In-silico Investigation of Ion-Pumping Rotary A- and V-type ATPases

Structural and Dynamical Aspects

Konstantinos Papachristos

Submitted in accordance with the requirements for the degree of

Doctor of Philosophy

The University of Leeds

School of Molecular and Cellular Biology

October 2014

The candidate confirms that the work submitted is his own, except where work which has formed part of jointly-authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others. Chapter 3 is based on the results and discussion presented in:

- *Song CF, Papachristos K, Rawson S, Huss M, Wieczorek H, Paci E, Trinick J, Harrison MA, Muench SP (2013) Flexibility within the rotor and stators of the vacuolar H^+-ATPase. PloS one 8: e82207*. Conceived and designed the experiments: CS KP SR MH HW EP JT MAH SPM. Performed the experiments: CS SR KP MH EP MAH SPM. Analyzed the data: CS KP SR MH HW EP JT MAH SPM. Contributed reagents/materials/analysis tools: CS KP MH HW EP MAH SPM. Wrote the manuscript: CS KP MH HW EP JT MAH SPM.

This page intentionally left blank

To the memory of my beloved brother,

Ανέστη

This page intentionally left blank

Acknowledgments

I am indebted to many people associated with the *Astbury Center for Structural Molecular Biology* for fruitful scientific discussions, including my PhD supervisors (Dr Emanuele Paci, Dr Stephen Muench and Prof Peter Knight) and assessors (Dr Sergei Krivov and Prof Arwen Pearson). Particularly, I would like to thank Dr Stephen Muench for introducing me into the exciting field of ion-pumping rotary ATPases, providing guidance and invaluable help during all the stages of my PhD research. Last but not least, I should thank my family, Athanasios Tachtsidis, Polina Banushkina, Alexandros Chatziagorakis, Andreas Chatzipantelis, James Gowdy, Diana Monteiro, Pinelopi Nikolaidou, Anna Polyakova and Gaël Radou for all the amazing moments we have spent together.

Abstract

Advances in *Molecular Biosciences* have revolutionised the way we perceive and pursue current biological research. Dynamic, complex biomacromolecules constitute the essential components of *Cells.* Particularly proteins have been characterised as the workhorse molecules of life. Either as single chains or complexes of associated units, proteins participate in every biological process with a specific structural and/or functional role. Ion-pumping rotary ATPases is a large family of important membrane-bound protein nanomachines. In the current work we investigate structural and dynamical aspects of the A- and V-type rotary ATPases, related to functional dynamics, and propose a multiscale computational framework for their *in-silico* biophysical characterisation and the interpretation of low-resolution experimental data from electron microscopy in **Chapter 3**. For the first time we present results from explicit-solvent atomistic molecular dynamics simulations of the prokaryotic A-type peripheral stator stalk and central rotor axle, both being critical subunits involved in the mechanical coupling of the rotary ATPases in **Chapter 4**. Our simulation data reveal the presence of flexibility heterogeneity and demonstrate the dynamic nature of the peripheral stator stalk as a source of intact ATPase particle conformational variability. In **Chapter 5** we show the presence of structural plasticity in the eukaryotic peripheral stator stalk of the V-ATPase and discuss possible implications for V-ATPase regulation. Overall, the wealth of information accessed with molecular-dynamics simulations allows the exploitation of atomistic information within the multiscale framework of **Chapter 3** to be applied for the mechanical characterisation of rotary ATPases in future studies. In particular, atomistic data could serve as high-resolution information for future parameterisation of simplified coarse-grain models for all ATPase subunits and the construction of molecular models for the intact ATPases. We anticipate that our approach will contribute to elucidating the molecular origin of rotary ATPases' conformational flexibility and its implications for the holoenzyme's function and kinetic efficiency.

Table Of Contents

List of Tables

List of Figures

Chapter 1

Ion-pumping Rotary ATPases: Structure and Function

1.1 Ion-pumping Rotary ATPases are Complex Biomolecular Nano-machines

Proteins are ubiquitous macromolecules that constitute the molecular workhorses of cells involved in every biological process. They exert their functional and structural roles due to their unique composition and structure that in general yield multi-state funnel-like folding and functional energy landscapes. Proteins, either as single molecules or supra-molecular assemblages of varying stoichiometry and architectural characteristics, can be analogised with man-made, macroscopic machines that carry out specific tasks. However, the great difference between macroscopic and protein machines is the space and time regime where proteins operate, that is the nanoscopic world and associated physics and chemistry. Moreover, the surrounding solvent molecules that continuously change position and velocity interact with protein molecules resulting in stochastic fluctuations and excursions on the protein folding/functional free-energy landscape. In other words, proteins are stochastic machines that participate in free-energy conversions from one form to another.

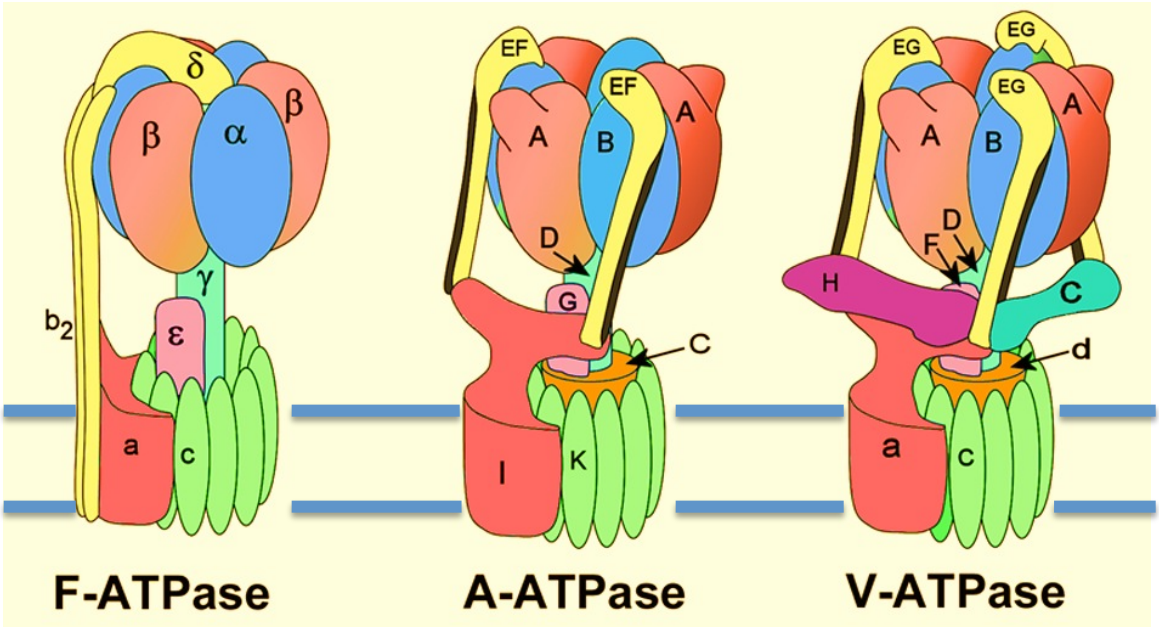


Figure 1.1 Schematic representation of the structural organisation of all three types of ion-pumping rotary ATPases. Nomenclature of the constituent subunits is adopted from (Muench et al, 2011).

An important class of *molecular machines* is that of molecular motors that interconvert (bio-) chemical energy into mechanical work and can be classified into linear and rotary motors depending on the directionality of the produced work. Examples of linear motors are myosins, kinesins and dyneins, while the best-characterised rotary motors is the *F_1F_0 ATP synthase*. Ion-pumping rotary ATPases are a family of large, integral-membrane bound protein nanomachines involved in various physiological processes and pathological conditions. Here we intend to give a short critical overview of our current understanding of rotary ATPases regarding their structure, dynamics and function. The interested reader is referred to some excellent review articles (Marshansky et al, 2014; Muench et al, 2011; Stewart et al, 2014).

All the members of the family share some common structural organisation: the holoenzymes consist of two rotary motors, a soluble chemical motor and an integral-membrane electrochemical one. The soluble chemical motor is a hexameric subcomplex that catalyses the hydrolysis or synthesis of adenosine triphosphate

(ATP) (subunits αβ of F-ATPase and AB of the A- & V-ATPases, **Figure 1.1**). The catalytic subunits form an internal cavity where a central rotor axle is rotated interacting specifically with the catalytic subunits (subunits γε of the F-ATPase, DFC of the A-ATPase and DFd of the V-ATPase, **Figure 1.1**). On the other hand, the integral-membrane electrochemical motor is a multi-subunit protein ring embedded in biological biomembranes that forms a *polar interface* with a second integral-membrane protein, which is part of the *stator* (subunits ac of the F-ATPase, IK of the A-ATPase and ac of the V-ATPase, **Figure 1.1**). Depending on the exact origin - organism and cell type -, each rotary ATPase is characterised by specific optimum conditions wherein the holoenzyme operates either in ATP hydrolysis or ATP synthesis mode. In the ATP-synthesis mode, the protein ring harvests the electrochemical-potential gradient across a biomembrane via the translocation of cations to produce torque through the central rotor axle to the catalytic subunits of the soluble chemical motor resulting in ATP synthesis. In the ATP-hydrolysis mode, catalytic subunits undergo tertiary and/or quaternary conformational changes due to ATP binding and hydrolysis exerting torque on the central rotor axle, which in turn biases the rotation of the membrane protein ring in the direction of cation translocation against an established electrochemical potential gradient across the biomembrane.

In terms of subunit stoichiometry and topological organisation, each ion-pumping rotary ATPase is classified into one of three types (see **Figure 1.1**):

1. *F*-A*TPases*: include bacterial and mitochondrial F_1F_0 ATP synthases
2. *A*-ATPases: include archeal and prokaryotic A_1A_0 ATPases/ synthases
3. *V*-ATPases: include the eukaryotic V_1V_0 ATPases

Structural elucidation of rotary ATPases has been achieved with a combination of high- and low-resolution structural biology techniques such as X-ray crystallography, NMR spectroscopy and electron microscopy (EM). In particular EM has permitted investigations of structural aspects of the intact holoenzyme, a task which is extremely challenging or currently impossible with X-ray crystallography or NMR. Combined with single-particle analysis and mathematical 3D reconstruction techniques, cryoEM has provided low-resolution electron density

maps for all types of rotary ATPases resolving some previous misconceptions. Overall, we can comment that structural/topological complexity increases from the F-type through the A-type to the V-type and that complexity coincides with the presence of complex regulatory mechanisms of assembly/disassembly of the V-ATPase. Topological complexity is reflected on three main structural features:

1. ***Central Rotor Axle***

   Both A- and V-ATPases have a cup-like subunit (subunit d) within the central rotor that forms a protein-protein interface with both the protein ring and the central rotor axle DF. Those interfaces are particularly polar with charged residues contributing to electrostatic stabilization. One of the consequences of the extra subunit within the rotor subcomplex is the *elongation* of the A- and V-ATPases compared to the F-ATPase. That elongation could result in increased conformational variability of the intact A/V-ATPase particles as opposed to the much more compact F-ATPase particles. The biophysical implications of the increased conformational flexibility are not clear. In addition, subunit d introduces an extra subunit-subunit interface with no characterised properties. For example, water molecules could be involved in the establishment of protein-protein contacts (even in the form of a water layer) with unknown effect on the properties of the interface; water molecules could act as a *glue* or *lubricant*. No matter what the properties of the protein-protein interface are, they must have an impact on the function and energy-conversion efficiency of rotary ATPases. For example, a *slippery* interface would result in a loss of efficiency since only part of the torque produced by one motor would be delivered to the other one.

2. ***Peripheral Stator Stalks***

   A characteristic structural feature of rotary ATPases is the number of copies of the peripheral stalks that attach to the catalytic subunits and couple the R_1-ATPase (where R=F, A, V) with the R_0 sector counteracting the torque produced and the rigid-body rotation of the catalytic subunits subcomplex. F-ATPases have one peripheral stalk ($b_2\delta$), A-ATPases have two peripheral

stalks (EF [1]) and V-ATPases (EG) have three peripheral stalks. One biophysical aspect to be clarified is how this stoichiometry was evolutionary adapted and what are the implications for ATPase flexibility and function. The transition from F- to A-type stoichiometry is likely related to conformational dynamics differences of the catalytic subunits between the F_1- and A_1-ATPases. The transition from A-type to V-type stoichiometry could be related to the presence of extra subunits (subunits C and H) with a regulatory role. In addition, it is unknown how the flexibility of the peripheral stalks compares among the different types of rotary ATPases. We could speculate that a decrease in the *stiffness/flexibility* of the peripheral stalks would be expected to allow conformational flexibility within the intact A- & V-ATPase particles.

3. ***Peripheral Stator Collar***

   The peripheral stator collar is a semicircular subcomplex where the peripheral stalks of A- and V-ATPases are attached to. Comparison between the A- and V-ATPases show increased complexity within the peripheral stator collar of V-ATPases with the presence of two subunits, C and H. Subunit H has been shown to form contacts with the central rotor axle DF, when the V-ATPase disassembles, preventing rotation and ATP consumption. Regarding subunit C, it has been hypothesised that the subunit is involved in the mechanism of V-ATPase disassembly, i.e., subunit C is under strain within the complex compared to its conformational flexibility in solution acting as an elastic energy buffering element that releases its energy upon receiving a disassembly signal. If such a mechanism exists, it is likely that the peripheral stalks are also implicated.

[1] In the rest of the Thesis we make no distinction between the A- and V-type homologous proteins and the V-type nomenclature is adopted.

1.2 Elastic Energy Transmission Between the two Constituent Motors

One important structural aspect of rotary ATPases with important functional implications is the asymmetry mismatch between the two motors; the chemical motor (R_1-ATPase, where R=F, A, V) has a pseudo-three-fold symmetry while the protein ring of the electromotor (R_0 sector, where R=F, A, V) is of variable stoichiometry. Especially for F_0F_1 ATP synthase, it has been suggested that some constituents of the holoenzyme behave as *elastic-energy buffering* components that transiently accumulate *strain free-energy* (Cherepanov et al, 1999). In subsequent stages of the rotary catalysis cycle, those subunits relax and the released conformational free-energy is harvested to complete the cycle with high kinetic efficiency. The elastic coupling between F_1 and F_0 is expected to smoothen the free energy landscape of the coupled enzyme increasing the kinetic efficiency. Following the theoretical analysis in (Panke et al, 2001) assuming that the central shaft subunit **γ** can be approximated as an elastic rod with r_γ radius and L_γ effective length, twisted by the angle θ_γ, the amount of conformational free-energy ΔF_γ accumulated is

$$\Delta F_\gamma = \frac{\pi R_\gamma^4 E_\gamma}{8 L_\gamma (1+\sigma)} \theta_\gamma^2 \qquad \textbf{(1.1)}$$

where E_γ is the longitudinal Young modulus of γ subunit, and σ the Poisson ratio. It was theoretically shown that the steady-state drift velocity is reduced when the stiffness of subunit γ increases. Therefore, a torsionally soft central shaft enhances the kinetic efficiency of the coupled enzyme. Further single-molecule experimental studies (Sielaff et al, 2008; Wachter et al, 2011) showed that the two motors in the F_0F_1 ATP synthase are coupled with a peripheral stator which is at least 10-fold stiffer that the central rotor shaft.

Further investigations are needed to elucidate the elastic energy transmission mechanism in F-ATPases and whether such a mechanism also exists in

the other two members of the rotary ATPase family, A- and V-ATPases. The latter can be achieved with a synergy between theory, computations and experimental studies.

Having summarised our current knowledge on the structure and function of the members of rotary ATPases –and particularly the A- and V-ATPases that are the focus of the present Thesis–, in the next Chapter we present the essential background on the computational techniques that have been developed for *in-silico* biophysical studies of proteins and protein complexes.

Chapter 2

Molecular Modelling and Simulation Approaches in Molecular Biophysics

2.1 Protein Mechanics and Dynamics

One of the long-term goals of Biochemistry and Molecular Biophysics is the establishment of the relationship between structure, dynamics and function of biomacromolecules, and particularly proteins. Crystallographic studies over the last decades have been instrumental in advancing our understanding of the microscopic determinants of protein organisation, function and regulation. However, the rather static picture of the protein universe emerging out from crystallographic investigations obscures the realisation of proteins as dynamic entities. Protein dynamical complexity as encoded in their free-energy landscape is important to function in the highly complex and dynamic environment of cells, adopting their structure to external signals or changes in temperature, osmolytes concentration, pH and direct mechanical force. Our understanding of the effects of the latter on protein dynamics and function has been advanced by the development of single-molecule force spectroscopy techniques, such as atomic force microscopy and optical/magnetic tweezers. These biophysical techniques have not only provided great insight into the microscopic basis of the mechanics of natural force-bearing proteins (e.g., mechanical unfolding of titin modules, Marszalec et al, 1999), but also triggered the development of new computational techniques contributing to the

interpretation of the experimental data, as well as complementing experimental studies as standalone *in-silico* techniques. Understanding protein mechanics is not only relevant to mechanical proteins, but all proteins in general. Recent studies pointed out to the correlation between protein mechanics and their importation inside mitochondria, or degradation by the proteasome, indicating that mechanical forces are exerted on proteins during all stages of protein life (Ingber, 1997; Ingber, 2006). Similarly, protein mechanics are related to the function of protein nanomachines, especially those operating as nanomotors, such as the F-ATP synthase.

From the above it is evident that studying protein mechanics advances our understanding of the biological effects of force on the function of biomolecular nanomachines. Recently, computational studies complementing single-molecule mechanical experiments have provided atomistic insight into the experimental observations. Molecular-dynamics simulations have already become a standard tool in simulating proteins under non-equilibrium conditions that mimic mechanical unfolding. Molecular systems are represented by a collection of classical particles that interact with each other through a potential function –known as *molecular-mechanics force field*- that captures the essential physicochemical interactions between atoms within and between molecules. Various functional forms for interactions may be employed but among all, pair-wise force fields with fixed-point atomic charges for electrostatic interactions are the most popular for biomolecular simulations such as the CHARMM (MacKerell et al, 1998), OPLS-AA (Jorgensen et al, 1996), AMBER (Cornell et al, 1995) and GROMOS (Christen et al, 2005) force fields.

In principle, all-atom protein models with explicit solvent can be employed to study protein systems with high spatial and time resolution. However, the tremendous computational cost of an explicitly solvated system limits the application of the technique to relatively small proteins, employing unrealistically large forces for mechanical unfolding compared to experiments to accelerate the simulation. Since solvent molecules constitute the largest part of a solvated system,

replacing those with an implicit model that accounts for the solvent environment (*implicit solvation*) can reduce the computational expense by several orders of magnitude, allowing the application of smaller forces for mechanical unfolding. Water molecules participate in hydrogen-bond interactions with the protein backbone, so their absence could hinder the actual microscopic mechanism. However, the implicit-solvent paradigm is not only computationally efficient but also free from solvent-relaxation artifacts and instabilities that may arise under large applied load (Paci & Karplus, 2000). Indeed, Paci and Karplus, using a Gaussian-like effective energy to account for solvent effects (Lazaridis & Karplus, 1999) and a polar-hydrogen atomistic model for protein (Brooks et al, 1983), investigated *in-silico* the mechanical unfolding mechanism of two β-sandwich proteins and two alpha proteins revealing significant differences in their mechanical response both within the same class of proteins and between different classes due to differences in topology and energetics.

Despite their success in protein mechanics applications, implicit-solvent simulations of large protein assemblies and nanomachines are still a challenging computational task. Drastic approximations of protein chemistry are introduced resulting in simplified coarse-grained models, with or without explicit solvent. Apart from being computationally efficient, coarse-grained models also represent coarse-grained realisations of our understanding of biomolecular physical chemistry.

2.1.1 Structure-based Coarse-Grain Modelling

The development and use of simplified protein models trace back to the first simulations of protein folding and dynamics in the seminal work of Levitt and Warshel (Levitt & Warshel, 1975). Since those first computer simulations, tremendous advances have been made in coarse-grain (CG) modelling of proteins and other biomolecules with several applications in Molecular Biosciences. In this section, two popular protein CG models are described, i.e., *Gō*-like and *Elastic*

Network Models (ENM). Despite their limitations due to the underlying assumptions, both models permit computer investigations of biomolecular systems with reasonable resources and allow comparison with experimental data.

2.1.2 Gō-like Coarse-Grained Protein Models

Gō-like models is an umbrella term encompassing explicit-chain polymer models for proteins where a reference experimental structure (the so-called native structure) serves as a bias towards the native configuration of the protein. The details of the actual model, i.e., the degree of coarse-graining and the potential energy functional form, may vary from model to model, all of them having though a common characteristic feature: only native residue contacts present in the reference structure contribute favorably in the system Hamiltonian, while non-native contacts are less favorable, usually modeled as repulsive forces between coarse-grain particles. Despite their simplicity and the apparent artificial character of the configurational bias towards a particular structure, Gō-like models partly obtain their legitimacy from the fact that evolutionary pressure has led proteins to adopt a minimally frustrated funnel-like energy landscape contrary to random hetero-polymers. Native protein structures lay on the basin of such an energy landscape, so that structure-based coarse-grain Go-like models constitute a parameterisation of that particular part of the energy landscape. The original Gō model (Gō & Taketomi, 1979a; Gō & Taketomi, 1979b) was developed for simulating protein folding dynamics adopting a native structure-centered potential for a lattice-based one-bed protein polymer, while the configurational space was sampled using the standard Metropolis Monte Carlo algorithm. Many extensions of the original Gō model have been developed. The problem of insufficient cooperativity of Gō-like models has been partly addressed by the inclusion of extra energetic terms in the model Hamiltonian, accounting for effective desolvation barriers (Cheung et al, 2002; Kaya & Chan, 2003) and sequence effects on protein folding (Karanicolas & Brooks, 2002). An example of a versatile Gō-like model with finite extensible nonlinear

elastic energetics terms for backbone chain connectivity is the *Self-Organised Polymer* (SOP) model introduced by Thirumalai and coworkers (Hyeon et al, 2006a) and successfully applied to the *in-silico* mechanical folding/unfolding of proteins and the dynamical simulation of allosteric transitions in ATP-dependent nanomachines, like chaperonin GroEL-GroES (Hyeon et al, 2006b) and myosin (Tehver & Thirumalai, 2010), while another nano-machine, namely hexameric helicase interacting with DNA double-helix was recently studied (Yoshimoto et al, 2010) by means of a Gō-like model for both the proteins and nucleic acid following (Hoang & Cieplak, 2000). Wu et al (Wu et al, 2008) and Onuchic and coworkers (Whitford et al, 2009) introduced all-atom Gō-like models to bridge the gap between CG models and all-atom empirical force fields.

2.1.3 Elastic Network Coarse-grain Protein Models

Elastic Network Models (ENM), a popular class of coarse-grain models for proteins, consider the folded state of proteins as an elastic body, where a set of nodes are interconnected by interacting with simplified potentials, giving rise to an elastic network (Bahar et al, 1997) following the ideas of Tirion (Tirion, 1996). In ENMs, the coarse-graining stage consists of mapping the nodes of the network onto protein's structural elements. On the basis of residue-based coarse-graining, there is one node corresponding to each C_α carbon atom. A cut-off distance is used to assign interacting nodes and a simple harmonic potential is applied between interacting nodes (Gaussian Network Model, GNM). In an important extension of the simple GNM, anisotropic residual fluctuations are taken into account resulting in the Anisotropic Network Model (Atilgan et al, 2001; Doruker et al, 2000), ANM.

Combined with normal mode analysis, the ENM's success in describing functionally relevant low-frequency motions has led to their application in studying functional dynamics of allosteric proteins and protein nanomachines. ENMs are a type of highly restricted CG models without even explicit representation of the

protein backbone. ENMs are non-transferable and therefore need to be re-parameterised for different proteins. In that sense, ENMs are an example of a multi-scale modeling approach, since the data used for training the model comes from either fine-grain molecular simulations or experimental data (X-ray/NMR), with the corresponding residual correlations being used to infer inter-residue elastic force constants. While standard ENMs can capture global rearrangements, high-frequency modes cannot be predicted accurately (Petrone & Pande, 2006). Such residual cross-correlations could lead to more accurate ENMs, explored in the work of Lezon and Bahar (Lezon & Bahar, 2010) who applied an entropy-maximization methodology to parameterise an ENM from a set of experimental NMR data for 68 proteins. More details about the ENM-NMA theory and applications can be found in (Bahar et al, 2010; Bahar & Rader, 2005).

2.1.4 Multistate CG protein models

Standard Gō-like and Elastic Network Models describe a single basin of attraction within the folding landscape of a protein. However, proteins may adopt different conformations to carry out their function, each corresponding to a different landscape basin. In order to study the conformational transition of calmodulin, a Ca^{2+} signaling protein, Zuckerman (Zuckerman, 2004) developed a double-basin united-atom CG model where both two-end conformations where incorporated into the system Hamiltonian. The implementation of a high-resolution lattice Monte Carlo algorithm allowed an ensemble of conformational transitions of calmodulin to be examined. Hummer and coworkers (Best et al, 2005) used a similar approach to simulate the helix-to-sheet transition in Arc Repressor NL11 protein. At the same time, Maragakis and Karplus (Maragakis & Karplus, 2005) introduced a multistate Elastic Network Hamiltonian, termed Plastic Network Model (PNM), and applied their model to the well-studied allosteric protein, adenylate kinase, identifying the minimum-energy transition pathway between the open and closed states. PNM was

also applied to probe the torque generation and transmission during the function of a molecular nano-machine (Pu & Karplus, 2008), the F_1-ATPase/synthase, ATP hydrolysis operation mode, re-producing the two $80^o/40^o$ rotation sub-steps of the total 120^o step in accord with experimental observations. F_1-ATPase dynamics had been previously studied by Takada and coworkers (Koga & Takada, 2006) with a switching Gō-like model, reproducing basic features of the F-ATPase rotation dynamics.

2.1.5 Mesoscopic Continuum Protein Models

Continuum mechanics (CM) is another modeling approach, particularly useful for large protein assemblies. Here a material field is considered as the continuum analog of a particulate elastic-network model, where a mass field and the elastic moduli replace the particles, connectivity and elastic force constants of the network. Although such a replacement leads to a system of infinite degrees of freedom (DOF), the assumptions of an elastically homogeneous protein and/or the presence of a finite number of elastically homogeneous regions within the protein effectively reduce the DOF, resulting in a coarse-grain model. Continuum models are quite appealling for biomolecular simulations when combined with established *Finite Element Method* for numerical solution of elliptic partial differential equations and can be easily integrated with Poisson-Boltzmann equation for continuum electrostatics and Stokes equation for hydrodynamic damping. Bathe (Bathe, 2008) explored the applicability of continuum-mechanics models with the finite-element method and normal-mode analysis on a set of proteins, assuming isotropic elastic stiffness and that the volume of the protein is delimited by the solvent-excluded surface. Cui and coworkers developed and applied a continuum-mechanics framework for studying mechano-sensitive membrane proteins in a series of papers (Chen et al, 2008; Tang et al, 2006; Tang et al, 2008a), reviewed in (Tang et al, 2008b). In (Ma et al, 2009), a computational framework for studying the mechanical

properties of macromolecules was developed where a continuum model was employed for modelling biomacromolecules while a continuum electrostatics model based on Debye-Hueckel theory of electrolyte solutions was used to treat solvation effects. Finally, CM is well suited for describing mechanical properties of large biomolecular assemblies and making connections with experimental results. In particular, during the last years a wealth of experimental data of nano-indentation of viral capsids has emerged. Coarse-grain computational studies have sought to simulate experiments probing the mechanical properties of capsids and the effects of mutations and capsid maturation. Gibbons and Klug (Gibbons & Klug, 2008), motivated by nano-indentation experiments of cowpea chlorotic mottle virus capsids, performed CM/FEM-based nano-indentation simulations within the framework of nonlinear elasticity theory allowing the assessment of the impact of geometric heterogeneity on softening between native and swollen states of the capsid and predicting an enhanced nonlinear character of the mechanical response upon the native-to-swallen transition.

2.1.6 Multiscale Modelling and Coarse-graining

Coarse-grain models suffer from several limitations, for example the lack of rigorousness in terms of atom-to-bead mapping and model parameterisation. Moreover CG models are not appropriate for some computational investigations where either atomistic resolution is necessary or coarser resolution is needed to study complex and large biomolecular systems. A more systematic computational framework is needed where coarse-graining of a system is performed at the desired level of resolution. Many multiscale methodologies have been developed (Voth, 2009), which are either sequential or concurrent simulation methods (Ayton et al, 2007). The most widely used are the sequential methods where several models of varying resolution do not interact with each other directly; rather one model serves as a template (*reference model*) upon which the other level is built (*training model*).

The reference model can be either a more detailed model (e.g., atomistic) or a thermodynamics-based coarse-grain one. As an example of the application of multi-scale methods in protein assemblies, Chu and Voth, based on large-scale explicit-solvent atomistic simulations of the F-actin (Chu & Voth, 2005), derived a coarse-grain model for actin monomers at the subdomain level (Chu & Voth, 2006) within the fluctuation-matching MS-CG framework. Gohlke and Thorpe proposed coarse-graining biomolecules on the basis of rigid clusters identified with a graph-theoretic approach (Gohlke & Thorpe, 2006; Jacobs & Thorpe, 1995). Schulten and co-workers used an artificial neural network algorithm (Martinetz & Schulten, 1994) to define coarse-grain sites that preserve the shape of proteins (Arkhipov et al, 2006a), while Voth and coworkers proposed an essential-dynamics coarse-graining method based on molecular dynamics trajectories (Zhang et al, 2008a) or alternatively low-frequency modes from ENM-NMA to derive low-resolution CG models (Zhang et al, 2009).

On the other hand, following a top-bottom approach, efforts to build a transferable coarse-grain model reproducing the free-energy of partition between polar and apolar phases of a set of chemical compounds resulted in the so-called MARTINI force field for lipids (Marrink et al, 2007) with explicit coarse-grain water beads, further extended later for protein (Periole & Marrink, 2013) and carbohydrate (Lopez et al, 2009) modelling and simulations. The MARTINI model allows efficient molecular-dynamics simulations of larger systems and timescales than those accessible with all-atom MD, although the exact connection between AA-MD and CG-MD timescales is not clear, and the rescaling is empirical. However, MARTINI model cannot accurately reproduce the protein backbone energetics, and restraints are applied to maintain the secondary structure of proteins. Various membrane proteins have already been studied by means of MARTINI-based MD, such as the mechanosensitive membrane channel in lipid bilayers (Yefimov et al, 2008) or pressurized liposomes (Louhivuori et al, 2010). A hybrid structure-based ENM/thermodynamics-based CG (MARTINI) was investigated (Periole et al, 2009) and shown to reproduce residual fluctuations, protein deformations and large-

amplitude collective motions of proteins and protein assemblies in accord to atomistic simulations, while Stansfeld and Sansom (Stansfeld & Sansom, 2011) developed and assessed a protocol for serial CG/AA simulations of membrane proteins.

2.2 The use of Electron Microscopy to Study Large Protein Complexes.

Understanding biological phenomena at the atomistic level is both of fundamental scientific and practical interest, e.g., for rational lead optimization and drug discovery. Many key biochemical processes involve interactions between various biomolecules, giving rise to the formation of complex biomolecular machines. Commonly used structural-biology techniques such as X-ray crystallography and NMR spectroscopy are not well suited for structural interrogation of large complexes; NMR is presently applicable to relatively small proteins, while X-ray Crystallography can obtain high-resolution structures of large complexes, but the crystallization process in highly challenging, the obtained crystals are not always of good-quality and/or homogeneous, and it may be even impossible to obtain crystals of the protein in a specific functional state. A powerful and versatile technique that can provide important structural information and insight into the functional mechanism of biomolecular nanomachines is *Electron Microscopy* (EM), and particularly single-particle cryo-electron microscopy (cryo-EM) (Frank, 2006; Frank & Spahn, 2006). In single-particle cryoEM, the sample is rapidly frozen in a native state and 2D projection images are collected in the microscope, where molecules of the complex are captured in various (random) orientations. Assuming a unique conformational state within the ensemble of particles, the 2D images are processed mathematically (in Fourier Space) and a final 3D map in real space is obtained at a particular resolution- routinely between 10-20 A with several examples of symmetrical complexes in the subnanometric regime. The electron density map

captures closely the complex mass distribution in real space. The resulting map is of low resolution-compared to those from X-ray crystallography-, which practically means that atomic positions are not resolved, except for the overall shape of the complex and the corresponding mass distribution. To obtain a near-atomistic model from the cryo-EM maps, the raw density data must be combined with prior structural information for the subunits and dedicated computational approaches are employed to deduce detailed models the macromolecular complex. The structural interpretation of low-resolution map can be generally expressed as an optimisation problem, where the optimised quantity is the agreement between an atomistic model and the cryoEM map. Several approaches have been developed and applied to different biophysical systems. The optimisation problem is outlined as follows:

1. **Structural data are used to initiate the refinement procedure driven by the (cryo-)EM map**

1. **A scoring function is employed to evaluate the instantaneous probe structure**

2. **A propagator produces new probe structures, each compared against the map and evaluated; every probe structure production stage is considered as one refinement cycle**

3. **When a criterion is fulfilled, the refinement procedure is terminated and the final model is accepted**

It is also possible to produce an (weighted) ensemble of structural models to interpret the experimental density map. For an interesting application to *small-angle X-ray scattering* (SAXS) data, the reader is referred to (Rozycki et al, 2011).

2.2.1 Initial Structure and Scoring Function

The combination of an atomistic - knowledge-based or physics-based - force field with an experimental density map cannot unambiguously lead to a unique and/or accurate structure for a protein (*over-fitting problem*). The funnel-like, locally frustrated protein energy landscape requires extensive conformational sampling. Even with the currently most efficient sampling algorithms, the computational cost can be considerably large. So, it is necessary to incorporate prior structural knowledge with the form of high-resolution structural data from X-ray crystallography/NMR experiments. When structural data is missing, comparative modelling can provide approximate structural models (Marti-Renom et al, 2000).

To quantify the relationship between a probe structure ("model") and an EM map, an appropriate measure of similarity must be used to assess the 'fitness' of the model by comparing the experimental density map with a theoretical one of the same resolution *d*. Experimental maps are represented as scalar fields on a real-space grid, so the synthetic density map value $\rho^{S}_{i,j,k}$ on the voxel (*i*, *j*, *k*) is calculated as

$$\rho^{S}(i,j,k) = \sum_{n=1}^{N} \int_{V_{ijk}} dxdydzg(x,y,z,x_n,y_n,z_n) \tag{2.1}$$

where *j* is the *j*th atom of the model at position r_j, Z_j is its atomic number and *g* a weighting Gaussian function positioned at the atomic position r_j depending on its atomic number. The resolution, *d*, equals 2.353σ using the half-maximum criterion. Having calculated the theoretical map, the *similarity* between the two maps is given from the correlation coefficient, *CC*

$$CC = \frac{\sum_{ijk} \rho^{S}(i,j,k)\rho^{E}(i,j,k)}{\sum_{ijk}[\rho^{S}(i,j,k)]^2 \sum_{ijk}[\rho^{E}(i,j,k)]^2} \tag{2.2}$$

The quasi-atomistic interpretation of low-resolution density maps is usually

expressed mathematically as an optimisation problem. One of the most crucial aspects of optimisation is the employed *cost function* to be optimised, since the latter dictates the accuracy of resulting model(s). The experimental density map, ρ^{E}, is inserted in the scoring function along with other terms accounting for several stereochemical and physicochemical restraints. As discussed below, there is a class of computational approaches where the conformational sampling is carried out using MD simulations with empirical FFs. In that case, the computational method represents a specific case of biased molecular-dynamics simulations, rather than a genuine optimisation problem. The bias (or prior-knowledge) from the experimental density map is inserted into the system Hamiltonian either directly (e.g., as an external field) or indirectly through the CC (or other similarity measure I general) between the experimental and theoretical density maps.

2.2.2 Propagators

2.2.2.1 Rigid Docking

The simplest approach to interpreting low-resolution density maps is by docking available structural models for constituent subunits directly into the density map without considering any internal conformational degrees of freedom for the subunits except for the translational and rotational ones, i.e., subunits are treated as rigid bodies. The fitting process is quite straightforward; it can even be performed manually and in cases where the subunits adopt the same overall conformation can be very successful. Several algorithms have been developed (Baker & Cheng, 1996; Chacon & Wriggers, 2002; Roseman, 2000; Rossmann et al, 2001), most of them based on combinatorial optimisation algorithms and stochastic space-search algorithms to find the correct protein-protein interfaces and localisation inside the density map.

One of the most successful and popular packages for rigid docking is the SITUS (Wriggers et al, 1999) software package, developed by Wriggers and

coworkers and serves as the gold standard for newly developed rigid docking algorithms. One interesting methodology, outperforming SITUS, has been implemented in the CHARMM biomolecular simulation package (EMAP module) (Brooks et al, 2009). The methodology makes use of a Monte-Carlo grid-based searching protocol to sample real-space position/orientation of the protein subunit(s). The algorithm is both faster and more accurate compared to SITUS due to the use of a core-weighted correlation function as the similarity measure between the probe structure and the density map.

Although rigid docking drastically reduces the DOF of the optimisation problem, it lacks high accuracy, and the outcome of that analysis should be used as a starting point for more sophisticated computational techniques, known as *flexible fitting* techniques. Since proteins are dynamic entities, their intrinsic flexibility cannot be ignored if accurate structural models are required.

2.2.2.2 Normal Mode-based Fitting

The use of *Normal Mode Analysis* (NMA) (Tama et al, 2004a) is practical and successful methodology to incorporating protein conformational changes into the fitting procedure. NMA is based on a harmonic approximation of the potential energy and is commonly used to study protein essential-dynamics (Rueda et al, 2007). That provides a means to access large-scale conformation changes, otherwise inaccessible by brute-force techniques, and interrogate mechanistic aspects of protein functional dynamics. NMA is also used in X-ray crystallography refinement (Kidera & Go, 1990; Kidera et al, 1992a; Kidera et al, 1992b)

During the refinement stage, the objective goal is to maximise the real-space cross-correlation (*CC*) between the experimental EM density map and the synthetic one based on the sampled protein conformations. For each conformation, normal modes and the gradient of the *CC* with respect to the modes q_l are calculated. The

absolute value of the gradient indicates the magnitude of the effect of the mode displacement on the *CC* value. The size of the mode displacement is proportional to the gradient, $F_l = \frac{\partial CC}{\partial q_l}$,

$$q_l = \lambda F_l \tag{2.3}$$

where the scalar, λ, is chosen so that the maximum displacement of any atom during an iteration cycle is less than 2 Å. The modes are ranked according to their corresponding *CC* gradient and the five most significant ones are used to distort the protein structure. If the *CC* is reduced, the new structure is rejected, otherwise accepted and a new iteration cycle is initiated. If the structure is rejected, the amplitude is decreased and a new attempt is performed.

The normal mode flexible fitting (NMFF) method shows very good performance in terms of accuracy and timing. In addition to all-atom normal mode analysis, an elastic network model (ENM) based only on C_α atoms shows similar results to the all-atom model. That makes ENM-NMFF a very efficient technique applicable to large complexes, which are the usual systems studied with EM. NMFF has been applied to real experimental maps of the Elongation Factor bound to the ribosome, Cowpea chlorotic mottle virus (CCMV) and *E. Coli* RNA polymerase using only C_α atoms (Tama et al, 2004b).

2.2.2.3 Dynamics-based Approaches

This class of methods treats the protein as a dynamic entity simulating its temporal evolution, biased or restrained by the experimental data (EM, FRET, NMR). Various approaches have been presented and implemented. One of those is the *Molecular Dynamics Flexible Fitting* (MDFF) introduced by Klaus Schulten and coworkers in (Trabuco et al, 2008). MDFF is based on molecular-dynamics simulations performed

in vacuo or implicit/explicit solvent, where the experimental map serves as an external field that drives the system from areas of low-density towards areas of larger density. To avoid extensive local conformational distortion and over-fitting, harmonic restraints are usually applied at the secondary-structure level. The total potential energy of the system reads as

$$U = U_{MD} + U_{EM} + U_{SS} \tag{2.4}$$

where U_{MD} is the protein force-field potential, U_{EM} the potential derived from the presence of the density map and U_{SS} the restraint potential applied to maintain the stereochemistry of the protein. The map potential energy U_{EM} takes the form

$$U_{EM}(\boldsymbol{R}) = \sum_j w_j V_{EM}(\boldsymbol{r}_j) \tag{2.5}$$

where **R** is the coordinates of the atoms and

$$V_{EM}(\boldsymbol{r}) = \begin{cases} \xi \left[1 - \dfrac{\Phi(\boldsymbol{r}) - \Phi_{\mathrm{thr}}}{\Phi_{\max} - \Phi_{\mathrm{thr}}}\right] & \text{if } \Phi(\boldsymbol{r}) \geq \Phi_{\mathrm{thr}} \\ \xi & \text{if } \Phi(\boldsymbol{r}) < \Phi_{\mathrm{thr}} \end{cases} \tag{2.6}$$

where $\Phi(\mathbf{r})$ is the map density, $\Phi_{\max}$ is the maximum value of the density, Φ_{thr} is a threshold density value used to account only for the contribution of the biomolecule to the map, ξ is a scaling positive value and w_{j} is a weight factor specific to each atom type and usually set equal to the atomic mass.

The presence of the U_{EM} term in the system Hamiltonian leads to the exertion of an external force on each atom given by the equation

$$\boldsymbol{F}_j^{\mathrm{EM}} = -\nabla_j U_{EM}(\boldsymbol{R}) = -w_j \nabla_j V_{EM}(\boldsymbol{R}) \tag{2.7}$$

The method as applied to various synthetic maps of proteins in different conformational states gives similar results with other methods, e.g., Flex-EM (Topf et al, 2008). The main feature of MDFF is that conformational sampling is not driven by the cross-correlation but the protein is steered by the density map instead. The

latter feature makes MDFF a biased MD scheme similar to TMD. The main drawback of the approach is that the presence of the map-derived external potential can cause severe distortions of the secondary structure of proteins giving rise to unrealistic models. Imposing harmonic restraints can overcome those problems, but their choice is rather arbitrary and may prevent biophysically important conformational changes. Therefore, the application of MDFF or other similar techniques is not always straightforward, and evaluation of the resulting models is needed.

Further studies used a similar MD-based method with a modified potential (Orzechowski & Tama, 2008). The substantive difference lies in the biasing potential included in the Hamiltonian; instead of a local density-based potential, a global correlation-based potential, U^{Fit} was used. The total potential energy, U, has two terms, the biasing potential U^{Fit} and the physical molecular mechanics potential U^{ff}. The extra term U^{Fit} is proportial to $1-CC$. The proportionality factor, k, is an empirical parameter that tunes the effect of the effective potential U^{Fit}. No restraints were applied and at least for the proteins studied, the method was successful. One of the conclusions of the study was that the choice of the form for the effective biasing potential is important and depending on the case some particular form may be preferential to others.

Finally, Tama and co-workers (Grubisic et al, 2010) tested the application of simplified coarse-grained models in MD-based EM fitting. The use of coarse-grained model for biomacromolecules is very appealing for many reasons. First, the total degrees of freedom of the model are reduced, which implies reduced computational cost and a way to overcome over-fitting problems. Secondly, coarse-grained models can retain the flexibility of the protein and stereochemistry, so that applying atomic reconstruction techniques, atomistic models can be recovered. Tama et. al. applied a well-studied Gō-like coarse grained model (Clementi, 2000), which represent each residue as a simple bead centered at their C_α positions. The Gō -potential considers only native attractive non-bonded interactions, while the presence of bond, angle and non-local dihedral energy terms between consecutive C_α atoms satisfy

stereochemical constraints. Using the open-close conformations of adenylate kinase, it was demonstrated that in terms of computational time, CGMD optimization is much faster than when using an all-atom protein model. However, the NMFF approach proves to be the most efficient computationally. Still, CGMD is more accurate than NMFF, which can provide only large-scale conformational changes. Those results show that a combined approach of NMFF followed by CGMD/AAMD method could be promising and efficient for obtaining quasi-atomistic models from experimental (cryo)EM maps.

At this point, we have provided a dense overview of the current state-of-the art computational methodologies that allow *in-silico* biophysical investigations of large biomolecular systems either as standalone techniques or combined with low-resolution experimental data for the elucidation of structural and dynamical aspects of biomolecular nanomachines. In the next Chapters, we make use of some of the above mentioned methodologies in investigating the inherent flexibility within the A- & V-ATPase complexes.

Chapter 3

Flexibility Within the Rotor and Stators of the Vacuolar H^+-ATPase[2]

3.1 Introduction

In the V-ATPase family, the free energy of ATP hydrolysis is used to move protons across biological membranes against an electrochemical potential gradient (Yoshida et al, 2001) using a rotational mechanism (Imamura et al, 2003; Noji et al, 1997). As already discussed, this linkage involves a symmetry mismatch where a 3-stroke motor in the V_1 domain is linked to a c-ring in V_o that can vary in size. It has been suggested that energy-storing elastic elements within the rotary ATPase include the central rotor axle and/or the peripheral stator. Crystal structures of F-ATPase have suggested the presence of conformational flexibility, as has normal mode analysis of the stators of the A-ATPase but no direct evidence had been provided so far. In the current chapter, electron-microscopy 2D classes for the yeast and *Manduca sexta* V-ATPases, which clearly show conformational variability within the V-ATPase

[2] The current chapter is based on the results and discussion presented in *Song CF, Papachristos K, Rawson S, Huss M, Wieczorek H, Paci E, Trinick J, Harrison MA, Muench SP (2013) Flexibility within the rotor and stators of the vacuolar H+-ATPase. PloS one* ***8:*** *e82207*. Conceived and designed the experiments: CS KP SR MH HW EP JT MAH SPM. Performed the experiments: CS SR KP MH EP MAH SPM. Analyzed the data: CS KP SR MH HW EP JT MAH SPM. Contributed reagents/materials/analysis tools: CS KP MH HW EP MAH SPM. Wrote the manuscript: CS KP MH HW EP JT MAH SPM.

particles for the first time, are presented and discussed under the light of normal-mode-analysis calculations for the intact V-ATPase complex using a coarse-grain particle-based model. Then, calculations are carried out for the intact A-ATPase complex, the evolutionary relative of the V-ATPase with distinctive topological features, and the results are compared to those for the eukaryotic V-ATPase. Finally, we give an overview of the factors that in principle are expected to dictate the observed conformational flexibility of rotary ATPase particles, the importance of flexibility to function and we suggest a hybrid computational/experimental framework that can be employed to tackle the above mentioned biophysical questions.

3.2 Materials and Methods

3.2.1 Electron microscopy and image processing

Electron microscopy studies were carried out by the group of Dr S.P. Muench at the University of Leeds. In brief, negative stain electron microscopy images of both yeast and *M. sexta* V-ATPase were taken on a Jeol 1200EX microscope. This data was then processed such that particles belonging to a certain projection of the V-ATPase (angular view) were reclassified to look for flexibility within the complex. Data were also collected in the presence of ATP and a cryo-EM data set was also analysed. Full details are given in Song et al., 2013.

3.2.2 Coarse-grain Modelling of the intact V-ATPase Particle

A coarse-grained (CG) representation of the ATPase complex was generated using the 3D cryoEM map of intact *M. Sexta* V-type ATPase (Muench et al, 2009b). The topology-preserving algorithm (Pogoryelov et al, 2007; Wriggers et al, 1998) as implemented in the SCULPTOR (Birmanns et al, 2011) visualization software, was

then used to construct a pseudo-particle model consisting of 250 beads. The subsequent CG model was used as a scaffold for a Bend-Twist-Stretch Elastic Rod Network Model (ERNM) (Stember & Wriggers, 2009a). Application of standard Elastic Network Models (ENMs) using coarse-grained representations of EM 3D maps has previously been reported as a means to investigating bio-macromolecular dynamics [52,53] and assist the EM 3D reconstruction process (Brink et al, 2004). However, drastic CG is not expected to give accurate EN models, and eigenvectors calculated with these models should be considered more as a reduced-space sub-basis for conformational description and/or sampling. On the other hand ERNM incorporates elements from basic linear-elasticity theory and provides a more realistic description of protein flexibility. The Python-based script suite MODEHUNTER (http://modehunter.biomachina.org/) was used to calculate the hessian matrix and corresponding eigenvectors. The first two non-trivial resulting eigenvectors, were used to visualize the intrinsic modes of the intact ATPase. 3D pseudo-maps were reconstructed from various structures along the first two eigenvectors. Due to the synthetic pseudo-maps being noisy, a Gaussian filter was used to smooth them using CHIMERA (Pettersen et al, 2004). These models along with the V-ATPase reconstruction in a "non-deformed" state were re-projected into 50 different views. The re-projections from the three models were then combined into one file and low pass filtered to 35 Å.

3.3 Results

3.3.1 Conformational Variability in the Negative-Stain Single-Particle Dataset

The internal flexibility of the V-ATPase was studied by electron microscopy with each single-particle image being aligned to the masked V_O domain and then classified using a mask covering only V_1 (**Figure 3.1A**). The resulting classes showed deviation from the axle-c-ring co-axis, consistent with flexing of the linkages between V_1 and V_O (**Figures 3.1B, C**). The data was then reprocessed in the opposite

way, with the image stack belonging to each projection being aligned to V_1 and with V_0 used for classification. This resulted in distinctive classes showing V_o being displaced from the central (**Figure 3.1D**). The above image-analysis was carried out on both the yeast and *M. sexta* V-ATPase revealing similar flexing behaviour.

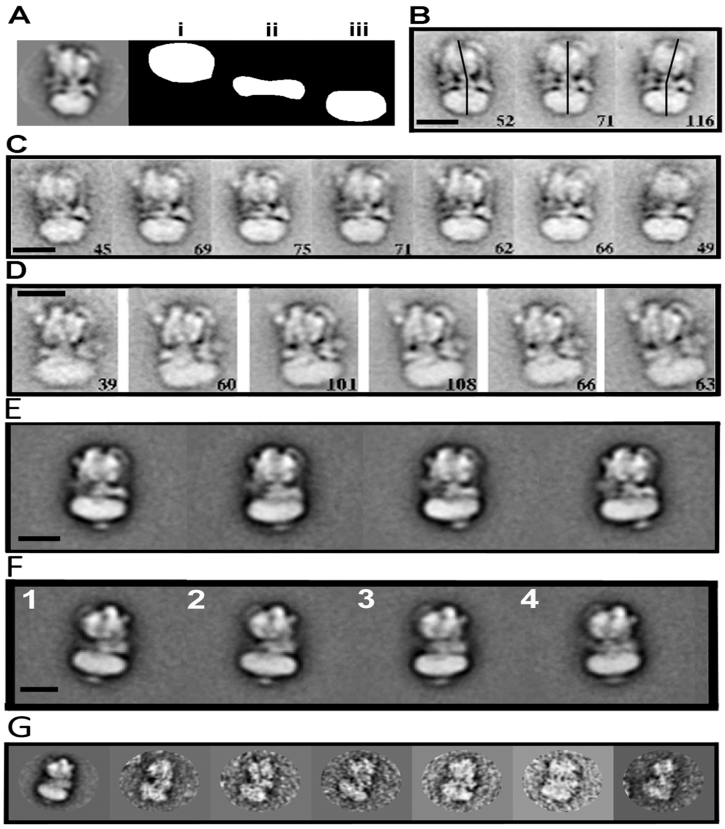


Figure 3.1 Negative stain electron microscopy of the *yeast* and *M. sexta* V-ATPases. (A) A representative class of yeast V-ATPase alongside the 3 masks used to extract the V_1 (i), central (ii) and V_0 domains (iii). (B, C) Yeast V-ATPase classes of particles belonging to the same orientation, as determined from multi-reference alignment, re-aligned to V_o and classified using a mask over V_1. (D) Yeast V-ATPase particles of particular views aligned against V_1 and classified by V_0. Numbers in the bottom right corner of B-D, are particle numbers in each class. (E, F) *M. sexta* V-ATPase classes of particles belonging to the same orientation and aligned against V_1 and classified around V_0. (G) *M. sexta* V-ATPase class (far left) and representative views of some of the particles making up the class. In all cases scale bars represent 120 Å.

3.3.2 Conformational Variability in the cryo-EM Single-Particle Dataset

A well-known shortcoming of negative-stain electron microscopy is the possible introduction of artefacts, such as flattening and distortion of particles. To remove this possibility the cryo-EM data for native state *M. sexta* V-ATPase in vitreous ice were also analysed. The proportion of side view particles showing flexing (721 of 3718 (~19%)) is slightly lower than that of the negative stain data (25%) but both negative stain and cryo-EM analysis shows flexing to a maximum of ~25° along the long axis of the complex (**Figure 3.2A, left panel**).

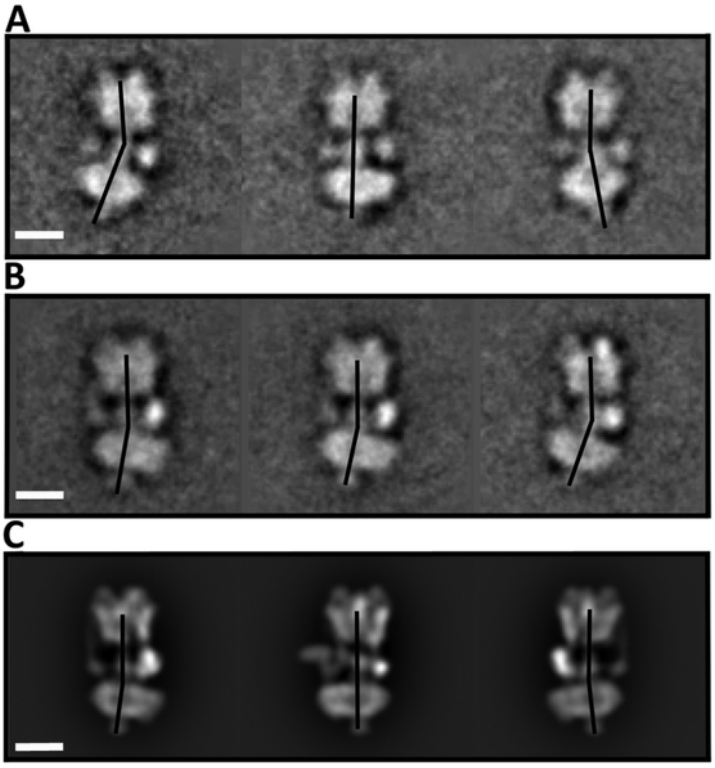


Figure 3.2 CryoEM single-particle analysis of flexibility in *M. Sexta* V-ATPase. (A, B) Single-particle analysis of the *M. sexta* V-ATPase enzyme aligned against V_1 and classified against V_O. (C) Re-projections from the *M. sexta* reconstruction showing the greatest displacement of V_1 relative to V_O. Scale bars represent 100 Å.

3.3.3 Coarse-grain Modelling and Mechanics of the Eukaryotic V-ATPase

The electron density of *M. sexta* V-ATPase was interpolated with 250 pseudo-particles. With such coarse-graining, application of standard elastic network models (ELN) without refinement/parameterisation specific to the system under investigation is not expected to give reliable results regarding the dynamics and mechanics of the complex. Therefore, a bent-stretch elastic network was employed as explained in the Materials and Methods, where topological information was combined with effective, empirical force constants between the beads and employing elements from linear elasticity theory. The first two non-trivial eigenmodes of the Hessian matrix of the BTS-EN model are depicted in **Figure 3.3**. The first eigenmode corresponds to the longitudinal flexing of V_1 relative to V_O (**Figure 3.3A, B**). The second mode corresponds to a twist of V_1 relative to V_O (**Figure 3.3C**). The eigenmodes of the model suggest that the holoenzyme is dynamic and deformable along specific directions representing putative functionally relevant motions. Those modes of motion are encoded in the topological organisation of the subunits in the complex (molecular architecture) indicating they are likely to be a universal characteristic of eukaryotic V-ATPases.

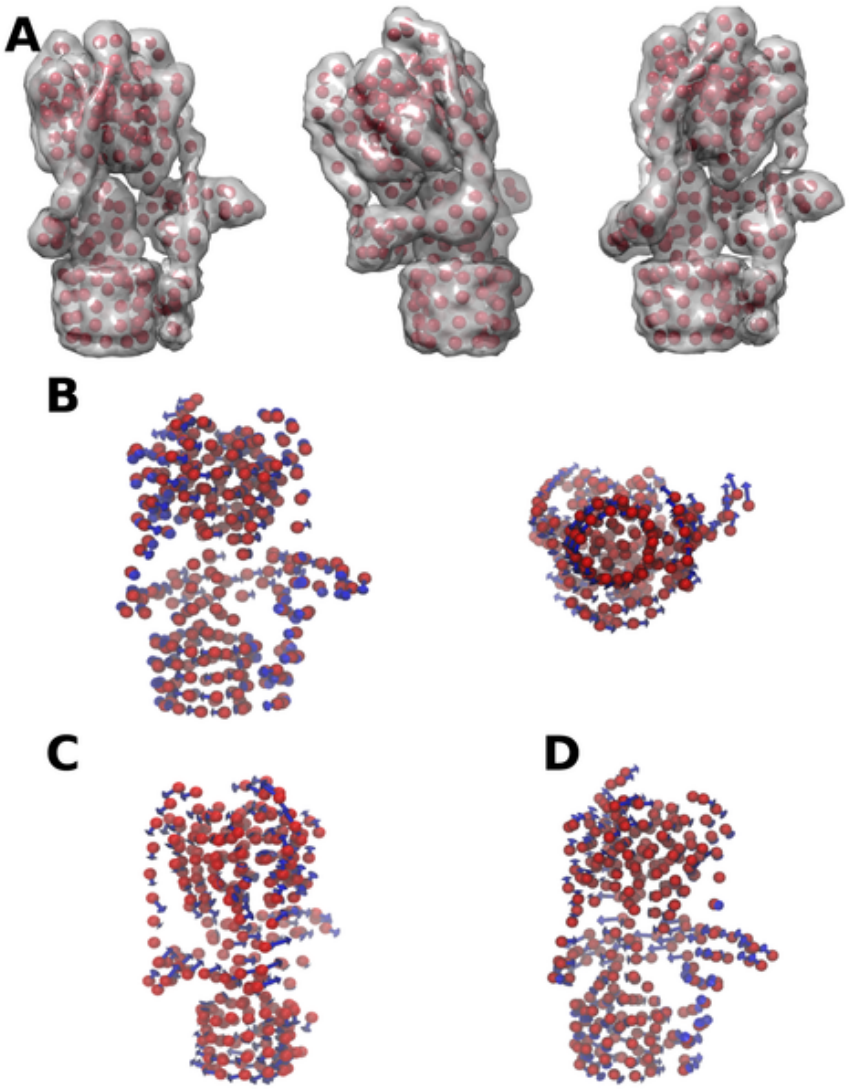


Figure 3.3 Deformation of the yeast V-ATPase along the first three non-trivial normal modes as calculated for the 250-bead ERNM. (A) Extreme conformers are depicted as a coarse-grained representation and as interpolated density maps, where the two motors are rotated against each other (left) or bended (middle, right). (B) Blue arrows represent the eigenvector corresponding to the first non-zero normal mode, which corresponds to twisting of the whole complex, consistent with the rotary mechanism of the V-ATPase. (C, D) The second and third modes of motion are bending motions with the soluble motor flexing (V_1) with respect to the membrane rotor domain (V_O) either back-forth (C) or side-by-side (D) suggesting that V-ATPase is laterally compliant.

3.3.4 Coarse-grain Modelling and Mechanics of the A-ATPase

As discussed for the case of the eukaryotic V-ATPase, conformational flexibility is an inherent property of rotary V-ATPases directly and/or indirectly linked to their function. Structural organisation of the constituent subunits is one of the main factors that dictate its mechanical behavior, a feature of all proteins as it has recently become apparent after the success of simple, topological models in predicting low-frequency, large-scale modes of motion that correlate with protein flexibility, plasticity and function.

Rotary A-ATPases are much simpler compared to eukaryotic V-ATPases in terms of structural organisation and metabolic regulation, and they have served as a prototypical system for biophysical and structural studies of the rotary ATPase family. However, it is still a very large system for *in-silico* investigations and resorting to simple, intuitive and empirical approaches is inevitable. To investigate A-ATPase mechanics and compare this with the V-ATPase, the same modelling approach was employed. Normal-mode analysis of a coarse-grain model of the *T. thermophilus* A-ATPase resulting in the decomposition of A-ATPase dynamics into harmonic modes of motion. The first three, low-frequency modes were examined and qualitatively characterised. In **Figure 3.4**, the first three eigenvectors are depicted as porcupine plots. The first mode of motion corresponds to a torsion movement along which the two motors rotate relatively to each other. This low-energy motion correlates with the biophysical mechanism of action of the system, that of producing torsion to catalyse ATP synthesis or pump ions against an electrochemical gradient across the membrane. The other two motions are bending motions of *orthogonal* directionality determined by the presence of the two peripheral stators that counteract the produced torque. One of the bending motions is *radial,* resembling the wobbling mechanism suggested by D. Stock and colleagues.

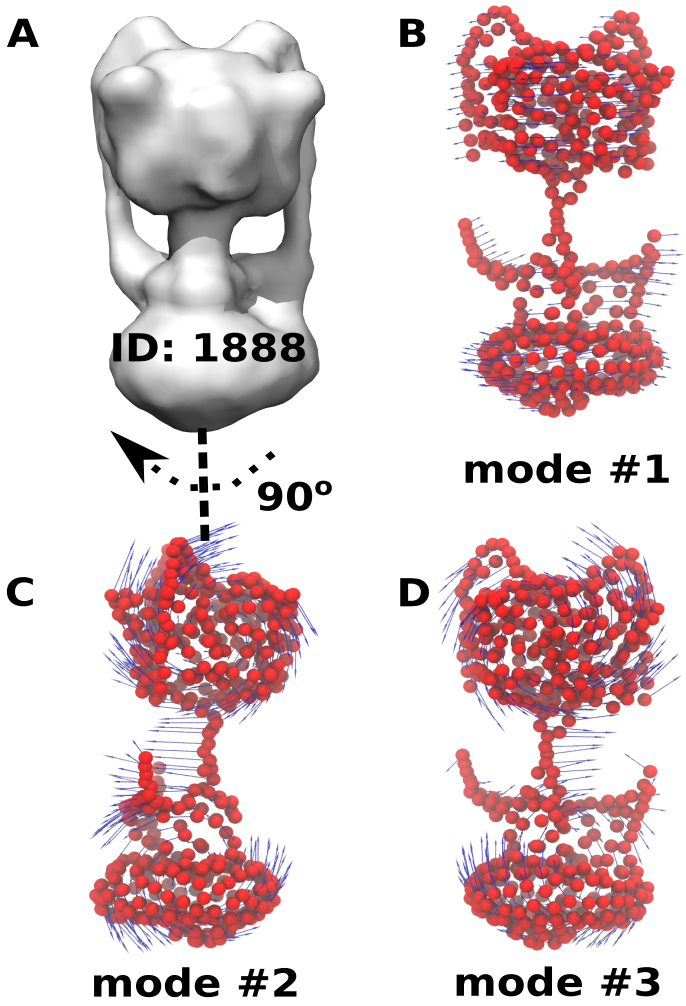


Figure 3.4 Bent-Stretch-Torsion Elastic Network Rod of *Thermus thermophilus* intact A/V-ATPases. (A) Electron cryomicroscopy density map of the *T. thermophilus* A/V-ATPase at 10-Å resolution (EMDB_ID: 1888). Normal mode analysis of a 250-bead BTS ER model based on the EMDB_ID: 1888 density map resulted in three non-trivial low-frequency eigenmodes that describe the essential dynamics of the intact complex under the quasi-harmonic approximation. (B) The first non-trivial mode (#1) is qualitatively characterised as torsional movement of the soluble motor relatively to the membrane-embedded sector. The second #2 (C) and third #3 (D) eigenmodes refer to the main global bending motions of the complex dictated by the topological organisation of the holo-enzyme and particularly the presence of the two peripheral stalks.

We should note that although Stock suggested the wobbling motion occurring during the rotary catalysis cycle, in our case the bending motion results from thermal agitation of the intact complex being in a straight-like, minimum-energy state with the catalytic subunits being in a specific occupancy sates. In other words, in out investigation the system is in a specific chemo-mechanical state, and the normal-mode analysis provides possible modes of motions, thermally excited and contribution to the overall conformational flexibility and variability. The agreement between out normal-mode analysis and the Stock analysis of rotary catalysis mechanism indicates that conformational dynamical changes during rotary catalysis are, to a large extent topologically dictated. Its also important to note that the mechanism reported by Stock and colleagues of A-ATPase dynamics implies a mechanical mechanism where the minimum-energy shape of the intact complex changes (bends) as a results of the chemo-mechanical stage of the catalytic cycle, although not explicitly claimed so. We suggest that such a mechanism is unlikely and that the complex harvests thermal fluctuations along its main modes of motions (excitations), mainly dictated by topological constraints of subunit organisation in the intact ATPase, to allow conformational re-arrangements within the catalytic subunits to occur.

An important deficiency of the BTS elastic rod network model and all other empirical topological models is the absence of physics-based, accurate energetics missing in that way any mechanistic implications of energetic frustration. In addition, the ranking of the various normal modes according to their eigenvalues is approximate and the magnitude of the motion cannot be inferred from first-principles within the framework of elastic network models. The results from normal-mode analysis for A-ATPase, however, provide a simple and intuitive picture of ATPase mechanics and conformational flexibility and demonstrate that mechanical linkages formed by peripheral stator and central rotor subunits are important determinants of the mechanical behavior. Therefore, detailed investigations of those subunits could provide a better understanding of their intrinsic mechanical behavior and their contributions into the mechanics and functional dynamics of the intact A-ATPase. In the next two sections, results from

molecular dynamics simulations of the peripheral stator stalk EG complex and central rotor axle DF are presented and discussed.

3.3.5 Understanding the Apparent Stator-Rotor Connections using Coarse-grain Protein Modelling and Computer Simulations

An interesting feature that emerges from inspection of available 3D density reconstruction of rotary A- and V-ATPases is the existence of apparent connections between subunits of the peripheral stator and central rotor axle. Naturally, such linkages restrict the rotational motion of the c-ring during the rotary catalysis cycle. It is not known what the biophysical importance and origin of those linkages are. We lack information about the strength of the corresponding protein-protein contacts implicated in the apparent linkages. To investigate the effect of the apparent stator-rotor connections on the intrinsic flexibility of intact rotary ATPases, the ERNM methodology and the *Fluctuating Finite Element Method* (Richardson et al, 2013) coarse-grain models were used and compared as described in detail in (Richardson et al, 2014). The computer simulations and normal mode analysis suggested that global conformational flexibility is significantly influenced by such interactions in the strong-interaction limit. The good qualitative agreement of the coarse-grain methodologies provides further support to the appropriateness of the simpler ERNM for dynamical investigations of large protein complexes and can be attributed to the fact both models are able to capture accurately topological features of the rotary ATPases.

3.3 Discussion

Rotary ATPases have extraordinary levels of efficiency of energy transduction and the intrinsic flexibility is likely a significant contributory factor. This flexibility has been postulated for the F- and (by analogy) V-ATPase mechanism but had previously not been directly shown. This flexibility has been linked to a spring-like

function, proposed to mediate elastic power transmission between the asymmetric ATPase and H^+ translocation motors in F-ATPase. Flexibility is also implicated in accommodating the large-scale conformational changes that occur sequentially in the soluble domain during the catalytic cycle. Since the membrane bound V_o domain is constrained by the membrane environment this would translate into 'wobble' of the ATPase motor as the rotor processes through its full cycle, accommodated by radial movement of the stator filament. The related F-ATPase is also suggested by crystallographic studies of F_1-c_{10} complex to show an apparent wobble of $\sim 11^o$ (Giraud et al, 2012; Stock et al, 1999), but these studies may be influenced by crystal lattice contacts. It is also noteworthy that in the crystal structure of the membrane extrinsic part of F-ATPase (Rees et al, 2009), the partially resolved stator filament comprising subunits OSCP/b/d/F_6 bends towards the central axle by a greater angle than that displayed in the cryo-EM reconstruction of the whole complex (Rubinstein et al, 2003). This implies that in the F-ATPase the stator is 'spring-loaded', clamping subunit *a* onto the c-ring whilst maintaining sufficient flexibility to accommodate changes in the surface of the $(\alpha\beta)_3$ complex during rotation and any eccentricity in the rotation of the c-ring.

This flexibility is linked to the properties of both the stator and central axle elements of the rotary ATPase (discussed in **Chapters 4 & 5**). The stator filaments are different both in composition and number within the rotary ATPase family. Moreover, the F-ATPase has a multiple helical fold (Dickson et al, 2006), whereas A- and V-ATPases contain respectively 2 and 3 right-handed coiled-coil helical filaments (Diepholz et al, 2008a; Diepholz et al, 2008b; Lau & Rubinstein, 2010; Lee et al, 2010b; Muench et al, 2009a; Vonck et al, 2009; Zhang et al, 2008b) only one of which may be directly linked to subunit *a* (Norgett et al, 2007). The central rotor axle axles of both A/V- and F-ATPases are similar in that they contain an extended helical coiled-coil (Gibbons et al, 2000b; Numoto et al, 2009), but differ significantly in both size and composition in the region that interacts with the *c*-ring (see **Figure 1.1**). Although the limited resolution of the EM classes makes it difficult to decide, the data is more consistent with a change in the angle at which the axle exits the

$(AB)_3$ headgroup rather than articulation at the D/*d*/*c*-ring coupling, for this reason the stator components were subjected to more rigorous analysis.

The absence of complete atomistic models for any rotary ATPase, in particular for the membrane bound subunit-*a*, introduces significant uncertainties and errors in atomistic modelling and molecular-dynamics simulations. However, topology and shape are often sufficient to predict the dynamic behaviour of proteins and their complexes using a coarse-graining *Elastic Network Model* (ENM), approach. The resulting simulations of the V-ATPase show two low-frequency modes. The first of the reliable models shows flexion along the long axis of the complex, consistent with the stator 'wobble' evident from the EM data analysis. The normal-mode analysis of the ENM also suggests twisting of V_1 relative to V_0 is possible. Since both motors work in a rotary fashion, this motion maybe more representative of an elastic storage mechanism whereby the torsional forces created in V_1 rotate it away from V_0, with the stators and central axle twisting in response. In principle, this motion could represent stator filament bending as part of an elastic power transmission mechanism, whereby the torsional forces created in V_1 cause counter-rotation with respect to V_0, with the central axle and stators, twisting in response. In order to understand how stator number may affect the flexibility of the rotary ATPase family the A-ATPase was also studied by the same coarse-grained approach as the V-ATPase. These results show a similar flexion feature which significant flexing along the complex with V_1 moving relative to V_o.

An interesting feature within the rotary ATPase family is an apparent linkage between a stator and the central axle in both the A and V-ATPase family. Flexibility in rotary ATPases had previously been predicted through incomplete crystal structures and molecular dynamic simulations of components of the complex. Here we reported the first normal-mode analysis of the intact A and V-ATPases and showed consistency of the results with electron-microscopy data. Both studies showed a flexing motion of V_1 relative to V_o to a maximum of 30° and rotation of the two domains relative to each other. Such flexibility has implications for elastic transmission and the dissociation mechanism. Overall this work has produced the

first study examining apparent flexing within the V-ATPase and by implication rotary ATPase family.

3.4 Multiscale Modelling and Simulation of Intact Rotary ATPases-Future work

The results generated from the negative stain and ENM analysis clearly demonstrated the presence of flexibility within the V-ATPase, which may be a general feature of rotary ATPases among divergent organisms. However, it is difficult to understand the biophysical consequences of such flexibility and its connection with rotary ATPase function and efficiency. Especially when conformational flexibility is itself poorly understood and its molecular origin unclear. The effect of flexibility on function could be *direct* or *indirect. Direct connection* simply means that flexibility has an active involvement in the mechanism of the complex, similarly to the suggestion of the existence of energy-buffering elements in F-ATPase. *Indirect connection* is simply established as a result of physico-chemical/molecular features that also determine the function of the complex.

Compared to empirical elastic network models, more systematic, rigorous and accurate *in-silico* models and computer simulations are needed to complement, interpret and inform future experimental studies on the structure, flexibility and function of intact rotary ATPases. The correspondence between *evolution*, *structure*, *flexibility* and *function* of rotary ATPases is not well understood; whereas it is a subject of paramount importance for understanding the biophysical underpinnings of rotary ATPase function, only recently researchers gained some insight into the structure and flexibility of intact ATPases, and the results discussed above in that chapter constitute one of the most important recent contributions. At this point, we clarify that the term ‘mechanics’ as used in this thesis refers to the characterisation of *shape*/*broadness* of a (meta-) stable state in the (free) energy landscape of the biomolecular system: the wider the free-energy basin, the more flexible the system.

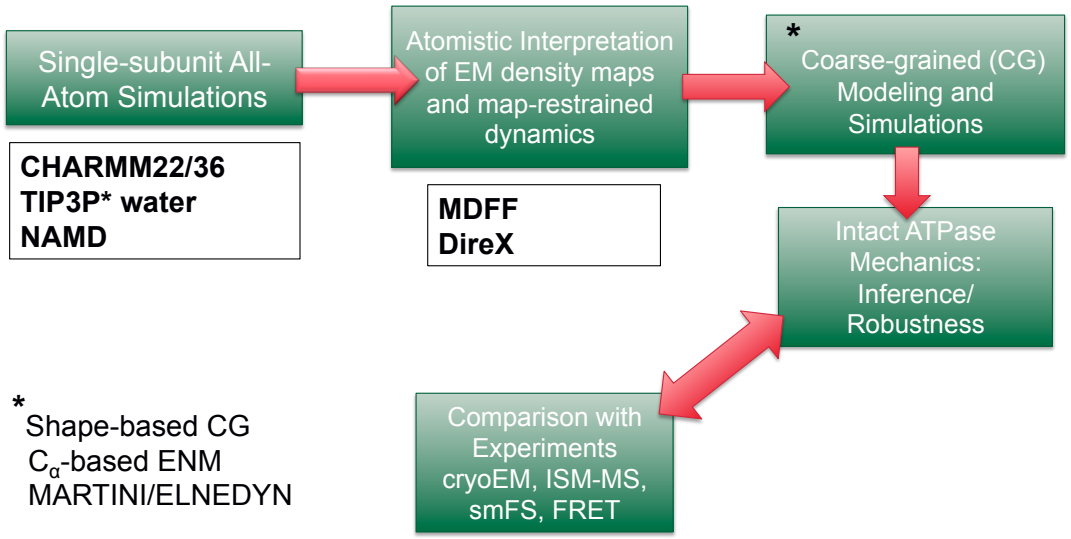


Figure 3.5 Multiscale integrated computational/experimental scheme proposed to investigate structural and functional dynamics of rotary ATPases.

The crucial questions that arise then and should be investigated are:

1. ***what are the equilibrium mechanical properties of an intact ATPase in a specific chemo-mechanical state***: before proceeding deeper into the structure-flexibility-function connection in rotary ATPases, we need to fully characterise the mechanical properties of the intact complex as the latter arise from equilibrium conformational variability of the complex at a specific stage of the catalytic cycle. The characterisation can be achieved by proper analysis of experimental data the come from electron microscopy, FRET and single-molecule force spectroscopy studies complemented with appropriate

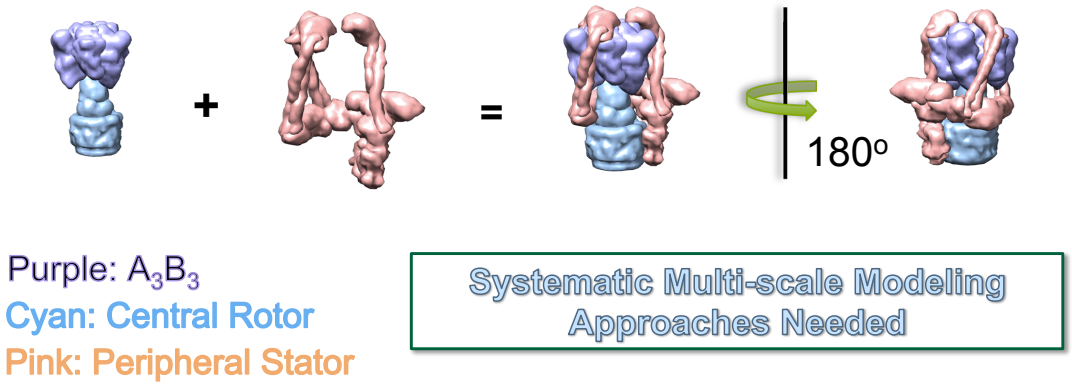


Figure 3.6 It is convenient to consider all rotary ATPases as a system consisted of two sub-systems, the 'central' subunits (*purple*, *cyan*) and the peripheral subunits that form a scaffold and part of the membrane rotor (*pink*). An essential prerequisite for an accurate CG model would be the accurate description of conformational flexibility for those subunits that couple the two motors (*pink* colour subunits along with central rotor subunits).

molecular modelling that reconcile experimental observations with high predictability of the system behavior under different conditions.

2. ***what is the origin of the mechanical behavior***: having characterised the flexibility of intact ATPases, the next question is to reveal the origins of such a nanoscopic behavior. Since ATPases are complexes, in principle, the factors that dictate the global behavior are the topological organisation of the subunits within the complex, the intrinsic flexibility of each individual subunit and the strength/plasticity of protein-protein contacts within the complex.

Indeed, arrangement of protein subunits within a complex poses topological constraints on the dynamics of the complex, in the same way that the 'network' of residues in space determine protein shape and, therefore, dynamics. Topological differences and subunit stoichiometry are important discriminatory factors

between the members of the rotary family. The apparent complexity of the rotary ATPase family increases from F-ATPase through A-ATPase to V-ATPase, characterised by an increase in number of subunits and peripheral stalks serving as a *topological classifier*. The subunits likely differ in shape and material properties, so that each subunit shows different mechanical characteristics (flexibility, plasticity) that would affect the global behavior of the intact complex and complex regulation. Especially, subunits involved in the *mechanical coupling* of the two motors transmitting torque generated (central rotor) or counteracting the rigid-body rotation movement of the soluble motor with regards to the membrane one. It would be interesting to understand how homologous subunits from different members of the family compare with each other and how those changes reflect on functional/regulatory characteristics of each family member. Finally, protein-protein interactions are another important aspect since the stability of those interactions has an impact on the stability of the complex.

To tackle computationally/theoretically the above fundamental questions, we need a framework that would allow computations to be feasible at an appropriate level of accuracy and theoretical abstraction. The proposed computational framework is presented in **Figure 3.5**. First, each subunit is investigated with detailed, atomistic simulations in an explicit-solvent environment to probe its conformational dynamics and landscape. Not all the subunits are expected to be equally important for intact complex mechanics. Rather subunits that connect the two motors are considered *mechanically critical*. Secondly, if low-resolution structural data are available for the isolated subunits in solution or the intact complex, it could be used to infer a quasi-atomistic model of the subunits within the complex and if any structural changes are observed compared to the solution ensemble of structures. Biasing methods such as MDFF are well suited for those structural investigations. Thirdly, the wealth of structural information obtained form atomistic simulations could become manageable with the application of dimensionality reductions techniques that would preserve structural and molecular intuition at the same time. Such a coarse-grain approach can result in a simplified representation of the protein that is in accordance with fine-grain

atomistic data, allowing faster computations. Having defined CG models for each subunit and their interactions, a CG model for the intact complex can be built and used for computer simulations of that larger system. A model for the intact ATPase can be further assessed and/or refined against available experimental data (cryo-EM, FRET, smFS, native IMS-MS) and used to guide the design of new experiments. Particularly for rotary ATPases, the whole complex can be considered as an association of the 'central' subunits with a sub-complex of peripheral subunits ('scaffold') as graphically shown in **Figure 3.6**. The results presented in the next chapter of the thesis essentially represent our investigations during the first two stages of the proposed scheme (**Figure 3.5)** as applied in the cases of rotary A- and V-ATPases.

3.5 Conclusions

In the current chapter, we presented the analysis of electron-microscopy 2D images of intact V-ATPase particles from the yeast and *M. sexta* V-ATPase that demonstrate for the first time that conformational variability (or *flexibility*) is intrinsic and conserved among the V-ATPase holoenzymes. The importance of conformational flexibility to function is less clear though. We believe that in order to clarify the importance of flexibility to enzymatic action, it is essential to characterise the mechanical properties of the holoenzyme at different chemo-enzymatic states of the catalytic subunits and determine the biophysical origin of the observed mechanical behavior. As a first step towards that goal, we discussed the experimental results under the light of normal-mode analysis results for the intact V-ATPase complex. We employed a coarse-grain particle-based model that allows the efficient calculation of normal mode eigenvectors for such a large biomolecular system and compared the computational results to those from 2D image analysis. The coarse-grain model is able to capture the basic modes of motion of the intact complex such as bending as a result of the overall topology of the complex. A torsional normal mode suggested from the calculations cannot be observed in the 2D images due to the limitations of

the experimental technique but it is likely to be physiologically important. The latter demonstrates the importance of such computations that can provide a first glimpse into the functionally important global motions of biomolecular complexes at minimal computational cost. Analogous computational analysis of the A-ATPase complex gave similar results with differences on the global bending motions as a result of differences in the peripheral-stalk stoichiometry and the topological organisation of the subunits. However, elastic-network-based coarse grain models are still simple enough to capture and/or explain the effect of protein-protein contacts on the flexibility of the intact complex. A multiscale hybrid computational/experimental methodology has been introduced at the end of the chapter as a promising framework to investigate the relationship between flexibility and function within the rotary ATPase family. The next two chapters of the thesis constitute the very first step towards that goal.

Chapter 4

Flexibility Within the Stator Filaments and Central Rotor Axle of the Prokaryotic A-ATPase

4.1 Introduction

The results and analysis discussed in Chapter 3 clearly demonstrated conformational flexibility of rotary V-ATPase particles, which is dictated in part by the stator elements and central rotor axle. This movement may accommodate ATP cycling and/or dictate the kinetic efficiency of these systems. Understanding rotary ATPase mechanics by computational means calls for a specialised framework that combines information from different sources, scales and levels of detail. To further understand the role that the stators and central rotor axle play in flexibility, we need to carry out detailed atomistic simulations of the peripheral stalk EG and the central rotor axle DF subcomplexes.

X-ray crystallography has provided static insight into the structure of the prokaryotic A-type peripheral stator stalk EG from *Thermus thermophilus* (**Figure 4.1**) (Lee et al, 2010a; Oot et al, 2012; Stewart et al, 2012). The two constituent subunits of the stator form a C-termini globular head domain that interacts with the catalytic A/B subunits of A_1-ATPase and a N-terminal right-handed coiled coil (rhcc) tail domain interacting with the peripheral collar subunit I. Based on two crystal structures of the *T. thermophilus* A/V-ATPase peripheral stalk, Stock and collaborators (Stewart et al, 2012) suggested that catalytic subunits A_3B_3 of A/V-ATPases wobble during the rotary catalysis cycle with the peripheral stators

accommodating the global conformational changes associated with ATP turnover, as a result of their intrinsic flexibility.

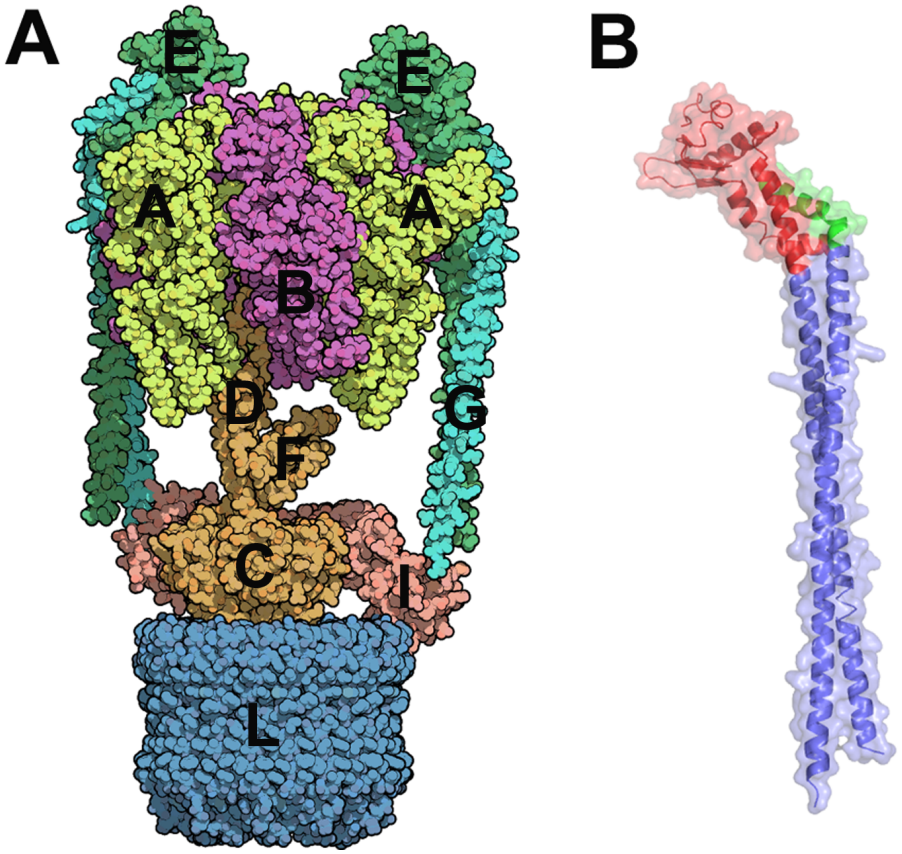


Figure 4.1 Structural organisation of the prokaryotic A-ATPase. (A) Two copies of the peripheral stator EG connect the catalytic subunits A/B with the membrane-bound subunit I. Along with the central rotor subunits DFC, EG subunits determine the overall flexibility of the intact holoenzyme. (B) The *T. thermophilus* peripheral stalk is shown in *new cartoon* representation highlighting specific domains. The two subunits form a right-handed coiled coil at the N-termini (blue) ending up in a globular head that interacts directly with the catalytic subunits of the A_1-ATPase (red). The C-terminal α-helix of subunit G interacts with the N-terminal region of subunit E to form a "neck" region at the boundaries between the globular and coiled-coil domains of the peripheral stalk.

For the central rotor of A/V-ATPases, most of our current knowledge comes from structural and biophysical studies on the *Enterococcus hirae* V_1-ATPase. The structure of the DF complex was determined both isolated (Saijo et al, 2011) and in the V_1-ATPase motor (Arai et al, 2013), revealing conformational changes induced by ATP binding and/or interactions of the catalytic subunits with the central rotor axle. However, the quasi-static structures are not particularly informative regarding the flexibility of the central rotor axle, the presence of an elastic-energy buffering mechanism and impact of the *rotor-stator* interactions on the dynamics and kinetic efficiency of the holoenzyme.

In the present chapter, atomistic molecular dynamics (MD) simulations of the isolated peripheral stalk and central rotor axle were employed to study their intrinsic conformational flexibility in solution conditions. Considering the currently limited structural studies of isolated A-type peripheral stalks and central rotor, our molecular dynamics simulations provide novel insight into their conformational dynamics. To simplify the analysis and enhance understanding of protein flexibility, we introduce a coarse-grain scheme for geometrical analysis of the conformational space sampled. In addition, our data allow an estimation of the persistence length of coiled-coil tail of the peripheral stalk for the first time. For the central rotor axle, analysis of torsional flexibility is carried out to allow comparison of conformational flexibility and heterogeneity with the F-type homologous γδε subcomplex. The chapter is organised into two parts: in the first part the results for the peripheral stator stalk are presented in detail along with various analysis methods to characterise conformational variability. In the second part, the results for the central rotor axle are discussed along with the analysis of its torsional flexibility. Finally, conclusions and future work are summarised at the end of the chapter.

4.2 Materials and Methods

4.2.1 Systems Preparation

Three different conformers of *T. thermophilus* A-ATPase EG stators (designated here as *PS1*, *PS2* and *PS3*) were used as initial structures for unbiased MD simulations. The first structure (PS1, PDB ID: 3K5B (Lee et al, 2010a)) was solved by crystallography to 3.1 Å resolution and comprises residues 21-120 of subunit G and residues 3-188 of subunit E. Missing residues on subunit E 143-144 were modelled with MODELLER (Webb & Sali, 2014). For crystallographic phasing the residues 134, 171 and 178 were engineered to selenomethionines but retained as wild-type leucines in the current study (PS1). The second crystal structure (PS2, PDB ID: 3V6I (Stewart et al, 2012)) (solved to 2.25 Å) exhibits a bending and twisting deformation with respect to PS1 (Kawasaki et al, 2010). PS3 conformer is a structural model (Stewart et al, 2012) that accounts for the density of the proximal peripheral stalk from an electron microscopy 3D reconstruction of the intact A/V-ATPase (Bernal & Stock, 2004). For the DF subcomplex, the crystal structure (PDB ID: 3AON (Saijo et al, 2011)) has been used (referred to as *Eh*DF). All protein structures were immersed in an orthorhombic water box and water molecules less than 2.5 Å from any protein atom were deleted. Sodium and chloride ions were added resulting in an electro-neutral solution of ~140 mM.

4.2.2 Molecular Dynamics Simulations

Atomistic, explicit-solvent, molecular dynamics simulations were performed with NAMD2.8 (Phillips et al, 2005). Three independent runs were carried out initiated from PS1, PS2 and PS3 for the peripheral stator and one run for the rotor axle. A summary of all simulations carried out is provided in **Table 4.1**. The CHARMM22 force-field with the CMAP correction (MacKerell et al, 2004a) was used for the protein, while the modified TIP3P* (Jorgensen et al, 1983) model was used for water molecules. Periodic boundary conditions were applied and long-range

electrostatics were calculated with particle-mesh Ewald (Lee et al, 1995) and grid spacing less than 1 Å. A cut-off distance of 12 Å was used to truncate long-range interactions with a switching function between 10 and 12 Å. The system was initially energy-minimized with 10,000 steps of conjugate-gradient minimization. Positional restraints were applied at the backbone atoms of the protein with a force constant of 50 kcal/mol $Å^2$. The system was then heated from 0 to 310 K using velocity rescaling, after which the temperature was maintained at 310 K with Langevin thermostatting while positional restraints were successively reduced until the protein was finally free to move. MD runs were performed in the NPT thermodynamic ensemble. Langevin dynamics (Brunger et al, 1984) with a damping constant of 1 ps^{-1} was employed to maintain a temperature of 310 K. Pressure control was achieved with a Langevin piston barostat (Feller et al, 1995) at 1 bar with a period of 200 fs and a decay rate of 100 fs. All covalent bonds involving hydrogens were constrained with the RATTLE algorithm (Andersen, 1983). An integration time-step of 2 fs was used; short-range interactions have been evaluated every time step, while long-range electrostatic interactions every second time step.

Table 4.1 Summary of all simulations carried out in this study and discussed in the main text with total simulation times indicated accordingly.

Simulation ID	Initial Structure	Total Time (ns)
sim1	PS1[a]	250
sim2	PS2[a]	140
sim3	PS3[a]	170
simDF	*Eh*DF	130

[a] Reference structures of the *T. thermophilus* peripheral stator stalk

4.2.3 Analysis

Average structures were calculated after aligning the proteins minimizing the C_α-based root-mean-square deviations from reference conformations. The covariance matrix C_{ij} of coordinates x_i and x_j was calculated and diagonalised. The resulting eigenvectors (or *principal components, PC*) were sorted by decreasing eigenvalue. Trajectories were projected on the first three eigenvectors to get the PC coordinates q_i. All PC analysis calculations were performed with WORDOM (Seeber et al, 2011), and all other geometrical calculations with VMD1.9 (Humphrey et al, 1996).

Local torsional flexibility for the *Eh*DF complex was characterized by aligning the first principal axis to be parallel to the Z-axis. The complex was then divided into 11 segments of 10 Å width along the *Z*-axis and the relative rotation of each segment relatively to the underlying one was calculated as the rotation angle around the axis that passes through the center of mass of the material segment in parallel to the Z-axis, so that the RMS deviation of segment *i* from the reference conformation is minimized. Circular standard deviations of angle fluctuations are reported as a measure of local torsional flexibility. A minimal model of elasticity (Czub & Grubmuller, 2011) has been constructed for the *Eh*DF complex, where residues are clustered into dynamic domains of residual coherent motion. The essential dynamics coarse-graining method (Zhang & Voth, 2010) has been adopted to partition residues into four clusters of atoms. The boundaries of the clusters were modified to result into four dynamic domains along the *Z*-axis. Bending and torsional angles were probed for each domain relatively to the underlying one by calculating the Euler angles of rigid-body rotation of each domain during MD simulations relatively to the reference structure. The rotation matrix was decomposed as

$$Q = Q_z(\varphi)Q_y(\vartheta)Q_x(\psi) \tag{4.1}$$

where $Q_{z,y,x}$ are the rotation matrices along Z, Y and X axis respectively.

4.3 Results and Discussion

4.3.1 Conformational Flexibility of the *T. thermophilus* Peripheral Stator Stalk through Atomistic Simulations

4.3.1.1 Conformational Dynamics and Residue Mobility

To examine conformational dynamics and stability along the three trajectories, C_α-based root-mean-square deviations (RMSDs) of the stator relative to the reference structures PS1/PS2/PS3 have been monitored for the whole complex (**Figure 4.2**).

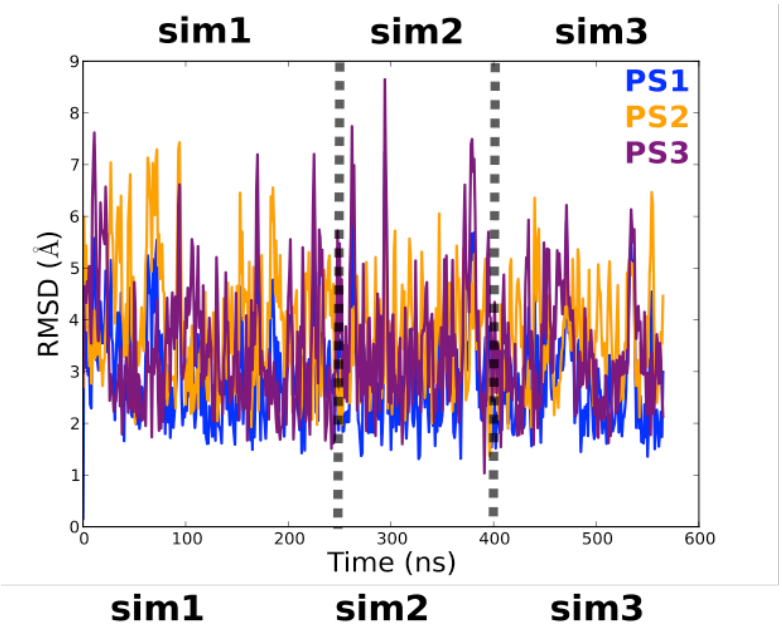


Figure 4.2 RMSD time series calculated for the whole heterodimer EG complex with PS1 (blue), PS2 (orange) and PS3 (purple) as reference structures. Results are depicted for all simulations with dashed vertical lines marking the boundaries between separate trajectories. RMSD values for the whole structure capture both local structural deviations and global, rigid-body motions between domains of the peripheral stalk. Sample statistics (averages, variances) are calculated for each trajectory excluding the first 40 ns.

Residues subE:123-150, which are part of the head domain, were excluded from RMSD calculations, since they form a disordered loop in PS1 where a β-strand appears in PS2/PS3. In all cases, global RMSD values show large fluctuations that are indicative of relative rigid-body motions between the peripheral stator domains (*head*, *rhcc*). Average $RMSD_{PS1}$ values are the smallest in all trajectories compared to $RMSD_{PS2}$/$RMSD_{PS3}$ suggesting that on average the peripheral stalk adopts in solution a global conformation closer to PS1. To eliminate the effect of rigid-body inter-domain motions, RMSD values for each dynamic domain (*head*, *rhcc*) were also calculated and shown in **Figures 4.3-4.4**.

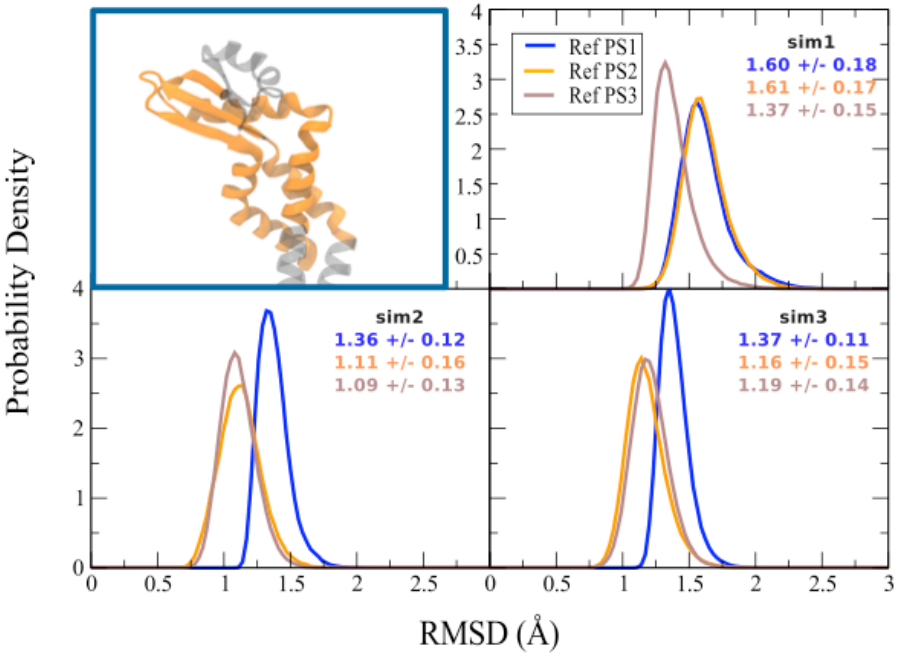


Figure 4.3 Probability density distribution of RMSD values calculated for the head domain (region highlighted in *orange* cartoon representation, *upper left box*). RMSD values are shown for each trajectory individually (sim1, sim2, sim3) using three different reference structures (PS1 in blue, PS2 in orange, PS3 in brown). Numerical values correspond to averages and standard deviations of the distributions.

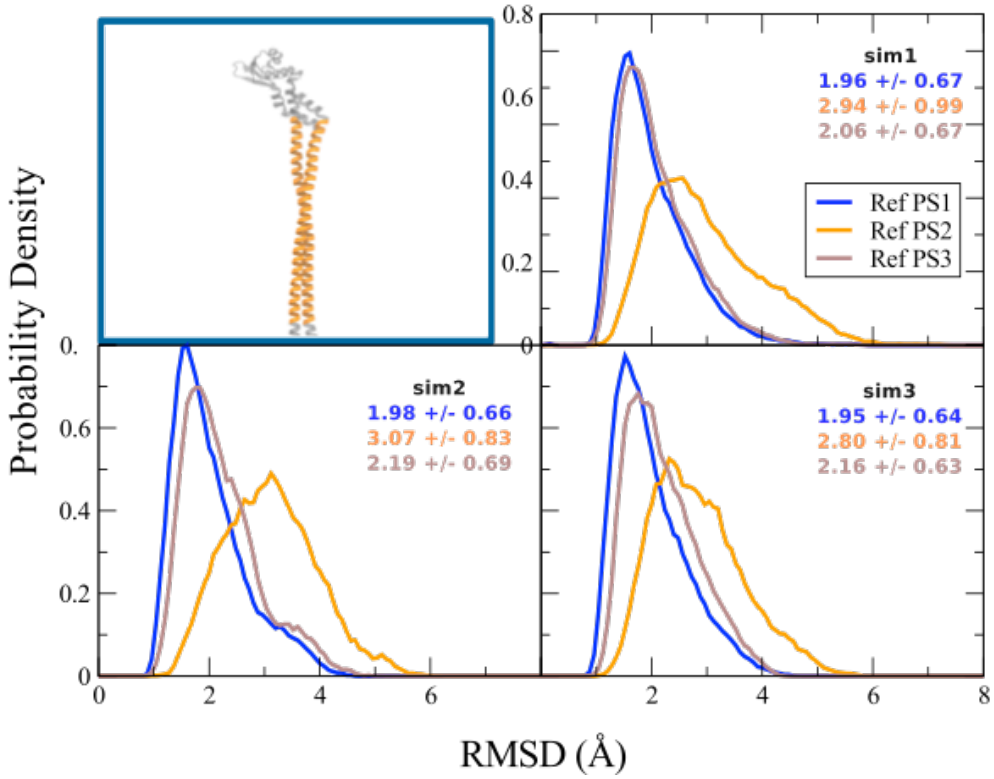


Figure 4.4 Probability density distribution of RMSD values calculated for the tail domain (region highlighted in *orange* cartoon representation, *upper left box*). . RMSD values are shown for each trajectory individually (sim1, sim2, sim3) using three different reference structures (PS1 in blue, PS2 in orange, PS3 in brown). Numerical values correspond to the average +/- standard deviation of the distributions. All three trajectories appear similar, and all deviate most from PS2, which has *curved* tail domain.

The head domain conformation remains very similar among the three reference structures. RMSD analysis of all simulations suggests similar stability of all three stator conformations in sim2/sim3, while differences between sim1 and sim2/sim3 are a result of the local secondary-structure differences of subunit E at residues 123-150 as previously discussed and/or the modelled leucines in *sim1*. The head domain dynamics appear to be sensitive to local changes as a result of local secondary-structure changes and residue modifications. On the contrary, analysis of *rhcc* RMSD values reveals that structural stability of the coiled coil is not affected.

The small values of RMSD for PS1/PS3 compared to PS2 demonstrate the free energy cost required for the coiled coil to adopt a PS2-like, bent configuration. Instead, on average the rhcc conformation is closer to PS1 and PS3.

To further characterize the flexibility of the peripheral stalk coiled coil the position of its N-terminal end relatively to the head domain was monitored. To align the trajectories for analysis, a reference structure was prepared as follows: an average structure was calculated in Cartesian space of C_α atoms from a concatenated trajectory of PS2 and PS3 excluding the first 40 ns of each constituent trajectory and then aligned so that the first principal moment of inertia of residues subE:42-51 and subG:65-73 was parallel to the z axis. Each trajectory was then superimposed onto the *head* domain of the reference structure mimicking the peripheral stalk when attached to the catalytic subunits. The XY trace of the center of mass of residues subE:6-13 and subG:32-37 at the N-terminal end of the coiled coil was monitored, and 2D probability density functions were constructed from histogramming (**Figure 4.5**). To characterize the above-mentioned fluctuations principal components analysis on the xy space was carried out. Eigenvectors calculated for the covariance matrix of x, y coordinates identify modes of motion around the average position, while the corresponding eigenvalues give the variance along those directions and they are therefore related to the flexibility along those directions. The data show that simulations of *sim2* and *sim3* essentially correspond to the same dynamical behaviour while *sim1* exhibits different dynamics. This reflects differences in the head domain, which affect the dynamical behaviour of the coiled-coil tail, as probed by the geometrical variable used in our analysis. Interestingly, the agreement between the directions of PCA and those of the principal moments of inertia (-0.92, 1.08) calculated from the average structure supports the view that the peripheral stator is stiffer along the direction of rotation than the perpendicular one (Stewart et al, 2014). However, the differences in flexibility as probed in our study are of the same order of magnitude. Defining effective elastic spring constants, k^{eff}, as $1/(2\sigma^2)$ in k_BT units, the ratio of k^{eff}_{radial} for the radial motion and k^{eff}_{perp} for the perpendicular one is ≈3 for *sim2* and ≈2 for *sim3.*

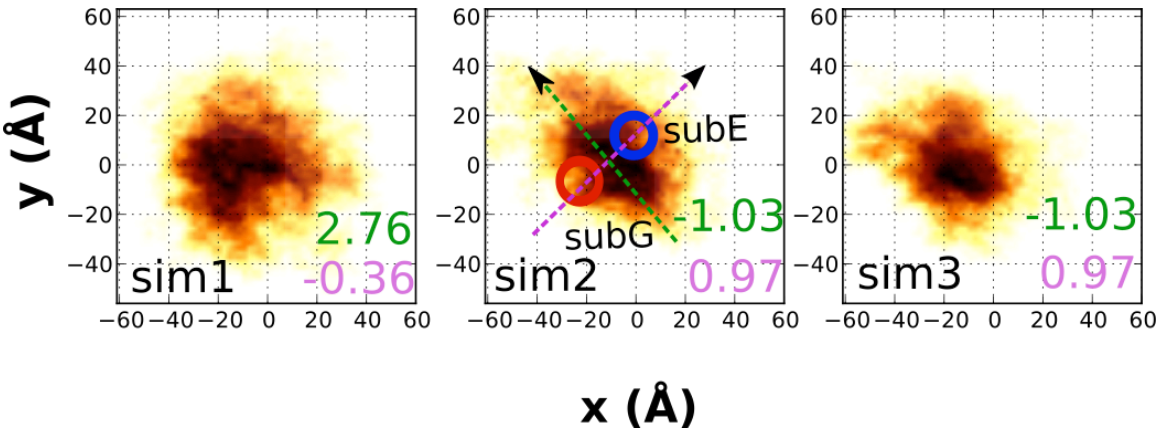


Figure 4.5 Probability density function of the N-terminal end as calculated for each trajectory separately. PS1 conformer is used as a reference structure aligned so that its first principal axis lies along the Z axis and the center-of-mass of the head domain is shifted to the origin of the axes. Then, each trajectory frame is transformed so that the RMSD of the head domain against the head domain of the reference structure is minimized. Finally, the position of the N-terminal end, defined as the center-of-mass (COM) of C_α atoms of subE:6-13 and subG:32-37 residues on the XY plane, is calculated and the XY coordinates retrieved. The 2D probability density function of the XY coordinates are depicted. Dark-coloured regions correspond to high-density regions, while light-coloured regions to low-density ones. Principal component analysis of the COM trace resulted in the eigenvectors of the covariance matrix, providing a means to characterize the coiled-coil apparent flexibility both quantitatively (flexibility along a mode related to their eigenvalues) and qualitatively (direction of the eigenvectors, numerical values in *green* and *purple* colours). *Blue* and *red* hollow circles (*middle*) graphically represent the two chains on the XY plane along with the two directions of the principal axes (dotted arrows in *purple* and *green* colours).

In **Figure 4.6**, the root-mean-square fluctuations (RMSFs) for the peripheral stalk complex are shown after global alignment of each trajectory to PS1 structure and the calculation of an average conformation in Cartesian space. The RMSF

indicates the presence of two relatively rigid regions connected by hinges. The first *hinge* comprises residues subE:19-31 and subG:44-51 located near the N-termini of subunits E and G where the peripheral stalk attaches to the soluble domain of the membrane subunit I. The second one consists approximately of residues subE:81-88 and subG:107-120 forming a '*neck*' region at the boundary of the globular head and the right-handed coiled coil. The stator complex preserves in that way its structural stability and rigidity bearing at the same time hinge-like regions that would allow the stator to accommodate tertiary displacement of A_3B_3 sub-complex relatively to the membrane sector as suggested in (Stewart et al, 2012). The rod-like shape of the tail suggests a level-arm movement where small-scale, local changes near hinge #2 segment are translated into large-amplitude changes at the N-termini end of the peripheral stalk. These results provide a dynamic view of the peripheral stalk in solution conditions. Simulations started from PS2 and PS3 tend to the same conformational ensemble in solution. Local differences in the head domain in terms of secondary structure and/or engineered residues between PS1 and PS2/PS3 appear to affect the head dynamics and be associated with differences in apparent flexibility of the coiled coil N- terminal end. However, in all cases except *sim1* and regardless of the degree of flexibility, the intact peripheral stalk complex preserves an average structure that is close to a PS1-like conformation with a straight coiled-coil tail arrangement. The latter contrasts with short MD simulations of the $A_3B_3E_2G_2$ 40^+ ion in gas phase performed to investigate structural dynamics under ion mobility spectrometry conditions. Absence of solvent screening in gas phase suggests that the structure of the EG is modulated in solvent-free conditions.

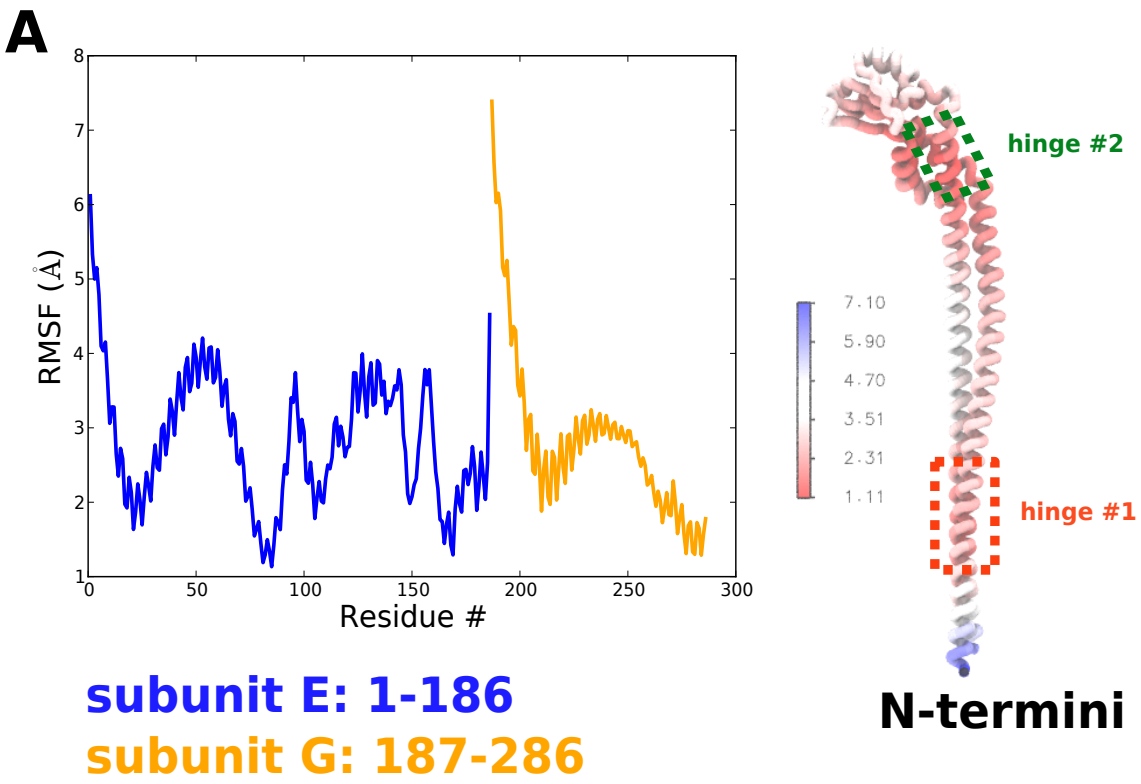


Figure 4.6 Residue mobility in the peripheral stator stalk. (A) Root-mean-squared-fluctuations (RMSF) calculated from the concatenated *sim2+sim3* trajectories; residues 1 to 186 refer to subunit E, residues 187 to 286 refer to subunit G. (B) The average structure from *sim2+sim3* coloured according to the RMSF values. Two hinge regions of low mobility are highlighted within dashed rectangular boxes. Small-scale local changes around *hinge* #2 region are amplified into large-scale movements of the N-termini end resembling a lever-arm motion.

4.3.1.2 Essential Dynamics Analysis

Given the similarity of *sim2*/*sim3* conformational ensembles, a concatenated trajectory consisted of the last 100 ns of *sim2* and the last 130 ns of *sim3* were subjected to principal component analysis (PCA). PCA decomposes conformational fluctuations around an *average structure* into *principal* modes of motion that best explain the observed conformational variability. Principal modes associated with large eigenvalues define *soft* modes of motion providing a coarse-grain means to

characterizing equilibrium nanomechanics of the EG stator when only a small subset of low-frequency eigenmodes are retained. The first three PC modes account for 80% of total conformational variability and projecting conformations sampled onto them reduce the conformational space into an essential dynamics subspace that is likely the most functionally relevant.

Inspection of the first PC eigenvector from the analysis of the whole stator shows that it corresponds to a bending motion where the two domains, head and tail, that are attached to the A_1 and A_o motors, respectively, move cooperatively allowing wobbling motions of the soluble catalytic subunits relatively to the membrane sector (**Figure 4.7A**). PC modes 2 and 3 also share some similarity with PC mode 1 but more complex motions are observed with twisting within the tail domain. To remove inter-domain motions and probe the flexibility within each domain individually, we carried out additional PC analysis separately for the two domains. Deformation of each domain is characterized in a similar way as above; PC analysis is first performed separately for each domain and projections on the first three PCA eigenvectors are used to characterize its deformation state from the average structure. In **Figure 4.7B**, 2D PMF of projections 1/2 and 2/3 are shown for the head domain. Reference structures PS2 (*triangle*) and PS3 (*rhomboid*) are also projected onto the PMF to show differences in stability between the two conformations. PC3 head domain is more stable compared to PS2 in accordance with the RMSD analysis. In **Figure 4.7C**, the same analysis is carried out only for the rhcc tail domain along with the three reference structures PS1 (*circle*), PS2 (*triangle*) and PS3 (*romboid*). Although, accurate free energy differences cannot be obtained from the three conformers, PC analysis suggests that the straight tail of PS1/PS3 most likely dominates in solution conditions, while the PS2 conformer adopts a higher-energy bent arrangement within its rhcc domain. The PS1/PS3 and PS2 conformations do not represent distinguishable states of the peripheral stator in solution but rather particular microstates within a broader ensemble of accessible microstates.

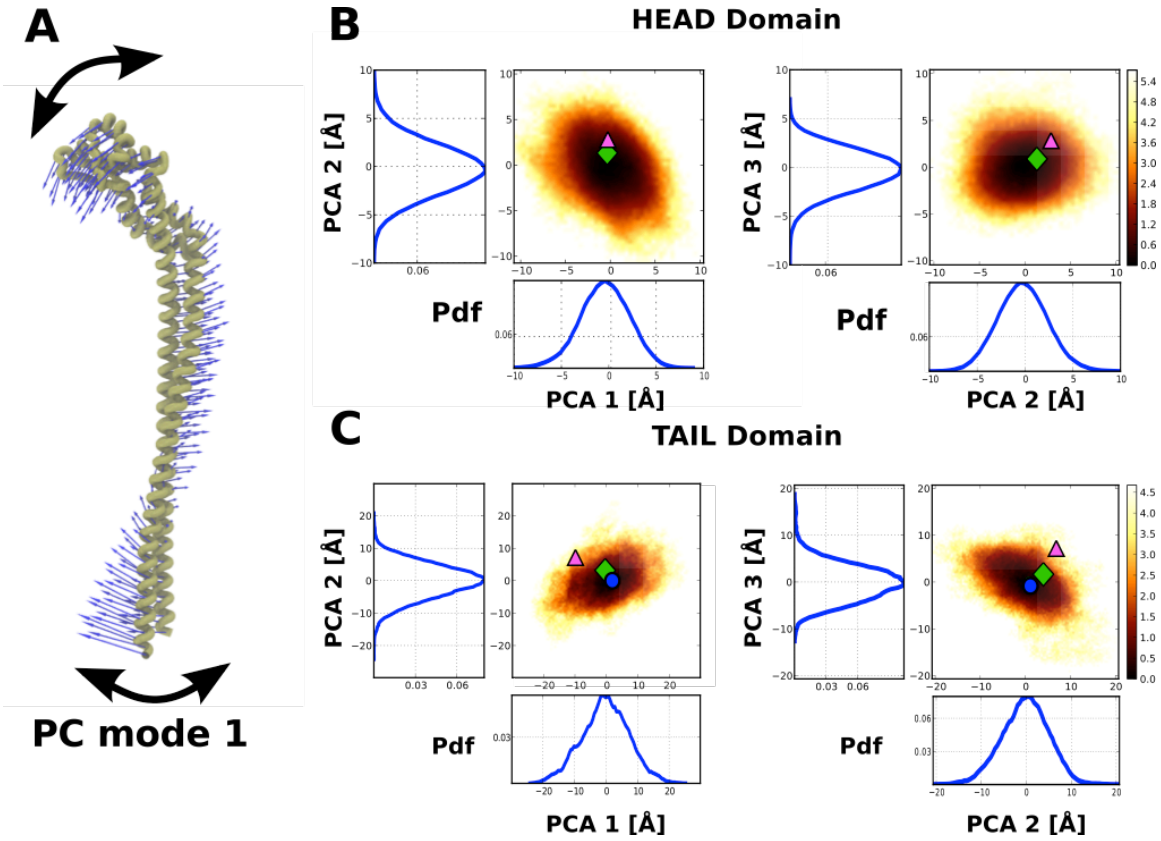


Figure 4.7 Essential-dynamics analysis of the peripheral stator stalk. (A) Visualization of the first eigenvector (PC mode 1) from PCA of sim2+sim3 trajectories depicted as *blue* arrows positioned on the C_α atoms of the complex. Arrows have arbitrarily been scaled up to enhance visibility. (B) Potential-of-mean-force, PMF, profile in k_BT units on the subspace spanned by the first two eigenvectors as calculated for the head (B) and tail domain (C) separately. The three reference conformers PS1, PS2 and PS3 are also depicted in (B) and (C) with a *circle*, *triangle* and *rhomboid* marker respectively.

4.3.1.3 Coarse Grain Analysis of the Peripheral Stalk

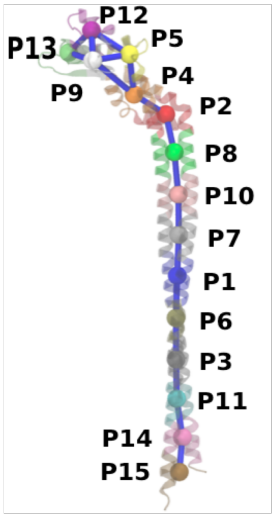
To further characterize stator mechanics, a coarse-grain (CG) pseudo-particle model for the hetero-complex was examined. The CG model captures the overall topology of the complex composed of supra-molecular beads, with residues coming from both stator subunits, justified by the tight coupling of the two chains through coiled-coil interactions. The use of CG particles in molecular dynamics simulations has

previously been applied in various systems (Baker & Voth, 2013; Saunders et al, 2014) and these served as a starting point to parameterize CG Hamiltonians for dynamical simulations of large biomolecular systems (Arkhipov et al, 2006a; Arkhipov et al, 2006b). The number of CG particles used in the study corresponds to an approximate ratio of 305 atoms per bead. Six beads were assigned to the head domain while the rest represent partitions of the right-handed coiled-coil domain (**Figure 4.8**). Having defined a CG representation of the EG complex, geometrical properties related to stretching, bending and torsional mechanics (distances, angles, and dihedrals respectively between CG beads) were calculated and sample statistics collected. Geometrical quantities provide a simple and intuitive way to characterize conformational flexibility avoiding the complications of dealing with a large number of degrees of freedom when preserving atomistic detail. Effective spring constants, k_{eff}, were calculated for all pseudo-bonded terms (**Figure 4.8B**) under the harmonic approximation according to the formula

$$k_{eff,i} = \frac{k_B T}{2\sigma_i^2} \tag{4.2}$$

where k_B is the Boltzmann constant, T the temperature of atomistic simulations and σ_i^2 the sample variance of the corresponding internal coordinate (IC). As **Equation 4.2** suggests effective spring constants are measures of stiffness as they are related to the inverse of an IC's variance. In the case of the EG complex, spring constants for bonds show a much "stiffer" coiled-coil tail compared to a "softer" globular head domain. The relative inextensibility of the *rhcc* domain results from interactions between the N-terminal domain of subunits E and G. The softness of the globular head is likely a result of its particular secondary-structure composition. Dihedral spring constants show a clear distinction between *head* and *tail* domain torsional flexibility. The difference between the average values is two orders of magnitude indicating that the *rhcc* domain is torsionally more flexible. It can be seen that on the basis of the above CG geometrical parameters, *sim2* and *sim3* sample the same conformational subspace (**Figure A1.1, Appendix**).

A



# CG Bead	Subunit E Residues	Subunit V Residues
1	42-51	65-73
2	79-83, 179-188	102-111
3	24-33	47-55
4	84-91, 167-178	112-119
5	92-104, 106-107	120
6	34-41	56-64
7	52-60	74-83
8	69-78	93-101
9	122-128, 143-153, 165-166	-
10	61-68	84-92
11	16-23	38-46
12	105, 108-109, 121, 129-142	-
13	110-120, 154-164	-
14	6-13	32-37
15	3-5	21-31

B

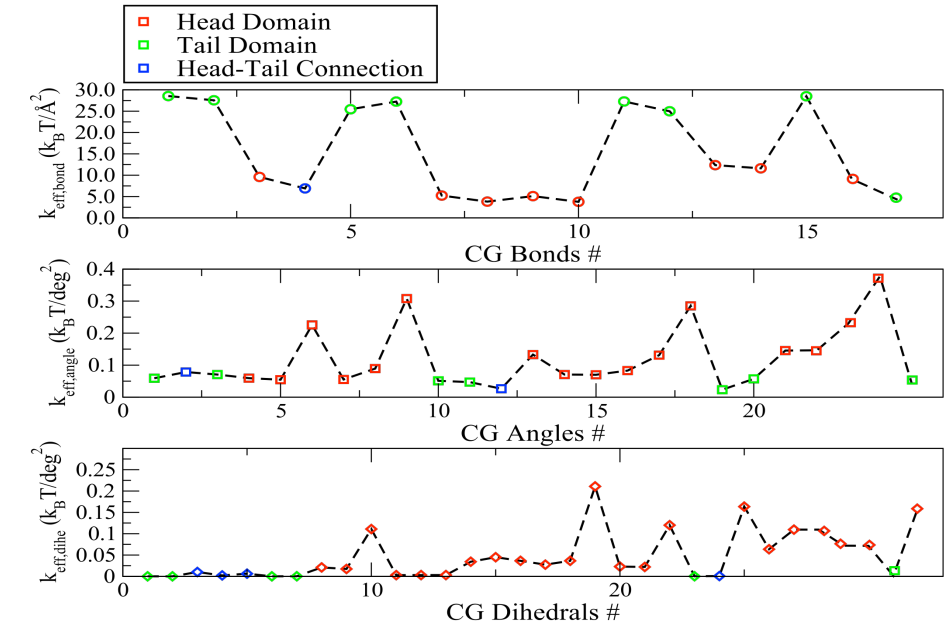


Figure 4.8 Geometric coarse-grain analysis of the peripheral stator stalk. (A) Visualisation of the coarse-grain beads selected to represent domains of the EG complex (*left*). The exact definitions of the CG particles in terms of residue membership is given in the table (*right*) (B) Apparent elastic force constants of the pseudo-bonded terms as calculated according to ***Eq. 4.2***. Terms that involve only head CG particles are shown in *red*, only rhcc tail CG particles in *green*, while terms that represent inter-domain linkages are shown in *blue* colour.

The head domain exhibits more complicated dynamical behaviour and initial conformations and/or mutations within that domain disturb the dynamics and flexibility of the domain, as previously indicated. It is not easy to comment on the mechanistic implications of these observations, but it is likely that altered dynamics would affect the cooperativity between the head and tail domain. Among all bonded terms involving *head-only* beads, one bond, angle and dihedral were selected to best demonstrate differences in the flexibility of the head domain and histograms of the sampled values are shown in **Figure A1.2, Appendix**. Some of the *reference values* although sampled during *sim2-sim3,* lay in lower-probability ranges of values, highlighting the deviations of PS1-PS3 structures from the ensemble in solution.

In addition, potential-of-mean-force (PMF) profiles extracted from the sampled distributions can be approximated by harmonic potentials (**Tables A1.1-3, Appendix**) with the exception of a few dihedrals that are better represented with CHARMM-like dihedral potentials. An interesting feature is the observed heterogeneity in terms of spring constant values. Interestingly, the quasi-harmonic approximation is valid for most of the terms examined except for a few dihedral terms (**Table A1.3, Appendix**). Quasi-harmonic behavior is one of the main approximations of the Bent-Twist-Stretch Elastic Rod Network model (BTS-ERNM) (Stember & Wriggers, 2009b), where a CG model is developed to more accurately describe protein dynamics using severe coarse-grain representations. Our current results point to the validity of BTS-ERNM's assumptions at least in the case of the peripheral stator EG complex. However, due to its complex architecture, the EG stator is expected to have a complicated mechanical behavior under mechanical load and is not fully described with simple, coarse-grain models or fundamental rod elasticity theory.

4.3.1.4 Right-handed Coiled-Coil Tail Persistence Length

Coiled coil is a ubiquitous structural motif consisting of two or more α-helices wrapped around each other. Tight packing of hydrophobic amino acids side chains at the core of the coiled-coil in a 'knobs-in-holes' fashion results in mechanical stability and rigidity, an advantageous property for structural proteins and molecular motors. Several experimental and theoretical studies have examined dynamical and mechanical properties of coiled coil domains of structural proteins and linear molecular motors (steppers). Single-molecule force spectroscopy has been employed to probe the force-extension diagram of myosin coiled coils (Schwaiger et al, 2002) while electron-microscopy and single particle analysis of 2D images are well suited for characterization of equilibrium thermal fluctuations of large coiled-coil proteins such as tropomyosin (Li et al, 2010). A property usually reported is the (apparent) persistence length as a measure of protein stiffness as it is related to the decay of spatial correlations along the contour length of the polymer.

The 'tail' domain of peripheral stalks of rotary ATPases is coiled-coil-like with some characteristic features compared to other known coiled-coil dimers. Specifically, peripheral stalk tail domains are right-handed with a combination of two different pitches. So, it is not *a priori* that they behave similarly to their left-handed counterparts. There are currently no reported values of the peripheral stalk persistence length for the A or V-ATPase. However, Junge and colleagues (Wachter et al, 2011) have estimated torsional and lateral elasticity for the b_2 complex from *E. coli* F_1F_O-ATPase recording thermally agitated fluctuations reporting lateral stiffness κ^{lat} = 77 pN/nm for a 5 nm long segment. Compared to measurements carried out for myosin II coiled coil, b_2 is much stiffer (Schwaiger et al, 2002).

In the absence of experimental estimations, lateral stiffness and *bending* persistence length, L_p was calculated using the end-to-end method of (apparent) persistence length estimation (Li et al, 2010). The average conformation of the coiled coil tail is rather straight, so no significant contribution of static persistence length to the total apparent persistence length is expected. For a worm-like chain,

end-to-end distance can be expressed as a function of the persistence length according to equation:

$$\langle R_{ee}^2 \rangle = 2LL_p - 2{L_p}^2 \left[1 - \exp\left(-\frac{L}{L_p}\right)\right] \quad (4.2)$$

where R_{ee}^2 is the end-to-end distance between the end segments of the coiled-coil, L is the contour length and L_p is the persistence length. Numerical solution of Eq. (4.2) gave a value of L_p=105±42 nm, comparable to estimations of conventional coiled-coil domains, e.g. myosins (Adamovic et al, 2008; Schwaiger et al, 2002). These results suggest a rather stiff peripheral stalk tail in accordance to cryo-EM density maps of $A_3B_3DF(EG)_3CH$ subcomplex where straight segments corresponding to EG subcomplexes were present in the density map (Muench et al, 2014).

4.3.2 Conformational Flexibility of the *E. hirae* Central Rotor Axle through Atomistic Simulations

4.3.2.1 Conformational Dynamics and Residue Mobility

The crystal structure of the *Eh*DF complex at 2.0 Å resolution revealed (Saijo et al, 2011) that N-terminal residues 9-69 and C-terminal residues 127-193 constitute a long coiled-coil of two α-helices (**Figure 4.9**). This coiled coil is inserted inside the cavity of A_3B_3 subcomplex and is involved in the transmission of the torque produced during the catalytic cycle. Interestingly, the coiled-coil structure is conserved in V- and A-ATPases and shares significant similarity with the rotor of mitochondrial F_1-ATPase (Gibbons et al, 2000a) and the coiled-coil segment of FliJ subunit of the flagellar type III protein export apparatus (Ibuki et al, 2011).

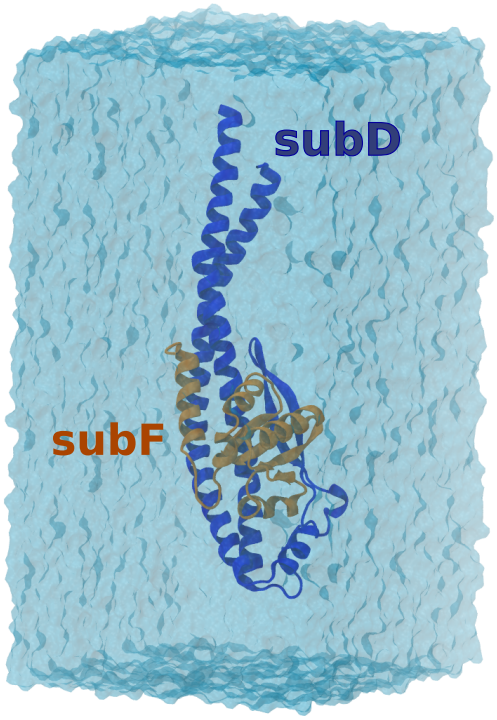


Figure 4.9 Visualisation of the *Eh*DF complex in a water box as prepared and used for a 130-ns unbiased molecular dynamics simulations.

The free *Eh*DF complex was simulated for 130 ns and the structural stability inspected for each subunit individually (**Figure 4.10A**). Both subunits appear to be stable with the elongated and larger subunit D being characterized by larger structural deviations and transient fluctuations due to conformational dynamics of the DF coiled coil segment.

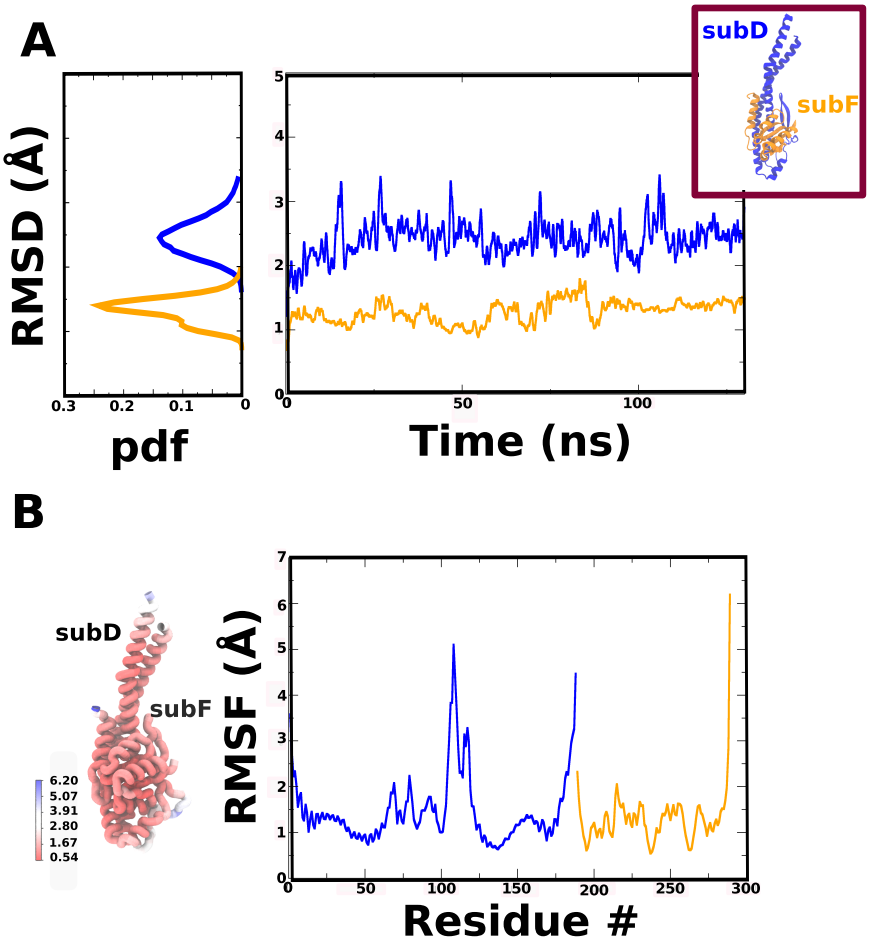


Figure 4.10 Conformational stability and residue mobility in the *Eh*DF complex during 130 ns of MD simulation. (A) RMSD calculations for each subunit demonstrate that the complex remains structurally intact and stable through our simulation with subunit D being characterized by larger RMSD value on average due to the presence of a dynamic coiled-coil segment (B) RMSF values calculated for each subunit from the last 90 ns of the simulation, show that subunit D is dynamically more heterogeneous compared to subunit D.

Root-mean-square fluctuations around the average structure (**Figure 4.10B**) were calculated excluding the first 40 ns of the trajectory and for each subunit individually to remove the contribution of quaternary structure fluctuations. Subunit F, whose role in the rotary ATPase is yet to be clearly resolved, fluctuates least, with the exception of the C-terminus. In comparison subunit D appears dynamically more heterogeneous, with low and high RMSF regions throughout the structure with dynamic loops having the highest RMSF values. The roles for this difference are yet to be resolved.

4.3.2.2 Essential Dynamics Analysis

PCA analysis of the last 90 ns resulted in the decomposition of atomic fluctuations around an average structure into a set of eigenvectors that correspond to *essential* modes of motion. The first three eigenvectors account for 50% of total conformational variability observed while the first 39 eigenvectors retain 90% of the total variance. The first eigenvector (PCA 1) (**Figure 4.11A**) captures a global bending motion where approximately 20% of the residues with the lowest mobility form a compact cluster at the core of subunit D and subunit F (subD: 54-55, 57-59, 136-147, 172-176 and subF: 6-14, 25-33, 48-58, 74-88) and a smaller cluster near the terminal helices of subunit D (subD: 20, 172-176). These could be identified as 'hinge-like' regions that connect higher-mobility protein segments. Similar analysis of the second PCA eigenvector shows the presence of a second bending motion towards different directions and the formation of two other clusters of residues, again on at the core of subunit D and the respective interacting subunit F residues (subD: 48-58, 83-85 and subF: 7, 14-24, 75-76, 80-87) and a second one near the termini of the helices of subunit D (subD: 23-32, 167, 170, 171). The third PCA eigenvector is very different and much more localized, since most of the residues are of low-mobility and instead of bending a twisting motion is observed between the terminal helical segments of subunit D. Since protein motions identified from PCA analysis are large-amplitude motions usually associated with functionally

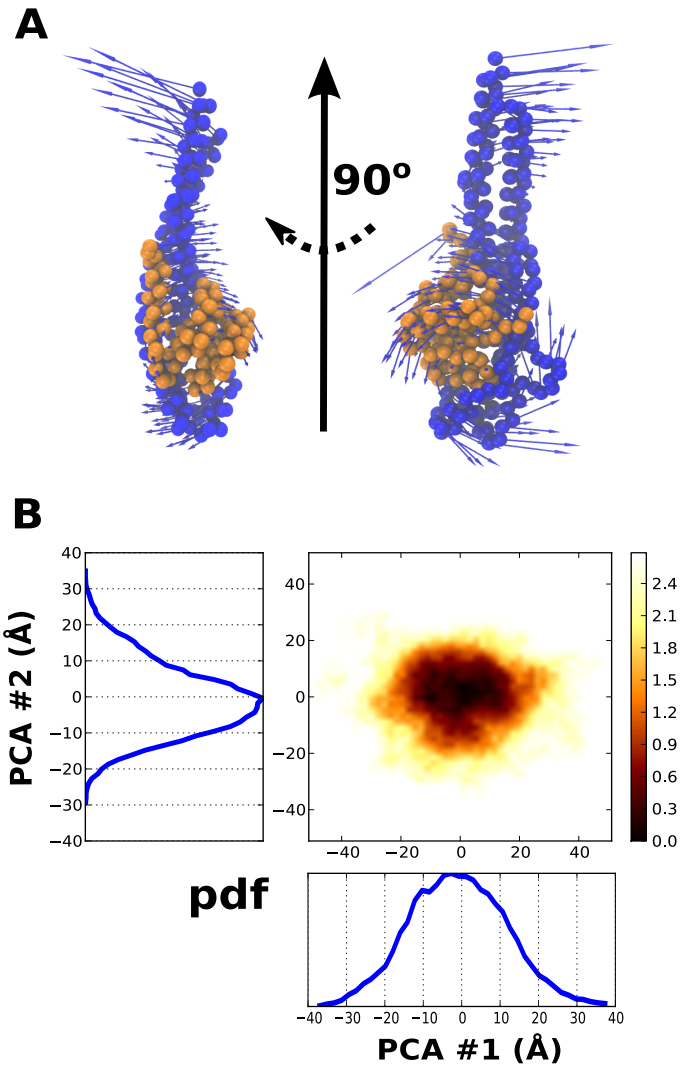


Figure 4.11 Essential-dynamics analysis of the central rotor axle DF. (A) Porcupine plot of the first eigenvector from PCA (PCA #1) of the last 90 ns of the MD simulation, representing a bending motion of the complex. C_α atoms of *Eh*DF are depicted as beads in *blue* (subunit D) and *orange* (subunit F) colour. (B) Projection of the MD trajectory on the first two eigenvectors from PCA analysis and the corresponding PMF profile in k_BT units. The corresponding 1D probability density functions (pdf) of each collective variable (projections on PCA eigenvectors) are also shown as blue lines. The presence of one relatively broad free-energy basin is in accordance with its mechanical role within the rotary ATPase complex.

important motions, the observation of bending instead of global twisting motions indicates the likely importance of these motions for the function of A/V-ATPases.

Normal mode analysis, molecular dynamics simulations and most biophysical and structural studies performed for the homologous F_1-ATPase have focused on the torsional flexibility within those complexes and neglected the role of flexibility in the central shaft. However, Nishizaka and colleagues (Sugawa et al, 2011) have demonstrated the tilting motion of the central shaft between intermediate states during ATP hydrolysis. The bending motion we observe in our simulations as revealed from essential dynamics analysis suggests that a tilting motion in the central shaft would be a conserved feature throughout the family of rotary ATPases.

Finally, a simplified free energy landscape was constructed from the projection of the trajectory onto the subspace spanned by the first two eigenvector (**Figure 4.11B**). *Eh*DF complex appears to sample a single, relatively broad basin. The variability along each eigenvector is related to the flexibility of the complex along that specific mode of motion. The absence of any other, distinct metastable state indicates that at least within the timescale of our simulations it is intrinsically related to the functional role of subunit *Eh*DF as a 'mechanical' component of the rotary A-ATPase implicated in the transmission of mechanical torque between the two motors.

4.3.2.3 Torsional and Flexular Elasticity of Central Axle *Eh*DF

To characterize equilibrium mechanics of the DF complex in solution and near ambient conditions, we can simplify the protein complex as rod-like, elongated material body with a varied cross-section. Instead of a continuum, a discrete representation is adopted, where the protein is divided into material slices along the first principal axis- here for convenience the latter axis is considered parallel to the Z axis (**Figure 4.12**). Due to thermal motion, every material slice experiences non-zero forces and torque from the environment (solvent molecules) and conformational fluctuations are observed. These fluctuations can be monitored and provide insight into the local flexibility of the system. The local segment fluctuations were approximated as effective torsional fluctuations around the axis that passes through the center of mass of the slice and is parallel to the *Z*-axis. Slices of 10 Å are

considered (eleven in total) and the fluctuations of each relatively to the underlying one were probed. Therefore, torsional fluctuations calculated by those means refer to local, differential torsional elasticity. In **Figure 4.12**, the circular standard deviation of rotation angles are presented as a function of the z coordinate that runs from 0 at the bottom of subunit D to approximately 100 Å at the tip of the termini α-helical segment of subunit D.

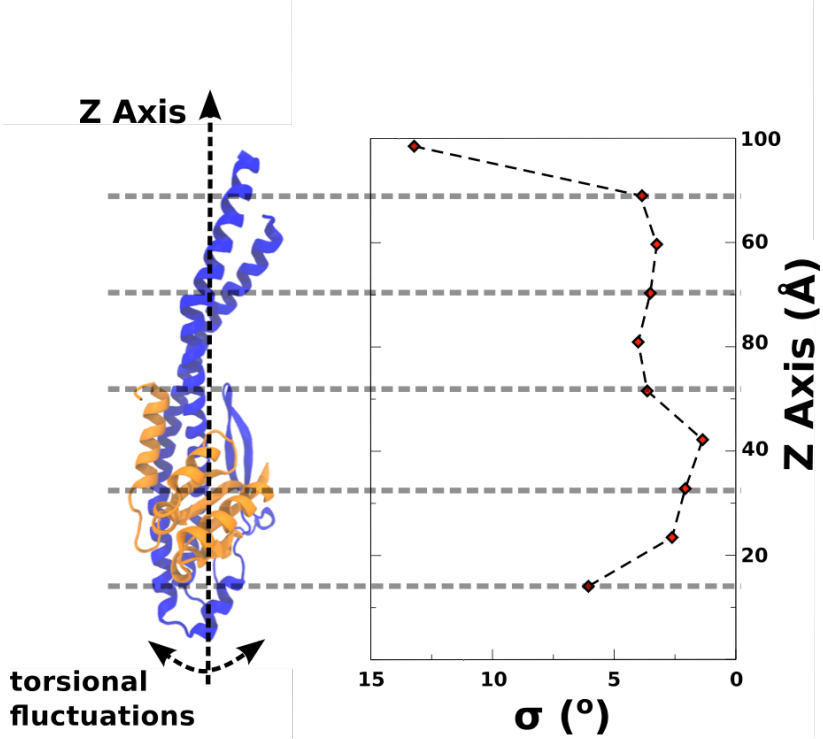
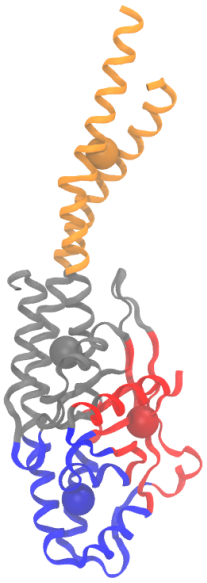


Figure 4.12 Apparent torsional flexibility of *Eh*DF along its principal axis. *Eh*DF complex is divided into small segments of 10 Å each along its principal axis, parallel to the Z axis. Rotations of each segment are calculated relatively to the underlying one along the Z coordinate. Fluctuations are reported as standard deviations, σ (in deg). The spatial profile of the angular standard deviation, σ, clearly demonstrates the heterogeneity of *Eh*DF torsional elasticity with the lowest values found within the region where subunits D and F tightly interact.

In that way, effective torsional elasticity is characterized spatially along the principal axis of the *Eh*DF complex, showing heterogeneity with the lower variability at the core region of subunit D where subunit F forms through extended interactions with subunit F a rigid, bulky body compared with the thinner edges of the complex that make contacts to subunit C of the rotor complex and the A_3B_3 catalytic subunits. Moving towards the tip of the coiled-coil domain, torsional flexibility increases, with the freely fluctuating tip being the most flexible segment.

Although the previous analysis gives some insight into the conformational



Domain	Residues
blue	subD:57-83,118-139 subF:9-19,26-29,78-81
red	subD:84-94,104-117,140-141,144 subF:1-8,20-25,31-33,35-45,70-71
grey	subD:36-56,95-103,142-143,145-161 subF:30,34,50-69,72-77,82-101
orange	subD:7-35,57-83,118-139 subF:9-19,26-29,78-81

Figure 4.13 Essential-dynamics based partitioning of the central rotor axle *Eh*DF. The center of mass of each domain is shown as bead particle overlaid on a cartoon representation of the whole complex (*left*). An exact definition of the domains in terms of residue membership is also given (*right*).

flexibility of the free *Eh*DF complex, it only approximates the complex as a series of segments that rotate along the Z axis, and it neglects bending motions taken into account only implicitly. Similar effective analysis has been presented in (Czub & Grubmuller, 2011 for the homologous γδε subcomplex of F_1-ATPase. To decompose conformational variability into torsional and bending motions, a minimal model of

elasticity has been constructed based on the dynamic-domain analysis of the MD trajectory, with the exception that the boundaries of the dynamic-domain are re-distributed along the Z-axis. Generalized rotation of each segment (dynamic-domain) relatively to the underlying one has been interpreted as a sequence of rotations along the Z-, X- and Y-axes. Z-rotation captures torsional flexibility of the corresponding segment while the other two rotation components capture the bending flexibility. In **Figure 4.14** the results of the rotation analysis are summarized. The free *Eh*DF complex is divided into four coarse dynamic domains with index *i* running from 1 to 4 as discussed in the ***Materials and Methods (Analysis)***. The dynamic domains resulting form the essential-dynamics domain coarse-graining are summarized in **Figure 4.13.** The boundaries of the dynamic domains and residue membership were re-distributed based on their Z-coordinate. The probability distributions of the sampled angles of rotation $\phi_{i,i+1}$, $\psi_{i,i+1}$, $\theta_{i,i+1}$ of segment *i*+1 with reference to the underlying one indexed *i* are also presented. For the dynamic domain *i*=2 (red color) we clearly notice anisotropic bending flexibility while the other two domains are characterized by similar distributions of ψ and ϕ angles. It can be concluded that bending flexibility is not negligible and contributes into total conformational flexibility, especially for the α-helical tip of subunit D which is more elongated and with smaller cross-section.

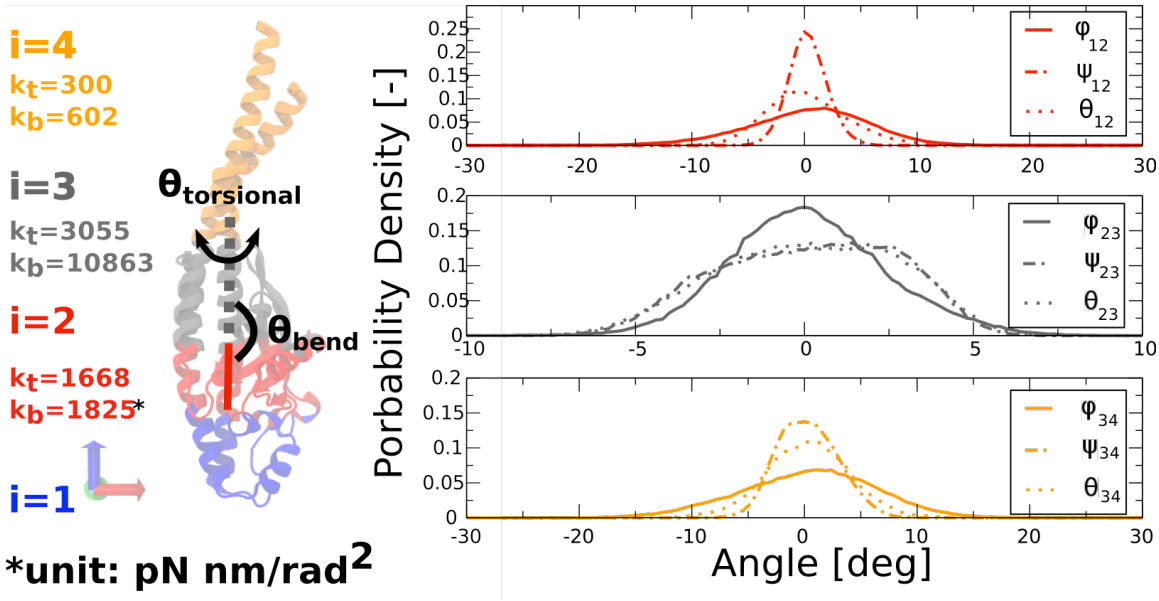


Figure 4.14 Minimal model of elasticity for the *Eh*DF complex. *Eh*DF is divided into four segments along its first principal axis (here parallel to the Z axis). The segments originate from the dynamic domains of **Fig. 4.13,** after redistributing the residues into self-excluded domains along the Z coordinate. Torsion and bending angles of each domain (i=2,3,4) are calculated relatively to the underlying one (i-1) and their 1D probability density functions are shown (*right*). Apparent elastic force constants of bending and torsion, k_b and k_t respectively, were calculated in pN nm/rad^2 units (*left*).

4.4 Conclusions

In the current study we reported results from explicit-solvent atomistic molecular dynamics simulations of the *T. thermophilus* EG complex initiated from different conformers and the central rotor axle *Eh*DF. Available crystallographic data had provided insight into the structural organization of the *Th*EG and the conformational flexibility of that important A-ATPase component. Stock and colleagues (Stewart et al, 2012) have proposed a tilting of the A_3B_3 axis relative to the membrane of 7^o magnitude during catalysis. However, for simplicity the central

rotor was considered rigid and the sequential mechanism described appears to fit better that of a macroscopic machine, rather than a stochastic nano-machine as it the case for rotary ATPases. Crystallographic data, however, due to their static nature, can hide the dynamic nature of the peripheral stator stalk in solution conditions, as clearly revealed by molecular dynamic simulations. Simulations show that the free peripheral stator EG is best represented by a broad continuum of structures, and not by a set of discrete conformers. Three conformers have previously been proposed to explain the dynamics associated with catalytic turnover of the rotary A-ATPase. However, of these three structures we have shown that PS1 is less representative for the wild type sequence regarding local dynamics of the head domain, and simulations started from there remain trapped in a metastable state. These deviations are due to the Leu/Met substitution within the crystal structure or the presence of an unstructured loop in PS1, whereas in PS2 and PS3 this region forms a β-sheet. In contrast simulations, starting from PS2 and PS3 converge rapidly in a broad region of the conformation space.

PS2 represents a state where the coiled coil adopts a bent arrangement that is marginally sampled in the simulations. The PS2 conformation within the holo-complex would result in elastic deformation that could serve as an elastic-energy storage element. The conformational flexibility of the peripheral stator is partly due to the presence of "hinges" within the coiled coil domain and the linkage of the head and coiled coil domains. The large-scale deformations within the peripheral stalk reported recently in (Zhou et al, 2014) in gas-phase short MD simulations and proposed to explain variability of experimental CCS values from ion mobility/mass spectrometry experiments are not represented in our simulations. This is most likely due to the gas phase simulations being conducted in a non-native environment, (no waters, highly charge) in order to best match those found in the mass spectrometry experiment. By conducting our simulations in a water environment we believe they better represent the "native" environment.

To our knowledge the persistence length of the stator elements has not previously been calculated and we show the EG stator to have a value close to those seen in conventional coiled coils (Adamovic et al, 2008; Li et al, 2010; Schwaiger et

al, 2002). The persistence length of 105±42 nm shows a rigid coiled coil element, with the hinge regions allowing for a number of defined states to be sampled. However, as coarse-grain analysis showed, right-handed coiled coil equilibrium mechanical behavior is heterogeneous and exhibits torsional compliance and mechanical behavior under load is expected to be more complicate than of a simple elastic rod.

Explicit-solvent atomistic simulation of the *E. hirae* DF complex, a prototypical central rotor axle of rotary A-ATPases and homologous to γδε subunit of F_1-ATPase. *Eh*DF is involved in the mechanical coupling of the soluble and membrane motor of A-ATPase and responsible for torque transmission during rotary catalysis. Torsional flexibility of γδε complex of F_1-ATPase has been proposed to be part of an elastic-energy buffering mechanism that accounts for the symmetry mismatch of the two motors generally observed in the rotary ATPase family members. Our simulation results allowed detailed analysis of spatial effective torsional flexibility revealing heterogeneity. Binding of F complex on subunit D results in a 'thicker', stiffer region essential for torque transmission. Proper analysis decomposed further observed conformational flexibility into torsional and bending flexibility. A minimal model of elasticity has been employed to allow spatial resolution of elasticity. Heterogeneity is highlighted with the calculation of effective torsional and bending stiffnesses of DF segments. Explicit account of bending flexibility into flexibility of a central DF complex is for the first time reported, to our knowledge in the framework of ion-pumping rotary ATPases in general and A/V-ATPases specifically.

Combined, thermal conformational fluctuations within the rotor and stator of rotary A-ATPases would preserve the same ground state in terms of global shape while accommodating any structural re-arrangements within A_1 motor during rotary catalysis. The true nature of the stators as scaffold or energy storage structures is yet to be conclusively established but this work would suggest a compliant structure with the ability to strain.

Although the current chapter focused on the presentation of results from explicit-solvent atomistic simulations, the microscopy data provided by those simulations is particularly useful for the refinement of simplified models of A-ATPase subunits that would allow the simulation of the intact ATPase with reasonable accuracy to draw biophysical conclusions. Currently, the atomistic information is used for the construction of Elastic Network models for the T. thermophilus EG and E. hirae DF subcomplexes at the residue-level; beads are used to represent the residual COMs. The parameterisation of the elastic network follows the spirit of previously published work on Heterogeneous Elastic Network Models (Zhang et al, 2008a) and re-parameterisation of the ElNeDyn/MARTINI model (Globisch et al, 2013). Various options to model the inter-subunits interactions and their impact on flexibility is tested. Those segments of the T. thermophilus A-ATPase not atavistically resolved are approximated as beads resulting from the application of the same neural-network topology preserving algorithm as used for our ERNM normal mode analysis on **Chapter 3**. Such a methodology is likely to result in a reliable CG model for the intact ATPase augmented with inter-protein plasticity since the subunits are in principle free to move relatively to each other when non-bonded-like interactions are used to describe protein-protein contact energetics. The ensemble obtained from Langevin Dynamics of the CG model will be compared against cryo-EM 2D images of the *T. thermophilus* A-ATPase particles (courtesy of John Rubinstein, University of Toronto) in a methodology similar to the one in (Song et al, 2013). We note that recent advances in the field (Cossio & Hummer, 2013; Jin et al, 2014)will be beneficial to the interpretation of the *T. thermophilus* A-ATPase 2D images.

Chapter 5

Conformational Dynamics and Plasticity of the Eukaryotic V-ATPase Peripheral Stator Stalk

5.1 Introduction

One of the distinctive structural features of the H^+-pumping V-ATPase is the presence of three elongated stalks as components of the peripheral stator connecting the catalytic subunits A/B of the V_1-ATPase motor with the membrane sector of the holoenzyme. The peripheral stalks are heterodimeric complexes of the subunits E and G, homologous to the prokaryotic A-ATPase EG complex (see **Chapter 4** for details). The peripheral stalks make direct contacts with the subunits C and H of the peripheral collar and the transmembrane subunit *a*. Despite its importance as a critical element for the mechanical coupling of the two motors in the V-ATPase, it was only recently that the first high-resolution structure of the EG complex came into light when Wilkens and co-workers (Oot et al, 2012) solved the crystal structure of the yeast EGC complex at 2.90-Å resolution. The overall crystallographic structure of the EG complex matches the corresponding density segments from the low-resolution electron cryomicroscopy 3D reconstructions of the intact V-ATPase from the yeast (Benlekbir et al, 2012) and *M. sexta* (Muench et al, 2009) as well as the *ab-initio* reconstruction from SAXS experiments of the EG complex in solution conditions (Diepholz et al, 2008b).

An interesting structural feature unveiled by the crystallographic studies was the presence of a disordered segment within subunit G (see **Table 5.1** and **Figure 5.1)**. The presence of three glycine residues disfavors the formation of a α-helical segment contrary to the prokaryotic A-ATPase counterpart. Structural disorder has been predicted to be an evolutionary conserved feature among all V-ATPase EG complexes (Oot et al, 2012). The biophysical implications of those structural features are not immediately obvious. Although the observation of structural disorder in the crystal environment implies thermodynamic instability of the helical conformation for the bulge region of subunit G, coexistence of disorder and helical populations cannot be excluded in solution conditions. In addition, a disorder/helical transition could also be possible in response to environmental changes/signals. The EG complex appears to adopt three different conformations within the intact V-ATPase. The observed heterogeneity of the peripheral stalk can be the result of the contacts between the latter and subunits of the peripheral collar. Each copy of EG within the holoenzyme may experience strain of varying magnitude. If EG behaves as an elastic object characterized by a single, free-energy basin, deformations of the EG complex are associated with a specific amount of (deformation) elastic energy, i.e., a free-energy penalty for being restrained in a contact-dictated conformation deviating from the ensemble of conformations populated in solution conditions. On the other hand, if the EG complex is characterized by a complex, multi-basin free-energy landscape, other metastable states can be sampled with relatively lower free-energy cost. It is interesting that a regulatory mechanism exists to control V-ATPase's activity that allows reversible disassembly/assembly of the holoenzyme. The *Regulator of ATPase of Vacuoles and Endosomes* (RAVE) complex is implicated in the Vph1p-containing V-ATPase assembly and activity in yeast (Smardon et al, 2014; Smardon et al, 2002). RAVE complex interacts with the EG subunits, but atomistic understanding of the interaction and regulation is lacking. Finally, alanine-scanning mutagenesis experiments have shown that residue Glu44 of subunit E of the yeast V-ATPase is important for assembly and catalysis efficiency but without providing mechanistic interpretation of those observations (Okamoto-Terry et al, 2013).

In the absence of structural and biophysical experimental data that could help answering the above stated questions, we have employed explicit-solvent molecular dynamics simulations of the yeast EG complex to probe its structural dynamics and stability in fully atomistic detail. An alternative conformer of the EG complex is successfully prepared by means of a molecular-mechanics force field known to favor helical formation. Additional MD simulations were set up to investigate the relative stability of the two conformers. Biased MD simulations initiated from both conformers (Wilkens' and the newly one reported here) were able to fit the EG complex in the corresponding density segments of the yeast V-ATPase cryo-EM reconstruction. The current chapter is organized as follows: details of all molecular dynamics simulations and calculations are given. Then, a newly obtained conformer is described and results from the explicit-solvent molecular dynamics simulations to investigate the relative stability of the Wilkens' and new conformer are presented for near ambient conditions in solution environment. The results from biased MD simulations to obtain quasi-atomistic models of the EG complex in the intact V-ATPase complex are discussed. Finally, a summary of the results and their importance along with directions for future work are provided.

Table 5.1 Amino-acid sequence of the disordered "bulge" of subunit G as solved and discussed in (Oot et al, 2012) (PDB ID: 4DL0). Highlighted is the segment that most disfavours α-helicity in the 'bulge' sequence.

Subunit	**Sequence**
subG	56-FEQKNAGGVGELEK KAEA-75

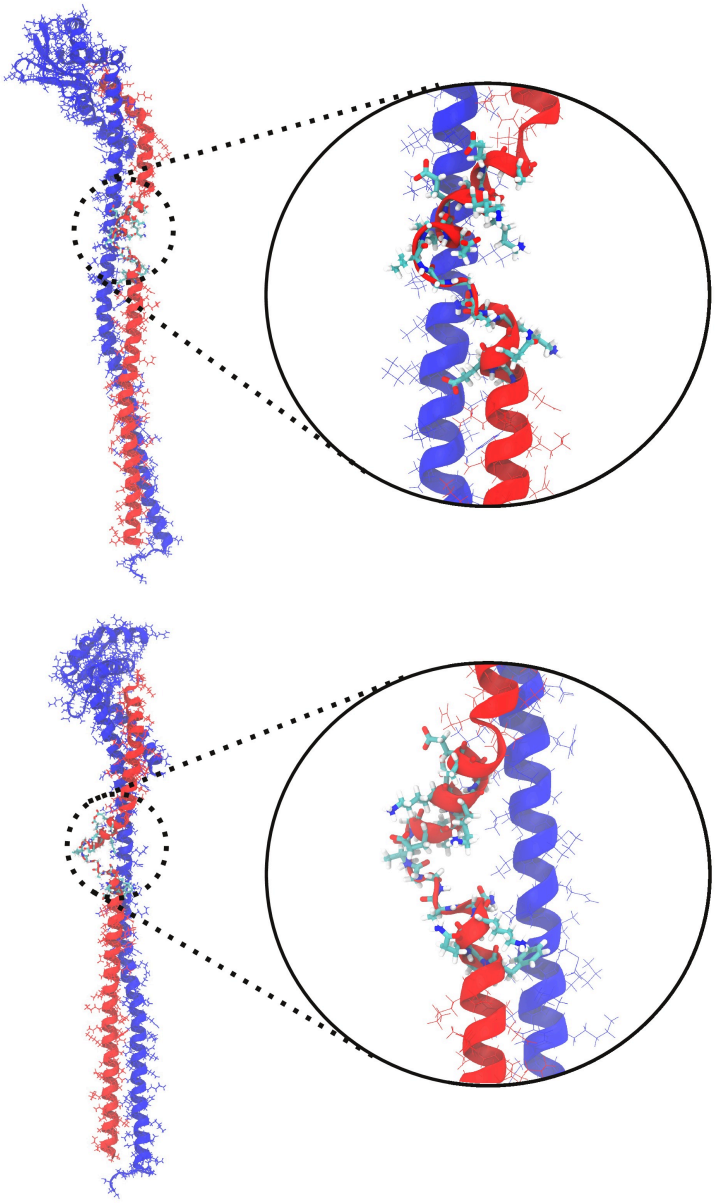


Figure 5.1 Conformation of the yeast EG complex (PDB ID: 4DL0) used to initiate the molecular dynamics simulations reported in the current chapter. Subunits E and G are depicted in blue and red colour respectively. The helical secondary structure of subunit G is interrupted by an unstructured segment-"bulge", a structural feature not observed in the prokaryotic A-type homologous subunit G.

5.2 Materials and Methods

5.2.1 System set-up and MD simulations

Wilkens and colleagues (Oot et al, 2012) reported the crystal structure of the heterotrimeric peripheral stator subcomplex EGC of the yeast rotary V-ATPase at 2.90 Å resolution (PDB ID: 4DL0). The EG complex coordinates served as the initial conformation to set up explicit-solvent atomistic MD simulations. The protein was immersed in an orthorhombic box of TIP3P (Jorgensen et al, 1983) water molecules and the system was neutralized with Na^+ and Cl^- with ionic strength of approximately 140 mM. CHARMM22 all-atom force field (MacKerell et al, 1998) for proteins with CMAP correction (MacKerell et al, 2004b) or CHARMM36 all-atom force field (Best et al, 2012) for proteins were used to model the protein molecules. Periodic boundary conditions were applied and longe-range electrostatics were calculated with the Particle Mesh Ewald summation (Essmann et al, 1995) with a grid spacing < 1 Å. A cut-off distance of 12 Å was used to truncate short-range interactions with a switching function between 10 and 12 Å. Neighbor-atom lists were constructed including all atoms being less than 14 Å away from a given atom.

The system was initially energy-minimized with 10,000 steps of conjugate-gradient minimization. Positional restraints were applied at the backbone atoms of the protein with a force constant of 50 kcal mol^{-1} $Å^{-2}$. Then, the system was heated up from 0 to 310 K using velocity rescaling. After the heating process was complete, the temperature was maintained at 310 K with langevin dynamics, while positional restraints were successively reduced until the protein was finally unrestrained. Equilibrium MD runs were performed in the NPT thermodynamic ensemble. Langevin dynamics (Brunger et al, 1984) with a damping constant of 1 ps^{-1} was employed to maintain a temperature of 310 K. Pressure control was achieved with a Langevin piston barostat (Feller et al, 1995) at 1 bar with a period of 200 fs and a decay rate of 100 fs. All covalent bonds involving hydrogens were constrained with the RATTLE algorithm (Andersen, 1983). An integration timestep of 2 fs was adopted with the short-range interactions being evaluated every time-step and the

long-range electrostatics being calculated every second time-step. MD simulations were carried out with NAMD2.8 (Phillips et al, 2005) biomolecular simulation package.

In addition to unbiased explicit-solvent MD simulations of the EG complex, biased atomistic simulations in the generalized Born implicit solvent (Tanner et al, 2011) and CHARMM22/CMAP force field for protein were carried out using the MDFF (Trabuco et al, 2008) technique to fit the EG complex into the corresponding segments of the 11-Å electron cryo-microscopy 3D reconstruciton of the yeast V-ATPase (EMDB_ID: 5476) (Benlekbir et al, 2012). A summary of all simulations carried out is shown in **Table 5.2**.

Table 5.2 Summary of all simulations carried out for the yeast EG peripheral stalk.

Simulation ID	Comment	Simulation Time
yEGw-C27	unbiased simulation of yeast EG (yEG) with C27	300 ns
yEGw	unbiased simulation of yEG with C36 starting from Wilkens' structure	300 ns
yEGh	unbiased simulation of yEG starting from 'helical' bulge resulting from *sim1*	300 ns
mdff-PS1	biased MD flexible fitting in combined map segments 08+11 of EMDB_ID: 5476 map starting from Wilkens' structure	50 ns
mdff-PS2	biased MD flexible fitting in combined map segments 09+12 of EMDB_ID: 5476 map starting from Wilkens' structure	50 ns
mdff-PS3	biased MD flexible fitting in combined map segments 10+13 of EMDB_ID: 5476 map starting from Wilkens'	50 ns

	structure.	
mdff_h-PS1	biased MD flexible fitting in combined map segments 08+13 of EMDB_ID: 5476 map starting from the 'helical' structure.	50 ns
mdff_h-PS2	biased MD flexible fitting in combined map segments 09+12 of EMDB_ID: 5476 map starting from the 'helical' structure.	50 ns
mdff_h-PS3	biased MD flexible fitting in combined map segments 10+13 of EMDB_ID: 5476 map starting from 'helical' structure.	50 ns

5.2.2 Analysis

Average structures were calculated after aligning the protein by minimizing the C_α-based root-mean-square deviations from the initial conformation. The covariance matrix C_{ij} of coordinates x_i and x_j was calculated and diagonalised. The resulting eigenvectors or principal modes were sorted by decreasing eigenvalue. Trajectories were projected on the first three eigenvectors to get the PC coordinates q_i. All PCA calculations were performed with WORDOM (Seeber et al, 2011).

5.3 Results and Discussion

5.3.1 Conformational dynamics of the peripheral-stator stalk's intrinsically disordered region

As previously mentioned, subunit G has an intrinsically disordered region (IDR) in striking contrast to the homologous subunit of A-type peripheral stalks. Intrinsically

disordered proteins (IDPs) or intrinsically disordered regions (IDR) are implicated in molecular recognition and signaling, mediated by modulation of the conformational free-energy landscape in response to environmental stimuli, e.g., disorder-to-order transitions (Moritsugu et al, 2012; Sigalov et al, 2008). It is unknown whether such a mechanism exists for the eukaryotic EG subcomplex. However, the presence of conformational plasticity would provide a mechanistic explanation of the EG involvement in V-ATPase regulation of assembly/disassembly through interactions with the RAVE complex. In addition, it is unclear how the unstructured segment can affect the mechanical properties of the coiled-coil tail domain and how the latter correlate with the characteristic stoichiometry of V-ATPases, i.e., having three copies of EG instead of the two copies in A-ATPases. To interrogate the dynamics of the whole heterocomplex and particularly the structural dynamics of the bulge region, unbiased MD simulations in explicit solvent were set up using the all-atom CHARMM22/CMAP force field for proteins starting from the Wilkens' conformation (**yEGw-C27**, see **Table 5.2**). Its important to note that the CHARMM22/CMAP force field has a helical bias and so is not accurate enough to study the free-energy landscape of disordered proteins/peptides. However, in this case the application of such a force field serves several purposes. Firstly, it can investigate whether using such a force field for a large protein with IDR is able to alter the conformational state of the segment against the intermolecular interactions with the ret of the protein and any topological constraint. Secondly, any observation of disorder-to-order transition within the complex would suggest that such a structural transition is feasible under the appropriate environmental conditions. In other words, it is feasible that enhancement of the helical propensity of the bulge region, either directly (ligand or protein binding) or indirectly (topological constraints within the intact V-ATPase), would allow the transition to well-defined secondary structure with the rest of the proteins accommodating such a transition. In that case CHARMM22/CMAP mimics the role of such a helical-enhancing environmental stimuli.

In **Figure 5.2**, the helical content of the bulge region is calculated as the percentage of residues being in the α-helical configuration as calculated with

WORDOM. Within the first 25 ns the segment folds into a α-helical state. The relatively short timescale of the transition reflects the strong helical bias introduced from the molecular-mechanics force field CHARMM22/CMAP. However, interestingly the hetero-complex easily accommodates the observed structural changes suggesting the feasibility of such a transition and the corresponding structural re-organisation.

Next, to test the stability/metastability of the new conformation a separate MD simulation was carried out starting from an EG conformation at 300 ns of the yEGw-C27 trajectory. To eliminate the helical-bias, the all-atom CHARMM32 force field was employed that has been shown to be more appropriate for intrinsically disordered proteins and regions. In **Figure 5.3**, the helical content has been calculated as previously for *yEGw (blue)* and *yEGh (orange)* for a total time of 300 ns.

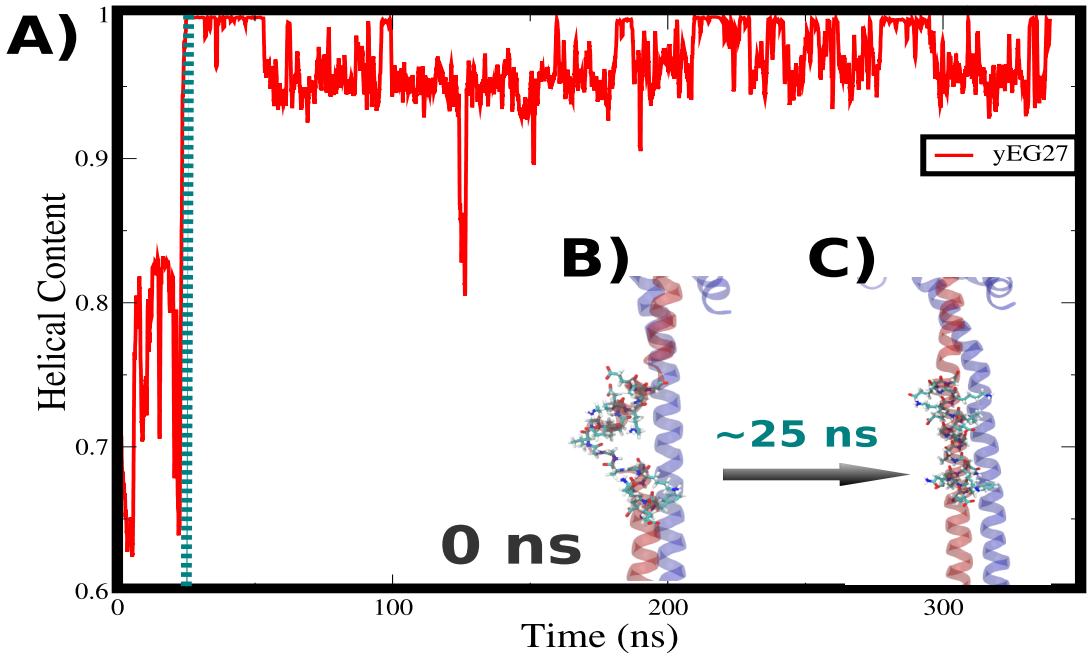


Figure 5.2 Temporal evolution of the *helical content* of the "bulge" sequence along the yEG-C27 trajectory starting from the *Wilkens'* structure. The disordered segment undergoes a *disorder-to-order* transition into a helical conformation resulting in a local and global conformational change.

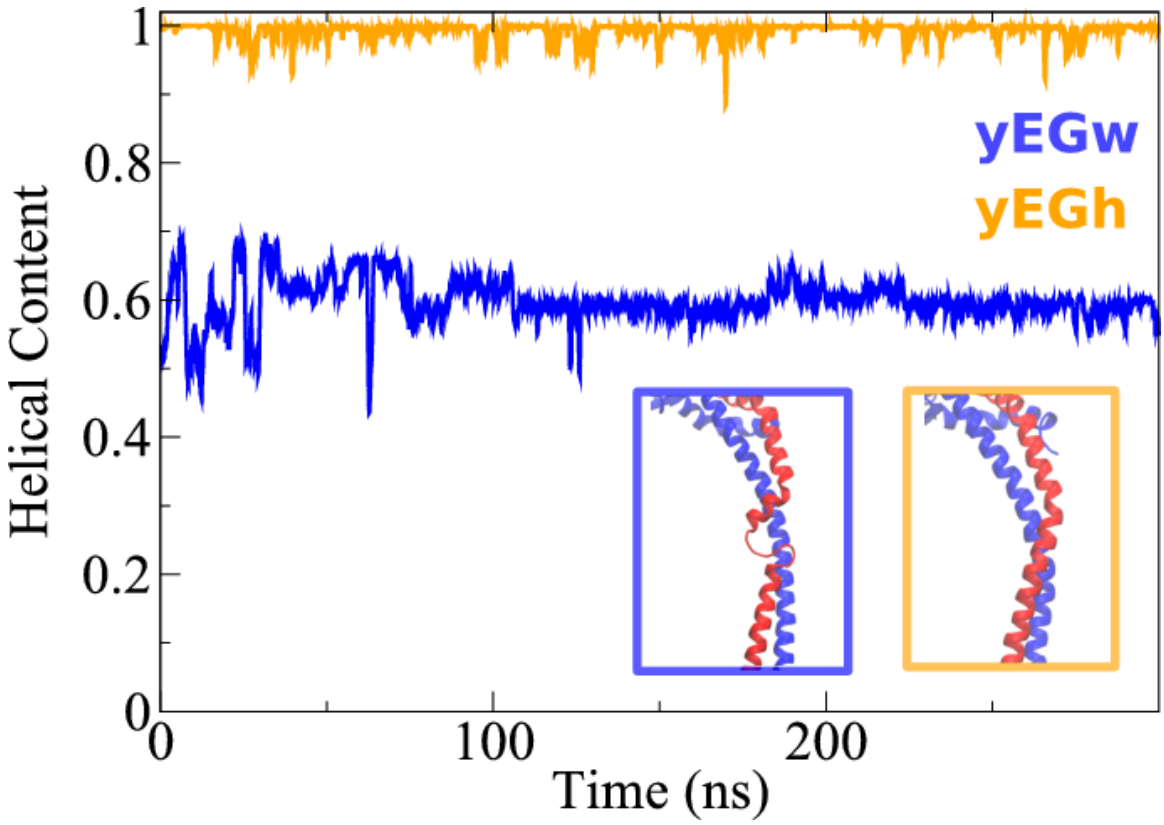


Figure 5.3 Temporal evolution of the helical content of the 'bulge' region as calculated for the trajectories *yEGw* (blue) and *yEGh* (orange) for a total simulation time of 300 ns. Within the simulated timescale both structural states ('disordered/helical') appear to be (meta-) stable.

5.3.2 Principal Component Analysis of the EG complex in near-ambient conditions in solution environment

As in **Chapter 4,** principal component analysis (PCA) is applied as a dimensionality-reduction tool for the analysis of the sampled molecular-dynamics trajectories. In particular, application of PCA on the last 150 ns of the *yEGw* and *yEGh* trajectories using only C_α atoms resulted in the projection of each trajectory in a low-dimensional (sub-) space. The empirical free-energy landscape (or PMF) on the first (PC1) and second (PC2) eigenvectors provides a simplified representation of the

conformational space as depicted in **Figures 5.4 & 5.5**. In addition, the reference structures (*yEGw* and *yEGh*, the initial conformations of the respective trajectories) were also projected on the PC1/PC2 subspace, marked with a *green circle* and *pink triangle* marker respectively. In both cases we observe increased complexity of the low-dimensional space compared to the prokaryotic A-ATPase counterpart.

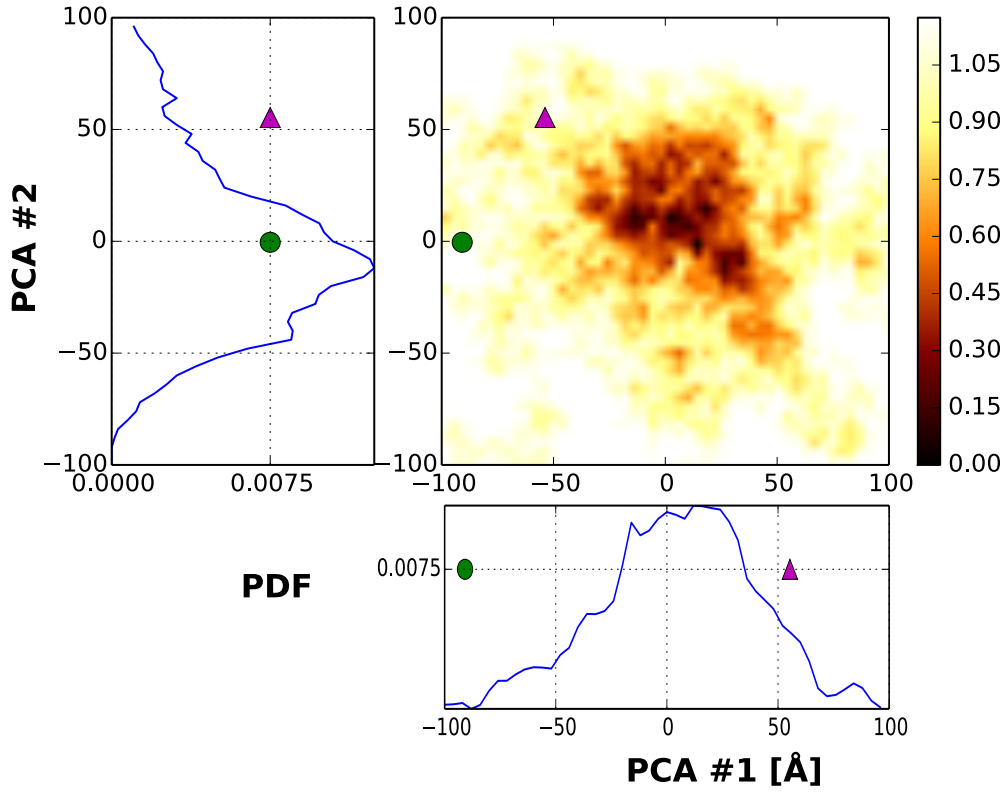


Figure 5.4 Principal component analysis of the last 150 ns of the *yEGw* simulation. The trajectory is projected on the first two PCA eigenvectors that describe most of the conformational variability. The empirical potential-of-mean-force, PMF, has been calculated in k_BT units for the low-dimensional space. For clarity the probability density of each PCA collective coordinate is also depicted. The corresponding projections of the reference structures of *yEGw* and *yEGh* are shown as a *circle* and *triangle* marker respectively.

The distribution of the projections deviates significantly from the Gaussian distribution. The motions are highly anharmonic with the presence of many local minima, separated by low-energy barriers as suggested from the small values of the empirical PMF. The large range of values in the PCA coordinates is related to the elongated shape of the complex. Both reference structures (Wilkens' and helical conformers) in **Figure 5.4** reside in low-populated areas of the landscape underlying the dynamic nature of the complex and suggesting deviations from the in-solution ensemble. **Figure 5.5** provides indications of bistability in the distribution of the first projection. The reference conformer *yEGh* resides near a highly-populated region while the Wilkens' *yEGw* conformer essentially is not populated.

We should note that the PCA eigenvectors and the associated free-energy landscape are only approximate and they mainly serve as abstract collective coordinates to allow the characterisation and comparison of conformational states. The two data sets have not been reweighted to obtain an accurate free-energy surface. However, the results of PCA provide the starting point for future rigorous and computationally expensive free-energy calculations on the eigenvector-based conformational space of the EG complex. Specifically, the approximate, rugged landscape spanned from the first two eigenvectors can be reweighted/refined with umbrella sampling (US) simulations. The computational burden of those calculations can be alleviated with the application of the *adaptive* version of US simulations. Alternatively, *Adaptive Biasing Force* (ABF) simulations could be used rather than the US technique. However, the abstract mathematical nature of the PC analysis does not provide an intuitive picture of the behavior of the EG complex.

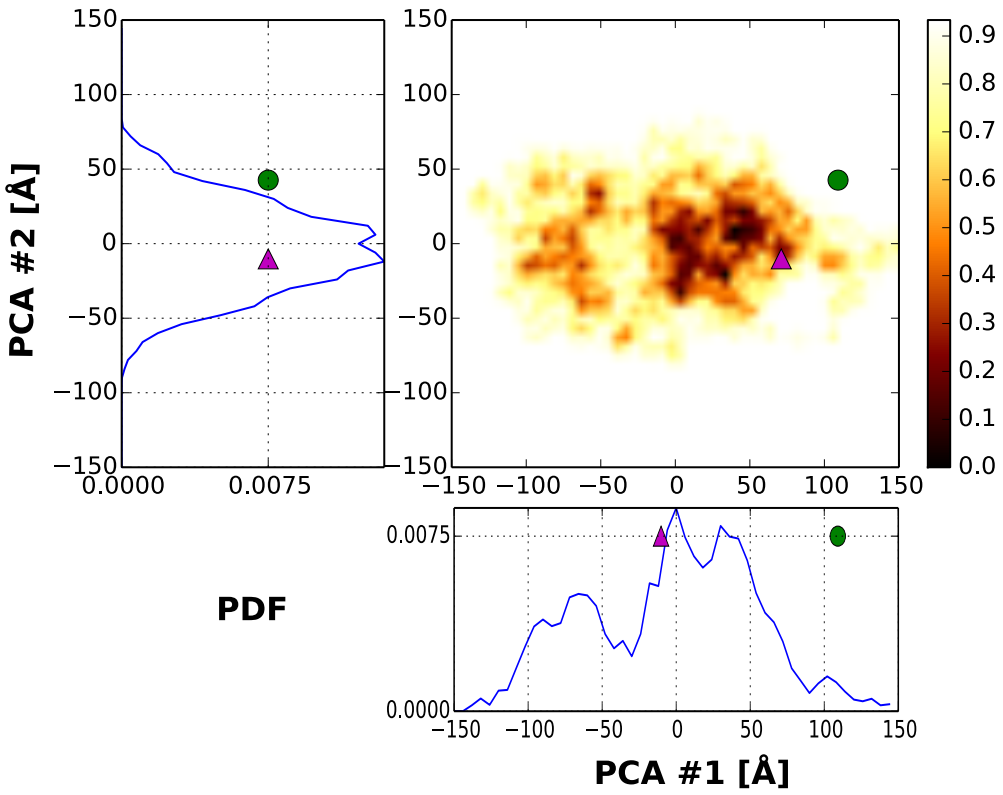


Figure 5.5 Principal component analysis of the last 150 ns of the *yEGh* simulation similarly to **Figure 5.4**. We observe that the PC1 collective variable is characterised by a more complicate bimodal distribution compared to the *yEGw* simulation and the Wilkens' conformer is only marginally sampled.

5.3.3 Molecular-Dynamics flexible fitting of the peripheral-stator stalk EG in electron cryo-microscopy 3D reconstruction of the intact yeast V-ATPase

To provide quasi-atomistic models for the three peripheral stalks in the intact yeast V-ATPase, biased MD simulations were carried out using the corresponding 3D densities from the 3D cryoEM reconstruction as *restraints* within the framework of the *Molecular-Dynamics Flexible Fitting* (MDFF) technique (see **Chapter 2** for details). In MDFF simulations a density map acts as an external biasing field, implemented as a 3D grid that exerts force on (specified) atoms of the biomolecular system. There are three density segments in the 3D reconstruction of the yeast V-ATPase (EMDB_ID: 5476), each adopting a different global conformation. To enhance conformational sampling and reduce computational cost, the Generalized-Born implicit-solvent model -as implemented in NAMD -was used instead of an explicit representation for water molecules; for the protein the same level of detail was retained as in the case of explicit-solvent simulations. The final conformations of the EG complex at the end of the 50-ns MDFF simulations are shown in **Figure 5.6**. To distinguish between the three peripheral-stalks, we adopt the name convention used in the work of (Oot et al, 2012). Each peripheral stalk is named as EG*X*, where *X* runs from 1 to 3 (**Figure 5.6A**). The correspondence between the peripheral stalks and the masked segments of the yeast density map is the following: EG1 corresponds to masked segments 8 & 11, EG2 to the masked segments 9 & 12 and finally EG3 corresponds to the masked segments 10 & 13. Overall, the resulting conformations of the EG complex for each density segment are very similar globally with the exception of the local structure of the bulge region. We note that in MDFF applications it is common practice to apply harmonic restraints to maintain stereochemistry and secondary-structure against the application of large map-bias forces. In our case, no harmonic restraints have been applied within the bulge region, therefore that segment is able to undergo structural changes such as

an order-disorder transition. As indicated in **Figure 5.6B,** local 'cracking' occurs in simulations *mdff_h-PS{1,3}***,** starting from the *helical-bulge* conformation.

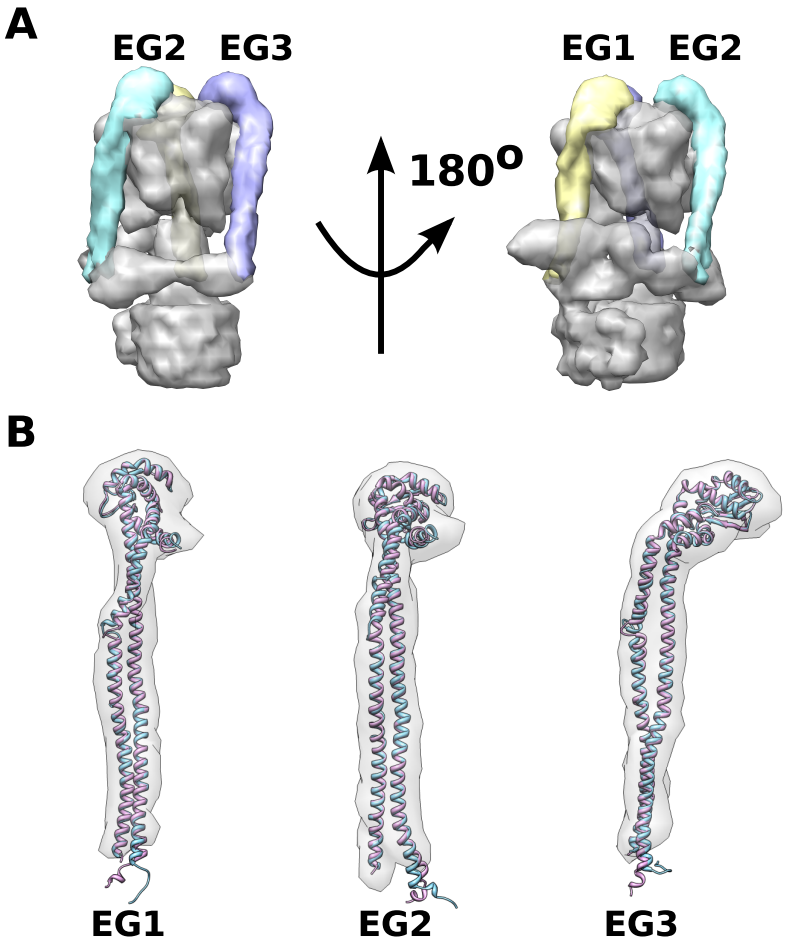


Figure 5.6 Structural models of the three EG complexes resulting from implicit-solvent MDFF simulations. (A) Visualization of the 3D density map of the intact yeast V-ATPase reconstructed at 11-Å resolution. The density segments that correspond to the three peripheral stalks are shown is different color from the rest ones. The same name convention (EG1, EG2, EG3) for each of the three EG complexes is used as in (Oot et al, 2012). (B) The conformation of EG complex fitted in the density segments of the three peripheral stalks are shown, resulting from two MDFF simulations, one started from the Wilkens' conformer (*cyan* colour) and the other from the helical conformer (*purple* colour).

Notably, in *mdff_h-PS2* the bulge helicity is preserved and we do not observe a disorder-helical transition in the *mdff-PS2* started from Wilken's structure. The three peripheral stators likely represent mechanically distinct states with enhanced local strain in EG1 and EG3 at the local bulge region. Further investigations are needed to assess the robustness of the results using explicit-solvent MD simulations and additional, independent methodologies to model the density segments. Another interesting observation, summarised in **Figure 5.7,** is that for the density segments EG2/EG3 the trajectories started from the 'helical' reference structure require smaller global conformational changes compared to those started from the Wilken's reference structure. EG1 stalk deviates equally from both reference structures. We note that RMSD values reported exclude the bulge region and the floppy N-terminal region of the subunit E, therefore they represent a (low-resolution) measure of global conformation changes. It is intriguing that the 'helical' reference conformer shows structural pre-organisation for EG2/EG3 positions and requires smaller structural changes to be incorporated inside the assembled complex. Provided that the RAVE complex is implicated in the assembly of the yeast holoenzyme through direct interactions with the E and G subunits, we hypothesise that RAVE modulates the in-solution conformational landscape of the EG complex stabilizing metastable states that fit into the holoenzyme (*population-shift mechanism*). The feasibility of the population-shift mechanism will be the focus of future computational and experimental investigations.

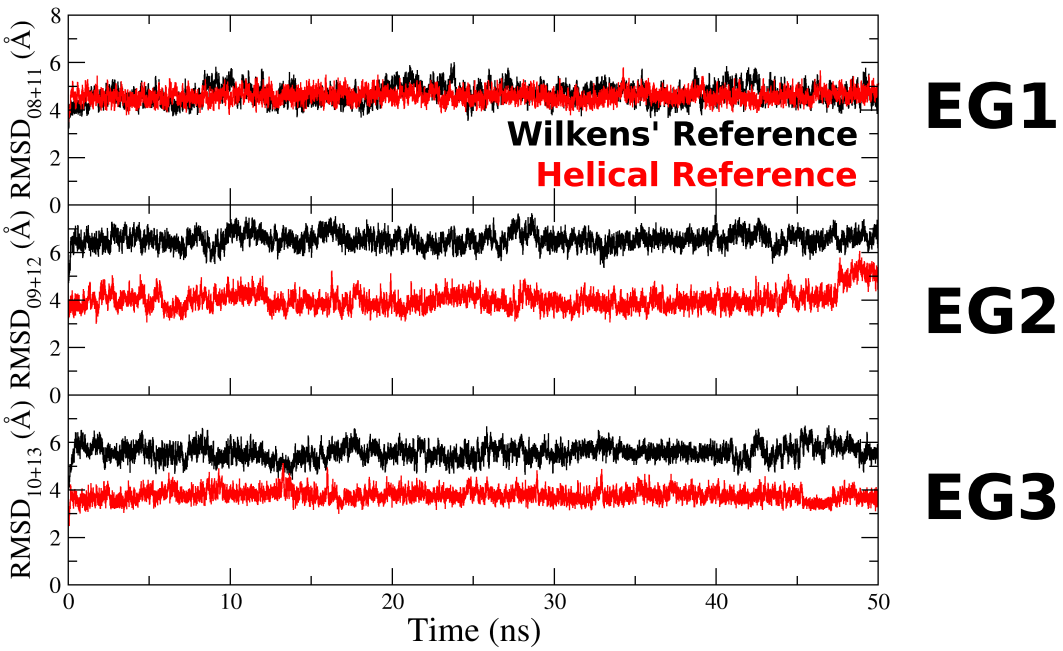


Figure 5.7 Time series of RMSD values during the course of MDFF simulations for a total time of 50 ns. Each panel refers to one of the three peripheral stalks of the yeast V-ATPase. Interestingly, the alternative helical conformer reported in the current chapter undergoes smaller-scale global structure re-arrangements compared to the Wilkens' conformer to fit into the EG2 and EG3 segments. For the EG1 segment, both conformers need to undergo similar-scale global conformational changes (albeit following different pathways) in order to fit inside the V-ATPase holoenzyme at that particular position.

5.4 Conclusions

In the current chapter, we reported the first atomistic molecular dynamics simulations of the eukaryotic peripheral stalk EG of the yeast V-ATPase in a realistic solution environment providing high-resolution information for the structural dynamics of the whole hetero-complex and the bugle region in particular. Although

crystallographic studies and 3D cryo-EM reconstructions have provided some insight into the structure of the EG stator, our calculations represent the first investigations for the free EG complex of the V-ATPase. Understating how EG behaves as an isolated complex in solution is essential to understanding the mechanical state of the EG complex within the holoenzyme and whether any of the three peripheral stators of the V-ATPase is under "*mechanical stress*". Our results clearly show that the Wilkens' conformer solved by X-ray crystallography globally deviates from the sampled in-solution conformations. An alternative conformer, where the bulge region adopts a helical structure, has also been obtained with a molecular-mechanics force field with known helical bias. The helical conformer samples different global states compared to the Wilken's structure. Unbiased MD simulations stating from both conformers and with a molecular-mechanics force field with reduced helical bias suggest that both states (disordered and helical bulge region) are metastable for the time scale studied (300 ns). Although no accurate free-energy differences can be reported from our study, our results suggest that the free-energy landscape of the eukaryotic EG complex is more complex than its prokaryotic counterpart with more than one metastable states. Environmental conditions and external stimuli (e.g., interaction with the RAVE complex) could modulate the free-energy landscape stabilising particular states. Implicit-solvent MD flexible fitting the EG complex in the corresponding density segments of the three peripheral stalks of the yeast V-ATPase starting from both reference structures suggest that the three stalks represent distinct mechanical states with the bulge region likely experiencing different levels of mechanical stress. Notably, starting from the newly proposed helical conformer, smaller scale global structural re-arrangements are necessary to dock the peripheral stalk inside the intact V-ATPase for positions EG2 and EG3 compared to the Wilkens' conformer. We hypothesise that RAVE-assisted V-ATPase assembly could possibly occur through the stabilisation of helical-conformer-like conformations resembling a population-shift paradigm.

Summary

Conclusions and Future Work

The focus of the *Thesis* was the investigation of the structural dynamics of the peripheral stator stalk and central rotor axle subunits as determinants of the observed conformational flexibility of the A- and V-ATPases, members of the family of ion-pumping rotary ATPases. The direct and/or indirect effect of conformational flexibility on the function of rotary ATPases is not clear. We believe that understanding the molecular origin of flexibility is a prerequisite for the elucidation of the structure-function relationship in that family.

In *Chapter* 3 we presented for the first time results from combined *experimental* and *computational* work that demonstrated the presence of conformational flexibility within single particles of the yeast and *Manduca sexta* V-ATPase. Analysis of 2D electron-microscopy images and normal mode calculations of an elastic-network model for biomolecular complexes suggest that the observed conformational flexibility is a general feature of V-ATPases. Calculated anisotropic bending modes match those identified in the EM data sets, while a torsional mode of motion captured from the elastic-network model cannot be observed by electron microscopy. The relative success of such a simple model for protein mechanics is due to its ability to capture the effects of topological constraints on protein mechanics and flexibility. However, in order to investigate the contribution of other factors in the observed flexibility of ATPases, we devised and discussed a general

multiscale, hybrid framework where molecular models of different resolution are combined to predict and characterise the mechanical properties of biomolecular complexes and compare the computational results with experimental data.

After setting up the general computational framework to study protein mechanics, we presented results from detailed atomistic molecular dynamics simulations of those subunits of the A- and V-ATPase that are critical for the mechanical coupling of the two molecular motors (soluble chemical motor and membrane electrochemical motor). In particular, *Chapter* 4 is devoted to the detailed study of the peripheral stator stalk and central rotor axle of the A-ATPase. Molecular dynamics simulations starting from different initial conformations showed that the peripheral stator explored a broad free-energy basin where initial conformers represent microstates residing in low-probability regions of the free-energy surface. The application of a coarse-grained, geometrically intuitive analysis showed the presence of flexibility heterogeneity within the hetero-complex and of two hinge regions that allow the peripheral stator to accommodate large-scale conformational changes of the intact ATPase during rotary catalysis. We also presented for the first time an estimation of the persistence length of the right-handed coiled coil (rhcc) `tail` region of the peripheral stalk. The semi-flexible rhcc tail allows the peripheral stalk to be stiffer compared to the central rotor axle, a feature suggested to be important for the kinetic efficiency of the F-ATP synthase. Molecular dynamics simulations of the central rotor axle provided a detailed picture of the torsional heterogeneity within the EG complex similarly to its F-type counterpart.

In *Chapter* 5 we carried out similar dynamical studies for the yeast V-ATPase peripheral stator stalk. An intriguing structural feature of the eukaryotic V-type EG complex is the presence of a disordered segment-referred to as the *bulge* region-contrary to its prokaryotic counterpart. To investigate the structural plasticity of that region, we prepared a new, helical-bulge conformer starting from the crystallographic Wilkens' conformer using a molecular-mechanics force field known to exhibit helical bias to accelerate the transition to unexplored regions of the conformational space. To assess the feasibility and relative stability of the new

conformer, we carried out unbiased MD simulations with a more accurate force field with reduced helicity bias and we observe that both states with helical and disordered bulge region are stable within the time of the trajectories. Although estimation of relative free energy differences between the two states would demand the application of specialised techniques, out results demonstrate that both states are feasible and structural plasticity maybe related to RAVE-assisted assembly/disassembly of V-ATPases. We also present results from biased MD simulations to obtain quasi-atomistic models of the three peripheral stators of the yeast V-ATPase from a low-resolution 3D electron-microscopy data. Interestingly, for two out of three density segments (EG2 and EG3) starting from the new conformer requires smaller-scale conformational re-arrangement to fit the density.

Overall, our investigations constitute the first stage of our proposed computational/experimental framework employing molecular dynamics simulations of isolated subunits to obtain atomistic-resolution insight into their conformational variability and flexibility. In the future, our results will provide the starting point to develop coarse-grain models that would allow dynamical simulation of intact A/V-ATPase particles and direct comparison with data from electron microscopy and/or force spectroscopy experiments.

References

Adamovic I, Mijailovich SM, Karplus M (2008) The elastic properties of the structurally characterized myosin II S2 subdomain: a molecular dynamics and normal mode analysis. *Biophys J* **94:** 3779-3789

Andersen HC (1983) RATTLE-a Velocity Version of the SHAKE Algorithm for Molecular-Dynamics Calculations. *J Comput Phys* **52:** 24-34

Arai S, Saijo S, Suzuki K, Mizutani K, Kakinuma Y, Ishizuka-Katsura Y, Ohsawa N, Terada T, Shirouzu M, Yokoyama S, Iwata S, Yamato I, Murata T (2013) Rotation mechanism of Enterococcus hirae V_1-ATPase based on asymmetric crystal structures. *Nature* **493:** 703-707

Arkhipov A, Freddolino PL, Imada K, Namba K, Schulten K (2006a) Coarse-grained molecular dynamics simulations of a rotating bacterial flagellum. *Biophys J* **91:** 4589-4597

Arkhipov A, Freddolino PL, Schulten K (2006b) Stability and dynamics of virus capsids described by coarse-grained modeling. *Structure* **14:** 1767-1777

Atilgan AR, Durell SR, Jernigan RL, Demirel MC, Keskin O, Bahar I (2001) Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophys J* **80:** 505-515

Ayton GS, Noid WG, Voth GA (2007) Multiscale modeling of biomolecular systems: in serial and in parallel. *Curr Opin Struct Biol* **17:** 192-198

Bahar I, Atilgan AR, Erman B (1997) Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Fold Des* **2:** 173-181

Bahar I, Lezon TR, Bakan A, Shrivastava IH (2010) Normal mode analysis of biomolecular structures: functional mechanisms of membrane proteins. *Chem Rev* **110:** 1463-1497

Bahar I, Rader AJ (2005) Coarse-grained normal mode analysis in structural biology. *Curr Opin Struct Biol* **15:** 586-592

Baker JL, Voth GA (2013) Effects of ATP and actin-filament binding on the dynamics of the myosin II S1 domain. *Biophys J* **105:** 1624-1634

Baker T, Cheng R (1996) A model-based approach for determining orien- tations of biological macromolecules imaged by cryoelectron microscopy. *J Struct Biol* **116:** 120-130

Bathe M (2008) A finite element framework for computation of protein normal modes and mechanical response. *Proteins* **70:** 1595-1609

Benlekbir S, Bueler SA, Rubinstein JL (2012) Structure of the vacuolar-type ATPase from Saccharomyces cerevisiae at 11-Å resolution. *Nat Struct Mol Biol* **19:** 1356-1362

Bernal RA, Stock D (2004) Three-dimensional structure of the intact Thermus thermophilus H^+-ATPase/synthase by electron microscopy. *Structure* **12:** 1789-1798

Best RB, Chen YG, Hummer G (2005) Slow protein conformational dynamics from multiple experimental structures: the helix/sheet transition of arc repressor. *Structure* **13:** 1755-1763

Best RB, Zhu X, Shim J, Lopes PE, Mittal J, Feig M, Mackerell AD, Jr. (2012) Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone phi, psi and side-chain chi_1 and chi_2 dihedral angles. *J Chem Theory Comput* **8:** 3257-3273

Birmanns S, Rusu M, Wriggers W (2011) Using Sculptor and Situs for simultaneous assembly of atomic components into low-resolution shapes. *J Struct Biol* **173:** 428-435

Brink J, Ludtke SJ, Kong Y, Wakil SJ, Ma J, Chiu W (2004) Experimental verification of conformational variation of human fatty acid synthase as predicted by normal mode analysis. *Structure* **12:** 185-191

Brooks B, Bruccoleri R, Olafson B, States D, Swaminathan S, Karplus M (1983) CHARMM: A Program for Macromolecular Energy, Minimization, and Molecular Dynamics Calculations. *J Comput Chem* **4:** 187-217

Brooks BR, Brooks CL, Mackerell AD, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodoscek M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor RW, Post CB, Pu JZ, Schaefer M, Tidor B, Venable RM, Woodcock HL, Wu X, Yang W, York DM, Karplus M (2009) CHARMM: The Biomolecular Simulation Program. *J Comput Chem* **30:** 1545-1614

Brunger A, Brooks CL, Karplus M (1984) Stochastic Boundary-Conditions for Molecular-Dynamics Simulations of ST2 Water. *Chem Phys Lett* **105:** 495-500

Chacon P, Wriggers W (2002) Multi-resolution contour-based fitting of macromolecular structures. *J Mol Biol* **317:** 375-384

Chen X, Cui Q, Tang Y, Yoo J, Yethiraj A (2008) Gating mechanisms of mechanosensitive channels of large conductance, I: a continuum mechanics-based hierarchical framework. *Biophys J* **95:** 563-580

Cherepanov DA, Mulkidjanian AY, Junge W (1999) Transient accumulation of elastic energy in proton translocating ATP synthase. *FEBS Lett* **449:** 1-6

Cheung MS, Garcia AE, Onuchic JN (2002) Protein folding mediated by solvation: water expulsion and formation of the hydrophobic core occur after the structural collapse. *Proc Natl Acad Sci USA* **99:** 685-690

Christen M, Hunenberger PH, Bakowies D, Baron R, Burgi R, Geerke DP, Heinz TN, Kastenholz MA, Krautler V, Oostenbrink C, Peter C, Trzesniak D, van Gunsteren WF (2005) The GROMOS software for biomolecular simulation: GROMOS05. *J Comput Chem* **26:** 1719-1751

Chu JW, Voth GA (2005) Allostery of actin filaments: molecular dynamics simulations and coarse-grained analysis. *Proc Natl Acad Sci USA* **102:** 13111-13116

Chu JW, Voth GA (2006) Coarse-grained modeling of the actin filament derived from atomistic-scale simulations. *Biophys J* **90:** 1572-1582

Cornell WD CP, Bayly CI, Gould IR, Merz KM Jr, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA (1995) A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J Am Chem Soc* **117:** 5179-5197

Cossio P, Hummer G (2013) Bayesian analysis of individual electron microscopy images: towards structures of dynamic and heterogeneous biomolecular assemblies. *J Struct Biol* **184:** 427-437

Czub J, Grubmuller H (2011) Torsional elasticity and energetics of F_1-ATPase. *Proc Natl Acad Sci USA* **108:** 7408-7413

Dickson VK, Silvester JA, Fearnley IM, Leslie AG, Walker JE (2006) On the structure of the stator of the mitochondrial ATP synthase. *EMBO J* **25:** 2911-2918

Diepholz M, Borsch M, Bottcher B (2008a) Structural organization of the V-ATPase and its implications for regulatory assembly and disassembly. *Biochem Soc Trans* **36:** 1027-1031

Diepholz M, Venzke D, Prinz S, Batisse C, Florchinger B, Rossle M, Svergun DI, Bottcher B, Fethiere J (2008b) A different conformation for EGC stator subcomplex in solution and in the assembled yeast V-ATPase: possible implications for regulatory disassembly. *Structure* **16:** 1789-1798

Doruker P, Atilgan AR, Bahar I (2000) Dynamics of proteins predicted by molecular dynamics simulations and analytical approaches: application to alpha-amylase inhibitor. *Proteins* **40:** 512-524

Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG (1995) A Smooth Particle Mesh Ewald Method. *J Chem Phys* **103:** 8577-8593

Feller SE, Zhang YH, Pastor RW, Brooks BR (1995) Constant-Pressure Molecular-Dynamics Simulation - the Langevin Piston Method. *J Chem Phys* **103:** 4613-4621

Frank J (2006) Three-dimensional electron microscopy of macromolecular assemblies, New York: Oxford University Press.

Frank J, Spahn C (2006) The ribosome and mechanism of proteins synthesis. *Rep Prog Phys* **69:** 1183-1417

Gibbons C, Montgomery MG, Leslie AG, Walker JE (2000) The structure of the central stalk in bovine F_1-ATPase at 2.4-Å resolution. *Nat Struct Biol* **7:** 1055-1061

Gibbons MM, Klug WS (2008) Influence of nonuniform geometry on nanoindentation of viral capsids. *Biophys J* **95:** 3640-3649

Giraud MF, Paumard P, Sanchez C, Brethes D, Velours J, Dautant A (2012) Rotor architecture in the yeast and bovine F_1-c-ring complexes of F-ATP synthase. *J Struct Biol* **177:** 490-497

Globisch C, Krishnamani V, Deserno M, Peter C (2013) Optimization of an elastic network augmented coarse-grained model to study CCMV capsid deformation. *PloS One* **8:** e60582

Go N, Taketomi H (1979a) Studies on protein folding, unfolding and fluctuations by computer simulation. III. Effect of short-range interactions. *Int J Pept Protein Res* **13:** 235-252

Go N, Taketomi H (1979b) Studies on protein folding, unfolding and fluctuations by computer simulation. IV. Hydrophobic interactions. *Int J Pept Protein Res* **13:** 447-461

Gohlke H, Thorpe MF (2006) A natural coarse graining for simulating large biomolecular motion. *Biophys J* **91:** 2115-2120

Grubisic I, Shokhirev MN, Orzechowski M, Miyashita O, Tama F (2010) Biased coarse-grained molecular dynamics simulation approach for flexible fitting of X-ray structure into cryo electron microscopy maps. *J Struct Biol* **169:** 95-105

Hoang TX, Cieplak M (2000) Molecular dynamics of folding of secondary structures in Go-like models of proteins. *J Chem Phys* **112:** 6851–6862

Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. *J Mol Graph* **14:** 33-38, 27-38

Hyeon C, Dima RI, Thirumalai D (2006a) Pathways and kinetic barriers in mechanical unfolding and refolding of RNA and proteins. *Structure* **14:** 1633-1645

Hyeon C, Lorimer GH, Thirumalai D (2006b) Dynamics of allosteric transitions in GroEL. *Proc Natl Acad Sci USA* **103:** 18939-18944

Ibuki T, Imada K, Minamino T, Kato T, Miyata T, Namba K (2011) Common architecture of the flagellar type III protein export apparatus and F- and V-type ATPases. *Nat Struct Mol Biol* **18:** 277-282

Imamura H, Nakano M, Noji H, Muneyuki E, Ohkuma S, Yoshida M, Yokoyama K (2003) Evidence for rotation of V_1-ATPase. *Proc Natl Acad Sci USA* **100:** 2312-2315

Ingber DE (1997) Tensegrity: the architectural basis of cellular mechanotransduction. *Annu Rev Physiol* **59:** 575-599

Ingber DE (2006) Cellular mechanotransduction: putting all the pieces together again. *FASEB J* **20:** 811-827

Jacobs DJ, Thorpe MF (1995) Generic rigidity percolation: The pebble game. *Phys Rev Lett* **75:** 4051-4054

Jin Q, Sorzano CO, de la Rosa-Trevin JM, Bilbao-Castro JR, Nunez-Ramirez R, Llorca O, Tama F, Jonic S (2014) Iterative elastic 3D-to-2D alignment method using normal modes for studying structural dynamics of large macromolecular complexes. *Structure* **22:** 496-506

Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of Simple Potential Functions for Simulating Liquid Water. *J Chem Phys* **79:** 926-935

Jorgensen WL MD, Tirado-Rives J (1996) Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J Am Chem Soc* **118:** 11225-11236

Karanicolas J, Brooks CL, 3rd (2002) The origins of asymmetry in the folding transition states of protein L and protein G. *Protein Sci* **11:** 2351-2361

Kawasaki Y, Chufan EE, Lafont V, Hidaka K, Kiso Y, Mario Amzel L, Freire E (2010) How much binding affinity can be gained by filling a cavity? *Chem Biol Drug Des* **75:** 143-151

Kaya H, Chan HS (2003) Solvation effects and driving forces for protein thermodynamic and kinetic cooperativity: how adequate is native-centric topological modeling? *J Mol Biol* **326:** 911-931

Kidera A, Go N (1990) Refinement of protein dynamic structure: normal mode refinement. *Proc Natl Acad Sci USA* **87:** 3718-3722

Kidera A, Inaka K, Matsushima M, Go N (1992a) Normal mode refinement: crystallographic refinement of protein dynamic structure applied to human lysozyme. *Biopolymers* **32:** 315-319

Kidera A, Inaka K, Matsushima M, Go N (1992b) Normal mode refinement: crystallographic refinement of protein dynamic structure. II. Application to human lysozyme. *J Mol Biol* **225:** 477-486

Koga N, Takada S (2006) Folding-based molecular simulations reveal mechanisms of the rotary motor F_1-ATPase. *Proc Natl Acad Sci USA* **103:** 5367-5372

Lau WC, Rubinstein JL (2010) Structure of intact Thermus thermophilus V-ATPase by cryo-EM reveals organization of the membrane-bound V_O motor. *Proc Natl Acad Sci USA* **107:** 1367-1372

Lazaridis T, Karplus M (1999) Effective energy function for proteins in solution. *Proteins* **35:** 133-152

Lee H, Darden T, Pedersen L (1995) Accurate Crystal Molecular-Dynamics Simulations Using Particle-Mesh-Ewald - RNA Dinucleotides - Apu and Gpc. *Chem Phys Lett* **243:** 229-235

Lee LK, Stewart AG, Donohoe M, Bernal RA, Stock D (2010) The structure of the peripheral stalk of Thermus thermophilus H^+-ATPase/synthase. *Nat Struct Mol Biol* **17:** 373-378

Levitt M, Warshel A (1975) Computer simulation of protein folding. *Nature* **253:** 694-698

Lezon TR, Bahar I (2010) Using entropy maximization to understand the determinants of structural dynamics beyond native contact topology. *PLoS Comput Biol* **6:** e1000816

Li XE, Holmes KC, Lehman W, Jung H, Fischer S (2010) The shape and flexibility of tropomyosin coiled coils: implications for actin filament assembly and regulation. *J Mol Biol* **395:** 327-339

Lopez CA, Rzepiela AJ, de Vries AH, Dijkhuizen L, Hulnenberger PH, Marrink SJ (2009) Martini coarse-grained force field: Extension to carbohydrates. *J Chem Theory Comput* **5:** 3195–3210

Louhivuori M, Risselada JJ, van der Giessen E, Marrink SJ (2010) Release of content through mechanosensitive gates in pressurized liposomes. *Proc Natl Acad Sci USA* **107:** 19856-19860

Ma L, Yethiraj A, Chen X, Cui Q (2009) A computational framework for mechanical response of macromolecules: application to the salt concentration dependence of DNA bendability. *Biophys J* **96:** 3543-3554

MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B* **102:** 3586-3616

MacKerell AD, Feig M, Brooks CL (2004~~a~~) Improved treatment of the protein backbone in empirical force fields. *J Am Chem Soc* **126:** 698-699

Maragakis P, Karplus M (2005) Large amplitude conformational change in proteins explored with a plastic network model: adenylate kinase. *J Mol Biol* **352:** 807-822

Marrink SJ, Risselada HJ, Yefimov S, Tieleman DP, de Vries AH (2007) The MARTINI force field: coarse grained model for biomolecular simulations. *J Phys Chem B* **111:** 7812-7824

Marshansky V, Rubinstein JL, Gruber G (2014) Eukaryotic V-ATPase: Novel structural findings and functional insights. *Biochim Biophys Acta* **1837:** 857-879

Marszalek PE, Lu H, Li H, Carrion-Vasquez M, Oberhauser AF, Schulten K, Fernandez JM (1999) Mechanical unfolding intermediates in titin modules. *Nature* **402:** 100-103

Marti-Renom MA, Stuart AC, Fiser A, Sanchez R, Melo F, Sali A (2000) Comparative protein structure modeling of genes and genomes. *Annul Rev Biophys Biomol Struct* **29:** 291-325

Martinetz T, Schulten K (1994) Topology-representing networks. *Neural Netw* **7:** 507-522

Moritsugu K, Terada T, Kidera A (2012) Disorder-to-order transition of an intrinsically disordered region of sortase revealed by multiscale enhanced sampling. *J Am Chem Soc* **134:** 7094-7101

Muench SP, Huss M, Song CF, Phillips C, Wieczorek H, Trinick J, Harrison MA (2009) Cryo-electron microscopy of the vacuolar ATPase motor reveals its mechanical and regulatory complexity. *J Mol Biol* **386:** 989-999

Muench SP, Scheres SH, Huss M, Phillips C, Vitavska O, Wieczorek H, Trinick J, Harrison MA (2014) Subunit positioning and stator filament stiffness in regulation and power transmission in the V_1 motor of the Manduca sexta V-ATPase. *J Mol Biol* **426:** 286-300

Muench SP, Trinick J, Harrison MA (2011) Structural divergence of the rotary ATPases. *Q Rev Biophys* **44:** 311-356

Clementi HN, and J. Onuchic (2000) Topological and energetic factors: what determines the structural details. *J Mol Biol* **298:** 937-953

Noji H, Yasuda R, Yoshida M, Kinosita K, Jr. (1997) Direct observation of the rotation of F_1-ATPase. *Nature* **386:** 299-302

Norgett EE, Borthwick KJ, Al-Lamki RS, Su Y, Smith AN, Karet FE (2007) V_1 and V_0 domains of the human H^+-ATPase are linked by an interaction between the G and a subunits. *J Biol Chem* **282:** 14421-14427

Numoto N, Hasegawa Y, Takeda K, Miki K (2009) Inter-subunit interaction and quaternary rearrangement defined by the central stalk of prokaryotic V_1-ATPase. *EMBO Rep* **10:** 1228-1234

Okamoto-Terry H, Umeki K, Nakanishi-Matsui M, Futai M (2013) Glu-44 in the amino-terminal alpha-helix of yeast vacuolar ATPase E subunit (Vma4p) has a role for V_oV_1 assembly. *J Biol Chem* **288:** 36236-36243

Oot RA, Huang LS, Berry EA, Wilkens S (2012) Crystal structure of the yeast vacuolar ATPase heterotrimeric EGC(head) peripheral stalk complex. *Structure* **20:** 1881-1892

Orzechowski M, Tama F (2008) Flexible fitting of high-resolution x-ray structures into cryoelectron microscopy maps using biased molecular dynamics simulations. *Biophys J* **95:** 5692-5705

Paci E, Karplus M (2000) Unfolding proteins by external forces and temperature: the importance of topology and energetics. *Proc Natl Acad Sci USA* **97:** 6521-6526

Panke O, Cherepanov DA, Gumbiowski K, Engelbrecht S, Junge W (2001) Viscoelastic dynamics of actin filaments coupled to rotary F-ATPase: angular torque profile of the enzyme. *Biophys J* **81:** 1220-1233

Periole X, Cavalli M, Marrink SJ, Ceruso MA (2009) Combining an elastic network with a coarse-grained molecular force field: structure, dynamics and intermolecular recognition. *J Chem Theory Comput* **5:** 2531-2543

Periole X, Marrink SJ (2013) The Martini coarse-grained force field. *Methods Mol Biol* **924:** 533-565

Petrone P, Pande VS (2006) Can conformational change be described by only a few normal modes? *Biophys J* **90:** 1583-1593

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF Chimera-a visualization system for exploratory research and analysis. *J Comput Chem* **25:** 1605-1612

Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K (2005) Scalable molecular dynamics with NAMD. *J Comput Chem* **26:** 1781-1802

Pogoryelov D, Reichen C, Klyszejko AL, Brunisholz R, Muller DJ, Dimroth P, Meier T (2007) The oligomeric state of c rings from cyanobacterial F-ATP synthases varies from 13 to 15. *J Bacteriol* **189:** 5895-5902

Pu J, Karplus M (2008) How subunit coupling produces the gamma-subunit rotary motion in F_1-atpase. *Proc Natl Acad Sci USA* **105:** 1192-1197

Oliver RC, Read DJ, Harlen OG, Harris SA (2013) A stochastic finite element model for the dynamics of globular macromolecules. *J Comput Phys* **239:** 147-165

Rees DM, Leslie AG, Walker JE (2009) The structure of the membrane extrinsic region of bovine ATP synthase. *Proc Natl Acad Sci USA* **106:** 21597-21601

Richardson RA, Papachristos K, Read DJ, Harlen OG, Harrison M, Paci E, Muench SP, Harris SA (2014) Understanding the apparent stator-rotor connections in the rotary ATPase family using coarse-grained computer modelling. *Proteins* **82:** 3298-3311

Roseman A (2000) Docking structures of domains into maps from cryo-electron microscopy using local correlation. *Acta Crystallogr D Biol Crystallogr* **56:** 1132-1340

Rossmann MG, Bernal R, Pletnev SV (2001) Combining electron microscopic with X-ray crystallographic structures. *J Struct Biol* **136:** 190-200

Rozycki B, Kim YC, Hummer G (2011) SAXS ensemble refinement of ESCRT-III CHMP3 conformational transitions. *Structure* **19:** 109-116

Rubinstein JL, Walker JE, Henderson R (2003) Structure of the mitochondrial ATP synthase by electron cryomicroscopy. *EMBO J* **22:** 6182-6192

Rueda M, Chacon P, Orozco M (2007) Thorough validation of protein normal mode analysis: a comparative study with essential dynamics. *Structure* **15:** 565-575

Saijo S, Arai S, Hossain KM, Yamato I, Suzuki K, Kakinuma Y, Ishizuka-Katsura Y, Ohsawa N, Terada T, Shirouzu M, Yokoyama S, Iwata S, Murata T (2011) Crystal structure of the central axis DF complex of the prokaryotic V-ATPase. *Proc Natl Acad Sci USA* **108:** 19955-19960

Saunders MG, Tempkin J, Weare J, Dinner AR, Roux B, Voth GA (2014) Nucleotide regulation of the structure and dynamics of G-actin. *Biophys J* **106:** 1710-1720

Schwaiger I, Sattler C, Hostetter DR, Rief M (2002) The myosin coiled-coil is a truly elastic protein structure. *Nat Mater* **1:** 232-235

Seeber M, Felline A, Raimondi F, Muff S, Friedman R, Rao F, Caflisch A, Fanelli F (2011) Wordom: A User-Friendly Program for the Analysis of Molecular Structures, Trajectories, and Free Energy Surfaces. *J Comput Chem* **32:** 1183-1194

Sielaff H, Rennekamp H, Wachter A, Xie H, Hilbers F, Feldbauer K, Dunn SD, Engelbrecht S, Junge W (2008) Domain compliance and elastic power transmission in rotary F_OF_1-ATPase. *Proc Natl Acad Sci USA* **105:** 17760-17765

Sigalov AB, Kim WM, Saline M, Stern LJ (2008) The intrinsically disordered cytoplasmic domain of the T cell receptor zeta chain binds to the nef protein of simian immunodeficiency virus without a disorder-to-order transition. *Biochemistry* **47:** 12942-12944

Smardon AM, Diab HI, Tarsio M, Diakov TT, Nasab ND, West RW, Kane PM (2014) The RAVE complex is an isoform-specific V-ATPase assembly factor in yeast. *Mol Biol Cell* **25:** 356-367

Smardon AM, Tarsio M, Kane PM (2002) The RAVE complex is essential for stable assembly of the yeast V-ATPase. *J Biol Chem* **277:** 13831-13839

Song CF, Papachristos K, Rawson S, Huss M, Wieczorek H, Paci E, Trinick J, Harrison MA, Muench SP (2013) Flexibility within the rotor and stators of the vacuolar H^+-ATPase. *PloS One* **8:** e82207

Stansfeld PJ, Sansom MSP (2011) From coarse-grained to atomistic: A serial multiscale approach to membrane protein simulations. *J Chem Theory Comput* **7:** 1157-1166

Stember JN, Wriggers W (2009) Bend-twist-stretch model for coarse elastic network simulation of biomolecular motion. *J Chem Phys* **131:** 074112

Stewart AG, Laming EM, Sobti M, Stock D (2014) Rotary ATPases - dynamic molecular machines. *Curr Opin Struct Biol* **25:** 40-48

Stewart AG, Lee LK, Donohoe M, Chaston JJ, Stock D (2012) The dynamic stator stalk of rotary ATPases. *Nat Commun* **3:** 687

Stock D, Leslie AG, Walker JE (1999) Molecular architecture of the rotary motor in ATP synthase. *Science* **286:** 1700-1705

Sugawa M, Okada KA, Masaike T, Nishizaka T (2011) A change in the radius of rotation of F_1-ATPase indicates a tilting motion of the central shaft. *Biophys J* **101:** 2201-2206

Tama F, Miyashita O, Brooks CL, 3rd (2004a) Flexible multi-scale fitting of atomic structures into low-resolution electron density maps with elastic network normal mode analysis. *J Mol Biol* **337:** 985-999

Tama F, Miyashita O, Brooks CL, 3rd (2004b) Normal mode based flexible fitting of high-resolution structure into low-resolution experimental data from cryo-EM. *J Struct Biol* **147:** 315-326

Tang Y, Cao G, Chen X, Yoo J, Yethiraj A, Cui Q (2006) A finite element framework for studying the mechanical response of macromolecules: application to the gating of the mechanosensitive channel MscL. *Biophys J* **91:** 1248-1263

Tang Y, Yoo J, Yethiraj A, Cui Q, Chen X (2008a) Gating mechanisms of mechanosensitive channels of large conductance, II: systematic study of conformational transitions. *Biophys J* **95:** 581-596

Tang Y, Yoo J, Yethiraj A, Cui Q, Chen X (2008b) Mechanosensitive channels: insights from continuum-based simulations. *Cell Biochem Biophys* **52:** 1-18

Tanner DE, Chan KY, Phillips JC, Schulten K (2011) Parallel Generalized Born Implicit Solvent Calculations with NAMD. *J Chem Theory Comput* **7:** 3635-3642

Tehver R, Thirumalai D (2010) Rigor to post-rigor transition in myosin V: link between the dynamics and the supporting architecture. *Structure* **18:** 471-481

Tirion MM (1996) Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Phys Rev Lett* **77:** 1905-1908

Topf M, Lasker K, Webb B, Wolfson H, Chiu W, Sali A (2008) Protein structure fitting and refinement guided by cryo-EM density. *Structure* **16:** 295-307

Trabuco LG, Villa E, Mitra K, Frank J, Schulten K (2008) Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. *Structure* **16:** 673-683

Vonck J, Pisa KY, Morgner N, Brutschy B, M ller V (2009) Three-dimensional structure of $A1A_O$-ATP synthase from the hyperthermophilic archeon Pyrococcus furiosus by electron microscopy. *J Biol Chem* **284:** 10110-10119

Voth GA (ed) (2009) *Coarse-graining of Condensed Phase and Biomolecular Systems*: CRC Press, Taylor & Francis Group

Wachter A, Bi Y, Dunn SD, Cain BD, Sielaff H, Wintermann F, Engelbrecht S, Junge W (2011) Two rotary motors in F-ATP synthase are elastically coupled by a flexible rotor and a stiff stator stalk. *Proc Natl Acad Sci USA* **108:** 3924-3929

Webb B, Sali A (2014) Protein structure modeling with MODELLER. *Methods Mol Biol* **1137:** 1-15

Whitford PC, Noel JK, Gosavi S, Schug A, Sanbonmatsu KY, Onuchic JN (2009) An all-atom structure-based potential for proteins: bridging minimal models with all-atom empirical forcefields. *Proteins* **75:** 430-441

Wriggers W, Milligan R, McCammon J (1999) Situs: A package for docking crystal structures into low-resolution maps from electron microscopy. *J Struct Biol* **125:** 185-195

Wriggers W, Milligan RA, Schulten K, McCammon JA (1998) Self-organizing neural networks bridge the biomolecular resolution gap. *J Mol Biol* **284:** 1247-1254

Wu L, Zhang J, Qin M, Liu F, Wang W (2008) Folding of proteins with an all-atom Go-model. *J Chem Phys* **128:** 235103

Yefimov S, van der Giessen E, Onck PR, Marrink SJ (2008) Mechanosensitive membrane channels in action. *Biophys J* **94:** 2994-3002

Yoshida M, Muneyuki E, Hisabori T (2001) ATP synthase-a marvellous rotary engine of the cell. *Nat Rev Mol Cell Biol* **2:** 669-677

Yoshimoto K, Arora K, Brooks CL, III (2010) Hexameric helicase deconstructed: interplay of conformational changes and substrate coupling. *Biophys J* **98:** 1449-1457

Zhang Z, Lu L, Noid WG, Krishna V, Pfaendtner J, Voth GA (2008a) A systematic methodology for defining coarse-grained sites in large biomolecules. *Biophys J* **95:** 5073-5083

Zhang Z, Pfaendtner J, Grafmuller A, Voth GA (2009) Defining coarse-grained representations of large biomolecules and biomolecular complexes from elastic network models. *Biophys J* **97:** 2327-2337

Zhang Z, Voth GA (2010) Coarse-Grained Representations of Large Biomolecular Complexes from Low-Resolution Structural Data. *J Chem Theory Comput* **6:** 2990-3002

Zhang Z, Zheng Y, Mazon H, Milgrom E, Kitagawa N, Kish-Trier E, Heck AJR, Kane PM, Wilkens S (2008b) Structure of the yeast vacuolar ATPase. *J Biol Chem* **283:** 35983-35995

Zhou M, Politis A, Davies RB, Liko I, Wu KJ, Stewart AG, Stock D, Robinson CV (2014) Ion mobility-mass spectrometry of a rotary ATPase reveals ATP-induced reduction in conformational flexibility. *Nat Chem* **6:** 208-215

Zuckerman DM (2004) Simulation of an ensemble of conformational transitions in a united-residue model of calmodulin. *J Phys Chem B* **108:** 5127-5137

Appendix

A.1 Supplementary Information for *Chapter 4*

CG Particles	K_b (kcal/mol/A^2)	b_0 (A)
P1-P6	18.50	13.46
P1–P7	16.97	14.17
P2-P4	6.76	12.93
P2-P8	3.88	12.97
P3-P6	17.29	13.59
P3-P11	18.27	13.58
P4-P5	1.90	13.80
P4-P9	1.02	20.43
P5-P9	0.55	14.37
P5-P12	0.78	16.06
P7-P10	18.64	13.43
P8-P10	18.51	13.34
P9-P12	4.93	11.28
P9-P13	2.74	13.66
P11-P14	14.94	12.90
P12-P13	2.13	13.93
P14-P15	1.79	12.85

Table A1.1 CHARMM-like elastic force constants of the CG pseudo-bonds calculated from the atomistic MD of the prokaryotic peripheral stator stalk, EG complex.

CG Particles	K_θ (kcal/mol/rad^2)	θ_0 (deg)
P6-P1-P7	124.34	173.26
P4-P2-P8	183.92	125.33
P6-P3-P11	163.41	171.01
P2-P4-P5	33.85	131.50
P2-P4-P9	43.75	164.55
P5-P4-P9	117.18	44.54
P4-P5-P9	58.64	92.95
P4-P5-P12	86.16	117.67
P9-P5-P12	330.18	42.95
P1-P6-P3	142.65	171.36
P1-P7-P10	150.50	171.59
P2-P8-P10	63.46	158.98
P4-P9-P5	229.48	42.32
P4-P9-P12	55.63	103.67
P4-P9-P13	61.70	84.39
P5-P9-P12	79.64	76.48
P5-P9-P13	154.19	102.55
P12-P9-P13	222.55	67.05
P7-P10-P8	54.41	168.09
P3-P11-P14	114.17	167.26
P5-P12-P9	67.81	60.31
P5-P12-P13	174.71	93.36
P9-P12-P13	324.67	64.58
P9-P13-P12	408.85	48.25
P11-P14-P15	126.77	141.57

Table A1.2 CHARMM-like elastic force constants of the CG pseudo-angles as calculated from the atomistic MD of the prokaryotic peripheral stator stalk, EG complex.

CG Particles	Kχ (kcal/mol/rad²)	n [-]	delta (deg)
P7-P1-P6-P3	0.64	1	-5.81
P6-P1-P7-P10	0.71	1	41.16
P11-P3-P6-P1	0.71	1	11.34
P6-P3-P11-P14	1.00	1	8.44
P2-P4-P5-P9	2.43	4	-109.95
P2-P4-P5-P12	2.46	4	104.99
P2-P4-P9-P12	0.87	2	81.41
P2-P4-P9-P13	0.88	2	211.21
P1-P7-P10-P8	1.46	1	-71.23
P3-P11-P14-P15	0.89	2	150.91
P8-P2-P4-P5	16.18	-	138.59
P8-P2-P4-P9	1.36	-	70.63
P4-P2-P8-P10	20.23	-	111.15
P9-P4-P5P12	101.16	-	-36.83
P2-P4-P9-P5	1.59	-	81.39
P5-P4-P9-P12	32.77	-	51.18
P5-P4-P9-P13	43.17	-	115.88
P4-P5-P9-P12	30.65	-	-129.06
P4-P5-P9-P13	26.19	-	-66.72
P12-P5-P9-P4	30.64	-	128.97
P12-P5-P9-P13	214.33	-	62.35
P4-P5-P12-P9	22.19	-	61.37
P4-P5-P12-P13	22.97	-	3.24
P9-P5 12 13	196.00	-	-58.17
P2-P8-P10-P7	3.43	-	-121.24
P4-P9-P12-P5	213.05	-	-32.61
P4-P9-P12-P13	66.50	-	77.61
P5-P9-P12-P13	133.94	-	110.13
P13-P9-P12-P5	133.94	-	-110.18
P4-P9-P13-P12	71.37	-	-107.66
P5-P9-P13-P12	81.48	-	-69.30
P5-P12-P13-P9	74.15	-	54.74

Table A1.3 CHARMM-like (*n, integer*) and harmonic (*n, no value indicated*) elastic force

constants of the CG pseudo-dihedrals as calculated from the atomistic MD of the prokaryotic peripheral stator stalk.

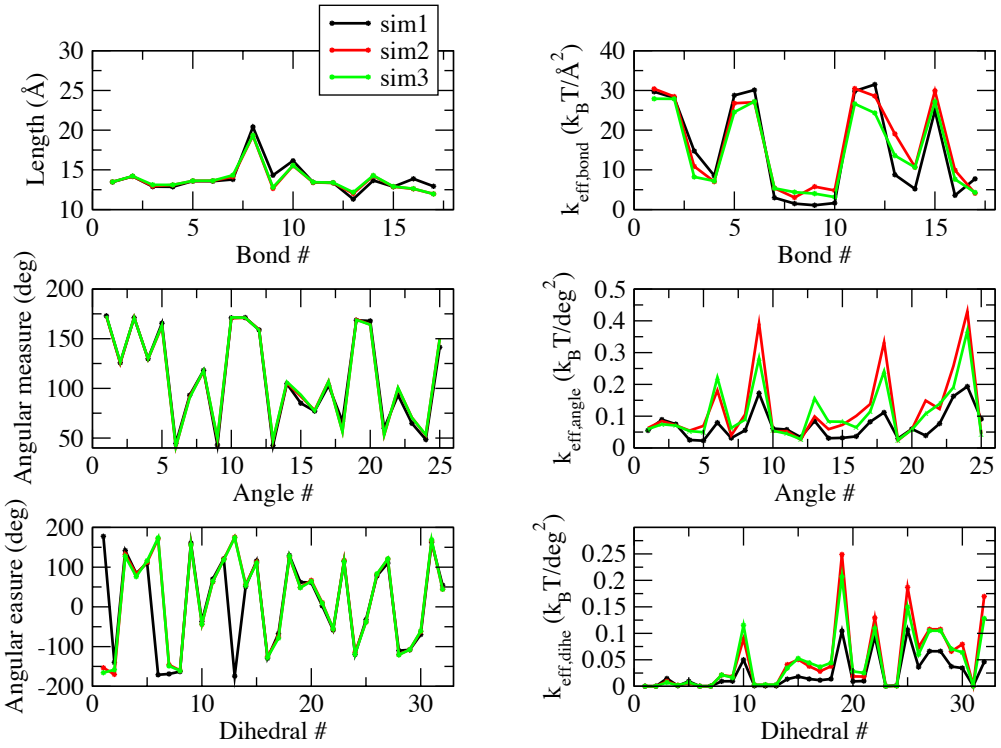


Figure A1.1 Averages and effective force constants of all the (pseudo)-bonded terms considered. Results are shown for individuals trajectories *sim1* to *sim3* to highlight the convergence of the properties shown over two trajectories started from structures PS2 and PS3.

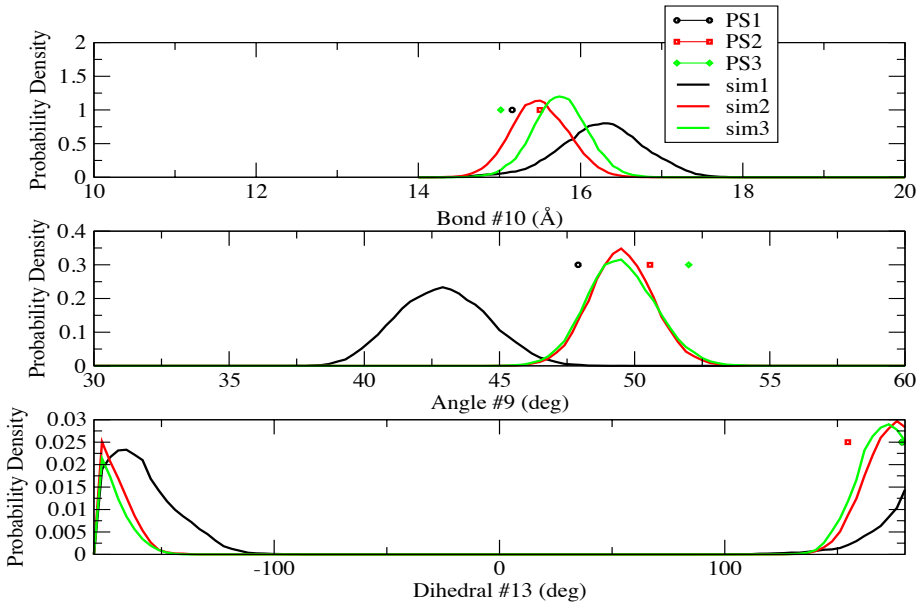


Figure A1.2: Selected head-domain CG bonded terms highlighting that the range of conformations sampled is similar in *sim2* and *sim3*, but different for *sim1* seeded with initial structure PS1. The values for the reference structures (PS1-3) are also shown.