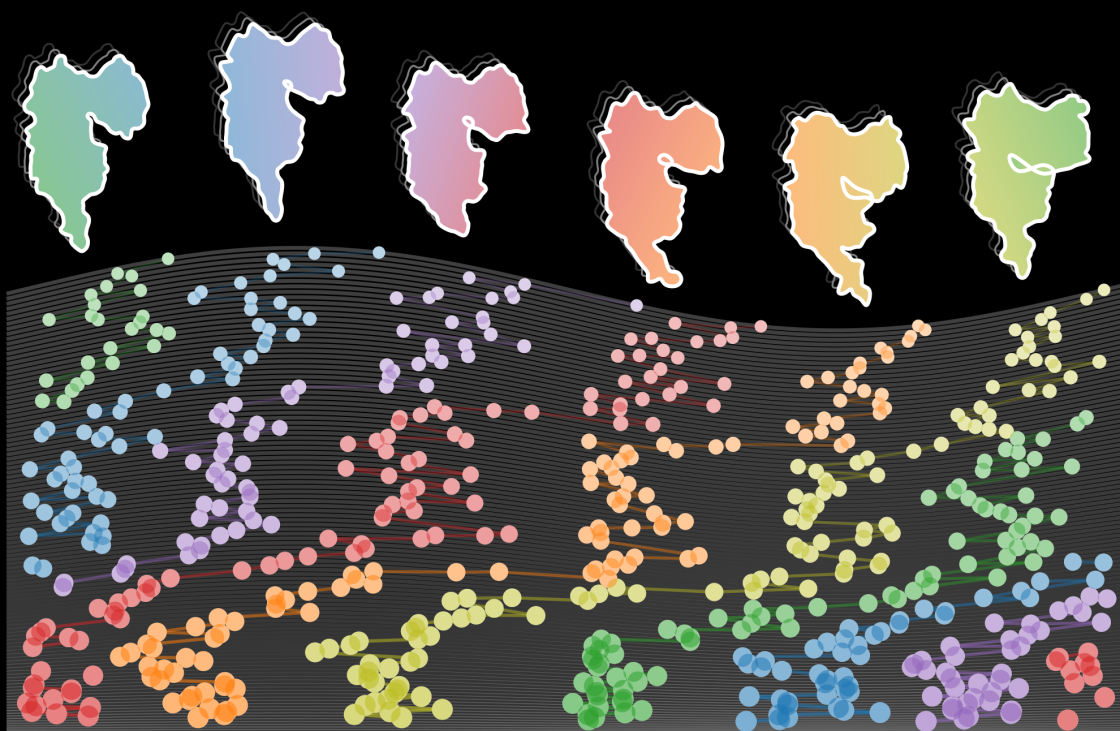


From mechanics to thermodynamics of hexameric molecular motors

Victor Hugo de Mello Pessoa



Dissertation presented to obtain the **Ph.D degree in Integrative Biology and Biomedicine**

Oeiras, November 2025

From mechanics to thermodynamics of hexameric molecular motors

Victor Hugo de Mello Pessoa

Dissertation presented to obtain the Ph.D. degree in In-
tegrative Biology and Biomedicine

Instituto de Tecnologia Química e Biológica António Xavier | Universidade Nova de
Lisboa

Research work coordinated by:



INSTITUTO
GULBENKIAN
DE CIÊNCIA



Oeiras, November 2025



Declaration

The work described in this PhD thesis received financial support from Fundação para a Ciência e a Tecnologia, through the grant UI/BD/152254/2021, awarded to Victor Hugo de Mello Pessoa.



Fundação
para a Ciência
e a Tecnologia

This work was conducted at Instituto Gulbenkian de Ciência (IGC), and Gulbenkian Institute for Molecular Medicine (GIMM), Oeiras, Portugal, under the supervision of Pablo Sartori.

All contributions have been acknowledged in the abstract of each research chapter.

Resumo

Motores moleculares são complexos proteicos que convertem energia química em movimento. Tal conversão de energia pode ser entendida como uma transferência entre processos químicos e mecânicos, mediada por mudanças de conformação altamente orquestradas na estrutura destas proteínas. Estes motores são indispensáveis para todas as formas de vida, sustentando tanto processos metabólicos quanto processos de organização espacial. Entre eles, ATPases hexaméricas formam um amplo e diverso grupo de motores moleculares que desempenham um papel central na manipulação de DNA e proteínas. Apesar de avanços importantes em bioquímica, biofísica e biologia estrutural, a relação entre mudanças de conformação e processamento de energia nestas proteínas ainda permanece parcialmente compreendida. Esta tese tem por objetivo preencher tal lacuna integrando dados de estrutura de proteínas a novas ferramentas analíticas interdisciplinares.

No Capítulo 2, começamos por conectar princípios de biofísica molecular a dados de biologia estrutural. Para alcançar este objetivo, foi desenvolvido um método para extrair informação energética diretamente da estrutura tridimensional de proteínas. Usando conceitos de elasticidade, o método *Protein Strain Analysis* foi estendido para a criação da métrica *pseudoenergia elástica*. Brevemente, esta métrica consiste em uma estimativa local da energia elástica associada à deformação entre duas conformações de uma proteína. Este método foi testado usando dados sintéticos e experimentais, estabelecendo-o como uma ferramenta geral para quantificar mecânica de proteínas.

No Capítulo 3, aplicamos esta metodologia ao ciclo conformacional de RuvB, um motor hexamérico responsável pela translocação de DNA. Ao computar a distribuição espacial da pseudoenergia elástica entre as conformações, dois padrões gerais foram identificados. Localmente, quatro regiões estruturais concentram a maior parte do conteúdo energético por subunidade ao mesmo tempo que executam interações críticas, sugerindo uma trajetória para a transferência de energia. Globalmente, o ciclo apresenta um perfil energético suave, caracterizado por uma única alta barreira associada à retomada de contato com o DNA. Estes resultados indicam que a comunicação entre subunidades é essencial para ultrapassar esta barreira, revelando o papel central da coordenação para a atividade desta proteína.

No Capítulo 4, nos baseamos nestes achados para construir um modelo cinético simplificado específico para o ciclo de RuvB. Restringindo o modelo com as pseudoenergias globais por conformação, foi estabelecida uma conexão direta entre mecânica

estrutural e cinética. As escalas de tempo entre mudanças conformacionais foram calibradas usando dados de experimentos bioquímicos e de experimentos *single-molecule*, mostrando que as energias inferidas são consistentes com uma cinética processiva. Este modelo foi aplicado posteriormente para avaliar velocidades de translocação, otimização do perfil energético e dissipação de energia neste sistema. O modelo provê uma perspectiva termodinâmica sobre como RuvB modula sua coordenação ao decorrer do ciclo conformacional.

No Capítulo 5, nós generalizamos este modelo para ATPases hexaméricas. Para isso, as conformações das subunidades foram mapeadas em variáveis de fase no formalismo de osciladores acoplados. Variando-se os parâmetros de acoplamento, foram encontrados distintos comportamentos emergentes destes motores, ou *modos de operação*, coincidindo com mecanismos de coordenação propostos previamente. Um destes modos, o regime sequencial, sempre apresenta uma maior velocidade de translocação que os demais, oferecendo, portanto, uma explicação física para a predominância deste modo entre motores hexaméricos. Ao aplicar esta metodologia para dados publicados de ClpX, um motor considerado probabilístico, foi encontrado que seu comportamento também é compatível com o modo sequencial, a depender de quais premissas bioquímicas são consideradas.

Em suma, esta tese descreve e aplica um conjunto de metodologias que relacionam conjuntos de estruturas, energia de conformações, cinética e termodinâmica de ATPases hexaméricas. Tais metodologias não só trazem novas perspectivas sobre a operação de motores específicos, como RuvB e ClpX, mas também estabelecem princípios gerais sobre coordenação e função nesta importante classe de máquinas moleculares.

Abstract

Molecular motors are protein assemblies that convert chemical energy into motion. This energy conversion can be understood as a transfer between chemical and mechanical processes, mediated by highly orchestrated conformational changes in protein structure. These motors are indispensable for all forms of life, sustaining both their metabolism and spatial organisation. Among them, hexameric ATPases form a broad and diverse group of molecular machines that play central roles in protein and DNA manipulation. Despite major advances in biochemistry, biophysics, and structural biology, the link between conformational changes and energy processing in these proteins remains poorly understood. This thesis aims to address this gap by integrating data from protein structures with novel interdisciplinary analytical tools.

In Chapter 2, we start this challenge by connecting principles of protein biophysics to structural biology datasets. To achieve this, we develop a method to extract energetic information directly from three-dimensional protein structures. Using concepts from elasticity theory, we extend the already established method of Protein Strain Analysis to devise a metric of *elastic pseudoenergy*. In a nutshell, this metric consists of a local estimate of the elastic energy associated with the deformation between two protein conformations. We probe this framework using both synthetic and experimental datasets, establishing it as a general tool to quantify protein mechanics.

In Chapter 3, we apply this approach to the conformational cycle of RuvB, a hexameric motor responsible for DNA translocation. By computing the distribution of elastic pseudoenergy across conformation and at a residue level, we identify two key patterns. Locally, four structural regions concentrate most of the energy content per subunit while mediating critical interactions, suggesting a pathway for energy transfer. Globally, the cycle forms a smooth energy landscape, characterised by a single large barrier associated with DNA reattachment. Our results indicate that inter-subunit communication is essential for overcoming this barrier, revealing the central role of coordination to protein activity.

In Chapter 4, we build on these findings to construct a coarse-grained kinetic model tailored to RuvB's conformational transitions. By constraining the model with global elastic pseudoenergy profiles, we establish a direct link between structural mechanics and kinetics. We calibrated the time scale of conformational changes using batch biochemical and single-molecule data, showing that the energetics we inferred is consistent with processive kinetics. We further applied this model to assess translocation speeds, landscape optimality, and energy dissipation in this system. This

model provides a thermodynamic perspective on how RuvB modulates its coordination throughout the conformational cycle.

In Chapter 5, we generalise this framework to hexameric ATPases. To do so, we map subunit conformations to phase variables within a coupled oscillator formalism. Strikingly, varying inter-subunit coupling parameters leads to distinct emergent behaviours of such motors, or *modes of operation*, matching previously proposed coordination mechanisms. One such mode, the sequential regime, consistently achieves the highest translocation speeds, offering a physics-based rationale for its widespread prevalence across hexameric motors. Applying this framework to published data on ClpX, long regarded as a probabilistic motor, shows that its behaviour may also be reconciled with a sequential model, depending on key biochemical assumptions.

Altogether, this thesis describes and applies a set of methodologies linking structural ensembles, conformational energetics, kinetics, and thermodynamics of hexameric ATPases. These approaches not only shed light on the operation of individual motors such as RuvB and ClpX, but also establish general principles of coordination and function within this important class of molecular machines.

Acknowledgements

In these years of PhD, I have observed changes in the world, changes in Portugal, changes in Lisbon, changes in the institute, changes in myself. All these repeated “changes”, despite being very different, are interconnected. I am convinced that science cannot be done without humanity, diversity of ideas, openness, curiosity, respect, funding, but above all, equity. Equity which is still so far to grasp, but at the same time, is always the first to be threatened. I begin this section by acknowledging this fact because, without equity in science (and beyond), we as a society and as individuals suffer several losses. I hope that starting by acknowledging this is a modest yet primordial step.

Seria impossível não agradecer uma família tão calorosa quanto a minha. Uma família que sempre me inspirou a acreditar em mim mesmo, apesar da minha relutância (e militância, às vezes). Obrigado a todos, que me apoiam incondicionalmente, ainda que por vezes pela memória, ainda que quase sempre à distância, ainda que muitas vezes em silêncio. Obrigado por todo amor que se manifesta em tão abrangentes espectros. Obrigado, mãe, por ser uma fonte inesgotável de inspiração e motivação. Obrigado aos meus avós, tão amavelmente doidos. Obrigado ao Alexandre, que é só alegria, que faz tão bem a todos sem nem perceber.

Obrigado, também em português, aos meus amigos que estão no Brasil. Vocês fazem a noção de tempo ser bem mais confusa que a do relógio. Estar em suas presenças é como se nem um dia tivesse passado. Um abraço a Anna, Larissa, Julia, Matheus, Lucas, Ricardo, Caio, Amanda, Giulia, Mateus, Cibele, Luisa, Janaína, Giovana, Davi, Carol, Guilherme e Cris. Um abraço também a todos a quem cometi injustiça em não citar!

Thanks for all the scientific (and only sometimes casual) conversations with my Living Physics buddies. Special thanks to Patroa Camila, for saving me at least twice. Many thanks as well to our extended group of Ponte Fellows. You are all nerds, and very much missed. Besides, I would like to thank a few people for whom I have great scientific and non-scientific admiration: my PhD cohort and IGC friends.

I am very grateful for the incredible past and ongoing collaborations within this period. Many thanks to Jiri Wald, Thomas Marlovits, and Wolfgang Lugmayr for the passionate discussions about an extremely complex molecular motor and for accepting the challenge of thinking about biophysics. I have also appreciated collaborating with Miguel Correa and Pedro Beltrão for their genuine scientific curiosity and flexibility. Thanks to Eric Lynch and Justin Kollman for recent yet very motivating interactions.

Thanks to my dear and hard-working friend Camila Hohenfeld. My gratitude to all the members of Maria João's group for the enriching experience. Also, thanks to Joe Howard for the sharp conversations, which helped me a lot in the beginning of my PhD. Moreover, I am hugely grateful to Ana Aranda and Jorge Carneiro, who are responsible for my presence here.

I dare to say that it won't be possible to thank Pablo enough. Thanks for the absolute trust and humanity, for the guidance, for the friendship, and for the rigour with figure aesthetics. Ah, and thank you for being an amazing scientist! I am sure that I could not be in better hands.

I would like to give a special thanks to the housekeeping genes of my (in)sanity Ravi, María, Damián, Estefanía, and Turing, who have taught me to challenge and extend the concept of home. I don't know who I would be without you all. I'm also very lucky to have met Samuel, Sonia, Romana, Felipe, Ramya, Yatharth, Camila, Slipa, and Artemis. You probably don't know how much I learn from you (which can be worrying). My Portuguese trajectory is intrinsically dependent on your support, which has changed me in many ways.

Contents

Resumo	v
Abstract	vii
Acknowledgements	ix
1 Introduction	1
1.1 An introduction to molecular motors	1
1.1.1 Biochemical cycle kinetics	3
1.1.2 Mechanochemical coupling	5
1.2 Hexameric motors	6
1.2.1 Hydrolysis mechanisms – the chemical component of the cycle	8
1.2.2 Hand-over-hand mechanism – the mechanical component of the cycle	9
1.3 Model homohexameric ATPase systems	10
1.3.1 RuvB and Holliday Junction branch migration	11
1.3.2 ClpX and protein unfolding before degradation	12
1.4 Toy models of molecular motors	13
1.5 Thesis objectives	14
Bibliography	15
2 Protein Strain Analysis and elastic pseudoenergy	23
2.1 Introduction	23
2.2 Protein strain analysis formalism	25
2.2.1 Elastic pseudoenergy	26
2.2.2 Elastic moduli of proteins	26
2.3 Elastic energy difference in ideal objects	27
2.3.1 Linear extension of a cylinder	27
2.3.2 Bulk expansion of a sphere	29
2.3.3 Shear of a cylinder	31
2.4 Energy changes of haemoglobin states using structural data	33
2.5 Discussion	36
2.6 Methods	38
2.6.1 PSA parameter choice for the analysis of haemoglobin dataset	38
Bibliography	39

3	Elastic deformations of an AAA+ motor reveal the structural basis for hand-over-hand coordination	45
3.1	Introduction	46
3.2	Results	47
3.2.1	Cycle reconstruction using minimal deformation principle . . .	47
3.2.2	Elastic pseudoenergy of the mechanochemical cycle at the sub-unit scale	49
3.2.3	Residues with high pseudoenergy variance map to four distinct functional regions	50
3.2.4	Energy is released upon transition to DNA-bound state	52
3.2.5	Inter-subunit coupling drives DNA hand-in	54
3.2.6	Nucleotide exchange is the energy source of the mechanochemical cycle	56
3.3	Discussion	57
3.3.1	New applications of elastic analysis to conformational ensembles	57
3.3.2	Biological aspects of hand-over-hand in RuvB	58
3.4	Methods	58
3.4.1	Structural data	58
3.4.2	Cycle reconstruction	59
3.4.3	PSA parameter choice for cycle reconstruction	60
3.4.4	PSA parameter choice for elastic pseudoenergy calculation . .	60
3.4.5	Computing polar interactions	61
	Bibliography	62
4	Kinetic model bridges structures and dynamics for RuvB assemblies	67
4.1	Introduction	68
4.2	A kinetic model tailored for RuvB mechanochemical cycle	69
4.3	Energy-related parameters draw conditions for processivity	72
4.4	Calibration of time scales with experimental data	73
4.5	Shape of the elastic pseudoenergy potential encodes an enhanced speed	75
4.6	Coupling profiles are consistent with structural observations	76
4.7	Trade-offs among dissipation, efficiency, and speed	77
4.8	Discussion	80
4.9	Appendix A: Definition of coupling energy and coupling force per state	82
	Bibliography	83

5	Modelling modes of operation displayed by hexameric ATPase motors	85
5.1	Introduction	86
5.2	A model for ring-shaped coupled ATPases	87
5.3	Retrieving trajectories with different modes of operation	90
5.3.1	Random initial conditions	93
5.3.2	Dependence on potential amplitude	94
5.4	Sequential mode operates faster in all conditions	95
5.5	ClpX as case study: probabilistic or sequential?	96
5.5.1	ATPase activity dataset and model assumptions	97
5.5.2	Hypothesis testing	98
5.6	Discussion	99
5.7	Appendix A: Implementation details	102
5.8	Appendix B: Equilibrium conditions for $V(\theta) = 0$ and $K > 0$	104
5.8.1	Global minima	105
5.8.2	Some local minima	106
	Bibliography	109
6	Conclusions	113
6.1	Summary	113
6.2	Conclusion	114

Introduction

Molecular motors are nanoscale protein assemblies that convert chemical energy into mechanical work. Their capacity to couple chemical and mechanical processes relies on cyclic structural deformations amid large thermal fluctuations, making their activity intrinsically stochastic. Among these systems, hexameric ATPases stand out for their broad and diverse scope, driving several processes related to protein and DNA manipulation. While cryogenic electron microscopy (cryo-EM) has brought momentum to the experimental research in the field, a general understanding of how chemical and mechanical cycles are coupled through these machines is still unresolved. Other techniques, such as single-molecule biophysics with tweezers and batch biochemical assays, provide quantitative resources to measure their activities. In contrast, theoretical studies have lagged behind in establishing novel insights into hexameric motors' properties. Here, we briefly introduce the current knowledge of their operation and highlight the existing contradictions and gaps yet to be explored.

1.1 An introduction to molecular motors

Living matter can be recognised by its incredible capacity to self-replicate, evolve, and resist perturbations at the expense of energy input. At the core of this dynamic behaviour are highly specialised molecular machines, which accelerate and control processes that would otherwise be too slow or error-prone. Like human-made machines, many of these systems are capable of converting energy from one form to another. Among them, molecular motors are protein assemblies that convert chemical energy into mechanical work.

From a cellular perspective, these motors are capable of performing a vast catalogue of functions that involve transport or molecular rearrangements. Two well-known examples that illustrate important aspects of energy conversion are ATP synthase and dynein. Strictly, the F_0F_1 ATP synthase is a notable exception to the defining attribute of a molecular motor, because it converts electric into chemical energy, where mechanical energy plays an intermediate role, see Fig. 1.1. This protein assembly can be seen as two coupled motors, F_0 and F_1 . While F_0 uses the flux of

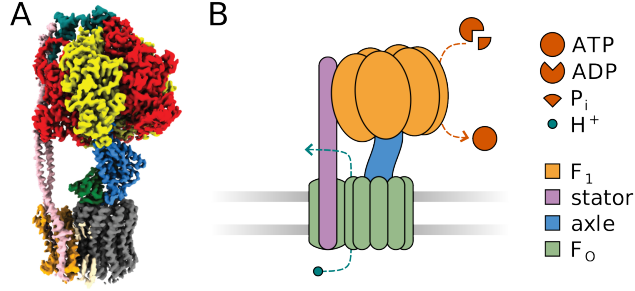


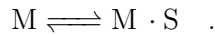
Figure 1.1: *Structure and function of F₀F₁ ATP synthase.* **A.** Cryo-EM electronic densities of *E. coli* ATP synthase. The multiple subunits that form the assembly are represented in different colors. Adapted from [3]. **B.** Simplified scheme of ATP synthase activity. The subunits were grouped into four categories: F₁ (catalytic component); F₀, (membrane-embedded proton-driver rotor); *stator*, (peripheral stalk); and *axle*, (central rotating stalk).

protons across a membrane in favour of its electrochemical gradient to generate torque – and therefore mechanical energy –, F₁ uses this torque to perform a chemical reaction, synthesising ATP from ADP and P_i [1, 2]. Due to the activity of ATP synthase, cellular ATP and ADP concentrations are kept far from thermodynamic equilibrium. This imbalance sustains a large free-energy gradient in favour of ATP hydrolysis, which drives many non-spontaneous cellular processes, including those performed by most molecular motors.

In contrast, dynein uses the chemical energy from the ATP hydrolysis cycle to generate mechanical work, similar to most motors. In its intracellular transport role, it forms dimers that move stepwise along microtubules towards its minus end [4]. To achieve the conversion of chemical energy into motion, each stepping cycle is composed of multiple enzymatic substeps



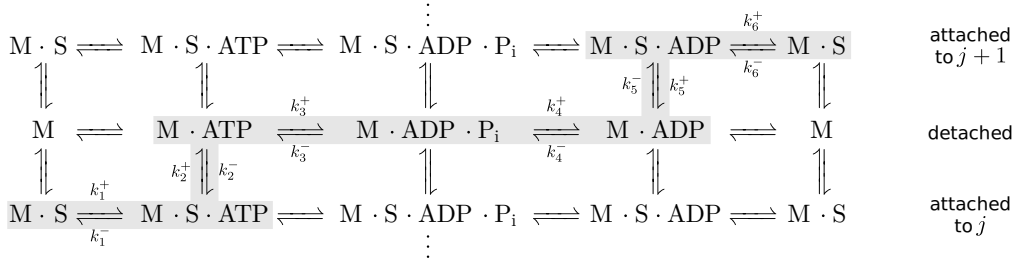
where M represents a subunit of the molecular motor. Similarly, each dynein subunit obeys an equilibrium between attached and detached states with respect to the substrate S.



Importantly, the two groups of transitions characterise the chemical and mechanical cycles, central not only to dynein but also to most molecular motors. As we mentioned earlier, the chemical cycle is driven forward because the intracellular concentrations of ATP, ADP, and P_i favour hydrolysis. The mechanical cycle, however, consists of a directional walk, thus requiring energy dissipation. Such coordination

between a spontaneous chemical reaction and a non-spontaneous mechanical process is a central aspect of molecular motors' function, often referred to as *mechanochemical coupling*. Therefore, energy must be transduced from the nucleotide-binding pocket to the substrate-interacting surface, a process that is mediated by conformational changes [5].

This coupling has been assessed experimentally for dynein using a combination of biochemical assays and structural snapshots at different states of its mechanochemical cycle [6, 7].



In dynein's kinetic diagram, transitions related to the chemical cycle are represented horizontally, where the right-most and left-most states are repeated to facilitate the visualisation. In the mechanical transitions, the directionality of the step is represented vertically, from the microtubule coordinate j to $j + 1$. Of the eight possible configurations of the system, only those shaded in gray are observed experimentally. This indicates that the rates that take dynein away from the “preferred track” are negligible compared to those in the highlighted region.

The task of understanding the system's operation becomes more challenging when accounting for the two subunits that compose the dynein dimer, because the number of possible states increases to $8^2 = 64$. For systems where the motor subunits are not independent, studying one of them in isolation does not provide a full picture of the assembly's kinetics. For hexameric motors, which are the focus of this thesis, the number of possible states is $8^6 > 2 \cdot 10^5$, making it nearly impossible to characterise substep rates for each subunit. Next, we formalise the concepts we discussed so far using the theory of biochemical kinetics.

1.1.1 Biochemical cycle kinetics

Transitions in the molecular motor's cycles are intrinsically stochastic and can be quantified by kinetic rates linking adjacent states [8]. Such kinetic rates, the k_i s in the kinetic diagram, are defined as the inverse of the mean time for the corresponding transition. Viewed from a statistical mechanics perspective, the operation of a motor

can therefore be represented as a random walk on a kinetic diagram, with the rates determining the steady-state probabilities of occupying each state.

If we consider only the shaded states, the diagram is simplified to a unidimensional cycle of six states and twelve kinetic rates. This particular cycle exhibits a tight coupling between chemical and mechanical cycles, as it comprises the hydrolysis of a single molecule of ATP and a single motor step. In a more general scenario, futile cycles – where ATP is hydrolysed without leading to a step – can also be present in a fraction of the cycle realisations. This effectively reduces the motors' efficiency. For simplicity, we will discuss the case where chemistry and mechanics are tightly coupled.

The magnitude of kinetic rates reflects characteristic time-scales of conformational changes and chemical reactions at each transition. Those rates involving binding to freely diffusing molecules are called pseudo-first-order rates, because they depend on the concentration of the ligand, such as k_1^+ , k_4^- , and k_6^- . They can be further modelled as $k_1^+ = k_1^{+*}[\text{ATP}]$, for example, where k_1^{+*} is a second-order rate. The remaining rates are first-order, as they do not depend on concentrations.

Under normal cellular conditions, dynein's cycle can be compared to a biased random walk, driven out of equilibrium by concentrations of ATP, ADP, and P_i . Therefore, the forward flux J^+ is significantly larger than the backward J^- , resulting in a net flux in the $+$ sense. If we consider the substrate S as part of the system and that $[\text{ATP}]$, $[\text{ADP}]$, and $[\text{P}_i]$ are chemostated, then the following relationship is valid

$$\frac{J^+}{J^-} = \frac{k_1^+ k_2^+ k_3^+ k_4^+ k_5^+ k_6^+}{k_1^- k_2^- k_3^- k_4^- k_5^- k_6^-} = e^{\frac{\Delta\mu}{k_B T}} \quad (1.1)$$

where $\Delta\mu$ is the free-energy released in one ATP hydrolysis cycle, k_B is the Boltzmann constant, and T the temperature. The free-energy change can be written as

$$\Delta\mu = k_B T \log \left(\frac{[\text{ATP}]K_{\text{eq}}}{[\text{ADP}][\text{P}_i]} \right) \quad (1.2)$$

where $K_{\text{eq}} = \frac{[\text{ADP}]_{\text{eq}}[\text{P}_i]_{\text{eq}}}{[\text{ATP}]_{\text{eq}}}$ is the equilibrium constant for the ATP hydrolysis reversible reaction. Because the concentrations of ATP and ADP are kept very far from equilibrium, the free-energy differences in *E. coli* are $\Delta\mu \approx 20k_B T$ [9] and in eukaryotic cells $\Delta\mu \approx 25k_B T$ [2]. Very importantly, substituting the lowest value of $\Delta\mu$ in Eq. 1.1 shows that the product of $+$ rates is $\approx 5 \cdot 10^8$ times larger than the product of $-$ rates, making the process effectively irreversible.

When a particular transition does not dissipate energy, it is possible to relate the energy of the system to the kinetic rates via the principle of detailed balance. Using

transition 2 as an example, we obtain

$$\frac{k_2^+}{k_2^-} = \exp\left(\frac{E(\text{M} \cdot \text{S} \cdot \text{ATP}) - E(\text{M} \cdot \text{ATP})}{k_{\text{B}}T}\right) \quad (1.3)$$

where E denotes the potential energy as a function of the cycle states. Similarly, this can be extended for dissipative transitions by including the energetic contribution of $\Delta\mu$ throughout the cycle. This tells us that knowing the shape of such an energy function, together with the information of where detailed balance is not satisfied, it is possible to model the energetics of biochemical cycles [10, 11, 12].

1.1.2 Mechanochemical coupling

The kinetic theory offers powerful tools to characterise and predict the behaviour of molecular motors at a coarse-grained level, yet it cannot, by itself, explain the microscopic mechanisms that give rise to such emergent kinetics. An example of that is the cycle kinetics of dynein we just described. Why do dyneins have less affinity to microtubules upon ATP binding, for instance? Is this pattern similar to all walking molecular motors? To answer these questions, it is necessary to consider the molecular details of how motors operate.

In practice, molecular motors transduce chemical energy into mechanical work through orchestrated conformational changes among their parts [5]. In dynein, for instance, binding of ADP is associated with a deformation in its structure, which can be measured by the difference in angle between the domains that bind the cargo and microtubule, see Fig. 1.2. This illustrates that changes in the chemical state of the motor correlate with changes in its mechanical state. Ultimately, the combined effect of conformational changes throughout the cycle allows for energetic mechanochemical coupling.

The example above constitutes a quantitative approach to characterise mechanochemical coupling. However, quantitative approaches are far from being the standard methodology for this purpose. Most structural biology studies employ visual-based comparisons between conformational states or relative displacements after superimposing conformations with structural alignment strategies [14]. While the second approach can provide numerical results, such as root-mean-squared displacement (RMSD), it relies on the arbitrary choice of how to align the multiple conformations being analysed [15].

Alignment-free approaches circumvent this problem, enabling the comparison of protein conformations without orientational bias. Among them, Protein Strain Analysis (PSA) is an emergent framework to directly compare pairs of conformational

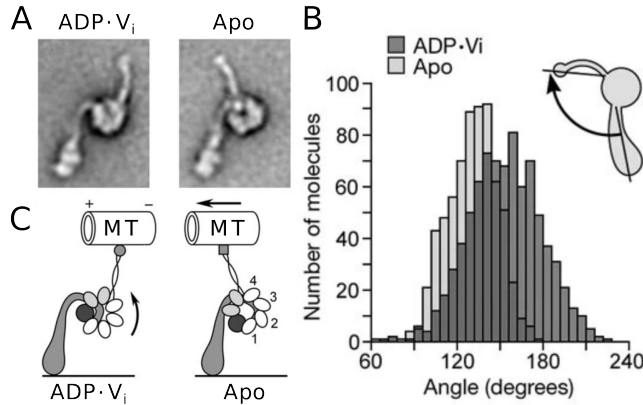


Figure 1.2: *Dynein molecules captured using negative stain in different conformations.* **A.** Average conformation of dynein of a representative class for each condition. In the first panel, dynein is incubated with ATP and a P_i analogue to mimic a long-lasting $M \cdot ADP \cdot P_i$ state. In the second, dynein is incubated in the absence of nucleotides. **B.** Distribution of angle difference between the stem (cargo-binding) and stalk (microtubule-binding). **C.** Speculative scheme depicting the conformational changes on dynein and its relative displacement to the microtubule. Adapted from [13].

states based on elasticity theory [15]. Recently, PSA was applied to a diverse group of F_1 -ATP synthase structures, showing that deformation patterns are kept even at large evolutionary scales [16]. However, despite being used to study a molecular motor, PSA has yet to be used as a tool to uncover the process of mechanochemical coupling.

In summary, this brief introduction highlights two examples of molecular motors that have been extensively studied from two well-established paradigms – kinetic theory at the coarse-grained level and structural analysis at the molecular level. Far from being antagonistic, many of the findings at the molecular scale provide insights into their thermodynamic and kinetic properties and *vice-versa*. In this thesis, we aim to achieve this link quantitatively by employing resources from both structural and biophysical paradigms. In the following, we introduce the model systems at the base of these developments.

1.2 Hexameric motors

Hexameric motors are pervasive across all domains of life, yet they do not form a monophyletic group. In other words, no single phylogenetic lineage contains only hexameric motors. Nevertheless, many are highly conserved and essential. A striking

example is the F_1 -ATP synthase compartment, whose ancestry predates the Last Universal Common Ancestor [17, 18].

Both motors introduced previously, F_1 compartment of the ATP synthase and dynein, share a close evolutionary relationship with hexameric motors. In F_1 , only half of the six subunits are catalytically active, but the complex shares close ancestry with homohexameric motors in which all subunits are active [19]. Dynein's catalytic core, in contrast, is formed by six fused subunits homologous to those of hexameric ATPases, from which only one can hydrolyse ATP. Other ring-shaped motors have evolved to operate in alternative oligomeric states, such as pentamers and heptamers [20, 21]. This remarkable diversity underscores a rich evolutionary plasticity, where hexameric motors often give rise to other assemblies that are not necessarily homohexameric.

Most known homohexameric motors belong to the vast family of P-loop NTPases, which also includes non-hexameric motors such as kinesin and myosin [22, 23]. Within this family, the ASCE proteins (Additional Strand Conserved E family) are prominent among hexameric motors, see Fig. 1.3. Here, each subunit displays only an incomplete ATP binding pocket, making either oligomerisation or motif duplication essential for activity. Examples include PilT, a main driver of bacterial pilus retraction [24], and FtsK, which translocates DNA in a step of bacterial replication [25].

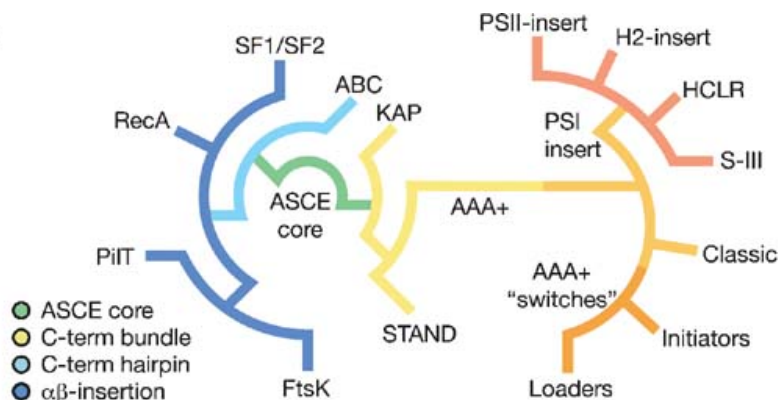


Figure 1.3: *ASCE family is rich in hexameric assemblies.* Scheme representing the phylogenetic relationships between ASCE and main subgroups, highlighting specific appendages acquired throughout the evolutionary process. Adapted from [26]

Among ASCE, the protein family of AAA+ (ATPases associated with a variety of cellular activities) has attracted substantial attention, given the rapidly growing number of solved structures. Members of this family are major translocators of protein and DNA substrates, which are typically transported through the central pore of the hexamer. The protein-translocating AAA+ are related to, among other functions, protein degradation, microtubule remodelling, and membrane fusion [27]; while

the DNA-based motors are responsible for functions such as transcription regulation, recombination, DNA repair, and are key for replication in certain viruses and transposons [28].

As with other molecular motors, hexameric ATPases operate through mechanochemical cycles. However, for homohexamers, all six subunits display the same potential to engage in both chemical and mechanical processes. This imposes consequences on their dynamics, depending on how the six subunits coordinate their conformational changes. Next, we discuss the current understanding of how these machines self-coordinate.

1.2.1 Hydrolysis mechanisms – the chemical component of the cycle

To better characterise the chemical cycle of hexameric motors, structural studies often begin by describing the distribution of bound nucleotides across subunits. To explain such distribution patterns, four types of the so-called *hydrolysis mechanisms* have been proposed (reviewed in [29]), see Fig. 1.4. In brief, they describe how temporally sustained patterns of hydrolysis and nucleotide exchange could give rise to observed nucleotide distributions.

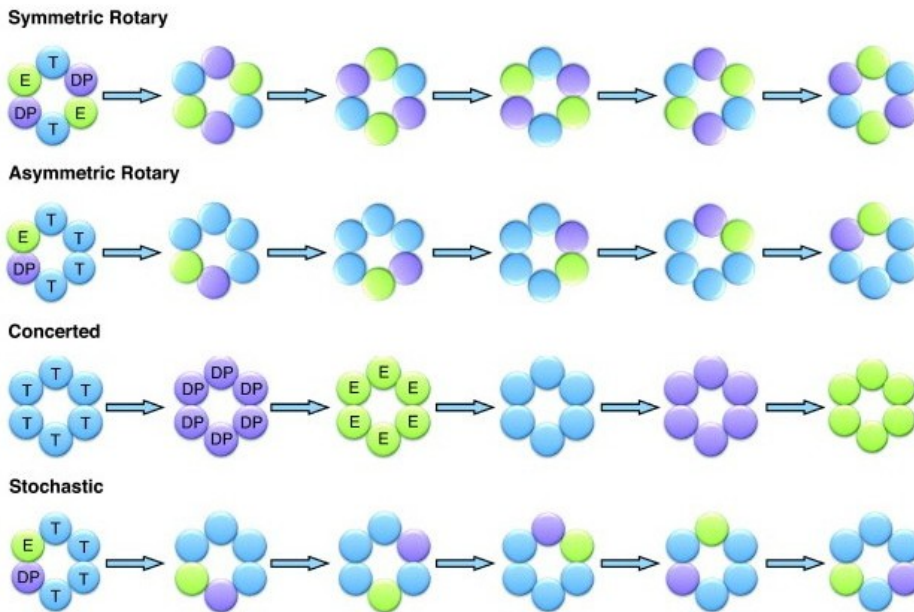


Figure 1.4: *Hydrolysis mechanisms proposed for hexameric motors.* Different models display the sequence of observed states for each subunit, colored and labelled according to the nucleotide bound: T for ATP, DP for ADP, and E for empty. Adapted from [29].

Studies using X-ray crystallography and cryo-EM on hexameric motors are typically applied to classify a given motor into one of these categories. In these studies, however, several classification inconsistencies for the same protein were observed, especially between substrate-engaged and disengaged structures [30, 31, 32, 33, 34, 35]. A likely explanation is that disengaged structures capture latent states that does not belong to the active mechanochemical cycle. Recent cryo-EM studies circumvent this issue by elucidating the structure of hexameric ATPases “in action”, while engaged on the substrate and actively translocating it. Under these circumstances, a growing body of evidence points towards a sequential hydrolysis mechanism (also known as Asymmetric Rotary in the literature, Fig. 1.4), in which ATP hydrolysis proceeds in a fixed order around the ring.

Two types of protein assemblies are clear outliers of this emerging trend: the bacterial pilus / secretion system molecular motors and the proteolysis-related translocase ClpX. In the first type, structures of molecular motors that promote the retraction and extension of bacterial pilus have been shown to adopt a two-fold symmetric structure, matching with the symmetric rotary model [36, 24, 37]. These motors also produce the largest average stall force known so far [38]. The second type consists of a single example, ClpX, that appears to operate via probabilistic hydrolysis (also known as Stochastic, Fig. 1.4), which we will discuss in a dedicated section.

1.2.2 Hand-over-hand mechanism – the mechanical component of the cycle

The term *hand-over-hand* was first coined for kinesin, which moves along microtubules by alternating between attached and detached states [39, 40], similar to dynein’s walk. In this framework, hand-over-hand was introduced as a stepping process for a two-headed motor in which both heads alternate in the frontal position (in the direction of the motion), instead of having a dominant head that always stays in front – an alternative model called the inchworm mechanism.

In a six-headed motor, the analogy is preserved, with the additional specificity that the head at the rear of the motion attaches to the front position at each step [27], see Fig. 1.5. This definition has two major implications: first, the order of attached-detached heads is compatible with a sequential mechanism, as described before for the chemical cycle; and second, the attached subunits surrounding the substrate at different positions form a spiral or staircase arrangement [41]. While the first implication requires a dynamic observation of the molecular motor at work, the second can be easily identified from snapshots, and it is widely reported [33, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53].

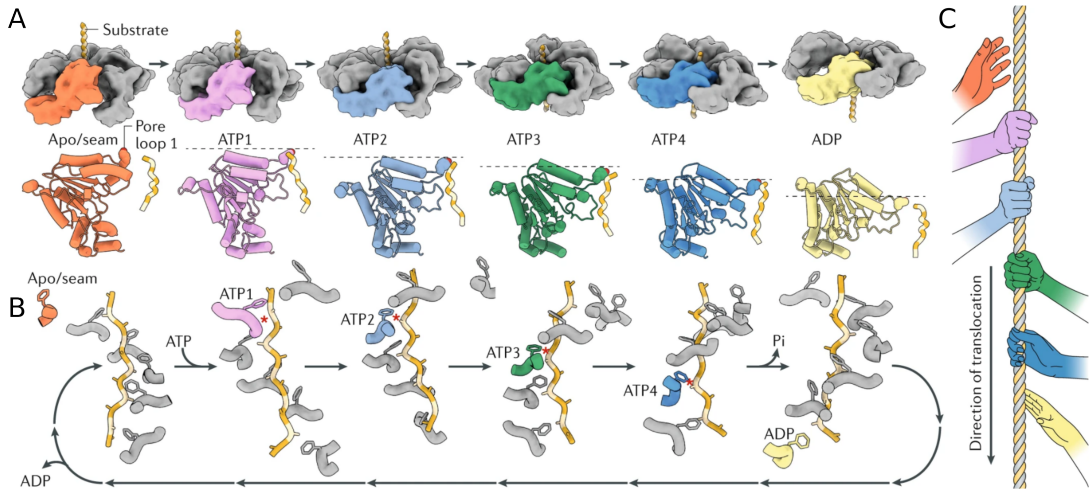


Figure 1.5: *Hand-over-hand model for substrate translocation.* **A** Structures of a generic AAA+ translocating a protein. Each colored subunit indicates a unique position in the assembly, associated to a unique conformation. In this cartoon, four subunits are bound to ATP, one is bound to ADP and one is empty. **B** The cycle depicts the structural motif interacting with the substrate at different stages of the cycle. **C**. Cartoon showing hands colored according to the previous subunits pulling on a rope. Each hand represents a subunit that stays attached to the substrate from top to bottom, until the next detaching event. Adapted from [27].

Within this framework, however, several contradictory interpretations can be found. For instance, while some references equate the sequential and hand-over-hand mechanisms [54], opposing studies claim to have identified spiral structures that do not follow a sequential hydrolysis [55]. In another study, an assembly displaying a staircase arrangement around the substrate has been claimed not to display hand-over-hand translocation because the substrate-interacting motifs differ from those typical of other AAA+ [56]. In addition, other translocation mechanisms have been proposed with essentially the same distinguishing features of hand-over-hand translocation but different terminologies [57]. Also, hand-over-hand has been omitted in favour of other translocation mechanisms [33]. In summary, these inconsistencies highlight a need for standardised definitions as well as a community-level effort to describe and interpret structural data consistently.

1.3 Model homohexameric ATPase systems

Up to this point, we have discussed molecular motors in general, with an emphasis on homohexameric ATPases. We now turn to two AAA+ motors, RuvB and ClpX,

which will serve as case studies throughout this thesis owing to their unique structural and biochemical datasets.

1.3.1 RuvB and Holliday Junction branch migration

RuvB is the active component of the RuvAB complex, which is essential for homologous recombination in bacteria. After the initiation of homologous recombination, a four-way DNA molecule is formed, a structure known as Holliday Junction. On its own, this DNA assembly is highly dynamic, displaying the capacity to shrink or grow opposing “arms” spontaneously by disrupting base-pair hydrogen bonds in the shrinking arms and forming on the growing ones. This process, however, is unbiased, as there is no net free-energy difference favouring the growth of a particular direction [58]. This scenario changes in the presence of RuvAB, which consists of an assembly formed at the center of the quadruple DNA bifurcation. While RuvA is a passive structural element, it interacts with two RuvB motors around opposing DNA arms, see Fig. 1.6. RuvB couples the free-energy of ATP hydrolysis to the unidirectional translocation of DNA, promoting the growth of DNA arms along the RuvB-RuvA-RuvB axis [59, 60].

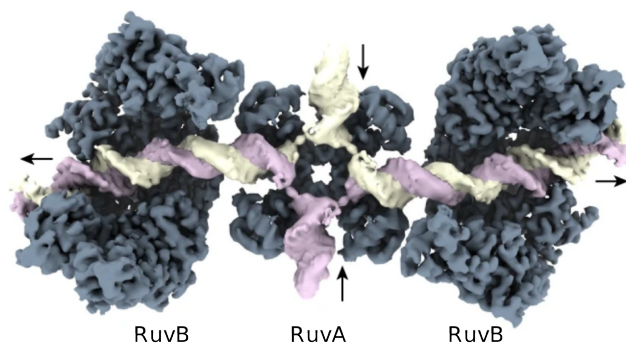


Figure 1.6: *RuvAB-Holliday Junction structure*. The figure depicts the electronic densities of two flanking RuvB motors (two subunits omitted), a central RuvA assembly (four subunits omitted) and the Holliday Junction structure. The arrows represent the direction of the DNA movement catalysed by RuvB activity. Adapted from [33].

RuvB provides a particularly suitable case study for homohexameric ATPases, as its mechanochemical cycle has been captured in a considerably large ensemble of conformational states. In a cryo-EM work by Wald and collaborators, using a slowly hydrolysing ATP analogue, multiple substeps of this cycle were resolved [33]. Complementarily, single-molecule pulling experiments have probed RuvAB activity under varying ATP concentrations and external forces [61, 62, 63]. This combination

of structural ensembles and quantitative single-molecule data makes RuvB a candidate system for integrative modelling approaches.

1.3.2 ClpX and protein unfolding before degradation

The Clp system is among the most versatile proteolytic machineries in bacteria and mitochondria. It typically consists of a tetradecameric assembly of the peptidase ClpP together with one or two AAA+ translocases that unfold and transport protein substrates into the degradation chamber. Remarkably, ClpP can form assemblies with a variety of translocases, shaping and controlling the repertoire of targeted proteins [20]. The most studied among these assemblies is ClpXP, where ClpX acts as the translocase component [64], see Fig. 1.7 for its structure.

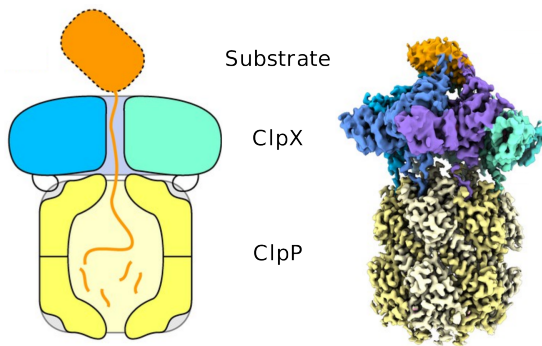


Figure 1.7: *Structure of ClpXP engaged with the substrate.* Left: scheme representing the substrate-engaged assembly. Right: composite of ClpP/ClpX/Substrate electronic densities. Adapted from [55].

Unlike other AAA+, ClpX is widely regarded as a probabilistic motor. Two central pieces of information support this view. First, a biochemistry-based assay using chimeric ClpX assemblies with fixed wild-type and ATPase-deficient subunit stoichiometries demonstrates that these assemblies retain translocation and hydrolytic capabilities [65]. This observation, together with the assumption that strictly sequential motors would stall in the presence of a knock-out mutation, suggests that only a probabilistic motor could be processive in these conditions. Second, single-molecule biophysical experiments revealed variable step sizes of substrate translocations [66]. The authors interpret such data as suggestive of near-synchronous and coordinated steps involving multiple subunits on some but not all steps.

While these studies point towards a unique mechanism among AAA+, ClpX structure adopts the canonical spiral staircase architecture, commonly found in sequential motors. In four recent cryo-EM studies, ClpXP structures bound to the translo-

cated substrate are arranged in a configuration that matches with hand-over-hand translocation [67, 55, 68, 69]. However, while Fei and colleagues propose that the data reconciles stochastic firing and strong intersubunit coordination [55], Ripstein and colleagues argue that the data, instead, supports a sequential hydrolysis model [69]. Such a contradiction between experimental approaches and their interpretation makes ClpX’s case suitable for a case study.

1.4 Toy models of molecular motors

While the study of hexameric motors has become an active field of experimental research, as discussed above, theoretical work remains significantly scarce. This is unexpected, given that well-established statistical mechanics formalisms, such as Langevin, Fokker-Planck, and Master Equation, have long been applied to molecular motors in other contexts [70]. In brief, Langevin formalism describes the dynamics of a particle through stochastic differential equations, Fokker-Planck equation models the time evolution of the probability density for ensembles of particles, and the Master Equation formalises transitions of systems in discrete states, as in the case of biochemical cycle kinetics. When combined with stochastic thermodynamics, these approaches allow for the calculation of energetic quantities such as work, efficiency, and dissipation along stochastic trajectories [71, 72].

Most theoretical studies describe motor systems with either a small numbers of active particles ($N = 1$ [73], $N = 2$ [12, 74], $N = 3$ [75, 76]) or in the limit where N is large [77, 78]. Among these, the extensive literature on modelling F_1 -ATP synthase has established it as one of the canonical model systems to study energy transduction through molecular motors [76, 79, 80]. Despite this, such studies typically focus on the coupling between the F_O rotation (or the central axle) and catalysis in F_1 , while largely overlooking the interactions among F_1 subunits themselves. Moreover, the uniqueness of ATP synthase, which hardly aligns with the formal definition of molecular motors, limits the extent to which the learned principles can be extrapolated to other hexameric ATPases. By contrast, the case of $N = 6$, where the subunits of hexameric motors are modelled directly, remains largely unexplored. We argue that this gap hinders our general understanding of the collective behaviour and emergent properties of this wide range of molecular motors.

1.5 Thesis objectives

So far, we have presented molecular motors as machines that can be seen through opposing yet interlinked perspectives: energetic-kinetic; biophysical-structural; macro-micro; mechano-chemical; theoretical-experimental. Guided by the several gaps raised in this brief review, we aspire to achieve two broader ambitions: one, to connect these contrasting perspectives, and two, to provide novel elements into the understanding of hexameric motors' dynamics. More specifically, we focus on reaching the following milestones that contribute to these major goals:

- Develop an elasticity-based method to quantify the energetics of protein deformation at the residue scale from structural data.
- Apply this method to characterise the structural details of an ensemble of RuvB conformations throughout its mechanochemical cycle.
- Elucidate the relationship between structural deformations and energy transduction from ATP hydrolysis to DNA translocation.
- Build a toy model for RuvB operation relying on the structural elements identified in the previous step and constrained by the computed energetics.
- Connect protein energetics (obtained from structural data analysis) and kinetics (from single-molecule and enzymatic assays) through the toy model.
- Generalise the toy model built for RuvB to other homohexameric ATPases.
- Use this generalised model as a framework to formalise the modes of operation across hexameric ATPases and explore its predictions.
- Test through simulations whether published enzymatic data from ClpX is consistent with particular modes of operation.

Bibliography

- [1] Y. M. Romanovsky and A. N. Tikhonov, “Molecular energy transducers of the living cell. Proton ATP synthase: a rotating molecular motor,” *Physics-Uspekhi*, vol. 53, no. 9, p. 893, 2010. 2
- [2] J. Howard, *Mechanics of Motor Proteins and the Cytoskeleton*. Sunderland, MA: Sinauer, 1 ed., 2001. 2, 4
- [3] M. Sobti, J. L. Walshe, D. Wu, R. Ishmukhametov, Y. C. Zeng, C. V. Robinson, R. M. Berry, and A. G. Stewart, “Cryo-EM structures provide insight into how E. coli F1Fo ATP synthase accommodates symmetry mismatch,” *Nature Communications*, vol. 11, no. 1, p. 2615, 2020. 2
- [4] A. P. Carter, A. G. Diamant, and L. Urnavicius, “How dynein and dynactin transport cargos: a structural perspective,” *Current Opinion in Structural Biology*, vol. 37, pp. 62–70, 2016. 2
- [5] P. M. Hoffmann, “How molecular motors extract order from chaos (a key issues review),” *Reports on Progress in Physics*, vol. 79, no. 3, p. 32601, 2016. 3, 5
- [6] A. J. Roberts, T. Kon, P. J. Knight, K. Sutoh, and S. A. Burgess, “Functions and mechanics of dynein motor proteins,” *Nature Reviews Molecular Cell Biology*, vol. 14, no. 11, pp. 713–726, 2013. 3
- [7] K. Imamula, T. Kon, R. Ohkura, and K. Sutoh, “The coordination of cyclic microtubule association/dissociation and tail swing of cytoplasmic dynein,” *Proceedings of the National Academy of Sciences*, vol. 104, no. 41, pp. 16134–16139, 2007. 3
- [8] T. L. Hill, *Free energy transduction and biochemical cycle kinetics*. Courier Corporation, 2013. 3
- [9] R. A. Alberty and R. N. Goldberg, “Standard thermodynamic formation properties for the adenosine 5'-triphosphate series,” *Biochemistry*, vol. 31, pp. 10610–10615, nov 1992. 4
- [10] H. Khataee and J. Howard, “Force Generated by Two Kinesin Motors Depends on the Load Direction and Intermolecular Coupling,” *Phys. Rev. Lett.*, vol. 122, no. 18, p. 188101, 2019. 5
- [11] H.-Y. Wang, T. Elston, A. Mogilner, and G. Oster, “Force Generation in RNA Polymerase,” *Biophysical Journal*, vol. 74, pp. 1186–1202, mar 1998. 5

- [12] C. S. Peskin and G. Oster, “Coordinated hydrolysis explains the mechanical behavior of kinesin.,” *Biophysical journal*, vol. 68, pp. 202S–210S; discussion 210S–211S, apr 1995. 5, 13
- [13] S. A. Burgess, M. L. Walker, H. Sakakibara, P. J. Knight, and K. Oiwa, “Dynein structure and power stroke,” *Nature*, vol. 421, no. 6924, pp. 715–718, 2003. 6
- [14] N. D. Thomsen, M. R. Lawson, L. B. Witkowsky, S. Qu, and J. M. Berger, “Molecular mechanisms of substrate-controlled ring dynamics and substepping in a nucleic acid-dependent hexameric motor,” *Proceedings of the National Academy of Sciences*, vol. 113, no. 48, pp. E7691–E7700, 2016. 5
- [15] M. R. Mitchell, T. Tlusty, and S. Leibler, “Strain analysis of protein structures and low dimensionality of mechanical allosteric couplings,” *Proceedings of the National Academy of Sciences*, vol. 113, no. 40, pp. E5847—E5855, 2016. 5, 6
- [16] P. Sartori and S. Leibler, “Evolutionary conservation of mechanical strain distributions in functional transitions of protein structures,” *Physical Review X*, vol. 14, no. 1, p. 11042, 2024. 6
- [17] T. A. Mahendrarajah, E. R. R. Moody, D. Schrempf, L. L. Szánthó, N. Dombrowski, A. A. Davín, D. Pisani, P. C. J. Donoghue, G. J. Szöllösi, T. A. Williams, and A. Spang, “ATP synthase evolution on a cross-braced dated tree of life,” *Nature Communications*, vol. 14, no. 1, p. 7456, 2023. 7
- [18] D. D. Leipe, Y. I. Wolf, E. V. Koonin, and L. Aravind, “Classification and evolution of P-loop GTPases and related ATPases11Edited by J. Thornton,” *Journal of Molecular Biology*, vol. 317, no. 1, pp. 41–72, 2002. 7
- [19] D. D. Majewski, L. J. Worrall, C. Hong, C. E. Atkinson, M. Vuckovic, N. Watanabe, Z. Yu, and N. C. J. Strynadka, “Cryo-EM structure of the homohexameric T3SS ATPase-central stalk complex reveals rotary ATPase-like asymmetry,” *Nature Communications*, vol. 10, no. 1, p. 626, 2019. 7
- [20] A. Illigmann, Y. Thoma, S. Pan, L. Reinhardt, and H. Brötz-Oesterhelt, “Contribution of the Clp Protease to Bacterial Survival and Mitochondrial Homeostasis,” *Microbial Physiology*, vol. 31, no. 3, pp. 260–279, 2021. 7, 12
- [21] J. R. Moffitt, Y. R. Chemla, K. Aathavan, S. Grimes, P. J. Jardine, D. L. Anderson, and C. Bustamante, “Intersubunit coordination in a homomeric ring ATPase,” *Nature*, vol. 457, no. 7228, pp. 446–450, 2009. 7

- [22] J. E. Walker, M. Saraste, M. J. Runswick, and N. J. Gay, “Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold,” *The EMBO Journal*, vol. 1, no. 8, pp. 945–951, 1982. 7
- [23] R. A. Cross, “The kinetic mechanism of kinesin,” *Trends in Biochemical Sciences*, vol. 29, pp. 301–309, jun 2004. 7
- [24] V. Solanki, S. Kapoor, and K. G. Thakur, “Structural insights into the mechanism of Type IVa pilus extension and retraction ATPase motors,” *The FEBS Journal*, vol. 285, no. 18, pp. 3402–3421, 2018. 7, 9
- [25] N. L. Jean, T. J. Rutherford, and J. Löwe, “FtsK in motion reveals its mechanism for double-stranded DNA translocation,” *Proceedings of the National Academy of Sciences*, vol. 117, no. 25, pp. 14202–14208, 2020. 7
- [26] J. P. Erzberger and J. M. Berger, “EVOLUTIONARY RELATIONSHIPS AND STRUCTURAL MECHANISMS OF AAA+ PROTEINS,” *Annual Review of Biophysics*, vol. 35, no. Volume 35, 2006, pp. 93–114, 2006. 7
- [27] C. Puchades, C. R. Sandate, and G. C. Lander, “The molecular principles governing the activity and functional diversity of AAA+ proteins,” *Nature Reviews Molecular Cell Biology*, vol. 21, no. 1, pp. 43–58, 2020. 7, 9, 10
- [28] A. Carver, B. Zhang, and X. Zhang, “Structures and mechanisms of AAA+ protein complexes in DNA processing,” *Current Opinion in Structural Biology*, vol. 92, p. 103056, 2025. 8
- [29] A. Y. Lyubimov, M. Strycharska, and J. M. Berger, “The nuts and bolts of ring-translocase structure and mechanism,” *Current Opinion in Structural Biology*, vol. 21, no. 2, pp. 240–248, 2011. 8
- [30] D. Thirumalai, C. Hyeon, P. I. Zhuravlev, and G. H. Lorimer, “Symmetry, Rigidity, and Allosteric Signaling: From Monomeric Proteins to Molecular Machines,” *Chemical Reviews*, vol. 119, pp. 6788–6821, jun 2019. 9
- [31] S. E. Glynn, A. Martin, A. R. Nager, T. A. Baker, and R. T. Sauer, “Structures of Asymmetric ClpX Hexamers Reveal Nucleotide-Dependent Motions in a AAA+ Protein-Unfolding Machine,” *Cell*, vol. 139, no. 4, pp. 744–756, 2009. 9

- [32] C. D. Putnam, S. B. Clancy, H. Tsuruta, S. Gonzalez, J. G. Wetmur, and J. A. Tainer, “Structure and mechanism of the RuvB holliday junction branch migration motor” Edited by T. Richmond,” *Journal of Molecular Biology*, vol. 311, no. 2, pp. 297–310, 2001. 9
- [33] J. Wald, D. Fahrenkamp, N. Goessweiner-Mohr, W. Lugmayr, L. Ciccarelli, O. Vesper, and T. C. Marlovits, “Mechanism of AAA+ ATPase-mediated RuvAB–Holliday junction branch migration,” *Nature*, vol. 609, no. 7927, pp. 630–639, 2022. 9, 10, 11
- [34] W. Lilyestrom, M. G. Klein, R. Zhang, A. Joachimiak, and X. S. Chen, “Crystal structure of SV40 large T-antigen bound to p53: interplay between a viral oncoprotein and a cellular tumor suppressor.,” *Genes development*, vol. 20, pp. 2373–2382, sep 2006. 9
- [35] T. Shahid, A. U. Danazumi, M. Tehseen, L. Alhudhali, A. R. Clark, C. G. Savva, S. M. Hamdan, and A. De Biasio, “Structural dynamics of DNA unwinding by a replicative helicase,” *Nature*, vol. 641, no. 8061, pp. 240–249, 2025. 9
- [36] M. McCallum, S. Tammam, A. Khan, L. L. Burrows, and P. L. Howell, “The molecular mechanism of the type IVa pilus motors,” *Nature Communications*, vol. 8, no. 1, p. 15091, 2017. 9
- [37] M. Hohl, E. J. Banks, M. P. Manley, T. B. K. Le, and H. H. Low, “Bidirectional pilus processing in the Tad pilus system motor CpaF,” *Nature Communications*, vol. 15, no. 1, p. 6635, 2024. 9
- [38] B. Maier, L. Potter, M. So, H. S. Seifert, and M. P. Sheetz, “Single pilus motor forces exceed 100 pN,” *Proceedings of the National Academy of Sciences*, vol. 99, no. 25, pp. 16012–16017, 2002. 9
- [39] J. Howard, A. J. Hudspeth, and R. D. Vale, “Movement of microtubules by single kinesin molecules.,” *Nature*, vol. 342, pp. 154–158, nov 1989. 9
- [40] A. Yildiz, M. Tomishige, R. D. Vale, and P. R. Selvin, “Kinesin Walks Hand-Over-Hand,” *Science*, vol. 303, no. 5658, pp. 676–678, 2004. 9
- [41] S. N. Gates and A. Martin, “Stairway to translocation: AAA+ motor structures reveal the mechanisms of ATP-dependent substrate translocation,” *Protein Science*, vol. 29, no. 2, pp. 407–419, 2020. 9

- [42] A. H. de la Peña, E. A. Goodall, S. N. Gates, G. C. Lander, and A. Martin, “Substrate-engaged 26 S proteasome structures reveal mechanisms for ATP-hydrolysis-driven translocation,” *Science*, vol. 362, no. 6418, p. eaav0725, 2018. 9
- [43] M. Shin, E. R. Watson, A. S. Song, J. T. Mindrebo, S. J. Novick, P. R. Griffin, R. L. Wiseman, and G. C. Lander, “Structures of the human LONP1 protease reveal regulatory steps involved in protease activation,” *Nature Communications*, vol. 12, no. 1, p. 3239, 2021. 9
- [44] E. Zehr, A. Szyk, G. Piszczek, E. Szczesna, X. Zuo, and A. Roll-Mecak, “Katanin spiral and ring structures shed light on power stroke for microtubule severing,” *Nature Structural Molecular Biology*, vol. 24, no. 9, pp. 717–725, 2017. 9
- [45] H. Han, H. L. Schubert, J. McCullough, N. Monroe, M. D. Purdy, M. Yeager, W. I. Sundquist, and C. P. Hill, “Structure of spastin bound to a glutamate-rich peptide implies a hand-over-hand mechanism of substrate translocation,” *Journal of Biological Chemistry*, vol. 295, no. 2, pp. 435–443, 2020. 9
- [46] M. Shin, C. Puchades, A. Asmita, N. Puri, E. Adjei, R. L. Wiseman, A. W. Karzai, and G. C. Lander, “Structural basis for distinct operational modes and protease activation in AAA+ protease Lon,” *Science Advances*, vol. 6, no. 21, p. eaba8404, 2020. 9
- [47] S. N. Gates, A. L. Yokom, J. Lin, M. E. Jackrel, A. N. Rizo, N. M. Kendersky, C. E. Buell, E. A. Sweeny, K. L. Mack, E. Chuang, M. P. Torrente, M. Su, J. Shorter, and D. R. Southworth, “Ratchet-like polypeptide translocation mechanism of the AAA+ disaggregase Hsp104,” *Science*, vol. 357, no. 6348, pp. 273–279, 2017. 9
- [48] C. Puchades, B. Ding, A. Song, R. L. Wiseman, G. C. Lander, and S. E. Glynn, “Unique Structural Features of the Mitochondrial AAA+ Protease AFG3L2 Reveal the Molecular Basis for Activity in Health and Disease,” *Molecular Cell*, vol. 75, no. 5, pp. 1073–1085.e6, 2019. 9
- [49] C. Puchades, A. J. Rampello, M. Shin, C. J. Giuliano, R. L. Wiseman, S. E. Glynn, and G. C. Lander, “Structure of the mitochondrial inner membrane AAA+ protease YME1 gives insight into substrate processing,” *Science*, vol. 358, no. 6363, p. eaao0464, 2017. 9

- [50] C. Deville, M. Carroni, K. B. Franke, M. Topf, B. Bukau, A. Mogk, and H. R. Saibil, “Structural pathway of regulated substrate transfer and threading through an Hsp100 disaggregase,” *Science Advances*, vol. 3, no. 8, p. e1701726, 2017. 9
- [51] K. I. White, M. Zhao, U. B. Choi, R. A. Pfuetzner, and A. T. Brunger, “Structural principles of SNARE complex recognition by the AAA+ protein NSF,” *eLife*, vol. 7, p. e38888, 2018. 9
- [52] C. Alfieri, L. Chang, and D. Barford, “Mechanism for remodelling of the cell cycle checkpoint protein MAD2 by the ATPase TRIP13,” *Nature*, vol. 559, no. 7713, pp. 274–278, 2018. 9
- [53] K. E. Lopez, A. N. Rizo, E. Tse, J. Lin, N. W. Scull, A. C. Thwin, A. L. Lucius, J. Shorter, and D. R. Southworth, “Conformational plasticity of the ClpAP AAA+ protease couples protein unfolding and proteolysis,” *Nature Structural Molecular Biology*, vol. 27, no. 5, pp. 406–416, 2020. 9
- [54] Y. A. Khan, W. K. Ian, , and A. T. Brunger, “The AAA+superfamily: a review of the structural and mechanistic principles of these molecular machines,” *Critical Reviews in Biochemistry and Molecular Biology*, vol. 57, pp. 156–187, mar 2022. 10
- [55] X. Fei, T. A. Bell, S. Jenni, B. M. Stinson, T. A. Baker, S. C. Harrison, and R. T. Sauer, “Structures of the ATP-fueled ClpXP proteolytic machine bound to protein substrate,” *eLife*, vol. 9, p. e52774, 2020. 10, 12, 13
- [56] F. Gao, F. Ye, M. Jovanovic, M. Buck, and X. Zhang, “Subunit specialization in AAA+ proteins and substrate unfolding during transcription complex remodeling,” *Proceedings of the National Academy of Sciences*, vol. 122, no. 17, p. e2425868122, 2025. 10
- [57] E. J. Enemark and L. Joshua-Tor, “Mechanism of DNA translocation in a replicative hexameric helicase,” *Nature*, vol. 442, no. 7100, pp. 270–275, 2006. 10
- [58] S. A. McKinney, A. D. J. Freeman, D. M. J. Lilley, and T. Ha, “Observing spontaneous branch migration of Holliday junctions one step at a time,” *Proceedings of the National Academy of Sciences*, vol. 102, no. 16, pp. 5715–5720, 2005. 11
- [59] H. Iwasaki, M. Takahagi, A. Nakata, and H. Shinagawa, “Escherichia coli RuvA and RuvB proteins specifically interact with Holliday junctions and promote branch migration,” *Genes development*, vol. 6, pp. 2214–2220, nov 1992. 11

- [60] S. C. West and B. Connolly, “Biological roles of the Escherichia coli RuvA, RuvB and RuvC proteins revealed,” *Molecular Microbiology*, vol. 6, no. 19, pp. 2755–2759, 1992. 11
- [61] Y.-W. Han, T. Tani, M. Hayashi, T. Hishida, H. Iwasaki, H. Shinagawa, and Y. Harada, “Direct observation of DNA rotation during branch migration of Holliday junction DNA by Escherichia coli RuvA–RuvB protein complex,” *Proceedings of the National Academy of Sciences*, vol. 103, no. 31, pp. 11544–11548, 2006. 11
- [62] A. Dawid, V. Croquette, M. Grigoriev, and F. Heslot, “Single-molecule study of RuvAB-mediated Holliday-junction migration,” *Proceedings of the National Academy of Sciences*, vol. 101, pp. 11611–11616, aug 2004. 11
- [63] R. Amit, O. Gileadi, and J. Stavans, “Direct observation of RuvAB-catalyzed branch migration of single Holliday junctions,” *Proceedings of the National Academy of Sciences*, vol. 101, pp. 11605–11610, aug 2004. 11
- [64] R. T. Sauer, X. Fei, T. A. Bell, and T. A. Baker, “Structure and function of ClpXP, a AAA+proteolytic machine powered by probabilistic ATP hydrolysis,” *Critical Reviews in Biochemistry and Molecular Biology*, vol. 57, pp. 188–204, mar 2022. 12
- [65] A. Martin, T. A. Baker, and R. T. Sauer, “Rebuilt AAA + motors reveal operating principles for ATP-fuelled machines,” *Nature*, vol. 437, no. 7062, pp. 1115–1120, 2005. 12
- [66] R. Maillard, G. Chistol, M. Sen, M. Righini, J. Tan, C. M. Kaiser, C. Hodges, A. Martin, and C. Bustamante, “ClpX(P) Generates Mechanical Force to Unfold and Translocate Its Protein Substrates,” *Cell*, vol. 145, pp. 459–469, apr 2011. 12
- [67] C. Gatsogiannis, D. Balogh, F. Merino, S. A. Sieber, and S. Raunser, “Cryo-EM structure of the ClpXP protein degradation machinery,” *Nature Structural Molecular Biology*, vol. 26, no. 10, pp. 946–954, 2019. 13
- [68] X. Fei, T. A. Bell, S. R. Barkow, T. A. Baker, and R. T. Sauer, “Structural basis of ClpXP recognition and unfolding of ssrA-tagged substrates,” *eLife*, vol. 9, p. e61496, 2020. 13

- [69] Z. A. Ripstein, S. Vahidi, W. A. Houry, J. L. Rubinstein, and L. E. Kay, “A processive rotary mechanism couples substrate unfolding and proteolysis in the ClpXP degradation machinery,” *eLife*, vol. 9, p. e52158, jan 2020. 13
- [70] D. Keller and C. Bustamante, “The Mechanochemistry of Molecular Motors,” *Biophysical Journal*, vol. 78, pp. 541–556, feb 2000. 13
- [71] T. Schmiedl and U. Seifert, “Efficiency of molecular motors at maximum power,” *Europhysics Letters*, vol. 83, p. 30005, jul 2008. 13
- [72] L. Peliti and S. Pigolotti, *Stochastic thermodynamics: an introduction*. Princeton University Press, 2021. 13
- [73] T. Elston, H. Wang, and G. Oster, “Energy transduction in ATP synthase,” *Nature*, vol. 391, no. 6666, pp. 510–513, 1998. 13
- [74] I. Derényi and T. Vicsek, “The kinesin walk: a dynamic model with elastically coupled heads,” *Proceedings of the National Academy of Sciences*, vol. 93, no. 13, pp. 6775–6779, 1996. 13
- [75] C. Doering, B. Ermentrout, and G. Oster, “Rotary DNA motors,” *Biophysical journal*, vol. 69, pp. 2256–2267, dec 1995. 13
- [76] H. Wang and G. Oster, “Energy transduction in the F1 motor of ATP synthase,” *Nature*, vol. 396, no. 6708, pp. 279–282, 1998. 13
- [77] F. Jülicher and J. Prost, “Cooperative molecular motors,” *Physical review letters*, vol. 75, no. 13, p. 2618, 1995. 13
- [78] F. Jülicher and J. Prost, “Spontaneous Oscillations of Collective Molecular Motors,” *Phys. Rev. Lett.*, vol. 78, pp. 4510–4513, jun 1997. 13
- [79] K.-i. Okazaki and G. Hummer, “Elasticity, friction, and pathway of γ -subunit rotation in FoF1-ATP synthase,” *Proceedings of the National Academy of Sciences*, vol. 112, no. 34, pp. 10720–10725, 2015. 13
- [80] S. Iranbakhsh and D. A. Sivak, “Effects of symmetry on coupled rotary molecular motors,” *arXiv preprint arXiv:2508.03869*, 2025. 13

Protein Strain Analysis and elastic pseudoenergy

Describing conformational changes is one of the most fundamental ways to understand protein function. However, most structural biology studies rely on visual inspection and alignment-dependent metrics, which overlook the energetic constraints that shape conformational transitions. In this chapter, we introduce the concept of elastic pseudoenergy as an elasticity-based metric to quantify local conformational changes. We first recapitulate the previously established formalism of Protein Strain Analysis and then build on it to propose a novel energy measure grounded in the Saint Venant-Kirchhoff model. To validate the approach, we apply it to both synthetic and experimental protein data, comparing our estimates with expected results from elasticity theory and biophysical-based methods. This framework thus provides a quantitative tool for analysing pairs of conformations applicable to a broad range of biological phenomena.

This chapter partially includes an adaptation of the publication “*Elastic analysis bridges structure and dynamics of an AAA+ molecular motor*” [1], Mello VH, Wald J, Marlovits TC, Sartori P. The chapter is further complemented with additional information not included in the manuscript. Author contributions: Conceptualisation, VHM and PS; Formal analysis, VHM; Writing – Original draft, VHM; Writing – Review and editing, VHM and PS; Visualisation, VHM; Supervision, PS. Also, I would like to thank J. Howard for illuminating discussions.

2.1 Introduction

The structure-function relationship is a central concept in life sciences in general, and more specifically at the molecular scale. Most proteins depend on changing their conformation to perform specific functions – including catalysis, transport, and regulation. In numerous cases, a protein’s activity is governed by transitions between a small number of discrete states, such as open/closed, tense/relaxed, or bent/stretched.

Crucially, these transitions are constrained by energetics. For proteins that operate at thermodynamic equilibrium, the relative abundance of these states is dictated solely by their free-energy differences at steady-state. In a remarkable experimental effort, Dashti and collaborators illustrate this principle by mapping single conformations of the ribosome at equilibrium in a low-dimensional manifold [2]. This work identified well-known biologically relevant motions in particular regions of the energy landscape, yet at very low resolution. Despite this compelling link between energetics and protein structures at a global scale, understanding the role of energetics at a local scale remains a major challenge.

From a bottom-up perspective, the free-energy of a protein conformation can be split into two: an internal energy component, or enthalpy, and an entropic component. While the entropic component can be difficult to infer from structural data, the internal energy can be modelled as a function of the protein coordinates. In molecular dynamics, this is achieved using empirical force fields that approximate internal energy as a sum of interatomic potentials [3]. However, this approach relies on the parameterisation and relative scaling of multiple interaction types, making the choice of an appropriate force field a nontrivial and system-dependent task. Alternatively, we propose to model the energetics of conformational changes as residue-scale elastic deformations. While elasticity was originally aimed to describe deformations in large objects, proteins have been shown to exhibit elastic behaviour [4, 5, 6, 7, 8, 9]. Furthermore, deformations such as bending [10, 11], stretching [12], and twisting [13] are routinely (yet informally) ascribed to protein structures.

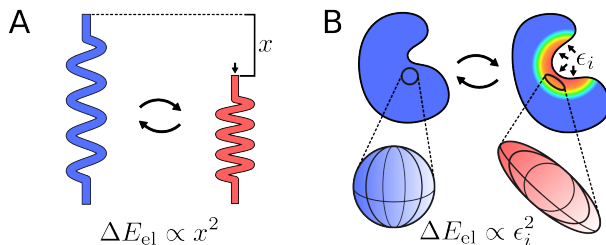


Figure 2.1: *A brief on Protein Strain Analysis.* **A.** The spring is the simplest example of an elastic system. Hooke’s law establishes that when a relaxed spring is compressed by a length x , the change in elastic energy behaves as $\Delta E_{\text{el}} \propto x^2$. **B.** For three-dimensional objects, like proteins, Hooke’s law generalises to continuum mechanics. A small region of a protein, depicted as a sphere in the reference conformation, deforms into an ellipsoid. This deformation is quantified by the local principal strains ϵ_i (with $i = 1, 2, 3$). The associated elastic energy change behaves as $\Delta E_{\text{el}} \propto \epsilon_i^2$.

Here, we provide a method to locally quantify a proxy for elastic energy from structural data, available in the publicly available python package Protein Strain

Analysis (PSA). This approach assumes proteins are homogeneous, isotropic elastic materials – a strong simplification that motivates our use of the term *pseudoenergy*. Building on previous work [14, 15], we define a measure of structural deformation at a residue-scale, the elastic pseudoenergy ΔE_{el} (Fig. 2.1A and B). Next, we test the behaviour of the elastic pseudoenergy in controlled settings, by deforming ideal objects *in silico*, and in published data for haemoglobin in two biologically meaningful conformations. We show that the computed pseudoenergies are compatible with theoretical expectations derived analytically. Moreover, we highlight that the metric we propose yields energy difference estimates comparable in magnitude to experimental free-energy differences, thus successfully bridging structural and biophysical energetics for this example.

2.2 Protein strain analysis formalism

To estimate the elastic pseudoenergy associated with protein conformational changes, we first need to introduce strain, a well-established physical concept that quantifies deformation. In order to do so, we briefly describe the mathematical framework of elasticity theory. This method is explained in detail in [15] and implemented in the PSA python package available at github.com/Sartori-Lab/PSA.

Consider the three-dimensional coordinates of a given atom i of a protein in two conformations, $\mathbf{x}_i$ and $\mathbf{X}_i$. A local deformation between these conformations is defined by the motion of the atom i relative to its surrounding atoms, and it is quantified by the deformation gradient tensor $\mathbf{F}_i$. This matrix can be estimated using the atoms in a neighbourhood of the given atom i as

$$\Delta \mathbf{x}_i^j \approx \mathbf{F}_i \Delta \mathbf{X}_i^j \quad , \quad (2.1)$$

where j are all atoms in a local neighbourhood of i , and $\Delta \mathbf{x}_i^j$ and $\Delta \mathbf{X}_i^j$ are the displacement vectors between the atoms i and j in the different conformations. The above equation is typically over-determined, and we thus estimate $\mathbf{F}_i$ minimising a quadratic error as proposed in [16], see also [15]. From the deformation gradient, we define the Lagrangian strain tensor $\mathbf{E}_i$, which captures the finite deformations around the atom i regardless of coordinate system and has the advantage of being independent of structural alignment:

$$\mathbf{E}_i = \frac{1}{2} \left(\mathbf{F}_i^T \mathbf{F}_i - \mathbf{I} \right) \quad . \quad (2.2)$$

The strain tensor can be viewed as compressions or expansions in three orthogonal

directions. This is represented schematically in Fig. 2.1B by the deformations that transform a sphere into an ellipsoid. Associated to this view we define the three principal strains $\epsilon_{i,a}$, with $a = 1, 2, 3$, the eigenvalues of the strain tensor given by $\mathbf{E}_i \mathbf{v}_{i,a} = \epsilon_{i,a} \mathbf{v}_{i,a}$, where the eigenvectors $\mathbf{v}_{i,a}$ are the directions of each principal strain. The principal strains are ordered as $\epsilon_{i,1} < \epsilon_{i,2} < \epsilon_{i,3}$.

2.2.1 Elastic pseudoenergy

In order to discuss protein energetics, we use a hyperelastic model for the energy density, $\Psi(\mathbf{E})$ [17]. Specifically, we use a Saint Venant-Kirchhoff model for which Ψ is given by:

$$\Psi(\mathbf{E}_i) = \frac{\lambda}{2} \text{tr}(\mathbf{E}_i)^2 + \mu \text{tr}(\mathbf{E}_i^2) \quad , \quad (2.3)$$

where λ and μ are the first and second Lamé constants, respectively. The advantage of using this form of Ψ is that λ and μ relate to the Young modulus, which has been measured experimentally and estimated computationally for proteins [18, 4, 19, 20, 21, 22]. This is not the case for most hyperelastic models. Also, this form of Ψ is a strictly non-negative function, unlike other simple elastic models that might yield negative energies for compressed regions.

Because energy is an extensive quantity, it is defined as the integral of the energy density over the protein's volume. We compute the elastic pseudoenergy ΔE_{el} contained in the deformations of the neighbourhood of an atom as

$$\begin{aligned} \Delta E_{\text{el},i} &= \Psi(\mathbf{E}_i) v_i \phi \\ \phi &= \frac{V}{\sum_{j=1}^N v_j} \quad , \end{aligned} \quad (2.4)$$

where v_i represents the van der Waals volume of the atom i , and ϕ is a scaling factor that depends on the van der Waals volume of the whole protein V and the sum of volumes of the N atoms considered in the analysis. The scaling factor ϕ is especially useful for adjusting the elastic pseudoenergy definition to different choices of atoms (only C_α , main chain, all atoms, etc).

2.2.2 Elastic moduli of proteins

Different elastic moduli are related to various forms of deformation exerted on materials, such as shear (shear modulus), compression (bulk modulus), or one-directional extensions (Young's modulus). Most measurements of the Young modulus of proteins were obtained for polymers (summarised in [20, 4]), which might not reflect the

elastic properties of single molecules due to relative displacement between protomers. These estimates range several orders of magnitude, spanning 1MPa to 10GPa. We then decided to base our stiffness constant on studies where the Young modulus was inferred using atomic force microscopy on single molecules of globular proteins. The results range from 2.5-9MPa in [23], 4-50MPa [21], 2-75MPa in [19], and 500MPa in [18]. Using a different nanorheology approach, Wang and Zocchi estimate Young's modulus of a globular protein to be around 11MPa [7].

The elastic pseudoenergy density in the Saint Venant-Kirchhoff model depends linearly on proteins' stiffness in parameters λ and μ , which are themselves functions of the Young modulus and Poisson's ratio. Because the estimates of protein stiffness vary by orders of magnitude, our estimations of elastic pseudoenergy also comprise a large uncertainty in its amplitude. The dependency of λ and μ on Poisson's ratio is non-linear and mostly flat for values close to $\frac{1}{3}$, typically assumed for proteins [18]. For our analysis, we considered $Y = 100\text{MPa}$ and $\nu = \frac{1}{3}$, which led respectively to $\lambda \approx 73\text{MPa}$ and $\mu \approx 38\text{MPa}$.

2.3 Elastic energy difference in ideal objects

2.3.1 Linear extension of a cylinder

We begin by estimating the elastic pseudoenergy difference ΔE_{el} between ideal objects, for which analytical expressions can be obtained due to their simple geometry. Our first example is the linear extension of a homogeneous elastic material. In elasticity theory, the regime in which stress and strain are linearly related is characterised by the Young's modulus. Here, we denote it as Y not to be confused with the Lagrange strain tensor $\mathbf{E}$. In this linear regime, the following approximation holds:

$$Y = \frac{\sigma_{yy}}{\varepsilon_{yy}} = \frac{F}{A} \frac{L_0}{L - L_0} \quad , \quad (2.5)$$

where σ_{yy} is the stress (also represented as force F orthogonal to the surface of area A) and ε_{yy} is the linear extensive strain, defined in terms of the length of the body in the relaxed L_0 and deformed L states (see Fig. 2.2A for a scheme). Here we arbitrarily assign the deformations along the y -axis, without any loss of generality. The potential elastic energy associated with the force F is then defined as

$$\Delta E_{\text{el}} = \frac{Y A L_0 \varepsilon_{yy}^2}{2} = \frac{Y \pi R^2 (L - L_0)^2}{2 L_0} \quad . \quad (2.6)$$

In our modelling, however, we assume that the strain depends non-linearly on

the relative displacement (recall Eq. 2.2), as it is habitually considered for finite deformations [17]. To correctly quantify the elastic energy using the Lagrangian strain tensor for a linear extension $\mathbf{E}_{\text{ext}}$, we consider its components for a simple axial extension. Thus, the strain tensor can be written as

$$\mathbf{E}_{\text{ext}} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & E_{yy} & 0 \\ 0 & 0 & 0 \end{bmatrix} ,$$

where (using Eq. 2.2)

$$E_{yy} = \frac{\left(\frac{L}{L_0}\right)^2 - 1}{2} = \frac{L^2 - L_0^2}{2L_0^2} .$$

Substituting this into Eqs. 2.3 and 2.4, we have

$$\begin{aligned} \Psi(\mathbf{E}_{\text{ext}}) &= \frac{\lambda}{2} E_{yy}^2 + \mu E_{yy}^2 = \left(\frac{\lambda}{2} + \mu\right) E_{yy}^2 \\ \Delta E_{\text{el}} &= \Psi(\mathbf{E}_{\text{ext}}) V_0 \\ &= \left(\frac{\lambda}{2} + \mu\right) \frac{(L^2 - L_0^2)^2}{4L_0^4} \pi R^2 L_0 \\ &= \left(\frac{\lambda}{2} + \mu\right) \frac{L^4 - 2L^2 L_0^2 + L_0^4}{4L_0^3} \pi R^2 . \end{aligned}$$

We use this result as a ground truth for assessing the elastic energy computed from PSA. To do so, we use simulated coordinates of a cylinder subjected to linear extensions using a Python script, as shown in Fig. 2.2B. In summary, we generated a set of three-dimensional coordinates forming concentric circular disks at various radii. We then stacked multiple disks along the y -axis, resulting in a cylinder of dimensions $L_0 = 180\text{\AA}$ and $R = 20\text{\AA}$. The resulting coordinates were subjected to a series of compressions and expansions by scaling the y coordinates by factors ranging from 0.8 to 1.2. We then used the PSA library to calculate the elastic pseudoenergy of each coordinate, and computed the total ΔE_{el} by summing the contributions from all points. While the number of points does not affect the results, due to the proportional decrease of each of their volumes, their density must be sufficient to have enough neighbours within the local radius when computing the deformation gradient. Therefore, care was taken to ensure that multiple elements from the same and adjacent disks were inside the local neighbourhood, set to a sphere of radius $r = 9\text{\AA}$.

By analysing Fig. 2.2C, it becomes clear that the elastic energies calculated with PSA accurately recover the analytical results. Another important observation is that

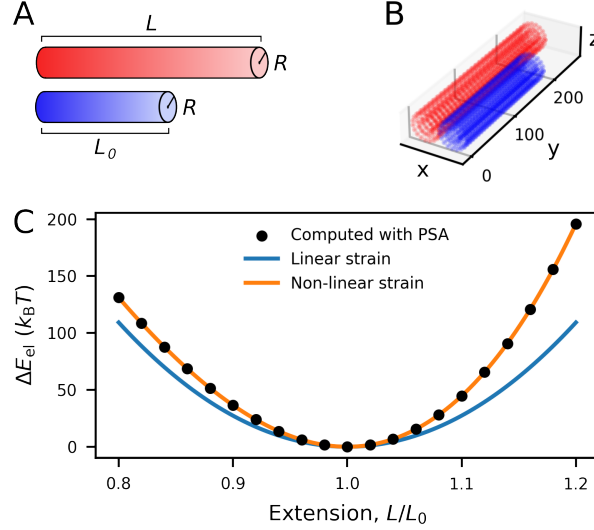


Figure 2.2: *Potential energy changes upon linear extension/contraction of a cylinder.* **A.** Scheme representing the geometry of a cylinder in two states: relaxed (blue) and extended (red). All axes are scaled in Å units. **B.** Example of simulated relaxed and extended states. **C.** Elastic energy at different extensions obtained using PSA is compared to theoretical results using linear and non-linear strain definitions.

the results for the linear regime are only similar to the non-linear at very low stretching, as the divergence between the curves grows fast away from contractions and even faster for extensions. In the next subsections, we apply analogous strategies to two other types of deformations.

2.3.2 Bulk expansion of a sphere

We next investigate the elastic energy that arises from bulk compression (or expansion) of a material. This type of deformation relates to a material property quantified by the bulk modulus K , formally defined as proportional to the derivative of pressure with respect to volume. However, in analogy to the previous example, we can define a discrete version of the bulk modulus in terms of a stress-strain relationship, yielding

$$K = -\frac{\sigma_v}{\varepsilon_v} = -P \frac{V_0}{V - V_0} \quad , \quad (2.7)$$

where σ_v and ε_v are the volumetric strain and stress, respectively, P is the pressure exerted on the material, and V_0 and V represent initial and final volumes, see Fig. 2.3A for schematics. Here, the minus sign is present to satisfy the convention of $K > 0$, because $V > V_0$ and $P < 0$ under expansions. From this relationship, we derive the

associated elastic energy under linear strain

$$\Delta E_{\text{el}} = \frac{KV_0\varepsilon_v^2}{2} = \frac{KV_0}{2} \left(\frac{V - V_0}{V_0} \right)^2 = \frac{2\pi K(R^3 - R_0^3)^2}{3R_0^3} . \quad (2.8)$$

To compute the same quantity from the Saint Venant–Kirchhoff model, we start by calculating the strain tensor for an isometric expansion

$$\mathbf{E}_{\text{exp}} = \begin{bmatrix} E_{xx} & 0 & 0 \\ 0 & E_{yy} & 0 \\ 0 & 0 & E_{zz} \end{bmatrix} = E_R \mathbf{I} .$$

Calculating the strain E_R for a varying radius R , we have

$$E_R = \frac{\left(\frac{R}{R_0}\right)^2 - 1}{2} = \frac{R^2 - R_0^2}{2R_0^2} ,$$

which leads to the energy density and elastic energy

$$\begin{aligned} \Psi(\mathbf{E}_{\text{exp}}) &= \frac{\lambda}{2}(3E_R)^2 + \mu(3E_R^2) = \left(\frac{3\lambda}{2} + \mu\right) 3E_R^2 \\ \Delta E_{\text{el}} &= \Psi(\mathbf{E}_{\text{exp}})V_0 \\ &= \left(\frac{3\lambda}{2} + \mu\right) 4\pi R_0^3 \left(\frac{R^2 - R_0^2}{2R_0^2}\right)^2 \\ &= \pi \left(\frac{3\lambda}{2} + \mu\right) \frac{R^4 - 2R^2R_0^2 + R_0^4}{R_0} . \end{aligned}$$

To test the numerical accuracy of our method, we followed a similar approach developed previously. We generated spheres of radius $R_0 = 20\text{\AA}$ with coordinates distributed in five layers of different radii and uniformly distributed polar and azimuthal angles. To apply a series of systematic compressions or expansions, we first converted the coordinate system to Cartesian and then scaled all the coordinates by factors ranging from 0.8 to 1.2. Then, we proceed to calculate the elastic energies as described before.

We observe that the elastic energy calculated numerically matches the analytical expression, see Fig. 2.3C. Interestingly, both curves for linear extension and bulk expansion display the same bi-quadratic form, despite depending on different pre-factors. The energy scale from both deformations is also comparable, ranging around $\sim 10^1 k_B T$ for small ratios and $\sim 10^2 k_B T$ for large ratios. We also note that the linear strain approximation remains close to the nonlinear prediction at intermediate strain

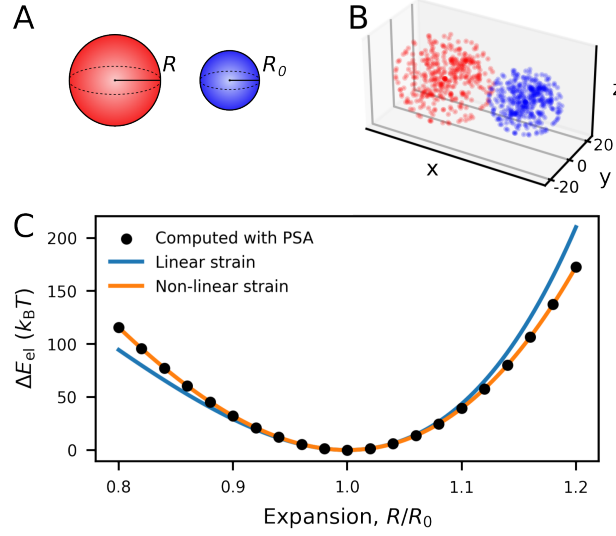


Figure 2.3: *Potential energy changes upon bulk expansion/contraction of a sphere.* **A.** Scheme representing the geometry of a sphere in two states: relaxed (blue) and expanded (red). **B.** Example of simulated relaxed and deformed states. **C.** Elastic energy as a function of the ratio between final and initial radii.

magnitudes. This might be explained by the higher order of the polynomial in R in the linear-strain elastic energy function.

2.3.3 Shear of a cylinder

Finally, we consider the case of a pure shear deformation, which, unlike the previous cases, implies no change in volume of a deformed body. In the linear regime, a material deforms in response to a shear stress σ_{xy} according to

$$\mu = \frac{\sigma_{xy}}{\varepsilon_{xy}} = \frac{F}{A} \frac{L}{\Delta x} = \frac{F}{A\gamma} \quad , \quad (2.9)$$

where μ is the shear modulus (note that it coincides with the second Lamé's constant), ε_{xy} is the shear strain – the response in x to a tangential force F applied to the surface of area A . The force F causes the surface to displace by a distance $\Delta x = \gamma L$, where γ represents the tangent of the deflection angle induced by shear (see Fig. 2.4A). The corresponding elastic energy associated with F is then

$$\Delta E_{\text{el}} = \frac{\mu A L \varepsilon_{xy}^2}{2} = \frac{\mu \pi R^2 L \gamma^2}{2} \quad . \quad (2.10)$$

While in the previous cases the formulation of the strain tensor was straightforward, here we derive it from the coordinate transformation under simple shear for clarity. To do so, we consider the coordinates in the relaxed state $\mathbf{X} = [x, y, z]$ and in the deformed state $\mathbf{x} = [x + \gamma y, y, z]$. This linear transformation corresponds to $\mathbf{F}_{\text{shear}}$ which is the deformation gradient for this particular deformation

$$\mathbf{F}_{\text{shear}} = \begin{bmatrix} 1 & \gamma & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} .$$

From Eq. 2.2, we calculate the Lagrange strain tensor $\mathbf{E}_{\text{shear}}$

$$\mathbf{E}_{\text{shear}} = \frac{1}{2} (\mathbf{F}^T \mathbf{F} - \mathbf{I}) = \frac{1}{2} \begin{bmatrix} 0 & \gamma & 0 \\ \gamma & \gamma^2 & 0 \\ 0 & 0 & 0 \end{bmatrix} .$$

Note that the typical off-diagonal terms are present due to shear, but also, a second-order main-diagonal correction appears.

This leads to an elastic energy

$$\begin{aligned} \Psi(\mathbf{E}_{\text{shear}}) &= \frac{\lambda}{2} \gamma^4 + \frac{\mu}{4} (2\gamma^2 + \gamma^4) \\ &= \frac{\mu}{2} \gamma^2 + \left(\frac{2\lambda + \mu}{4} \right) \gamma^4 \\ \Delta E_{\text{el}} &= \Psi(\mathbf{E}_{\text{shear}}) V \\ \Delta E_{\text{el}} &= \frac{\pi R^2 L}{2} \left[\mu \gamma^2 + \left(\frac{2\lambda + \mu}{2} \right) \gamma^4 \right] . \end{aligned}$$

Simulations were performed similarly to the previous sections, by generating a cylindrical object, applying shear transformations, and calculating the resulting energy from local deformations. In contrast to linear extension and bulk expansion, the elastic energy obtained for linear strain is exactly a second-order approximation of the nonlinear case. As a consequence, both models remain compatible up to larger magnitudes of shear (note that the magnitudes we tested vary up to 40%). Moreover, we observe that energy differences in shear are considerably smaller than the other two deformation types at similar strain magnitudes, likely due to the absence of volume changes.

Altogether, two main conclusions can be taken from the results we just presented: *(i)* the algorithm we implemented to analyse spatial coordinates upon deformation is extremely precise in retrieving the expected values of elastic energy difference between

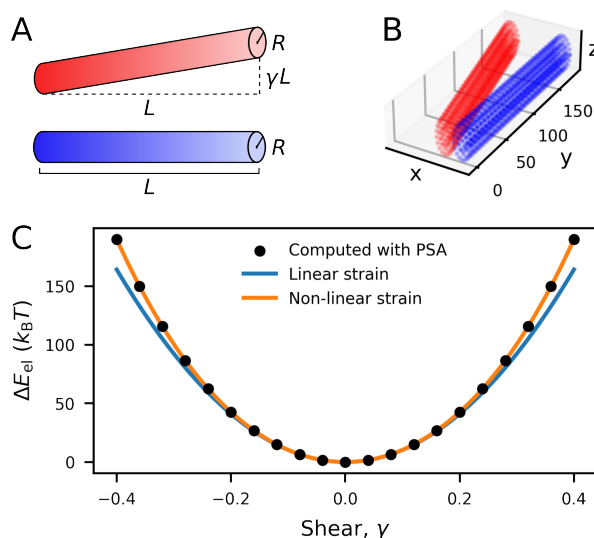


Figure 2.4: *Potential energy changes upon shear of a cylinder.* **A.** Scheme representing the geometry of a cylinder in two states: relaxed (blue) and sheared (red). **B.** Example of simulated relaxed and deformed states. **C.** Elastic energy at different shear rates γ , computed analytically and numerically with PSA.

deformed and relaxed states; and (ii), the three canonical types of deformation differ significantly between linear and nonlinear strain definitions for large strains. Since protein deformations often occur at scales comparable to their size, linear approximations may fail to correctly estimate elastic energies from structural data. This observation highlights the importance of using nonlinear elastic models such as the Saint Venant-Kirchhoff. In the next section, we apply this framework to real protein data.

2.4 Energy changes of haemoglobin states using structural data

While computing pseudoenergy differences between protein conformations requires little computational effort, identifying suitable examples to validate the proposed metric is considerably more challenging. An appropriate candidate for a test case must meet two key criteria: (i) the existence of an experimental estimate of free-energy difference between two conformations; and (ii) the availability of at least one example of solved structure for the same two conformations. To meet these requirements, we chose haemoglobin as a test case, a classical model system for studying allostery and

cooperativity at equilibrium that has been studied for over a century [24]. Nonetheless, as we will discuss later, different biophysical models yield free-energy estimates that can differ by as much as an order of magnitude.

Among the extensive literature on oxygen binding at haemoglobin’s four heme-containing sites [25], a notable model is the Monod-Wyman-Changeux (MWC) model, which introduced the causal relationship between conformational changes and cooperativity [26]. Also, this model was the first to refer to distinct conformational states of the haemoglobin tetramer, T (tense) and R (relaxed), with different affinities for binding oxygen. The nomenclature remained established in the field after experimental verification of distinct conformational states. We collected estimates of free-energy difference derived from various types of models for the O₂ saturation curves for haemoglobin, see Table 2.1.

Model type	Reference	Energy difference (kcal/mol)
Adair’s model	Imai, 1982 [27]	1.3 - 3.6*
Molecular code	Smith & Ackers, 1985 [28]	5.9 ± 0.2
Molecular code	Parrella et al., 1990 [29]	7
Molecular code	Huang et al., 1996 [30]	6.3 ± 0.2
Concerted (MWC)	Yonetani et al., 2002 [31]	0.7 - 3.1*
Elastic pseudoenergy	This study ($\alpha_2\beta_2$)	6.7 ± 0.5
Elastic pseudoenergy	This study ($2 \alpha\beta$)	5.2 ± 0.4

Table 2.1: *Energy difference between haemoglobin R and T states from biophysical models and elastic pseudoenergy.* The values for estimated energy difference were derived from fitting experimental data to cooperative models in various experimental approaches. *Estimates of free-energy obtained from experiments in different buffer conditions.

Taking advantage of multiple structures of haemoglobin assemblies solved at T and R states [32], we computed the pairwise elastic pseudoenergy using the same parameters described in *Methods*. To be precise, the dataset we used consists of crystallographic structures in the fully oxygenated state (R) and in the fully deoxygenated state (T), Fig. 2.5A. The R structures display a single $\alpha\beta$ pair per unit cell, whereas the T structure has two full tetramers per unit cell. With these structures, we performed two types of pairwise comparisons: first, considering $\alpha\beta$ chains, resulting in four R states and four T states (Fig. 2.5B); and second, considering the full tetramers $\alpha_2\beta_2$, resulting in four R states (applying symmetry operations) and two T states (Fig. 2.5C). Using this approach, we calculated the elastic pseudoenergy difference of R $\times$ R pairs, T $\times$ T pairs, and T $\times$ R pairs. The advantage of using both approaches is that we can account (or not) for elastic deformations in the interface between two

$\alpha\beta$ subunits. We also included the mean $\pm$ standard deviation for each group in Table 2.1 in units of kcal/mol to facilitate the comparison with the other observations.

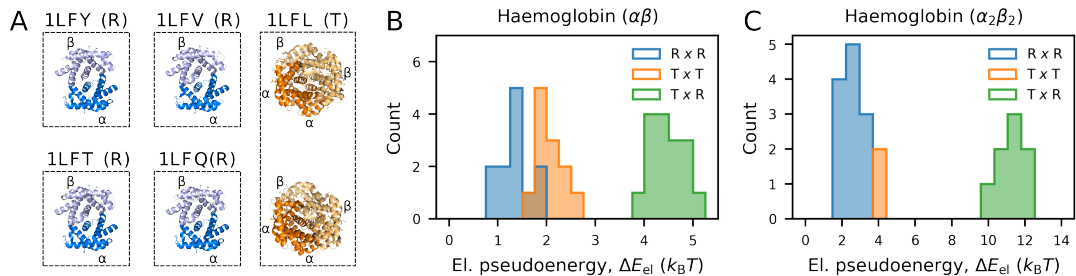


Figure 2.5: *Haemoglobin elastic pseudoenergy difference from pairwise comparisons.* **A.** Haemoglobin conformation in the R state (PDBs: 1LFQ, 1LFT, 1LFV, and 1LFY) and in the T state (PDB: 1LFL), depicted according to the unit cell from crystallographic structures. **B.** Comparison of individual $\alpha\beta$ pairs. With four pairs within each group, there is a total of 16 pairwise comparisons for R $\times$ R, T $\times$ T, and T $\times$ R. **C.** Comparison of $\alpha_2\beta_2$ tetramers. With four R and two T tetramers, the number of comparisons is 16 for R $\times$ R, 2 for T $\times$ T, and 8 for T $\times$ R.

Analysing the magnitude of elastic pseudoenergy in the dataset we used, we can pose interesting observations regarding haemoglobin conformational changes:

- Our results for T $\times$ R comparisons are in the same order of magnitude as those obtained with different methods and probed using very different experimental approaches. Besides, our estimate of elastic pseudoenergy falls within the equivalence range for the molecular code model. However, instead of supporting the validity of a specific model, we would rather interpret that our energy estimates match the order of magnitude of multiple biophysical models. Using strain analysis to corroborate a specific biophysical model of oxygen binding with haemoglobin requires a dedicated study for this purpose.
- Our observations reveal a consistent ordering of elastic pseudoenergy difference between types of comparison, namely, R $\times$ R < T $\times$ T < T $\times$ R. This suggests that (a) the energy difference between T and R states is significantly larger than those observed for T and R individually; and (b) the R states are less variable than T states, which is in remarkable agreement with the analysis performed by the authors who solved these structures [32]. While the first observation comes with no surprise, the second is counterintuitive given the labels Relaxed and Tense. This suggests that having oxygen bound to haemoglobin has a stabilising effect on its structure.

- Comparing the same groups using only two or all four chains provides an interesting pattern: whereas $R \times R$ and $T \times T$ comparisons with $\alpha\beta$ have exactly half of the elastic pseudoenergy of $\alpha_2\beta_2$, that is not the case for $T \times R$ comparisons. This is illustrated in the last two items of Table 2.1. This result implies that there is a significant signal of deformation localised in the interface between the two $\alpha\beta$ dimers exclusively in $R \rightleftharpoons T$ transitions. This particular observation is known since 1968, when the first structures of oxyhaemoglobin and deoxyhaemoglobin were compared [33].

In summary, here we present a system where both biophysical measures of cooperative energy and structural information were available for a pair of conformational states. Using protein elasticity, we find that the elastic pseudoenergy difference between these states corresponds to the free-energy difference derived from biophysical models, within the same order of magnitude.

2.5 Discussion

In this chapter, we extend the concept of elastic energy of deformation to protein structures by introducing the elastic pseudoenergy as a metric to quantify conformational changes at the residue level. We validated this framework through two complementary tests: first, aiming to recover the analytical results for the simplest deformation types in ideal bodies; and second, focusing on matching biophysical estimates of energy difference using real haemoglobin structures [32]. Beyond reaching a satisfactory validation on both tasks, we were also able to capture features of haemoglobin’s conformational transitions, such as localised deformation patterns and the level of dissimilarity between states, corroborating previous findings [32, 33].

Concerning energetics of protein structures, previous approaches have focused on the construction of energy landscapes from the raw ensemble of single-molecule cryo-EM images [2, 34, 35, 36] and molecular dynamics [37]. The key distinctive feature of our method is that it provides energetic information at the residue-scale. To achieve this, we relied on a simple yet strong assumption: that the deformation of a protein relative to a reference state is well captured by a homogeneous and isotropic elastic model. While proteins have long been ascribed elastic properties [19, 21, 7], our model assumptions are a clear oversimplification. For example, proteins are generically anisotropic and non-homogeneous materials, as they deform more easily in specific directions and on specific regions. Despite this critical disadvantage, we show that the elastic pseudoenergy is consistent with biophysical estimates for the case of haemoglobin. A broader benchmarking across larger datasets of pairs of conformations

for which the energy difference is known would not only constitute a more solid validation of the approach but also be useful to calibrate the stiffness constants used in our model.

A natural extension of our method is to incorporate Elastic Network Models (ENM) to estimate protein stiffness. In ENM, atoms in a single structure and within a cutoff distance are connected by springs, with coupling strengths inferred from β -factors [38]. While this framework does not directly translate to 3D continuum mechanics, it can inform analogous stiffness calibration strategies for PSA. The same idea applies to ENM variants that account for heterogeneity [39] and anisotropy [40]. Also, reference structures for deformations are likely pre-stressed [41, 42], which implies that simplistic elastic models such as the one considered here may under or overestimate the energy of a deformed conformation. A potential way of addressing this is to consider plastic deformations in addition to the standard elasticity theory as a framework to model proteins [43]. Finally, it is possible that further experimental advances in single-molecule mechanics can provide more detailed local information on protein stiffness.

By directly comparing pairs of structures, our method provides a unique procedure for exploring conformational changes. We argue that, when combined with powerful experimental methods that yield ensembles of conformations, this framework has the potential to support the interpretation of such datasets by providing quantitative analytical tools. Examples of that are applications to active research areas, such as mapping allosteric pathways [14], identifying the mechanical effect of mutations [44], sorting sets of conformations from ensembles (as we will discuss in the next chapter), and analysing the effect of ligands and post-translational modifications [45]. With the increasing influence of machine learning approaches to proteins, the elastic pseudoenergy could be integrated as a tool for protein dissimilarity instead of RMSD in structural comparison tasks [46]. Moreover, PSA could be coupled to generative tools to support the design of novel proteins that require large deformations to operate, such as molecular motors and molecular switches [11, 47, 48]. Finally, by being based on an energetic description, this interdisciplinary framework offers a bridge between structural biology and single-molecule biophysics datasets, thus supporting a physically grounded understanding of protein dynamics.

2.6 Methods

2.6.1 PSA parameter choice for the analysis of haemoglobin dataset

We considered the following PSA parameter choice to perform structural comparisons between (and within) oxy and deoxyhaemoglobin states:

1. *Set of atoms*: heavy atoms of the main chain were selected, namely C_α , C, N, and O; the elastic pseudoenergy per residue consists of the sum of elastic pseudoenergy of its atoms.
2. *Neighbourhood radius*: for a given assembly α , the set of atoms belonging to the local neighbourhood of i was defined as $\mathcal{N}_{\alpha,i} = \{j \mid \|\Delta \mathbf{x}_i^j\| < r\}$, with the cutoff radius $r = 9\text{\AA}$;
3. *Neighbourhood method*: to have a unified neighbourhood list per atom i , we considered the intersection of neighbourhoods of i for the two structures being compared, $\mathcal{N}_i = \mathcal{N}_{\alpha_1,i} \cap \mathcal{N}_{\alpha_2,i}$. Thus, all neighbouring atoms in $\mathcal{N}_i$ have weight $w_j = 1$;
4. *Protein volume*: To estimate the total van der Waals volume V , we used the software ProteinVolume [49] to estimate the volume of $\alpha\beta$ dimers of each PDB structure. As the variation of volume was small ($< 0.1\%$), we consider V to be the average of all estimates.

Bibliography

- [1] V. H. Mello, J. Wald, T. C. Marlovits, and P. Sartori, “Elastic analysis bridges structure and dynamics of an AAA+ molecular motor,” *PLOS Computational Biology*, vol. 21, no. 10, pp. 1–25, 2025. 23
- [2] A. Dashti, P. Schwander, R. Langlois, R. Fung, W. Li, A. Hosseinizadeh, H. Y. Liao, J. Pallesen, G. Sharma, V. A. Stupina, and Others, “Trajectories of the ribosome as a Brownian nanomachine,” *Proceedings of the National Academy of Sciences*, vol. 111, no. 49, pp. 17492–17497, 2014. 24, 36
- [3] C. Tian, K. Kasavajhala, K. A. A. Belfon, L. Raguette, H. Huang, A. N. Migue, J. Bickel, Y. Wang, J. Pincay, Q. Wu, and C. Simmerling, “ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution,” *Journal of Chemical Theory and Computation*, vol. 16, pp. 528–552, jan 2020. 24
- [4] J. Howard, *Mechanics of Motor Proteins and the Cytoskeleton*. Sunderland, MA: Sinauer, 1 ed., 2001. 24, 26
- [5] K. Kawaguchi and S. Ishiwata, “Nucleotide-Dependent Single- to Double-Headed Binding of Kinesin,” *Science*, vol. 291, no. 5504, pp. 667–669, 2001. 24
- [6] C. Cecconi, E. A. Shank, C. Bustamante, and S. Marqusee, “Direct Observation of the Three-State Folding of a Single Protein Molecule,” *Science*, vol. 309, no. 5743, pp. 2057–2060, 2005. 24
- [7] Y. Wang and G. Zocchi, “The folded protein as a viscoelastic solid,” *Europhysics Letters*, vol. 96, no. 1, p. 18003, 2011. 24, 27, 36
- [8] A. E. M. Beedle and S. Garcia-Manyes, “The role of single-protein elasticity in mechanobiology,” *Nature Reviews Materials*, vol. 8, no. 1, pp. 10–24, 2023. 24
- [9] E. Weinreb, J. M. McBride, M. Siek, J. Rougemont, R. Renault, Y. Peleg, T. Unger, S. Albeck, Y. Fridmann-Sirkis, S. Lushchekina, J. L. Sussman, B. A. Grzybowski, G. Zocchi, J.-P. Eckmann, E. Moses, and T. Thust, “Enzymes as viscoelastic catalytic machines,” *Nature Physics*, 2025. 24
- [10] G. M. Alushin, G. C. Lander, E. H. Kellogg, R. Zhang, D. Baker, and E. Nogales, “High-resolution microtubule structures reveal the structural transitions in $\alpha\beta$ -tubulin upon GTP hydrolysis,” *Cell*, vol. 157, no. 5, pp. 1117–1129, 2014. 24

- [11] J.-H. Ha and S. N. Loh, “Protein Conformational Switches: From Nature to Design,” *Chemistry – A European Journal*, vol. 18, no. 26, pp. 7984–7999, 2012. 24, 37
- [12] O. Keskin, S. R. Durell, I. Bahar, R. L. Jernigan, and D. G. Covell, “Relating Molecular Flexibility to Function: A Case Study of Tubulin,” *Biophysical Journal*, vol. 83, pp. 663–680, aug 2002. 24
- [13] S. Maki-Yonekura, K. Yonekura, and K. Namba, “Conformational change of flagellin for polymorphic supercoiling of the flagellar filament,” *Nature Structural Molecular Biology*, vol. 17, no. 4, pp. 417–422, 2010. 24
- [14] M. R. Mitchell, T. Tlusty, and S. Leibler, “Strain analysis of protein structures and low dimensionality of mechanical allosteric couplings,” *Proceedings of the National Academy of Sciences*, vol. 113, no. 40, pp. E5847—E5855, 2016. 25, 37
- [15] P. Sartori and S. Leibler, “Evolutionary conservation of mechanical strain distributions in functional transitions of protein structures,” *Physical Review X*, vol. 14, no. 1, p. 11042, 2024. 25
- [16] P. M. Gullett, M. F. Horstemeyer, M. I. Baskes, and H. Fang, “A deformation gradient tensor and strain tensors for atomistic simulations,” *Modelling and Simulation in Materials Science and Engineering*, vol. 16, no. 1, p. 15001, 2007. 25
- [17] G. A. Holzapfel, “Nonlinear solid mechanics: a continuum approach for engineering science,” 2002. 26, 28
- [18] M. Radmacher, M. Fritz, J. P. Cleveland, D. A. Walters, and P. K. Hansma, “Imaging adhesion forces and elasticity of lysozyme adsorbed on mica with the atomic force microscope,” *Langmuir*, vol. 10, no. 10, pp. 3809–3814, 1994. 26, 27
- [19] R. Afrin, M. T. Alam, and A. Ikai, “Pretransition and progressive softening of bovine carbonic anhydrase II as probed by single molecule atomic force microscopy,” *Protein science*, vol. 14, no. 6, pp. 1447–1457, 2005. 26, 27, 36
- [20] M. Guthold, W. Liu, E. A. Sparks, L. M. Jawerth, L. Peng, M. Falvo, R. Superfine, R. R. Hantgan, and S. T. Lord, “A comparison of the mechanical and structural properties of fibrin fibers with other protein fibers,” *Cell biochemistry and biophysics*, vol. 49, pp. 165–181, 2007. 26
- [21] M. Dong, S. Husale, and O. Sahin, “Determination of protein structural flexibility by microsecond force spectroscopy,” *Nature Nanotechnology*, vol. 4, no. 8, pp. 514–517, 2009. 26, 27, 36

- [22] H. J. Lessen, P. J. Fleming, K. G. Fleming, and A. J. Sodt, “Building blocks of the outer membrane: calculating a general elastic energy model for β -barrel membrane proteins,” *Journal of chemical theory and computation*, vol. 14, no. 8, pp. 4487–4497, 2018. 26
- [23] A. P. Perrino and R. Garcia, “How soft is a single protein? The stress–strain curve of antibody pentamers with 5 pN and 50 pm resolutions,” *Nanoscale*, vol. 8, no. 17, pp. 9151–9158, 2016. 27
- [24] C. Bohr, K. Hasselbalch, and A. Krogh, “Ueber einen in biologischer Beziehung wichtigen Einfluss, den die Kohlensäurespannung des Blutes auf dessen Sauerstoffbindung übt,” *Skandinavisches Archiv Für Physiologie*, vol. 16, no. 2, pp. 402–412, 1904. 34
- [25] W. A. Eaton, “A retrospective on statistical mechanical models for hemoglobin allostery,” *The Journal of Chemical Physics*, vol. 157, no. 18, p. 184104, 2022. 34
- [26] J. Monod, J.-P. Changeux, and F. Jacob, “Allosteric proteins and cellular control systems,” *Journal of Molecular Biology*, vol. 6, no. 4, pp. 306–329, 1963. 34
- [27] K. Imai, *Allosteric effects in haemoglobin*. Cambridge University Press, 1982. 34
- [28] F. R. Smith and G. K. Ackers, “Experimental resolution of cooperative free energies for the ten ligation states of human hemoglobin,” *Proceedings of the National Academy of Sciences*, vol. 82, pp. 5347–5351, aug 1985. 34
- [29] M. Perrella, A. Colosimo, L. Benazzi, M. Ripamonti, and L. Rossi-Bernardi, “What the intermediate compounds in ligand binding to hemoglobin tell about the mechanism of cooperativity,” *Biophysical Chemistry*, vol. 37, no. 1, pp. 211–223, 1990. 34
- [30] Y. Huang, M. L. Doyle, and G. K. Ackers, “The oxygen-binding intermediates of human hemoglobin: evaluation of their contributions to cooperativity using zinc-containing hybrids,” *Biophysical Journal*, vol. 71, no. 4, pp. 2094–2105, 1996. 34
- [31] T. Yonetani, S. Park, A. Tsuneshige, K. Imai, and K. Kanaori, “Global allostery model of hemoglobin. Modulation of O(2) affinity, cooperativity, and Bohr effect by heterotropic allosteric effectors,” *Journal of Biological Chemistry*, vol. 277, pp. 34508–34520, sep 2002. 34

- [32] B. K. Biswal and M. Vijayan, “Structures of human oxy- and deoxyhaemoglobin at different levels of humidity: variability in the T state,” *Acta Crystallographica Section D*, vol. 58, pp. 1155–1161, jul 2002. 34, 35, 36
- [33] W. Bolton, J. M. Cox, and M. F. Perutz, “Structure and function of haemoglobin: IV. A three-dimensional Fourier synthesis of horse deoxyhaemoglobin at 5.5 Å resolution,” *Journal of Molecular Biology*, vol. 33, no. 1, pp. 283–297, 1968. 36
- [34] J. Frank and A. Ourmazd, “Continuous changes in structure mapped by manifold embedding of single-particle data in cryo-EM,” *Methods*, vol. 100, pp. 61–67, 2016. 36
- [35] A. Dashti, G. Mashayekhi, M. Shekhar, D. Ben Hail, S. Salah, P. Schwander, A. des Georges, A. Singharoy, J. Frank, and A. Ourmazd, “Retrieving functional pathways of biomolecules from single-particle snapshots,” *Nature communications*, vol. 11, p. 4734, sep 2020. 36
- [36] J. Giraldo-Barreto, S. Ortiz, E. H. Thiede, K. Palacio-Rodriguez, B. Carpenter, A. H. Barnett, and P. Cossio, “A Bayesian approach to extracting free-energy profiles from cryo-electron microscopy experiments,” *Scientific Reports*, vol. 11, no. 1, p. 13657, 2021. 36
- [37] R. G. Viegas, I. B. S. Martins, M. N. Sanches, A. B. Oliveira Junior, J. B. de Camargo, F. V. Paulovich, and V. B. P. Leite, “ELViM: Exploring Biomolecular Energy Landscapes through Multidimensional Visualization,” *Journal of Chemical Information and Modeling*, vol. 64, pp. 3443–3450, apr 2024. 36
- [38] M. M. Tirion, “Large Amplitude Elastic Motions in Proteins from a Single-Parameter, Atomic Analysis,” *Phys. Rev. Lett.*, vol. 77, no. 9, pp. 1905–1908, 1996. 37
- [39] M. Li, J. Z. H. Zhang, and F. Xia, “Heterogeneous elastic network model improves description of slow motions of proteins in solution,” *Chemical Physics Letters*, vol. 618, pp. 102–107, 2015. 37
- [40] A. R. Atilgan, S. R. Durell, R. L. Jernigan, M. C. Demirel, O. Keskin, and I. Bahar, “Anisotropy of Fluctuation Dynamics of Proteins with an Elastic Network Model,” *Biophysical Journal*, vol. 80, no. 1, pp. 505–515, 2001. 37
- [41] M. Krieg, A. R. Dunn, and M. B. Goodman, “Mechanical control of the sense of touch by β -spectrin,” *Nature Cell Biology*, vol. 16, no. 3, pp. 224–233, 2014. 37

- [42] R. G. Parra and D. U. Ferreira, “Frustration, dynamics and catalysis,” 2025. 37
- [43] E. Efrati, E. Sharon, and R. Kupferman, “The metric description of elasticity in residually stressed soft materials,” *Soft Matter*, vol. 9, no. 34, pp. 8187–8197, 2013. 37
- [44] J. M. McBride, K. Polev, A. Abdirasulov, V. Reinharz, B. A. Grzybowski, and T. Thusty, “AlphaFold2 Can Predict Single-Mutation Effects,” *Phys. Rev. Lett.*, vol. 131, p. 218401, nov 2023. 37
- [45] M. Correa Marrero, V. H. Mello, P. Sartori, and P. Beltrao, “Global comparative structural analysis of responses to protein phosphorylation,” *Nature Communications*, vol. 16, no. 1, p. 9407, 2025. 37
- [46] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Židek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, and D. Hassabis, “Highly accurate protein structure prediction with AlphaFold,” *Nature*, vol. 596, no. 7873, pp. 583–589, 2021. 37
- [47] T. Kortemme, “De novo design of protein structure and function with RFdiffusion,” *Cell*, vol. 187, pp. 526–544, feb 2024. 37
- [48] A. Courbet, J. Hansen, Y. Hsia, N. Bethel, Y.-J. Park, C. Xu, A. Moyer, S. E. Boyken, G. Ueda, U. Nattermann, D. Nagarajan, D.-A. Silva, W. Sheffler, J. Quispe, A. Nord, N. King, P. Bradley, D. Veisler, J. Kollman, and D. Baker, “Computational design of mechanically coupled axle-rotor protein assemblies,” *Science*, vol. 376, no. 6591, pp. 383–390, 2022. 37
- [49] C. R. Chen and G. I. Makhatadze, “ProteinVolume: calculating molecular van der Waals and void volumes in proteins,” *BMC Bioinformatics*, vol. 16, no. 1, p. 101, 2015. 38

Elastic deformations of an AAA+ motor reveal the structural basis for hand-over-hand coordination

Proteins carry out cellular functions by changing their structure among a few conformations, each characterised by a different energy level. Therefore, structural changes, energy transformations, and protein function are intimately related. Despite its central importance, this relationship remains elusive. For example, while many AAA+ assemblies are known to function using a hand-over-hand alternation of subunits, how energy transduction throughout the assembly structure drives hand-over-hand is not known. In this work, we unravel the energetic basis of hand-over-hand in a model AAA+ motor, RuvB. To do so, we apply our elastic pseudoenergy approach to an ensemble of RuvB structures, recently resolved through cryo-EM. This allows us to quantify how progression through RuvB’s mechanochemical cycle translates into residue-scale energy transduction. In particular, we find that DNA binding is associated with overcoming a high energy barrier. This is possible through inter-subunit transmission of energy, and is ultimately driven by binding ATP. Overall, our work elucidates the energetic basis for the hand-over-hand mechanism in RuvB’s cycle. Furthermore, it presents a generally applicable methodology for studying the energetics of conformational cycles in other proteins.

This chapter partially includes an adaptation of the publication “*Elastic analysis bridges structure and dynamics of an AAA+ molecular motor*” [1], Mello VH, Wald J, Marlovits TC, Sartori P. Author contributions: Conceptualisation, VHM, JW, TCM, and PS; Formal analysis, VHM; Writing – Original draft, VHM; Writing – Review and editing, VHM, JW, TCM and PS; Visualisation, VHM; Supervision, TCM and PS.

3.1 Introduction

AAA+ is a superfamily of protein assemblies that catalyse a wide range of cellular functions, including protein degradation and DNA recombination. These motors typically operate by translocating a polymeric substrate through the central pore of the assembly, a process driven by the ATP hydrolysis cycle [2]. Recent cryo-EM structures of hexameric AAA+ have consistently shown that they adopt an asymmetric assembly architecture, where individual subunits organise as a staircase around the translocated substrate [3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15]. Based on this structural insight, it was suggested that subunits coordinate to move relative to the substrate by sequentially gripping and releasing it in a *hand-over-hand* fashion [2]. However, the underlying basis for such orchestrated motion remains unclear.

From a thermodynamic perspective, hand-over-hand coordination arises from energy transduction. Energy transduction is fundamental to all molecular motors because it couples the energy released from a chemical cycle – ATP hydrolysis and nucleotide exchange – to the motion produced on a mechanical cycle – substrate translocation, Fig. 3.1A. Crucially, chemical energy is not only needed to sustain motion but also to ensure coordination of different components [16, 17, 18, 19, 20]. However, despite the central role of energy transduction in protein function [21, 22, 23], how this energy is distributed throughout the structure of proteins is largely unknown. Consequently, how to connect the energetics of protein function with detailed structural information remains an open question. In this chapter, we make a step towards characterising the energetic basis of the hand-over-hand mechanism in an AAA+. Our main assumption is that elastic deformations on the protein structure, quantified using the elastic pseudoenergy ΔE_{el} , are a key element to transiently store and later transduce energy throughout the mechanochemical cycle.

Here we use PSA, the framework described in the previous chapter, to study how energy transduction promotes hand-over-hand in RuvB, an AAA+ hexameric assembly. RuvB is a molecular motor that translocates DNA in the Holliday Junction, hence driving branch migration in bacteria, Fig. 3.1A. We base our analysis on RuvB structures engaged in its mechanochemical cycle, recently visualised by time-resolved cryo-EM [3]. We first show that a simple principle of minimal elastic deformation can sort the conformational cycle of RuvB, in agreement with the result of manual sorting. We then identify the spatial distribution of elastic pseudoenergy along the conformational cycle. With this information, we propose that two distinct steps of the hand-over-hand cycle have specific energetic requirements. The *hand out*, or detachment from DNA, does not involve significant changes in elastic pseudoenergy,

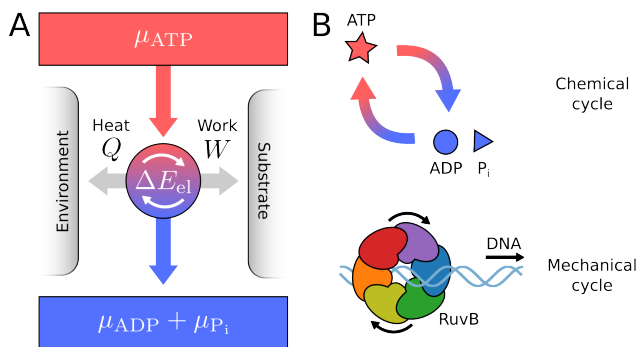


Figure 3.1: *Elastic deformations mediate energy transduction.* **A.** Energy transduction through a molecular motor occurs when the difference in free-energy between ATP (μ_{ATP}), and ADP and P_i ($\mu_{\text{ADP}} + \mu_{\text{P}_i}$) drives cyclic deformations in the motor (ΔE_{el}), resulting in work performance and heat release. **B.** RuvB is a hexameric AAA+ molecular motor which transduces chemical energy into the motion of a DNA molecule through its central pore.

whereas the DNA *hand in* is associated with overcoming a large barrier. Altogether, these findings suggest that the energy needed to overcome the energetic barrier comes from nucleotide exchange, and is transmitted across assembly monomers.

3.2 Results

3.2.1 Cycle reconstruction using minimal deformation principle

To describe energy transduction of a protein at the structural level it is necessary to resolve the multiple conformations it adopts along its mechanochemical cycle. These “structural snapshots” can be inferred, for example, from structural variability in the cryo-EM ensemble [24]. Once the conformations are obtained, a key task consists of finding their correct time order in the cycle, typically through visual inspection. Can an elasticity-based approach provide a systematic solution to this task?

For the case of the RuvB homohexamer, we consider five structures of the assembly ($s_1, \dots, s_5$) and a putative reference not engaged on the cycle s_0 , which were obtained and manually ordered in [3], Fig. 3.2A and *Methods*. Upon the completion of one cycle, each of the RuvB subunits switches to the next position relative to the DNA molecule. Altogether, the five cycle hexamers and the position switching rule imply that RuvB subunits perform a cycle with 30 conformations, Fig. 3.2B. From now on, we refer to each of these conformations by a letter, indicating a position relative to the DNA (with A in the top of the staircase, followed by B, C, $\dots$), and a number,

indicating the numbering of the structure, e.g., F1, F2, ..., A5. Crucially, ordering the five structural snapshots as done in [3] involved extensive and time-consuming geometrical analysis, visual inspection, and careful evaluation of nucleotide-binding sites' electronic densities. We now provide an alternative approach for time ordering based on protein elasticity.

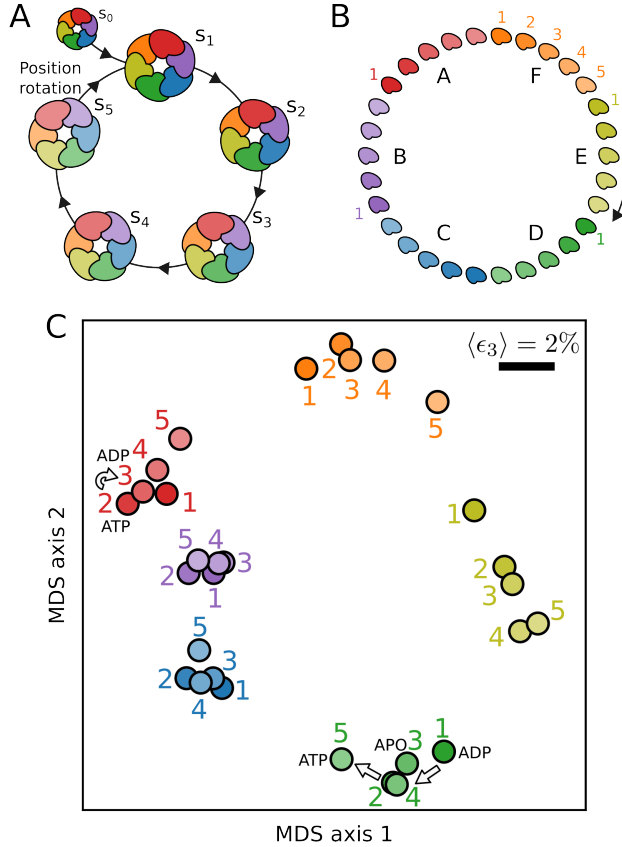


Figure 3.2: *Minimal deformation principle reconstructs RuvB monomer cycle.* **A.** Sketch of RuvB hexameric cycle (states $s_1, \dots, s_5$) together with the initial state (s_0). In the final transition, $s_5 \rightarrow s_1$, subunits switch to the next position. **B.** The ordered assembly states together with the position switching rule imply that each RuvB subunit undergoes a cycle of 30 conformations. **C.** Low-dimensional embedding of pairwise comparisons between RuvB subunits is arranged in a cycle. Conformations cluster according to their position and a parsimonious order from 1 to 5 can be deduced for positions F, A, E and D. We use Multidimensional Scaling on the mean strain values, $\langle \epsilon_3 \rangle$, for each pairwise comparison. The length of the scale bar indicates a mean strain of 2% for both axes. The plot depicts the mean of 100 individual MDS realisations.

Specifically, we use a principle of minimal elastic deformation, by which subsequent

structures in the cycle correspond to the smallest deformations. First, we calculate the residue-scale strain tensor – a quantity foundational to elasticity theory that measures local deformation [25] – over all possible pairs of conformations. Second, from this tensor we extract the local principal strain ϵ_3 , compute its average over all residues, and construct a matrix of size 30×30 with all pairwise comparisons. Pairs of conformations to which correspond a large matrix element are very different, as they relate to each other by a large deformation. Third, to interpret the matrix of deformations, we project it in two dimensions using Multidimensional Scaling (MDS) [26, 27]. The resulting projection shows that subunits cluster in agreement with their position in the assembly (different colors), Fig. 3.2C. Furthermore, within these clusters, we can identify the order of the cycle substeps (1 to 5), which corresponds to the temporal order of the hexameric assemblies. Overall, our unbiased approach independently validates the ordering of structural snapshots in the mechanochemical cycle of RuvB.

3.2.2 Elastic pseudoenergy of the mechanochemical cycle at the subunit scale

Having shown the importance of elasticity in recovering the RuvB cycle, we now compute the elastic pseudoenergy of all thirty conformations. To do so, we use as reference an assembly structure, s_0 , that is not engaged in the mechanochemical cycle [3] and compute the elastic pseudoenergy of each subunit compared to the reference, see *Methods* for details. Figure 3.3A shows the magnitude of elastic pseudoenergy as a heat map on each conformation. We find that the spatial distribution of elastic pseudoenergy is highly heterogeneous, with small regions accumulating most of the pseudoenergy. We now quantitatively analyse these energetic patterns.

To obtain a global view of RuvB energetics we compute the total elastic pseudoenergy per subunit, Fig. 3.3B. The resulting pseudoenergy landscape is smooth and consists of a steady build-up of an energetic barrier in DNA free states and a decrease to a flat landscape in DNA bound states. Note that the highly energetic conformations (D, E and F) correspond to the “converter module” hypothesised in [3] to mediate energy transduction. In the following, we focus on the spatial distribution of elastic energy at a local scale.

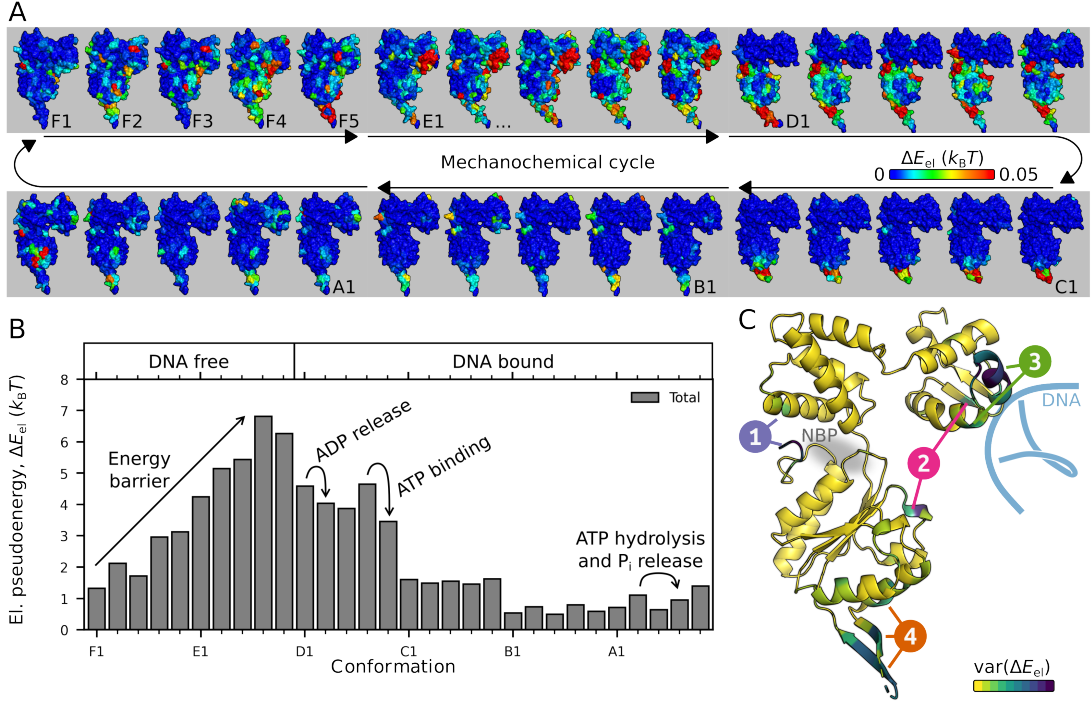


Figure 3.3: *Elastic pseudoenergy content of a RuvB subunit along its mechanochemical cycle.* **A.** We calculate the elastic pseudoenergy content in the RuvB's protein structure at a residue level for its conformational cycle in units of $k_B T$, with k_B the Boltzmann constant and T the temperature. **B.** The sum of all elastic pseudoenergy is represented throughout the cycle, representing RuvB's pseudoenergy landscape. On top, the conformations are labelled according to their nucleotide content and engagement or not with the substrate. **C.** RuvB subunit (A1 conformation in the figure) color-coded with the variance of elastic pseudoenergy across conformations. The most variable residues, grouped into four mechanically active regions, are marked by dark blue colors. Note that region 1 localises close to the nucleotide-binding pocket, and region 3 localises close to the DNA molecule. The regions are numbered and colored consistently throughout the manuscript.

3.2.3 Residues with high pseudoenergy variance map to four distinct functional regions

At a local scale, we describe the elastic pseudoenergy for each residue by a trajectory over conformations. These trajectories show that different residues are more deformed in specific states of the mechanochemical cycle. We identify four mechanically active regions based on the variability of the pseudoenergy trajectories, Fig. 3.3C. Interestingly, all four regions are discontinuous in sequence but perform similar interaction patterns (see Fig. 3.4), suggesting that despite containing relatively distant residues,

each region is consistently related to the same function.

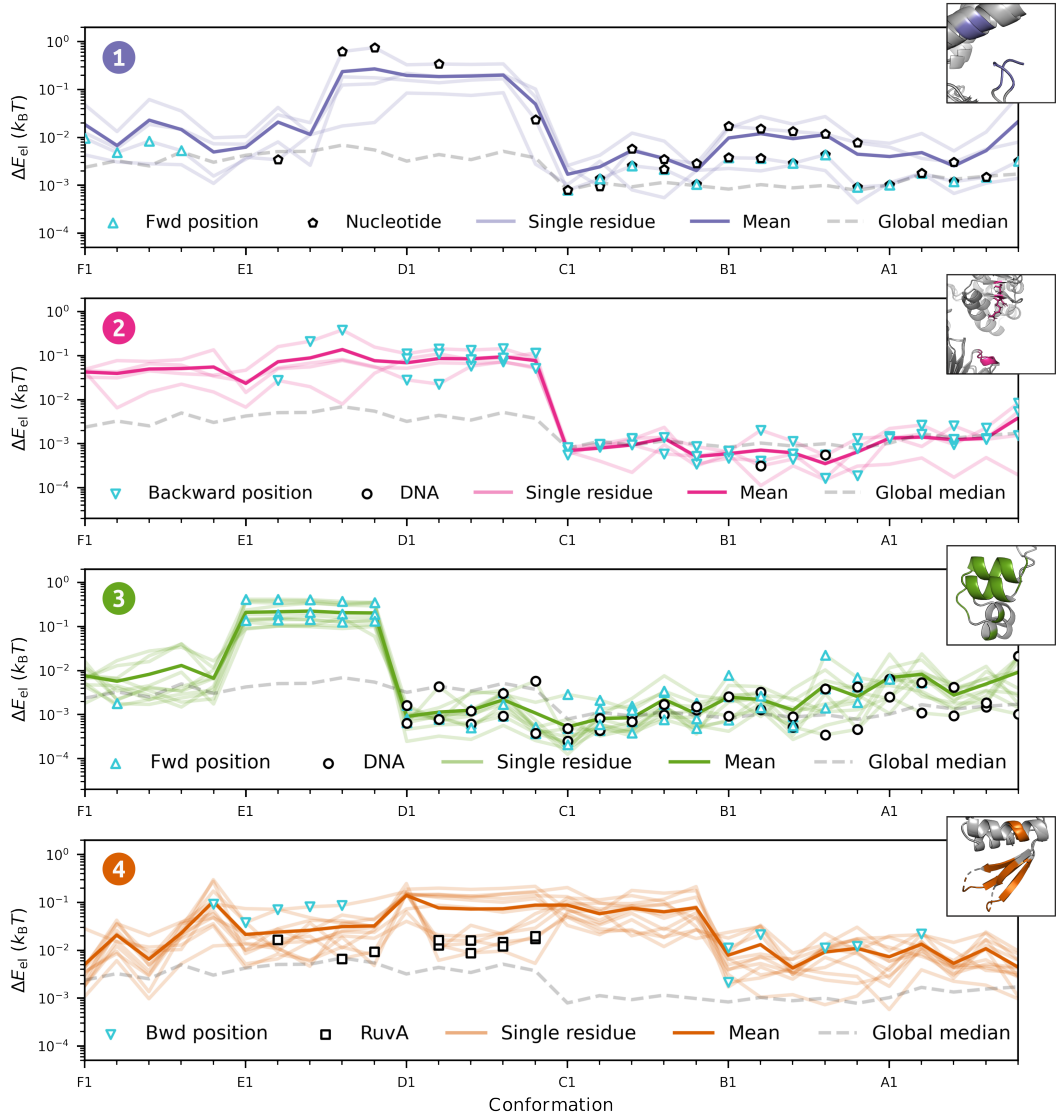


Figure 3.4: *Elastic pseudoenergy profiles and intermolecular interactions performed by the mechanically active regions.* Elastic pseudoenergy profile of regions 1 to 4 and their specific intermolecular interactions. On the top right corner of each panel, the structure of the four regions is displayed in two conformations with different elastic pseudoenergy contents.

Region 1 comprises part of the nucleotide-binding pocket and connects the small and large ATPase domains. Region 2 localises mostly to sensor 1, a loop motif conserved in all AAA+ [28] whose specific role is critical for function but not consensual

among studies [29, 30, 2, 31]. Region 3 belongs to the head domain (C-terminal), which interacts with the DNA substrate. Region 4 largely coincides with a conserved structural motif named pre-sensor I β -hairpin, an insertion common to PS1 clade of the AAA+ superfamily, crucial for function [32]. In the context of RuvB, this motif provides direct and transient protein-protein interactions between RuvB and RuvA [33].

To summarise, we have identified the pseudoenergy landscape of RuvB and key mechanically active regions in the structure. In the following, we study energy transduction through these regions and its connection to hand-over-hand. To this end, we analyse energetic exchanges with the DNA substrate (region 3, related to translocation), across RuvB subunits (regions 2 and 3, related to inter-subunit coupling), and with the nucleotide (region 1, related to chemical driving). Deformations in region 4 relate to the dynamic interactions with RuvA, therefore playing a role in the assembly cohesion, which is outside the scope of this paper. Altogether, these elements compose the energetic contributions for the function of a molecular motor, Fig. 3.1A.

3.2.4 Energy is released upon transition to DNA-bound state

A central aspect of molecular motor energetics is that they perform mechanical work by moving relative to a substrate [34]. For some AAA+, the motion of the substrate through the pore arises from a sequential hand-over-hand action of the different subunits [2], as is apparent by the staircase arrangement of the subunits around DNA, Fig. 3.5A. To quantify the energetics of this translocation process, we identify all the residues in RuvB that establish polar interactions with DNA across each conformation, (see *Methods* for the procedure to compute polar interactions). The RuvB/DNA interface is highly dynamic and comprises six residues, some of which belong to regions 2 and 3. Because most of these interactions are established in conformation D1 and are lost in conformation A5, we name *hand-in* the transition E $\rightarrow$ D, and *hand-out* the transition A $\rightarrow$ F.

Figures 3.5B and C show the energy content of region 3 residues in the RuvB/DNA interface. Remarkably, we find that these residues interact with DNA when they are in a low pseudoenergy state, colored in blue. Conversely, they exhibit high elastic pseudoenergy when they do not interact with the DNA, Fig. 3.5D. When compared to other interactions, such as with the nucleotide and across RuvB subunits, interactions with DNA are characterised by a low elastic pseudoenergy, Fig. 3.5E. Overall, this suggests that after DNA hand-out the interface accumulates energy that is then released upon DNA hand-in. Therefore, DNA stabilises the interface, see Fig. 3.5F for schematics. Next, we investigate how inter-subunit interactions play a role in trans-

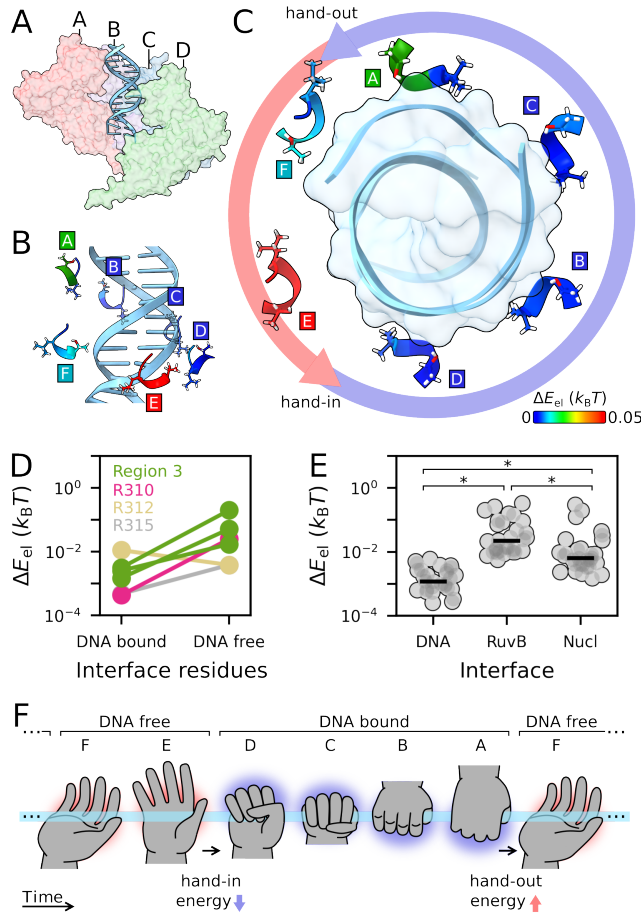


Figure 3.5: *Energetics of the RuvB/DNA interface.* **A.** Global view of the RuvB subunits A, B, C and D arranged as a staircase around the DNA (assembly state s_5). **B/C.** Residues G279-T282, belonging to region 3, colored by elastic pseudoenergy (side view in B and top view in C). **D.** Mean elastic pseudoenergy content of residues in DNA bound and DNA free states. Except for R312, all residues display higher elastic pseudoenergy when not interacting with the DNA, which corresponds mostly to positions F and E. **E.** Total elastic pseudoenergy of different interfaces, with each data point corresponding to a conformation. RuvB/DNA interface is less energetic than RuvB/RuvB and RuvB/Nucleotide interfaces throughout the mechanochemical cycle (Kruskal-Wallis test $p < 0.05$; Mann-Whitney U test, $p < 0.05$, Bonferroni corrected). **F.** Sketch representing “hand-over-hand” with the energy content of RuvB/DNA interface appearing as a red (high energy) or blue (low energy) glow. Each hand represents the same RuvB subunit progressing over time through its conformational cycle (from left to right). For visual clarity, only six of the thirty conformations are shown.

mitting energy to this interface during the DNA free conformations, i.e. between *hand-out* and *hand-in*.

3.2.5 Inter-subunit coupling drives DNA hand-in

Above, we showed that in DNA free states RuvB increases its energy, i.e., it “climbs” an energy barrier. How can these energetically unfavourable dynamics unfold? To address this question, we now investigate energy transduction across neighbouring RuvB subunits by characterising their interface. As shown in Fig. 3.6A, we define the forward subunit (pink) as the subunit forward to the interface according to the cycle orientation, and correspondingly for the backward subunit (gray). We find that regions 2 and 4 dominate the elastic pseudoenergy at the forward subunit (Fig 3.6B), whereas region 3 accounts for most of the elastic pseudoenergy at the backward subunit (Fig 3.6C). In particular, elastic pseudoenergy is high when the forward subunit is in the DNA bound position D and the backward subunit in the DNA free position E. Therefore, the interface between subunits D/E has increased elastic pseudoenergy on both sides.

We now focus on the interaction between regions 2 and 3, which come in contact with each other from opposite sides of the interface throughout the cycle (Fig. 3.6D-F, and Fig. 3.4 for the complete interaction profile). From the inspection of these regions at a molecular level, we observe two noticeable patterns: *(i)* in region 3 at the backward subunit, the highly energetic states correspond only to conformation D1 to D5, where a large displacement of the whole head domain is observed; and *(ii)* in region 2 at the forward subunit, the increase and decrease of elastic pseudoenergies correlate with switches in the secondary structure from a loop (high energy, F1 to D5) to a 3_{10} helix conformation (low elastic pseudoenergy, C1 to A5). Note that 3_{10} helices are sterically unfavourable [35, 36], suggesting that folding can be involved in transient energy storage. In summary, these deformation patterns indicate that energy storage and release in region 2 resemble a spring-like mechanism, by which one subunit assists the displacement and repositioning of its neighbour. Such rearrangement accommodates RuvB within the double-helix minor groove, a contact that establishes the basis of hand-in for RuvB, see Fig. 3.6G.

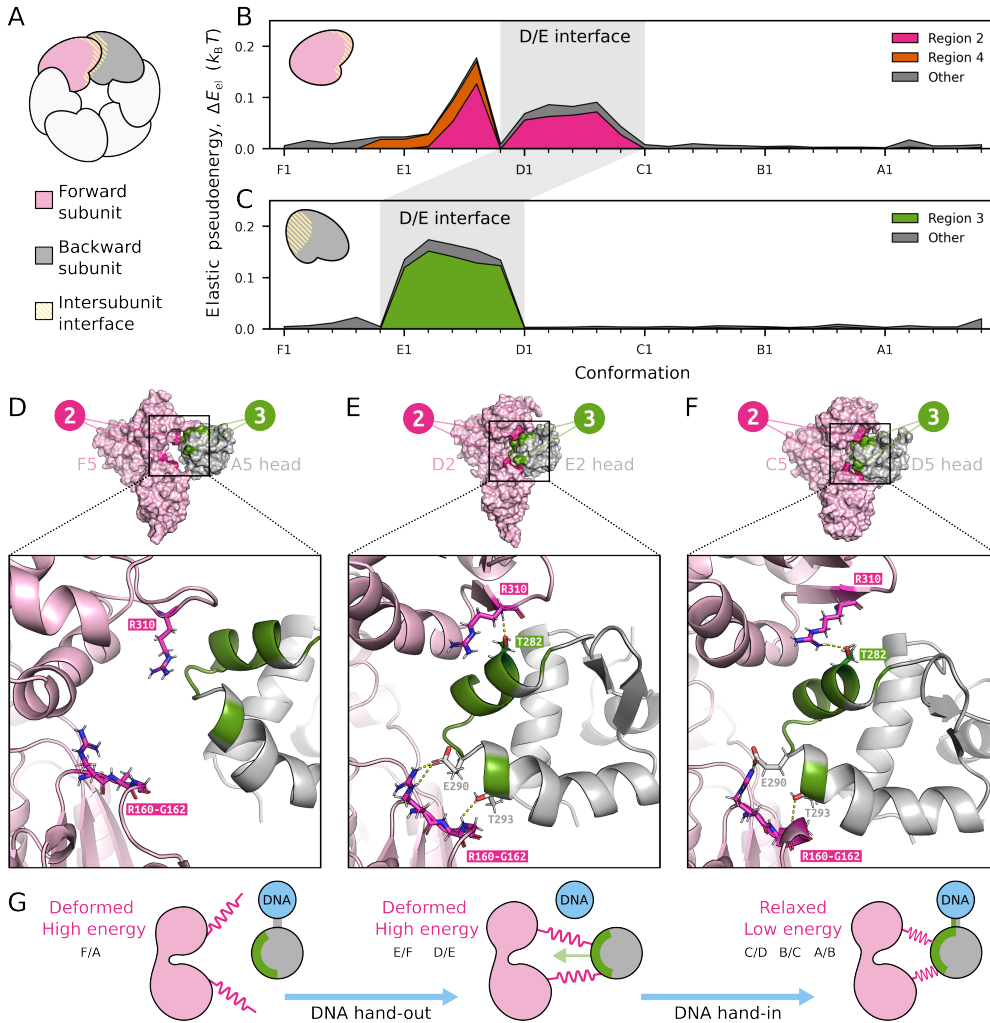


Figure 3.6: *Energy transduction across adjacent RuvB subunits.* **A.** Schematic of the interface between the forward and the backward subunits. The nomenclature *backward* and *forward* refers to their order on the conformational cycle. **B/C.** Total elastic pseudoenergy at the interface on the forward subunit (panel B) and backward subunit (panel C). Regions 2 and 4 represent a large fraction of the total elastic pseudoenergy of the former and region 3 of the latter. **D/E/F.** Representative structures of the interface dynamics: F5/A5 in panel D, where no interactions are found across subunits; D2/E2 in panel E, where these regions display polar interactions and residues R160-G162 adopt a loop structure; and C5/D5 in panel F, where regions 2 and 3 interact, but R160-G162 adopt a 3_{10} helix structure. **G.** Schematic of the interaction and energy patterns between neighbouring RuvB monomers: forward subunit is in a high-energy state, it then approaches and binds the backward subunit, finally it releases energy as the backward subunit moves and binds to DNA.

3.2.6 Nucleotide exchange is the energy source of the mechanochemical cycle

So far we have shown that the mechanochemical cycle of RuvB is characterised by an energy build-up just before DNA hand-in, and that this build-up is associated with inter-subunit energy coupling. What is the ultimate source of this energy?

Following the framework in Fig. 3.1A, it is clear that the source of energy is in the nucleotide cycle, and thus we turn our attention to the nucleotide binding pocket (NBP). The NBP is near region 1, identified in Fig. 3.3. This small region presents two different conformations, closed and open, acting as a gate, Fig. 3.7A. In the open conformation, both ADP release and ATP uptake take place; in the closed conformation, hydrolysis takes place.

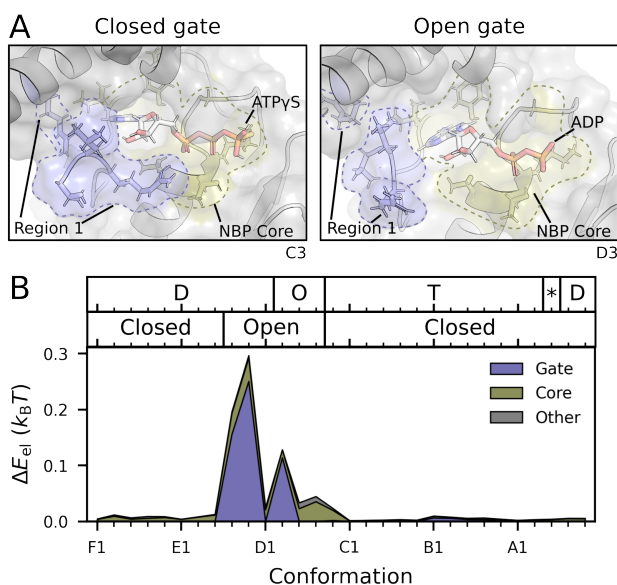


Figure 3.7: *Nucleotide exchange drives hand-over-hand.* **A.** Region 1, which surrounds the NBP, adopts two distinct conformations: closed, when the nucleotide is ATP γ S (a slowly hydrolysable ATP analogue); and open, when the nucleotide is ADP. **B.** The elastic pseudoenergy of residues interacting with the nucleotide is high and dominated by region 1 when this region adopts its open conformation. Top labels indicate nucleotide state in each conformation (D for ADP, O for empty, T for ATP, and * for ADP + P_i).

We find that region 1 displays a high pseudoenergy while in the open conformation, when it also interacts with the nucleotide, see Fig 3.7B. In fact, in its open conformation region 1 accounts for 60% of the elastic pseudoenergy of residues interacting with the nucleotide, largely surpassing the core of the NBP, which is mostly

composed by the Walker A motif residues, a conserved motif responsible for binding ATP. In contrast, region 1 has low pseudoenergy in its closed conformation.

This implies that the imbalance in chemical potential, which fuels RuvB dynamics, has a direct consequence on structural deformations during nucleotide exchange, but not hydrolysis. Furthermore, the open conformation of region 1 co-occurs with the conformations in which the interface between subunits is highly energetic, Fig. 3.6. Taken together, these observations suggest that the free-energy of ADP binding is transduced across subunits to mediate the highly energetic hand-in interaction. This completes the energy transduction cycle behind hand-over-hand and substrate repositioning.

3.3 Discussion

Protein function depends on structural deformations. Despite much of these deformations being caused by spontaneous thermal fluctuations, coordinated motion requires energy input [37]. This is particularly notable for motors, such as those in the AAA+ superfamily, in which function is directly driven by the free-energy released from the nucleotide cycle. Despite this intimate relationship, protein energetics and function remain largely disconnected at the structural scale. In this chapter, we have linked these two aspects of proteins for a particular AAA+ assembly, RuvB.

3.3.1 New applications of elastic analysis to conformational ensembles

Elastic deformations have previously been shown to identify allosteric sites [38], exhibit evolutionary conservation across species [39], and quantify the structural effect of point mutations [40] and phosphorylations [41]. Here, we provide two new applications of PSA using RuvB as a model system. First, we demonstrate that a principle of minimal deformation can recover the temporal order of multiple conformations belonging to a mechanochemical cycle. Second, we use the variability of deformation patterns across a conformational ensemble to identify functionally relevant regions. While we apply these methodologies to RuvB, they are completely general and can be deployed to any protein for which multiple structures have been experimentally or computationally resolved.

Several established methods also aim to extract insights from structural variability. Examples are methods relying on Ensemble PCA or NMA [42, 43, 44, 45], as well as methods aimed at heterogenous cryo-EM reconstructions [46, 24, 47]. Compared to the first group, our approach analyses dissimilarities between pairs of structures,

instead of structure-specific features such as atomic coordinates (for ensemble PCA) or their vibrational modes (for ensemble NMA). An additional difference is that PSA is independent of structural alignment, thus better characterising motions between rigid domains [38, 39]. Compared to the second group, our approach presents the key difference that it relies on atomic maps, not raw EM micrographs. On the one hand, this makes PSA suitable to assess structures regardless of their acquisition approach (cryo-EM, X-ray diffraction, NMR, or computational structure prediction). On the other hand, our method is sensitive to the accuracy of the atomic-map reconstruction. A natural next step is to adapt the elastic approach to process raw experimental data, e.g. electron densities. Overall, our method provides a readily interpretable low-dimensional representation of a set of structures.

3.3.2 Biological aspects of hand-over-hand in RuvB

In oligomeric AAA+ ATPases, hand-over-hand establishes that each subunit within the assembly sequentially alternates its conformation as it interacts with a substrate. Through a series of structural snapshots depicting actively processing RuvB in the RuvAB-Holliday Junction complex, the hand-over-hand mechanism was observed for RuvB [3]. After each cycle, each RuvB subunit adopts the conformation of its preceding neighbour. These conformations, characterised by their helical positioning relative to the DNA, play distinct roles in the DNA hand-over-hand process.

Our energetic analysis provides insights into these roles and how they interrelate. We found that the progression of the mechanochemical cycle is associated with overcoming an energy barrier during DNA-free conformations (F and E), see Fig. 3.8. Our findings suggest that mechanical coupling between subunits D and E plays a crucial role in assisting with the crossing of this barrier. Furthermore, we found that elastic deformations in the NBP during nucleotide exchange, rather than ATP hydrolysis, supply the energy that is transduced through inter-subunit coupling. This is in line with observations in other molecular motors, see [48, 49, 50]. In conclusion, an energetic cascade that starts in the NBP and is transmitted across chains leads to DNA hand-in, and constitutes the energetic basis of hand-over-hand in RuvB.

3.4 Methods

3.4.1 Structural data

The cryo-EM structures were downloaded from the Protein Data Bank (PDB) and are designated throughout the text as s_0 : 7PBR, s_1 : 7PBL, s_2 : 7PBM, s_3 : 7PBN,

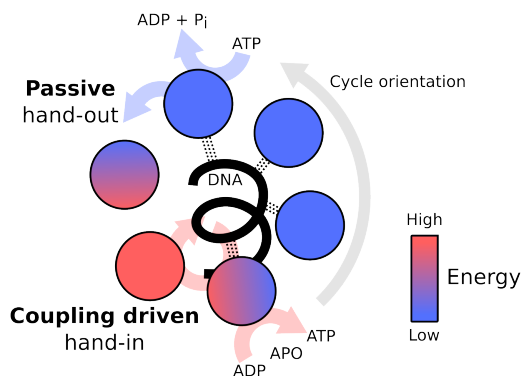


Figure 3.8: *An energy-inspired biological model for RuvB operation.* RuvB subunits (circles) cycle between six positions relative to DNA. Curved arrows represent five different processes: cycle orientation, through which subunits change position at the end of each cycle; ATP hydrolysis; DNA hand-out; DNA hand-in; and nucleotide exchange. Colors indicate the energetic content of regions associated with each process and conformation.

s_4 : 7PBO, and s_5 : 7PBP, following the nomenclature used in [3]. The RuvB subunits are labelled according to the original PDB file annotations, where chain A represents the top of the staircase interacting with the substrate, followed by chains B, C, and D as substrate-engaged, and chains E and F as substrate-disengaged. Additionally, residues not modelled in at least one conformation were excluded from all subsequent analyses.

3.4.2 Cycle reconstruction

To perform a low-dimensional embedding of the conformational ensemble, strain is used as a measure of dissimilarity between two conformations. Specifically, we computed strain for 900 pairs of structures (30×30), corresponding to the five hexamers in the mechanochemical cycle ($s_1, \dots, s_5$).

After performing all pairwise comparisons, we constructed a pairwise strain matrix M , where each element represents the mean strain $\langle \epsilon_3 \rangle$ across all N atoms included in the analysis, $\langle \epsilon_3 \rangle = \sum_i^N \epsilon_{i,3}/N$. Since M is not necessarily symmetric, we symmetrised it as $\frac{1}{2}(M + M^T)$ to obtain a suitable dissimilarity matrix for multidimensional scaling (MDS) (see Fig. 3.9 for a schematic overview). MDS then identifies a set of coordinates in a lower-dimensional space (here, two dimensions: (x, y)) for each row of the dissimilarity matrix.

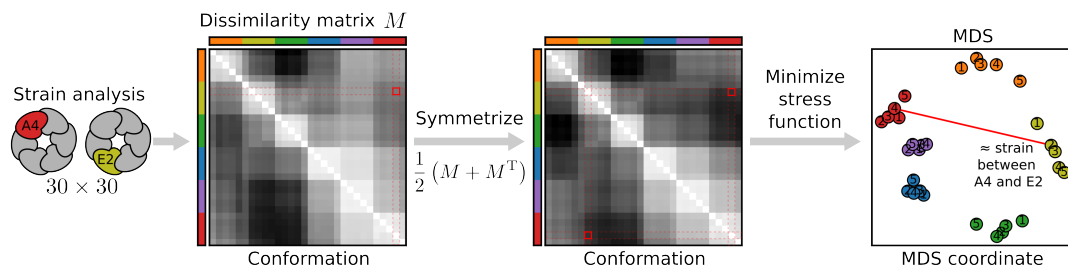


Figure 3.9: *Workflow to build a low-dimensional embedding from structural comparisons.* The approach begins by calculating the local strain for all possible pairs of conformations. In the figure, we highlight the structural comparison between conformations A4 and E2 as an example. We build a matrix M with rows and columns representing particular conformations, whose elements correspond to the mean strain over the whole structure. Darker shades of gray represent high mean strain. Because such matrix is not symmetric, we apply a symmetrisation operation. Finally, we compute the MDS for the symmetrised M as a dissimilarity matrix. This procedure is carried out by minimising a stress function, for which we considered an Euclidean distance on strain space.

3.4.3 PSA parameter choice for cycle reconstruction

We considered the following PSA parameter choice for all 900 structure pairs:

1. *Set of atoms:* heavy atoms of the main chain were selected, namely C_α , C, N, and O.
2. *Neighbourhood radius:* for a given conformation α , the set of atoms belonging to the local neighbourhood of i was defined as $\mathcal{N}_{\alpha,i} = \{j \mid \|\Delta \mathbf{x}_i^j\| < r\}$, with the cutoff radius $r = 9\text{\AA}$;
3. *Neighbourhood method:* to enhance the local difference between conformations, we used a linear weight method among all conformations $w_j = n(j)/N$, where n is a function that counts the number of occurrences of $j \in \mathcal{N}_{\alpha,i} \forall \alpha \in \{F1, F2, \dots, A5\}$, in the $N = 30$ conformations. These weights represent the relative contribution of each atom in the neighbourhood to calculate the local deformation gradient.

3.4.4 PSA parameter choice for elastic pseudoenergy calculation

We considered the following PSA parameter choice to perform structural comparisons between cycle structures $(s_1, s_2, s_3, s_4, s_5)$ and reference (s_0) as follows:

1. *Set of atoms:* heavy atoms of the main chain were selected, namely C_α , C, N, and O; the elastic pseudoenergy per residue consists of the sum of elastic

pseudoenergy of its atoms.

2. *Neighbourhood radius*: for a given assembly α , the set of atoms belonging to the local neighbourhood of i was defined as $\mathcal{N}_{\alpha,i} = \{j \mid \|\Delta \mathbf{x}_i^j\| < r\}$, with the cutoff radius $r = 9\text{\AA}$;
3. *Neighbourhood method*: to have a unified neighbourhood list per atom i , we considered the intersection of all neighbourhoods of i in each assembly, $\mathcal{N}_i = \mathcal{N}_{s_0,i} \cap \mathcal{N}_{s_1,i} \cap \dots \cap \mathcal{N}_{s_5,i}$. Thus, all neighbouring atoms in $\mathcal{N}_i$ have weight $w_j = 1$;
4. *Elastic constants*: for the material properties of proteins, we considered the Young modulus $Y = 100\text{MPa}$.
5. *Protein volume*:. To estimate the total van der Waals volume V we used the software ProteinVolume [51] for each of the 36 RuvB conformations individually.

3.4.5 Computing polar interactions

We identified the polar interactions of each atom using a script adapted from pymol [52] core functions, `show_contacts()`. We computed all inter and intramolecular H-bonds and optimal polar interactions (cutoff $3.5\text{\AA}$) and suboptimal polar interactions (cutoff $3.9\text{\AA}$). To calculate the total elastic pseudoenergy of intermolecular interactions, we sum the elastic pseudoenergy of all interacting residues weighted by the fraction of such interactions established with that particular molecule.

Bibliography

- [1] V. H. Mello, J. Wald, T. C. Marlovits, and P. Sartori, “Elastic analysis bridges structure and dynamics of an AAA+ molecular motor,” *PLOS Computational Biology*, vol. 21, no. 10, pp. 1–25, 2025. 45
- [2] C. Puchades, C. R. Sandate, and G. C. Lander, “The molecular principles governing the activity and functional diversity of AAA+ proteins,” *Nature Reviews Molecular Cell Biology*, vol. 21, no. 1, pp. 43–58, 2020. 46, 52
- [3] J. Wald, D. Fahrenkamp, N. Goessweiner-Mohr, W. Lugmayr, L. Ciccarelli, O. Vesper, and T. C. Marlovits, “Mechanism of AAA+ ATPase-mediated RuvAB–Holliday junction branch migration,” *Nature*, vol. 609, no. 7927, pp. 630–639, 2022. 46, 47, 48, 49, 58, 59
- [4] A. H. de la Peña, E. A. Goodall, S. N. Gates, G. C. Lander, and A. Martin, “Substrate-engaged 26 S proteasome structures reveal mechanisms for ATP-hydrolysis-driven translocation,” *Science*, vol. 362, no. 6418, p. eaav0725, 2018. 46
- [5] M. Shin, E. R. Watson, A. S. Song, J. T. Mindrebo, S. J. Novick, P. R. Griffin, R. L. Wiseman, and G. C. Lander, “Structures of the human LONP1 protease reveal regulatory steps involved in protease activation,” *Nature Communications*, vol. 12, no. 1, p. 3239, 2021. 46
- [6] E. Zehr, A. Szyk, G. Piszczek, E. Szczesna, X. Zuo, and A. Roll-Mecak, “Katanin spiral and ring structures shed light on power stroke for microtubule severing,” *Nature Structural Molecular Biology*, vol. 24, no. 9, pp. 717–725, 2017. 46
- [7] H. Han, H. L. Schubert, J. McCullough, N. Monroe, M. D. Purdy, M. Yeager, W. I. Sundquist, and C. P. Hill, “Structure of spastin bound to a glutamate-rich peptide implies a hand-over-hand mechanism of substrate translocation,” *Journal of Biological Chemistry*, vol. 295, no. 2, pp. 435–443, 2020. 46
- [8] M. Shin, C. Puchades, A. Asmita, N. Puri, E. Adjei, R. L. Wiseman, A. W. Karzai, and G. C. Lander, “Structural basis for distinct operational modes and protease activation in AAA+ protease Lon,” *Science Advances*, vol. 6, no. 21, p. eaba8404, 2020. 46
- [9] S. N. Gates, A. L. Yokom, J. Lin, M. E. Jackrel, A. N. Rizo, N. M. Kendersky, C. E. Buell, E. A. Sweeny, K. L. Mack, E. Chuang, M. P. Torrente,

- M. Su, J. Shorter, and D. R. Southworth, "Ratchet-like polypeptide translocation mechanism of the AAA+ disaggregase Hsp104," *Science*, vol. 357, no. 6348, pp. 273–279, 2017. 46
- [10] C. Puchades, B. Ding, A. Song, R. L. Wiseman, G. C. Lander, and S. E. Glynn, "Unique Structural Features of the Mitochondrial AAA+ Protease AFG3L2 Reveal the Molecular Basis for Activity in Health and Disease," *Molecular Cell*, vol. 75, no. 5, pp. 1073–1085.e6, 2019. 46
- [11] C. Puchades, A. J. Rampello, M. Shin, C. J. Giuliano, R. L. Wiseman, S. E. Glynn, and G. C. Lander, "Structure of the mitochondrial inner membrane AAA+ protease YME1 gives insight into substrate processing," *Science*, vol. 358, no. 6363, p. eaao0464, 2017. 46
- [12] C. Deville, M. Carroni, K. B. Franke, M. Topf, B. Bukau, A. Mogk, and H. R. Saibil, "Structural pathway of regulated substrate transfer and threading through an Hsp100 disaggregase," *Science Advances*, vol. 3, no. 8, p. e1701726, 2017. 46
- [13] K. I. White, M. Zhao, U. B. Choi, R. A. Pfuetzner, and A. T. Brunger, "Structural principles of SNARE complex recognition by the AAA+ protein NSF," *eLife*, vol. 7, p. e38888, 2018. 46
- [14] C. Alfieri, L. Chang, and D. Barford, "Mechanism for remodelling of the cell cycle checkpoint protein MAD2 by the ATPase TRIP13," *Nature*, vol. 559, no. 7713, pp. 274–278, 2018. 46
- [15] K. E. Lopez, A. N. Rizo, E. Tse, J. Lin, N. W. Scull, A. C. Thwin, A. L. Lucius, J. Shorter, and D. R. Southworth, "Conformational plasticity of the ClpAP AAA+ protease couples protein unfolding and proteolysis," *Nature Structural Molecular Biology*, vol. 27, no. 5, pp. 406–416, 2020. 46
- [16] F. Jülicher and J. Prost, "Cooperative molecular motors," *Physical review letters*, vol. 75, no. 13, p. 2618, 1995. 46
- [17] J.-C. Liao, Y.-J. Jeong, D.-E. Kim, S. S. Patel, and G. Oster, "Mechanochemistry of T7 DNA Helicase," *Journal of Molecular Biology*, vol. 350, no. 3, pp. 452–475, 2005. 46
- [18] S. Liu, G. Chistol, and C. Bustamante, "Mechanical Operation and Intersubunit Coordination of Ring-Shaped Molecular Motors: Insights from Single-Molecule Studies," *Biophysical Journal*, vol. 106, no. 9, pp. 1844–1858, 2014. 46

- [19] Y. Cao, H. Wang, Q. Ouyang, and Y. Tu, “The free-energy cost of accurate biochemical oscillations,” *Nature Physics*, vol. 11, no. 9, pp. 772–778, 2015. 46
- [20] E. Lathouwers, J. N. E. Lucero, and D. A. Sivak, “Nonequilibrium energy transduction in stochastic strongly coupled rotary motors,” *The Journal of Physical Chemistry Letters*, vol. 11, no. 13, pp. 5273–5278, 2020. 46
- [21] T. L. Hill, *Free energy transduction and biochemical cycle kinetics*. Courier Corporation, 2013. 46
- [22] D. C. Rees and J. B. Howard, “Structural bioenergetics and energy transduction mechanisms,” *Journal of Molecular Biology*, vol. 293, no. 2, pp. 343–350, 1999. 46
- [23] A. I. Brown and D. A. Sivak, “Theory of Nonequilibrium Free Energy Transduction by Molecular Machines,” *Chemical Reviews*, vol. 120, pp. 434–459, jan 2020. 46
- [24] E. D. Zhong, T. Bepler, B. Berger, and J. H. Davis, “CryoDRGN: reconstruction of heterogeneous cryo-EM structures using neural networks.,” *Nature methods*, vol. 18, pp. 176–185, feb 2021. 47, 57
- [25] A. J. M. Spencer, *Continuum mechanics*. Courier Corporation, 2004. 49
- [26] W. S. Torgerson, “Multidimensional scaling: I. Theory and method,” *Psychometrika*, vol. 17, no. 4, pp. 401–419, 1952. 49
- [27] N. Saeed, H. Nam, M. I. U. Haq, and D. B. Muhammad Saqib, “A Survey on Multidimensional Scaling,” *ACM Comput. Surv.*, vol. 51, no. 3, 2018. 49
- [28] A. F. Neuwald, L. Aravind, J. L. Spouge, and E. V. Koonin, “AAA+: A class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes,” *Genome research*, vol. 9, no. 1, pp. 27–43, 1999. 51
- [29] G. J. Steel, C. Harley, A. Boyd, and A. Morgan, “A screen for dominant negative mutants of SEC18 reveals a role for the AAA protein consensus sequence in ATP hydrolysis,” *Molecular biology of the cell*, vol. 11, no. 4, pp. 1345–1356, 2000. 52
- [30] D. A. Hattendorf and S. L. Lindquist, “Cooperative kinetics of both Hsp104 ATPase domains and interdomain communication revealed by AAA sensor-1 mutants,” *The EMBO journal*, vol. 21, no. 1-2, pp. 12–21, 2002. 52

-
- [31] T. V. Seraphim and W. A. Houry, “AAA+ proteins,” *Current Biology*, vol. 30, no. 6, pp. R251—R257, 2020. 52
- [32] M. Jessop, J. Felix, and I. Gutsche, “AAA+ ATPases: structural insertions under the magnifying glass,” *Current Opinion in Structural Biology*, vol. 66, pp. 119–128, 2021. 52
- [33] Y.-W. Han, H. Iwasaki, T. Miyata, K. Mayanagi, K. Yamada, K. Morikawa, and H. Shinagawa, “A unique β -hairpin protruding from AAA+ ATPase domain of RuvB motor protein is involved in the interaction with RuvA DNA recognition protein for branch migration of Holliday junctions,” *Journal of Biological Chemistry*, vol. 276, no. 37, pp. 35024–35028, 2001. 52
- [34] J. Howard, *Mechanics of Motor Proteins and the Cytoskeleton*. Sunderland, MA: Sinauer, 1 ed., 2001. 52
- [35] D. Pal and P. Chakrabarti, “On residues in the disallowed region of the Ramachandran map,” *Biopolymers*, vol. 63, no. 3, pp. 195–206, 2002. 54
- [36] A. V. Finkelstein and O. Ptitsyn, *Protein physics: a course of lectures*. Elsevier, 2016. 54
- [37] R. J. D. Miller, “Energetics and Dynamics of Deterministic Protein Motion,” *Accounts of Chemical Research*, vol. 27, pp. 145–150, may 1994. 57
- [38] M. R. Mitchell, T. Thlusty, and S. Leibler, “Strain analysis of protein structures and low dimensionality of mechanical allosteric couplings,” *Proceedings of the National Academy of Sciences*, vol. 113, no. 40, pp. E5847—E5855, 2016. 57, 58
- [39] P. Sartori and S. Leibler, “Evolutionary conservation of mechanical strain distributions in functional transitions of protein structures,” *Physical Review X*, vol. 14, no. 1, p. 11042, 2024. 57, 58
- [40] J. M. McBride, K. Polev, A. Abdirasulov, V. Reinharz, B. A. Grzybowski, and T. Thlusty, “AlphaFold2 Can Predict Single-Mutation Effects,” *Phys. Rev. Lett.*, vol. 131, p. 218401, nov 2023. 57
- [41] M. Correa Marrero, V. H. Mello, P. Sartori, and P. Beltrao, “Global comparative structural analysis of responses to protein phosphorylation,” *Nature Communications*, vol. 16, no. 1, p. 9407, 2025. 57

- [42] A. Bakan and I. Bahar, “The intrinsic dynamics of enzymes plays a dominant role in determining the structural changes induced upon inhibitor binding,” *Proceedings of the National Academy of Sciences*, vol. 106, no. 34, pp. 14349–14354, 2009. 57
- [43] L. Skjærven, X.-Q. Yao, G. Scarabelli, and B. J. Grant, “Integrating protein structural dynamics and evolutionary analysis with Bio3D,” *BMC Bioinformatics*, vol. 15, no. 1, p. 399, 2014. 57
- [44] X.-Q. Yao, L. Skjærven, and B. J. Grant, “Rapid Characterization of Allosteric Networks with Ensemble Normal Mode Analysis,” *The Journal of Physical Chemistry B*, vol. 120, pp. 8276–8288, aug 2016. 57
- [45] V. Lombard, S. Grudinin, and E. Laine, “Explaining Conformational Diversity in Protein Families through Molecular Motions,” *Scientific Data*, vol. 11, no. 1, p. 752, 2024. 57
- [46] J. Frank and A. Ourmazd, “Continuous changes in structure mapped by manifold embedding of single-particle data in cryo-EM,” *Methods*, vol. 100, pp. 61–67, 2016. 57
- [47] W. S. Tang, E. D. Zhong, S. M. Hanson, E. H. Thiede, and P. Cossio, “Conformational heterogeneity and probability distributions from single-particle cryo-electron microscopy,” *Current Opinion in Structural Biology*, vol. 81, p. 102626, 2023. 57
- [48] R. G. Wolcott and P. D. Boyer, “The reversal of the myosin and actomyosin ATPase reactions and the free energy of ATP binding to myosin,” *Biochemical and Biophysical Research Communications*, vol. 57, no. 3, pp. 709–716, 1974. 58
- [49] R. Yasuda, H. Noji, M. Yoshida, K. Kinosita, and H. Itoh, “Resolution of distinct rotational substeps by submillisecond kinetic analysis of F1-ATPase,” *Nature*, vol. 410, no. 6831, pp. 898–904, 2001. 58
- [50] M. Dittrich, S. Hayashi, and K. Schulten, “On the Mechanism of ATP Hydrolysis in F1-ATPase,” *Biophysical Journal*, vol. 85, no. 4, pp. 2253–2266, 2003. 58
- [51] C. R. Chen and G. I. Makhatadze, “ProteinVolume: calculating molecular van der Waals and void volumes in proteins,” *BMC Bioinformatics*, vol. 16, no. 1, p. 101, 2015. 61
- [52] Schrödinger, LLC, “The PyMOL Molecular Graphics System, Version~1.8.” 2015. 61

Kinetic model bridges structures and dynamics for RuvB assemblies

RuvB is a homohexameric ATPase that drives branch migration of Holliday Junctions through a mechanochemical cycle powered by ATP hydrolysis. In the previous chapter, we analysed in detail how structural deformations couple these two processes energetically at the residue scale. Here, we introduce a simplified model that coarse-grains the rich structural information into conformational states, inspired by biochemical cycle kinetics. This toy model of RuvB’s cycle is built on three energetic components: a structural potential energy derived from conformational ensembles, an inter-subunit coupling term, and a nonequilibrium driving force. This minimal framework, directly informed by structural data, captures essential features of RuvB’s activity, including processivity, coordinated subunit dynamics, and energy transduction. Simulations further suggest that RuvB’s conformational landscape is optimised for faster translocation, while the magnitude of the nonequilibrium drive minimises energy dissipation. Together, these results provide a physically grounded description of RuvB’s mechanochemistry and demonstrate how structural data can be used to constrain the thermodynamics of conformational cycles.

This chapter partially includes an adaptation of the publication “*Elastic analysis bridges structure and dynamics of an AAA+ molecular motor*” [1], Mello VH, Wald J, Marlovits TC, Sartori P. The chapter is further complemented with additional information not included in the manuscript. Author contributions: Conceptualisation, VHM and PS; Formal analysis, VHM; Writing – Original draft, VHM; Writing – Review and editing, VHM and PS; Visualisation, VHM; Supervision, PS.

4.1 Introduction

Molecular motors perform mechanical tasks such as transport, cell motility, and remodelling of other molecules by converting free-energy of ATP into directed motion. To study their function, two intrinsic aspects of their activity must be taken into account: first, they act far from thermodynamic equilibrium; and second, they are largely subjected to thermal fluctuations [2]. These factors pose several constraints to the approaches used to understand how molecular motors work. Developments in stochastic thermodynamics have brought a significant advance in this direction by providing tools to quantify kinetics [3], efficiency [4, 5], and information transfer [6], for example. Importantly, this theoretical framework evolved hand-in-hand with experimental techniques such as single-molecule manipulations, leading to a significant transfer of knowledge between these fields [7, 8]. However, a limiting aspect is that it still remains largely disconnected from structural aspects of molecular motors, often studied in structural biology.

A notable example of such effort is the work of Okazaki and Hummer [9], where a stochastic model coupled to molecular dynamics simulations was used to study the cycle of ATP synthase. In this framework, dynamical information was harnessed from molecular dynamics simulations and then used to parameterise a Langevin model for the rotor component of ATP synthase. While this work showcases how molecular simulations can be coupled to stochastic models, similar integrative efforts towards experimental structural data are much more limited, especially at nonequilibrium. This is particularly relevant in the present, given the rise of time-resolved experimental techniques [10].

Here, we devise a stochastic kinetic model tailored to the RuvB mechanochemical cycle, using structure-derived energetic profiles as input. Rather than imposing an arbitrary potential energy landscape – a common feature among theoretical studies on molecular motors –, we use the potential energies obtained directly from conformational ensembles. This allows us to constrain dynamics based on experimentally resolved protein structures. Moreover, we show that both the shape and magnitude of the structurally derived energy landscape are consistent with a processive motor, probed with single-molecule and biochemical experiments [11]. Finally, we explore three consequences of the model: *(i)* we show that the shape of the energy landscape leads to enhanced translocation speed; *(ii)* that the inter-subunit coupling landscape emerges naturally from the model’s dynamics; and *(iii)* that trade-offs between dissipation, speed, and efficiency arise as intrinsic features of the system. Therefore, this work shows how structure-based energetic information from a conformational ensemble

ble can be harnessed to build physically grounded models of nonequilibrium molecular machines.

4.2 A kinetic model tailored for RuvB mechanochemical cycle

Consider a system with M molecules, labelled $m \in \{1, 2, \dots, M\}$, representing each of the RuvB subunits. These molecules populate a discrete and ordered set of states, representing the protein conformations, $i \in \{1, 2, \dots, N\}$. To match these states with RuvB's mechanochemical cycle, we set $M = 6$ and $N = 30$. Transitions in this system are caused by random interactions with a thermal reservoir of temperature T , inducing a molecule m in state i to change to one of the neighbour states ($i + 1$ or $i - 1$). We set periodic boundary conditions, allowing transitions between states 1 and N . We assume that a transition changes the state of a single molecule at a time.

Each of the N states is associated with a reaction coordinate $\theta(i) = 2\pi i/N$. To simplify our notation, we denote as θ_m the reaction coordinate of the subunit m . In our analyses, we assume that these reaction coordinates correspond to the conformations observed in RuvB's mechanochemical cycle, $\theta(1) \rightarrow \text{F1}$, $\theta(2) \rightarrow \text{F2}$, $\dots$, $\theta(30) \rightarrow \text{A5}$. Thus, a hexameric assembly state can be fully described by a vector with the reaction coordinates of each subunit, for instance, $\mathbf{a} = [\theta_1(4), \theta_2(9), \theta_3(14), \theta_4(18), \theta_5(24), \theta_6(29)]$ corresponds to the assembly shown in Fig. 4.1.

The internal energy of the whole assembly can be decomposed in two terms. First, each state is characterised by a potential elastic energy $\Delta E_{\text{el}}(\theta)$. Second, a coupling energy E_c links pairs of molecules with consecutive indices, see Fig. 4.1A. We define the coupling energy inspired by the Kuramoto model [12], a classical mathematical model for describing coupled phase oscillators:

$$E_c(\theta_m, \theta_{m+1}) = \frac{K}{M} (1 - \cos(\theta_{m+1} - \theta_m - \delta)) \quad , \quad (4.1)$$

where K and δ respectively represent the coupling strength and the phase-lag in reaction coordinate between consecutive molecules at equilibrium. Because we are interested in modelling RuvB, which operates as a sequential motor, we fix $\delta = \pi/3$. This ensures that the five possible arrangements in which subunits are evenly spaced in conformation – matching the states $s_1, s_2, \dots, s_5$ introduced in the previous chapter

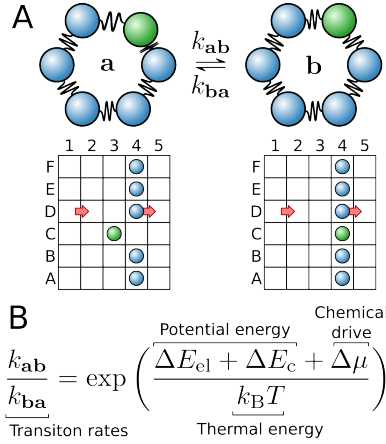


Figure 4.1: *Kinetic model formulation*. **A**. Representation of a particular transition between assembly states. A RuvB assembly consists of six coupled subunits, represented by circles connected through springs. In a given hexameric state, each subunit occupies one of the 30 conformations of the mechanochemical cycle, see tables below schematics for the example states **a** and **b**. Transitions among adjacent states are controlled by the kinetic rates k . In the example, the green circle transitions between the adjacent conformations C3 and C4, where the other subunits remain in the same configuration. Red arrows highlight the transitions driven by the chemical potential $\Delta\mu$, which are associated with ADP release and ATP binding **B**. The detailed-balance condition enforces the ratio of back-and-forth transition rates to respect the energy difference of the assembly configurations. The driven transitions, however, are biased towards one direction.

– are minima of E_c . The energy of the assembly can be thus summarised as

$$E(\boldsymbol{\theta}) = \sum_{m=1}^M [\Delta E_{el}(\theta_m) + E_c(\theta_m, \theta_{m+1})] \quad . \quad (4.2)$$

The energetics and dynamics of the system are connected by local detailed-balance (Fig. 4.1B), which links transition rates between pairs of states and their respective energy difference [13]. As we mentioned previously, transitions are allowed between assembly states for which one of the molecules m changes to a neighbouring state. From the allowed transitions, a few have an additional energetic contribution $\Delta\mu$, representing the external chemical drive coming from the ATP cycle. We set the

kinetic rates between two arbitrary neighbouring states $\mathbf{a} = \boldsymbol{\theta}$ and $\mathbf{b} = \boldsymbol{\theta}'$ as such:

$$k_{\mathbf{ab}} = \begin{cases} \tau_{\mathbf{ab}}^{-1} \Gamma_{\mathbf{ab}} & \text{if } \mathbf{a} \leftrightarrow \mathbf{b} \text{ is not driven} \\ \tau_{\mathbf{ab}}^{-1} \Gamma_{\mathbf{ab}} \exp\left(\frac{\Delta\mu}{2k_{\text{B}}T}\right) & \text{if } \mathbf{a} \rightarrow \mathbf{b} \text{ is driven} \\ \tau_{\mathbf{ab}}^{-1} \Gamma_{\mathbf{ab}} \exp\left(\frac{-\Delta\mu}{2k_{\text{B}}T}\right) & \text{if } \mathbf{a} \leftarrow \mathbf{b} \text{ is driven,} \end{cases} \quad (4.3)$$

with $\Gamma_{\mathbf{ab}} = (1 + \exp((E(\mathbf{b}) - E(\mathbf{a}))/k_{\text{B}}T))^{-1}$, k_{B} the Boltzmann constant, and $\tau_{\mathbf{ab}}$ the characteristic time scale of the transition (note that $\tau_{\mathbf{ab}} = \tau_{\mathbf{ba}}$).

Next, we introduce the emergent translocation speed v of the system based on trajectories simulated using the Gillespie algorithm [14] according to the model described above. To do so, we first consider that a translocation step occurs upon the completion of a conformational cycle by any of the six subunits. This corresponds to the transitions from $i = 30$ with a jump $i \rightarrow i + 1$ for a positive displacement and from $i = 1$ with a jump $i \rightarrow i - 1$ for a negative displacement. Assuming that each cycle completion results in a translocation step of two DNA base-pairs, of length $d \approx 0.68\text{nm}$ [15] and that there is no “slippage” between conformational cycles and translocation (futile cycles), we can compute v as displacement over time for a given trajectory, see Fig. 4.2 for examples.

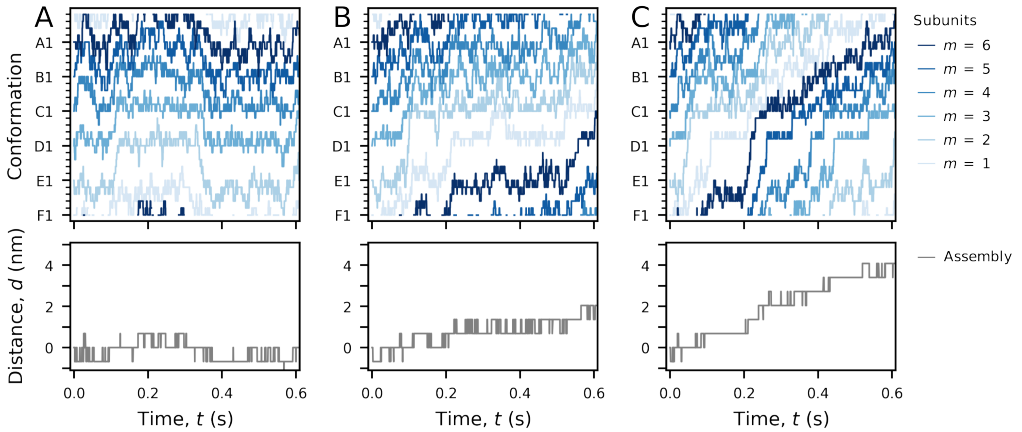


Figure 4.2: *Examples of trajectories constrained by energetics from structural data.* Trajectories of RuvB conformations simulated with the model (top, subunits in shades of blue) and distance (total DNA length) translocated by the assembly (bottom, in gray). The panels show trajectories of coupled subunits ($K = 100k_{\text{B}}T$) subjected to different chemical potentials: **A.** $\Delta\mu = 0$. **B.** $\Delta\mu = 5k_{\text{B}}T$. **C.** $\Delta\mu = 20k_{\text{B}}T$.

Altogether, we develop a kinetic model that connects the microscopic conformational cycle of RuvB to its mesoscopic biophysical behaviour. The model integrates three energy contributions governing conformational transitions: an energy potential,

defined by a state-dependent elastic pseudoenergy landscape derived from protein mechanics; intersubunit coupling, modelled as a Kuramoto-inspired synchronisation model between subunits; and a non-equilibrium driving force, representing the chemical potential of the ATP hydrolysis cycle. Next, we study how the interplay between these three energetic parameters shapes the emergent translocated speed.

4.3 Energy-related parameters draw conditions for processivity

We start by highlighting the key parameters that control the amplitude of each energetic term. To explore the role of potential energy amplitude, we define the potential V as

$$V = V_\alpha \frac{\Delta E_{\text{el}}}{\max(\Delta E_{\text{el}})} \quad ,$$

where V_α scales the amplitude of the potential energy landscape ΔE_{el} while maintaining its shape. The other parameters are K , representing the inter-subunit coupling strength, and $\Delta\mu$, denoting the magnitude of the non-equilibrium drive. Figure 4.3 illustrates how these energetic parameters affect translocation speed.

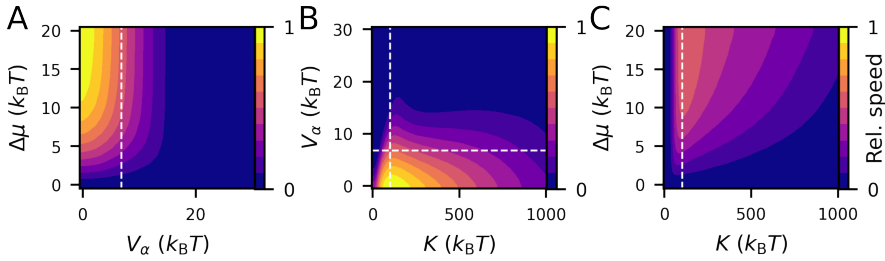


Figure 4.3: *Dependence of relative processive speed on energy-related model parameters.* All figures display the processive speed in colors normalised to the maximum speed achieved. The magnitude of the speed is displayed on the colorbar at the right of each heatmap. **A.** Relative speed in function of potential energy amplitude and driving magnitude, at fixed coupling strength ($K = 100k_B T$). **B.** Relative speed in function of coupling strength and the potential energy amplitude, at fixed driving energy ($\Delta\mu = 20k_B T$). **C.** Relative speed in function of coupling strength and driving magnitude at fixed potential energy amplitude ($V_\alpha = \max(\Delta E_{\text{el}})$). White dashed lines represent values used in the coming sections. All results show the mean speed of ten simulations, which were then smoothed with a Gaussian filter.

As shown in Fig. 4.3A, the emergent speed of the system increases monotonically with $\Delta\mu$ and decreases with V_α . The driving energy biases the system towards an unidirectional progression, thereby increasing the motor's speed. For $K = 100k_B T$,

the speed saturates for $\Delta\mu > 15k_B T$, as indicated by the vertical parallel level curves. Conversely, lower potential amplitudes reduce the energy barriers that individual particles must overcome to complete a full cycle, resulting in higher translocation speeds.

Figures 4.3B and C demonstrate that the translocation speed varies non-monotonically with the coupling strength, exhibiting a maximum around $100k_B T$. At low coupling strengths, the trajectories of neighbouring subunits become uncoupled, preventing energy transduction between them. In the absence of coupling, particles rely solely on thermal fluctuations to cross the potential energy barrier. Conversely, at very high coupling strengths, translocation speed also declines. This decrease likely arises from a combination of finite-size effect (limited N) and the constraints imposed by Master-equation dynamics. As the coupling strength K increases, the system must overcome a larger collective energy barrier on E_c to perform a coordinated transition from state i to $i + 1$ across all subunits. Such an increased barrier inflates the time scale of successful forward transitions, thereby impairing overall processive speed.

In the upcoming sections, we fix these three parameters to simplify our analyses, unless stated otherwise. For the potential, we consider $V_\alpha = \max(\Delta E_{el})$, to constrain the system to the elastic pseudoenergy obtained from the structural analyses. We set the chemical potential $\Delta\mu = 20k_B T$, which nearly matches the experimentally free-energy of ATP hydrolysis in *E. coli* ($\approx 19k_B T$). While the coupling strength K , a model parameter that cannot be estimated from structural analyses, was fixed at $100k_B T$ to match the value that maximises speed for the previous two energy parameters.

4.4 Calibration of time scales with experimental data

The kinetic model we introduced has the principle of detailed-balance as a key thermodynamic constraint, see Fig. 4.1 for its formulation. However, the actual form of the kinetic rates cannot be fully described by the system's energetics, as detailed-balance only dictates the behaviour of back-and-forth kinetic rates. We thus propose that a time scale variable controls the temporal dependency on kinetic rates, Eq. 4.3.

To calibrate such time scale, we use previously published data from two independent datasets that relate ATP concentration to translocation speeds [11]. In brief, Han and colleagues measured the [ATP]-speed curves with a bulk biochemical approach, hereby denoted as the biochemical data, and with a single-molecule experiment, resulting in what we denote biophysical data. We digitised both datasets using the web application apps.automeris.io/wpd/, reproduced in Fig. 4.4A.

We define the time scales τ_{ab} in two forms to account for the specificities of each experimental approach. To model the biochemical dataset, we consider that all transitions have the same time scale, $\tau_{ab} = \tau_F$. For the biophysical dataset, we model τ_{ab} as a step function to account for the slowed transitions due to the friction of the bead present in the experimental assay:

$$\tau_{ab} = \begin{cases} \tau_F & \text{if } 0 < \theta_m, \theta'_m < \frac{2\pi}{3} \\ \tau_B & \text{if } \frac{2\pi}{3} < \theta_m, \theta'_m \leq 2\pi \\ \frac{1}{2}(\tau_F + \tau_B) & \text{otherwise,} \end{cases}$$

with m representing the molecule which underwent a state transition in $\mathbf{a} \rightarrow \mathbf{b}$. θ_m and θ'_m are the coordinates of the molecule m before and after the transition, respectively. This particular choice of intervals in the reaction coordinate (0 and $\frac{2\pi}{3}$) for τ_F and τ_B corresponds to the RuvB conformations between free and bound, detailed in the previous chapter. Therefore, this choice assumes that the viscous drag present in the experiment affects the transitions among DNA bound conformations. The transitions that display $\tau_{ab} = \frac{1}{2}(\tau_F + \tau_B)$ are required to respect the symmetry condition $\tau_{ab} = \tau_{ba}$ for arbitrary neighbouring assembly states $\mathbf{a}$ and $\mathbf{b}$.

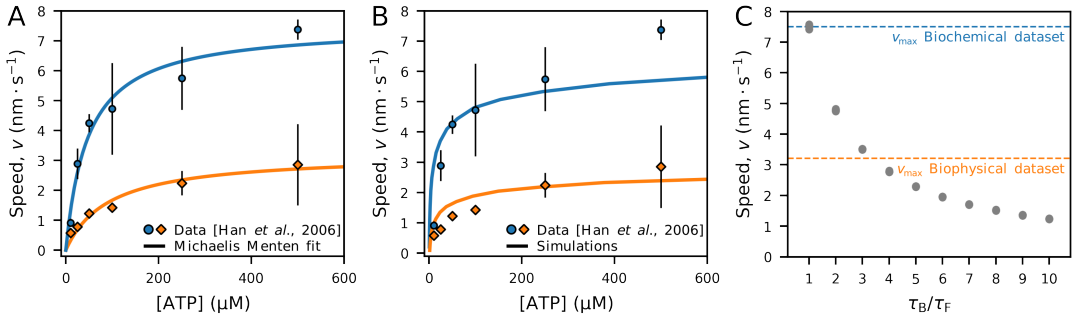


Figure 4.4: *Fitting of characteristic time scale of conformational changes, τ .* **A.** Reproduction of data obtained experimentally in [11], where branch migration speed v is measured in two experimental approaches, namely *biochemical* and *biophysical*. **B.** Simulations adjusted to the experimental data. We sampled conformational trajectories with varying $\Delta\mu$ ranging from 0.2 to $9k_B T$, each simulated with $n = 10$ different random seeds. **C.** Speed calculation for the biophysical dataset, with fixed τ_F and varying τ_B ($n = 10$ simulation replicates).

We first calibrate τ_F , because it can be obtained directly from the biochemical dataset and our simulations. To do so, we extract the speed at saturated ATP concentration $v_{\max}$ fitting the biochemical dataset to the Michaelis-Menten curve $v = v_{\max}(1 + \frac{K_M}{[ATP]})^{-1}$, with v the translocation speed and K_M the Michaelis con-

stant. We then adjusted τ_F to match the emergent translocation speed, simulating trajectories at saturated driving force ($\Delta\mu = 500k_B T$, $n = 10$ simulation replicates). Next, we simulate multiple trajectories also at saturated chemical drive, with fixed τ_F and varying τ_B , in order to find τ_B that best reproduces $v_{\max}$ from the biophysical dataset, see Fig. 4.4C. With this approach, we obtain the time scales estimates $\tau_F = 1.13\text{ms}$ and $\tau_B \approx 3\tau_F = 3.39\text{ms}$.

Finally, to adjust the speed-[ATP] curves, we derive from the definition of chemical potential $\Delta\mu = \log([ATP]/C)$, where C depends on ATP, ADP, and P_i chemical potentials and concentrations. Then, we fit curves of processive speed simulated from a range of chemical potentials ($\Delta\mu \in [0.2, 9]k_B T$) to the experimental data, optimising C and K_M , see Fig. 4.4B. Our fittings reveal a poorer match to the experimental data when compared to a Michaelis-Menten fitting, as the shape of the [ATP]-speed curve for the simulations presents a sharper transition to a saturated regime. Still, the fitted time scales for transitions between states lie in the order of milliseconds, which matches the time scales of processes such as catalysis [10]. In the next sections, we discuss three properties that emerge from the model constrained by the elastic pseudoenergy we obtained from our previous analyses.

4.5 Shape of the elastic pseudoenergy potential encodes an enhanced speed

We next examine the impact of altering the order of the energy landscape in the conformational cycle. Two ensembles of randomised conformational landscapes were generated: one in which the elastic pseudoenergies of the thirty conformations were fully shuffled, and another in which only the order of the assemblies (s_1 to s_5) was shuffled while maintaining the subunit order (F to A), see Fig. 4.5A for examples.

The resulting speed distributions reveal that the native order of conformations yields significantly higher processive speeds compared to randomised cycles in average, Fig. 4.5B. This outcome is intuitive, as the native elastic pseudoenergy landscape is relatively smooth, facilitating efficient progression. In contrast, shuffled potentials are generally more rugged, increasing the likelihood of RuvB subunits to dwell in local energy minima.

Overall, these results suggest that our model, constrained by structures, results in robust assembly dynamics. Furthermore, one may speculate that the conformations visited by RuvB subunits on the translocation process have undergone evolutionary tuning to optimise the shape of such energetic profile. Notably, the presence of a single significantly large barrier rather than multiple, is an example of that.

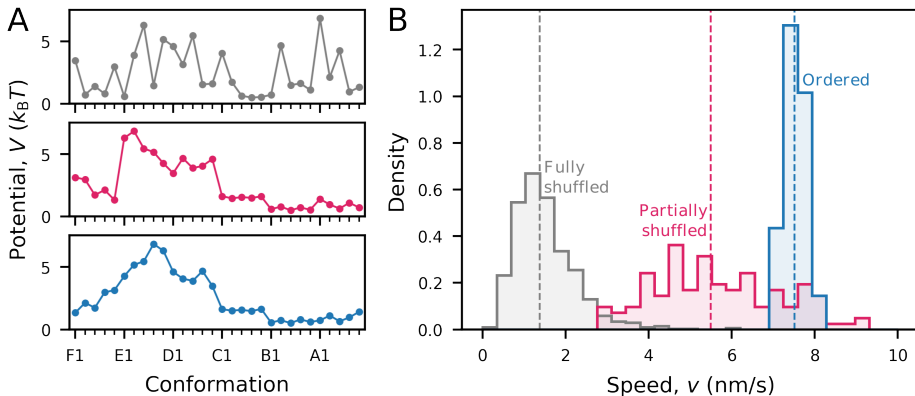


Figure 4.5: *Native order of energy landscape promotes faster translocation kinetics than random.* **A.** Examples of fully shuffled (gray) and partially shuffled energy landscapes; below, the unique native order of the energy landscape is displayed (blue). **B.** Distribution of speeds simulated from the model with shuffled and ordered elastic pseudoenergy landscapes. The sample sizes are $n = 1000$ (fully shuffled); $n = 120$ (partially shuffled); and $n = 20$ random seeds (ordered).

4.6 Coupling profiles are consistent with structural observations

In our structural analyses presented in the previous chapter, we could focus on specific protein regions to infer local energy variations. One of such regions was at the inter-subunit interface (see Fig. 3.6B and C), which we used as a rough indicator of energy transduction across the RuvB hexamer. A similar approach cannot be used using the current model. Here, the spatial dimension is largely neglected, as we coarse-grain all the elastic energy of each conformation into a scalar. To draw a comparison with these specific structural insights, we turn our attention to the coupling energy E_c instead.

Our method inherently considers that adjacent subunits are allowed to transfer energy, but it does not set *a priori* the magnitude of this coupling at each conformation. To quantify the magnitude of the inter-subunit coupling throughout the cycle, we compute the coupling energy between adjacent subunits, see Appendix A for details. Figure 4.6A shows that forward and backward coupling energies are localised to conformations before and after the largest energy barrier. We emphasise that this pattern is not imposed, but rather emerges from the self-organisation of the system, shaped by the interplay between potential energy landscapes and driving forces applied to each subunit.

We then calculate the mean coupling force across subunits in each cycle confor-

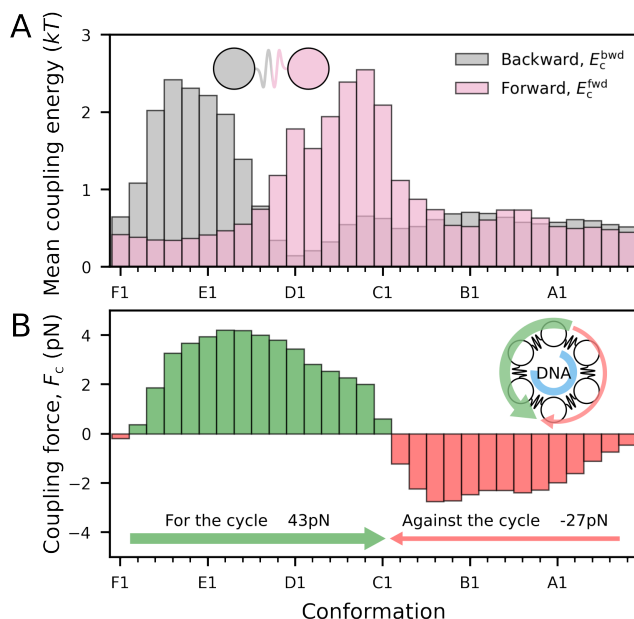


Figure 4.6: *Coupling energy prediction of kinetic model.* **A.** Mean coupling energy at backward and forward interfaces (see schematic). The coupling energy is high before and after the energy barrier. **B.** Total coupling forces, summed effect of the spatial derivative of backward and forward coupling energies. Bars in green represent forces aligned to the cycle direction (F → E) and in red, in the opposite direction (E → F). The model predicts a net positive force over the whole cycle. The cartoon represents the RuvB hexamer, with green and red curved arrows for the positive and negative forces, respectively.

mation. The force, which corresponds to the spatial derivative of the coupling energy, determines the direction and intensity with which subunits push or pull each other. Figure 4.6B shows that the mean resulting force is positive (aligned to the cycle) in positions F, E, and D, and negative (opposed to the cycle) otherwise. The assisting role of positive forces in crossing the energy barrier is reminiscent of the previously proposed role of the converter module [15, 16]. Altogether, these results support the notion that inter-subunit coupling plays a key role in efficiently overcoming the energy barrier, consistent with conclusions drawn from structural analyses.

4.7 Trade-offs among dissipation, efficiency, and speed

With elementary tools from stochastic thermodynamics, we now explore the relationship between kinetics extracted from conformational trajectories and free-energy dissipation. In this model, the system operates at nonequilibrium due to a chemical

driving force associated with ATP hydrolysis, quantified by the chemical potential difference $\Delta\mu$. Note that the DNA interactions are not explicitly modelled; instead, their effect is implicitly captured in the equilibrium components of the energy landscape.

To quantify dissipation, we compute the mesoscopic heat of a trajectory (here denoted as $\{\boldsymbol{\theta}\}$), which reflects energy changes linked to transitions between assembly states. This quantity is naturally defined within stochastic thermodynamics [17] as

$$q(\{\boldsymbol{\theta}\}) = \sum_{u=1}^U [E(\boldsymbol{\theta}(t_{u-1})) - E(\boldsymbol{\theta}(t_u)) + \Delta\mu_{t_{u-1},t_u}] = E(\boldsymbol{\theta}(t_0)) - E(\boldsymbol{\theta}(t_U)) + \sum_{u=1}^U \Delta\mu_{t_{u-1},t_u}, \quad (4.4)$$

where

$$\Delta\mu_{t_{u-1},t_u} = \begin{cases} 0 & \text{if } \boldsymbol{\theta}(t_u) \leftrightarrow \boldsymbol{\theta}(t_{u-1}) \text{ is not driven} \\ \Delta\mu/2 & \text{if } \boldsymbol{\theta}(t_u) \leftrightarrow \boldsymbol{\theta}(t_{u-1}) \text{ is driven,} \end{cases} \quad (4.5)$$

and U counts the number of time steps in the trajectory. Importantly, each driven step consists of half of the chemical potential because there are two driven transitions per cycle (check Fig. 4.1).

To facilitate the comparison between trajectories, we introduce the quantity dissipation per cycle Q , defined as

$$Q(\{\boldsymbol{\theta}\}) = \frac{q(\{\boldsymbol{\theta}\})d}{vt_U}, \quad (4.6)$$

where v is the emergent translocation speed, t_U is the trajectory duration, and d is the length of one translocation step.

Then, we computed the dissipation per cycle from simulated trajectories across a range of chemical potentials. Fig. 4.7A shows that dissipation per cycle exhibits a non-monotonic dependence on $\Delta\mu$. At low driving, the system requires large amounts of free-energy to perform a cycle. Dissipation per cycle reaches a minimum at $\Delta\mu \approx 18k_B T$ (highlighted with a star in the figure), remarkably close to the experimentally observed free-energy difference in *E. coli* under exponential growth, $\Delta\mu \approx 19k_B T$ [18]. At large driving, Q converges asymptotically to $\Delta\mu$, which implies that a single ATP molecule per subunit is sufficient to drive the conformational cycle, in such cases.

Motivated by this lower bound on dissipation, we define an efficiency metric as the ratio $\Delta\mu/Q$, see Fig. 4.7B. Curiously, the chemical driving that minimises dissipation displays an intermediate value of efficiency (≈ 0.79). This suggests that achieving a single ATP molecule cycled per subunit is not necessarily advantageous, as it can lead to higher overall energy dissipation per cycle.

To test whether a similar trade-off exists with respect to kinetics, we examined

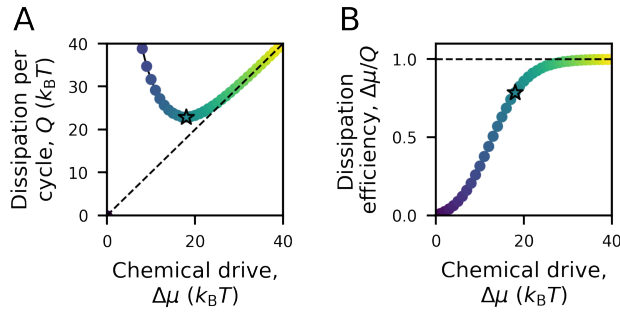


Figure 4.7: *Energy dissipation per cycle is minimal at physiological conditions.* **A.** Dissipation per cycle in function of chemical drive, tested at every $k_B T$ unit, shows a minimum at $18k_B T$, which we highlight with a star. **B.** Dissipation efficiency in function of chemical drive. The results shown are the means of $n = 5$ simulations with different random seeds.

the relationship between driving force and translocation speed. Here, we observe that the speed-drive curve saturates near the minimal values of Q , Fig. 4.8A. In Fig. 4.8B, it becomes clear that, for values of $\Delta\mu > 18k_B T$, there is an increase in dissipation that does not lead to improvement of translocation speed.

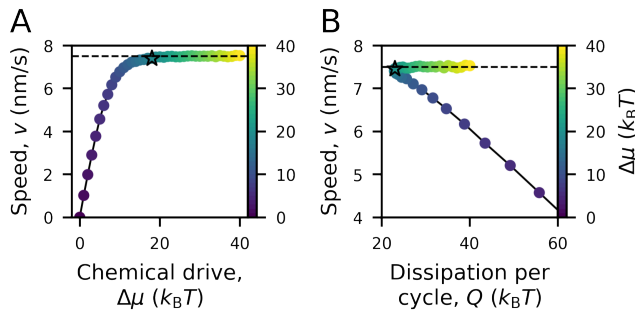


Figure 4.8: *Translocation speed is maximal at physiological conditions.* **A.** Translocation speed in function of chemical drive plateaus at values close to those that minimise dissipation per speed, highlighted with a star. **B.** Scatter plot of speed *vs.* dissipation per cycle. Color gradient represents data points simulated with different values of chemical potential. The dashed lines correspond to $v_{\max}$. All the results shown represent the mean of $n = 5$ simulations.

Summarising, these findings reveal that there is an optimal range of chemical drive that can simultaneously maximise translocation speeds and minimise energy dissipation per cycle. This value is in close agreement with experimental data, strongly suggesting that RuvB has evolved its energy landscape to achieve such a precise match.

4.8 Discussion

In this chapter, we build and implement a toy model of RuvB’s mechanochemical cycle based on three energetic components: a potential energy derived from our previous structural analysis; a coupling energy that links adjacent ATPase subunits; and a nonequilibrium drive from the cycle of ATP hydrolysis. Importantly, although grounded in fundamental physical principles, these three energetic components were specifically designed based on insights from our structural analyses and the structural data itself. Remarkably, this minimal model captures essential features of molecular motor function, including processivity and energy transduction. Moreover, it can be used as an integrative tool, bridging data from different experimental setups, such as single particle cryo-EM reconstructions [15], batch biochemical assays, and single-molecule pulling experiments [11]. Next, we comment on specific findings that we learn as model outcomes.

We find that the simple model of inter-subunit coupling we proposed gives rise to informative dynamics. Though E_c is agnostic to the absolute coordinate in the conformational landscape, its magnitude varies with the spacing between neighbouring subunits, which is further affected by the potential and driving force. As a consequence, we find that coupling energy peaks around the major energy barrier in the cycle. This indicates that the “conformational springs” are either compressed or stretched near these positions, suggesting that, despite energy transduction being always allowed, it concentrates in particular parts of the mechanochemical cycle. This finding corroborates the observations in the previous chapter, where we identified that the interface between subunits is deformed only in specific conformations.

Our simulations also suggest that RuvB’s conformational landscape exhibits signs of evolutionary optimisation. One line of evidence comes from the elastic energy profile, which enables processive speeds more than four times faster than bootstrapped orders. Although the landscape we identified is not globally optimal, it is likely shaped by structural and biochemical constraints. This led us to hypothesise that evolutionary pressures have optimised the energy landscape for maximal speed, while remaining constrained by physical and biochemical feasibility. Further studies of the energy landscapes in other AAA+ motors might reveal a broader picture regarding their potential optimality.

Finally, our results suggest that RuvB has also evolved to reach maximum speed, while dissipating as little energy as possible. To explore this idea, let us hypothesise that a species has a fixed free-energy of ATP hydrolysis $\Delta\mu_h$, by having ATP, ADP and P_i concentrations chemostated. Also, consider that a particular process mediated

by an ATPase requires an energy of E_p to operate sufficiently close to its maximum speed (recall that the speed of chemical reactions under Michaelis-Menten dynamics saturates). Then, processes for which $E_p < \Delta\mu_h$ will dissipate an extra amount of $\Delta\mu_h - E_p$ per cycle. Conversely, if $E_p > \Delta\mu_h$, multiple hydrolysis events might be needed to complete a single cycle, also increasing dissipation. Therefore, minimal dissipation is achieved when $E_p = \Delta\mu_h$. Our results suggest that this is the case of RuvB: its minimal dissipation value ($\approx 18k_B T$) remarkably matches $\Delta\mu_h \approx 19k_B T$ [18], while displaying maximum speed. This correspondence between minimal dissipation and maximal speed suggests that RuvB works at a regime optimal for both thermodynamics and kinetics.

4.9 Appendix A: Definition of coupling energy and coupling force per state

To calculate the coupling energy as a function of conformation, we first need to introduce the probability density of the assembly states $P(\boldsymbol{\theta})$, with $\boldsymbol{\theta} = [\theta_1, \dots, \theta_6]$. Because we are interested in describing the energetics of operational cycles, we consider here the probability density given that the initial condition $\boldsymbol{\theta}^\circ$, which we define as $\boldsymbol{\theta}^\circ = [\theta_1(1), \theta_2(6), \theta_3(11), \theta_4(16), \theta_5(21), \theta_6(26)]$. We use this configuration because it represents an assembly state with conformations that correspond to s_1 . Then, we define $P^\circ(\boldsymbol{\theta}) = P(\boldsymbol{\theta} | \boldsymbol{\theta}(t=0) = \boldsymbol{\theta}^\circ)$. Other particular choices of initial state can lead to local minima of the energy landscape which are not processive. The rationale for this choice is that RuvB assembly starts the mechanochemical cycle in s_1 .

With this, we introduce the procedure to calculate F , the mean value of an arbitrary function f evaluated in the reaction coordinates ϕ

$$F(\phi) = \langle f(\phi) \rangle_{\boldsymbol{\theta}} = \sum_{\boldsymbol{\theta}} P^\circ(\boldsymbol{\theta}) \left[\frac{1}{M} \sum_{m=1}^M f(\theta_m) \delta(\theta_m - \phi) \right],$$

where δ is the Kronecker delta.

Using this definition, we then calculate the mean backward coupling E_c^{bwd} as the mean elastic pseudoenergy $E_c(\theta_m, \theta_{m+1})$ with the first argument fixed, and the mean forward coupling E_c^{fwd} with the second argument fixed. Thus, we obtain

$$E_c^{\text{bwd}}(\phi) = \sum_{\boldsymbol{\theta}} P^\circ(\boldsymbol{\theta}) \left[\frac{1}{M} \sum_{m=1}^M E_c(\theta_m, \theta_{m+1}) \delta(\theta_m - \phi) \right],$$

$$E_c^{\text{fwd}}(\phi) = \sum_{\boldsymbol{\theta}} P^\circ(\boldsymbol{\theta}) \left[\frac{1}{M} \sum_{m=1}^M E_c(\theta_{m-1}, \theta_m) \delta(\theta_m - \phi) \right].$$

The coupling forces are defined as the spatial derivative of the coupling energy. To compute the spatial derivative, we assume a linear relationship between conformational and physical spaces, $x_m = R\theta_m$, with the conformations arranged along a circumference of radius $R = 6\text{nm}$. Therefore, the coupling force can be computed as

$$F_c = \frac{\partial}{\partial x} (E_c^{\text{bwd}} - E_c^{\text{fwd}}) = \frac{\partial \theta}{\partial x} \frac{\partial}{\partial \theta} (E_c^{\text{bwd}} - E_c^{\text{fwd}}) = R \frac{\partial}{\partial \theta} (E_c^{\text{bwd}} - E_c^{\text{fwd}})$$

Bibliography

- [1] V. H. Mello, J. Wald, T. C. Marlovits, and P. Sartori, “Elastic analysis bridges structure and dynamics of an AAA+ molecular motor,” *PLOS Computational Biology*, vol. 21, no. 10, pp. 1–25, 2025. 67
- [2] P. M. Hoffmann, “How molecular motors extract order from chaos (a key issues review),” *Reports on Progress in Physics*, vol. 79, no. 3, p. 32601, 2016. 68
- [3] S. Leibler and D. A. Huse, “Porters versus rowers: a unified stochastic model of motor proteins.,” *The Journal of cell biology*, vol. 121, pp. 1357–1368, jun 1993. 68
- [4] F. Jülicher, A. Ajdari, and J. Prost, “Modeling molecular motors,” *Reviews of Modern Physics*, vol. 69, no. 4, p. 1269, 1997. 68
- [5] H. Wang and G. Oster, “The Stokes efficiency for molecular motors and its applications,” *Europhysics Letters*, vol. 57, no. 1, p. 134, 2002. 68
- [6] M. P. Leighton and D. A. Sivak, “Flow of Energy and Information in Molecular Machines,” *Annual Review of Physical Chemistry*, vol. 76, no. Volume 76, 2025, pp. 379–403, 2025. 68
- [7] G. Hummer and A. Szabo, “Free energy reconstruction from nonequilibrium single-molecule pulling experiments,” *Proceedings of the National Academy of Sciences*, vol. 98, no. 7, pp. 3658–3661, 2001. 68
- [8] G. Hummer and A. Szabo, “Kinetics from Nonequilibrium Single-Molecule Pulling Experiments,” *Biophysical Journal*, vol. 85, pp. 5–15, jul 2003. 68
- [9] K.-i. Okazaki and G. Hummer, “Elasticity, friction, and pathway of γ -subunit rotation in FoF1-ATP synthase,” *Proceedings of the National Academy of Sciences*, vol. 112, no. 34, pp. 10720–10725, 2015. 68
- [10] A. Banari, A. K. Samanta, A. Munke, T. Laugks, S. Bajt, K. Grünewald, T. C. Marlovits, J. Küpper, F. R. N. C. Maia, H. N. Chapman, D. Oberthür, and C. Seuring, “Advancing time-resolved structural biology: latest strategies in cryo-EM and X-ray crystallography,” *Nature Methods*, vol. 22, no. 7, pp. 1420–1435, 2025. 68, 75
- [11] Y.-W. Han, T. Tani, M. Hayashi, T. Hishida, H. Iwasaki, H. Shinagawa, and Y. Harada, “Direct observation of DNA rotation during branch migration of

- Holliday junction DNA by Escherichia coli RuvA–RuvB protein complex,” *Proceedings of the National Academy of Sciences*, vol. 103, no. 31, pp. 11544–11548, 2006. 68, 73, 74, 80
- [12] Y. Kuramoto, “Self-entrainment of a population of coupled non-linear oscillators,” in *International Symposium on Mathematical Problems in Theoretical Physics: January 23–29, 1975, Kyoto University, Kyoto/Japan*, pp. 420–422, Springer, 1975. 69
- [13] T. L. Hill, *Free energy transduction and biochemical cycle kinetics*. Courier Corporation, 2013. 70
- [14] D. T. Gillespie, “A general method for numerically simulating the stochastic time evolution of coupled chemical reactions,” *Journal of Computational Physics*, vol. 22, no. 4, pp. 403–434, 1976. 71
- [15] J. Wald, D. Fahrenkamp, N. Goessweiner-Mohr, W. Lugmayr, L. Ciccarelli, O. Vesper, and T. C. Marlovits, “Mechanism of AAA+ ATPase-mediated RuvAB–Holliday junction branch migration,” *Nature*, vol. 609, no. 7927, pp. 630–639, 2022. 71, 77, 80
- [16] J. Wald and T. C. Marlovits, “Holliday junction branch migration driven by AAA+ ATPase motors,” *Current Opinion in Structural Biology*, vol. 82, p. 102650, 2023. 77
- [17] L. Peliti and S. Pigolotti, *Stochastic thermodynamics: an introduction*. Princeton University Press, 2021. 78
- [18] Q. H. Tran and G. Unden, “Changes in the proton potential and the cellular energetics of Escherichia coli during growth by aerobic and anaerobic respiration or by fermentation,” *European journal of biochemistry*, vol. 251, no. 1-2, pp. 538–543, 1998. 78, 81

Modelling modes of operation displayed by hexameric ATPase motors

Hexameric motors are molecular machines that couple ATP hydrolysis to substrate translocation across a broad range of biological contexts. Recent cryogenic electron microscopy (cryo-EM) studies provide growing evidence that many of these motors coordinate subunit hydrolysis in a sequential fashion. However, this behaviour does not appear to be universal, as other coordination mechanisms were also identified. To date, such patterns have not been systematically studied through theoretical methods. Here, we develop a minimal stochastic model of homohexameric motors inspired by the Sakaguchi–Kuramoto formalism for coupled phase-oscillators, associating subunit conformations with a continuous, periodic phase variable. Distinct modes of operation – probabilistic, concerted, sequential, two-fold, and three-fold symmetric – emerge naturally as dynamical regimes of the model. We find that the sequential mode consistently achieves the highest speed across conditions, helping to explain its prevalence among experimentally characterised hexameric ATPases. We further apply the framework to ClpX chimeric mutants and show that their biochemical behaviour can be reconciled with either probabilistic or sequential dynamics, depending on whether mutations disrupt ATPase activity or inter-subunit coupling. This analysis provides a new interpretation of ClpX biochemical data, suggesting that the sequential model is more robust to perturbations than previously assumed. Altogether, the model introduces a systematic, physics-based definition of hexameric motor operation modes and shows that the sequential mechanism holds a clear kinetic advantage.

Author contributions: Conceptualisation, VHM and PS; Formal analysis, VHM; Writing – Original draft, VHM; Writing – Review and editing, VHM and PS; Visualisation, VHM; Supervision, PS.

5.1 Introduction

Hexameric motors are ubiquitous and functionally diverse protein assemblies present across all domains of life. While some have evolved subunits with specialised roles, like dynein or the F_1 sector of ATP synthase, many hexameric motors are homomeric, comprising identical subunits with the same functional potential. These molecular machines typically perform two interconnected tasks: nucleotide hydrolysis (chemical) and substrate translocation (mechanical), where the free-energy from the former is transduced to the latter. What remains partially understood is how, or whether, subunits coordinate their activities during this process.

Empirical evidence has led to the proposal of different “hydrolysis mechanisms”, linking ATP hydrolysis dynamics to patterns of inter-subunit synchronisation [1, 2, 3, 4]. Despite variations in terminology, these mechanisms classify hexameric motors according to the configuration of bound nucleotides (ATP, ADP, or empty) at each subunit over time. Numerous studies – including biochemical assays [5], single-molecule pulling experiments [6], mutational perturbations [7, 8], and especially X-ray crystallography and single-particle cryo-EM [9, 10, 11, 12, 13, 14, 15, 16] – have provided evidence for and against specific hydrolysis mechanisms in multiple instances of ATPases. Altogether, they reveal a prevalence of a “sequential mechanism”, where subunits hydrolyse ATP one at a time in a defined spatial order [17, 18]. Despite this growing body of examples, currently, there are no studies that either formalise the description of the hydrolysis mechanism or investigate whether (dis)advantages are inherent to each type.

Because it is unclear how to perturb the mechanism in which molecular motors work, we sought to build a toy model that allows their direct comparison. While models for coupled molecular motors are widespread in biophysics [18], very few modelling attempts focus on homohexameric ATPases [19, 20, 21]. To the best of our knowledge, only one study compares different hydrolysis mechanisms through simulations [22]. Yoshimoto and colleagues used molecular dynamics simulations of the E1 helicase engaged in a single-stranded DNA, and then imposed a fixed time interval in which the subunits assume experimentally resolved ATP, ADP, and empty conformations. Using this approach, they were able to devise protocols for different hydrolysis mechanisms and compare them directly. Still, this approach is constrained to the architecture of a particular example of hexameric ATPase.

Instead of relying on a nucleotide-centric definition, which is the basis to justify hydrolysis mechanisms, we argue that a subunit’s conformation is a more fundamental variable to understand the dynamics of hexameric motors. This association

between conformations and function is inspired by the diverse range of symmetry types experimentally observed for homohexameric ATPases, see Fig. 5.1. To model such symmetry types, we identify each subunit conformation with a continuous phase variable, representing its reaction coordinate along the mechanochemical cycle. Thus, we incorporate the formalism of coupled phase-oscillators [23, 24] into the classical stochastic approaches to model molecular motors [25, 26, 27, 28, 29, 30]. By exploring the model parameters, we identify five emergent dynamic regimes – or *modes of operation* – analogous to the previously hypothesised hydrolysis mechanisms. We find that the sequential mode consistently outperforms other modes in terms of speed. Finally, leveraging published data on ClpX chimeric mutants, we show that experimental results can be reconciled with either probabilistic or sequential dynamics depending on whether mutations disrupt ATPase activity or inter-subunit coupling. Using this framework, we provide a systematic characterisation of the different modes of operation and a physics-based argument for the widespread prevalence of the sequential mode in hexameric ATPases.

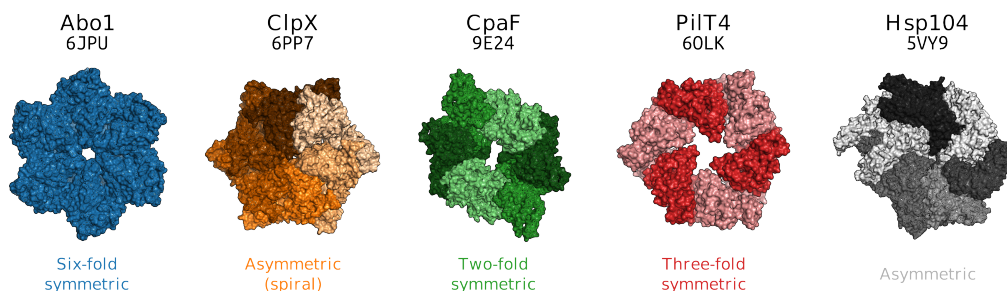


Figure 5.1: *Symmetry types of hexameric motors.* Examples of homohexameric ATPase motors for which their structures were obtained through cryogenic electron microscopy.

5.2 A model for ring-shaped coupled ATPases

We start by introducing a toy model for coupled ATPases using the Langevin formalism. We first develop a general framework that describes any homomeric ring-shaped motor with M subunits, and, in the following, we study the specific case of hexamers (where $M = 6$). Notice that the model introduced below resembles the Master-Equation introduced in the previous chapter with a discrete number of conformational states, denoted by N . The model proposed below can be recovered by taking the continuum limit $N \rightarrow \infty$, while rescaling the length of a conformational step and its respective timescale.

Consider a system composed of M interacting particles labelled by $m \in \{1, 2, \dots, M\}$, which represent ATPase subunits. Each subunit displays a conformation that belongs to a continuous and periodic conformational space $\theta_m \in [0, 2\pi[$. Therefore, the state of the system is the conformation of the whole assembly and corresponds to a vector $\boldsymbol{\theta} = [\theta_1, \dots, \theta_M]$. The system is in contact with a thermal reservoir at fixed temperature T , which causes the particles to change conformations stochastically, thus allowing the subunits to diffuse in the conformational space.

The energetics of the assembly can be divided into two components, a potential $V(\theta)$, that depends on the conformation of each particle; and a coupling energy $E_c(\theta_m, \theta_{m+1})$, that links particles with adjacent indices. We impose that $V(\theta)$ is a continuous and 2π -periodic function, which represents the energetic cost to deform the protein structure throughout the mechanochemical cycle. The main energetic component, which defines the coupling patterns across subunits, is defined as

$$E_c(\theta_m, \theta_{m+1}) = \frac{K}{M} [1 - \cos(\theta_{m+1} - \theta_m - \delta)] \quad , \quad (5.1)$$

with K the coupling strength and δ a phase-lag between adjacent subunits. This particular form of coupling energy is inspired by the Sakaguchi-Kuramoto model [23]. The total energy of the system is

$$E(\boldsymbol{\theta}) = \sum_{m=1}^M [V(\theta_m) + E_c(\theta_m, \theta_{m+1})] \quad . \quad (5.2)$$

Importantly, the coupling strength and the phase-lag play a critical role in this model. As it will become clear later, they control the dynamical regimes of the system, which, in this case, corresponds to the different modes of operation.

In order to observe persistent dynamics, such as a unidirectional processive walk, this system must be driven out of thermodynamic equilibrium, where the driving force comes from the cycle of ATP hydrolysis. We consider that the ATP, ADP, and P_i molecules are chemostated with chemical potential difference $\Delta\mu$ favouring ATP hydrolysis. We assume that this chemical drive biases only a finite interval in the conformational space, which can be interpreted microscopically as the protein conformations which allow ATP binding and ADP release. Thus, we model this energy input as a chemical force $F(\theta_m)$ acting on the subunit m as a function of its conformation coordinate

$$F(\theta_m) = \begin{cases} \Delta\mu/\alpha & \text{if } \theta_m \in [\theta_\mu - \frac{\alpha}{2}, \theta_\mu + \frac{\alpha}{2}], \\ 0 & \text{if } \theta_m \notin [\theta_\mu - \frac{\alpha}{2}, \theta_\mu + \frac{\alpha}{2}], \end{cases} \quad (5.3)$$

where θ_μ corresponds to the conformational coordinate around which the drive occurs and α to the width of the driven interval.

According to Langevin dynamics, the conformational trajectory of the particle m can be expressed as

$$\dot{\theta}_m = \frac{D}{k_B T} \left[-\frac{\partial E}{\partial \theta_m} + F \right] + \sqrt{2D} \xi_m(t) \quad , \quad (5.4)$$

with D the diffusion coefficient, k_B the Boltzmann constant, and $\xi_m(t)$ a stochastic process. For simplicity, we consider $\xi_m(t)$ as a white Gaussian noise, thus satisfying the following conditions for its first and second moments: $\langle \xi_m(t) \rangle = 0$ and $\langle \xi_i(t) \xi_j(t') \rangle = \delta_{ij} \delta(t - t')$. We fix $D = 1 \text{ s}^{-1}$ to set a timescale for the simulations.

By modelling conformations with Langevin dynamics, we can obtain stochastic conformational trajectories for the assembly (see Appendix A for implementation details). While the formulation so far describes the conformational state of individual motor subunits, hexameric motors ultimately operate by displacing relative to a polymeric substrate in a stepwise fashion. Each step typically corresponds to engagement with the next available amino acid or base pair along the substrate axis.

To capture this behaviour, we assume that flux in the continuous conformation space maps to a discrete physical step of the assembly along the substrate track, see Fig. 5.2. This assumption implies a tight coupling between the completion of the conformational cycle and a physical step, thus neglecting potential slippage between subunits and the substrate. In order to quantify such a stepping process, we introduce an intermediate counting function C_m , which discretises the passage of the m -th subunit through a fixed coordinate $\phi \in [0, 2\pi[$

$$C_m(t) = \begin{cases} \delta(t) & \text{if } \theta_m(t) < \phi \text{ and } \theta_m(t + dt) > \phi, \\ -\delta(t) & \text{if } \theta_m(t) > \phi \text{ and } \theta_m(t + dt) < \phi, \\ 0 & \text{otherwise,} \end{cases} \quad (5.5)$$

where $\delta(t)$ is Dirac's delta. The assembly's spatial coordinate x along the substrate is given by

$$x(t) = L \int_0^t \sum_{m=1}^M C_m(t') dt' \quad , \quad (5.6)$$

with L the length of a step and the processive speed $\bar{v}(t) = x/t$. Note that this formulation provides a coarse-grained position of the whole assembly with respect to the substrate, rather than modelling each subunit's position independently.

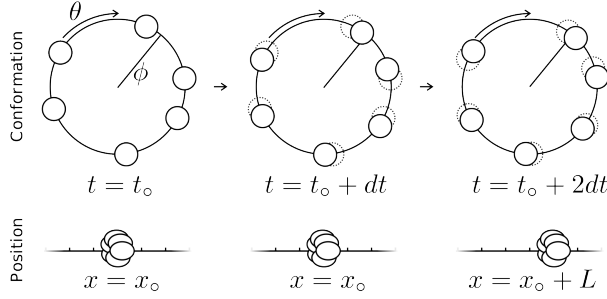


Figure 5.2: *Coupling conformational cycle and ‘walking’ over the substrate.* The scheme represents the relationship between conformation coordinates θ and the assembly’s position x in three states of the system separated by an infinitesimal time interval dt . The spatial coordinate is modelled as a discrete variable updated when a subunit crosses a fix coordinate ϕ .

5.3 Retrieving trajectories with different modes of operation

To characterise the regimes in which the motors operate, we evaluate two emergent behaviours: (i) a sustained maintenance (or absence) of symmetry over time, and (ii) a processive speed, $\bar{v}$. To analyse the first emergent behaviour, we define a mesoscopic quantity $\Delta\theta_{i,j}(t)$, which measures the relative dephase between two subunits with indexes i and j at time t . Because of the 2π -periodicity of the phase space, we formally introduce the relative dephase as the absolute value of the arc-length between θ_i and θ_j as in

$$\Delta\theta_{i,j} = \min(|\theta_i - \theta_j|, |2\pi - \theta_i + \theta_j|) \quad . \quad (5.7)$$

By measuring the dephase between adjacent subunits $\Delta\theta_{m,m+1}$, we can quantitatively track how spaced two subunits along their trajectories, see Fig. 5.3A and B.

When $\Delta\theta_{m,m+1} \approx \Delta\theta \approx 0$, all subunits display approximately the same conformation, and hence the assembly is in a six-fold symmetric state Fig. 5.3C. Given the periodicity of the conformational space, there are few alternative arrangements where the dephase can be nearly the same for all $(m, m+1)$ pairs simultaneously. Some examples are: $\Delta\theta \approx \frac{\pi}{3}$ where the assembly is asymmetric, but displays sequentially ordered subunits in conformation; $\Delta\theta \approx \frac{2\pi}{3}$, where the assembly is two-fold symmetric; and $\Delta\theta \approx \pi$, where the assembly is three-fold symmetric. Also, the system can display a disordered and asymmetric configuration, where $\Delta\theta_{m,m+1}$ are different for each m . Note that the stochastic settings of the model provide an approximated symmetry, as strictly symmetric states are those that fulfil the condition $\theta_{m,m+1} = \Delta\theta$ for all m .

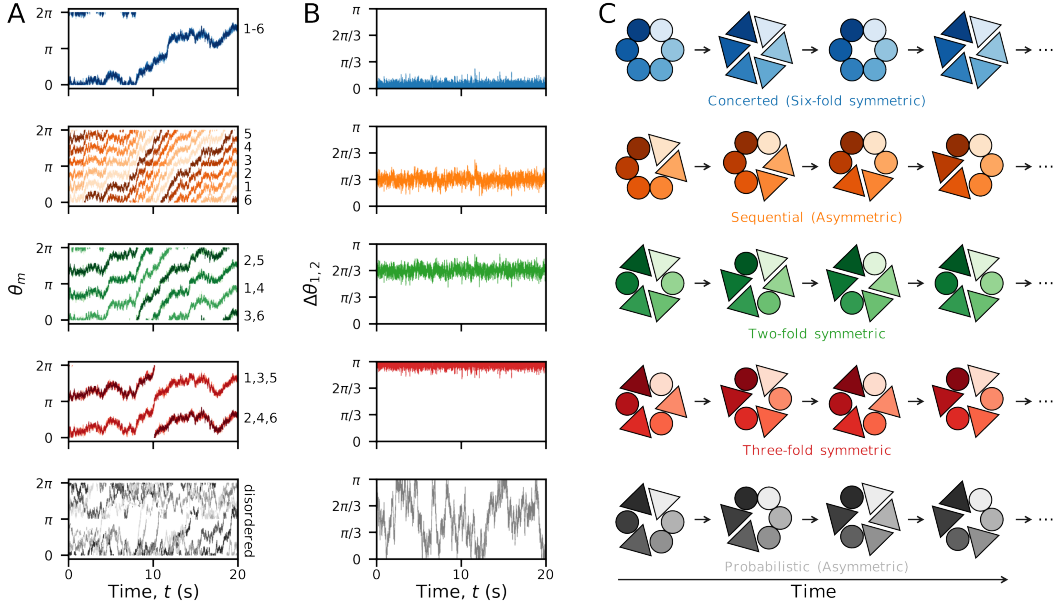


Figure 5.3: *Stochastic trajectories reveals five modes of operation.* Trajectories within each mode of operation are shown in each row **A**. Trajectory θ_m for $m \in [1, \dots, 6]$ over time represented in color shades. At the right of each plot, the index of each subunit is depicted except for the disordered case. **B**. Dephase profile for particles 1 and 2 over time. **C**. Cartoon representing the dynamical modes of operation. The shades of each color depict the identity of each subunit, whereas the geometrical shape represents distinct conformations, which were discretised for a simpler visualisation. Simulation parameters (from top to bottom for K and δ): $\tau = 20\text{s}$, $K = [200, 200, 200, 200, 0]k_{\text{B}}T$, $\Delta\mu = 20k_{\text{B}}T$, $\delta = [0, \pi/3, 2\pi/3, \pi, 0]$, flat potential.

With this operational description of the emergent symmetry of the assembly, we proceed to characterise the impact of the coupling parameters, namely strength K and phase-lag δ , on $\Delta\theta_{m,m+1}$. In Appendix B, we show that the different modes of operation can be encoded as local and global energy minima by changing these two parameters for a flat potential. However, such analytical result is constrained to few regions of the parametric space. We then quantify the mean and standard deviation of relative dephase $\Delta\theta_{1,2}$ from an ensemble of trajectories at a time $t = \tau \gg 2\pi^2/D$ varying K and δ , Fig. 5.4. We start by describing the regime diagram under two simplifying assumptions: (i) that the initial conditions are pre-ordered at $\theta(t=0)$, as in $\Delta\theta_{m,m+1} = \delta$, except for $m=6$, and (ii) that $\partial_\theta V = 0$.

The heatmaps obtained for mean and standard deviation of $\Delta\theta_{1,2}$ depict a clear pattern of self-organisation of hexameric ATPases in non-equilibrium conditions. First, increasing K leads to a regime of lower variance with an inflection around $K \sim 10k_{\text{B}}T$,

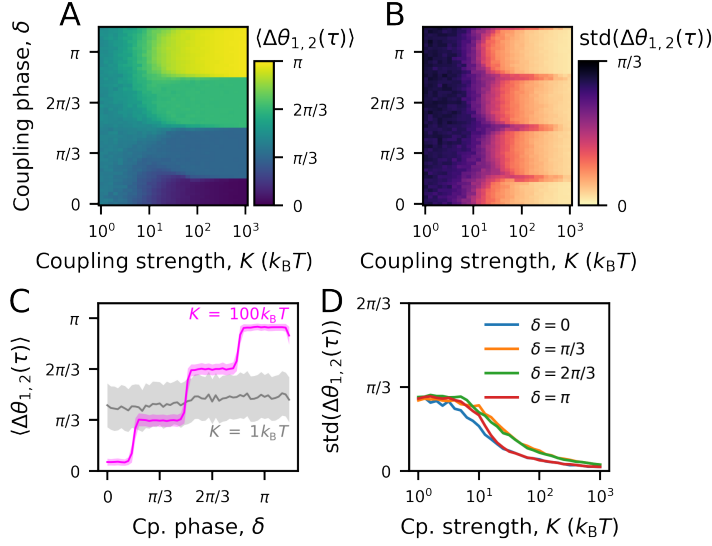


Figure 5.4: *Coupling parameters controls the dynamical regimes of hexameric motors.* **A.** Relative dephase between subunits $m = 1$ and $m = 2$ assessed at the end time of the simulation, at $\tau = 10^3$ s. Mean of $n = 500$ different trajectories with different random seeds. **B.** Standard deviation of relative dephase. **C.** Mean (curve) and standard deviation (shaded area) of relative dephase for low and intermediate coupling strengths. **D.** Standard deviation of relative dephase decays with increasing coupling. Simulation parameters: $\Delta\mu = 20k_B T$, flat potential, ordered initial condition.

identified in Fig. 5.4B and D. The ordered region is further subdivided into four sharp stripes for $\Delta\theta_{1,2}$, see Fig. 5.4A and C. The mean dephase in each stripe matches the closest δ values that can satisfy simultaneously $E_c = 0$ and periodic boundary conditions (see Appendix B). These characteristic stripes are consistent with the concept of phase-locking [31], for which the relative dephase is kept nearly constant, despite the presence of thermal fluctuations and far from equilibrium conditions. In the context of hexameric ATPases, these regions correspond to the concerted, sequential, two-fold, and three-fold symmetric modes of operations.

In summary, we demonstrate that our model recovers the modes of operation typical of hexameric ATPases that can either be observed experimentally or have been hypothesised to exist. We next evaluate how the coupling parameters control in which dynamic mode the assemblies operate using different initial conditions and a non-flat potential $V(\theta)$.

5.3.1 Random initial conditions

The initial conditions, namely the state of the system $\theta(t=0)$, can play a significant role in the assembly's kinetics. As we discuss in detail in Appendix B, each of the ordered modes displays at least one family of local energy minima. For some carefully selected states, we identify that the assembly might not recover the regime encoded by the phase-lag δ , remaining trapped in local minima instead. To test the influence of the initial conditions on the system's regime diagram, we perform simulations with random initial condition as opposed to the previous one, where the subunits were pre-ordered (Fig. 5.5A).

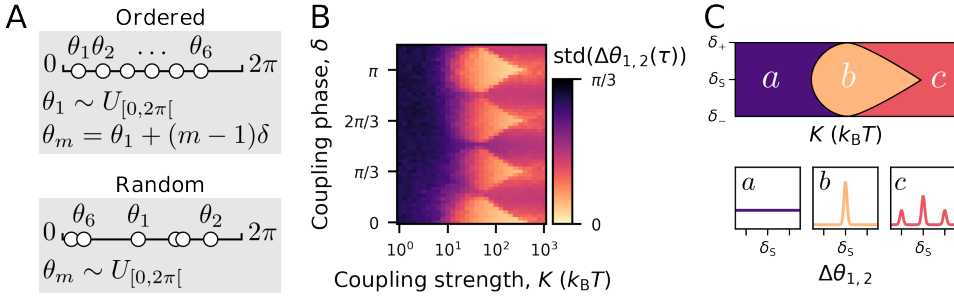


Figure 5.5: *Effect of random initial conditions on the regime diagram* **A**. Process to generate initial conditions at $t=0$ for ordered (top) and the random (bottom) assemblies. **B**. Standard deviation of relative dephase using random initial conditions. Sample of $n=500$ different trajectories with different random seeds and different $\theta(t=0)$. **C**. The assemblies display three distinct behaviours for each δ_s : (a) Probabilistic, (b) Single ordered mode, and (c) Multiple ordered modes. The alternative local minima are situated at $\delta_{\pm} = \delta \pm \pi/3$. Simulation parameters: $\tau = 10^3$ s, $\Delta\mu = 20k_B T$, flat potential.

In the Figure 5.5B, we observe a qualitative difference from the previous variability profile for the relative dephase $\Delta\theta_{1,2}$. For large coupling strengths K , a new region of intermediate standard deviation emerges when compared to the previous profile in Fig. 5.5B. These regions comprise a pool of trajectories, consisting of some trajectories that reached the relative dephase of the closest stable δ , here denoted as δ_s , and some where the assembly is kinetically trapped in the neighbouring modes, see Fig.5.12 for examples.

In summary, a more general description follows the scheme presented in Fig. 5.5C. For low K , thermal fluctuations dominate the behaviour of the assembly, and the assembly operates in a probabilistic mode. For intermediate K , a balance between thermal and coupling energies allows the system to relax to its energy minima, which corresponds to a unique mode of operation, encoded by δ . For large K , the thermal

energy is lower than the energy barriers that separate the global minimum from other local minima, allowing the assembly to stay for phase-locked into distinct modes of operation for longer time-scales, depending on the initial conditions.

5.3.2 Dependence on potential amplitude

So far, we have neglected the effect of the potential energy associated with each conformation by considering $\partial_\theta V = 0$. In more realistic cases, transitions between conformational states can display energetic barriers due to mechanical constraints within the subunit and interactions with the substrate. Importantly, we are aware that both shape and amplitude of the $V(\theta)$ can have a significant impact on the outcomes of a motor's operation, but still, we consider a simple yet numerically tractable example to illustrate how an energy barrier can affect the behaviour of hexameric motors.

In the previous chapter, we inferred the shape and magnitude of this barrier from structural data. Here, inspired by our previous results, we assume that the conformational cycle displays a single energy barrier with a sinusoidal shape.

$$V(\theta) = V_{\max} \cos(\theta - \theta_b) \quad , \quad (5.8)$$

with $V_{\max}$ a parameter that controls the height of the energy barrier and θ_b conformational coordinate where the energy barrier is centred. Here we explore the case $V_{\max} = 10k_B T$, as it represents a barrier significantly larger than thermal fluctuations and is in the same order of magnitude as the elastic energy we estimated for RuvB in its mechanochemical cycle [32]. The choice of θ_b matters, as its relative position compared to the coordinate of driving force, θ_μ , can have a significant impact on the emergent speed, as we will discuss later.

In Fig. 5.6, we explore the effect of $V(\theta)$ on the previously characterised coupling profile. It is clear from Fig. 5.6A and B that the probabilistic mode has a much narrower distribution of $\Delta\theta_{m,m+1}$ and that the boundaries between modes are modified. At low coupling strengths, the subunit conformations will tend to accumulate at the minimum of energy of V , which is identical for each subunit, and therefore approximates the behaviour of probabilistic and concerted modes, Fig. 5.6C. Moreover, the transitions between ordered and disordered modes (except for the concerted mode) display a peak of high variance, a distinctive feature not observed for a flat potential, see Fig. 5.6D. This is probably due to the interplay between V , which favours the grouping of subunits in the conformational landscape, and E_c , which favours their spread.

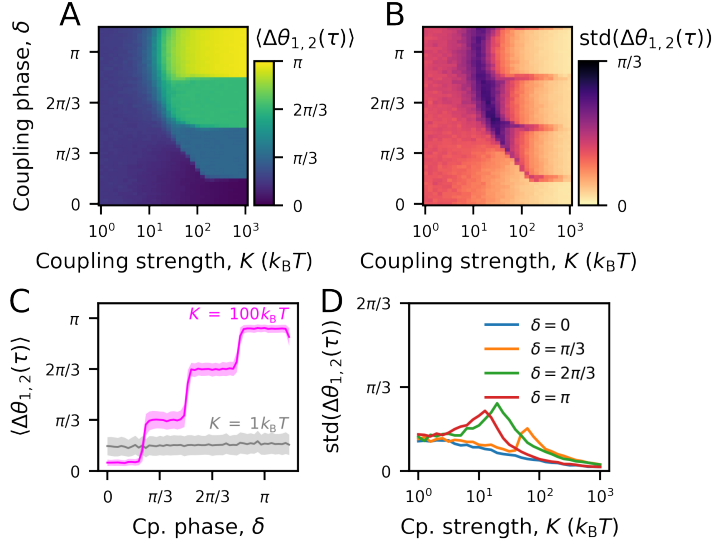


Figure 5.6: *Regime diagram of coupled ATPases in the presence of an energy barrier* **A.** Relative dephase heatmap for varying coupling strength and phase-lag (mean of $n = 500$ trajectories). **B.** Heatmap of standard deviation of relative dephase. **C.** Mean (curve) and standard deviation (shaded area) of relative dephase for low and intermediate coupling strengths. **D.** Standard deviation profiles are non-monotonous for $\delta \neq 0$. Simulation parameters: $\tau = 10^3\text{s}$, $\Delta\mu = 20k_B T$, sinusoidal potential with $V_{\max} = 10k_B T$, ordered initial condition.

5.4 Sequential mode operates faster in all conditions

After identifying the conditions for the hexameric motors to operate in each mode, we evaluate their processive speeds, $\bar{v}$. In Fig. 5.7, we explore the interplay of coupling and potential energies in shaping the motor's speed. Because we set the same diffusion constant D in all simulations, we can directly compare the emergent speeds for different parameter sets.

In the absence of potential energy, all phase-lags δ display a non-monotonic behaviour of $\bar{v}$ when increasing K except for the concerted mode, which is monotonically decreases with K . The sequential, three-fold, and two-fold symmetric modes display a maximum of speed at intermediate values of coupling strength, all of them surpassing the probabilistic mode. The presence of a peak in speed suggests that there is a trade-off between coupling and thermal energy in all cases, therefore, tightly coupled subunits are not optimal for maximising speed.

When we consider the case of a potential $V(\theta)$ with an energy barrier $V_{\max} = 10k_B T$, the conformational coordinate of the non-equilibrium drive plays a major role. For instance, the motor is processive in the probabilistic mode when θ_μ localises to the

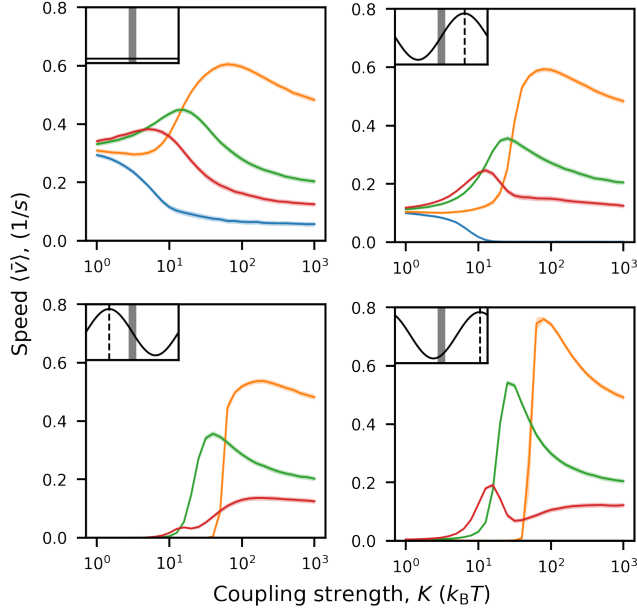


Figure 5.7: *Speed vs coupling strength curves in different potentials.* The panels show the behaviour of speed (mean $\pm$ standard deviation, $n = 500$ simulations) with increasing coupling strengths with phase-lags $\delta = 0$ in blue, $\delta = \pi/3$ in orange, $\delta = 2\pi/3$ in green, and $\delta = \pi$ in red. Insets show the interplay of the potential V (curve) and the chemical force F (shaded area). Top left: flat potential, $V(\theta) = 0$. Top right: $\theta_b - \theta_\mu = \pi/2$. Bottom left: $\theta_b - \theta_\mu = -\pi/2$. Top right: $\theta_b - \theta_\mu = 5\pi/6$. For those with sinusoidal potential, $V_{\max} = 10k_B T$. Other simulation parameters: $\tau = 10^3 \text{s}$, $\Delta\mu = 20k_B T$, ordered initial condition.

region of maximum slope of V . In all the other cases we show, there is no detectable speed for the probabilistic and concerted modes. Remarkably, the maximum speeds reached by the other modes display the same hierarchy regardless of the energy barrier coordinate, with the sequential mode being the fastest. We also note that, counter-intuitively, the sequential (as well as the three-fold symmetric) mode can achieve speeds even higher than those in flat potentials. This suggests that such arrangement can benefit from the interplay of drive and potential to boost speed even further.

5.5 ClpX as case study: probabilistic or sequential?

ClpX, a hexameric AAA+ ATPase, has long been considered a paradigmatic example of a probabilistic motor. A central piece of evidence comes from the work of Martin and collaborators, who engineered covalently linked ClpX assemblies containing specific mutant subunits [7]. The observation that the system was processive in the

presence of multiple mutant subunits with inhibited hydrolysis led to the conclusion that both sequential and concerted mechanisms were unfeasible, as a single inactive subunit would stall the motor. This seems to be, however, not fully consistent with the cryo-EM structures of ClpX [33], which display a staircase-like structure, typically found in sequential assemblies. Still, the authors do not rule out other mechanistic processes that ensure probabilistic hydrolysis while maintaining a sequential coupling. Motivated by this discrepancy, we reanalyse their data in the light of the toy model we proposed to test whether ClpX behaves as a probabilistic or sequential motor.

5.5.1 ATPase activity dataset and model assumptions

In the experimental setup, Martin and colleagues used a synthetic biology approach to generate multiple ClpX subunits connected by linkers [7]. The resulting assemblies consisted of hexamers with subunits of three types: *W* (wildtype), *E* (mutant E185Q), and *R* (mutant R370K). They represented the constructs as a sequence of six letters corresponding to the arrangement of the subunits in the assembly. For example, *WRE/WRE* corresponds to an assembly composed of two repeated molecules of *WRE*, with the three subunits linked covalently. The experimental data we use consists of ATPase activities for 13 combinations of subunit types, including a fully wild-type assembly. We further simplify the dataset by normalising all ATPase activities to the mean of wild-type assemblies and neglecting the effect of the linkers. Next, we introduce how we define ATPase activity inferred from simulations.

First, we must consider the set of assumptions underlying the effect of the mutations on the emergent ATPase activity of ClpX. In the article, the authors argue that the mutation *E*, in the Walker B motif, allows ClpX to bind ATP but not to hydrolyse it, both corroborated experimentally for the case of full mutant assemblies [34]. The mutation *R*, in the sensor-2 motif, was shown to also bind ATP without hydrolysing it in a full mutant, and, in addition, it is not capable to bind the substrate [35]. This led the authors to hypothesise that the *R* mutant is trapped in the “ADP conformation” [35, 7]. To accommodate these assumptions in our model, we simulate trajectories of assemblies distinguishing each type of subunit. While *W* subunits work as described earlier, we impose *E* to be insensitive to the drive $\Delta\mu$, and *R* to remain at rest in its initial conformation, see Fig. 5.8A. We then consider the ATPase activity as the number of cycles completed by *W* subunits.

In addition to the set of assumptions above, we propose that the mutation *R* has a different consequence on ClpX activity. Instead of locking ClpX in a specific conformation, we suggest that this mutation disrupts inter-subunit coupling and has no detrimental effect on hydrolysis, see Fig. 5.8B. The following pieces of evidence

motivate this interpretation: *(i)* in an evolutionary related hexameric ATPase, MCM, the sensor-2 motif affects hydrolysis in *trans*, but not in *cis* [8]; *(ii)* in ClpX, the sensor-2 motif bridges nucleotide binding-pocket and the neighbouring subunit [33]; *(iii)* also in another related hexameric ATPase, RuvB, the sensor-1 motif mediates coupling between adjacent subunits [32]. Because sensor-1 is not functional in ClpX [36], this function might be mediated by sensor-2 instead. Importantly, this assumption is not inconsistent with the observation that a full *R* mutant does not hydrolyse ATP [35]. In this case, a full *R* mutant would effectively uncouple all subunits, and therefore enforce the probabilistic mode, which can be inactive depending on other model parameters (see Fig. 5.7 for examples). Under this alternative scenario, we compute the ATPase activity as the number of cycles completed by *W* and *R* subunits.

5.5.2 Hypothesis testing

To test which mode of operation better explains the experimental data, we simulated ensembles of hexameric assemblies under the two contrasting assumptions described above. Specifically, we simulate trajectories in a range of coupling strengths K and heights of the energy barrier $V_{\max}$, while fixing the phase-lag $\delta = \pi/3$. This ensures we test a gradient of cases where the wildtype assembly behaves from a probabilistic (low K) to a sequential motor (large K). We then compare the results of normalised ATPase activities in simulations to the experimental data using the coefficient of determination R^2 , see Fig. 5.8.

We observe that, by following the assumptions proposed by Martin and colleagues, the data shows a modest agreement with the probabilistic mode, but a poor agreement with the sequential, Fig. 5.8C. For totally uncoupled subunits and in a flat potential, we obtain $R^2 = 0.5 \pm 0.1$. Also, the R^2 profiles are roughly unchanged across different $V_{\max}$. As K becomes sufficiently large, most mutant arrangements become stalled and therefore the correspondence between data and simulations vanishes, Fig. 5.8E. By contrast, when we assume that the mutation *R* disrupts coupling, we observe an inverted trend. At low K , when subunits behave independently, the simulated trajectories overestimate the catalytic activity of the mutants, thereby failing to match the experimental data, Fig. 5.8F. However, when coupling is large enough, simulations and data show strong agreement, reaching up to $R^2 = 0.87 \pm 0.02$ (for $K \approx 30k_B T$ and $V_{\max} = 0$). Interestingly, the dependence of R^2 on the energy barrier is non-monotonic with $V_{\max}$, with two distinct optima at $V_{\max} = 0$ and $V_{\max} = 8k_B T$. For completeness, we verify that all other ordered modes show no agreement with the data regardless of the assumption we consider (Fig. 5.9).

In summary, we find that the experimental data are better explained by a sequen-

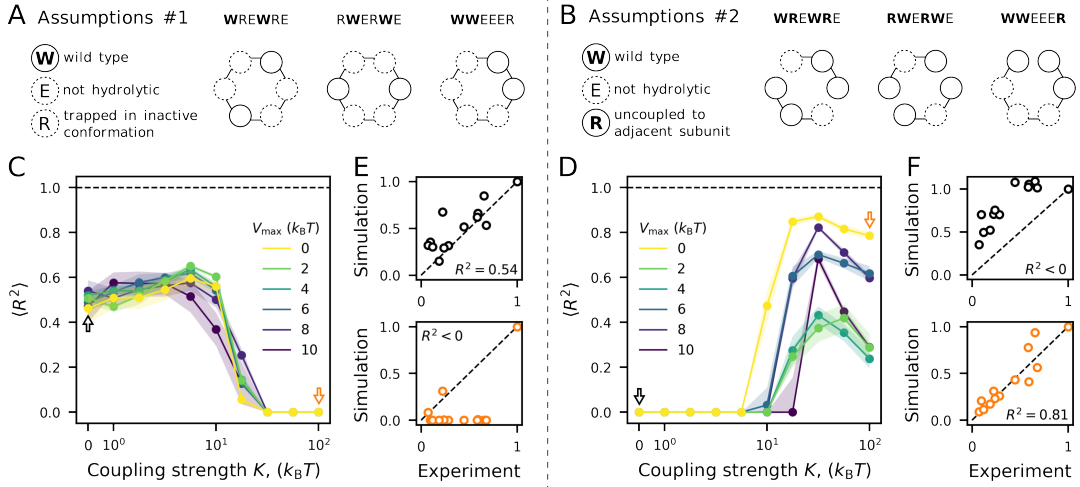


Figure 5.8: *Experimental data on ClpX ATPase activity compared to simulations.* **A.** Set of assumptions proposed by [7] to interpret the effect of ClpX mutations E185Q (mutant *E*) and R370K (mutant *R*). On the right, we highlight a few examples of assemblies distinguishing subunits that are able or unable to hydrolyse ATP. **B.** Set of assumptions we propose, which differs from the previous by removing coupling connectivity instead of hydrolytic capability in the mutant *R*. **C.** R^2 calculated between experimentally measured ATPase activity and simulations following Assumptions #1, normalised by the mean ATPase activity of WWWWW constructs. The values represent a mean of $n = 20$ sets of simulations for each mutant arrangement. **D.** R^2 calculated following Assumptions #2. **E.** Examples of ATPase activity in experiment and simulations for uncoupled $K = 0$ (black) and coupled $K = 100 k_B T$ (orange) subunits, considering Assumptions #1. **F.** Examples simulations considering Assumptions #2. Simulation parameters: $\tau = 10^3 s$, $\delta = \pi/3$, $\Delta\mu = 20 k_B T$, $\theta_b - \theta_\mu = \pi/2$, ordered initial condition.

tial model, provided that the R370K mutation disrupts inter-subunit coupling. This interpretation reconciles the biochemical activity measurements with the staircase structure observed in ClpX, suggesting that ClpX represents yet another instance of a sequential motor, a pattern increasingly observed across several hexameric ATPases.

5.6 Discussion

Despite their stark importance to a vast variety of cellular activities, homo-hexameric ATPases remain largely unexplored from a theoretical perspective. In particular, no previous work has systematically characterised the kinetics of different modes of operation of hexameric ATPases using a biophysical framework. To address this gap, we proposed a simple stochastic model that has at its core an inter-subunit coupling

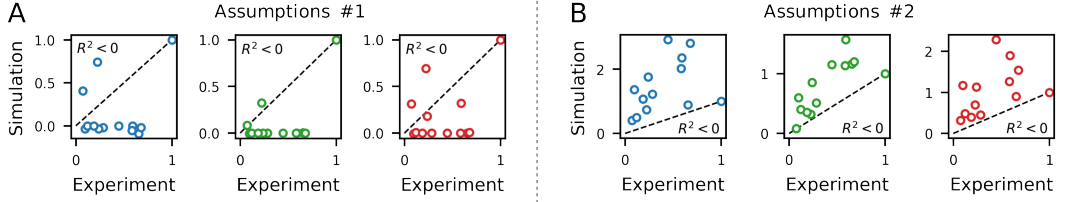


Figure 5.9: *Concerted, two-fold, and three-fold symmetric modes are incompatible with observed ClpX ATPase activity.* **A.** Simulations run considering Assumptions #1: concerted ($\delta = 0$, blue); two-fold symmetric ($\delta = 2\pi/3$, green); three-fold symmetric ($\delta = \pi$, red); **B.** Simulations run considering Assumptions #2: concerted mode ($\delta = 0$, blue); two-fold symmetric ($\delta = 2\pi/3$, green); three-fold symmetric ($\delta = \pi$, red). Simulation parameters: $\tau = 10^3\text{s}$, $K = 100k_{\text{B}}T$, $\Delta\mu = 20k_{\text{B}}T$, flat potential, ordered initial condition.

term inspired by the Sakaguchi-Kuramoto formalism for coupled phase-oscillators [23]. While the idea of using phase-oscillators to describe molecular motors is not novel [37, 38], these studies have considerably different models and goals.

Using this model, we identify the regions of the parametric space which lead to five distinct dynamic regimes, or modes of operation. When subunits are weakly coupled or uncoupled (low K), their conformations are independent, characteristic of the probabilistic mode. The intermediate to strongly coupled regime is divided into four regions, depending on the phase-lag δ , corresponding to the concerted ($\delta = 0$), sequential ($\delta = \pi/3$), two-fold symmetric ($\delta = 2\pi/3$), and three-fold symmetric ($\delta = \pi$) modes. We further show that a motor can be kinetically trapped but still operational at a mode different from its encoded phase-lag, depending on the assembly conditions.

A key observation from the model is that the sequential mode outperforms all others in speed. Furthermore, in all tested conditions, we observed that the maximum speeds at each mode follow the hierarchy: sequential $>$ two-fold symmetric $>$ three-fold symmetric $>$ probabilistic $\geq$ concerted. This hierarchical order of speeds is in line with the observed prevalence of sequential motors, suggesting that evolution might have favoured the maintenance of sequentiality due to its kinetic advantage. The second fastest mode is also populated by a few examples of assemblies, in particular as components of bacterial pili and secretion systems [13, 15]. Conversely, the protein assembly LTag, thought to operate as a concerted motor [9], is predicted to work at comparatively lower speeds, thus agreeing with predictions from molecular dynamics [22]. However, the model also predicts that six-fold symmetric structures can be at local or global energy minima but still be capable of operating sequentially. According

to our model, this is possible because trajectories whose phase-lags promote concerted mode can be kinetically trapped in the sequential mode and vice versa.

Lastly, we show an application of the model to a dataset which combines ClpX mutants and an *in vitro* characterisation of their ATPase activity. By testing two contrasting interpretations of mutational experiments, we conclude that the data can be reconciled with either a probabilistic or a sequential mode. While the statistical analysis points to a better agreement with the sequential mode, our results are contingent on fundamental assumptions of ClpX biochemistry that require further investigation. Future experimental assays, such as solving the structure of the chimeric mutants, for instance, could resolve the present ambiguities. Nevertheless, our work challenges the traditional view that the inhibition of a single subunit disagrees with a sequential mechanism [6, 7], and also provides a means to test the hypothesis of which mode of operation is more consistent with experimental data.

In summary, we present a simple and general framework to describe the coordinated operation of hexameric motors. With this model, we identify a possible explanation for the prevalence of sequential motors and evaluate the case of ClpX by comparing the model predictions to biochemical experiments. Besides this application, this model can be used to uncover fundamental properties of hexameric motors coupled to other methods, such as single-molecule biophysical assays. Moreover, it could also be employed to explore evolutionary pathways leading to the functionally diverse families of hexameric ATPases.

5.7 Appendix A: Implementation details

To simulate stochastic trajectories, we build an in-house script to discretise the stochastic differential equations using the Euler method. We rewrite the Eq. 5.4 in the differential form

$$d\theta_m = \frac{D}{k_B T} \left[-\frac{\partial E}{\partial \theta_m} + F \right] dt + \sqrt{2D} dW_m \quad , \quad (5.9)$$

where dW_m is a Wiener process.

To test whether the time discretisation dt does not introduce any systematic bias in the trajectories of $\theta(t)$, we tested a range of values of dt . In particular, we investigate whether the distributions of $d\theta$ steps and speed $\bar{v}$ converge as dt decreases. We observe that, as the system is driven increasingly far from equilibrium, systematic biases on the step size $d\theta$ for $dt = 10^{-2}$ s and to a smaller extent on $dt = 10^{-3}$ s, see Fig. 5.10A. At $dt \leq 10^{-4}$ s, the distributions of step-size collapse to a single curve at different chemical drives $\Delta\mu$, matching the step-size distribution expected for a stochastic diffusive process. A similar pattern is observed for the speeds $\bar{v}$, which stabilise at $dt \leq 10^{-3}$ s, Fig. 5.10B.

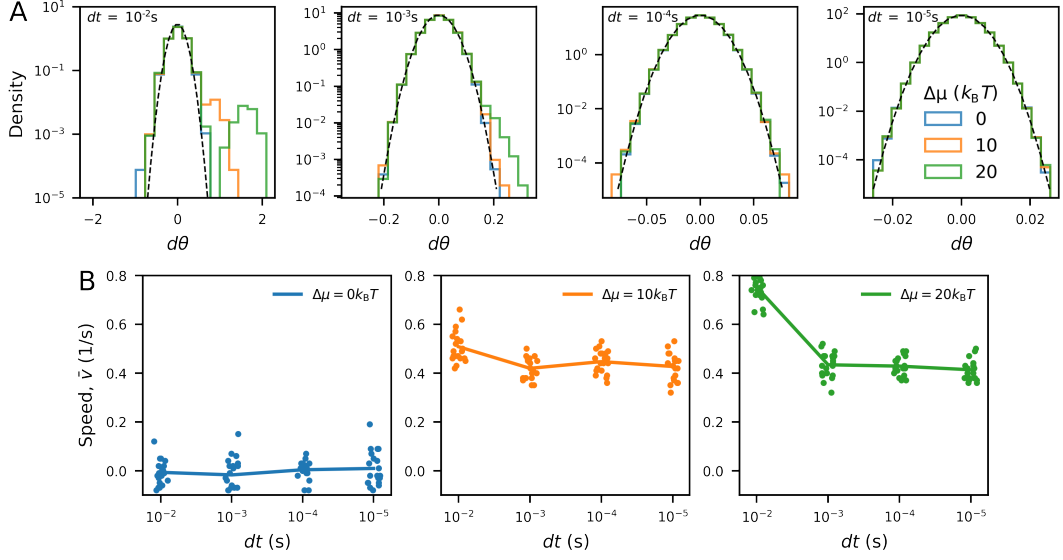


Figure 5.10: *Step-size $d\theta$ and speeds of simulations using different discretisation times.* **A.** Density plots of conformational step-sizes $d\theta_m$ for all m , varying the time increment dt , and the drive $\Delta\mu$. Dashed lines represent a normal distribution $\mathcal{N}(0, \sqrt{2Ddt})$, which corresponds to the distribution of step-sizes from the diffuse component of $d\theta_m$. **B.** Speed calculated for $n = 10$ repetitions per drive and time increment with different random seeds. Simulation parameters: $\tau = 100s$, $K = 100k_B T$, $\delta = \pi/3$, $V_{\max} = 10k_B T$, $\theta_b = \theta_\mu$, ordered initial conditions.

5.8 Appendix B: Equilibrium conditions for $V(\theta) = 0$ and $K > 0$

We now focus on identifying the system configurations that are at thermodynamic equilibrium. In particular, we consider the case of a flat potential to explore how the coupling energy influences the distribution of the subunits in the conformational landscape. We first introduce the variable phase difference $\varphi_m = \theta_{m+1} - \theta_m$ to simplify the calculations. Then, we investigate the conditions where all particles satisfy

$$\frac{\partial E}{\partial \theta_m} = 0, \text{ for } m = \{1, 2, \dots, M\} \quad .$$

For a flat potential, the partial derivative consists of

$$\begin{aligned} \frac{\partial E}{\partial \theta_m} &= \frac{\partial}{\partial \theta_m} \left[\sum_{m=0}^M E_c(\theta_m, \theta_{m+1}) \right] \\ \frac{M}{K} \frac{\partial E}{\partial \theta_m} &= \sin(\theta_m - \theta_{m-1} - \delta) - \sin(\theta_{m+1} - \theta_m - \delta) \\ &= \sin(\varphi_{m-1} - \delta) - \sin(\varphi_m - \delta) \quad . \end{aligned}$$

Therefore, the critical points are those satisfying

$$\sin(\varphi_{m+1} - \delta) = \sin(\varphi_m - \delta) \quad .$$

This equation admits two solutions,

$$\begin{aligned} \varphi_{m+1} - \varphi_m &= 0 \quad \text{and,} \\ \varphi_{m+1} - \varphi_m &= (2n+1)\pi + 2\delta \quad , \end{aligned} \tag{5.10}$$

with $n \in \mathbb{Z}$. Because there are M instances for this equilibrium of forces, the system can display a combination of those solutions for the different particles.

To evaluate the stability of the solutions, we compute the Hessian matrix of the energy function $\mathcal{H}(E)$. Given the symmetry of the coupling terms, the matrix elements of the Hessian matrix are zero for non-adjacent indices, therefore $\mathcal{H}(E)$ is a cyclic tridiagonal matrix. The elements of the main diagonal and lower/upper diagonals are

$$\begin{aligned} \frac{M}{K} \mathcal{H}(E)_{m,m} &= \cos(\varphi_{m-1} - \delta) + \cos(\varphi_m - \delta) \quad , \\ \frac{M}{K} \mathcal{H}(E)_{m,m+1} &= \frac{M}{K} \mathcal{H}(E)_{m+1,m} = -\cos(\varphi_m - \delta) \quad , \end{aligned}$$

From this, it is clear that $\mathcal{H}(E)$ is singular independently from the state θ , because the sum of the row m is

$$\sum_i^M \mathcal{H}(E)_{m,i} = \mathcal{H}(E)_{m,m-1} + \mathcal{H}(E)_{m,m} + \mathcal{H}(E)_{m,m+1} = 0 \quad ,$$

for $m = 1, \dots, M$. Thus, $\mathcal{H}(E)$ has $\lambda = 0$ as an eigenvalue regardless of the system's configuration. This implies that all states of the system are saddle points, and therefore, the Hessian matrix does not help us to determine the minima of E . However, for some specific cases, it is possible to determine the global minima of the system, which we describe next.

5.8.1 Global minima

For a flat potential, the energy function $E(\theta)$ consists of a sum of M coupling energies E_c which are non-negative. Therefore, those states for which $E(\theta) = 0$ are necessarily the global minima of E , or

$$E_c(\theta_m, \theta_{m+1}) = 0 \quad \text{for all } m,$$

which implies

$$\cos(\varphi_m - \delta) = 1 \quad \text{for all } m.$$

To satisfy the condition above, we take into account the periodicity of the space Θ . We set this geometrical constraint by imposing

$$\sum_m^M \varphi_m = 2\kappa\pi \quad ,$$

with $\kappa \in \mathbb{Z}$. Because of the minimal energy condition, we can write $\varphi_m = \delta$, which leads us to

$$\delta = \frac{2\kappa\pi}{M} \quad , \tag{5.11}$$

with the phase difference at rest δ limited to $[0, 2\pi[$, there are only M solutions for which $E(\theta) = 0$. In particular, when $M = 6$, the solutions for $\varphi_m = \delta$ are $\left\{0, \frac{\pi}{3}, \frac{2\pi}{3}, \pi, \frac{4\pi}{3}, \frac{5\pi}{3}\right\}$. We highlight that, in terms of modes of operation, these six solutions correspond to only four qualitatively different modes, for the cases $\delta = \frac{\pi}{3}$ and $\delta = \frac{5\pi}{3}$ correspond to a sequential order whereas $\delta = \frac{2\pi}{3}$ and $\delta = \frac{4\pi}{3}$ correspond to a two-fold symmetric case.

5.8.2 Some local minima

As discussed previously in this session, we cannot rely on the Hessian matrix to identify local minima. We then turn our attention to the constraints imposed in the values of φ_m (Eq. 5.10) and δ (Eq. 5.11).

To do so, we evaluate if the states which satisfy the condition $\varphi_m = \varphi_{m+1}$ but not $\varphi_m = \delta$ also configure as local minima of E . Restricting our attention to the interval $[0, \pi/2]$ and $M = 6$ limits our search to four values of φ_m . We then construct an arbitrary θ for which all coordinates respect the aforementioned condition, and then perturb the coordinate of an arbitrary subunit. Because of the periodicity of the boundary conditions and the sequential coupling rule, the choice of the perturbed subunit does not lead to any loss of generality. Therefore, we consider:

$$\theta_m = (m - 1)\psi$$

with $\psi \in \left\{0, \frac{\pi}{3}, \frac{2\pi}{3}, \pi\right\}$ for $m \neq 1$ and θ_1 varying freely. This leads to $\varphi_m = \psi$ for $m \in \{2, 3, 4, 5\}$. Clearly, the values of φ_1 and φ_6 are not fixed. In these settings, the condition $\varphi_1 = \varphi_6 = \psi$ is only met when $\theta_1 = 0$. We then evaluate how $E(\theta)$ varies with θ_1 for different values of δ , Fig. 5.11.

As we predicted earlier, the panels in the diagonal (purple curves) reveal that all instances where $\varphi_m = \varphi_{m+1} = \delta$ configure the global minima of the system, because $E(\theta) = 0$. Besides, all curves display two critical points, one of which has positive concavity. For the brown curves ($|\psi - \delta| = \pi/3$), the energy minima occurs for $\theta_1 = 0$, and therefore satisfies the condition Eq. 5.10. Because the choice of another m would not result in any change in the result above, this configuration corresponds to a local energy minimum of the whole system.

For the remaining two cases (pink and olive curves), the local minimum identified with this approach does not correspond to $\theta_1 = 0$. Therefore, the condition Eq. 5.10 is not necessarily satisfied, and the symmetry argument used above is no longer applicable. We then calculate the first and second derivatives of total energy for each case directly. In the first group, for which $|\psi - \delta| = 2\pi/3$ (pink), the coordinate $\theta_1 = \pi$ is not a local energy minimum for all m . For $|\psi - \delta| = \pi$ (olive), the tested coordinate has null first and second derivatives for the other θ_m . We also tested whether these secondary energy minima could be observed through simulations of trajectories, supporting our analytical observations (Fig. 5.12).

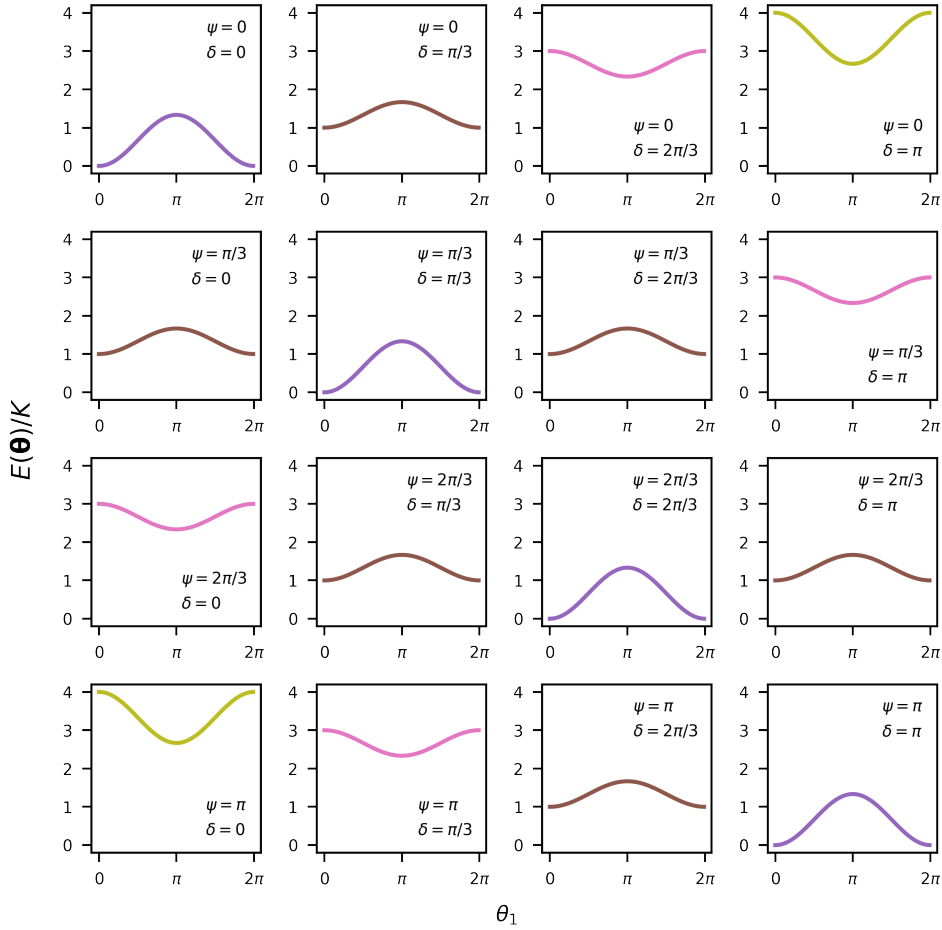


Figure 5.11: *Local and global energy minima.* Evaluation of $E(\theta)$ for a system with a single degree of freedom θ_1 . There are four distinct shapes for the curves, one for each possible combination of $|\psi - \delta|$. Colors represent each of these groups of shape.

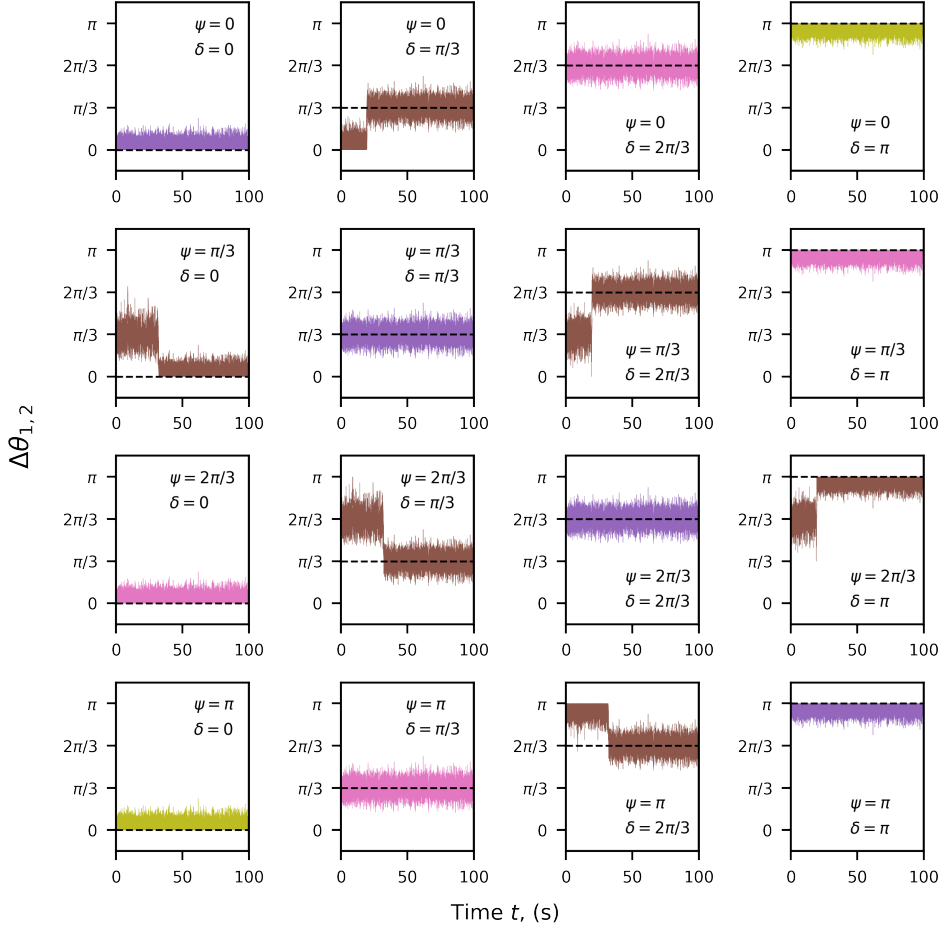


Figure 5.12: *System states remain transiently trapped in local minima at equilibrium.* Simulation of conformational trajectories assessed by the relative dephase between subunits 1 and 2. Initial conditions follow $\varphi_m = \psi$ for all m . Dashed lines represent the dephase for which $E_c(\theta_m, \theta_{m+1}) = 0$. Simulation parameters: $\tau = 100\text{s}$, $K = 200k_B T$, $\Delta\mu = 0$, flat potential.

Bibliography

- [1] M. R. Singleton, M. S. Dillingham, and D. B. Wigley, "Structure and Mechanism of Helicases and Nucleic Acid Translocases," *Annual Review of Biochemistry*, vol. 76, no. Volume 76, 2007, pp. 23–50, 2007. 86
- [2] A. Y. Lyubimov, M. Strycharska, and J. M. Berger, "The nuts and bolts of ring-translocase structure and mechanism," *Current Opinion in Structural Biology*, vol. 21, no. 2, pp. 240–248, 2011. 86
- [3] A. J. Fernandez, , and J. M. Berger, "Mechanisms of hexameric helicases," *Critical Reviews in Biochemistry and Molecular Biology*, vol. 56, pp. 621–639, nov 2021. 86
- [4] Y. A. Khan, W. K. Ian, , and A. T. Brunger, "The AAA+superfamily: a review of the structural and mechanistic principles of these molecular machines," *Critical Reviews in Biochemistry and Molecular Biology*, vol. 57, pp. 156–187, mar 2022. 86
- [5] B. L. Stitt and Y. Xu, "Sequential Hydrolysis of ATP Molecules Bound in Interacting Catalytic Sites of Escherichia coli Transcription Termination Protein Rho*," *Journal of Biological Chemistry*, vol. 273, no. 41, pp. 26477–26486, 1998. 86
- [6] M. Sen, R. Maillard, K. Nyquist, P. Rodriguez-Aliaga, S. Pressé, A. Martin, and C. Bustamante, "The ClpXP Protease Unfolds Substrates Using a Constant Rate of Pulling but Different Gears," *Cell*, vol. 155, pp. 636–646, oct 2013. 86, 101
- [7] A. Martin, T. A. Baker, and R. T. Sauer, "Rebuilt AAA + motors reveal operating principles for ATP-fuelled machines," *Nature*, vol. 437, no. 7062, pp. 1115–1120, 2005. 86, 96, 97, 99, 101
- [8] M. J. Moreau, A. T. McGeoch, A. R. Lowe, L. S. Itzhaki, and S. D. Bell, "ATPase Site Architecture and Helicase Mechanism of an Archaeal MCM," *Molecular Cell*, vol. 28, no. 2, pp. 304–314, 2007. 86, 98
- [9] D. Gai, R. Zhao, D. Li, C. V. Finkielstein, and X. S. Chen, "Mechanisms of Conformational Change for a Replicative Hexameric Helicase of SV40 Large Tumor Antigen," *Cell*, vol. 119, no. 1, pp. 47–60, 2004. 86, 100
- [10] E. J. Enemark and L. Joshua-Tor, "Mechanism of DNA translocation in a replicative hexameric helicase," *Nature*, vol. 442, no. 7100, pp. 270–275, 2006. 86

- [11] C. Deville, M. Carroni, K. B. Franke, M. Topf, B. Bukau, A. Mogk, and H. R. Saibil, “Structural pathway of regulated substrate transfer and threading through an Hsp100 disaggregase,” *Science Advances*, vol. 3, no. 8, p. e1701726, 2017. 86
- [12] E. Zehr, A. Szyk, G. Piszczek, E. Szczesna, X. Zuo, and A. Roll-Mecak, “Katanin spiral and ring structures shed light on power stroke for microtubule severing,” *Nature Structural Molecular Biology*, vol. 24, no. 9, pp. 717–725, 2017. 86
- [13] V. Solanki, S. Kapoor, and K. G. Thakur, “Structural insights into the mechanism of Type IVa pilus extension and retraction ATPase motors,” *The FEBS Journal*, vol. 285, no. 18, pp. 3402–3421, 2018. 86, 100
- [14] J. Wald, D. Fahrenkamp, N. Goessweiner-Mohr, W. Lugmayr, L. Ciccarelli, O. Vesper, and T. C. Marlovits, “Mechanism of AAA+ ATPase-mediated RuvAB–Holliday junction branch migration,” *Nature*, vol. 609, no. 7927, pp. 630–639, 2022. 86
- [15] M. Hohl, E. J. Banks, M. P. Manley, T. B. K. Le, and H. H. Low, “Bidirectional pilus processing in the Tad pilus system motor CpaF,” *Nature Communications*, vol. 15, no. 1, p. 6635, 2024. 86, 100
- [16] I. Cooney, H. L. Schubert, K. Cedeno, O. N. Fisher, R. Carson, J. C. Price, C. P. Hill, and P. S. Shen, “Visualization of the Cdc48 AAA+ ATPase protein unfolding pathway,” *Nature Communications*, vol. 15, no. 1, p. 7505, 2024. 86
- [17] C. Puchades, C. R. Sandate, and G. C. Lander, “The molecular principles governing the activity and functional diversity of AAA+ proteins,” *Nature Reviews Molecular Cell Biology*, vol. 21, no. 1, pp. 43–58, 2020. 86
- [18] S. N. Gates and A. Martin, “Stairway to translocation: AAA+ motor structures reveal the mechanisms of ATP-dependent substrate translocation,” *Protein Science*, vol. 29, no. 2, pp. 407–419, 2020. 86
- [19] H. Liu, Y. Shi, X. S. Chen, and A. Warshel, “Simulating the electrostatic guidance of the vectorial translocations in hexameric helicases and translocases,” *Proceedings of the National Academy of Sciences*, vol. 106, no. 18, pp. 7449–7454, 2009. 86
- [20] N. Koga and S. Takada, “Folding-based molecular simulations reveal mechanisms of the rotary motor F₁/F₀-ATPase,” *Proceedings of the National Academy of Sciences*, vol. 103, no. 14, pp. 5367–5372, 2006. 86

- [21] D. Zhang, Y. Cao, Q. Ouyang, and Y. Tu, “The energy cost and optimal design for synchronization of coupled molecular oscillators,” *Nature Physics*, vol. 16, no. 1, pp. 95–100, 2020. 86
- [22] K. Yoshimoto, K. Arora, and C. L. Brooks III, “Hexameric Helicase Deconstructed: Interplay of Conformational Changes and Substrate Coupling,” *Biophysical Journal*, vol. 98, pp. 1449–1457, apr 2010. 86, 100
- [23] H. Sakaguchi, S. Shinomoto, and K. Yoshiki, “Mutual Entrainment in Oscillator Lattices with Nonvariational Type Interaction,” *Progress of Theoretical Physics*, vol. 79, no. 5, pp. 1069–1079, 1988. 87, 88, 100
- [24] J. A. Acebrón, L. L. Bonilla, C. J. Pérez Vicente, F. Ritort, and R. Spigler, “The Kuramoto model: A simple paradigm for synchronization phenomena,” *Rev. Mod. Phys.*, vol. 77, no. 1, pp. 137–185, 2005. 87
- [25] A. I. Brown and D. A. Sivak, “Theory of Nonequilibrium Free Energy Transduction by Molecular Machines,” *Chemical Reviews*, vol. 120, pp. 434–459, jan 2020. 87
- [26] J. Xing, H. Wang, and G. Oster, “From Continuum Fokker-Planck Models to Discrete Kinetic Models,” *Biophysical Journal*, vol. 89, no. 3, pp. 1551–1563, 2005. 87
- [27] F. Jülicher and J. Prost, “Spontaneous Oscillations of Collective Molecular Motors,” *Phys. Rev. Lett.*, vol. 78, pp. 4510–4513, jun 1997. 87
- [28] C. Doering, B. Ermentrout, and G. Oster, “Rotary DNA motors.,” *Biophysical journal*, vol. 69, pp. 2256–2267, dec 1995. 87
- [29] H. Wang and G. Oster, “Energy transduction in the F1 motor of ATP synthase,” *Nature*, vol. 396, no. 6708, pp. 279–282, 1998. 87
- [30] C. S. Peskin and G. Oster, “Coordinated hydrolysis explains the mechanical behavior of kinesin.,” *Biophysical journal*, vol. 68, pp. 202S–210S; discussion 210S–211S, apr 1995. 87
- [31] J. C. Bronski, T. Carty, and L. DeVille, “Configurational stability for the Kuramoto–Sakaguchi model,” *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 28, no. 10, 2018. 92

- [32] V. H. Mello, J. Wald, T. C. Marlovits, and P. Sartori, “Elastic analysis of structural ensemble reveals the energetic basis of hand-over-hand in a AAA+ motor,” *bioRxiv*, 2024. 94, 98
- [33] X. Fei, T. A. Bell, S. Jenni, B. M. Stinson, T. A. Baker, S. C. Harrison, and R. T. Sauer, “Structures of the ATP-fueled ClpXP proteolytic machine bound to protein substrate,” *eLife*, vol. 9, p. e52774, 2020. 97, 98
- [34] G. L. Hersch, R. E. Burton, D. N. Bolon, T. A. Baker, and R. T. Sauer, “Asymmetric Interactions of ATP with the AAA+ ClpX₆/ClpX₆ Unfoldase: Allosteric Control of a Protein Machine,” *Cell*, vol. 121, pp. 1017–1027, jul 2005. 97
- [35] S. A. Joshi, G. L. Hersch, T. A. Baker, and R. T. Sauer, “Communication between ClpX and ClpP during substrate processing and degradation,” *Nature Structural Molecular Biology*, vol. 11, no. 5, pp. 404–411, 2004. 97, 98
- [36] P. Wendler, S. Ciniawsky, M. Kock, and S. Kube, “Structure and function of the AAA+ nucleotide binding pocket,” *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, vol. 1823, no. 1, pp. 2–14, 2012. 98
- [37] G. Costantini and A. Puglisi, “Thermodynamic precision of a chain of motors: the difference between phase and noise correlation,” *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2024, no. 2, p. 24003, 2024. 100
- [38] J. Buceta, J. M. Parrondo, C. den Broeck, and F. J. de la Rubia, “Negative resistance and anomalous hysteresis in a collective molecular motor,” *Phys. Rev. E*, vol. 61, pp. 6287–6293, jun 2000. 100

Conclusions

In this final chapter, we review and connect the main contributions of the thesis. We began by developing a general framework to study the mechanical deformations of protein structures and gradually advanced towards a theoretical description of the emergent properties of molecular motor assemblies. Along this trajectory, we integrated concepts from elasticity theory, structural biology, enzyme kinetics, and stochastic thermodynamics, introducing new computational tools and models to elucidate the activity of hexameric ATPases. Overall, we established a unified framework that bridges structure and kinetics through energetics for these molecular motors.

6.1 Summary

In this thesis, we developed an energy-based multiscale framework to investigate the activity of protein complexes, with an emphasis on RuvB (chapters 3 and 4), a DNA-translocating hexameric motor. The approaches devised here drew inspiration from prior work, literature, and our own analyses, leading to four main contributions:

- In Chapter 2 we proposed a hyperelastic model, grounded in non-linear elasticity, to quantify elastic energy differences from proteins in different conformations. This model was then implemented into an open-source and publicly available package that can be applied to any pair of protein atomic coordinates. We further performed two types of validation, using synthetic and real data, showing both numerical accuracy and agreement with known energy magnitudes for the case of haemoglobin.
- In Chapter 3, we used the method discussed above to characterise the RuvB's mechanochemical cycle. First, we devised a quantitative approach to reconstruct order of conformations in the cycle, recovering with striking agreement the previously known order, which was based on intensive and qualitative visual inspections. We then characterised the energetics of RuvB's cycle at two scales: *global*, consisting of the elastic energy landscape across RuvB conformations; and *local*, which led us to identify candidate regions responsible for energy

transduction between different components of the system. Together, these results suggest that energy transduction across RuvB subunits is key to its motor activity.

- In Chapter 4, we used insights from our previous structural analyses to build a toy model for RuvB’s kinetics. We first showed that the energy landscape we characterised earlier is consistent with processivity and kinetically advantageous compared to randomised landscapes. Moreover, we further adjusted model simulations to fit both batch biochemical and single-molecule biophysical data, establishing a timescale for conformational transitions. Finally, our model predicted that RuvB operates near an optimal chemical driving, which simultaneously minimises dissipation while maximising processive speed.
- In Chapter 5, we generalised our previous model using a continuous conformation variable, which was better suited to describe the behaviour of a generic hexameric ATPase. We first showed that, by controlling the coupling between adjacent subunits, we obtained different emergent kinetic behaviours of hexameric ATPases, which we called *modes of operation*. We then characterised the emergent speed the motors could reach, showing that the most abundant mode observed in nature – the sequential mode – is the fastest in all tested scenarios. We then used published data from a motor believed to operate in the probabilistic mode, ClpX, to test whether our model would recover such classification. We identified that the best matching mode was contingent on assumptions about the structural details of this motor. Still, the highest-scoring scenario supports that ClpX operates in a sequential mode, contrary to the current consensus.

6.2 Conclusion

Altogether, our work can be seen from a two-fold perspective. First, we have provided new insights on the operation of RuvB, ClpX, and homohexameric ATPase motors in general. While inter-subunit coordination is known to be essential for motor function, here we show this through two completely unexplored approaches.

Using protein strain analysis, we identified structural elements likely involved in inter-subunit coupling. Leveraging the large structural ensemble of RuvB, comprising 30 cycle conformations, allowed us to detect patterns of elastic pseudoenergies oscillations throughout the cycle, indicative of mechanical coordination. Although such large conformational ensembles among hexameric motors are rare (to date), the number of resolved hexameric ATPase structures has grown considerably over the last

decade. These datasets could potentially offer a rich source of information to uncover commonalities and differences in terms of inter-subunit coupling, which can be further explored using the approach developed here.

Using kinetic models, we show that coupling is a central feature that was likely tuned through evolution to sustain coordination, a major contributor to the motor's speed. This provides a rationale for the seemingly prevalent sequential mode of operation, which we also proposed for ClpX as a possible mechanism, yet to be experimentally tested. As the amount of structural data grows fast, comparing structures manually becomes not only impractical but also hinders the understanding of global patterns across hexameric ATPases. The toy models developed here are a step further in this direction by offering a flexible framework that can easily incorporate the specificities of each system.

Second, we introduced simple yet useful integrative tools, which have allowed us to establish extensive links across fields. Overall, we have used a combination of multiple sources of information, such as protein structures, free-energy differences, catalytic activities, and processive speeds. While these sources of information are often confronted at the level of discussion, it is uncommon to find them combined in the same framework, aiming to answer a common question. Thus, the different methodological approaches presented throughout the thesis also represent an effort to overcome disciplinary boundaries that are currently intrinsic to the scientific communities studying proteins.



ITqb nova