011): 947– 58. https://doi.org/10.1021/ci100458f.
- 24. Gillet, Natacha, Bernard Lévy, Vicent Moliner, Isabelle Demachy, and Aurélien de la Lande. "Theoretical Estimation of Redox Potential of Biological Quinone

Cofactors." *Journal of Computational Chemistry* 38, no. 18 (May 3, 2017): 1612–21. https://doi.org/10.1002/jcc.24802.

- 25. Grubmüller, H., H. Heller, A. Windemuth, and K. Schulten. "Generalized Verlet Algorithm for Efficient Molecular Dynamics Simulations with Long-Range Interactions." *Molecular Simulation* 6, no. 1–3 (March 1991): 121–42. https://doi.org/10.1080/08927029108022142.
- Hendrix, D. A., and C. Jarzynski. "A 'Fast Growth' Method of Computing Free Energy Differences." *The Journal of Chemical Physics* 114, no. 14 (April 8, 2001): 5974–81. https://doi.org/10.1063/1.1353552.
- 27. Hess, Berk, Henk Bekker, Herman J. C. Berendsen, and Johannes G. E. M. Fraaije. "LINCS: A Linear Constraint Solver for Molecular Simulations." *Journal of Computational Chemistry* 18, no. 12 (September 1997): 1463–72. https://doi.org/10.1002/(sici)1096-987x(199709)18:12<1463::aid-jcc4>3.0.co;2-h.
- Huang, Jing, and Alexander D. MacKerell Jr. "CHARMM36 All-Atom Additive Protein Force Field: Validation Based on Comparison to NMR Data." *Journal of Computational Chemistry* 34, no. 25 (July 6, 2013): 2135–45. https://doi.org/10.1002/jcc.23354.
- 29. Ishikita, Hiroshi. "14 Contributions of Protein Environment to the Reduction Potentials of Flavin-Containing Proteins." In *Handbook of Flavoproteins*, 321– 34. DE GRUYTER, 2013. http://dx.doi.org/10.1515/9783110298345.321.
- Ishikita, Hiroshi. "Influence of the Protein Environment on the Redox Potentials of Flavodoxins from Clostridium Beijerinckii." *Journal of Biological Chemistry* 282, no. 35 (August 2007): 25240–46. https://doi.org/10.1074/jbc.m702788200.
- 31. Iyanagi, Takashi. "Molecular Mechanism of Metabolic NAD(P)H-Dependent Electron-Transfer Systems: The Role of Redox Cofactors." *Biochimica et Biophysica Acta (BBA) - Bioenergetics* 1860, no. 3 (March 2019): 233–58. https://doi.org/10.1016/j.bbabio.2018.11.014.
- Jarzynski, C. "Nonequilibrium Equality for Free Energy Differences." *Physical Review Letters* 78, no. 14 (April 7, 1997): 2690–93. https://doi.org/10.1103/physrevlett.78.2690.
- 33. Jorgensen, William L., Jayaraman Chandrasekhar, Jeffry D. Madura, Roger W. Impey, and Michael L. Klein. "Comparison of Simple Potential Functions for Simulating Liquid Water." *The Journal of Chemical Physics* 79, no. 2 (July 15, 1983): 926–35. https://doi.org/10.1063/1.445869.
- 34. Kasim, Mumtaz, and Richard P. Swenson. "Alanine-Scanning of the 50's Loop in the Clostridium Beijerinckii Flavodoxin: Evaluation of Additivity and the Importance of Interactions Provided by the Main Chain in the Modulation of the Oxidation–Reduction Potentials." *Biochemistry* 40, no. 45 (October 19, 2001): 13548–55. https://doi.org/10.1021/bi011587c.
- 35. Kılıç, Murat, and Bernd Ensing. "First and Second One-Electron Reduction of Lumiflavin in Water—A First Principles Molecular Dynamics Study." *Journal of Chemical Theory and Computation* 9, no. 9 (August 15, 2013): 3889–99. https://doi.org/10.1021/ct400088g.
- Lawrenz, Morgan, Jeff Wereszczynski, Juan Manuel Ortiz-Sánchez, Sara E. Nichols, and J. Andrew McCammon. "Thermodynamic Integration to Predict Host-Guest Binding Affinities." *Journal of Computer-Aided Molecular Design* 26, no. 5 (February 16, 2012): 569–76. https://doi.org/10.1007/s10822-012-9542-5.

- Legrand, Yves-Marie, Mark Gray, Graeme Cooke, and Vincent M. Rotello.
 "Model Systems for Flavoenzyme Activity: Relationships between Cofactor Structure, Binding and Redox Properties." *Journal of the American Chemical Society* 125, no. 51 (November 27, 2003): 15789–95. https://doi.org/10.1021/ja036940b.
- 38. Leys, David, and Nigel S Scrutton. "Sweating the Assets of Flavin Cofactors: New Insight of Chemical Versatility from Knowledge of Structure and Mechanism." *Current Opinion in Structural Biology* 41 (December 2016): 19–26. https://doi.org/10.1016/j.sbi.2016.05.014.
- 39. Li, Wenjin, Ilona B. Baldus, and Frauke Gräter. "Redox Potentials of Protein Disulfide Bonds from Free-Energy Calculations." *The Journal of Physical Chemistry B* 119, no. 17 (April 21, 2015): 5386–91. https://doi.org/10.1021/acs.jpcb.5b01051.
- 40. Ludwig, Martha L., Katherine A. Pattridge, Anita L. Metzger, Melinda M. Dixon, Mesut Eren, Yucheng Feng, and Richard P. Swenson. "Control of Oxidation–Reduction Potentials in Flavodoxin from Clostridium Beijerinckii: The Role of Conformation Changes,." *Biochemistry* 36, no. 6 (February 1, 1997): 1259–80. https://doi.org/10.1021/bi962180o.
- Ludwig, Martha L., Lawrence M. Schopfer, Anita L. Metzger, Katherine A. Pattridge, and Vincent Massey. "Structure and Oxidation-Reduction Behavior of 1-Deaza-FMN Flavodoxins: Modulation of Redox Potentials in Flavodoxins." *Biochemistry* 29, no. 45 (November 13, 1990): 10364–75. https://doi.org/10.1021/bi00497a011.
- 42. Madaj, Rafal, Bartlomiej Gostynski, Roza Pawlowska, and Arkadiusz Chworos. "Tissue-Nonspecific Alkaline Phosphatase (TNAP) as the Enzyme Involved in the Degradation of Nucleotide Analogues in the Ligand Docking and Molecular

Dynamics Approaches." *Biomolecules* 11, no. 8 (July 27, 2021): 1104. https://doi.org/10.3390/biom11081104.

- 43. Mueller, Robyn M., Michael A. North, Chee Yang, Sanchita Hati, and Sudeep Bhattacharyya. "Interplay of Flavin's Redox States and Protein Dynamics: An Insight from QM/MM Simulations of Dihydronicotinamide Riboside Quinone Oxidoreductase 2." *The Journal of Physical Chemistry B* 115, no. 13 (April 7, 2011): 3632–41. https://doi.org/10.1021/jp1107922.
- Munro, Andrew W, Paul Taylor, and Malcolm D Walkinshaw. "Structures of Redox Enzymes." *Current Opinion in Biotechnology* 11, no. 4 (August 2000): 369–76. https://doi.org/10.1016/s0958-1669(00)00112-9.
- North, Michael A., Sudeep Bhattacharyya, and Donald G. Truhlar. "Improved Density Functional Description of the Electrochemistry and Structure–Property Descriptors of Substituted Flavins." *The Journal of Physical Chemistry B* 114, no. 46 (October 20, 2010): 14907–15. https://doi.org/10.1021/jp108024b.
- Nosé, Shuichi, and M.L. Klein. "Constant Pressure Molecular Dynamics for Molecular Systems." *Molecular Physics* 50, no. 5 (December 10, 1983): 1055– 76. https://doi.org/10.1080/00268978300102851.
- 47. Parrinello, M., and A. Rahman. "Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method." *Journal of Applied Physics* 52, no. 12 (December 1, 1981): 7182–90. https://doi.org/10.1063/1.328693.
- 48. Sattelle, Benedict M., and Michael J. Sutcliffe. "Calculating Chemically Accurate Redox Potentials for Engineered Flavoproteins from Classical Molecular Dynamics Free Energy Simulations." *The Journal of Physical Chemistry A* 112, no. 50 (October 2, 2008): 13053–57. https://doi.org/10.1021/jp803859j.

- Seeliger, Daniel, and Bert L. de Groot. "Protein Thermostability Calculations Using Alchemical Free Energy Simulations." *Biophysical Journal* 98, no. 10 (May 2010): 2309–16. https://doi.org/10.1016/j.bpj.2010.01.051.
- 50. Shirts, Michael R., Eric Bair, Giles Hooker, and Vijay S. Pande. "Equilibrium Free Energies from Nonequilibrium Measurements Using Maximum-Likelihood Methods." *Physical Review Letters* 91, no. 14 (October 2, 2003). https://doi.org/10.1103/physrevlett.91.140601.
- 51. Silvestri, Giuseppe, Federica Arrigoni, Francesca Persico, Luca Bertini, Giuseppe Zampella, Luca De Gioia, and Jacopo Vertemara. "Assessing the Performance of Non-Equilibrium Thermodynamic Integration in Flavodoxin Redox Potential Estimation." *Molecules* 28, no. 16 (August 11, 2023): 6016. https://doi.org/10.3390/molecules28166016.
- Smith, Ward W., Roger M. Burnett, George D. Darling, and Martha L. Ludwig.
 "Structure of the Semiquinone Form of Flavodoxin from Clostridium MP." Journal of Molecular Biology 117, no. 1 (November 1977): 195–225. https://doi.org/10.1016/0022-2836(77)90031-6.
- 53. Steiner, Denise, Chris Oostenbrink, and Wilfred F. van Gunsteren. "Calculation of the Relative Free Energy of Oxidation of Azurin at pH 5 and pH 9." *Journal of Computational Chemistry* 33, no. 17 (April 24, 2012): 1467–77. https://doi.org/10.1002/jcc.22972.
- 54. Stockman, Brian J., Thomas E. Richardson, and Richard P. Swenson. "Structural Changes Caused by Site-Directed Mutagenesis of Tyrosine-98 in Desulfovibrio Vulgaris Flavodoxin Delineated by 1H and 15N NMR Spectroscopy: Implications for Redox Potential Modulation." *Biochemistry* 33, no. 51 (December 1994): 15298–308. https://doi.org/10.1021/bi00255a011.

- 55. Swenson, Richard P., and Grigorios D. Krey. "Site-Directed Mutagenesis of Tyrosine-98 in the Flavodoxin from Desulfovibrio Vulgaris (Hildenborough): Regulation of Oxidation-Reduction Properties of the Bound FMN Cofactor by Aromatic, Solvent, and Electrostatic Interactions." *Biochemistry* 33, no. 28 (July 1, 1994): 8505–14. https://doi.org/10.1021/bi00194a015.
- 56. Ullmann, G. Matthias, Verónica I. Dumit, and Elisa Bombarda. "15 Methods Based on Continuum Electrostatics and Their Application to Flavoproteins – a Review." In *Handbook of Flavoproteins*, 335–60. De Gruyter, 2013. http://dx.doi.org/10.1515/9783110298345.335.
- 57. Van Der Spoel, David, Erik Lindahl, Berk Hess, Gerrit Groenhof, Alan E. Mark, and Herman J. C. Berendsen. "GROMACS: Fast, Flexible, and Free." *Journal of Computational Chemistry* 26, no. 16 (October 6, 2005): 1701–18. https://doi.org/10.1002/jcc.20291.
- 58. Watt, William, Alexander Tulinsky, Richard P. Swenson, and Keith D. Watenpaugh. "Comparison of the Crystal Structures of a Flavodoxin in Its Three Oxidation States at Cryogenic Temperatures." *Journal of Molecular Biology* 218, no. 1 (March 1991): 195–208. https://doi.org/10.1016/0022-2836(91)90884-9.
- 59. Zhou, Zhimin, and Richard P. Swenson. "Electrostatic Effects of Surface Acidic Amino Acid Residues on the Oxidation-Reduction Potentials of the Flavodoxin from Desulfovibrio Vulgaris (Hildenborough)." *Biochemistry* 34, no. 10 (March 1, 1995): 3183–92. https://doi.org/10.1021/bi00010a007.
- 60. Zhou, Zhimin, and Richard P. Swenson. "The Cumulative Electrostatic Effect of Aromatic Stacking Interactions and the Negative Electrostatic Environment of the Flavin Mononucleotide Binding Site Is a Major Determinant of the Reduction Potential for the Flavodoxin from Desulfovibrio Vulgaris [Hildenborough]." *Biochemistry* 35, no. 50 (January 1, 1996): 15980–88. https://doi.org/10.1021/bi962124n.

Chapter 3

Computational Characterization of the Ring Shuttling Process in Amidebased [2]Rotaxanes: the Impact of the Solvent

3.1 Abstract

In this study, we investigate the dynamics and stability of amide-based [2]rotaxanes, molecular machines capable of harnessing chemical reactions to produce unidirectional motion. These systems consist of a macrocycle that shuttles along an axle, stabilized by bulky terminal groups, with the shuttling mechanism strongly influenced by the surrounding solvent environment. Through computational analysis of rotaxanes in different solvents (DMSO, ACN, and CHCl₃) we reveal how solvent polarity and hydrogen bonding capacity modulate the energy landscape and dynamic behavior of these molecular machines. In polar solvents such as DMSO and ACN, the axle adopts a more extended conformation, which makes the binding stations more accessible, facilitating faster shuttling. Conversely, although the extended conformation of the axle is energetically favored, in CHCl₃ the axle tends to collapse more readily in order to maximize intramolecular interactions between the two stations, using the macrocycle as a bridge. This collapsed conformation is promoted by the solvent's inability to sequester hydrogen bond donors and create steric hindrance between the ring and the axle. Furthermore, in CHCl₃, the distal binding station is more stabilized compared to the proximal station, due to a stronger interaction between the distal fumaramide and the carboxylate group. While this interaction occurs in all solvents, it is particularly stable in CHCl₃, where the absence of hydrogen bond acceptors allows the amide donors on the macrocycle to interact directly with the carboxylate, stabilizing it in an optimal geometry for binding the distal station. By leveraging solvent effects and employing well-tempered metadynamics (WT-MTD), this work provides valuable insights for future optimization of rotaxanes as molecular machines and their potential applications in nanotechnology and molecular-scale devices.

3.2 Introduction

The interest in synthetic molecular machines dates back almost two centuries, when Robert Brown, through experiments, discovered that molecular motion is governed by random movements. The behavior of these motions is governed by thermodynamic laws, which explain how both synthetic and natural molecular motors acquire, process, and release energy, and how they perform work. A key role is played by the second law of thermodynamics, as it provides a connection between the physical laws of reversibility and the irreversible nature of the universe [22]. This law is critical in molecular machines, as it supplies the theoretical framework necessary to design molecular machines that can harness Brownian motion. This motion cannot be avoided, as molecular objects will always experience it at specific temperatures, determined by their kinetic energy. Even an isolated molecule would undergo Brownian motion due to heat transfer from surrounding walls via electromagnetic radiation.

Molecular machines are chemical systems where an external input (whether chemical, electrical, or thermal) induces controlled motion of molecular or submolecular components, thereby generating work. This category includes not only molecular motors, which are the focus of this project, but also molecular switches [21]. A molecular switch is a machine where the relative positions of components change in response to the switch state, while molecular motors cause movement based on the components' trajectory. The key distinction lies in the fact that molecular switches do not cumulate work, whereas molecular motors, by cumulating work through their cycles, can drive systems away from equilibrium. Inspired by biological molecular motors, which play critical roles in life processes, synthetic molecular motors have been the focus of research for the past few decades. Many biological motors, including ATP synthase and photosynthetic systems, use energy derived from ATP hydrolysis, proton gradients, or sunlight to drive stochastic



Figure 3.1: Kinetic and thermodynamic control of a general catalized reaction [7].

processes far from equilibrium. From studying these biological systems, we gain fundamental insights that help in designing synthetic molecular motors capable of optimal operation [7, 21].

In 2016, the Nobel Prize in Chemistry was awarded to Jean-Pierre Sauvage, Fraser Stoddart, and Ben Feringa for their contributions to molecular motor synthesis. These motors perform useful work, such as controlled motion, and are key to moving systems far from equilibrium. However, molecular machines operate in environments where classical parameters such as gravity and inertia are negligible, while viscosity and Brownian motion are dominant [21]. Molecular motors leverage thermal fluctuations to produce work through a mechanism known as ratcheting. Chemical reactions are typically under thermodynamic or kinetic control (fig. 3.1). Under thermodynamic control, the product distribution depends on the difference in Gibbs free energy of the species involved. Under kinetic control, the reaction proceeds towards the free energy minimum separated by the lowest free energy activation barrier. Reactions under this control require an additional energy input such as light, a chemical reaction, or other sources, which acts orthogonally to the target reaction. This allows for movement across the potential energy



Figure 3.2: Generic example of PES with ratchet mechanism [7]. Along reaction coordinate 1, the target reaction is present, while along coordinate 2, the orthogonal reaction is present

surface (PES) along reactive pathways that would be inaccessible under typical thermodynamic or kinetic control. In other words, a ratchet mechanism harnesses energy from a *fuel-to-waste* reaction (as represented by reaction coordinate 2 in figure 3.2) to influence the distribution of products in the target reaction (reaction coordinate 1 in Figure 3.2). In this case, the energy released by the fuel-to-waste reaction is used to drive the target reaction forward. Essentially, the orthogonal reaction (coordinate 2) modulates the distribution of products in the target reaction (coordinate 1). In synthetic chemistry, examples of reactions leveraging ratchet mechanisms are scarce, but in nature, there are many cases that employ this strategy. For instance, ATP synthase (Figure 3.3) utilizes a proton gradient as an orthogonal input to synthesize ATP, an otherwise thermodynamically and kinetically unfavorable reaction.

The *information ratchets* generate directional motion by regulating the relative rates of forward and reverse processes based on molecular conformations or co-conformations. This phenomenon is known as *kinetic gating* [2, 6, 7, 26]. In these systems, directional motion is achieved through the interplay between stochastic processes (random, thermally-driven fluctuations) and energy injection processes (i. e. a chemical reaction). The idea is that random motion alone, such as Brownian motion, does not inherently have



Figura 3.3: ATP synthase as biological molecular ratchet [7].

a preferred direction; molecules move randomly in all directions. However, when this random motion is coupled with a process that supplies energy to the system, the randomness can be biased or rectified, leading to directional movement. In the context of an information ratchet, the stochastic process could involve

the random fluctuations of a molecule between different conformational states (shapes or positions of a molecule), or the movement of a molecular component, such as a macrocycle on an axle in a rotaxane system [4]. Normally, these fluctuations would have no preferred direction. However, by coupling this random movement to an energy supply, it influences which molecular states are more likely to occur. This information transfer refers to the fact that the energy supply process "transfers" its influence to the stochastic process, biasing the random motion and allowing it to become directional. In other words, the energy input from the dissipative process guides the stochastic process, ensuring that the random fluctuations are rectified into a specific, useful motion, such as moving the macrocycle from one station to another on the axle of a rotaxane. The key here is that directionality is not purely a result of the input energy directly driving the motion, as in a classical motor. Instead, the energy modulates the probabilities of different outcomes (i.e., it influences which states the molecule is more likely to adopt after a random fluctuation),

thus rectifying the random motion into a net forward or backward direction. This is why it is referred to as an information ratchet: the system uses the energy input to "process" information about the current state of the system and biases the subsequent molecular transitions to achieve directional movement.

The introduction to ratchet mechanisms provides a foundational framework for understanding how molecular machines operate by harnessing stochastic processes to achieve directional work. This concept becomes particularly relevant when analyzing the behavior of synthetic molecular machines, such as rotaxanes, in different solvent environments. The [2]rotaxane system developed by S. Borsley and colleagues [8, 10], which operates as an information ratchet, exemplifies how the principles of ratcheting can be applied to guide molecular motion. In this system, the movement of a macrocycle along an axle is driven by a fuel-to-waste reaction and involves precise control over the molecular interactions and kinetics of the system. The axle features two fumaramidic binding site where the ring can reside, stabilized by hydrogen bonding (fig. 3.4). A key feature of this system is a carboxylate group, which catalyzes fuel-to-waste reactions critical for ratcheting activity. The fuel-to-waste reaction involves carbodiimides, which provide robustness and minimize side reactions unfavorable to the ratchet's activity (fig. 3.5). The energy released by this reaction drives the ring toward the distal binding site. This directional movement is driven by kinetic asymmetry, due exclusively to the position of the carboxylate group being unequally distant from the binding sites.

The solvent environment plays a crucial role in dictating the behavior of molecular machines like the Borsley's rotaxane, particularly in how it influences the kinetic gating and diffusion processes [15, 17, 23, 30]. Each solvent can modulate factors such as viscosity, polarity, and hydrogen bonding capabilities, which in turn affect the energy landscape and the rates of transition between molecular states. This interaction between the molecular machine and its solvent environment directly relates to the principles of information ratcheting, where the control over stochastic processes is key to achieving directed, progressive motion. For instance, in solvents with higher viscosity, the diffusion coefficient decreases, which slows the rate at which the macrocycle can move along the



Figure 3.4: Structure of machine showing dynamic exchange of the macrocycle between the two fumaramide binding sites on the axle (proximal station on the left, distal station on the right) [8]

axle. This affects the timescale of the ratcheting process, and the kinetic differences between solvent environments could lead to observable changes in the overall efficiency and speed of the molecular machine. Additionally, solvents that stabilize or destabilize certain conformations of the rotaxane may influence the energy barriers the system encounters, thus altering the directionality and success of the fuel-to-waste reaction in driving the macrocycle away from equilibrium.

By studying the rotaxane's behavior in various solvents with different properties, researchers can gain insights into how solvent-dependent factors like hydrogen bonding interactions (between the macrocycle and axle), polar interactions, and solvation energy affect the system's ability to act as a molecular motor. This solvent-dependent behavior can be directly tied to the broader concept of ratchet mechanisms, where the efficiency of rectifying stochastic motion into useful work is dependent not only on the design of the molecular machine but also on the external environment it operates in. The ability to leverage the asymmetry in its structure, which can be enhanced or inhibited depending on the solvent environment. Understanding these solvent effects is essential for optimizing the design and performance of molecular machines based on similar principles. To systematically investigate these effects, MD simulations combined with WT-MTD were employed. This computational approach allows for the exploration of the thermodynamics of the system in the very initial step under equilibrium conditions, in the absence of fuel. By simulating the rotaxane in different solvent environments, we aim to

assess how each solvent influences the free energy landscape and the transition states between the binding sites.



Figure 3.5: The chemomechanical cycle of the machine's operation involves transitions driven by the attachment and removal of a transient, hydrolyzable barrier. The bottom structures represent this barrier, which influences the machine's motion. The macrocycle's movement is dictated by differences in the rate of barrier attachment or removal depending on its position. The light gray arrows highlight reaction pathways that are less probable. Each step in the rotaxane-catalyzed reaction (fuel to waste) is shown as reversible to align with the principle of microscopic reversibility. However, in practice, the thermodynamic equilibrium between the carbodiimide fuel and the urea waste significantly favors the urea formation, driving the macrocycle toward the distal station on the right, resulting in a ratcheting motion [8].

3.3 Materials and methods

3.3.1 Simulation setup

The initial structure was derived from X-ray crystallography data (CCDC deposition code: 2191007), which features the macrocycle in a proximal position and a t-butyl ester derivative on the carboxylate moiety. After refining the structure, which involved removing residual solvent molecules and the t-butyl derivative, the model system was parameterized using the Generalized Amber Force Field 2 (GAFF2) [29]. Partial charges for the atomistic models were calculated using the RESP method [5], computed at the Hartree-Fock (HF) level of theory with the 6-31G basis set, as implemented in Gaussian16 [16]. The AmberTools23 package [13] was utilized for system parameterization. Solvent parameters were sourced from the virtualchemistry.org database [12, 28]. A potassium ion was added to the simulation box to neutralize the negative charge of the carboxylate. To prevent artificial electrostatic interactions between the ion and the carboxylate, a potential wall was applied around the carboxylate, with a radius of 2 nm and a force constant of 2 kcal/mol·Å², which exclusively affected the potassium ion. All systems were initially minimized using the steepest-descent algorithm and subsequently equilibrated under NPT conditions (constant number of particles, pressure, and temperature). Production runs were performed in the NPT ensemble for up to 1 µs. Temperature and pressure were maintained at 300 K and 1 bar using the v-rescale thermostat [11] and the Parrinello-Rahman barostat [24] (tab 3.1). Long-range electrostatic interactions were handled using the PME method [14] with a cutoff of 1.2 nm. Van der Waals interactions were truncated at 1.2 nm. The LINCS algorithm [18] was used to constrain all hydrogen-involving bonds, and the leapfrog integrator was used to propagate dynamics with a 2 fs time step. All simulations were carried out using GROMACS 2022.3 [1], with the system patched using PLUMED version 2.8.1 [25].

3.3.2 WT-MTD parameters

Early simulations revealed that the motion of this rotaxane system is highly complex, involving at least three coupled movements: the translational motion of the macrocycle along the axis, conformational changes in the macrocycle, and conformational changes in the axis itself. This complexity arises primarily from the numerous degrees of freedom associated with the pentane linker between the fumaramide stations. Given the complexity of the system and the limitations of unbiased molecular dynamics in thoroughly sampling the phase space, we employed well-tempered metadynamics to compute free energy landscapes.

To better understand the translational process of the rotaxane, we selected two main collective variables (CVs) to accurately represent the reaction coordinate: distance X and distance AB. The AB variable measures the distance between two key atoms, Cp (on the stopper) and Cd (on the distal station), as shown in fig. 3.6A. Distance X provides the relative position of the macrocycle along the axis as a function of the axis extension, using the following equation:

$$X = \frac{P_{d1} \cdot AB_{max}}{AB}$$

Here, d1 represents the distance between the center of mass (COM) of the four nitrogen atoms in the macrocycle and the Cp atom, so P_{d1} is its projection along AB distance; AB_{max} (set to 2.7 nm, based on preliminary simulations) is the maximum achieved value of distance AB. Additionally, the conformational properties of the macrocycle during the simulations were analyzed using the dihedral angle ϕ , which has been successfully used in prior studies [17] to discriminate between significant macrocycle configurations (fig. 3.6B).

The metadynamics bias was updated every 500 time steps (equivalent to 1 ps of simulation time) with Gaussians of initial height 0.5 kJ/mol and widths of 0.04 nm for X

and 0.08 nm for AB, with a bias factor of 20 for all systems. Each system was simulated for at least 3 μ s, with simulations extending up to 11 μ s in CHCl₃ (tab. 3.1), allowing us to observe multiple shuttling events between binding stations. This approach allowed for an efficient exploration of the free energy landscape, capturing the key dynamics of the rotaxane's shuttling process across various solvent environments.

Solvent	H-bond acceptor ability (β)	Number of atoms	Box size (nm)	WT-MTD simulation time (µs)
DMSO	0.76	17434	5.95 x 5.95 x 5.95	6
ACN	0.4	10484	5.45 x 5.45 x 5.45	3
CHCl ₃	0.1	6069	5.70 x 5.70 x 5.70	11

Table 3.1: Details of molecular systems simulated in this work with the H-bond acceptor ability of each solvent.



Figure 3.6: **A.** *Structure of the rotaxane and the coarse variables* (P_{d1} , AB) *chosen to define the transition coordinate.* **B.** *Isomerization of the macrocycle.*

3.4 Results and discussion

Three solvents with different properties were used in this study: dimethylsulphoxide (DMSO), acetonitrile (ACN) and trichloromethane (CHCl₃), so that we could assess how solvent properties influence the dynamics of rotaxane. The properties of these solvents are highlighted in table 3.1. Given that the behavior of the rotaxane is heavily influenced by the hydrogen bonds formed between the macrocycle and the binding sites, we selected solvents with varying β parameters (indicating their hydrogen-bond acceptor ability) [20] for comparison: chloroform (CHCl₃, $\beta = 0.10$), acetonitrile (ACN, $\beta = 0.40$), and dimethyl sulfoxide (DMSO, $\beta = 0.76$). Since ACN, and especially DMSO, have a significant hydrogen-bond acceptor capacity, they may interact locally with the axle by forming hydrogen bonds, thereby competing with the macrocycle's binding interactions. Conversely, CHCl₃ is relatively "inert" in this context, as his lower β value suggest minimal interaction with the axle, and their contribution to the shuttling dynamics is primarily due to residual solvophobic effects.

For clarity and ease of understanding, all representative structures discussed in the following section are provided in the *Supplementary figures and tables section*, as well as the table (tab. S3.1) of energy value related the aforementioned structures. This allows for a more comprehensive visualization of the molecular configurations referenced in the analysis.

3.4.1 Dimethylsulphoxide

As can be seen in fig 3.7A, in dimethylsulphoxide, rotaxane has two main minima, A1 and A3, at an X value of about 0.6 and 2.0 nm respectively, where the ring resides on the two fumaramide stations. The two structures have approximately the same energy, where the extention of the axis AB is about 1.9 nm. In the distal station, the macrocycle interacts directly with the carboxylate, which, due to a twist in the aliphatic chain, binds via hydrogen bond to an amine of the distal fumaramide, bridging the interaction with the macrocycle. Thus, two amines of the ring interact with the carboxylate, while the other two anchor the ring in this position by binding the closest fumaramidic carbonyl to the center of the axle. In both minima the ring is mainly stabilised in the *boat* configuration, transiently exploring the chair one (fig. 3.7A). The ring oscillates between these two configurations, passing over the carboxylate with which it forms transient hydrogen bonds: this configuration (A2) is less stable than those on the binding stations of about 2.5 kcal/mol. As it passes through the aliphatic chain, the ring also explores metastable 'chair' configurations, maximising the apolar interactions between the axis and phenylene moiety of the ring and the extension of the aliphatic chain itself. When the macrocycle is at the distal station, the axis can occur in a metastable U-shape fold (AB = 1.2 nm), where the carboxylate moiety points toward the bulk (A3a). In this conformation, the rotaxane is destabilized by 1.6 kcal/mol respect to the absolute minimum.

Interestingly, as the macrocycle transitions between A1 and A3, solvophobic interactions between the ring and the axle play a significant role, encouraging closer contacts between these two components. At the same time, interactions with DMSO favors a more extended conformation of the axle. This balance between solvophobic effects (driving the interaction between the ring and axle) and solvent-mediated stabilization leads to a more stretched configuration of the axle at the binding stations, where hydrogen bonds further stabilize the macrocycle's position.

At the binding stations, the macrocycle forms an average of 2-4 hydrogen bonds with the axle, depending on the extent of hydrogen bonding between the macrocycle, the axle, and

DMSO (fig. 3.7B). This dynamic equilibrium results in no clear preference for either binding station.



Figure 3.7: **A.** Free energy landscape of the ring position as a function of the axis extension (left) and ring conformation (right). The metastable states in the left plot are highlighted, as well as the path connecting the two binding stations. The reaction path was computed using the Nudged Elastic Band (NEB) algorithm [3], which identifies the minimum energy path between the proximal and distal binding stations, providing insight into the transition states and energy barriers encountered during the shuttling process of the macrocycle. **B.** Free energy landscape of the ring position as a function of the average number of H-bond between ring and axis (left), ring and solvent (middle), axis and solvent (right).

3.4.2 Acetonitrile

The metastable states of rotaxane along the reaction pathway connecting the two binding sites in acetonitrile are similar to those in DMSO due to the polarities of the two solvents, as reflected by the two-dimensional free energy landscape. The system in this solvent exhibits similar behaviour to that in a solvent with a more pronounced polarity such as DMSO, except for a small destabilisation near the transition state. The minima associated with representative spatial configurations of the system in acetonitrile are highlighted in fig. 3.8A. The B1, B2, and B2a structures in ACN are directly analogous to A1, A3, and A3a in DMSO. However, in ACN, the macrocycle exhibits a metastable proximal position with a contracted axle configuration (B1a). Overall, the energy difference between the extended and contracted axle conformations is smaller in ACN, a less polar solvent compared to DMSO. The free energy difference between these conformations in ACN is approximately 1.0 kcal/mol, reflecting the reduced impact of solvent polarity on the stability of the extended structure.

With regard to the ring conformation, the chair structure is destabilised, with the boat structure preponderant at both stations and, in general, along the entire axis (fig. 3.8B). The solvophobic interactions are less preponderant during shuttling, making the free energy trend of the path steeper close to the aliphatic chain than in DMSO, accompanied by a slight contraction of the axis and a reduction of H-bond interaction between axle and solvent. As a result, the overall energy landscape in ACN supports a dynamic system where both hydrogen bonding and solvophobic interactions influence the rotaxane's behavior, albeit with a reduced impact compared to more polar solvents like DMSO.



Figure 3.8: **A.** Free energy landscape of the ring position as a function of the axis extension (left) and ring conformation (right). The metastable states in the left plot are highlighted, as well as the path connecting the two binding stations. The reaction path was computed using the Nudged Elastic Band (NEB) algorithm [3], which identifies the minimum energy path between the proximal and distal binding stations, providing insight into the transition states and energy barriers encountered during the shuttling process of the macrocycle. **B.** Free energy landscape of the ring position as a function of the average number of H-bond between ring and axis (left), ring and solvent (middle), axis and solvent (right).

3.4.3 Trichloromethane

As can be seen in fig. 3.9, in trichloromethane, the behaviour of rotaxane differs significantly in several aspects from the other two polar solvents. Although the configurations for the proximal and distal stations (C1 and C3) are similar to those in the other solvents, the free energy between the two is about 2.2 kcal/mol. This notable difference is largely due to the more significant role of hydrogen bonding in the less polar environment of chloroform. In this solvent, the system tends to maximize hydrogen bonding interactions, as chloroform does not effectively compete for hydrogen bonds with the macrocycle or the axle. At the distal station, the macrocycle forms up to 4 stable hydrogen bonds, 2 of which are with the carboxylate group. The carboxylate is further stabilized by an additional hydrogen bond with the amine of the distal fumaramide. This results in a higher stabilization of the distal position compared to the proximal one. The lack of a polar solvent to mediate compensatory hydrogen bonds between the macrocycle and the axle forces the system to maximize its internal hydrogen bonding network. This contributes to the stronger stabilization of the distal station, where the macrocycle interacts more robustly with both the carboxylate and the axle, in contrast to the more competitive environment found in polar solvents like DMSO or ACN. In chloroform, without the interference of solvent-mediated interactions, these hydrogen bonds are not transient but remain well-defined, reinforcing the stability of the macrocycle at the distal station (fig. 3.9).

The pathway connecting the C1 and C3 minima in chloroform presents significant differences compared to that in more polar solvents, indicating an alternative shuttling mechanism that is primarily driven by hydrogen bonds. In particular, the increased tendency of the axle to adopt a contracted conformation (fig. 3.9A) is observed throughout the transition pathway, resulting in the formation of at least three metastable states: C2, C3a, and C3b. The C2 state, located near the transition state, involves the macrocycle positioned between the proximal station and the carboxylate. This configuration is less stable than C1 by approximately 1 kcal/mol, likely due to entropic constraints that restrict both the axle and the macrocycle to specific spatial conformations. During the transition

from C2 to C3 (the distal station with an extended axle), the macrocycle initially forms three hydrogen bonds at C2 (compared to the two observed in other solvents) reflecting the impact of the apolar chloroform environment on hydrogen bonding. As the macrocycle progresses through the transition, it adopts structures with a contracted axle, increasing the number of hydrogen bonds to up to four, as shown in the H-bond distribution (fig. 3.9). At the distal station, however, the macrocycle can explore configurations with fewer hydrogen bonds between the ring and the axle. In these contracted structures, stabilization arises not only from the solvent's low hydrogen bonding capability but also from hydrogen bonds between the macrocycle and the proximal fumaramide. A rotation of one of the amide groups of the macrocycle bridges the two stations, stabilizing the contracted form, where the carboxylate is no longer bound to the macrocycle but is instead exposed to the solvent. Interestingly, despite the average formation of only about 2.5 hydrogen bonds between the macrocycle and the axle in this contracted configuration, the energy is comparable to that of the extended structure, which forms four hydrogen bonds directly with the carboxylate. This suggests a potential entropic compensation mechanism, where the reduction in hydrogen bond numbers is offset by greater conformational freedom in the system. The balance between enthalpic and entropic contributions may therefore play a critical role in the stability of the contracted conformations in chloroform. Finally, the unfolding of the axis itself allows the system to reach the C3 minimum, with the carboxylate then able to form stable interactions with the axis and macrocycle. A further conformational analysis reveals that the macrocycle undergoes a boat-chair interconversion several times, as shown in fig. 3.9: when the axis folds the ring tends to assume a chair-like conformation in the C3a and C3b minima, while in the cases of the C1a, C1b, C1c minima the ring is forced into more 'distorted' configurations ($\phi = 50^\circ$), compensated by a slight increase in the average Hbond number that makes these minima energetically similar to C1. In summary, the transition from one station to the other involves a large axis folding, which, in an apolar solvent, makes the shuttling pathway different from that seen in more polar solvents such as DMSO and ACN, aimed at maximising the hydrogen bonds between axis and ring.



Figure 3.9: Free energy landscape of the ring position as a function of the axis extension (upper left), ring conformation (upper right), and average number of H-bond between ring and axis (bottom). The metastable states in the upper left plot are highlighted, as well as the path connecting the two binding stations. The reaction path was computed using the Nudged Elastic Band (NEB) algorithm [3], which identifies the minimum energy path between the proximal and distal binding stations, providing insight into the transition states and energy barriers encountered during the shuttling process of the macrocycle.

3.4.4 Effects on the shuttling mechanism

The behaviour of the rotaxane in DMSO, ACN, and CHCl₃ highlights the profound impact of solvent properties on the shuttling mechanism. The lubricant effect of the solvent is particularly notable in DMSO and ACN, where the high polarity reduces the contribution of hydrogen bonds between the ring and the chain, effectively lowering the free-energy barrier for translocation (fig. 3.10A). This effect is especially pronounced in DMSO, which, due to its strong hydrogen bonding capacity, weakens the macrocycleaxle interactions and facilitates faster shuttling. In CHCl₃, however, the lower polarity increases the strength of these hydrogen bonds, resulting in a higher energy barrier and slower shuttling. This can be seen in the steeper energy profile for CHCl₃ (Fig. 3.10A), where the solvent's inability to efficiently weaken hydrogen bonds leads to a more constrained pathway for the macrocycle. The flexibility of the axle and the corresponding conformational changes in the macrocycle play a crucial role in shaping the energy landscape. In both DMSO and ACN, the axle is less flexible and more extended, which allows the macrocycle to follow a relatively straightforward pathway between the proximal and distal stations. The macrocycle predominantly adopts boat and chair conformations along this path (Fig. 3.10C), and the weaker hydrogen bonding in these solvents results in a moderate free-energy barrier. In contrast, CHCl₃ introduces greater flexibility to both the axle and the macrocycle. In fig. 3.10B, the trend clearly demonstrates a progressive increase in the stability of the contracted conformation relative to the extended form as solvent polarity decreases. This is evident in the free energy difference between the two conformations, which, when integrated over the entire simulation, decreases from 1.5 kcal/mol in DMSO to 0.5 kcal/mol in CHCl₃. This observation highlights the influence of solvent polarity on the relative stability of the contracted and extended forms, with less polar solvents favoring the contracted state. This increased flexibility alters the shuttling pathway, forcing the rotaxane to pass through transition states where it becomes constrained in entropically unfavourable conformations. These constraints, along with stronger hydrogen bonding interactions, raise the activation energy, particularly near the distal station, where the macrocycle faces

more significant interactions to transition along the axle. Consequently, the more complex energy profile in CHCl₃ reflects the difficulty of overcoming these additional barriers compared to the smoother paths in DMSO and ACN. Regarding the equilibrium between the two stations, both DMSO and ACN display a relatively similar energy landscape, with comparable stability at the proximal and distal stations, reflecting the experimental data [10]. In contrast, in CHCl₃, the distal station becomes significantly more stable, as the macrocycle forms stronger interactions with the chain in this position. This difference in stability suggests that in CHCl₃, the system is biased toward maintaining the macrocycle in the distal station, potentially affecting the overall efficiency of the rotaxane's catalytic cycle. Such stabilization of the distal station could slow down catalytic cycles that depend on efficient shuttling between stations.

The carboxylate group plays a crucial role in stabilizing the macrocycle, especially in CHCl₃. In this non-polar solvent, the carboxylate forms an average of 1.5 additional hydrogen bonds at the distal station compared to the proximal one (fig. 3.11A-B). This stronger interaction stabilizes the distal position in CHCl₃ much more than in DMSO or ACN, where only 0.5 additional hydrogen bonds are formed in the distal state. The negative charge of the carboxylate is effectively shielded in CHCl₃, which enhances its interaction with the macrocycle, whereas in more polar solvents like DMSO and ACN, this effect is less pronounced due to better solvation of the charged species.



Figure 3.10: A. Free energy related to the ring path (inset), discretized with 100 bin to allow a comparization between solvents. **B.** Free energy related to the axis flexibility, integrating out the other variables. **C.** Free energy related to the ring conformation, integrating out the other variables.



Figure 3.11: A. Boxplot on the average number of H-bonds in each solvent between the carboxilate moiety and the macrocycle when it resides in the proximal or distal station. B. Boxplot on the average number of H-bonds in each solvent between the carboxylate moiety and the distal fumaramide when the macrocycle resides in the proximal or distal station.

3.5 Conclusions

Ratchet mechanisms are emerging as a central concept in chemistry, providing insights into molecular processes that could drive significant advancements in fields such as biotechnology, pharmaceuticals, and materials science. These systems, capable of performing work under non-equilibrium conditions, offer an alternative to traditional thermodynamic or kinetic constraints, opening the door to novel applications. However, a deeper understanding of their molecular dynamics, including the role of diffusion and reactive pathways on the free energy surface, is required to fully harness their potential. In this study, computational simulations of a [2]rotaxane system have provided valuable insights into the intrinsic energy barriers and binding affinities that define its behavior at equilibrium, without external energy inputs. This equilibrium state serves as a foundation for understanding how the system responds to the introduction of fuel and for identifying key factors that may either facilitate or hinder the ratchet's performance.

One of the key findings of this investigation is the significant role that solvent plays in determining the shuttling mechanism of the macrocycle along the axle. In DMSO and ACN, the polar nature and hydrogen bonding capabilities of the solvents act as molecular lubricants, reducing the energy barriers and promoting faster transitions between the binding stations. The higher polarity of these solvents weakens the hydrogen bonds between the macrocycle and the fumaramide groups, facilitating smoother movement along the axle. In contrast, CHCl₃, a less polar solvent, stabilizes the distal station significantly. This stabilization results in a higher free-energy barrier for the macrocycle's translocation and slows down the shuttling process. The carboxylate group, in particular, plays a crucial role in this stabilization by forming stronger hydrogen bonds with the macrocycle in the distal station, a phenomenon that is more pronounced in CHCl₃ has several important implications. First, it indicates that the macrocycle may spend more time in the distal station, which could negatively impact the efficiency of the ratchet mechanism by increasing the energy required to dislodge the macrocycle and restart the

cycle. The slower shuttling process in CHCl₃, caused by the more stabilized distal station, suggests that this solvent may not be ideal for applications where fast, efficient cycling is essential. Moreover, the kinetic asymmetry created by the stronger stabilization of the distal station could lead to a scenario where the fuel input is used less efficiently, as more energy is required to drive the macrocycle back toward the proximal station, potentially reducing the overall performance of the molecular machine.

Beyond this specific system, the results from this study offer important insights for the design of molecular machines that operate in different solvent environments. In systems like ATPase in biological contexts, the solvent environment is critical in modulating interactions between the moving components. Similar principles can be applied to abiological machines, where solvent selection can be used to finely tune the balance between hydrogen bonding and hydrophobic interactions. By optimizing the solvent environment, it may be possible to control the speed, directionality, and efficiency of molecular machines, making them more suitable for a wide range of technological applications. Future work should focus on further exploration of different solvent environments, particularly those that act as hydrogen bond donors. These solvents may provide additional stabilization to the transition states, further reducing the energy barriers and enhancing the efficiency of the shuttling mechanism. Additionally, more detailed investigations into other steps of the catalytic cycle, such as the fuel-to-waste conversion reactions, could be carried out using QM/MM calculations. This would allow for a deeper understanding of the interplay between the macrocycle's movement and the chemical reactions driving the rotaxane's function, providing a more comprehensive view of how to optimize these systems for specific applications, with potential implications across biotechnology, materials science, and nanotechnology.

Supplementary figures and tables

Table S3.1: Energetic values of minima in each solvent esspressed in kcal/mol. Nomenclature used for minima reflects the same used in the main text.

	DMSO	ACN	CHCl3
1	0.523	0.328	2.223
2	2.427	-	3.225
3	0	0	0
1 a	_	1.176	2.819
1b	_	-	2.567
1c	_	-	2.863
3 a	1.632	1.018	0.59
3 b	-	-	0.691



Figure S3.1: Structural representation of rotaxane's minima in DMSO solvent.



Figure S3.2: Structural representation of rotaxane's minima in ACN solvent.



Figure S3.3: Structural representation of rotaxane's minima in CHCl₃ solvent.
Bibliography

- Abraham, Mark James, Teemu Murtola, Roland Schulz, Szilárd Páll, Jeremy C. Smith, Berk Hess, and Erik Lindahl. "GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers." *SoftwareX* 1–2 (September 2015): 19–25. https://doi.org/10.1016/j.softx.2015.06.001.
- Alvarez-Pérez, Mónica, Stephen M. Goldup, David A. Leigh, and Alexandra M. Z. Slawin. "A Chemically-Driven Molecular Information Ratchet." *Journal of the American Chemical Society* 130, no. 6 (January 18, 2008): 1836–38. https://doi.org/10.1021/ja7102394.
- Ásgeirsson, Vilhjálmur, Benedikt Orri Birgisson, Ragnar Bjornsson, Ute Becker, Frank Neese, Christoph Riplinger, and Hannes Jónsson. "Nudged Elastic Band Method for Molecular Reactions Using Energy-Weighted Springs Combined with Eigenvector Following." *Journal of Chemical Theory and Computation* 17, no. 8 (July 18, 2021): 4929–45. https://doi.org/10.1021/acs.jctc.1c00462.
- Astumian, R. Dean. "Kinetic Asymmetry Allows Macromolecular Catalysts to Drive an Information Ratchet." *Nature Communications* 10, no. 1 (August 23, 2019). https://doi.org/10.1038/s41467-019-11402-7.
- Bayly, Christopher I., Piotr Cieplak, Wendy Cornell, and Peter A. Kollman. "A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: The RESP Model." *The Journal of Physical Chemistry* 97, no. 40 (October 1993): 10269–80. https://doi.org/10.1021/j100142a004.
- 6. Binks, Lorna, Stefan Borsley, Todd R. Gingrich, David A. Leigh, Emanuele Penocchio, and Benjamin M.W. Roberts. "The Role of Kinetic Asymmetry and

Power Strokes in an Information Ratchet." *Chem* 9, no. 10 (October 2023): 2902–17. https://doi.org/10.1016/j.chempr.2023.05.035.

- Borsley, Stefan, James M. Gallagher, David A. Leigh, and Benjamin M. W. Roberts. "Ratcheting Synthesis." *Nature Reviews Chemistry* 8, no. 1 (December 15, 2023): 8–29. https://doi.org/10.1038/s41570-023-00558-y.
- Borsley, Stefan, David A. Leigh, and Benjamin M. W. Roberts. "A Doubly Kinetically-Gated Information Ratchet Autonomously Driven by Carbodiimide Hydration." *Journal of the American Chemical Society* 143, no. 11 (March 11, 2021): 4414–20. https://doi.org/10.1021/jacs.1c01172.
- Borsley, Stefan, David A. Leigh, Benjamin M. W. Roberts, and Iñigo J. Vitorica-Yrezabal. "CCDC 2191007: Experimental Crystal Structure Determination." *Journal of the American Chemical Society* 144, no. 17241 (September 8, 2022). https://doi.org/10.5517/ccdc.csd.cc2cjxnh.
- Borsley, Stefan, David A. Leigh, Benjamin M. W. Roberts, and Iñigo J. Vitorica-Yrezabal. "Tuning the Force, Speed, and Efficiency of an Autonomous Chemically Fueled Information Ratchet." *Journal of the American Chemical Society* 144, no. 37 (September 8, 2022): 17241–48. https://doi.org/10.1021/jacs.2c07633.
- Bussi, Giovanni, Davide Donadio, and Michele Parrinello. "Canonical Sampling through Velocity Rescaling." *The Journal of Chemical Physics* 126, no. 1 (January 3, 2007). https://doi.org/10.1063/1.2408420.
- 12. Caleman, Carl, Paul J. van Maaren, Minyan Hong, Jochen S. Hub, Luciano T. Costa, and David van der Spoel. "Force Field Benchmark of Organic Liquids: Density, Enthalpy of Vaporization, Heat Capacities, Surface Tension, Isothermal Compressibility, Volumetric Expansion Coefficient, and Dielectric Constant."

Journal of Chemical Theory and Computation 8, no. 1 (December 20, 2011): 61–74. https://doi.org/10.1021/ct200731v.

- Case, David A., Hasan Metin Aktulga, Kellon Belfon, David S. Cerutti, G. Andrés Cisneros, Vinícius Wilian D. Cruzeiro, Negin Forouzesh, et al. "AmberTools." *Journal of Chemical Information and Modeling* 63, no. 20 (October 8, 2023): 6183–91. https://doi.org/10.1021/acs.jcim.3c01153.
- Darden, Tom, Darrin York, and Lee Pedersen. "Particle Mesh Ewald: An N·log(N) Method for Ewald Sums in Large Systems." *The Journal of Chemical Physics* 98, no. 12 (June 15, 1993): 10089–92. https://doi.org/10.1063/1.464397.
- Douarre, Maxime, Vicente Martí-Centelles, Cybille Rossy, Isabelle Pianet, and Nathan D. McClenaghan. "Regulation of Macrocycle Shuttling Rates in [2]Rotaxanes by Amino-Acid Speed Bumps in Organic–Aqueous Solvent Mixtures." *European Journal of Organic Chemistry* 2020, no. 36 (August 18, 2020): 5820–27. https://doi.org/10.1002/ejoc.202000997.
- Frisch, M. J. et al. Gaussian16 Revision C.01. 2016; Gaussian Inc. Wallingford CT.
- 17. Fu, Haohao, Xueguang Shao, Christophe Chipot, and Wensheng Cai. "The Lubricating Role of Water in the Shuttling of Rotaxanes." *Chemical Science* 8, no. 7 (2017): 5087–94. https://doi.org/10.1039/c7sc01593c.
- Hess, Berk, Henk Bekker, Herman J. C. Berendsen, and Johannes G. E. M. Fraaije. "LINCS: A Linear Constraint Solver for Molecular Simulations." *Journal of Computational Chemistry* 18, no. 12 (September 1997): 1463–72. https://doi.org/10.1002/(sici)1096-987x(199709)18:12<1463::aid-jcc4>3.3.co;2l.

- Jónsson, Hannes, Greg Mills, and Karsten W. Jacobsen. "Nudged Elastic Band Method for Finding Minimum Energy Paths of Transitions." In *Classical and Quantum Dynamics in Condensed Phase Simulations*. WORLD SCIENTIFIC, 1998. http://dx.doi.org/10.1142/9789812839664_0016.
- Kamlet, Mortimer J., and R. W. Taft. "The Solvatochromic Comparison Method. I. The .Beta.-Scale of Solvent Hydrogen-Bond Acceptor (HBA) Basicities." *Journal of the American Chemical Society* 98, no. 2 (January 1976): 377–83. https://doi.org/10.1021/ja00418a009.
- Kassem, Salma, Thomas van Leeuwen, Anouk S. Lubbe, Miriam R. Wilson, Ben L. Feringa, and David A. Leigh. "Artificial Molecular Motors." *Chemical Society Reviews* 46, no. 9 (2017): 2592–2621. https://doi.org/10.1039/c7cs00245a.
- 22. Kay, Euan R., David A. Leigh, and Francesco Zerbetto. "Synthetic Molecular Motors and Mechanical Machines." *Angewandte Chemie International Edition* 46, no. 1–2 (December 12, 2006): 72–191. https://doi.org/10.1002/anie.200504313.
- 23. Leanza, Luigi, Claudio Perego, Luca Pesce, Matteo Salvalaglio, Max von Delius, and Giovanni M. Pavan. "Into the Dynamics of Rotaxanes at Atomistic Resolution." American Chemical Society (ACS), March 23, 2023. http://dx.doi.org/10.26434/chemrxiv-2023-zzt1v.
- Parrinello, M., and A. Rahman. "Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method." *Journal of Applied Physics* 52, no. 12 (December 1, 1981): 7182–90. https://doi.org/10.1063/1.328693.
- "Promoting Transparency and Reproducibility in Enhanced Molecular Simulations." *Nature Methods* 16, no. 8 (July 30, 2019): 670–73. https://doi.org/10.1038/s41592-019-0506-8.

- 26. Serreli, Viviana, Chin-Fa Lee, Euan R. Kay, and David A. Leigh. "A Molecular Information Ratchet." *Nature* 445, no. 7127 (2007): 523–27. https://doi.org/10.1038/nature05452.
- Spoel, David van der, Mohammad Mehdi Ghahremanpour, and Justin A. Lemkul. "Small Molecule Thermochemistry: A Tool for Empirical Force Field Development." *The Journal of Physical Chemistry A* 122, no. 45 (October 26, 2018): 8982–88. https://doi.org/10.1021/acs.jpca.8b09867.
- Spoel, David van der, Paul J. van Maaren, and Carl Caleman. "GROMACS Molecule & Liquid Database." *Bioinformatics* 28, no. 5 (January 11, 2012): 752–53. https://doi.org/10.1093/bioinformatics/bts020.
- 29. Wang, Junmei, Romain M. Wolf, James W. Caldwell, Peter A. Kollman, and David A. Case. "Development and Testing of a General Amber Force Field." *Journal of Computational Chemistry* 25, no. 9 (April 13, 2004): 1157–74. https://doi.org/10.1002/jcc.20035.
- 30. Wu, Zhen, Shuangshuang Wang, Zilin Zhang, Yanjun Zhang, Yanzhen Yin, Haixin Shi, and Shufei Jiao. "Solvent Effects on the Motion of a Crown Ether/Amino Rotaxane." *RSC Advances* 12, no. 47 (2022): 30495–500. https://doi.org/10.1039/d2ra05453a.

Chapter 4

Exploring the Effect of Protonation Events at the S_{out} Site in the Fluc Ec2 Channel: Structural Insights from MD Simulations

4.1 Abstract

Fluc channels exhibit a remarkable selectivity for fluoride ions (F^-), a unique feature among anion channels that ensures efficient fluoride transport while excluding other more abundant anions like chloride. This property is crucial for maintaining membrane potential and cellular homeostasis in microorganisms exposed to fluoride-rich environments. Despite recent advances in elucidating the structural and functional aspects of these channels, key details of the permeation mechanism remain unresolved, particularly concerning the roles of specific residues in modulating ion coordination and channel gating.

This study focuses on *Fluc-Ec2*, a bacterial fluoride channel characterized by its dualpore architecture with two antiparallel pores. Unlike other bacterial Fluc channels, *Fluc-Ec2* contains a unique and non-conserved histidine residue, His106, which is crucial for its function. This histidine is hypothesized to coordinate the fluoride ion near a highly conserved glutamate residue (Glu86) at the S_{out} binding site, a key region at the entrance of one of the pores. The presence of this histidine sets *Fluc-Ec2* apart from other Fluc channels that lack a similar functional requirement, making it an intriguing subject for investigation. The roles of His106 and Glu86 in coordinating fluoride at S_{out} remain a central topic of scientific debate. Previous studies have proposed conflicting models regarding the necessity of their protonation states and their contribution to fluoride binding and permeation. Positioned adjacent to this site, Loop 1 seems to play a role in pore's function, as observed in other Flucs, likely by acting as a flexible gate that can modulate access to the fluoride-binding site at S_{out}. However, the exact role and dynamics of Loop 1 remain unclear.

In this study, we employed MD simulations to investigate the influence of the protonation states of Glu86 and His106 on the conformational flexibility and ion coordination at the S_{out} binding site of *Fluc-Ec2*. Six systems were modeled to capture different protonation configurations of these residues, both in the presence and absence of F⁻ at the S_{out} site. Our findings demonstrate that the protonation state of His106 is critical in modulating the

stability and dynamics of Loop 1, which, in turn, influences the channel's ability to accommodate and stabilize fluoride ions. Protonation of His106 enhanced Loop 1 flexibility and solvation, leading to greater pore openness and improved fluoride retention, especially when Glu86 was also protonated. In this state, the presence of charged species disrupted hydrophobic interactions at the pore entrance, further widening the pore and increasing solvation. In non-protonated states, or when only Glu86 was protonated, fluoride stability was reduced, highlighting the importance of electrostatic interactions for effective ion coordination.

These results highlight the complex interplay between protonation states and electrostatic effects in regulating channel behavior, adding new pieces to the mechanistic puzzle of selective fluoride transport in *Fluc-Ec2* and paves the way for future researches, with broader implications for mitigating fluoride toxicity and developing anion-selective transport systems.

4.2 Introduction

Fluoride is a ubiquitous environmental xenobiotic found at base concentrations of 20–100 mM in nature [24]. While fluoride is beneficial at low concentrations, at higher levels, it becomes toxic by inhibiting key phosphoryl-transfer enzymes involved in energy metabolism and nucleic acid synthesis [8, 9]. Many organisms have evolved defense mechanisms to mitigate fluoride toxicity, primarily through two distinct efflux systems: F^-/H^+ antiporters of the *CLC^F* family [21] and the Fluc family of fluoride ion channels [7]. These systems help reduce the intracellular accumulation of fluoride and ensure cellular survival in fluoride-rich environments. Fluoride channels (Flucs) are membrane proteins present in all classes of life except higher animals. They expel fluoride from the bacterial cytoplasm, thus protecting microorganisms from the toxic effects of this anion on a broad range of metabolic processes [12]. These small, homodimeric ion channels are remarkable proteins because of their unusual 'dual-topology' architecture and their extreme substrate selectivity [10].

Flucs consist of two homodimeric subunits (each with 120–130 residues, arranged in 4 transmembrane helices) that assemble antiparallel with respect to each other, yielding a double-barreled pair of pores related by twofold symmetry. The Fluc protein possesses two deep, aqueous vestibules with an electropositive character due to an absolutely conserved arginine sidechain and a deeply buried sodium ion at the center of the protein [11, 23]. Based on X-ray crystallographic data from *Escherichia coli* and *Bordetella pertussis* Flucs (*Fluc-Ec2* and *Fluc-Bpe*, respectively), it has been proposed that (i) the Fluc permeation pathway contains four main fluoride binding sites (S_{int}, S_{cen}, S_{ext} and S_{out}) and (ii) fluoride conduction involves a multi-ion permeation mechanism, in which fluoride binding in one site pushes its neighbor anion into the next site [10]. Recent NMR data, in combination with classical molecular dynamics simulations [26], provided further evidence for such a loosely coupled knock-on mechanism. Nonetheless, some details of the proposed fluoride conduction mechanism still need further validation, given the limitations of X-ray crystallography to distinguish between the isoelectronic fluoride and

water, as well as to assign protonation states. In particular, the S_{out} binding site contains a glutamate residue (Glu86 in *Fluc-Ec2*, see fig. 4.1) that is extremely well-conserved among fluoride channels [12].

Crystal structures (e. g. pdb id: 6BX4) have shown that Glu86 is in close contact with the putative fluoride ion occupying S_{out} (distance ~2.6 Å), which seems to be at odds with the carboxylate and the halide anion both carrying a negative charge. Hence, the protonation state of Glu86 has been investigated using both experiments and computation. On one hand, electrophysiology experiments suggested that Glu86 does not change protonation state between pH 7 and 8.7, i.e, the pH range at which the channel is conductive [10]. On the other, a recent computational study [25] predicted a pKa for Glu86 of 3.8; the subsequent classical MD simulations showed that such a negatively charged glutamate does not maintain the close contact with F⁻. Two possible explanations have been put forward to explain this apparent discrepancy. The Stockbridge lab [10] proposed that the electrostatic conflict between the negatively charged Glu86 and F⁻ could be resolved if Glu86 swings out into the solution. Such side chain rotation would then allow the fluoride ion at Sout to exit the channel, after which Glu86 could swing back inside Sout. However, such conformational change of Glu86 has not been observed in the X-ray structures of Flucs, most likely because they have been solved in the presence of monobodies plugging the permeation pores and thus restricting Glu86 flexibility. Instead, the Voth lab [25] suggested that either Glu86 or the permeating F^{-} ion could (transiently) protonate, so that they could be in close contact (i.e. hydrogen bonded) at the Sout position. However, the MD simulations of the Voth lab were performed at the force field level and thus the protonation state of glutamate and fluoride were fixed, in particular to their negatively charged state, precluding the investigation of such hydrogen bond. Moreover, this state was reached with fluoride far from Sout. During the determination of the pKa values, fluoride remained near the S_{cen} site, approximately 1 nm away from the S_{out} site. At this distance, the electrostatic and van der Waals interactions between the residues and the fluoride ion would be minimal. Therefore, it is unreasonable to exclude additional protonation events, potentially involving Glu86, with a closer fluoride. Stated that,

discerning between the two aforementioned hypotheses requires further study of the conformational preferences of Glu86, as well as the protonation state of Glu86 and F⁻. Besides Glu86, Fluc-Ec2 contains another titratable residue in the Sout site, His106 (fig. 4.1), located at ~6 Å from Glu86. Site-directed mutagenesis experiments have shown that His106 mutations, even to H-bond substitutes such as Asn and Ser, are not tolerated and abolish channel activity [5]. A previous computational study [25] showed that, when His106 is in its protonated state, it mediates ion-pairing with F⁻, thus seemingly explaining the His106 irreplaceability in Fluc-Ec2. However, His106 is suprisingly not strictly conserved across the Fluc family, e.g. being replaced by a non-titratable Ser in Fluc-Bpe. Such variability across Flucs suggests that either no titration of His106 is needed for F⁻ binding to Sout in *Fluc-Ec2* or that the binding determinants of the Sout site can differ across the Fluc family. Therefore, the protonation state of His106 also requires further investigation due to the lack of definitive experimental evidence, together with the aforementioned Glu86 and fluoride. Several important questions persist, including whether there is electrostatic coupling between Glu86 and His106 that could modulate their pKa values, potentially leading to transient configurations that alter the channel's gating state at physiological pH. Additionally, it is unclear how the proximity of fluoride may influence the acid-base properties of these residues, possibly establishing a specific equilibrium between their protonation states within the ternary system of fluoride, Glu86, and His106. Moreover, the structural dynamics of Loop 1, which is positioned adjacent to the S_{out} site, may play a significant role in channel conductivity [26].

This study is focused on addressing these complexities through MD simulations, aiming to evaluate the behavior of the channel in various configurations related to the protonation states of Glu86 and His106, both in the presence and absence of fluoride at the S_{out} binding site. By investigating these configurations, the work seeks to provide key insights into the influence of protonation states on channel behavior, potentially offering experimentally relevant parameters. The findings could help define conformational states that represent early stages of fluoride permeation or reflect the channel's resting state, thus contributing to a deeper understanding of the mechanisms driving selective fluoride transport through one of the channel's two pores.



Figure 4.1: 3D representation of the S_{out} site of Fluc-Ec2, based on the crystal structure with pdb code 6BX4. Protein residues are shown as sticks, whereas the putative fluoride anion and two water molecules are displayed as green and red spheres, respectively. Hydrogen atoms are not shown for the sake of clarity.

4.3 Materials and methods

The wild-type *Fluc-Ec2* protein (pdb id: 6BX4) [23] was selected as the starting configuration due to the presence of fluoride ions occupying both the S_{cen} and S_{out} sites in each pore. At the time of writing, no crystal structure is available with fluoride ions exclusively occupying the S_{out} site, necessitating the use of this configuration for the present study. Monobodies used for co-crystallization were removed and the putative fluoride electron densities in S_{cen} was filled with a water molecule, retaining only the ion in S_{out} (one per pore). In order to emulate the system in its resting state, even the fluoride in S_{out} was substituted with water.

The protein was embedded in a 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) lipid bilayer solvated with TIP3P water and neutralized with Cl⁻ counterions (fig. 4.2). The Cl- counterions were kept away from the centre of mass of the protein by about 4 nm (2 nm from the S_{out} site) by means of a potential wall with a force constant of 1 kcal/mol·Å², thus avoiding potential interference of the counterions on the behaviour of the channel and the fluoride-protein interaction. All titratable protein residues were modeled in the protonation state predicted by the Propka program at pH 7 [13, 18], except for residues Glu86 and His106. For these two residues, we built six independent systems using different protonation states, each with and without F⁻ in S_{out} (fig. 4.3):

1) NN(D): no extra protons, with only the nitrogen ND1 protonated on His (Glu⁻/His_{δ});

2) NN(E): no extra protons, with only the nitrogen NE2 protonated on His (Glu⁻/His_{ε});

3) *NP* : double protonated His (Glu⁻/His⁺);

4) PN(D): protonated Glu, with only the nitrogen ND1 protonated on His (Glu/His_{δ});

5) PN(E): protonated Glu, with only the nitrogen NE2 protonated on His (Glu/His_{ε});

6) *PP* : protonated Glu, double protonated His (Glu/His⁺).

In order to keep the interactions between the protein and fluoride stable and to observe the behaviour of the binding site in the presence of the ion, further simulations were conducted on systems with F^- in S_{out} , where a mild restraint (force constant: 2 kcal/mol ·Å²) was applied to the initial position of the fluoride in both pores throughout the entire

trajectory. Three simulations for each system (18 in total, 6 with fluoride, 6 with restrained fluoride and 6 with water in S_{out}), each 500 ns long (1.5 µs / system) after 10 ns equilibration, were run with *GROMACS* 2021.5 [1] using the CHARMM36 force field for protein and lipids [2, 6] and the TIP3P model for water [4]. Parameters for fluoride were taken from reference [16]. The same setup was used in an other recent work [26]. At the end of the simulations, the trajectories for each system were concatenated and the properties of the two subunits were considered cumulatively: this was done because the ability of the individual pores to work independently of each other was taken into account [5], as well as the fact that the S_{out} sites in the two pores are arranged anti-symmetrically in the channel (each faces a different side of the structures obtained were conducted with *GROMACS* 2021.5, while the application of restraints during the simulations and the remaining analyses discussed below were conducted with *PLUMED* version 2.8.1 [14].



Figure 4.2: Fluc channel, with the two antiparallel monomers in blue and red, embedded in a lipid bilayer (cyan surface) and solvated with water and chloride counterions (cyan spheres). The sodium ion (yellow sphere), located between the two subunits, stabilizes the dimer and contributes to the electropositivite character to the two permeation pores [11].

CHAPTER 4. EXPLORING THE EFFECT OF PROTONATION EVENTS AT THE S_{out} site in fluc-ec2 channel: structural insight from MD simulations



Figure 4.3: Schematic representation of model systems produced from the crystallographic structure.

4.4 Results and discussion

This section provides a detailed comparison of the various protonation states of Fluc-Ec2 under different conditions, focusing on key structural and dynamic features across simulations. Systems were analyzed both in the presence and absence of fluoride ions at the S_{out} site, highlighting critical structural fluctuations, such as pore opening, Loop 1 movements, and fluoride retention. To ensure a fair comparison, systems where fluoride binding was not stable were excluded from the analysis. Specifically, the PNs systems expelled fluoride early in both unbiased and restrained simulations, leading to their exclusion from further fluoride analyses. For the remaining systems, a restraint was necessary to retain F⁻ at Sout for at least 30% of the overall trajectory, except for the PP system, where fluoride remained stable without external bias, exhibiting similar behavior under restrained conditions. For the assessment of fluoride stability at Sout, a cutoff distance of 0.6 nm between His106 and F- was applied. Consequently, subsequent analyses included all systems in water and NN(D), NN(E), NP with restraint on fluoride, and PP unbiased in the presence of F^- . The analysis of the results begins with a thorough inspection of the Sout site in each investigated system, followed by a broader examination to identify common and contrasting aspects among the systems and ultimately discussing their implications. Detailed plot on relevant interactions and features for each system investigated can be found in the Supplementary figures and tables section of this chapter.

4.4.1 Structural features by each system

NN(E) and NN(D) in Water. Both NN(E) and NN(D) in water displayed structural behavior closely resembling the crystal structure, maintaining low RMSD values (fig. 4.2B). In both cases, Glu86 oscillated between pointing towards Sout and towards the bulk solution, without forming strong interactions with nearby residues. In NN(E), Loop 1 predominantly remained in a closed conformation, similar to the crystal structure, which resulted in low solvation of the S_{out} site. However, in NN(D), Loop 1 explored not only closed but also partially open conformations. This difference is likely due to the electrostatic influence of the negatively charged, non-protonated Glu86, which resides within less than 1 nm of the three hydrophobic residues (Leu33, Leu38, and Val103). This proximity could destabilize their hydrophobic interactions, leading to a slight opening of Loop 1. As a result, this partial opening in NN(D) allowed for more significant water ingress, enabling the complete solvation of Glu86's carboxyl moiety. Additionally, the main distinction between NN(E) and NN(D) was the interaction involving His106: in NN(D), the hydrogen on the ND1 atom of His106 interacted with the carbonyl oxygen of the Ser102 backbone, whereas in NN(E), no such interaction was observed. Despite these subtle differences, both systems generally exhibited a relatively rigid configuration, although NN(D) showed a greater tendency towards major loop flexibility.

NN(E) and NN(D) with Fluoride in S_{out}. When fluoride was restrained in the S_{out} site, both NN(E) and NN(D) systems remained structurally similar to the crystal, with some notable changes: in NN(E) system, Glu86 predominantly pointed toward the bulk, and fluoride was stabilized by His106 through an interaction with the hydrogen on NE2 atom. The fluoride ion remained hydrated, completing a coordination shell of six water molecules. Loop 1 opened slightly compared to the water-only simulation, with Leu33 losing hydrophobic contacts with Leu38 and Val103, thereby allowing greater water entry. On the other hand, in NN(D), similarly to NN(E), the ND1 atom of His106 coordinated the fluoride ion, while Glu86 oscillated between different positions. Loop 1 opened, and the fluoride ion remained solvated, contributing to an increase in water

entering the site. As in the NN(E) system, the disruption of hydrophobic contacts among Leu33, Leu38, and Val103 facilitated this opening.

NP in Water. The NP system, which had protonated His106 and deprotonated Glu86, showed significant structural divergence from the crystal structure, with an higher RMSD (fig. 4.4B). Loop 1 was notably more open, leading to increased solvation of the vestibule, including water molecules reaching as deep as Asn41. This was accompanied by the loss of hydrophobic contacts between Leu33, Leu38, and Val103, facilitating a wide pore opening. An electrostatic interaction between protonated His106 and Glu86 contributed to the movement of Glu86 away from Thr37, further diverging from the crystal configuration.

NP with Fluoride in Sout. With fluoride restrained in Sout, the NP system exhibited two distinct configurations:

- 1. Fluoride coordinated by His106 and Ser102: in this configuration, Glu86 pointed inward, interacting with Thr37, which resulted in a relatively closed Loop 1 conformation.
- Fluoride coordinated by His106 and Asn41: in this alternate configuration, Glu86 formed an electrostatic interaction with His106, leading to a much more open Loop 1 and increased pore solvation. As in the water simulation, the hydrophobic contacts between Leu33, Leu38, and Val103 were lost, allowing substantial water ingress.

PN(E) and PN(D) in Water. Both PN(E) and PN(D) systems closely resembled the crystal structure, with low RMSD values (fig. 4.4B). In both cases, Glu86 remained inside the pore, in close vicinity of Thr37, maintaining a closed Loop 1 and resulting in low solvation of the S_{out} site. His106 in PN(D) adopted a conformation where hydrogen on ND1 interacted with the backbone carbonyl of Ser102, while NE2 pointed outward, differing slightly from the crystal structure.

PP in Water. The PP system, which had both Glu86 and His106 protonated, behaved similarly to the crystal structure, with the exception of Loop 1, which alternated between closed and open states. In its closed state, Glu86 was positioned similarly to the crystal structure, and Thr37 was rotated, exposing its methyl group toward Pro34 in Loop 1.

His106 was slightly displaced, with proton on ND1 forming a hydrogen bond with the backbone of Ser102. When Loop 1 opened, hydrophobic contacts between Leu33, Leu38, and Val103 were disrupted, allowing increased water ingress, and Glu86 interacted with Thr37 while His106 moved away from Asn41, leaving space for a line of water molecules to reach Asn41.

PP with Fluoride in S_{out}. In the presence of fluoride at S_{out}, the PP system was similar to the crystal structure but showed dynamic fluctuations in Loop 1, which alternated between closed and open conformations. In the closed state, His106 remained near its crystal position, with NE2 coordinating the fluoride ion and ND1 interacting with the backbone of Ser102. Protonated Glu86 also coordinated fluoride, together with Thr37. In the open state, His106 was dislocated from the crystal position, with ND1 interacting with Ser102 and NE2 coordinating fluoride, while Glu86 coordinated the fluoride without the involvement of Thr37. The space vacated by His106 was occupied by a line of water molecules extending to Asn41, further increasing solvation.

4.4.2 Loop 1 dynamics are driven by electrostatic effects

Across all systems, the overall structural integrity of *Fluc-Ec2* remained stable, with relatively low RMSD values compared to the starting crystal structure (fig. 4.4B). However, significant conformational changes were observed in regions surrounding the S_{out} site, particularly involving Loop 1, Glu86, and His106. The protonation states of these two key residues strongly influenced the channel's behavior, affecting both pore opening and the degree of solvation (fig. 4.5). The dynamics of Loop 1 and the overall pore opening in *Fluc-Ec2* were heavily influenced by the protonation states of His106 and Glu86. Nevertheless, the presence of electrostatic charges at S_{out} , such as the negatively charged fluoride ion, also played a critical role in modulating these structural changes.

Water in S_{out}. In non-protonated His106 systems (NNs and PNs), Loop 1 predominantly remained in a closed state (fig. 4.4C), with closure occurrences of 77.70% in NN(E), 98.48% in PN(D), and 94.79% in PN(E). These closed configurations restricted the solvation of the S_{out} site and limited the accessibility of fluoride-binding residues. Notably, in the NN(D) system, Loop 1 exhibited a greater tendency to open, with a lower occurrence of closure at 35.07%. This difference suggests that the ND1 tautomer of His106 in NN(D) favors a more open Loop 1 configuration, especially when Glu86 is oriented towards S_{out}. This arrangement allows for cooperative water coordination, increasing local solvation. This observation indicates that even the tautomeric state of His106 can influence Loop 1 dynamics. Similarly, systems with fully protonated His106 (NP and PP) displayed a more open Loop 1 configuration, with occurrences of closure reduced to 7.77% in NP and 35.91% in PP. This underscores the role of His106 protonation in enhancing Loop 1 flexibility, thereby facilitating greater channel openness and solvation under water-only conditions.

With Fluoride in S_{out} . The presence of F⁻ further highlighted the importance of electrostatic interactions in regulating Loop 1 behavior. In both NN(E) and NN(D) systems with restrained F⁻, Loop 1 exhibited an increase in opening frequency (fig. 4.4C), as indicated by a significant reduction in Loop 1 closure. While Loop 1 remained

relatively closed in NN(E) with a closure occurrence of 29.60%, the NN(D) system showed a much lower closure rate of 0.32%. This suggests that the binding geometry of His106 with the ion significantly affects the movement of Loop 1 away from S_{out} . In both systems, the fluoride ion at the binding site increases solvation primarily through electrostatic effects. However, in NN(D), the binding geometry between His106 and F⁻ positions the ion in a way that replaces the interaction with the backbone of Ser102 and approaches Leu33, Leu38, and Val103, disrupting their hydrophobic interactions and promoting Loop 1 opening (fig. 4.4A). Protonated His106 systems (NP and PP) with F⁻ displayed a smaller reduction in Loop 1 closure compared to the NNs systems, indicating that the presence of the ion does not significantly alter Loop 1 flexibility or the solvation of S_{out} . For the NP system, this could be due to the fact that it is already extensively solvated in the absence of ions. However, in the PP system, this result is more notable, suggesting that the ion's charge is effectively stabilized and neutralized by the surrounding residues even when Loop 1 remains closed around S_{out} .

The data indicate that His106 and Glu86 play pivotal roles in mediating the stability and flexibility of Loop 1. Systems with charged species in S_{out} displayed the greatest Loop 1 flexibility and pore openness, allowing the channel to adopt more dynamic conformations. In these systems, the strong electrostatic effects destabilized the hydrophobic interactions between residues Leu33, Leu38, and Val103, as previously mentioned. This disruption widened the entrance to the vestibule, permitting greater water ingress and stabilizing the vestibule environment.



Figure 4.4: **A.** On the left, Fluc channel (pdb id: 6BX4). The approximative location of the S_{out} site for both subunits (A and B) is indicated by a red circle. Fluoride and Sodium ions showed as pink and yellow spheres, while Glu86 and His106 in green licorice. Loop 1 showed as blue ribbon, close to S_{out} site. On the right, loop 1 movement observed across the investigated systems, with the involved idrophobic residues highlighted in purple and loop 3 indicated by a red circle. **B.** RMSD boxplot of the investigated systems, with a water molecule or fluoride in S_{out}. **C.** Distance boxplot between COM of residues 31-34 backbone (loop 1) and COM of residues 93-95 backbone (loop 3), with a water molecule or fluoride in S_{out}.

NOTE: data with fluoride not shown for PN(D) and PN(E) systems due to lacking of data, while data from the unbiased simulations with fluoride were used for PP system. For the remaining systems, data from restrained fluoride simulations were used.



Figure 4.5: Water density, expressed in number of molecules, within the channel along two different directions. The center of the box is set to zero.

NOTE: data with fluoride not shown for PN(D) and PN(E) systems due to lacking of data, while data from the unbiased simulations with fluoride were used for PP system. For the remaining systems, data from restrained fluoride simulations were used.

4.4.3 His106 protonation enhances F⁻ retention in Sout site

The stability of the fluoride ion at the S_{out} site in *Fluc-Ec2* varied significantly across different protonation states and restraint conditions (fig. 4.6), highlighting the crucial role of electrostatic interactions in influencing F⁻ retention.

In non-protonated systems (NN(E) and NN(D)) without restraints, fluoride displayed minimal stability at S_{out} . The application of restraints dramatically increased F⁻ retention, particularly in NN(E), where the occurrence rose to 99.98%. In NN(D), although the restraint also enhanced retention, it was relatively lower, suggesting that the ND1 tautomer of His106 may present a slightly less favorable geometry for ion coordination. These results can be interpreted by noting that the opening of Loop 1 is favored by the solvation of the carboxylate group of Glu86. This indicates that coexistence between two negatively charged species (F⁻ and deprotonated Glu86) is only possible if sufficient space is created between them. Additionally, His106 remains stably coordinated with either Asn41 or Ser102, suggesting that the fluoride ion alone is not sufficient to displace His106, which is crucial for ion permeation.

In contrast, systems with a protonated His106 (NP and PP) showed significantly higher retention of F^- even without restraints. In the NP system, the occurrence of F^- at S_{out} without restraint was 26.85%, increasing to 99.77% under restrained conditions. Similarly, in the PP system, F^- retention was high at 88.03% without restraints, indicating strong stabilization at S_{out}. This highlights the combined effects of hydrogen-bond coordination and electrostatic interactions in stabilizing F^- at S_{out}, which aids in the necessary desolvation for ion permeation into the pore. This observation might seem contradictory to findings in *Fluc-Bpe*, where the E88Q mutant remains functional but less efficient in fluoride permeation. The absence of a histidine at position 108 in *Fluc-Bpe* prevents a direct analogy, but the activity of the E88Q mutant, which involves a hydrogen bond donor instead of a charged residue, may suggest that fluoride desolvation could also occur via (transient) protonation of glutamate to facilitate permeation, as proposed in

another work [25]. This aligns with previous studies that propose a role in the anion recognition for Glu residues [10].

Interestingly, in the PN systems, the lack of His106 protonation led to almost negligible retention of F^- at S_{out} , even under restraint conditions. This suggests that the presence of a charged species at S_{out} is highly destabilized by the relatively apolar environment, and hydrogen bonds with protonated Glu86 and unprotonated His106 are not sufficient for ion retention. The structures of these systems closely resemble the crystal structure, highlighting the intramolecular interactions between His106, Glu86, Thr37, and Ser102. This resemblance may indicate that this is the protonation state of the crystallized structure, induced into a closed, minimally solvated conformation by co-crystallization antibodies. This would imply that the electron density attributed to F^- in the crystal might, in fact, correspond to a water molecule.

The coordination of F^- with Asn41, deeper in the pore towards the S_{cen} site, was rare in most systems. In all NN and PN systems, both with and without restraints, the occurrence of F⁻ coordination with Asn41 was essentially zero, indicating these systems' limited ability to guide the ion inward. This lack of inward movement is likely due to insufficient electrostatic attraction or the absence of necessary structural rearrangements to facilitate fluoride transition. However, in the NP system, where His106 was fully protonated, the occurrence of F⁻ coordination with Asn41 increased to 1.20% without restraint and 11.80% with restraint. This indicates that His106 protonation creates favorable conditions for the fluoride ion to move towards the channel center. Similarly, in the PP system, while the occurrence was lower than in NP (0.04% without restraint and 1.46\% with restraint), the combined protonation of Glu86 and His106 still provided some level of attraction towards Asn41. This aligns with experimental observations in Fluc-Bpe, where dual protonation might serve as a selective filter for recognizing the correct ion. However, persistent coordination with Glu86 could slow down the ion release and permeation process. Notably, the PP system exhibited both open and closed Loop 1 states, suggesting that this transition may allow Glu86 to move away from F⁻ and eventually deprotonate, switching to an NP state.



Figure 4.6: Occurrence of fluoride bound at S_{out} site or in close proximity of Asn41, toward S_{cen} site, over the total sampling (1.5 μ s / system). Both the sets with free or restrained fluoride in S_{out} site are considered.

4.5 Conclusions

The evolutionary advantage of Fluc channels lies in their remarkable selectivity for fluoride over other anions, particularly chloride. This unique feature is attributed to sophisticated ion coordination mechanisms that allow efficient fluoride permeation while preventing the passage of other anions and cations. This selectivity preserves the membrane potential and maintains cellular homeostasis, distinguishing Fluc channels from most other anion channels, which are generally more permissive. Nevertheless, despite recent advances in our comprehension of these channels [5, 25, 26], the detailed permeation mechanism remains incompletely understood.

This in silico study underscores the significant impact of specific residues, His106 and Glu86, on Loop 1's stability and flexibility, as well as their role in securing fluoride at the Sout location. The Sout site may function not only as an exit route but also as a potential entrance, alongside the S_{int} site. This is due to the structural antiparallel arrangement of the two pores, which allows for independent ion movement through each channel [5]. The presence of charged species at the Sout site was observed to disrupt hydrophobic interactions at the pore entrance, resulting in wider pore openings and enhanced solvation. The combined effects of protonation and electrostatic interactions with fluoride ions revealed a complex interplay in regulating channel dynamics and influencing the movement of fluoride within Fluc-Ec2. Moreover, the study indicates that the low tendency of fluoride to approach Asn41 in negatively charged (NNs) or neutral (PNs) environments suggests that strong electrostatic forces and structural flexibility are essential for effective fluoride coordination and movement. Systems with a fully protonated His106 (NP and PP) created a more favorable environment for fluoride transition towards Asn41. This is likely due to the alleviation of repulsive interactions and the stabilization of fluoride movement towards the channel interior. Nevertheless, it remains crucial to recognize that a rigorous evaluation of fluoride stability at Sout requires complementary quantum mechanical modeling. Such advanced approaches could better clarify the subtle effects of Glu86 and His106, especially in contexts where polarization

and electronic effects play a more substantial role [3, 6, 25]. Incorporating these insights into future studies could refine the understanding of fluoride permeation and the interplay between protonation states and ion stability.

Additional research on Fluc channels, especially *Fluc-Ec2*, is essential to uncover the specific mechanisms involved in fluoride recognition and transport. By examining the interactions between fluoride ions and critical residues along the channel's pathway, such as the ones in S_{out}, scientists can enhance their knowledge of how these channels achieve remarkable efficiency and selectivity. These findings may have wide-ranging applications in addressing fluoride toxicity issues in environmental and industrial settings, and could potentially inspire innovative approaches in the design of anion-selective channels and transport systems.

Supplementary figures and tables



Figure S4.1: 2D density plots illustrating the relative occurrence of Glu86-His106 distance relatively to loop 1-loop 3 distance in NN(D) system: water in S_{out} (upper center), fluoride in S_{out} (left bottom) and restrained fluoride in S_{out} (right bottom) at the beginning of simulations.



Figure S4.2: 2D density plots illustrating the relative occurrence of two variables in NN(D) system: fluoride-Glu86 distance vs. fluoride-His106 distance (upper row), loop1-loop3 distance vs. fluoride-His106 distance (middle row) and Glu86-His106 distance vs. fluoride-His106 distance (bottom row). Values were computed for Fluc with fluoride (left column) or restrained fluoride (right column) in S_{out} at the beginning of simulations.



Figure S4.3: 2D density plots illustrating the relative occurrence of Glu86-His106 distance relatively to loop 1-loop 3 distance in NN(E) system: water in S_{out} (upper center), fluoride in S_{out} (left bottom) and restrained fluoride in S_{out} (right bottom) at the beginning of simulations.



Figure S4.4: 2D density plots illustrating the relative occurrence of two variables in NN(E) system: fluoride-Glu86 distance vs. fluoride-His106 distance (upper row), loop1-loop3 distance vs. fluoride-His106 distance (middle row) and Glu86-His106 distance vs. fluoride-His106 distance (bottom row). Values were computed for Fluc with fluoride (left column) or restrained fluoride (right column) in S_{out} at the beginning of simulations.



Figure S4.5: 2D density plots illustrating the relative occurrence of Glu86-His106 distance relatively to loop 1-loop 3 distance in NP system: water in Sout (upper center), fluoride in Sout (left bottom) and restrained fluoride in Sout (right bottom) at the beginning of simulations.



Figure S4.6: 2D density plots illustrating the relative occurrence of two variables in NP system: fluoride-Glu86 distance vs. fluoride-His106 distance (upper row), loop1-loop3 distance vs. fluoride-His106 distance (middle row) and Glu86-His106 distance vs. fluoride-His106 distance (bottom row). Values were computed for Fluc with fluoride (left column) or restrained fluoride (right column) in S_{out} at the beginning of simulations.



Figure S4.7: 2D density plots illustrating the relative occurrence of Glu86-His106 distance relatively to loop 1-loop 3 distance in PN(D) system: water in S_{out} (upper center), fluoride in S_{out} (left bottom) and restrained fluoride in S_{out} (right bottom) at the beginning of simulations.



Figure S4.8: 2D density plots illustrating the relative occurrence of two variables in PN(D) system: fluoride-Glu86 distance vs. fluoride-His106 distance (upper row), loop1-loop3 distance vs. fluoride-His106 distance (middle row) and Glu86-His106 distance vs. fluoride-His106 distance (bottom row). Values were computed for Fluc with fluoride (left column) or restrained fluoride (right column) in S_{out} at the beginning of simulations.
CHAPTER 4. EXPLORING THE EFFECT OF PROTONATION EVENTS AT THE S_{OUT} SITE IN FLUC-EC2 CHANNEL: STRUCTURAL INSIGHT FROM MD SIMULATIONS



Figure S4.9: 2D density plots illustrating the relative occurrence of Glu86-His106 distance relatively to loop 1-loop 3 distance in PN(E) system: water in S_{out} (upper center), fluoride in S_{out} (left bottom) and restrained fluoride in S_{out} (right bottom) at the beginning of simulations.



Figure S4.10: 2D density plots illustrating the relative occurrence of two variables in PN(E) system: fluoride-Glu86 distance vs. fluoride-His106 distance (upper row), loop1-loop3 distance vs. fluoride-His106 distance (middle row) and Glu86-His106 distance vs. fluoride-His106 distance (bottom row). Values were computed for Fluc with fluoride (left column) or restrained fluoride (right column) in S_{out} at the beginning of simulations.

CHAPTER 4. EXPLORING THE EFFECT OF PROTONATION EVENTS AT THE S_{OUT} SITE IN FLUC-EC2 CHANNEL: STRUCTURAL INSIGHT FROM MD SIMULATIONS



Figure S4.11: 2D density plots illustrating the relative occurrence of Glu86-His106 distance relatively to loop 1-loop 3 distance in PP system: water in S_{out} (upper center), fluoride in S_{out} (left bottom) and restrained fluoride in S_{out} (right bottom) at the beginning of simulations.



Figure S4.12: 2D density plots illustrating the relative occurrence of two variables in PP system: fluoride-Glu86 distance vs. fluoride-His106 distance (upper row), loop1-loop3 distance vs. fluoride-His106 distance (middle row) and Glu86-His106 distance vs. fluoride-His106 distance (bottom row). Values were computed for Fluc with fluoride (left column) or restrained fluoride (right column) in S_{out} at the beginning of simulations.

Bibliography

- Abraham, Mark James, Teemu Murtola, Roland Schulz, Szilárd Páll, Jeremy C. Smith, Berk Hess, and Erik Lindahl. "GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers." *SoftwareX* 1–2 (September 2015): 19–25. https://doi.org/10.1016/j.softx.2015.06.001.
- Huang, Jing, and Alexander D. MacKerell Jr. "CHARMM36 All-Atom Additive Protein Force Field: Validation Based on Comparison to NMR Data." *Journal of Computational Chemistry* 34, no. 25 (July 6, 2013): 2135–45. https://doi.org/10.1002/jcc.23354.
- Jing, Zhifeng, Chengwen Liu, Sara Y. Cheng, Rui Qi, Brandon D. Walker, Jean-Philip Piquemal, and Pengyu Ren. "Polarizable Force Fields for Biomolecular Simulations: Recent Advances and Applications." *Annual Review of Biophysics* 48, no. 1 (May 6, 2019): 371–94. https://doi.org/10.1146/annurev-biophys-070317-033349.
- Jorgensen, William L., Jayaraman Chandrasekhar, Jeffry D. Madura, Roger W. Impey, and Michael L. Klein. "Comparison of Simple Potential Functions for Simulating Liquid Water." *The Journal of Chemical Physics* 79, no. 2 (July 15, 1983): 926–35. https://doi.org/10.1063/1.445869.
- Last, Nicholas B, Ludmila Kolmakova-Partensky, Tania Shane, and Christopher Miller. "Mechanistic Signs of Double-Barreled Structure in a Fluoride Ion Channel." *eLife* 5 (July 23, 2016). https://doi.org/10.7554/elife.18767.
- 6. Lemkul, Justin A., Jing Huang, Benoît Roux, and Alexander D. MacKerell Jr. "An Empirical Polarizable Force Field Based on the Classical Drude Oscillator Model:

Development History and Recent Applications." *Chemical Reviews* 116, no. 9 (January 27, 2016): 4983–5013. https://doi.org/10.1021/acs.chemrev.5b00505.

- Macdonald, Christian B, and Randy B Stockbridge. "A Topologically Diverse Family of Fluoride Channels." *Current Opinion in Structural Biology* 45 (August 2017): 142–49. https://doi.org/10.1016/j.sbi.2017.04.003.
- Marquis, Robert E. "Antimicrobial Actions of Fluoride for Oral Bacteria." *Canadian Journal of Microbiology* 41, no. 11 (November 1, 1995): 955–64. https://doi.org/10.1139/m95-133.
- Marquis, Robert E., Sarah A. Clock, and Marilaine Mota-Meira. "Fluoride and Organic Weak Acids as Modulators of Microbial Physiology." *FEMS Microbiology Reviews* 26, no. 5 (January 2003): 493–510. https://doi.org/10.1111/j.1574-6976.2003.tb00627.x.
- McIlwain, Benjamin C, Roja Gundepudi, B Ben Koff, and Randy B Stockbridge.
 "The Fluoride Permeation Pathway and Anion Recognition in Fluc Family Fluoride Channels." *eLife* 10 (July 12, 2021). https://doi.org/10.7554/elife.69482.
- McIlwain, Benjamin C., Kamirah Martin, Elizabeth A. Hayter, and Randy B. Stockbridge. "An Interfacial Sodium Ion Is an Essential Structural Feature of Fluc Family Fluoride Channels." *Journal of Molecular Biology* 432, no. 4 (February 2020): 1098–1108. https://doi.org/10.1016/j.jmb.2020.01.007.
- McIlwain, Benjamin C., Michal T. Ruprecht, and Randy B. Stockbridge. "Membrane Exporters of Fluoride Ion." *Annual Review of Biochemistry* 90, no. 1 (June 20, 2021): 559–79. https://doi.org/10.1146/annurev-biochem-071520-112507.
- 13. Olsson, Mats H. M., Chresten R. Søndergaard, Michal Rostkowski, and Jan H. Jensen. "PROPKA3: Consistent Treatment of Internal and Surface Residues in

Empirical pKaPredictions." *Journal of Chemical Theory and Computation* 7, no. 2 (January 6, 2011): 525–37. https://doi.org/10.1021/ct100578z.

- 14. "Promoting Transparency and Reproducibility in Enhanced Molecular Simulations." *Nature Methods* 16, no. 8 (July 30, 2019): 670–73. https://doi.org/10.1038/s41592-019-0506-8.
- Rapp, Mikaela, Erik Granseth, Susanna Seppälä, and Gunnar von Heijne. "Identification and Evolution of Dual-Topology Membrane Proteins." *Nature Structural & amp; Molecular Biology* 13, no. 2 (January 22, 2006): 112–16. https://doi.org/10.1038/nsmb1057.
- Senn, Hans Martin, David O'Hagan, and Walter Thiel. "Insight into Enzymatic C–F Bond Formation from QM and QM/MM Calculations." *Journal of the American Chemical Society* 127, no. 39 (September 13, 2005): 13643–55. https://doi.org/10.1021/ja053875s.
- 17. Smith, Kathryn D., Patricia B. Gordon, Alberto Rivetta, Kenneth E. Allen, Tetyana Berbasova, Clifford Slayman, and Scott A. Strobel. "Yeast Fex1p Is a Constitutively Expressed Fluoride Channel with Functional Asymmetry of Its Two Homologous Domains." *Journal of Biological Chemistry* 290, no. 32 (August 2015): 19874–87. https://doi.org/10.1074/jbc.m115.651976.
- Søndergaard, Chresten R., Mats H. M. Olsson, Michał Rostkowski, and Jan H. Jensen. "Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pKa Values." *Journal of Chemical Theory and Computation* 7, no. 7 (June 9, 2011): 2284–95. https://doi.org/10.1021/ct200133y.
- Stockbridge, Randy B., Akiko Koide, Christopher Miller, and Shohei Koide.
 "Proof of Dual-Topology Architecture of Fluc F– Channels with Monobody Blockers." *Nature Communications* 5, no. 1 (October 7, 2014). https://doi.org/10.1038/ncomms6120.

- 20. Stockbridge, Randy B., Ludmila Kolmakova-Partensky, Tania Shane, Akiko Koide, Shohei Koide, Christopher Miller, and Simon Newstead. "Crystal Structures of a Double-Barrelled Fluoride Ion Channel." *Nature* 525, no. 7570 (September 2015): 548–51. https://doi.org/10.1038/nature14981.
- 21. Stockbridge, Randy B., Hyun-Ho Lim, Renee Otten, Carole Williams, Tania Shane, Zasha Weinberg, and Christopher Miller. "Fluoride Resistance and Transport by Riboswitch-Controlled CLC Antiporters." *Proceedings of the National Academy of Sciences* 109, no. 38 (September 4, 2012): 15289–94. https://doi.org/10.1073/pnas.1210896109.
- Stockbridge, Randy B, Janice L Robertson, Ludmila Kolmakova-Partensky, and Christopher Miller. "A Family of Fluoride-Specific Ion Channels with Dual-Topology Architecture." *eLife* 2 (August 27, 2013). https://doi.org/10.7554/elife.01084.
- Turman, Daniel L., and Randy B. Stockbridge. "Mechanism of Single- and Double-Sided Inhibition of Dual Topology Fluoride Channels by Synthetic Monobodies." *Journal of General Physiology* 149, no. 4 (March 3, 2017): 511– 22. https://doi.org/10.1085/jgp.201611747.
- 24. Weinstein, L. H., and A. Davison. *Fluorides in the Environment: Effects on Plants and Animals*. UK: CABI Publishing, 2003. http://dx.doi.org/10.1079/9780851996837.0000.
- Yue, Zhi, Zhi Wang, and Gregory A. Voth. "Ion Permeation, Selectivity, and Electronic Polarization in Fluoride Channels." *Biophysical Journal* 121, no. 7 (April 2022): 1336–47. https://doi.org/10.1016/j.bpj.2022.02.019.
- 26. Zhang, Jin, Dan Song, Florian Karl Schackert, Juan Li, Shengqi Xiang, Changlin Tian, Weimin Gong, Paolo Carloni, Mercedes Alfonso-Prieto, and Chaowei Shi.

"Fluoride Permeation Mechanism of the Fluc Channel in Liposomes Revealed by Solid-State NMR." *Science Advances* 9, no. 34 (August 25, 2023).

Final Remarks

This thesis employed a range of computational approaches to address key questions at the molecular level. These methods have proven to be powerful and versatile tools for describing various aspects of biologically relevant macromolecules, such as redox properties, molecular structure and dynamic behavior. As research continues to focus on molecular details within physical, chemical, and biological phenomena, computational methods increasingly stand out not only as efficient and cost-effective strategies but also as the only way to gain detailed insights at the atomic level.

The application of non-equilibrium thermodynamic integration calculations in this study highlighted the method's potential for predicting subtle thermodynamic properties, without using costly quantum mechanical frameworks. Specifically, this approach was successfully used to differentiate redox potentials between wild-type and single-point mutant flavodoxins. This capability is significant because it indicates that NE-TI could become a reliable computational tool for designing flavoproteins with tailored redox properties, a critical objective for biotechnological applications. However, the broader applicability of this technique must be evaluated in other enzyme families, particularly those that involve proton-coupled redox reactions where nearby amino acids play direct roles. Ongoing studies are expected to shed more light on the effectiveness of NE-TI calculations for such complex systems, further broadening the method's scope.

Molecular dynamics simulations of a [2]rotaxane system revealed valuable insights into how solvent properties impact the shuttling mechanism of the macrocycle along the axle. Solvent effects were found to act as "molecular lubricants", reducing energy barriers and enhancing the efficiency of the system's transitions between binding stations. These findings offer significant implications for the design of molecular machines, highlighting the importance of optimizing solvent conditions to control speed, directionality, and efficiency in such systems. Additionally, the study underscores the need for further exploration of catalytic cycles and their dependence on solvent interactions, which could lead to the development of more advanced molecular machines for applications in biotechnology, materials science, and nanotechnology.

Finally, the in silico investigation of the bacterial *Fluc-Ec2* channel focused on the role of key residues, such as His106 and Glu86, in modulating the stability and dynamics of Loop 1, as well as in stabilizing fluoride ions at the S_{out} binding site. The study revealed a complex interplay between protonation states and electrostatic interactions, which together influence channel behavior and the selective transport of fluoride ions. Systems with fully protonated His106 were shown to provide more favorable environments for fluoride retention and movement, while the presence of charged species at S_{out} disrupted hydrophobic interactions and enhanced pore solvation. These findings contribute new insights to the mechanistic puzzle of fluoride transport in Fluc channels, offering a foundation for future studies aimed at refining our understanding of ion stability and selective permeation in this family of channels.

Across all projects, molecular dynamics simulations played a crucial role in exploring structural dynamics, identifying key interactions, and uncovering mechanistic details. MD simulations have proven to be indispensable in visualizing molecular behavior over time and gaining insights that complement experimental findings. These computational approaches have the power to model large-scale systems and long-timescale events that are often challenging to capture with experimental techniques alone. By leveraging the strengths of MD simulations, this work provides new perspectives on diverse biological systems, from enzyme redox mechanisms and synthetic molecular machines to ion transport channels. Building on the findings of this thesis, several promising directions for future research emerge. In flavodoxin studies, the extension of NE-TI calculations to more complex systems could provide deeper insights into the effects of proton-coupled electron transfer. Further exploration of ratchet mechanisms should focus on fine-tuning solvent effects and catalytic cycles to achieve greater efficiency in molecular machines. Finally, additional studies on Fluc channels, particularly with the integration of QM/MM approaches, could refine the understanding of fluoride transport and reveal the roles of key residues in unprecedented detail. By advancing our understanding of redox

properties, molecular ratchets, and ion channel dynamics, this work paves the way for innovative applications in biotechnology, materials science, and beyond.

List of Publications

The thesis is based on the following publications:

- Silvestri, Giuseppe, Federica Arrigoni, Francesca Persico, Luca Bertini, Giuseppe Zampella, Luca De Gioia, and Jacopo Vertemara. "Assessing the Performance of Non-Equilibrium Thermodynamic Integration in Flavodoxin Redox Potential Estimation." *Molecules* 28, no. 16 (August 11, 2023): 6016. https://doi.org/10.3390/molecules28166016.
- Silvestri, Giuseppe, Mattia Fossati, Federica Arrigoni, Luca Bertini, Giuseppe Zampella, Luca De Gioia, and Jacopo Vertemara. "Computational Characterization of the Ring Shuttling Process in Amide-based [2]Rotaxanes: the Impact of the Solvent." In preparation.
- Silvestri, Giuseppe, Emiliano Ippoliti, Mercedes Alfonso-Prieto, and Paolo Carloni. "Exploring the Effect of Protonation Events at the Sout Site in the Fluc Ec2 Channel: Structural Insights from MD Simulations." In preparation.

Additional publications not discussed in this thesis:

Panzeri, Davide, Elisa Toini, Jacopo Vertemara, Giuseppe Silvestri, Victor Vladut Bunea, Giovanni Zecca, Werther Guidi Nissim, et al. "Small Proteins, Great Promises: Geographic Bioprospecting of Bowman–Birk Protease Inhibitors and Domestication Side-effects in African Cowpea (Vigna Unguiculata L.)." *PLANTS, PEOPLE, PLANET* 6, no. 4 (April 9, 2024): 921–34. https://doi.org/10.1002/ppp3.10507.