

DISSERTATION

Submitted to the
Combined Faculties for Natural Sciences and Mathematics
of the Ruprecht-Karls-University of Heidelberg, Germany

for the degree of

Doctor of Natural Sciences

Presented by:

Dipl.-Chem. Sampath Kumar Koppole

Born in Nidamanuru, India

Oral Examination: 12th January 2007

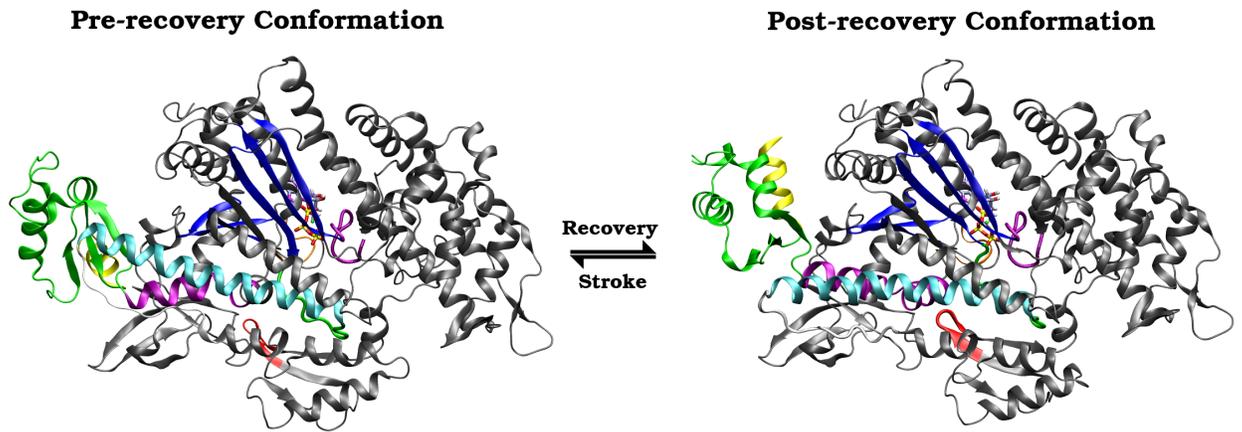
Structural mechanism of the
recovery-stroke in Myosin II molecular
motor at atomic detail.

Referees:

Prof. Dr. Joachim P. Spatz

Prof. Dr. Jeremy C. Smith

Structural mechanism of the recovery-stroke in Myosin II molecular motor at atomic detail.



Sampath Kumar Koppole

Dedicated to my Family

*Asatho Maa Sad Gamaya;
Thamaso Maa Jyothir Gamaya;
Mrithyor Maa Amritham Gamaya.
–Atharvana Veda*

*Lead me from untruth to truth;
Lead me from darkness to light;
Lead me from death to immortality.
–Atharvana Veda*

The work in this thesis was performed under the guidance of
Dr. Stefan Fischer, Head of the Computational Biochemistry group at
IWR (Interdisziplinäres Zentrum für Wissenschaftliches Rechnen),
University of Heidelberg, Germany.

CONTENTS

List of Figures	viii
List of Tables	ix
Summary	xi
Zusammenfassung	xiii
Acknowledgements	xv
Articles from this PhD Thesis	xvii
1 Introduction	1
1.1 Myosins: A Super Family	1
1.2 The Myosin II Motor	2
1.3 General structure of Myosin S1	3
1.4 Atomic structure of F-Actin	4
1.5 Strategies for Movement	6
1.6 Lymn-Taylor Cycle	7
1.7 Myosin II motor in <i>Dictyostelium discoideum</i>	7
1.8 Nucleotide binding site	9
1.9 Crystal Structure classification	9
1.10 Coupling mechanisms in Myosin II	12
1.11 Recovery stroke coupling mechanism	13
1.12 Questions addressed in this thesis	15
1.13 Outline for the thesis	16
2 Theoretical Aspects	17
2.1 A Molecular Mechanics Force Field	18
2.1.1 Treatment of the non-bonded energy terms	20
2.1.2 Long-range electrostatic interactions	21
2.1.3 The Extended Electrostatics Model	21
2.1.4 Treatment of Solvent in Computer Simulations	22

2.2	Continuum Electrostatics	22
2.2.1	Simple Coulomb Electrostatics	23
2.2.2	The Poisson-Boltzmann Equation	24
2.2.3	Focusing	26
2.2.4	Generalized Born Methods	27
2.2.5	The Generalized Born Equation (GB Equation)	27
2.2.6	Analytical Continuum Electrostatics (ACE)	30
2.2.7	Solvation Free Energy	32
2.2.8	Sub-division of Solute volume	33
2.2.9	Coulomb Field Approximation to the the Self Energy	36
2.2.10	Self Energy Approximation	36
2.2.11	Interaction Energy Approximation:	38
2.2.12	Optimization of ACE	39
2.3	Classical Mechanics	41
2.3.1	Integration Algorithms	42
2.4	Statistical Mechanics	43
2.4.1	Simulating in different Ensembles	45
2.4.2	Micro-canonical Ensemble	45
2.4.3	Canonical Ensemble	46
2.4.4	Nosé-Hoover constant temperature algorithm	46
2.4.5	Isobaric-Isothermal Ensemble	47
2.4.6	Constant temperature and pressure	47
2.5	Determination of Reaction Paths	48
2.5.1	Determination of a MEP in systems with many degrees of freedom	48
2.5.2	Conjugate Peak Refinement, CPR	48
2.5.3	Generation of an initial path	49
2.5.4	Reaction Coordinate	50
2.6	Analyzing conformational transitions using Computer Simulations	50
3	The structural coupling between ATPase activation and recovery-stroke in the Myosin II motor.	53
3.1	Summary	53
3.2	Introduction	54
3.3	Results	58
3.3.1	Phase I mechanism.	60
3.3.2	Phase II mechanism.	60
3.4	Discussion	63

3.5	METHODS	68
3.5.1	Protein Modeling	68
3.5.2	Preparation of End State Structures	68
3.5.3	Molecular Kinematics	71
4	Simulations of the myosin II motor reveal a nucleotide-state sensing element that controls the recovery stroke	73
4.1	Summary	73
4.2	Introduction	74
4.3	Results	80
4.3.1	Flexibility of the catalytic site	80
4.3.2	The hydrogen bonding of Asn475	81
4.3.3	Behavior of Asn475 upon hydrolysis	83
4.3.4	Arg238-Glu459 Salt bridge	86
4.4	Discussion	88
4.5	Methods	91
4.5.1	Protein Modeling	91
4.5.2	Molecular Dynamics (MD)	92
4.5.3	Perturbation of the electrostatic interactions with ATP.	94
4.5.4	Acknowledgments	94
5	Outlook	95
5.1	Development of accurate Continuum Electrostatics methods.	95
5.2	Free Energy Calculations.	96
5.3	Transition pathway sampling.	96
5.4	Modeling State IV in the Lymn-Taylor Cycle	97
5.5	Myosin II rigor conformation to Myosin II pre-recovery conformation	97
5.6	Normal Mode Analysis	97
	References	99

LIST OF FIGURES

1.1	Un-rooted phylogenetic tree obtained from the myosin motor domain sequences (This picture was taken from reference ¹).	2
1.2	Myosin-S1: The regulatory light chain (magenta) and the essential light chain (yellow). The proteolytic fragments are color coded as follows: 25K (N terminal), green; 50K, red; and 20K (C terminal), blue. The 50K fragment spans two domains: the 50K upper domain and the 50K lower domain or actin-binding domain. The actin-binding domain has been colored grey. Figure taken from. ² . . .	4
1.3	F-Actin. Figure taken from reference. ²	5
1.4	Lymn-Taylor Cycle.	6
1.5	The Nucleotide Binding Site. Legend: P-loop in Orange; Switch-2 in green; Switch-1 in purple. Phosphorus atoms of ATP in yellow and magnesium in green.	10
1.6	Structural elements of myosin involved in the recovery-stroke. Legend: P-loop in Orange; Switch-2 in green; Relay helix in cyan; SH-1 helix in purple; SH-2 helix in pink; converter domain in green; lever arm in yellow. In the right panel, the important residues and bond distances involved in the recovery stroke are indicated. The See-saw and the wedge phases movements are shown with solid arrows.	12
1.7	Three different views of the Myosin in the end-states of the recovery stroke. Legend: Switch-2 in green; Relay helix in cyan; SH-1 helix in purple; SH-2 helix in pink; lever arm in yellow. A. See-saw view; B. Wedge view; C. P-loop view. . .	14

2.1	Hypothetical atoms to illustrate the energetic terms included in the force-field. An Molecule comprises of atoms 1-5. Internal terms that occur in molecule are the bonds, b , between atoms 1 & 2, 2 & 3, 3 & 4; angles θ , involving atoms 1-2-3 and 2-3-4 and a dihedral or a torsional angle χ , described by atoms 1-2-3-4. Atom 5 is involved in a "Non-bonded" interaction with atoms 1, 2, 3, & 4. The energy of this interaction is determined by the distance, r_{ij} between atom 5 and rest of the molecule.	19
2.2	Lennard-Jones potential to account for van Der Waals (VDW)	20
2.3	Effect of truncation schemes on the energy of a given conformation	21
2.4	A molecule in a heterogeneous dielectric medium.	23
2.5	Part of the grid used to solve the PBEQ	26
2.6	Optimization of ACE parameters for use with minimum energy pathway calculations using CPR ³	40
3.1	Lymn-Taylor Cycle.	54
3.2	Structural elements of myosin involved in the recovery-stroke. Legend: P-loop in Orange; Switch-2 in green; Relay helix in cyan; SH-1 helix in purple; SH-2 helix in pink; converter domain in green; lever arm in yellow. In the right panel, the important residues and bond distances involved in the recovery stroke are indicated. The See-saw and the Crank-shaft phases movements are shown with solid arrows.	56
3.3	Plots of internal rearrangements during the recovery-stroke transition. A. Rotation of the converter-domain and lever arm; B. RMSD profile for the Switch-2 (in blue) and the Wedge loop (red); C. Hydrogen bond distances. Gly457: N_{amide} - ATP: $O_{\gamma-P}$ (red); Phe458: $O_{carbonyl}$ - Ser181: N_{amide} (Blue); D. See-saw angle of the relay-helix (red) and the SH1-helix longitudinal translation (blue).	59
3.4	Cartoon representing the sequence of coupling events during the return-stroke mechanism. Wedge loop in red; Switch-2 in green; Relay helix in cyan; SH1-helix in purple; converter domain in green; lever arm in yellow.	61
3.5	Packing between the converter domain (in green) and the C-terminal-third of the relay-helix (in brown). A. Salt-bridges; B. Van der Waals spheres.	62
3.6	Converter domain is held by just two helices, the relay helix and the SH1-helix and can move independent of the rest of the head.	64
3.7	The hydrophobic cradle of Phe458.	65
3.8	Longitudinal translation of the SH1-helix (in purple) relative to the relay helix (in cyan).	69

4.1	Lymn-Taylor Cycle. The structural domains of myosin are: Lever arm (yellow); myosin head (red); converter domain (green). The actin filament is shown as white spheres. Solid arrows: Standard Lymn-Taylor cycle. Dotted arrows: non-force-generating release of hydrolysis products.	75
4.2	End-states of the recovery stroke. A,B,C) Zoomed views of the Pre-recovery conformation (State II in Figure 4.1). D,E,F) Corresponding views of the Post-recovery conformation (State III in Figure 4.1). Structural elements: Relay helix in cyan; SH-1 helix in purple; converter domain in green; lever arm in yellow 7-stranded β -sheet in blue; P-loop in orange; Switch-1 in purple; Switch-2 in green; Wedge loop in red. Phosphorus atoms of ATP in yellow and Magnesium in green. The pre-recovery structure is PDB entry 1MMD ⁴ and the post-recovery conformation is essentially identical to 1VOM ⁵ (see Methods). Pictures created using Visual Molecular Dynamics (VMD) ⁶ and POVRAY ⁷	77
4.3	Root-Mean-Square coordinate fluctuations of structural elements. A) P-loop; B) Switch-1; C) Switch-2; D) N-terminal half of the relay helix (residues 466-476). White bars: pre-recovery conformation of myosin; grey bars: post-recovery conformation. The state of the nucleotide is indicated in each bar. .	80
4.4	Protein-protein interactions in response to the nucleotide state. Hydrogen bond distances: Asn475:ND-Gly457/Ser456:O _{peptide} (Blue), Arg238:N _Z -Glu459:O _E (Red), Asn475:ND-Tyr573:OH (Green). A) Pre-recovery with ADP. B) Pre-recovery without nucleotide (apo). C) Pre-recovery with ATP. D) Post-recovery with ATP. E) Post-recovery with ADP·Pi.	82
4.5	Nucleotide sensing by Asn475 in the post-recovery-stroke conformation. Rotation (black arrow) of Asn475 around χ_1 , switching from a hydrogen bond with the Gly457/Ser456 peptide group on Switch-2 to a hydrogen bond with Tyr573 on the wedge loop (residues 571 to 575).	84
4.6	Effect of shifting charges on the nucleotide, in the post-recovery conformation. The start and end of the electrostatic perturbations are indicated (modified electrostatics interactions of the Gly457/Ser456 peptide group with the β and γ -Phosphate groups, see text). Hydrogen bond distances: Asn475:N-Gly457/Ser456:O _{peptide} (blue), Asn475:N-Tyr573:O (green) and Gly457- γ -phosphate (yellow).	85

4.7	Movement of Switch-2 upon hydrolysis in the post-recovery conformation. The post-hydrolysis state III (Crystal structure) is shown in color: P-loop in orange; Switch-1 in purple; Switch-2 in green. The post-hydrolysis state III' (after 5 ns of MD) is shown in grey: ADP, Pi and Switch-2 only. The P-loop and ADP were used for the best-fit overlap of these two conformers. The two panels are for wall-eyed stereo-viewing.	86
4.8	Number of water molecules in the catalytic site. Water found within 6.0Å radius from the geometric center defined over the C α atoms of Ser181 (P-loop), Ser237 (Switch-1) and Gly457 (Switch-2). Bars are labeled as in Figure 4.3.	87
4.9	Root-Mean-Square fluctuations of C α atoms. Calculated from either the MD trajectory or the experimental B-factors (See Methods).	93

LIST OF TABLES

1.1	Crystal structures for <i>Dictyostelium discoideum</i> myosin II motor.	8
1.2	Consensus sequences.	10
1.3	Assignment to the states in the Lymn-Taylor cycle.	11
3.1	Consensus sequences.	57
3.2	Symmetric chemical moieties	71
4.1	Consensus sequences.	76
4.2	Tip of the “Wedge” loop ^a	90

SUMMARY

The molecular motor protein, **myosin**, converts chemical energy from ATP hydrolysis into useful mechanical work that is used to translocate the myosin filament along an actin filament during muscle contraction. The mechanism involved in the chemo-mechanical coupling necessary for myosin function is poorly understood. In this research work, a computational attempt is being made to understand the recovery-stroke mechanism in a myosin molecule which is one of the fundamental processes that occurs during muscle contraction in living organisms.

During the recovery stroke, the myosin motor is primed for the next power stroke by a 60° rotation of its converter domain (which bears the lever arm). This reversible motion is coupled to the activation of its ATPase function through conformational changes along the relay helix, which runs from the Switch-2 loop near the ATP to the converter domain. This coupling mechanism is determined by computing minimum energy pathways (MEP) between the crystallographic end-states of the recovery stroke, yielding a continuous series of optimized intermediates in atomic detail. The MEP reveals a two-phase mechanism, in which the successive formation of two hydrogen bonds by the Switch-2 loop is correlated with the successive movement of the two helices that hold the converter domain: the relay helix and the SH1-helix. The first phase involves the formation of a hydrogen bond (between Gly457, on the N-terminal of the relay helix and the γ -phosphate of ATP) which causes a “See-Saw” like motion of the relay helix. The second phase is triggered by the formation of another hydrogen bond (between Switch-2 and the Ser181 of the P-loop) which causes the wedging of a loop against the N-terminal end of the SH1-helix, resulting in the longitudinal translation of the SH1-helix relative to the relay helix. The converter domain first responds to the “See-Saw” motion of the relay helix by rotating $\sim 20^\circ$, then to the translation of the SH1-helix by rotating a further 40° . The proposed coupling mechanism is consistent with

the existing mutational data and explains the role of a highly conserved loop structure, called here as the “Wedge loop”, which was recognized for the first-time.

Molecular dynamics simulations of *Dictyostelium discoideum* myosin II in the two end conformations of the recovery stroke with different nucleotide states (ATP, ADP·Pi, ADP) reveal that the side-chain of Asn475 (which initiates the first-phase of the recovery-stroke) switches away from Switch-2 upon ATP hydrolysis to make a hydrogen bond with Tyr573 (on the Wedge loop). This sensing of the nucleotide state is achieved by a small displacement of the cleaved γ -phosphate towards Gly457 which in turn pushes Asn475 away. The sensing plays a dual role by (i) preventing the wasteful reversal of the recovery stroke while the nucleotide is in the ADP·Pi state, and (ii) decoupling the relay helix from Switch-2, thus allowing the power stroke to start upon initial binding to actin while Gly457 of Switch-2 keeps interacting with the Pi (known to be released only later after tight actin binding). The catalytically important salt bridge between Arg238 (on Switch-1) and Glu459 (on Switch-2), which covers the hydrolysis site, is seen to form rapidly when ATP is added to the pre-recovery stroke conformer and remains stable after the recovery stroke, indicating that it has a role in shaping the ATP binding site by induced fit.

ZUSAMMENFASSUNG

Das molekulare Motorprotein Myosin wandelt chemische Energie aus der ATP Hydrolyse in mechanische Arbeit um, die dazu genutzt wird um Myosin- und Aktin-Filamente gegeneinander zu verschieben und so z.B. die Muskelkontraktion zu ermöglichen. Der Mechanismus dieser chemisch-mechanischen Kopplung, der für die Funktion von Myosin essenziell ist, ist nur in Ansätzen verstanden. In dieser Arbeit wird ein rechnergestützter Ansatz verwendet um den Mechanismus des "recovery stroke" zu verstehen. Der "recovery stroke" ist einer der fundamentalen Prozesse bei der Muskelkontraktion in lebenden Organismen.

Während des "recovery stroke" wird der Myosin Motor für den nächsten Kraftschlag vorbereitet indem der Myosin-Kopf um 60° relativ zur Konverter-Domäne und dem Hebelarm gedreht wird. Der Drehpunkt ist mit der Bindetasche, in der die ATP Hydrolyse stattfindet, durch die sogenannte Relais-Helix verbunden. Während des "recovery stroke" finden eine Reihe von strukturellen Änderungen längs dieser Helix statt. In der vorliegenden Arbeit wird der Kopplungsmechanismus zwischen der ATP Hydrolyse und der Drehbewegung mit Hilfe eines Minimum-Energie Pfades (MEP) simuliert. Der MEP verbindet die Röntgenkristallographischen End-Zustände des Prozesses durch eine Kette von geometrioptimierten intermediären Strukturen. Der "recovery stroke" beruht auf der Bildung zweier Wasserstoffbrückenbindungen durch die "switch-2" Schleife, in Korrelation mit der Bewegung zweier Helices welche die Konverter-Domäne halten: der Relais-Helix und der SH1-Helix. Der MEP zeigt dass dieser Prozess aus zwei Phasen besteht. In der ersten Phase bildet sich eine Wasserstoffbrückenbindung zwischen Gly457 am N-terminalen Ende der Relais-Helix und dem γ -Phosphat des ATP, was eine Kipp-Bewegung der Relais-Helix zur Folge hat. Die zweite Phase wird durch die Bildung einer Wasserstoffbrückenbindung zwischen der "switch-2" Schleife und Ser181 der P-Schleife initiiert.

Dadurch wird eine weitere Schleife ähnlich einem Keil gegen das N-terminale Ende der SH1-Helix geschoben, wodurch letztere parallel zur Relais-Helix verschoben wird. Die Kippbewegung der ersten Phase bewirkt eine Drehung der Konverter-Domäne um 30°, während die Verschiebung der SH1-Helix eine Drehung um weitere 40° zur Folge hat. Der hier vorgeschlagene Kopplungsmechanismus ist konsistent mit verfügbaren Mutations-Experimenten und erklärt zum ersten Mal die Rolle der hochgradig sequenz-konservierten Schleife, die hier "Keil"-Schleife genannt wird.

In einem weiteren Teil der Arbeit werden Molekulardynamik-Simulationen von Myosin II des Organismus *Dictyostelium Discoideum* in beiden End-Zuständen des "recovery stroke" mit verschiedenen Nukleotid-Zuständen (ATP, ADP·Pi, ADP) durchgeführt. Diese Simulationen zeigen dass die Seitenkette von Asn475 (welche die erste Phase des "recovery stroke" initiiert) sich durch die ATP-Hydrolyse von "switch-2" wegbewegt und eine Wasserstoffbrückenbindung mit Tyr573 auf der Keilschleife bildet. Diese Abhängigkeit vom Nukleotid-Zustand wird erklärt durch eine kleine Verschiebung des abgespaltenen β -Phosphats hin zu Gly457 welches seinerseits Asn475 verschiebt. Die Sensitivität bezüglich des Nukleotid-Zustandes ist wichtig für (i) die Vermeidung einer unproduktiven Umkehrung des "recovery strokes" während des ADP·Pi Zustandes, und (ii) die Entkopplung der Relais-Helix vom "switch-2", wodurch erreicht wird, dass der Kraftschlag nach der initialen Bindung an Aktin ausgelöst wird, wobei Gly457 von "switch-2" weiterhin mit dem Pi interagiert, welches bekanntermaßen erst nach der Bindung an Aktin freigelassen wird. Es wird beobachtet dass die katalytisch wichtige Salzbrücke zwischen Arg238 (in "switch-1") und Glu459 (in "switch-2"), welche die Bindetasche an der Hydrolysestelle bedeckt, durch die Bindung von ATP an die Struktur vor dem "recovery stroke" schnell gebildet wird. Diese Salzbrücke bleibt auch nach dem "recovery stroke" stabil, was darauf hindeutet dass sie die Rolle hat die ATP Bindetasche durch "induced fit" zu formen.

ACKNOWLEDGMENTS

From experience, these acknowledgement pages of a PhD thesis are the most widely read pages of the entire publication. I guess, I would like to take the opportunity to highlight here the most important people who made this possible.

There are no words to express my gratitude to my research supervisors, Prof. Jeremy C. Smith and Dr. Stefan Fischer for giving me this opportunity to be a part of their esteemed groups. Their great scientific insights were the source of inspiration for this work. Without their guidance, this thesis would not have taken this present form. They has been extremely supportive during my studies here in Heidelberg and both have been excellent mentors during my study. I am very grateful to them for the academic freedom which has brought about independent thinking during my research. Jeremy had been a great company during fun times while we were in the lab and while we were on the group's "Winter Retreats" & "Summer Picnics". Jeremy was extremely helpful during the times I had accidents while on the Ski-slopes or while playing and fixing bulbs in my apartment.

I would like to thank Ms. Ellen Vogel who helped me a lot with administrative matters when I just arrived here in Heidelberg from Delhi and for the numerous German translations of all administrative letters. My friend and my colleague, Bogdan Costescu has been a real 'dream' system administrator and I would like to thank him for always taking the time to answer all of my "stupid" computing questions and helping me in the initial days of my "System administration" of my laptop. His insight into the hardware and software are truly beyond compare.

I extremely grateful to my parents and my sister for their unconditional love and encouragement. I am extremely lucky to have found my soul-mate. I treasure, her love, support and encouragement all through my PhD which brought about a new person in me. I also thank all my

close friends: KK, Abhishek and Vishnu who made my stay in Germany a really memorable one. Many thanks to all the members of the Computational Molecular Biophysics and Computation Biochemistry for providing an excellent environment for making research real fun and helping me out while I was stuck solving the most simple problems during my research. Last but not the least I would like to thank GOD for giving me all the faculties of knowledge and wisdom and I bow down to HIS grace and seek his blessings in all my endeavours.

Finally, I gratefully acknowledge the Bundesministerium für Bildung und Forschung and the Deutsche Forschungsgemeinschaft, which financially supported a large part of this research.

ARTICLES FROM THIS PHD THESIS

1. Sampath Koppole, Jeremy C. Smith, Stefan Fischer.
Simulations of myosin II motor reveals a nucleotide state sensing element that controls the recovery stroke. **Journal of Molecular Biology (J. Mol. Biol.)**, 2006, 361, pages: 604-616.
<http://dx.doi.org/10.1016/j.jmb.2006.06.022>.
2. Sampath Koppole, Jeremy C. Smith, Stefan Fischer.
The structural coupling between ATPase activation and recovery-stroke in the Myosin II motor. **To be submitted to The EMBO Journal**
3. Sidonia Mesentean, Sampath Koppole, Jeremy C. Smith, Stefan Fischer
Principal motions involved in the coupling mechanism of the recovery stroke of the Myosin motor. **Submitted to Journal of Molecular Biology.**

INTRODUCTION

Molecular motors are biological "nano-machines" that are the essential agents of movement in living organisms. Generally speaking, a motor is defined as a device that consumes energy in one form and converts it into mechanical work. Protein-based molecular motors convert the chemical energy present in ATP into mechanical energy and movement.^{8,9} This movement is essential to the life of a biological cell, and it takes many forms, from cytoplasmic streaming and the growth of neurones to the long distance flight of the albatross or the explosive performance of a sprinter.¹⁰ From the vantage point of biophysics, motor proteins provide fascinating systems for understanding how proteins use energy from ATP to power non-spontaneous events like muscle contraction. Three types of cytoplasmic molecular motors are known: myosins, which move along actin filaments, dyneins and kinesins, which use microtubules as tracks. The mechanism they use to convert chemical energy into mechanical work is both simple and ingenious.¹¹ Interest in motor proteins has expanded enormously in recent years. One of the most extensively studied molecular motors are the myosins, which drives muscle contraction (reviewed in refs^{2,12,13}). In myosin, ATP binding and later hydrolysis cause small conformational changes in a globular motor domain that are amplified and translated into movement with the aid of accessory structural motifs.^{2,12}

1.1 MYOSINS: A SUPER FAMILY

Myosins constitute a large super-family of proteins that share a common domain which has been shown to interact with actin, hydrolyze ATP and produce movement in all cases examined to date.^{1,14,15} The un-rooted phylogenetic tree of the myosin super-family (Figure 1.1) is derived from an alignment of 139 members of the myosin super-family. The alignment compared the

(or tail) which usually terminates in a short non-helical segment.¹⁸ Myosin II molecules have a two-headed structure, due to the dimerization of the heavy chain in the tail. The tails of myosin self-associate to form filaments both *in vivo* and *in vitro* at low ionic strength.

Myosin II can be proteolytically cleaved into discrete functional domains. One site of cleavage is at the junction between the head and the tail which produces a soluble fragment termed subfragment one (S1) and the tail fragment which remains a coiled-coil dimer and retains the solubility properties of the parent molecules. S1 binds to actin and nucleotides and contains two light chains. Another cleavage site of myosin II of the tail fragment produces two fragments, heavy meromyosin (HMM) and light meromyosin (LMM). HMM, which is soluble even at low ionic strength, contains the head region and a portion of the coiled-coil forming sequence (termed subfragment 2 or S2), which further dimerizes to produce a two-headed fragment. LMM retains the solubility properties of the parent molecule. Both S1 and HMM have been critical to kinetic, biophysical and structural studies of the myosin molecule.^{2, 12, 19, 20}

1.3 GENERAL STRUCTURE OF MYOSIN S1

The first X-ray crystal structure of S1 was from chicken muscle without nucleotide but with a sulfate ion in place of the β -phosphate in the active site.²¹ The structure shows the S1 in a tadpole-like conformation with an elongated head consisting of a 7-stranded- β -sheet and a C-terminal tail (Figure 1.2).

All three fragments (25K, 50K, and 20K) contribute to the 7-stranded β -sheet. Numerous α -helices that surround the β -sheet form a deep cleft extending from the nucleotide-binding site to the actin-binding site. In Figure 1.2, the proteolytic fragments are color coded as follows: 25K (N terminal), green; 50K, red; and 20K (C terminal), blue. The 50K fragment actually spans two domains, called as the 50K upper domain and the 50K lower domain or actin-binding domain.²¹ The actin-binding domain has been colored grey. The rest of the 25K fragment, together with the 50K upper fragment, residues 81-486, form one large domain that accounts for 6 of the 7 strands of the β -sheet and constitutes the bulk of the molecule. The ATP-binding site is in this large domain near the 25K-50K fragment boundary and contains a characteristic P-loop similar to that found in many ATPases and G-proteins.²² The ATP-binding site is about 4.0 nm from the

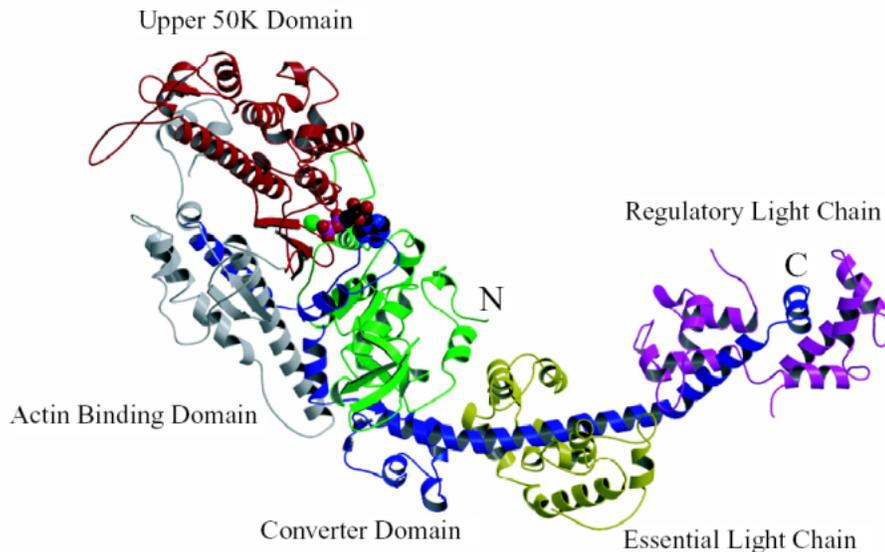


Figure 1.2: Myosin-S1: The regulatory light chain (magenta) and the essential light chain (yellow). The proteolytic fragments are color coded as follows: 25K (N terminal), green; 50K, red; and 20K (C terminal), blue. The 50K fragment spans two domains: the 50K upper domain and the 50K lower domain or actin-binding domain. The actin-binding domain has been colored grey. Figure taken from.²

actin-binding site.

1.4 ATOMIC STRUCTURE OF F-ACTIN

Actin has two conformations, G (globular) and F (fibrous). Crystal structures are only available for the G-form. The only high resolution data available for F-actin is from fiber diffraction patterns.²³ A number of attempts have been made to fit a helix of G-actin monomers to the observed F-actin, fiber diffraction pattern by refinement methods.^{24,25} Since the fiber-diffraction patterns are of limited resolution (6 to 8Å), the refinement is under determined. As a result, these various methods produce related, but different structures. The four sub-domains of G-actin were allowed to move as independent solid bodies so as to minimize the difference between the calculated fiber diffraction pattern and the observed fiber diffraction pattern.²⁶ The sub-domains were constrained to remain connected. After the refinement, the stereo-chemistry of residues in the junction regions

was optimized using the CNS software.²⁷

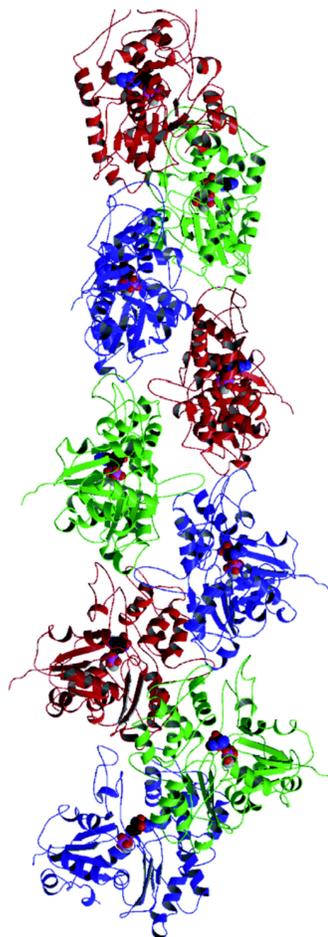


Figure 1.3: F-Actin. Figure taken from reference.²

Thus constituted, thin filaments (F-actin) are helical polymers which have 13 actin molecules (42 kDa) arranged on six left-handed turns repeating every 36 nm. The rise per subunit is 2.75 nm.²⁸ The morphology of the actin helix is, rather, two intertwined, steep right-handed helices. Along each of the morphological helices the actin monomers are spaced by 5.5 nm.²⁸ The structure of the monomer (G-actin) was solved by protein crystallography as a complex with DNAase I²³ and has since been solved in two other complexes.^{29,30} The structure shows actin to consist of two similar domains each of which contains a 5-stranded- β -sheet and associated α -helices. The phosphate moiety of a nucleotide (ATP or ADP), together with Mg^{2+} or Ca^{2+} (Mg^{2+} is physiological), is bound between the two β -sheet domains. Each of the domains carries a sub-

domain; one is involved in actin-actin interactions, and the other in addition forms the top of the nucleotide-binding pocket.

1.5 STRATEGIES FOR MOVEMENT

Different motors can adopt different strategies on how to move along their tracks. Which strategy is used depends on the function of the motor. Motor proteins can move along their tracks either processively or non-processively. Processivity is defined as the average number of steps taken per diffusional encounter between a motor and its track.³¹ Another key feature characteristic of a specific motor is its directionality, i.e., in which direction it moves along its track.¹⁰ Myosin II is a classical example of a non-processive motor that detaches from its track at each ATP turnover. This is adequate for myosin II because many myosin II monomers self-assemble into filaments that slide relative to its track along actin filaments. Unlike Myosin V, the cargo in the case of Myosin II is the myosin filament itself.

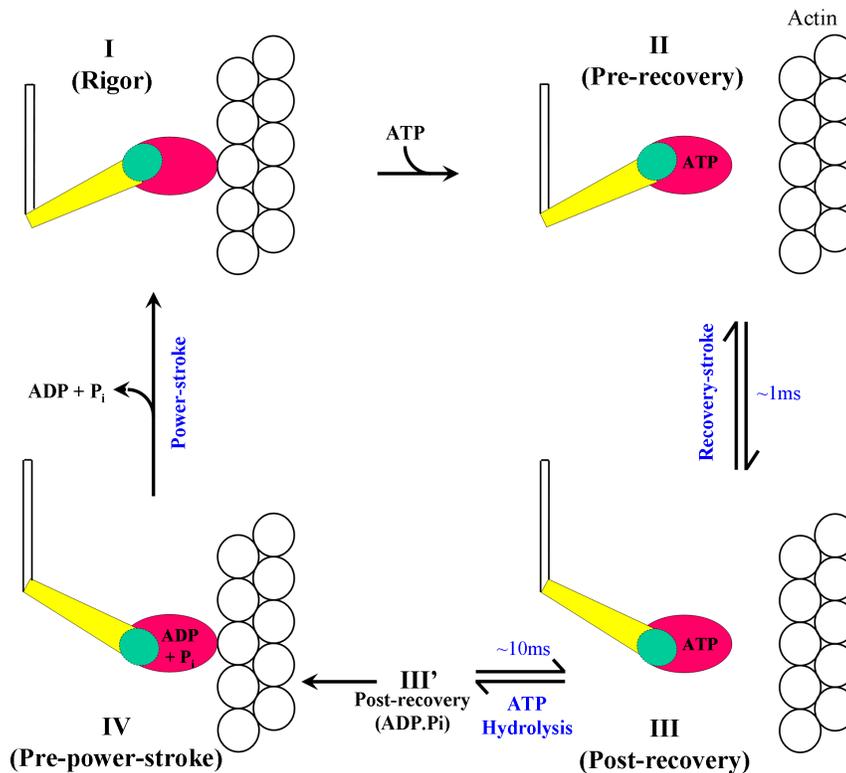


Figure 1.4: Lymn-Taylor Cycle.

1.6 LYMN-TAYLOR CYCLE

All members of the myosin family interact cyclically with actin, thereby moving past the actin filaments. The current view of how myosin and actin interact to produce movement non-processively is envisaged as the Lymn-Taylor cycle^{2,19,32} depicted in Figure 1.4 which outlines several coupling mechanisms that link structural changes in different parts of myosin. Structural changes associated with ATP binding to the myosin-rigor conformation of (State I) substantially reduces myosin-actin binding affinity^{33,34} thus dissociating it from actin (Figure 1.4, Step I → Step II). Myosin, then activates its ATPase function which is coupled to the 60° rotation of the converter domain that is associated with the “Recovery stroke” (Step II → Step III).^{2,12,35–37} ATP hydrolysis (State III → State III’) increases myosin-actin binding affinity which causes myosin to rebind to actin (State III’ → State IV). Actin binding triggers the conformational change associated with the back-rotation of the converter domain which brings about the rowing-like stroke referred to as the “Power stroke” (State IV → State I) which returns myosin to its rigor-conformation.^{2,19}

1.7 MYOSIN II MOTOR IN *Dictyostelium discoideum*

Dictyostelium discoideum is a slime mold that grows as a unicellular amoebae. Because of its unique features combining unicellular and multicellular characteristics, *Dictyostelium discoideum* is chosen as a model organism since they form a multicellular mound consisting of up to 100,000 cells, quite easily. In addition, the localization of specific proteins in the living cell at different stages of the cell cycle can be monitored using fluorescence confocal microscopy.³⁸ This organism has been used extensively to study the functionality of myosins.

The head of the *Dictyostelium discoideum* Myosin II motor (henceforth, myosin) has been crystallized in the absence of actin with different ATP analogues bound to the ATPase site, (Mg·ATP· γ S,³⁹ Mg·AMP·PNP,³⁹ Mg·ADP·BeF_x,⁴ Mg·ADP·AlF₄⁻,⁴ Mg·ADP·V_i⁵ etc.) under various conditions (Please see, Table 1.1 for a list of all crystal structures of *Dictyostelium discoideum* myosin) The converter-domain, which bears the lever arm is found in two orientations, in which the domain rotates by about 60° with respect to the rest of the head (Figures 1.4, 1.6A & 1.6C, left panels). The two orientations of the converter domain are thought to be the end orientations of the power-stroke (or the recovery-stroke).^{2,4,5} Myosin is also crystallized in State

Table 1.1: Crystal structures for *Dictyostelium discoideum* myosin II motor.

PDB ID	Resolution [Å]	Construct ^a	Ligand	State ^b
1D0X	2.00	759	m-nitrophenyl aminoethyl.PPi.BeF3	C/O, State II
1D0Y	2.00	759	o-nitrophenyl aminoethyl.PPi.BeF3	C/O, State II
1D0Z	2.00	759	p-nitrophenyl aminoethyl.PPi.BeF3	C/O, State II
1D1A	2.00	759	o,p-dinitrophenyl aminoethyl.PPi.BeF3	C/O, State II
1D1B	2.00	759	o,p-dinitrophenyl aminoethyl.PPi.BeF3	C/O, State II
1D1C	2.30	759	N-methyl-o-nitrophenyl aminoethyl.PPi.BeF3	C/O, State II
1FMV	2.10	759	Empty	C/O, State II
1FMW	2.15	759	Mg.ATP	C/O, State II
1G8X	2.80	761.	Mg.ADP	C/O, State II
		α -actinin-R238E		
1LVK	1.90	759. Q760L, R761P, I762N	Mg.BeF3.mantADP	C/O, State II
1MMA	2.10	759. Q760L, R761P, I762N	Mg.ADP	C/O, State II
1MMD	2.00	759. Q760L, R761P, I762N	Mg.ADP.BeF3	C/O, State II
1MMG	1.90	759. Q760L, R761P, I762N	Mg.ATP. γ S	C/O, State II
1MMN	2.10	759. Q760L, R761P, I762N	Mg.AMPPNP	C/O, State II
1MND	2.60	690. Q760L, R761P, I762N	Mg.ADP.AIF4	C/C, State III
1MNE	2.70	759 Q760L, R761P, I762N	Mg.PPi	C/O, State II
1Q5G	1.90	Myosin fused to dynamin	Empty	O/O, State I
1VOM	1.90	747	Mg.ADP.VO4	C/C, State III
REF1	-	759	Mg.ADP.BeF3	C/C, State III

^aThe last resolved amino acid is given. Mutations are indicated.

^bThe conformational states of switch-1/switch-2 are indicated. “O” refers to the open state, while “C” refers to the closed state.

I (Rigor conformation or near-rigor conformation) in both Myosin V³³ and *Dictyostelium discoideum* Myosin II³⁴ with no nucleotide bound to the ATP site (apo state). The change in orientation of the converter domain (residues 692-748) is associated to a change in the conformation of the “Relay-helix” (residues 466-498) which spans from the converter domain to the ATP binding site. The N-terminus end of the relay-helix responds to the presence of ATP which in turn triggers the C-terminus end (which is tightly bound to the converter-domain) to undergo a tilt in its helical axis and cause a local unwinding between the two ends of the recovery stroke (compare Figure 1.6 A & Figure 1.6C).

1.8 NUCLEOTIDE BINDING SITE

The nucleotide binding site is located at the interface between the 50 kDa and the N-terminal subdomains (Figures 1.2, 1.5 & 1.6). It is composed of three loops that are conserved not only among motor proteins but also among the G-proteins. These are the P-loop that is a common feature of a large number of enzymes that bind nucleotides,²² the switch-1 loop and the switch-2 loop. Together they form the so-called “Phosphate tube” (Table 1.2). The switch-1 and switch-2 loops are located in the upper and lower 50 kDa domains, respectively, whereas the P-loop belongs to the N-terminal 25 kDa domain. The two switch loops got their names from the observation that they can adopt different conformations, thus serving as a switch for information transduction by changing their conformation.³⁷ The nucleotide binding site undergoes conformational changes upon ATP binding.^{33,34} Between the two ends of the recovery stroke, the switch-2 loop closes upon the γ -phosphate of ATP so that the amide group of a well-conserved Gly457⁴⁰ can make a hydrogen bond with one of the oxygens of the γ -phosphate which is considered to be essential for enabling ATPase function.^{2,36,37}

1.9 CRYSTAL STRUCTURE CLASSIFICATION

The positions of the switch-1 and switch-2 can be used to classify different conformational states of myosin in the Lymn-Taylor scheme.¹⁹ Conformations corresponding to switch-1 open / switch-2 open (O/O), is assigned to State I, switch-1 closed / switch-2 open (C/O), to State II and switch-1 closed / switch-2 closed (C/C) to State III^{33,34,41} (refer Table 1.3).

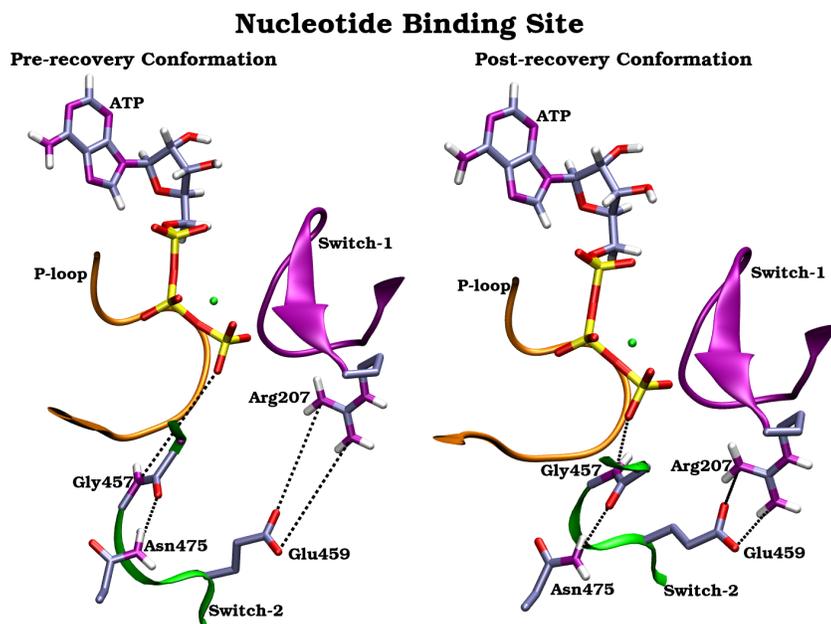


Figure 1.5: The Nucleotide Binding Site.

Legend: P-loop in Orange; Switch-2 in green; Switch-1 in purple. Phosphorus atoms of ATP in yellow and magnesium in green.

Table 1.2: Consensus sequences.

Loop	Consensus ^a	<i>Dictyostelium discoideum</i>	residues ^b
P-loop	GESGAGKT	GESGAGKT	179-186
Switch-1	NxNSSR	NNSSR	233-238
Switch-2	DxSGFE	DISGFE	454-459

^a Conserved in at least 80 out of 82 myosins¹

^b Residue numbers of *Dictyostelium discoideum* myosin II are used here throughout.

All states, except State IV (Pre-power-stroke conformation) have been identified by crystallography. The states are characterized by a well-defined γ -phosphate binding site that can be occupied either by phosphate itself or by γ -phosphate analogs such as vanadate, beryllium fluoride, or magnesium fluoride.⁴² Table 1.1, lists crystal structures with different nucleotides bound to the active site and the conformations adopted by switch-1 and switch-2 loops.

Recently, Holmes and co-workers⁴¹ comprehensively classified all the states with the available data from electron-density maps,⁴³ cryo-electron micrographs⁴¹ and X-ray crystallography. They included data from P-loop conformation along with the conformations of the relay

helix, central- β -sheet, converter domain and whether the actin-binding cleft is open or closed (See Table 1.3). They propose that two conditions are required for attaining the pre-power stroke conformation: switch 2 closed and β -sheet not twisted to have the relay helix kinked. Relaxing either of these conditions allows the relay helix to straighten and rotate the converter domain. Thus both the rigor-like and the post-rigor structures have a straight relay helix. This ability of the myosin cross-bridge to respond to two (or maybe more) input parameters has prompted Coureux and co-workers³³ to refer to the the β -sheet, which according to their notation is distorted (twisted) in rigor and relieved in post rigor, along with other structural elements that allow this distortion as “The Transducer” and propose that this is the “central processor” of the motor.^{33,44}

Table 1.3: Assignment to the states in the Lymn-Taylor cycle.

State	Switch-1	Switch-2	Actin-binding Cleft	β -Sheet	P-loop	Relay-Helix	Converter domain
I (Rigor) (apo-state)	Open	C ₁	Shut	Maximum twist	Up	Straight	Down
II (Pre-recovery) (ATP bound)	Closed	Open	Open	No-twist	Down	Straight	Down
III (Post-recovery) (ATP bound)	Closed	C ₂	Open	No-twist	Down	Kink	Up
IV (Strongly attached to actin) (ADP·Pi)	<i>closed</i>	C ₂	<i>closed</i>	No-twist	Down	Kink	Up

The values of the seven movable elements that have been identified for the states defined in the Lymn-Taylor cycle. C₁ and C₂ are two forms of closed conformation found in the rigor and the pre-power-stroke states. The assumed properties of conformation (shown in “bold” italic) are extrapolated from the adjoining states in the Lymn-Taylor cycle.

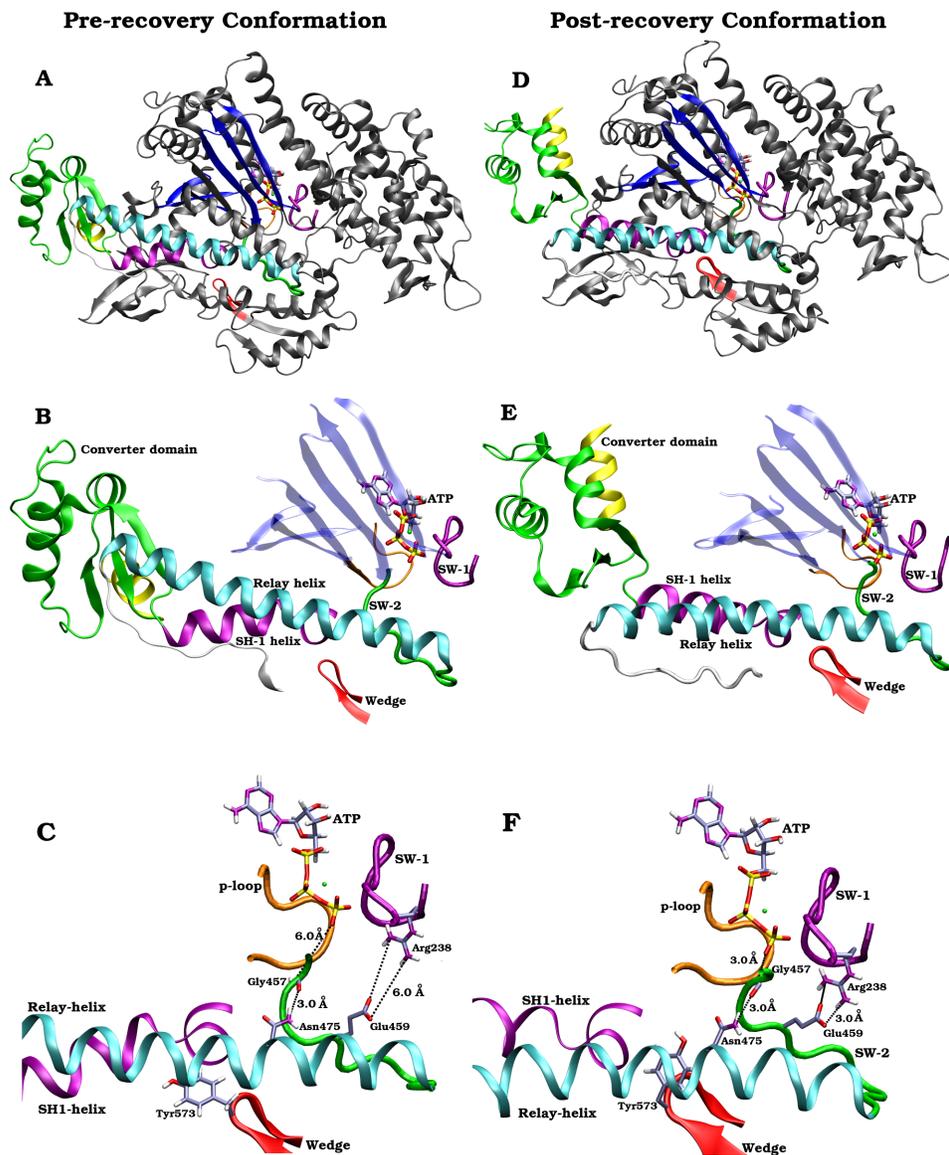


Figure 1.6: Structural elements of myosin involved in the recovery-stroke.

Legend: P-loop in Orange; Switch-2 in green; Relay helix in cyan; SH-1 helix in purple; SH-2 helix in pink; converter domain in green; lever arm in yellow. In the right panel, the important residues and bond distances involved in the recovery stroke are indicated. The See-saw and the wedge phases movements are shown with solid arrows.

1.10 COUPLING MECHANISMS IN MYOSIN II

The Lymn-Taylor cycle is made functional by several mechanisms that link structural changes in different parts of myosin. A small structural difference between having ATP/Mg²⁺ versus

ADP·Pi/Mg²⁺ bound to the nucleotide binding site, myosin controls the conformation of its actin binding region that determines whether the binding affinity for actin is low in State I → II or high in State III' → IV.^{13,33,34,45} The ATPase function in myosin is shown to be activated only when the Gly457/Ser456 peptide group, belonging to the Switch-2 loop, closes to make a hydrogen bond with the γ -phosphate.^{45,46} This hydrogen bond is absent in State II and present in State III and is thought to occur during the recovery-stroke transition.^{2,12,36} However, it would be wasteful if the Switch-2 loop could freely close and re-open with the lever arm still in the pre-recovery orientation, allowing ATP hydrolysis and an unproductive product release. This suggests that a mechanism is required to couple the closing of Switch-2 loop with the orientation of the converter domain to ensure that ATP is hydrolyzed only when the converter domain/lever arm is in the post-recovery orientation. Once ATP is hydrolyzed, a “locking mechanism” must prevent the reversal of the recovery stroke while the nucleotide is in the ADP·Pi state in State III' and not bound to actin. Rebinding to actin after the reversal of the recovery stroke would be wasteful since it could lead to the release of the hydrolysis products without generating force.

It is extremely difficult to study the mechanisms involved from State I → State II and State III' → State IV since State I and State IV require crystal structures with myosin and actin bound together. There are no such high-resolution X-ray crystal structures with myosin and actin bound together and data from fluorescence experiments,⁴⁷ electron-density maps⁴³ and cryo-electron micrographs⁴¹ are unreliable due to their low resolution. However, there are 2 X-ray crystal structures available for Myosin II³⁴ and Myosin V³³ which are thought to represent State I (Rigor conformation). On the other hand there are ~20 structures solved for states II and III (the ends of the recovery stroke). Hence, a lot of research work has been done to understand the recovery stroke mechanism which is the topic of interest in this thesis.^{36,37}

1.11 RECOVERY STROKE COUPLING MECHANISM

The Myosin head undergoes a reversible conformational transition called the “Recovery Stroke” or the Return stroke (State II \rightleftharpoons State III), in which the converter domain (which bears the lever arm) rotates by $\sim 60^\circ$. Simultaneously, myosin also activates its ATPase function, since hydrolysis of the bound ATP is required for strong actin binding.⁴⁵ The ATPase function in myosin is shown

Myosin II - Two end-states Three different views

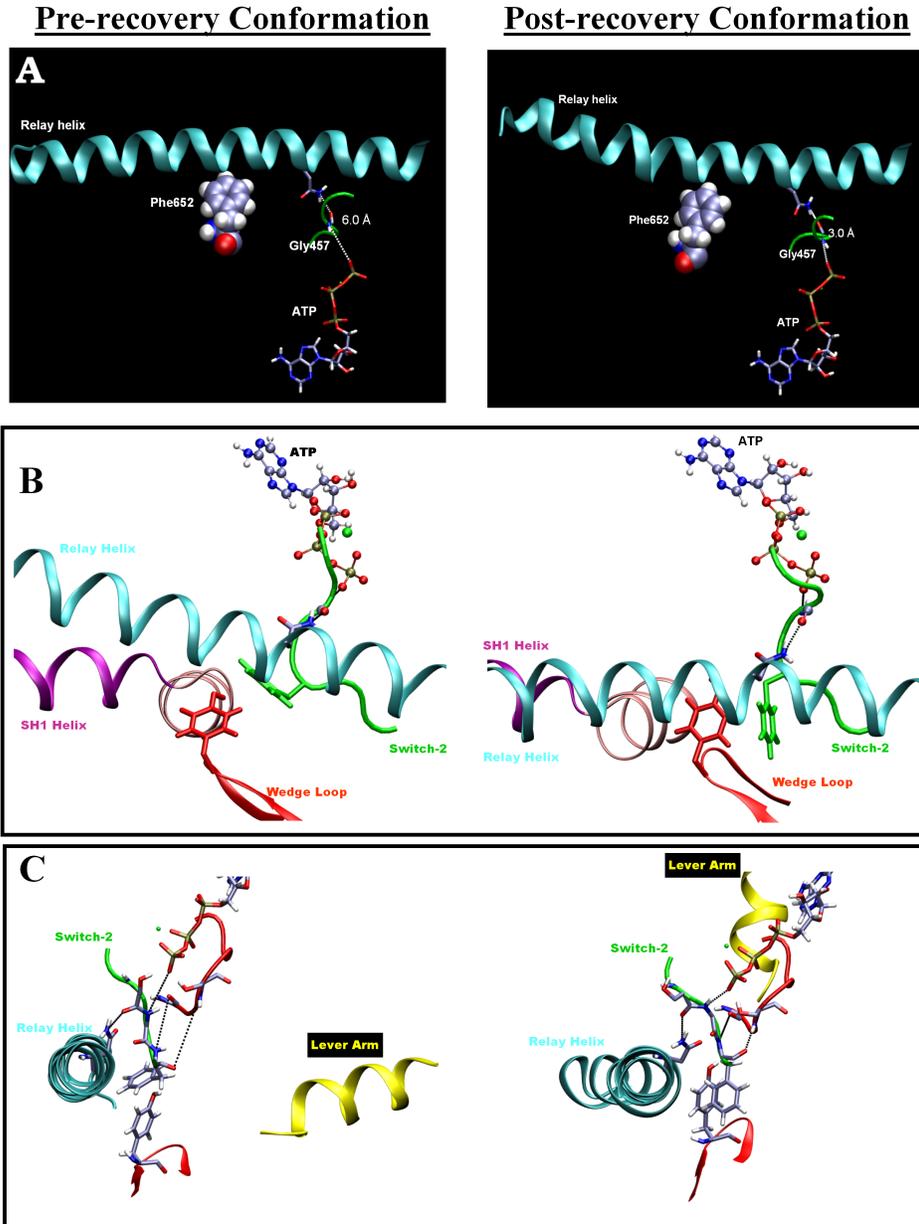


Figure 1.7: Three different views of the Myosin in the end-states of the recovery stroke. Legend: Switch-2 in green; Relay helix in cyan; SH-1 helix in purple; SH-2 helix in pink; lever arm in yellow. A. See-saw view; B. Wedge view; C. P-loop view.

to be activated when the Gly457/Ser456 peptide group, belonging to the Switch-2 loop, closes to make a hydrogen bond with the γ -phosphate.^{45,46} This hydrogen bond is absent in State II and present in State III and is thought to occur during the recovery-stroke transition.^{2,12,36} However, it would be wasteful if the Switch-2 loop could freely close and re-open with the lever arm still in the pre-recovery orientation, allowing ATP hydrolysis and an unproductive product release. This suggests that a mechanism is required to couple the closing of Switch-2 loop with the orientation of the converter domain to ensure that ATP is hydrolyzed only when the converter domain/lever arm is in the post-recovery orientation (i.e., ready to perform the power stroke).

In Figure 1.7, we have three different orientations of State II and State III (the ends of the recovery stroke. Figure 1.7A is the See-saw view where the N-terminus of the relay helix is near the ATP site while the C-terminus is connected to the converter domain. The relay helix is seen to have a “kink” in the post-recovery conformation while in the pre-recovery conformation its a straight helix. Figure 1.7B, is the Wedge view, where the Wedge loop (red) pushes into the N-terminal end of the SH1-helix (purple) and longitudinally translates it outward. Figure 1.7C, is the P-loop view, where the formation of two key hydrogen bonds around the ATP site which rotate the lever-arm (yellow) at the back.

1.12 QUESTIONS ADDRESSED IN THIS THESIS

In the present work, the return-stroke mechanism in myosin is investigated by performing minimum energy pathway calculations using an semi-empirical molecular mechanics force-field. Based on the calculations, unresolved questions regarding the return-stroke mechanism and its coupling to the mechanical events will be addressed.

The following questions will be addressed:

1. How do the interactions with ATP in State II contribute to the coupling between ATPase activation and the swinging of the lever arm during the recovery stroke? (Addressed in Chapter 3)
2. How does ATP hydrolysis in the post-recovery conformation affect the structure of the active site? (Addressed in Chapter 4)

3. How do these changes prevent the reversal of the recovery stroke while the nucleotide is in the ADPPi state? (Addressed in Chapter 4)
4. Can the available mutational data be rationalized ? (Addressed in Chapters 3 & 4)

1.13 OUTLINE FOR THE THESIS

In the following chapters, the theory and methods that were applied to answer the above question will be first introduced. Once the background is well-understood, the first question will be then addressed to rationalize a model as to how ATPase activation and the recovery-stroke of the converter domain are coupled. The remaining questions will be addressed in the following chapter which elaborates about a well-conserved loop structure which was first the first time recognized to explain nucleotide-sensing mechanism that controls the recovery stroke. The thesis is concluded with an outlook.

THEORETICAL ASPECTS

Ideally, if one were to address projects involving understanding protein function using *in-silico* methods, quantum mechanics would be the first choice. Unfortunately, quantum mechanical methods that deal with electrons between each nuclei in the system could be highly time consuming and not practical. On the other-hand, force-field methods (also called Molecular Mechanics) ignore the electronic motions and calculate the energy of the system as a function of nuclear positions only. These methods work quite well due to the validity of several assumptions. The first and the most important is the “Born-Oppenheimer Approximation”.^{48,49} The basis to this approximation that the nuclei are so much more massive than the electrons, the electrons must instantaneous relax or rearrange to the configuration of the nuclear positions.

Hence, under the Born-Oppenheimer approximation the total wave-function for the molecule can be written in the following form:

$$\Psi_{molecule}(\vec{r}_i, \vec{R}_j) = \Psi_{electrons}(\vec{r}_i, \vec{R}_j) \cdot \Psi_{nuclei}(\vec{R}_j) \quad (2.1)$$

Further, molecular mechanics is based upon a simple model of the interactions within a system with contributions from processes such as stretching of bonds, the opening and closing of angles, the rotations about single bonds and non-bonded interactions like the electrostatic and van der Waals forces between each pair of atoms. Simple functions (like the Hooke’s law) are often used to describe these contributions and the method performs quite acceptably. Another feature is the transferability as a set of parameters developed and tested on a relatively small molecules can be directly applied to study large molecules such as polymers and proteins.

2.1 A MOLECULAR MECHANICS FORCE FIELD

A Molecular mechanics force field can be interpreted in terms of a relatively simple two-component picture of bonded and non-bonded interactions within the system. The bonded-interactions are modeled as energetic penalties that are associated with the deviation of bonds and angles away from their “equilibrium” or reference values. A function describes the change in energy when bonds are rotated or stretched from the equilibrium position. In simple force-fields, non-bonded interactions are modeled as a Coulombic potential term for electrostatic interactions and Lennard-Jones potential for van der Waals interactions.

$$V(\vec{R}) = V(\vec{R})_{bonded} + (\vec{R})_{non-bonded} \quad (2.2)$$

$$V(\vec{R})_{bonded} = \sum_{bonds} k_b(b - b_0)^2 + \sum_{angles} k_\theta(\theta - \theta_0) + \sum_{dihedral} k_\chi[1 + \cos(n\chi - \sigma)] + \sum_{impropers} k_\omega(\omega - \omega_0)^2 + \sum_{Urey-Bradley} k_u(u - u_0)^2 \quad (2.3)$$

$$V(\vec{R})_{nonbonded} = \sum_{nonbonded} (\epsilon_{ij}[(\frac{R_{min,ij}}{r_{ij}})^{12} - (\frac{R_{min,ij}}{r_{ij}})^6] + \frac{q_i q_j}{\epsilon_D r_{ij}}) \quad (2.4)$$

The first term in the energy function accounts for the bond stretches where k_b is the bond force constant and $b - b_0$ is the distance from equilibrium that the atom has moved (Figure 2.1). The second term in the equation accounts for the bond angles where k_θ is the angle force constant and $\theta - \theta_0$ is the angle from equilibrium between 3 bonded atoms. The third term is for the dihedral (a.k.a. torsion angles) where k_χ is the dihedral force constant, n is the multiplicity of the function, χ is the dihedral angle and σ is the phase shift. The fourth term accounts for the impropers, that is out of plane bending, where k_ω is the force constant and $\omega - \omega_0$ is the out of plane angle. The Urey-Bradley component (cross-term accounting for angle bending using 1,3 non-bonded interactions) comprises the fifth term, where k_u is the respective force constant and $u - u_0$ is the distance between the 1,3 atoms in the harmonic potential. Non-bonded interactions between pairs of atoms (i,j) are

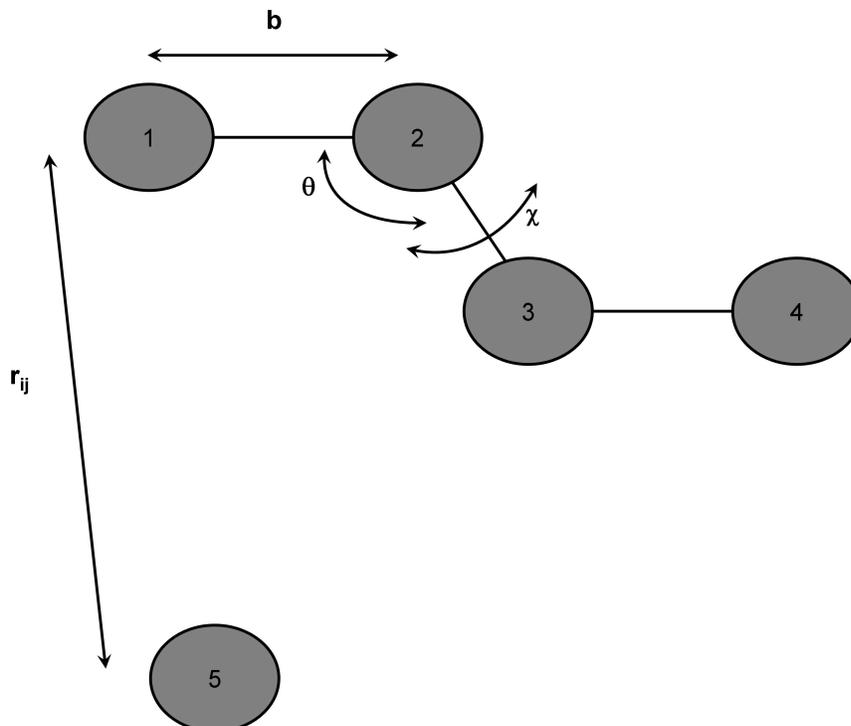


Figure 2.1: Hypothetical atoms to illustrate the energetic terms included in the force-field. An Molecule comprises of atoms 1-5. Internal terms that occur in molecule are the bonds, b , between atoms 1 & 2, 2 & 3, 3 & 4; angles θ , involving atoms 1-2-3 and 2-3-4 and a dihedral or a torsional angle χ , described by atoms 1-2-3-4. Atom 5 is involved in a "Non-bonded" interaction with atoms 1, 2, 3, & 4. The energy of this interaction is determined by the distance, r_{ij} between atom 5 and rest of the molecule.

represented by the last two terms. By definition, the non-bonded forces are only applied to atom pairs separated by at least three bonds. The van Der Waals (VDW) energy is calculated with a standard Lennard-Jones potential and the electrostatic energy with a Coulombic potential. In the Lennard-Jones potential above, the $R_{min,ij}$ term is not the minimum of the potential, but rather where the Lennard-Jones potential crosses the x-axis (i.e. where the Lennard-Jones potential is zero) (Figure 2.2). In order to perform minimization, calculate minimum energy pathways (MEP) or run the MD simulations, we need to be sure that a parameter file that has all the parameters specified in the energy function we are using (i.e. the equation specified above) and the topology file describing the distribution of different atoms around each amino acid is well-defined.

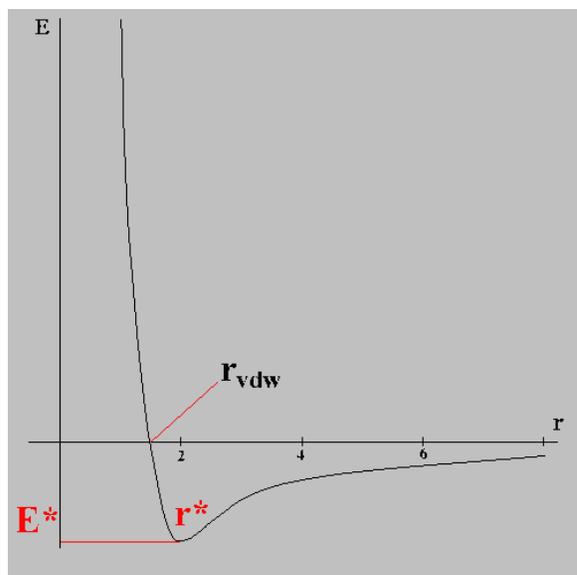


Figure 2.2: Lennard-Jones potential to account for van Der Waals (VDW)

2.1.1 TREATMENT OF THE NON-BONDED ENERGY TERMS

The most time consuming part of an energy minimization or a molecular dynamics simulation is the calculation of the non-bonded terms in the potential energy function, e.g., the electrostatic and van der Waals forces. In principle, the non-bonded energy terms between every pair of atoms should be evaluated; in this case, the number of increases as the square of the number of atoms for a pairwise model (N^2). To speed up the computation, the interactions between two atoms separated by a distance greater than a pre-defined distance, the cutoff distance, are ignored. Several different ways to terminate the interaction between two atoms have been developed over the years; some work better than others (see Figure 2.3).

Truncation: the interactions are simply set to zero for inter-atomic distances greater than the cutoff distance. This method can lead to large fluctuations in the energy. This method is not often used.

The SHIFT cutoff method: this method modifies the entire potential energy surface such that at the cutoff distance the interaction potential is zero. The drawback of this method is that equilibrium distances are slightly decreased.

The SWITCH cutoff method: This method tapers the interaction potential over a pre-defined range of distances. The potential takes its usual value up to the first cutoff and is then

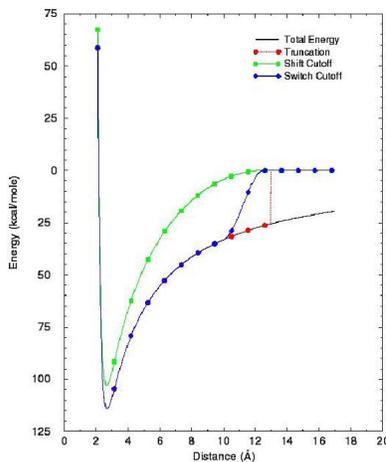


Figure 2.3: Effect of truncation schemes on the energy of a given conformation

switched to zero between the first and last cutoff. This model suffers from strong forces in the switching region which can slightly perturb the equilibrium structure. The SWITCH function is not recommended when using short cutoff regions.

2.1.2 LONG-RANGE ELECTROSTATIC INTERACTIONS

In recent years, a number of models have been introduced which permit the inclusion of long-range electrostatic interactions in molecular dynamics simulation. For simulations of proteins and enzymes in a crystalline state, the Ewald summation is considered to be the correct treatment for long range electrostatic interactions.⁵⁰ Variations of the Ewald method for periodic systems include the particle-mesh Ewald method.^{51,52}

2.1.3 THE EXTENDED ELECTROSTATICS MODEL

The Extended Electrostatics model approximates the full electrostatic interaction by partitioning the electric potential and the resulting forces on the atom at R_i into a "Near" and an "Extended" contribution. The "Near" contribution arises from the charged particles which fall within the sphere defined by the cutoff distance R_{cut} , while the "Extended" contribution, arises from the particles which are beyond the cutoff distance R_{cut} . The "Near" contribution is calculated by a conventional pairwise sum and the "Extended" contribution to the potential at R_i is calculated using a multi-pole approximation.⁵³

2.1.4 TREATMENT OF SOLVENT IN COMPUTER SIMULATIONS

Solvent, usually water, has a fundamental influence on the structure, dynamics and thermodynamics of biological molecules, both locally and globally. One of the most important effects of the solvent is the screening of electrostatic interactions. The electrostatic interaction between two charges is given by Coulomb's law,

$$V_{elec} = \frac{q_i \cdot q_j}{\epsilon_{eff} \cdot r_{ij}} \quad (2.5)$$

where q_i, q_j are the partial atomic charges, ϵ_{eff} is the effective dielectric constant and r_{ij} is the relative distance between the two particles. It is important to include solvent effects in an computer simulations. This can be done at several levels. The simplest treatment is to simply include a dielectric screening constant in the electrostatic term of the potential energy function. In this implicit treatment of the solvent, water molecules are not included in the simulation but an effective dielectric constant is used. Often the effective dielectric constant is taken to be distance dependent, $\epsilon_{eff} = r_{ij} \cdot \epsilon$, where ϵ ranges from 4 to 20. Although this is a crude approximation, it is still much better than using unscreened partial charges. Other implicit solvent models have been developed that range from the relatively simple distance-dependent dielectric constants to models that base the screening on the solvent exposed surface area of the protein. The distance-dependent dielectric coefficient is the simplest way to include solvent screening without including explicit water molecules and it is available in most simulation programs. Recently, several implicit solvent models based on continuum electrostatic theory have been developed.⁵⁴⁻⁶¹

2.2 CONTINUUM ELECTROSTATICS

The Solvation free energy (ΔG_{sol}) is the free energy change, that occurs, when you transfer a molecule from vacuum to a solvent of high dielectric ϵ . This Solvation free energy (ΔG_{sol}) has three components

$$\Delta G_{sol} = \Delta G_{elec} + \Delta G_{vdw} + \Delta G_{cav} \quad (2.6)$$

The electrostatic component ΔG_{elec} , is particularly important for polar and charged

solutes due to the polarization of the solvent which is modeled as a uniform medium of constant dielectric ϵ . The Van der Waals interaction term ΔG_{vdw} can further be split into attractive term, ΔG_{att} and a repulsive term, ΔG_{rep} . The ΔG_{cav} is the free energy required to form the solute cavity within the solvent. ΔG_{cav} is positive and describes the entropic penalty associated with the reorganization of the solvent molecules around the solute together with the work done against the solvent pressure in creating the cavity.

To get the accurate evaluation of ΔG_{elec} (The electrostatic component of Solvation free energy) for example the molecule in figure 1, one can follow one of the following methods:

1. Using simple Coulomb electrostatics
2. Solution to the Poisson-Boltzmann Equation.
3. Generalized Born Methods.

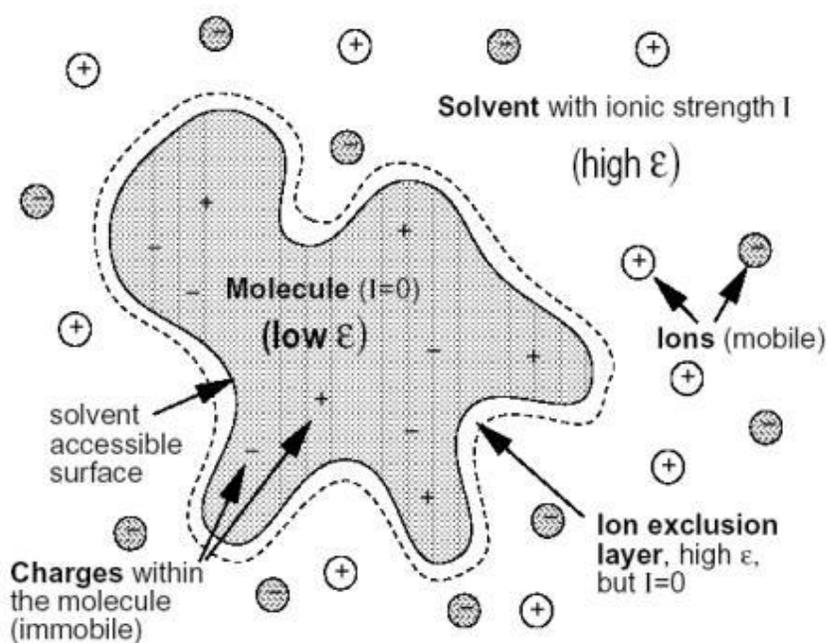


Figure 2.4: A molecule in a heterogeneous dielectric medium.

2.2.1 SIMPLE COULOMB ELECTROSTATICS

Coulomb's law describes the interaction force between two static point charges. Suppose we have two charges q_1 and q_2 , then these two charges interact with each other with a force given by the Coulomb equation:

$$F = C \frac{q_1 q_2}{r^2} \quad (2.7)$$

Where F is the force of interaction between the charges q_1 and q_2 , C is a constant which is given by $\frac{1}{4\pi\epsilon_0\epsilon}$ and r is the magnitude of the distance between the two charges. The permittivity of vacuum, $\epsilon = 1$, i.e., when the charges are placed in vacuum.

For Evaluating ΔG_{elec} , using the Coulomb model, we do not account for the change in dielectric when one moves from the solute environment to that of the solvent. Usually the charges are screened by solvent and this effect is not considered in this simple treatment. This model also ignores the reaction field of the solvent.

2.2.2 THE POISSON-BOLTZMANN EQUATION

A Brief Derivation

To describe the electrostatic component of Solvation energy ΔG_{elec} in molecules, like the one depicted in figure 2.4, with complete accuracy, the Poisson Equation can be used. The Poisson equation relates the variation of the potential ϕ within a medium of uniform dielectric constant ϵ to the charge density ρ .

$$\nabla^2 \phi(\vec{r}) = -\frac{4\pi\rho(\vec{r})}{\epsilon} \quad (2.8)$$

The Poisson Equation is thus a second-order partial differential equation. For a set of point charges in constant dielectric medium, the Poisson equation (equation 2.8) reduces to the simple Coulombs law equation (equation 2.7).

If the dielectric ϵ is not constant, but changes with position \vec{r} then the Poisson equation adopts the following form:

$$\nabla \cdot [\epsilon(\vec{r}) \nabla \phi(\vec{r})] = -4\pi\rho(\vec{r}) \quad (2.9)$$

If mobile ions are introduced into the medium, their distribution in response to the electric potential can be accounted by the Poisson Equation. The ions are prevented from congregating at the locations of extreme potential due to the repulsive interactions with the each other and natu-

ral thermal motion in the solvent. The ion distribution in the solvent is described by the Boltzmann distribution of the following form:

$$n(\vec{r}) = \mathcal{N} \exp(-\mathcal{V}(\vec{r})/k_B T) \quad (2.10)$$

Where $n(\vec{r})$ is the number density of ions at a particular point \vec{r} , \mathcal{N} is the bulk number density and $\mathcal{V}(\vec{r}) = \phi(\vec{r})$ in the mean-field approximation; ($\phi(\vec{r})$: the mean electric field produced by the charge distribution) is the energy change to bring the ion from infinity to the position \vec{r} , k_B is the Boltzmann Constant and T is the temperature.

When these effects are incorporated into the Poisson equation (equation 2.9), we obtain the Poisson-Boltzmann Equation

$$\nabla \cdot \epsilon(\vec{r}) \nabla \phi(\vec{r}) - \kappa' \sinh[\phi(\vec{r})] = -4\pi\rho(\vec{r}) \quad (2.11)$$

Where κ' is related to the Debye-Hückel inverse length κ , by:

$$\kappa' = \frac{\kappa^2}{\epsilon} = \frac{8\pi N_A e^2 I}{1000\epsilon k_B T} \quad (2.12)$$

Where I is the ionic strength of the solution and N_A is the Avogadro's Number.

Equation 2.11 is a non-linear partial differential equation and can be written in an alternative form by expanding the hyperbolic sine function as a Taylor series

$$\nabla \cdot \epsilon(\vec{r}) \nabla \phi(\vec{r}) - \kappa' \phi(\vec{r}) \left[1 + \frac{\phi(\vec{r})^2}{6} + \frac{\phi(\vec{r})^4}{120} + \dots \right] = -4\pi\rho(\vec{r}) \quad (2.13)$$

The linearized Poisson-Boltzmann Equation can be written by taking only the first term in the Taylor expansion, as

$$\nabla \cdot \epsilon(\vec{r}) \nabla \phi(\vec{r}) - \kappa' \phi(\vec{r}) = -4\pi\rho(\vec{r}) \quad (2.14)$$

Equation 2.14 is a partial differential equation and cannot be solved analytically for complex geometries. Hence, the Poisson-Boltzmann Equation is solved by a numerical method called "The Finite Difference Method", in which the protein (solute) is put in a cubic grid along with the "implicit solvent". Values of the electrostatic potential, charge density, dielectric constant and ionic strength are assigned to each grid point. The atomic charges usually don't

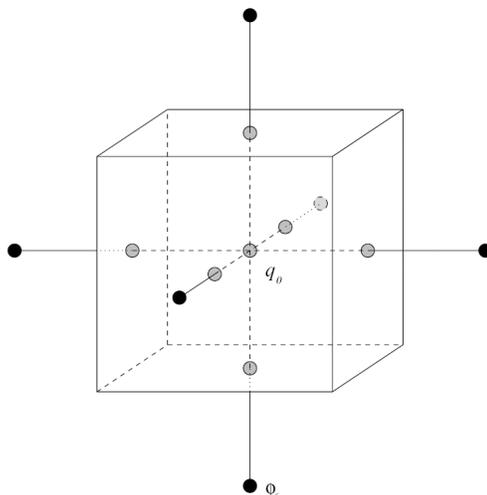


Figure 2.5: Part of the grid used to solve the PBEQ

coincide with the grid points. The charge allocation to each of the eight grid points is done in such a way that the closer the charge to the grid point the greater proportion of its total charge that is allocated. The derivatives in the Poisson-Boltzmann equation are then determined by the finite difference formula. Many Poisson-Boltzmann solvers are now available^{61–65}

The potential at a grid point shown in figure 2.5 having a charge q_0 associated with the grid point is given by

$$\phi_0 = \frac{(\sum_{i=1}^6 \epsilon_i \phi_i) + 4\pi \frac{q_0}{h}}{(\sum_{i=1}^6 \epsilon_i) + \kappa_0'^2 f(\phi_0)} \quad (2.15)$$

Where h is the grid spacing and the function $f(\phi_0)$ in the denominator has the value 1 for the linearized Poisson-Boltzmann Equation and is equivalent to the Taylor series expansion $(1 + \phi_0^2/6 + \phi_0^4/120 + \dots)$ for the non-linear case.

2.2.3 FOCUSING

There is a problem at the borders of the grid, since the grid points at the border have less than six neighboring points. So if the grid is much larger than the molecule the border grid points are far away from the molecule, that ϕ outside of the grid can be set to zero. If one has a huge grid the computational cost gets very expensive because more the number of grid points, more is the computation. So one could use a low resolution, huge grid and interpolate the electrostatic

potential from this calculation to a high resolution, small grid. Such focusing steps can be repeated, if necessary.

2.2.4 GENERALIZED BORN METHODS

BORN-ONSAGER MODELS

Important contributions to the electrostatic contribution to the free energy of Solvation were made by Born⁶⁶ and Onsager.⁶⁷ Born derived the electrostatic component of free energy of Solvation by placing a charge within a spherical solvent cavity. In Born's model, $\Delta G_{(elec)}$ of the ion (a charge in the a spherical solvent cavity) is equal to the work done in transferring the charge from vacuum to medium of high dielectric, ϵ . In other words, $\Delta G_{(elec)}$ is equal to the difference in work done to charge the ion in these two different environments. Onsager extended this model to a dipole in a spherical cavity.

Born showed that the work to charge the ion in a medium of dielectric constant ϵ is equal to $q^2/2\epsilon a$, where q is the charge on the ion, a is the radius of the ion (the radius of the cavity in the medium) and ϵ is the permittivity of the medium. The electrostatic component of free energy of Solvation is the difference in the work done in charging the ion in medium of high dielectric ϵ and in vacuum.

$$\Delta G_{elec} = \frac{q^2}{2a\epsilon} - \frac{q^2}{2a} \quad (\epsilon = 1, vacuum) \quad (2.16)$$

$$\Delta G_{elec} = -\frac{q^2}{2a} \left(1 - \frac{1}{\epsilon}\right) \quad (2.17)$$

2.2.5 THE GENERALIZED BORN EQUATION (GB EQUATION)

Consider a system of particles, each with a radii a_i and charge q_i . The total electrostatic free energy of such a system is given by the sum of coulomb energy and the Born energy of Solvation in the medium of relative permittivity ϵ . (This is just the correction to the coulomb energy in which the charges are screened according to the Born equation. If $\epsilon = 1$ as in vacuum, we end with normal Coulomb electrostatics in vacuum)

$$G_{elec} = \sum_{i=1}^N \sum_{j=i+1}^N \frac{q_i q_j}{\epsilon r_{ij}} - \frac{1}{2} \left(1 - \frac{1}{\epsilon}\right) \sum_{i=1}^N \frac{q_i^2}{a_i} \quad (2.18)$$

The first term of the above equation can be written as

$$\sum_{i=1}^N \sum_{j=i+1}^N \frac{q_i q_j}{\epsilon r_{ij}} = \sum_{i=1}^N \sum_{j=i+1}^N \frac{q_i q_j}{\epsilon r_{ij}} + \sum_{i=1}^N \sum_{j=i+1}^N \frac{q_i q_j}{r_{ij}} - \sum_{i=1}^N \sum_{j=i+1}^N \frac{q_i q_j}{r_{ij}} \quad (2.19)$$

On rearranging this equation we get,

$$\sum_{i=1}^N \sum_{j=i+1}^N \frac{q_i q_j}{\epsilon r_{ij}} = \sum_{i=1}^N \sum_{j=i+1}^N \frac{q_i q_j}{r_{ij}} - \left(1 - \frac{1}{\epsilon}\right) \sum_{i=1}^N \sum_{j=i+1}^N \frac{q_i q_j}{r_{ij}} \quad (2.20)$$

Substituting Equation 2.20 back into Equation 2.18, we obtain the GB Equation as a sum of three terms,

$$G_{elec} = \sum_{i=1}^N \sum_{j=i+1}^N \frac{q_i q_j}{r_{ij}} - \left(1 - \frac{1}{\epsilon}\right) \sum_{i=1}^N \sum_{j=i+1}^N \frac{q_i q_j}{r_{ij}} - \frac{1}{2} \left(1 - \frac{1}{\epsilon}\right) \sum_{i=1}^N \frac{q_i^2}{a_i} \quad (2.21)$$

Now we take the difference of the above equation in a solvent with permittivity of ϵ and in vacuum ($\epsilon = 1$). For the situation in vacuum the second and the third term in Equation 2.21 go to Zero because $\epsilon = 1$ in vacuum and we will have only the first term of Equation 2.21 left behind for the situation in vacuum.

So we can write,

$$\Delta G_{elec} = \left[\sum_{i=1}^N \sum_{j=i+1}^N \frac{q_i q_j}{r_{ij}} - \left(1 - \frac{1}{\epsilon}\right) \sum_{i=1}^N \sum_{j=i+1}^N \frac{q_i q_j}{r_{ij}} - \frac{1}{2} \left(1 - \frac{1}{\epsilon}\right) \sum_{i=1}^N \frac{q_i^2}{a_i} \right] - \left[\sum_{i=1}^N \sum_{j=i+1}^N \frac{q_i q_j}{r_{ij}} \right] \quad (2.22)$$

Which results in,

$$-\left(1 - \frac{1}{\epsilon}\right) \sum_{i=1}^N \sum_{j=i+1}^N \frac{q_i q_j}{r_{ij}} - \frac{1}{2} \left(1 - \frac{1}{\epsilon}\right) \sum_{i=1}^N \frac{q_i^2}{a_i} \quad (2.23)$$

If we notice equation 2.23 carefully, for just one charge (ion), $r_{ij} = 0$ and the above equation reduces to the familiar Born equation (Equation 2.17). So the Born equation for single ion is generalized to a system having N charges. So the name **“Generalized Born Equation”**

Still and co-workers,⁵⁷ combined the two terms in equation 2.23 into one term and re-wrote Equation 2.23 as follows

$$\Delta G_{elec} = -\frac{1}{2} \left(1 - \frac{1}{\epsilon}\right) \sum_{i=1}^N \sum_{j=1}^N \frac{q_i q_j}{f(r_{ij}, a_{ij})} \quad (2.24)$$

The function $f(r_{ij}, a_{ij})$ depends on the inter-charge distance, r_{ij} and the Born radii, a_{ij} of each ion. One can seek this function to be used in the Equation 2.24, such that in the self ($i = j$) terms, it acts as the “Effective Born Radii” and in the pair-wise terms ($i \neq j$), it acts as an “Effective Interaction distance”. The most common form chosen is

$$f(r_{ij}, a_{ij}) = [r_{ij}^2 + R_i R_j \exp(-r_{ij}^2/4R_i R_j)]^{\frac{1}{2}} \quad (2.25)$$

Where, R_i and R_j is the “Effective Born Radii” of atom i and j respectively, which not only depends on a_i and a_j but also on the relative positions of all other atoms. If we have $r_{ij} = 0$, i.e., we are looking at the case where $i = j$, then the function interpolates to radius of that particular ion. ($(R_i^2)^{\frac{1}{2}} = R_i$) and when we have two ions far apart then $\exp(-r_{ij}^2/4R_i R_j)$ tends to 0 when r_{ij} becomes large, so the function interpolates in this case to $(r_{ij}^2)^{\frac{1}{2}} = r_{ij}$

Ideally, R_i (effective Born radii) should be so chosen that if one were to solve the Poisson Equation for a single charge q_i placed at position of atom i , and the dielectric boundary determined by all the molecule’s atoms and their respective radii with no charge on them, then the self energy of the charge i in this reaction field, $q_i \phi_{reac}(r_i)/2$ would be equal to $-(q_i^2/2R_i)(1 - 1/\epsilon)$. i.e.,

$$\Delta G_{elec} = q_i \frac{\phi_{reac}(\vec{r}_i)}{2} = -\frac{q_i^2}{2R_i} \left(1 - \frac{1}{\epsilon}\right) \quad (2.26)$$

Obviously, this procedure per se would not have a practical advantage over the GB methods if one were to solve the Poisson-Boltzmann Equation to get the effective

Born radii. To find a more rapid way to calculate the effective born radii, we can turn to the formulation of electrostatics in terms of energy density $u(\vec{r})$.

2.2.6 ANALYTICAL CONTINUUM ELECTROSTATICS (ACE)

ACE⁶⁸ approach to the description of electrostatic component of Solvation free energy ΔG_{elec} is defining Electrostatic free energy in a particular medium, G_{elec} , is in terms of energy density $u(\vec{r})$. The energy density $u(\vec{r})$ of an electrostatic field generated by a charge distribution $\rho(\vec{r})$ can be expressed in terms of the electric displacement vector $D(\vec{r})$, where $D(\vec{r})$ is defined as

$$D(\vec{r}) = \epsilon(\vec{r})\vec{E}(\vec{r}) \quad (2.27)$$

Writing $u(\vec{r})$ in terms of the electric displacement vector $D(\vec{r})$, we have,

$$u(\vec{r}) = \frac{1}{8\pi\epsilon(\vec{r})}\vec{D}^2(\vec{r}) \quad (2.28)$$

By integrating the energy density over full space,^{69,70} the electrostatic energy G_{elec} can be written as

$$G_{elec} = \int_V u(\vec{r})d\vec{r} = \frac{1}{8\pi} \int_V \frac{1}{\epsilon(\vec{r})}\vec{D}^2(\vec{r})d\vec{r} \quad (2.29)$$

Now we can split this integral expression into two parts, one over the volume of the solute (for eg. protein) V_P with a low dielectric ϵ_p and the other is the integral over the remaining volume of the solvent V_S with a high dielectric ϵ_s .

$$G_{elec} = \frac{1}{8\pi\epsilon_s} \int_{V_s} \vec{D}^2(\vec{r})d\vec{r} + \frac{1}{8\pi\epsilon_p} \int_{V_p} \vec{D}^2(\vec{r})d\vec{r} \quad (2.30)$$

We can introduce, what is called the ‘‘Reduced Dielectric’’, $\Delta\epsilon$ and define it as,

$$\frac{1}{\Delta\epsilon} = \frac{1}{\epsilon_p} - \frac{1}{\epsilon_s} \quad (2.31)$$

Since $\epsilon_p < \epsilon_s$, the reduced dielectric constant $\Delta\epsilon$ is always positive. Using

this “Reduced Dielectric constant” in Equation 2.30, we can rewrite it extending the first integral over the full space V by adding and subtracting $\frac{1}{8\pi\epsilon_s} \int_{V_p} \vec{D}^2(\vec{r}) d\vec{r}$ to Equation 2.30, we have,

$$G_{elec} = \frac{1}{8\pi\epsilon_s} \int_{V_s} \vec{D}^2(\vec{r}) d\vec{r} + \frac{1}{8\pi\epsilon_p} \int_{V_p} \vec{D}^2(\vec{r}) d\vec{r} + \frac{1}{8\pi\epsilon_s} \int_{V_p} \vec{D}^2(\vec{r}) d\vec{r} - \frac{1}{8\pi\epsilon_s} \int_{V_p} \vec{D}^2(\vec{r}) d\vec{r} \quad (2.32)$$

We could combine the first and the third term and write the integral over the whole volume, V . The second and the fourth term can be combined to give the expression with the “Reduced Dielectric Constant”, $\Delta\epsilon$ over the solute volume V_p . Then we have,

$$G_{elec} = \frac{1}{8\pi\epsilon_s} \int_V \vec{D}^2(\vec{r}) d\vec{r} + \frac{1}{8\pi\Delta\epsilon} \int_{V_p} \vec{D}^2(\vec{r}) d\vec{r} \quad (2.33)$$

This expression is still exact. No approximations are done so far. An approximation can be introduced by assuming that the first term corresponds to the situation in a homogeneous dielectric medium and describe the dielectric displacement $\vec{D}(\vec{r})$ by simple Coulomb field. This is what is called “**Coulomb Field Approximation**”. This assumption is not exact since the dielectric displacement \vec{D} in both the integrals of equation 2.33 should satisfy the boundary conditions on the electric field at the interface between the solute and the solvent, i.e., the tangential component of the electric field $\vec{E} = \vec{D}/\epsilon$ and the normal component of the dielectric displacement \vec{D} do not change when passing through the solute-solvent boundary.

This assumption is shown to introduce a small percentage of error of at most a few percent in the Electrostatic free energy.⁷¹ Qualitatively, we can justify this assumption. If the solute is small, all the charges are highly exposed to the solvent. So there is a very small change from the Coulomb field. If, on the other hand, the solute is large, say a protein, most of the contribution to the Electrostatic free energy comes from the second integral and the error introduced is small in the calculation of Electrostatic free energy G_{elec} . Such a kind of approximation is not possible for the second integral in Equation 2.33, which is more difficult to evaluate.

One can now write Equation 2.33 using the Coulomb Field Approximation. We could now split the Electrostatic free energy G_{elec} into Self Energy term, G_i^{Self} , the interaction of atoms with themselves and Interaction Energy term, G_{ij}^{Int} , the interaction of atom i with atom j

$$G_{elec} = G_E = \sum_i \left(G_i^{self} + \sum_{j>i} G_{ij}^{Int} \right) \quad (2.34)$$

Where both the terms run up to the total number of atoms, N . Now rewriting Equation 2.33 using the coulomb field approximation for the first integral,

$$G_i^{self} = \frac{q_i^2}{2\epsilon_s R_i} + \frac{1}{8\pi\Delta\epsilon} \int_{V_p} \vec{D}_i^2(\vec{r}) d\vec{r} \quad (2.35)$$

$$G_{ij}^{int} = \frac{q_i q_j}{\epsilon_s |\vec{r}_{ij}|} + \frac{1}{8\pi\Delta\epsilon} \int_{V_p} \vec{D}_i(\vec{r}) \vec{D}_j(\vec{r}) d\vec{r} \quad (2.36)$$

The self energy G_i^{self} for a point charge yields a diverging energy contribution. So the atom is no longer considered a point charge, but a sphere with radius R_i . This point charge is evenly distributed over the surface of this sphere and the Electrostatic free energy is given by the Born energy term as in equation 2.17.

2.2.7 SOLVATION FREE ENERGY

We know that Solvation free energy, ΔG_{elec} is formally written as the difference between the electrostatic free energy in the heterogeneous dielectric medium G_{elec} (both the solute and the solvent), according to equation 2.34 and the electrostatic free energy in homogeneous dielectric medium, G_{hom} . Formally one can write it as:

$$\Delta G_{elec} = G_{elec} - G_{hom} \quad (2.37)$$

The partitioning of the electrostatic free energy into vacuum and solvated case is simple if one assumes that the solute interior has a dielectric constant of 1 such that the solute in vacuum constitutes a system with homogeneous dielectric where the Born's Self

energy and Coulomb's law are valid. So we can write

$$G_{hom} = \sum_i \left(\frac{q_i^2}{2\epsilon_p R_i} + \sum_{j>i} \frac{q_i q_j}{\epsilon_p |\vec{r}_{ij}|} \right) \quad (2.38)$$

Again we can split the electrostatic free energy into a self energy term and a interaction energy term:

$$\Delta G_{elec} = \Delta G^{self} + \Delta G^{int} = \sum_i \left(\Delta G_i^{self} + \sum_{j>i} \Delta G_{ij}^{int} \right) \quad (2.39)$$

Using Equations 2.35, 2.36 & 2.38 in the equation 2.37 we have,

$$\Delta G_i^{self} = \frac{q_i^2}{2\epsilon_s R_i} + \frac{1}{8\pi\Delta\epsilon} \int_{V_p} \vec{D}_i^2(\vec{r}) d\vec{r} - \frac{q_i^2}{2\epsilon_p R_i} \quad (2.40)$$

and

$$\Delta G_{ij}^{int} = \frac{q_i q_j}{\epsilon_s |\vec{r}_{ij}|} + \frac{1}{8\pi\Delta\epsilon} \int_{V_p} \vec{D}_i(\vec{r}) \vec{D}_j(\vec{r}) d\vec{r} - \frac{q_i q_j}{\epsilon_p |\vec{r}_{ij}|} \quad (2.41)$$

Using the expressing for reduced dielectric $\Delta\epsilon$ as in Equation 2.31, we can write the above equation as

$$\Delta G_i^{self} = -\frac{q_i^2}{2\Delta\epsilon R_i} + \frac{1}{8\pi\Delta\epsilon} \int_{V_p} \vec{D}_i^2(\vec{r}) d\vec{r} \quad (2.42)$$

$$\Delta G_{ij}^{int} = \frac{q_i q_j}{\Delta\epsilon |\vec{r}_{ij}|} + \frac{1}{8\pi\Delta\epsilon} \int_{V_p} \vec{D}_i(\vec{r}) \vec{D}_j(\vec{r}) d\vec{r} \quad (2.43)$$

We will use the Generalized Born equation for calculating the interaction energies of the charges (equation 2.43). We now need a integral approach to calculate the Self energies in Equation 2.42.

2.2.8 SUB-DIVISION OF SOLUTE VOLUME

To evaluate the integral in Equation 2.42, we introduce a molecular density function $P_S(\vec{r})$ describing the solute volume according to

$$P_S(\vec{r}) = \begin{cases} 1 & \text{if, } \vec{r} \text{ is inside the solute volume} \\ 0 & \text{Otherwise} \end{cases} \quad (2.44)$$

and rewrite Equation 2.42, with this Molecular density function $P_S(\vec{r})$.

$$G_i^{self} = -\frac{q_i^2}{2\epsilon_s R_i} + \frac{1}{8\pi\Delta\epsilon} \int_{V_p} \vec{D}_i^2(\vec{r}) P_S(\vec{r}) d\vec{r} \quad (2.45)$$

Now we introduce a atomic density function $P_k(\vec{r})$ by splitting the Molecular density function $P_S(\vec{r})$ into a sum of atomic density functions. For each k , there is one density function $P_k(\vec{r})$ describing the volume distribution of the atom k . We write,

$$P_S(\vec{r}) = \sum_k P_k(\vec{r}) \quad (2.46)$$

The contribution of each atom k to the self energy of atom i , G_{ik}^{self} is given by:

$$G_{ik}^{self} = \frac{1}{8\pi\Delta\epsilon} \int_{V_p} \vec{D}_i^2(\vec{r}) P_k(\vec{r}) d\vec{r} \quad (2.47)$$

In Equation 2.47, the self energy of atom i is given by its Born self energy in the solvent dielectric ϵ_s plus a sum of $N - 1$ integral terms G_{ik}^{self} involving all other solute atoms $k \neq i$. This integral is inversely proportional to the factor $\Delta\epsilon$, which is characteristic of a transfer from the dielectric ϵ_s to ϵ_i . In other words, G_{ik}^{self} is the energy required to replace the solvent dielectric by the solute dielectric within the volume of atom k , provided that there is only the charge of atom i available to generate the electric field.

In the normal case of a solvent with a high dielectric constant than the solute, both $\frac{1}{\Delta\epsilon}$ and G_{ik}^{self} are positive. Since the dielectric displacement $\vec{D}(\vec{r}) = \epsilon\vec{E}$ decreases with increasing distance from the charge i , G_{ik}^{self} represents an effective, repulsive interaction between the charge i and atom k of the solute because the atom k prevents the charge i from an energetically favorable interaction with the solvent dielectric within the volume of atom k .

There are two limiting cases to the self energy, G_i^{self} : the first corresponds to

a single ion in solution and the second to a charged atom embedded in an infinite solute with uniform dielectric ϵ_p . For an ion in solution, only the first term remains in equation 2.45. In the second case, if the charge on the atom i is uniformly distributed on the surface of the sphere with the Van der Waals radius R_i and the sum of the integration volumes $\sum_{k \neq i} P_k$ represents all the space except the Van der Waals sphere of atom i , the integral $\frac{1}{8\pi\Delta\epsilon} \int_{V_p} \vec{D}_i^2(\vec{r}) P_k(\vec{r}) d\vec{r} = \sum_{k \neq i} G_{ik}^{self}$ in equation 2.45 & 2.47, must yield the Born transfer energy, $q_i^2/(2\Delta\epsilon R_i)$. When this is added to the first term, $q_i^2/(2\epsilon_s R_i)$ results in the correct Born self energy term $q_i^2/(2\epsilon_p R_i)$ in the solute dielectric..

In practice, its not easy to find a suitable way to find the atomic density functions $P_k(\vec{r})$. By representing the atomic volumes by sphere, leads to overlap of bonded pairs and there are cavities between neighbored non-bonded atoms. Instead of having a very discontinuous solute dielectric medium, both the charge distribution of atom i , $\rho_i(\vec{r})$ and the atomic density function of atom k , $P_k(\vec{r})$, are represented by 3-Dimensional Gaussian.

$$\rho_i(\vec{r}) = q_i \pi^{-3/2} \hat{R}_i^{-3} \exp\left(-\frac{(\vec{r} - \vec{r}_i)^2}{\hat{R}_i^2}\right) \quad : \quad \hat{R}_i = \frac{R_i}{\sqrt{\pi/2}} \quad (2.48)$$

$$P_k(\vec{r}) = \frac{4}{3\sqrt{\pi}\alpha^3} \exp - \frac{(\vec{r} - \vec{r}_i)^2}{(\alpha\tilde{R}_k)^2} \quad (2.49)$$

Where R_i is the Van der Waals radius of atom i , the width parameter \hat{R}_i of the charge distribution is chosen such that the Born self energy of the Gaussian-distributed charge is the same as if the charge was evenly distributed on the sphere of radius R_i . The width parameter \tilde{R}_k of the volume distribution is the effective atom radius derived from the average solvent-inaccessible volume contribution \tilde{V}_k of different atom types in the solute.

The Gaussian are normalized such that

$$q_i = \int_V \rho_i(\vec{r}) d\vec{r} \quad (2.50)$$

and

$$\tilde{V}_k = \frac{4\pi \tilde{R}_k^3}{3} = \int_V P_k(\vec{r}) d\vec{r} \quad (2.51)$$

The volume parameter \tilde{V}_k does not depend on the smoothing factor α , which is introduced to control the width of the individual atomic functions $P_k(\vec{r})$.

2.2.9 COULOMB FIELD APPROXIMATION TO THE THE SELF ENERGY

To evaluate the volume integral in Equation 2.47, the dielectric displacement $\vec{D}(\vec{r})$ at a given point \vec{r} due to a charge i at \vec{r}_i can be approximated by the Coulomb field $\vec{D}_i(\vec{r}) = q_i/|\vec{r} - \vec{r}_i|$ for a point charge. The advantage of this approximation is that we can integrate the energy density of the electric field without prior knowledge of the electrostatic potential and introduce the electrostatic potential by a iterative procedure.⁷²

Since the reaction field is defined as the non-Coulombic contribution in an inhomogeneous dielectric, the Coulomb field approximation assumes that the reaction field contribution to the energy density of the electric field can be neglected. In the usual case the polarizability of the solvent is higher than the solute polarizability ($\because \epsilon_s > \epsilon_p$), the reaction field leads to the focusing of the field lines towards the solvent and decrease in magnitude of the dielectric displacement in the solute interior. Therefore, omission of the reaction field when integrating \vec{D}_i over the volume of each atom k in the interior of a large solute can be expected to an over-estimation of the energy G_{ik}^{self} .

In applications to biological molecules, in particular small solutes and solutes with high degree of solvent exposure, the over estimation error is expected to be small.

2.2.10 SELF ENERGY APPROXIMATION

In a homogeneous dielectric, ϵ , the potential ϕ_i of the Gaussian charge distribution ρ_i , using Equation 2.48, is

$$\phi_i(\vec{r}) = q_i \frac{\text{erf}(|\vec{r} - \vec{r}_i|/\tilde{R}_i)}{\epsilon|\vec{r} - \vec{r}_i|} \quad (2.52)$$

where "erf" denotes the error function. Based on the Coulomb field approxi-

mation, the dielectric displacement of the Gaussian charge distribution is

$$\vec{D}_i(\vec{r}) = -q_i \vec{\nabla} \frac{\text{erf}(|\vec{r} - \vec{r}_i|)/\tilde{R}_i}{|\vec{r} - \vec{r}_i|} \quad (2.53)$$

Using Equation 2.53 and Equation 2.48 into Equation 2.47, we have,

$$G_{ik}^{self} = \frac{q_i^2}{6\pi^{3/2}\Delta\epsilon\alpha^3} \int_V \left(\vec{\nabla} \frac{\text{erf}(|\vec{r} - \vec{r}_i|)/\tilde{R}_i}{|\vec{r} - \vec{r}_i|} \right)^2 \exp\left(-\frac{(\vec{r} - \vec{r}_k)^2}{(\alpha\tilde{R}_k)^2}\right) d\vec{r} \quad (2.54)$$

Again, this integral is not analytically solvable. However, a suitable approximation can be used if G_{ik}^{self} is finite and monotonically decreasing with the distance $|\vec{r}_{ik}|$. This assumption fails in the limit $R_i \rightarrow 0$ and if the ratio $R_i/(\alpha\tilde{R}_k)$ is larger than a critical value that is close to unity. By numerically integrating the above equation, it turns out that at short range ($|\vec{r}_{ik}| < \alpha\tilde{R}_k$), G_{ik}^{self} decreases like a Gaussian while it approaches $q_i^2 \vec{\nabla}_k / (8\pi\Delta\epsilon|\vec{r}_{ik}|^4)$ at long range. This behavior leads to the following *Ansatz*.

$$G_{ik}^{self} = \frac{q_i}{\Delta\epsilon\omega_{ik}} \exp\left(-\frac{|\vec{r}_{ik}|^2}{\sigma_{ik}^2}\right) + \frac{q_i^2 \vec{\nabla}_k}{8\pi\Delta\epsilon} \left(\frac{|\vec{r}_{ik}|^3}{|\vec{r}_{ik}|^4 + \mu_{ik}^4} \right)^4 \quad (2.55)$$

The parameters ω_{ik} and σ_{ik} determine the height and width of the Gaussian that approximates G_{ik}^{self} in the short-range domain. The first term of Equation 2.55 becomes negligible for large $|\vec{r}_{ik}|$ and the second term vanishes at $|\vec{r}_{ik}| = 0$ due to the parameter μ_{ik} . The parameters in Equation 2.55 are determined by calculating analytically the exact value for G_{ik}^{self} and the second derivative $\partial^2 G_{ik}^{self} / \partial |\vec{r}_{ik}|^2$ at $|\vec{r}_{ik}| = 0$ using Equation 2.54. This is possible since $|\vec{r}_{ik}| = 0$ means that $\vec{r}_i = \vec{r}_k$. At $|\vec{r}_{ik}| = 0$, the second term in Equation 2.55 as well as its second derivative vanish, so that the parameters ω_{ik} and σ_{ik} can be determined using the following relations:

$$\omega_{ik} = \frac{3\pi\alpha_{ik}^4 \tilde{R}_k}{4(Q_{ik} - \arctan Q_{ik})} \quad (2.56)$$

$$\sigma_{ik} = \sqrt{\frac{3\alpha_{ik}^2 \tilde{R}_k^2 (Q_{ik} - \arctan Q_{ik})}{(3 + f_{ik})Q_{ik} - 4\arctan Q_{ik} - ik}} \quad (2.57)$$

with

$$Q_{ik} = \frac{q_{ik}^2}{\sqrt{2q_{ik}^2 + 1}} : f_{ik} = \frac{2}{q_{ik}^2 + 1} - \frac{1}{2q_{ik}^2 + 1} : q_{ik}^2 = \frac{\pi}{2} \left(\frac{\alpha_{ik} \tilde{R}_k}{R_i} \right)^2 : \alpha_{ik} = \max \left(\alpha, \frac{R_i}{\tilde{R}_k} \right) \quad (2.58)$$

The long range parameter μ_{ik} is determined by making sure that the value of the Born energy term of a charge q_i in a homogeneous dielectric medium with the dielectric constant ϵ_p is maintained. this yields

$$\mu_{ik} = \frac{77\pi\sqrt{2}R_I}{512 \left(1 - \frac{2\pi^{3/2}\sigma_{ik}^3 R_i}{\omega_{ik} V_k} \right)} \quad (2.59)$$

2.2.11 INTERACTION ENERGY APPROXIMATION:

For calculating the interaction energy between two solute charge i and j , the Generalized Born equation is used. This method takes into account of the effect of both charge-charge distance, r_{ij} and the degree of solvent exposure of the interacting charges. One could define a parameter b_i which reflects the degree of charge burial of solute charge i . This value is chosen such that the atomic Solvation energy ΔG_i^{self} of the charge is equal to the Born Solvation energy of a sphere with identical charge, q_i , and radius b_i called the ‘‘Born Radii’’. Since ΔG_I^{self} is already calculated, the effective born radii b_i can be written as:

We know that

$$\Delta G_i^{self} = -\frac{q_i^2}{2\Delta\epsilon b_i} \quad (2.60)$$

rearranging, we have

$$b_i = -\frac{q_i^2}{2\Delta\epsilon\Delta G_I^{self}} \quad (2.61)$$

In the same way as above, we could write ΔG_{ij}^{self} can be written in terms of an effective interaction distance R_{ij}^{solv} taking into account the solute volume as:

$$\Delta G_{ij}^{int} = -\frac{q_i q_j}{\Delta \epsilon R_{ij}^{int}} \quad (2.62)$$

putting together Equations 2.60 & 2.61 we get to our familiar Generalized Born Equation along with the self energy of the solute charge i :

$$\Delta G_{elec} = -\sum_i \left(\frac{q_i^2}{2\Delta \epsilon b_i} + \sum_{j>i} \frac{q_i q_j}{\Delta \epsilon R_{ij}^{solv}} \right) \quad (2.63)$$

As before, the effective interaction distance R_{ij}^{solv} is given by the function (refer to the Generalized Born Methods) $R_{ij}^{solv} = \sqrt{|\vec{r}_{ij}|^2 + b_i b_j \exp(-|\vec{r}_{ij}|^2/4b_i b_j)}$.

2.2.12 OPTIMIZATION OF ACE

The Born radius thus, defines the extent of solvation of the charge while calculating the electrostatic solvation energy for a collection of charges. The smaller the Born radius, the higher the solvation.^{57,58,68} Previous studies have indicated that the accuracy of Born radius is the limiting factor for accurate calculation of electrostatic solvation energies.⁷³ Default ACE parameters which were previously obtained using small peptides give errors in the born-raddi evaluation. So the parameters were optimized here by benchmarking the interaction energies for each pair of charges in the ACE model with the Poisson-Boltzmann (PBEQ) model.⁶¹ The benchmarking was done by comparing the Interaction energy of PBEQ with that of ACE using the following protocol.

To get the Interaction energy in PBEQ,

1. First an atom, i , with a charge q_i is chosen.
2. All the other charges in the protein are assigned to zero and the potential ϕ_i is calculated at every point in space.

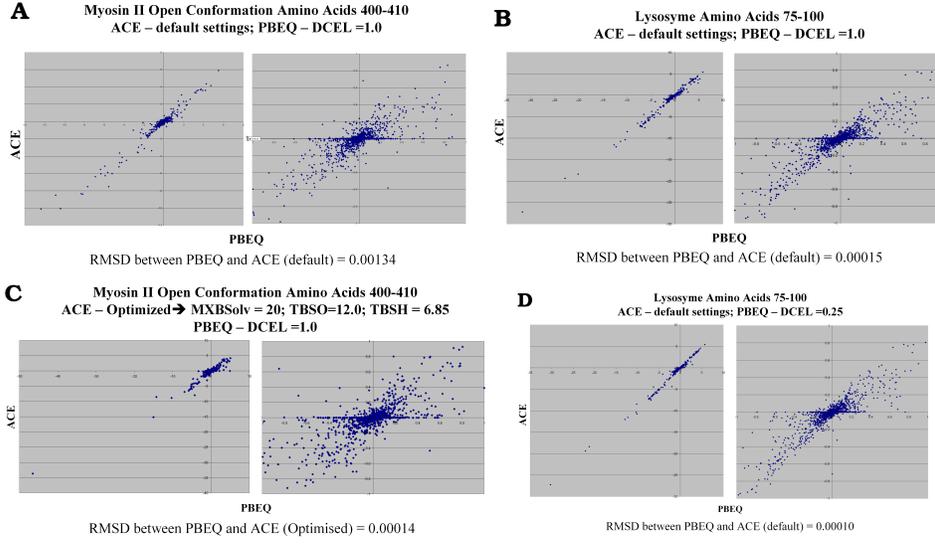


Figure 2.6: Optimization of ACE parameters for use with minimum energy pathway calculations using CPR³

3. Then another charge on an atom, j , is put on and the interaction energy (IE_{PBEQ}) between atom, j and atom, i is calculated using the relation:

$$IE_{PBEQ} = \phi_i(r_j) \cdot q_j \quad (2.64)$$

To get the interaction energy in ACE,

1. First the self-energy of atom, i with a charge q_i is chosen and its self-energy is computed by putting “off” all other charges in the system (IE_{q_i}).
2. Then the self-energy of atom, j with a charge q_j is chosen and its self-energy is computed by putting “off” all other charges in the system (IE_{q_j}).
3. Next, the self-energy is computed with both charges, i and j “on” while rest of the charges are put “off” (IE_{q_i, q_j}).

Then interaction energy is computed using the relation:

$$IE_{ACE} = IE_{q_i, q_j} - (IE_{q_i} + IE_{q_j}) \quad (2.65)$$

Now the difference between IE_{PBEQ} and IE_{ACE} was minimized while changing the variable `MXBSOLV`, for defining the maximum born radii allowed during the electrostatic energy evaluation using MATLAB⁷⁴ (This constraint was included to prevent singularities during the evaluation of self energy for each charge and was optimized for small peptides⁶⁸). We obtained a higher `MXBSOLV` value of 20 Å (default = 14 Å) (see Figure 2.6).

2.3 CLASSICAL MECHANICS

The molecular dynamics simulation method is based on Newton's second law or the equation of motion, $F = m \cdot a$, where F is the force exerted on the particle, m is its mass and a is its acceleration. From a knowledge of the force on each atom, it is possible to determine the acceleration of each atom in the system. Integration of the equations of motion then yields a trajectory that describes the positions, velocities and accelerations of the particles as they vary with time. From this trajectory, the average values of properties can be determined. The method is deterministic: which implies that once the positions and velocities of each atom are known, the state of the system can be predicted at any time in the future or the past.

Newton's equation of motion is given by

$$F_i = m_i \cdot a_i \quad (2.66)$$

where F_i is the force exerted on each particle i , m_i is the mass of particle i and a_i is the acceleration of particle i .

The force can also be expressed as the gradient of the potential energy V as

$$F_i = -\nabla_i V. \quad (2.67)$$

Combining equations 2.66 & 2.67, yields

$$-\frac{dV}{dr_i} = m_i \frac{d^2 r_i}{dt^2} \quad (2.68)$$

where V is the potential energy of the system. Newton's equation of motion can then relate the derivative of the potential energy to the changes in position as a function of time.

2.3.1 INTEGRATION ALGORITHMS

The potential energy is a function of the atomic positions ($3N$) of all the atoms in the system. All the integration algorithms assume the positions, velocities and accelerations can be approximated by a Taylor series expansion:

$$r(t + \delta t) = r(t) + v(t)\delta t + \frac{1}{2}a(t)\delta t^2 + \dots \quad (2.69)$$

Due to the complicated nature of this function, there is no analytical solution to the equations of motion and they must be solved numerically. Numerous numerical algorithms have been developed for integrating the equations of motion. Some of the most important are listed here.

- i Verlet algorithm
- ii Leap-frog algorithm
- iii Velocity Verlet
- iv Beeman's algorithm

In choosing which algorithm to use, the following criteria were considered

1. The algorithm should conserve energy and momentum.
2. It should be computationally efficient
3. It should permit a long time step for integration.
4. Must be stable and allow computations upto nano-second time-scales

2.4 STATISTICAL MECHANICS

In a molecular dynamics simulation, one often wishes to explore the macroscopic properties of a system through microscopic simulations, for example, to calculate changes in the binding free energy of a particular drug candidate, or to examine the energetics and mechanisms of conformational change. The connection between microscopic simulations and macroscopic properties is made via statistical mechanics which provides the rigorous mathematical expressions that relate macroscopic properties to the distribution and motion of the atoms and molecules of the N-body system; molecular dynamics simulations provide the means to solve the equation of motion of the particles and evaluate these mathematical formulas. With molecular dynamics simulations, one can study both thermodynamic properties and/or time dependent (kinetic) phenomenon. The goal is to understand and to predict macroscopic phenomena from the properties of individual molecules making up the system. The system could range from a collection of solvent molecules to a solvated protein-DNA complex.

In statistical mechanics, average values are defined as ensemble averages. The ensemble average is given by

$$\langle A \rangle_{ensemble} = \int \int dp^N dr^N A(p^N, r^N) \rho(p^N, r^N) \quad (2.70)$$

where, $A(p^N, r^N)$ is the observable of interest and it is expressed as a function of the momenta, p , and the positions, r , of the system. The integration is over all possible variables of r and p . while, $\rho(p^N, r^N)$ is the probability density of the ensemble, given by

$$\rho(p^N, r^N) = \frac{1}{Q} \exp \left[\frac{-H(p^N, r^N)}{k_B \cdot T} \right] \quad (2.71)$$

where H is the Hamiltonian, T is the temperature, k_B is Boltzmann's constant and Q is the partition function

$$Q = \int \int dp^N dr^N \exp \left[\frac{-H(p^N, r^N)}{k_B \cdot T} \right] \quad (2.72)$$

This integral is generally extremely difficult to calculate because one must calculate all possible states of the system. In a molecular dynamics simulation, the points in the ensemble are calculated sequentially in time. So, to calculate an ensemble average, the molecular dynamics simulations must pass through all possible states corresponding to the particular thermodynamic constraints.

Another way, as done in an MD simulation, is to determine a time average of A , which is expressed as

$$\langle A \rangle_{time} = \lim_{\tau \rightarrow \infty} \int_{t=0}^{\tau} A(p^N(t), r^N(t)) dt \approx \frac{1}{M} \sum_{t=1}^M A(p^N, r^N) \quad (2.73)$$

where t is the simulation time, M is the number of time steps in the simulation and $A(p^N, r^N)$ is the instantaneous value of A .

The problem now is to calculate time averages by molecular dynamics simulation while the experimental observables are assumed to be ensemble averages. Resolving this leads us to one of the most fundamental axioms of statistical mechanics, **The Ergodic Hypothesis**, which states that the time average equals the ensemble average i.e.,

$$\langle A \rangle_{ensemble} = \langle A \rangle_{time}$$

or

$$Ensemble\ average = Time\ average$$

The basic idea is, if one allows the system to evolve in time indefinitely, then that system will eventually pass through all possible states. One goal, therefore, of a molecular dynamics simulation is to generate enough representative conformations such that this equality is satisfied. If this is the case, experimentally relevant information concerning structural, dynamic and thermodynamic properties may then be calculated using a feasible amount of computer resources. Because the simulations are of fixed duration, one must be certain to sample a sufficient amount of phase space.

Some examples of time averages:

Average potential energy

$$V = \langle V \rangle = \frac{1}{M} \sum_{i=1}^M V_i \quad (2.74)$$

Average kinetic energy

$$K = \langle K \rangle = \frac{1}{M} \sum_{k=1}^M \left\{ \sum_{i=1}^N \frac{m_i \cdot v_i^2}{2} \right\}_k \quad (2.75)$$

where M is the number of configurations in the simulation, N is the number of atoms in the system, m_i is the mass of the particle i and v_i is the velocity of particle i .

2.4.1 SIMULATING IN DIFFERENT ENSEMBLES

MD simulations generate information on the microscopic level. The microscopic state of a system is defined in a $6N$ -dimensional space, consisting of the momenta, \mathbf{p} , and spatial coordinates \mathbf{r} of the N particles. The thermodynamic state of a system is usually defined by a small set of parameters, for example the temperature T , the pressure, P , and the number of particles, N . Other thermodynamic properties can be derived from the equations of state and other fundamental thermodynamic equations.⁷⁵

Statistical Mechanics relates the microscopic information to macroscopic properties (*i.e.* pressure, internal energy, etc.). This is achieved through the use of statistical ensembles. An ensemble is a collection of points in phase space satisfying the conditions of a particular thermodynamic state. Ensembles describe a collection of all possible systems that can have different microscopic states but are identical macroscopically (or thermodynamically) state. A summary of different statistical ensembles with different characteristics are given below.⁷⁶

2.4.2 MICRO-CANONICAL ENSEMBLE

In the micro-canonical ensemble the three constant parameters are N, V, E , *i.e.* number of particles, volume and energy of the system. This is the *natural* ensemble for MD

simulations, where the Newtonian equations of motion can be applied unchanged to the system. To convert to other statistical ensembles one has to integrate other equations in place of Newton's equations in such a way that sampling is performed in another statistical ensemble.

2.4.3 CANONICAL ENSEMBLE

In the canonical ensemble the fixed parameters are N, V, T , *i.e.* number of particles, volume and temperature of the system. Since the temperature has to be kept constant in this ensemble, a thermostat has to be introduced and the Hamiltonian of the system becomes:

$$H = K + V + K_s + V_s \quad (2.76)$$

where K is the kinetic energy of the system, V is the potential energy of the system and K_s and V_s are the kinetic and potential energies coupled to the thermostat.

2.4.4 NOSÉ-HOOVER CONSTANT TEMPERATURE ALGORITHM

The equations of motion for constant temperature or Nosé-Hoover thermostat equations are the following:⁷⁷

$$\begin{aligned} \dot{\mathbf{r}}_i &= \frac{\mathbf{p}_i}{m_i} \\ \dot{\mathbf{p}}_i &= \mathbf{F}_i - \frac{p_\eta}{Q} \mathbf{p}_i \\ \dot{\eta} &= \frac{p_\eta}{Q} \\ \dot{p}_\eta &= \sum_i \frac{\mathbf{p}_i^2}{m_i} - dNkT \end{aligned} \quad (2.77)$$

where η and p_η are the thermostat position and momentum and T is the temperature at which the system is to be regulated. N being the number of atoms in the system, k Boltzmann's constant and d the number of spatial dimensions. The parameter Q , given by $Q = dNkT\tau^2$, determines the time scale of the thermostat motion via the time scale parameter τ , which should be chosen in correspondence to a characteristic time scale of

the system, *e.g.*, a vibrational period.

The Nosé-Hoover scheme has the advantageous feature that it approximates the canonical distribution of temperature present in physical temperatures, *i.e.* the temperature of the system is not fixed at a given temperature but oscillates about it, as expected for small systems.

2.4.5 ISOBARIC-ISOTHERMAL ENSEMBLE

In this ensemble, pressure and temperature are kept constant in the system, *i.e.* the constant parameters are N, P, T . The Newtonian equations of motion become even more complicated in this ensemble because of the introduction of an additional barostat to the system.

2.4.6 CONSTANT TEMPERATURE AND PRESSURE

Simultaneous regulation of temperature and pressure can also be taken care of through the equations of motions: the isothermal-isobaric equations of motion:⁷⁸

$$\begin{aligned}
 \dot{\mathbf{r}}_i &= \frac{\mathbf{p}_i}{m_i} + \frac{p_\epsilon}{W} \mathbf{r}_i \\
 \dot{\mathbf{p}}_i &= \mathbf{F}_i - \left(1 + \frac{1}{N}\right) \frac{P_\epsilon}{W} \mathbf{p}_i - \frac{P_\eta}{Q} \mathbf{p}_i \\
 \dot{V} &= \frac{dV p_\epsilon}{W} \\
 \dot{p}_\epsilon &= dV (P_{\text{int}} - P_{\text{ext}}) + \frac{1}{N} \sum_i \frac{\mathbf{p}_i^2}{m_i} - \frac{p_\eta}{Q} p_\epsilon \\
 \dot{\eta} &= \frac{P_\eta}{Q} \\
 \dot{p}_\eta &= \sum_i \frac{\mathbf{p}_i^2}{m_i} + \frac{p_\epsilon^2}{W} - (dN + 1)kT
 \end{aligned} \tag{2.78}$$

where p_ϵ is a momentum conjugate to the logarithm of the volume, W is its associated mass parameter, $\epsilon = \ln(V/v_0)$, P_{ext} is the externally applied pressure, and P_{int} is the instantaneous internal pressure of the system given by:

$$P_{\text{int}} = \frac{1}{dV} \left[\sum_i \frac{\mathbf{p}_i^2}{m_i} + \sum_i \mathbf{r}_i \mathbf{F}_i - (dV) \frac{\partial U}{\partial V} \right] \quad (2.79)$$

Thus, the variable p_ϵ acts as a 'barostat', which drives the system to the steady state $\langle P_{\text{int}} \rangle = P_{\text{ext}}$. In this way, both temperature and pressure are regulated so as to reproduce exact canonical distributions.

2.5 DETERMINATION OF REACTION PATHS

A conformational change involves a change of the relative positions of the nuclei involved and the associated change in the structure of the system. To describe the motion of the nuclei it is necessary to know the potential energy surface (PES) on which the nuclei move. Mathematically, the PES is a function of the $3N - 6$ internal degrees of freedom of a non-linear molecule. If the multidimensional PES is completely and accurately known the conformational dynamics can be determined by solving the equations of motion of the nuclei moving along the PES. In such a scenario, the reactant conformation and product conformation are described by regions around minima on the PES that are thermally accessible at a given temperature. The transition state is defined as the dividing surface that separates the edduct and the product states which are defined by a first-order saddle points connecting the minima.

2.5.1 DETERMINATION OF A MEP IN SYSTEMS WITH MANY DEGREES OF FREEDOM

The optimization of a point into a minimum on a multidimensional surface is a mathematically well-described problem that can be solved using a number of algorithms (see Ref.⁷⁹). The optimization of a reaction path to a MEP, however, requires the simultaneous optimization of all path points subject to the constraint that the path remains continuous.

2.5.2 CONJUGATE PEAK REFINEMENT, CPR

CPR is a heuristic method for refining an initial path into an Minimum Energy Path (MEP).³ Its basic idea is to identify those points along the path where the energy is high-

est (the “peaks”) and to move these points closer to the MEP by a controlled conjugate-gradient minimization. The main advantage of CPR is that it automatically handles paths with many saddle points, each of which can be refined to any desired accuracy. The algorithm does not evaluate second derivatives but uses only the energy (which must be continuous) and its gradient. It starts from an initial continuous path, which is defined by linear interpolation along the series of points $P := [r_0, r_1, \dots, r_M]$ (Refer to the original CPR article³ for further details)

2.5.3 GENERATION OF AN INITIAL PATH

A continuous transition path is described by a series of discrete supporting points in conformational space $P = [r_0, r_1, \dots, r_{M-1}, r_M]$ that are connected by some defined interpolation method (e.g. linear or spline interpolation) in some defined coordinate system (e.g. Cartesian or internal). In the absence of a better guess for transition intermediates, a set of points P for the initial path can be generated by linear interpolation between the reactant and product end states. Interpolation in internal coordinates accurately describes the torsional transitions between rotameric states of the side chains. However, when applied to a whole protein, it can lead to severe disruption of the backbone fold. In contrast, Cartesian interpolation approximately preserves the backbone fold in most cases (if the protein remains compact during the transition) but often leads to extreme deformation of the side chains. Thus, it is sensible to combine the two interpolation methods. First, the backbone atoms are interpolated in Cartesian coordinates so as to preserve the backbone fold, and then the side chain atoms are built onto the interpolated backbone, using internal coordinate values that are interpolated between the internal coordinates of the end states.

This “combined interpolation” can still produce initial paths that lie in a reaction channel that includes unrealistic events like crossing of side chains and deformation of backbone structures. We found that this can be effectively avoided by generating the intermediate points of the initial path with all of the side chains shrunk. This is achieved by reducing all bond lengths of the side chains (here to half of the original size) before building them onto the backbone. At first glance, this approach would seem to be un-

physical and would result in even more energetically unfavorable paths. However, the minimization process applied to the path points during the CPR computation (see below) rapidly restores the shrunken side chains to their normal size, while undesirable events such as bond crossing or ring penetration no longer occur in the resulting MEP. This is because during path optimization the CPR procedure preserves the side-chain avoidance introduced in the initial path.

2.5.4 REACTION COORDINATE

Once an MEP, $P = [r_0, r_1, \dots, r_M]$ has been found, it defines the normalized curvilinear reaction coordinate, λ , which can be used to measure the progress along the reaction up to a given path point r_i :

$$\lambda(i) = \frac{\sum_{k=1}^i |r_k - r_{k-1}|}{\sum_{k=1}^M |r_k - r_{k-1}|} \quad (2.80)$$

$\lambda(i)$ is the normalized sum of the RMS-change in all coordinates along the path P up to r_i .

All paths were computed with the CPR³ implemented in the TReK module of CHARMM⁸⁰ using the default CPR settings for flagging a path point as a saddle point: With these settings, the gradient at a saddle point is required to be smaller than $g_{sad} = 0.05$ kcal.mol⁻¹.Å⁻¹ for an uninterrupted number $C = N^{0.5}$ of conjugate line minimizations (here: $C = 26$). These settings are stringent enough to obtain a path whose barriers are mostly within 1-2 kcal/mol of the first-order saddle points. For verification, the saddle points of the paths were further optimized with $g_{sad} = 0.01$ kcal.mol⁻¹.Å⁻¹ and $C = 1000$.

2.6 ANALYZING CONFORMATIONAL TRANSITIONS USING COMPUTER SIMULATIONS

In minimum energy pathway calculations (MEP) or molecular dynamics (MD) simulations, the behavior of the molecular system is obtained by using the potential energy function. The result of these simulation gives a series of conformations (either time de-

pendent (MD) or low-energy intermediates (MEP); which is called a trajectory or the path followed by each atom to go from one conformation to another.

MEP methods present a mechanism for the relevant conformational change by exploring the potential energy surface. This presents us a putative mechanism which would occur at 0° Kelvin which may not be realistic for systems in which there could be a huge entropic contribution to the conformational change (like in Myosin II, where the converter domain which amplifies motions near the ATP site to a large 60° rotation during which the lever-arm swings through the solvent, see the following chapters for a clear description). Nevertheless, MEP presents us a mechanism that could be plausible since only the relevant degrees of freedom are explored and it's much easier to conceptualize the mechanism. On the other hand, MD simulations performed under conditions of constant N, T and P (to better mimic experimental conditions) present a thermodynamically complete description of the system and give a true representation of the dynamics of the system in real-time. On the negative side, they suffer from a high computational cost and it's not possible to simulate a protein-system for more than a few nano-seconds (at the present computational power). So, here we combine MEP calculations along with the MD simulations to understand the conformational transitions during the recovery-stroke in Myosin II molecule.

In the following chapters, the usage of MEP method, CPR³ and MD simulations are described in detail along with the steps taken to setup and run a CPR calculation or MD simulations for a Myosin II system.

THE STRUCTURAL COUPLING BETWEEN ATPASE ACTIVATION AND RECOVERY-STROKE IN THE MYOSIN II MOTOR.

3.1 SUMMARY

The Myosin motor head undergoes a large conformational transition called the “Recovery Stroke” in which the converter domain (which bears the lever arm) rotates by $\sim 65^\circ$ before it can bind to the actin filament and perform the next power-stroke. Simultaneously, myosin must also activate its ATPase function, since hydrolysis of the bound ATP is required for strong actin binding. The activation is achieved by a closing of the residues 457-458 (belonging to the Switch-2 loop) over the ATP. The coupling between the motions of the converter domain and of the ~ 40 Å distant Switch-2 loop is essential for a productive motor cycle. The coupling mechanism is determined by computing a minimum energy pathway between the crystallographic end-states of the recovery stroke, yielding a continuous series of optimized intermediates in atomic detail. This reveals a two-phase mechanism, in which the successive formation of two hydrogen bonds by the Switch-2 loop is correlated with the successive movement of the two helices that hold the converter domain: the relay helix and the SH1-helix. The first hydrogen bond (between Gly457, on the N-terminal of the relay helix and the γ -phosphate of ATP) causes a see-saw motion of the relay helix. The second hydrogen bond (between Switch-2 and the Ser181 of the P-loop) causes the wedging of a loop against the SH1-helix end, resulting in a longitudinal translation of the SH1-helix relative to the relay helix. The converter domain responds to the “See-Saw” motion by rotating 25 degrees, then to the translation of the SH1-helix by rotating a further 40 degrees. The proposed coupling

mechanism is consistent with the existing mutational data and explains the role of the highly conserved wedge loop.

3.2 INTRODUCTION

Cells undergo a variety of motile processes that are driven by molecular motors that transduce the chemical energy of ATP hydrolysis into mechanical force and displacement. One of the most extensively studied molecular motors is myosin II, which drives muscle contraction (reviewed in refs^{2, 12, 13}) and other essential motile processes in eukaryotic cells by cyclically interacting with actin filaments.⁸¹ Myosin II is composed of a heavy chain and two light chains.²¹ The N-terminal globular domain of the heavy chain (the head) contains both the catalytic site and the actin-binding region.^{13, 33, 34} It has been shown that the N-terminal domain alone is able to hydrolyze ATP and move along an actin filament.⁸² Myosin and actin interact to produce movement as described in the Lymn-Taylor cycle (Figure 3.1).^{2, 19}

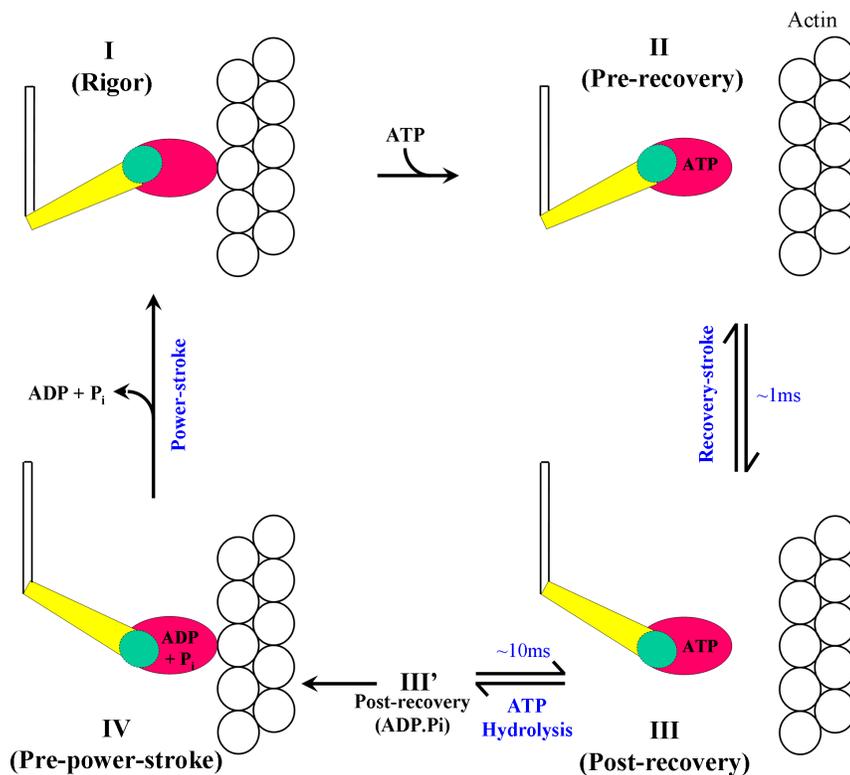


Figure 3.1: Lymn-Taylor Cycle.

This cycle is made possible by several coupling mechanisms that link structural changes in different parts of myosin. For instance, changes associated with ATP binding to the myosin-rigor conformation (State I) are coupled with structural changes in the actin binding region that substantially reduce binding affinity for actin,^{33,34} thus dissociating myosin from actin (Figure 3.1, State I → State II). In the next step, further changes in the ATP binding site that activate its ATPase function are coupled to a 65° rotation of the converter domain that carries the lever-arm (“The Recovery-Stroke” (State II ⇌ State III)).^{2,12,35–37} This primes myosin for the remaining steps: ATP hydrolysis (State III → State III′) increases myosin-actin binding affinity, causing myosin to rebind to actin (State III′ → State IV), this rebinding triggering the back-rotation of the converter domain that rows the myosin fibril past the actin filament. (“The Power stroke” (State IV → State I)).^{2,19} To understand the structural mechanisms of the motor function, it is necessary to understand the mechanisms by which the different domains of myosin are coupled. The goal of the present work is to understand the coupling mechanism involved during the recovery stroke at the level of atomic detail. This is done by computing a minimum energy pathway that links the crystallographically determined structures of States II & III. This computational approach, known as molecular kinematics, has already led to the understanding of other complex conformational transitions in proteins.^{36,83–86}

The head of the *Dictyostelium discoideum* Myosin II motor (henceforth, myosin) has been crystallized in the absence of actin with different ATP analogues bound to the ATPase site, (Mg·ATP· γ S,³⁹ Mg·AMP·PNP,³⁹ Mg·ADP·BeF_x,⁴ Mg·ADP·AlF₄⁻,⁴ Mg·ADP·V_i⁵ etc.). The converter-domain, which bears the lever arm is found in two orientations, in which the domain rotates by about 65° with respect to the rest of the head (Figures 3.2A & 3.2C, left panels). These two conformations have been assigned to State II and State III in the Lymn-Taylor cycle i.e., the end-states of the recovery stroke.^{2,4,5} The change in orientation of the converter domain (residues 692-748) is associated to a change in the conformation of the “Relay-helix” (residues 466-498) which undergoes a tilt in its helical axis and a “kink”, a quarter-turn unwinding (near residue 486), between the two ends of the recovery stroke (compare Figure 3.2 A & Figure 3.2C). The relay helix spans from the converter domain to the ATP binding site: at its C-terminus, it is tightly bound to the converter-domain, at its N-terminal end, the relay helix interacts with the switch-2 loop (residues 454-459), which changes during the recovery stroke from an “open” to a “closed”

conformation. Upon closing, the switch-2 forms two-key hydrogen bonds with ATP and the P-loop, a conserved loop in many nucleotide binding proteins. One hydrogen bond is between the Gly457 amide group and the γ -phosphate of ATP, while the other is between the Phe458 carbonyl group and the amide group of Ser181 on the P-loop. The P-loop together with the switch-1 and switch-2 loops form the so-called “Phosphate tube” (Table 3.1).

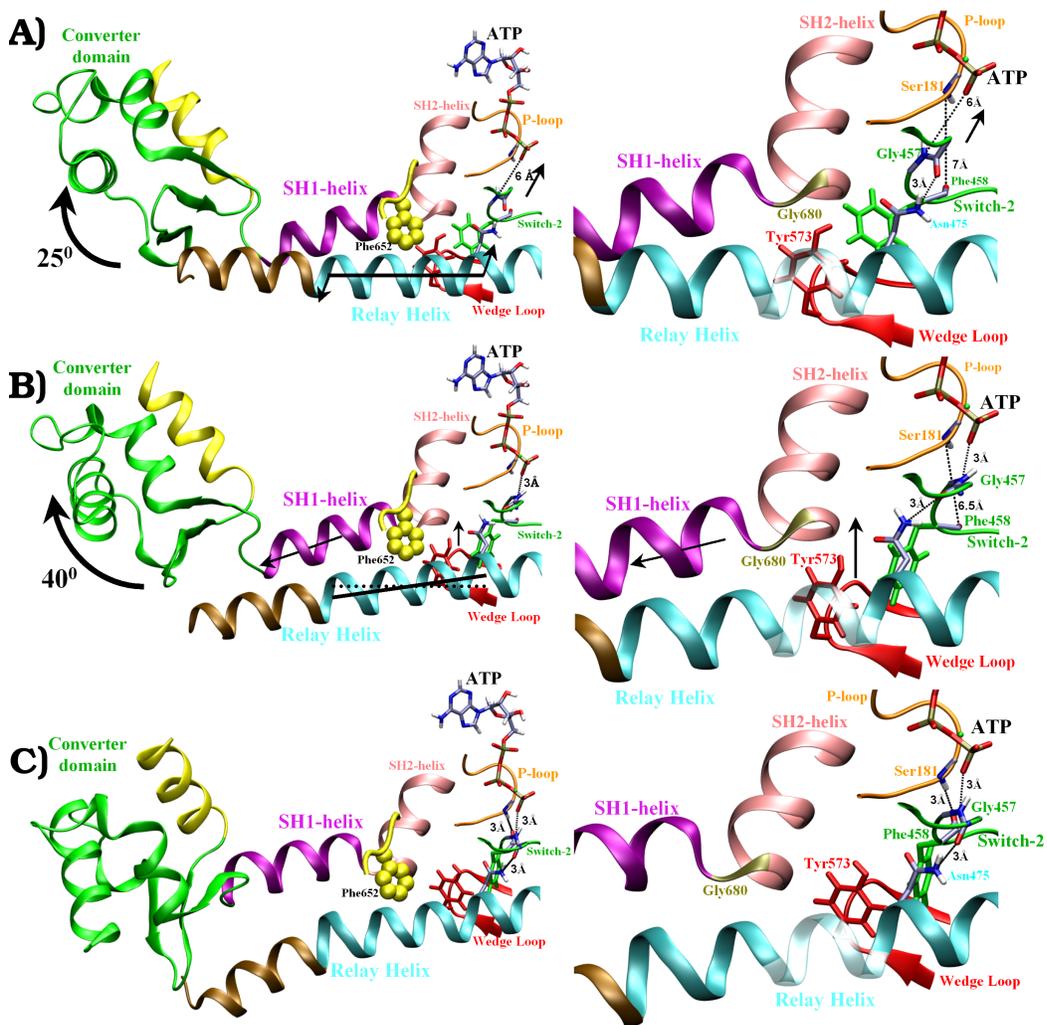


Figure 3.2: Structural elements of myosin involved in the recovery-stroke.

Legend: P-loop in Orange; Switch-2 in green; Relay helix in cyan; SH-1 helix in purple; SH-2 helix in pink; converter domain in green; lever arm in yellow. In the right panel, the important residues and bond distances involved in the recovery stroke are indicated. The See-saw and the Crank-shaft phases movements are shown with solid arrows.

The ATPase function in myosin is activated when the switch-2 loop closes and its Gly457/Ser456 peptide group, makes a hydrogen bond with the γ -phosphate.^{2,45,46} It would

Table 3.1: Consensus sequences.

Loop	Consensus ^a	<i>Dictyostelium discoideum</i>	residues ^b
P-loop	GESGAGKT	GESGAGKT	179-186
Switch-1	NxN SSR	NN SSR	233-238
Switch-2	DxSGFE	DISGFE	454-459

^a Conserved in at least 80 out of 82 myosins¹

^b Residue numbers of *Dictyostelium discoideum* myosin II are used here throughout.

be wasteful if the Switch-2 loop could freely close and re-open with the lever arm still in the pre-recovery orientation, allowing ATP hydrolysis and an unproductive release of the hydrolysis products. Thus, a mechanism is required to couple the closing of the Switch-2 loop with the orientation of the converter domain to ensure that ATP is hydrolyzed only when the converter domain/lever arm is in the post-recovery orientation (i.e., ready to perform the power stroke). A structural model was recently proposed for this coupling mechanism providing useful first insights on how small rearrangements at the ATP binding site can be gradually amplified through a network of coupling elements into a large-scale rotation of the converter domain.³⁶ It showed that the movement of Gly457 amide hydrogen on switch-2 towards the γ -phosphate is transmitted as a pull on the relay-helix via a hydrogen bond between the Gly457/Ser456 peptide group and the side chain of Asn475, which is located in the N-terminal end of the relay-helix (Figure 3.2A \rightarrow 3.2B, right panels). This pull results in a see-saw movement of the relay-helix (Figure 3.4A) which causes an initial rotation of the converter domain of about 25° (Figure 3.4A \rightarrow 3.4B). However, some important questions still remain concerning the later part of the mechanism. They are: (i) How is the further rotation by 40° of the converter domain coupled to the closing of the switch-2 loop? (ii) Is it possible to rationalize a simple model suggesting how the relay helix and the SH1-helix control the converter domain rotation? (iii) Why is the relay-helix “kinked” in the post-recovery conformation (Figure 3.2C)? Here, we address these questions so as to derive a comprehensive view of the coupling mechanisms involved in the recovery stroke.

The minimum-energy pathway (MEP) was obtained by “Conjugate Peak Refinement” (CPR),³ a computational methods that generates a continuous series of optimized intermediates connecting two given crystallographic end-state conformations of a protein without applying ex-

ternal driving constraints. Analysis of this pathway shows that the “See-saw” phase is followed by a second phase which is initiated by the movement of Switch-2 towards the P-loop to form a hydrogen bond between the carbonyl group of Phe458 and the amide group of Ser181 (Figure 3.2B → 3.2C, right panel). The phenyl ring of Phe458 remains in a tight hydrophobic packing with residues 572-574 on an adjacent loop (Figure 3.7), which thus follows the movement of switch-2 and wedges against the N-terminal end of the SH1-helix. The “Wedge loop” (Figure 3.2) is highly conserved structure in the myosin family.^{1,15,37,40} In response to the Wedge loop movement towards the ATP, the SH1-helix translates longitudinally in relation to the relay helix in a piston-like fashion (Figure 3.4B). This results in a “See-saw” like motion of the SH1-helix, due to the fact that the converter domain is suspended by the two helices: the SH1 helix and the relay helix and has minimal interaction with the rest of the myosin-head (Figure 3.6). This longitudinal motion of the SH1-helix pushes the converter domain outward which is covalently attached to the C-terminal ends of the SH1-helix (Figure 3.2C) and through salt-bridges to the relay helix (Figure 3.5). The local kink of the relay helix that appears during this second-phase (Figure 3.2B → C) is due to the fact that the C-terminus third of the relay helix is tightly bound to (Figure 3.5) and moves together with the converter domain, whose 40° rotation forces the relay helix to partially unwind locally by a quarter of a turn. Thus, the present results give an essentially complete picture of the coupling mechanics involved during the recovery stroke, a fundamental step in muscle contraction.

3.3 RESULTS

The conformational changes of the structural elements involved in the computed recovery-stroke transition can be divided into two distinct phases, called here the “Phase I” and the “Phase II”. Early in the Phase I, the lever-arm undergoes a ~30° rotation (Figure 3.3A) followed by a more gradual rotation until the end of the Phase II, where the final 20° occurs relatively quickly. Phase I is initiated by the formation of a hydrogen bond between the amide group of Gly457 and the γ -phosphate. While the Phase II is characterized by the movement of the Wedge loop between the N-terminal end of SH1-helix and Switch-2 during the second-half of the conformational pathway.

The RMS coordinate deviation (RMSD) profile for the Switch-2 (Figure 3.3B) exhibits a large initial change (~1.5Å) then remains relatively constant for the rest of the See-saw phase

before increasing by a further $\sim 1\text{\AA}$ during the Phase II. The Wedge loop changes its conformation along with the Switch-2 initially during Phase I and later undergoes a larger conformational change ($\sim 2\text{\AA}$) during Phase II. The Switch-2 loop and the Wedge loop interact via a hydrophobic cluster comprising Phe458 of Switch-2 together with His572, Tyr573 and Ala574 on the tip of the Wedge loop (Figure 3.7). During Phase II, a small change in the Switch-2 conformation ($\sim 1\text{\AA}$) brings about a larger conformational change in the Wedge loop (Figure 3.3B). This movement of the Wedge loop causes further rotation of the converter domain and completes the recovery stroke.

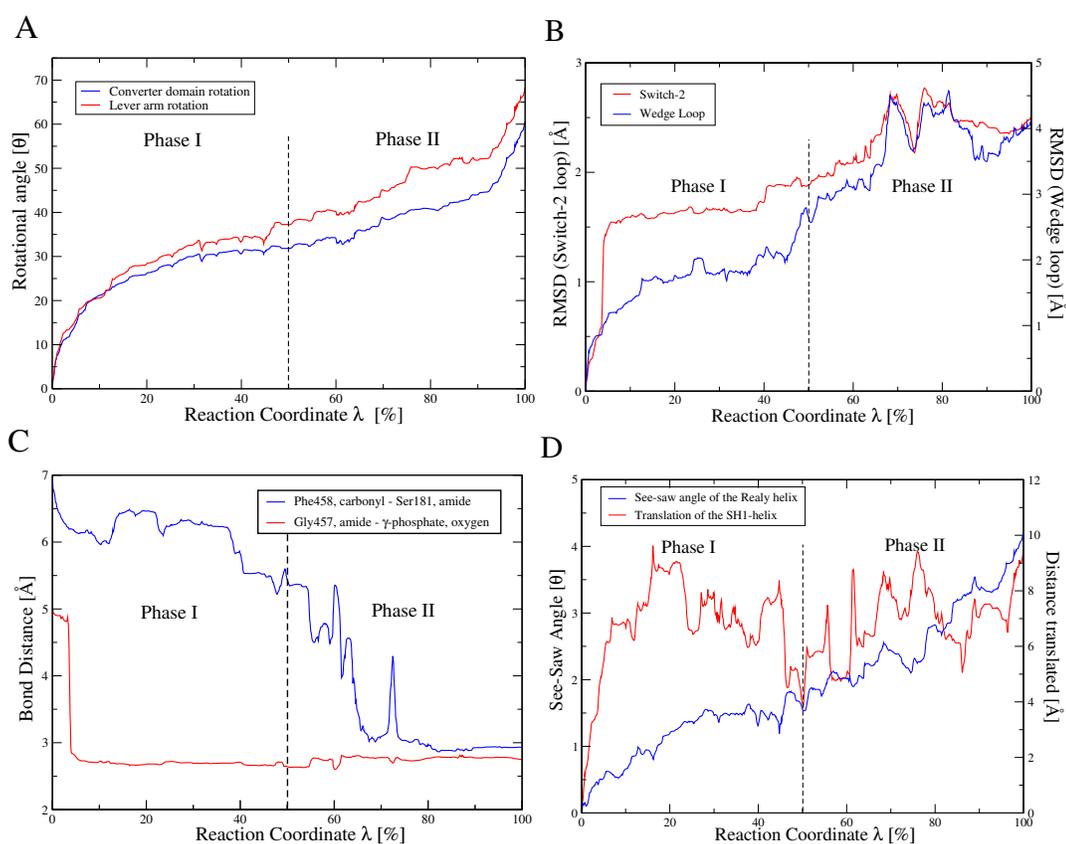


Figure 3.3: Plots of internal rearrangements during the recovery-stroke transition. A. Rotation of the converter-domain and lever arm; B. RMSD profile for the Switch-2 (in blue) and the Wedge loop (red); C. Hydrogen bond distances. Gly457:N_{amide} - ATP:O _{γ -P} (red); Phe458:O_{carbonyl} - Ser181:N_{amide} (Blue); D. See-saw angle of the relay-helix (red) and the SH1-helix longitudinal translation (blue).

3.3.1 PHASE I MECHANISM.

Phase I is initiated by the spontaneous formation of a hydrogen bond between the Gly457 amide group on Switch-2 and one of the oxygen atoms of the γ -phosphate. This hydrogen bond is believed to be essential for myosin ATPase function,^{45,46} because it positions the γ -phosphate in the active site and may facilitate the attack by a nucleophilic group that leads to the cleavage of the γ P- β P bond.⁴⁵

The series of conformational changes that occur during the recovery stroke transition are depicted as a cartoon in Figure 3.4 as a cartoon. The Gly457/Ser456 peptide group is engaged in a hydrogen bond with the side chain of Asn475 (represented by a triangle) which is located on the N-terminal half of the relay helix. This hydrogen bond, which is present in both end states, is thought to be maintained at all times when ATP is bound to the active site.^{4,36,37} Thus, when the Gly457/Ser456 peptide group (red rectangle on switch-2 loop in Figure 3.4) moves toward the γ -Phosphate, it pulls on Asn-475, which moves in the same direction. The net effect of the hydrogen bond formed between Gly457 amide group and γ -Phosphate is thus to pull on the relay helix at position 475, via the hydrogen-bonded-bridge, γ P-Gly457/Ser456-Asn-475-relay-helix. This pull on Asn475 initiates a series of structural changes in the myosin head. The first response is the relay helix motion resembling that of a see-saw (Figure 3.4A \rightleftharpoons 3.4B) in which the relay helix is pulled towards the nucleotide. The fulcrum for the see-saw motion is Phe652 (yellow hexagon in Figure 3.4) which is attached to the strand-3 of the 7-strand- β -sheet and positioned by Phe481 and Phe482 on the relay helix. The relay helix extremities move by $\sim 4\text{\AA}$ and rotate the converter domain and lever arm by $\sim 30^\circ$ (Figure 3.3A) by its contacts with converter domain (Figure 3.5). This completes Phase I of the recovery stroke transition. During this phase there is relatively little movement of the SH1 or SH2-helices (Figure 3.2A & B, right panel and Figure 3.4).

3.3.2 PHASE II MECHANISM.

The formation of a hydrogen bond between the Gly457 amide and the γ -Phosphate during Phase I, reduces the distance between Phe458 on Switch-2 and Ser181 on the P-loop by $\sim 1\text{\AA}$ (Figure 3.3C & Figure 3.4B \rightleftharpoons 3.4C). This also pulls the Wedge loop towards the ATP site by $\sim 1\text{\AA}$ since the phenyl ring of Phe458 is part of the hydrophobic cluster involving His572, Tyr573 and Ala574,

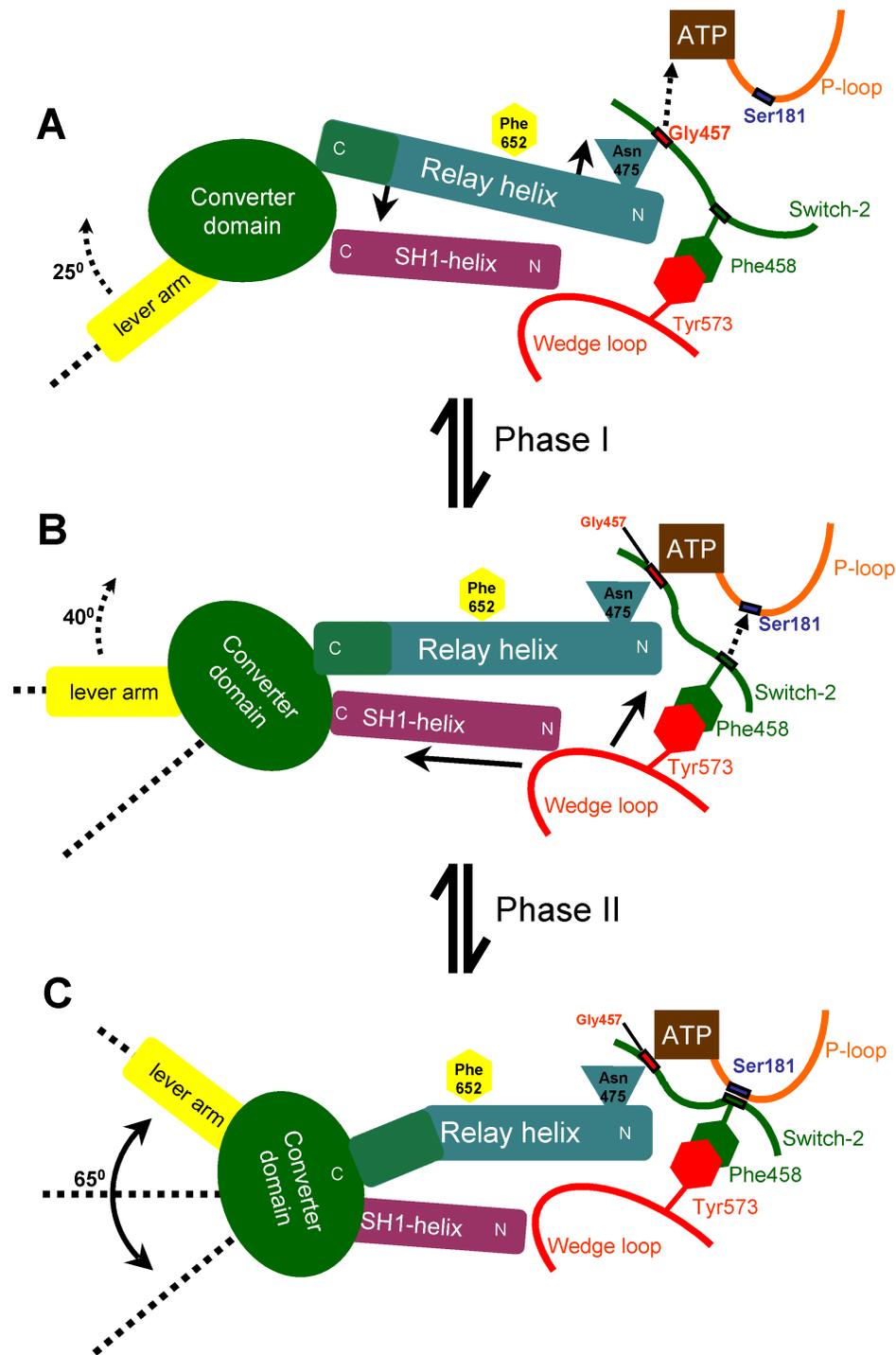


Figure 3.4: Cartoon representing the sequence of coupling events during the return-stroke mechanism.

Wedge loop in red; Switch-2 in green; Relay helix in cyan; SH1-helix in purple; converter domain in green; lever arm in yellow.

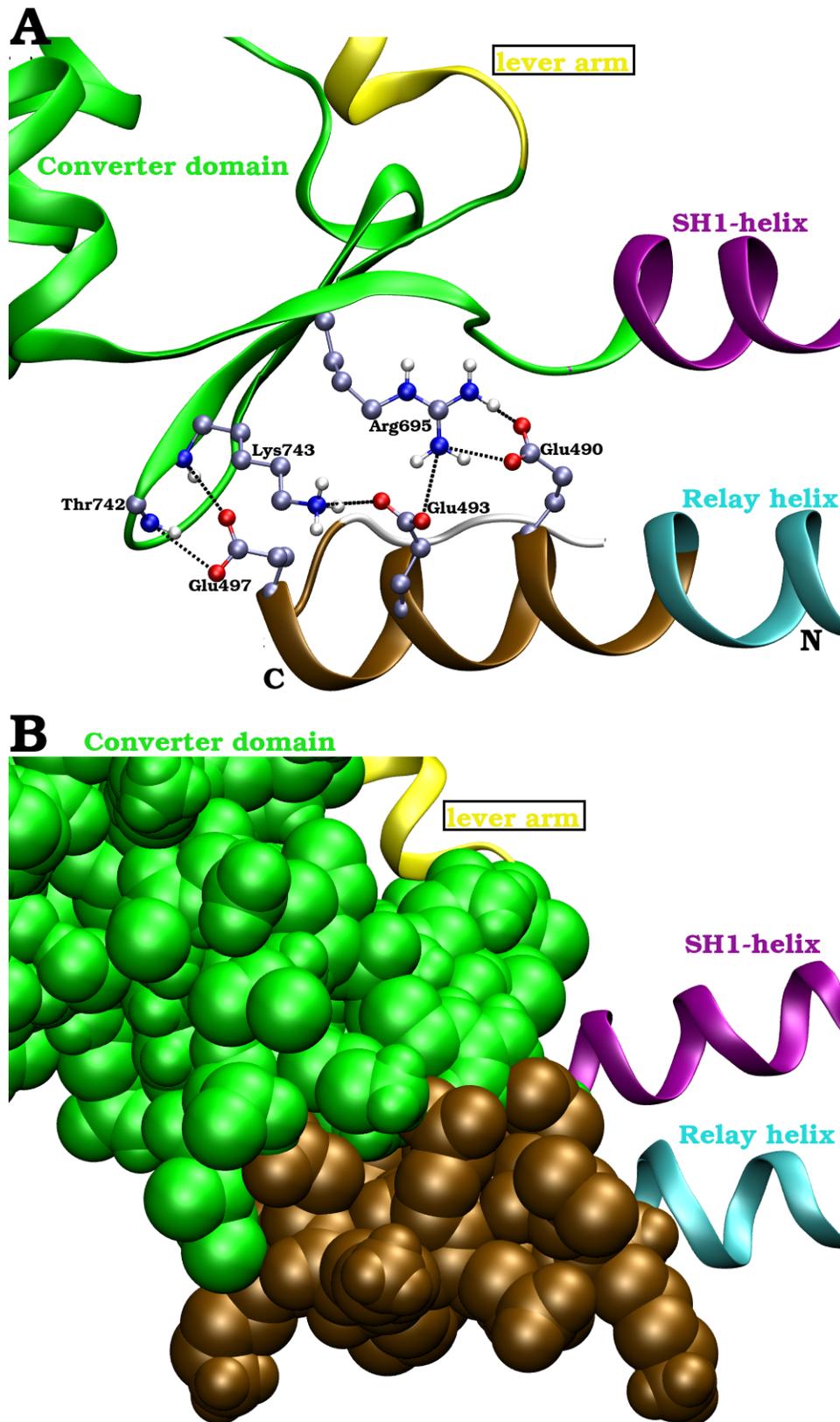


Figure 3.5: Packing between the converter domain (in green) and the C-terminal-third of the relay-helix (in brown). A. Salt-bridges; B. Van der Waals spheres.

which triggers Phase II (Figure 3.4B \rightleftharpoons 3.4C & 3.7). The formation of a hydrogen bond between the Phe458 carbonyl group and Ser181 amide on the P-loop (Figure 3.3C) pulls the Wedge loop towards ATP by ~ 4 Å (Figure 3.3D). This forces the wedge loop to wedge against the N-terminus of SH1 helix (Figure 3.4) and longitudinally translate it by one turn relative to the relay helix (Figure 3.2B & 3.2C, Figure 3.3F and 3.8). The wedging motion thus, pushes the converter domain outward since the converter domain is covalently attached to the SH1-helix.

The converter domain cannot move out-ward since it is strongly anchored at the C-terminus of the relay helix by salt-bridges and non-bonded interactions (Figure 3.5A & B). The converter domain instead rotates and pushes into the C-terminus of the relay-helix through the interactions along the interface between them. The relay helix is anchored to strand-3 of the 7-stranded- β -sheet by the phenyl rings of Phe481 and Phe482 which interact with Phe652 on strand-3. During the recovery stroke, the phenyl ring of Phe456 also rotates and re-packs between the relay helix and the SH1-helix³⁶ and holds the two helices together thorough hydrophobic interactions. The relay helix feels the greatest amount of pressure from the converter domain on Glu490 which is involved in a salt-bridge with Arg695 of the converter domain (Figure 3.5). The relay helix thus breaks (forms a kink^{2,12}) by the breakage of the inter-helical hydrogen bond between Met486 and Glu490. The kink allows further rotation of the converter domain due to the translation of the SH1-helix and completes the recovery stroke transition.

At the end of the recovery-stroke (in State III), phenyl group of Tyr573 belonging to the wedge loop is placed right below the Asn475. Upon ATP hydrolysis, Asn475, which plays a major role during Phase I breaks its interaction with the Gly457/Ser456 peptide group, rotates around χ_1 and switches to the side-chain hydroxyl of Tyr573 to make a hydrogen bond.³⁷ Thus, Asn475 and wedge loop movement provide a mechanism by which myosin is “locked” in State III’ (Figure 3.1) and prevent the reversal of the recovery stroke after ATP hydrolysis and before actin binding.

3.4 DISCUSSION

Regulation of ATP hydrolysis by protein conformations plays an important role for energy transduction in biological systems.⁸⁷⁻⁸⁹ In myosin II, the coupling between the converter domain orientation with the closing of Switch-2 over the ATP makes it unlikely that ATP is hydrolyzed without

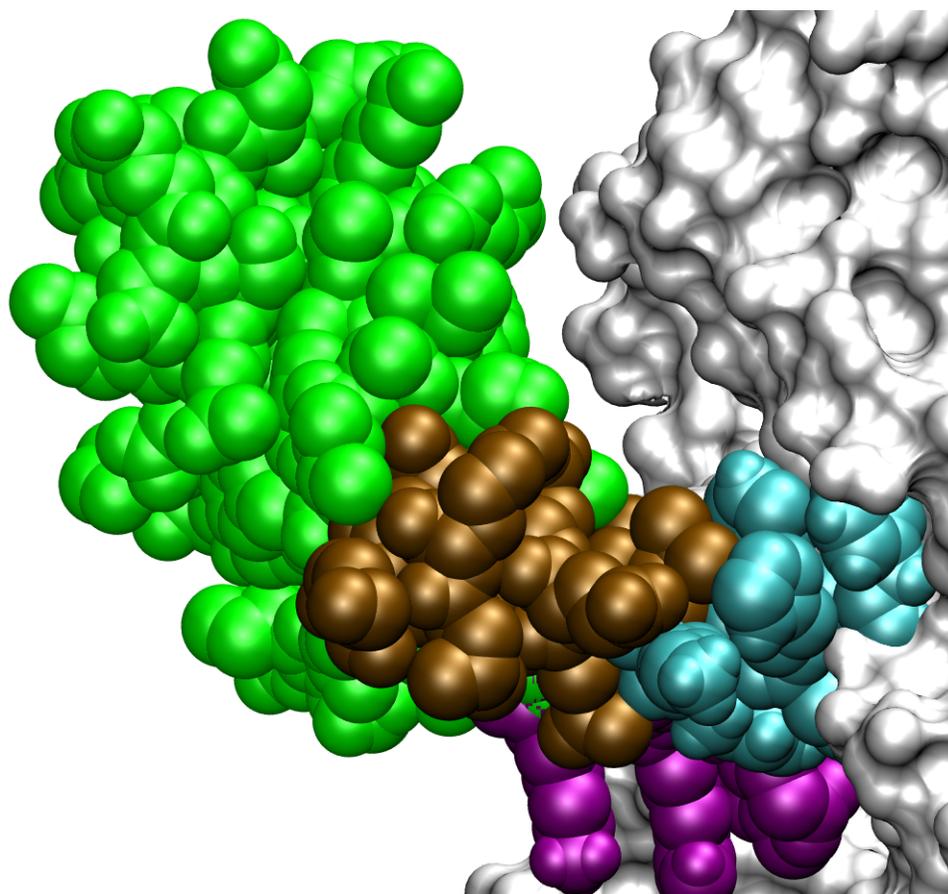


Figure 3.6: Converter domain is held by just two helices, the relay helix and the SH1-helix and can move independent of the rest of the head.

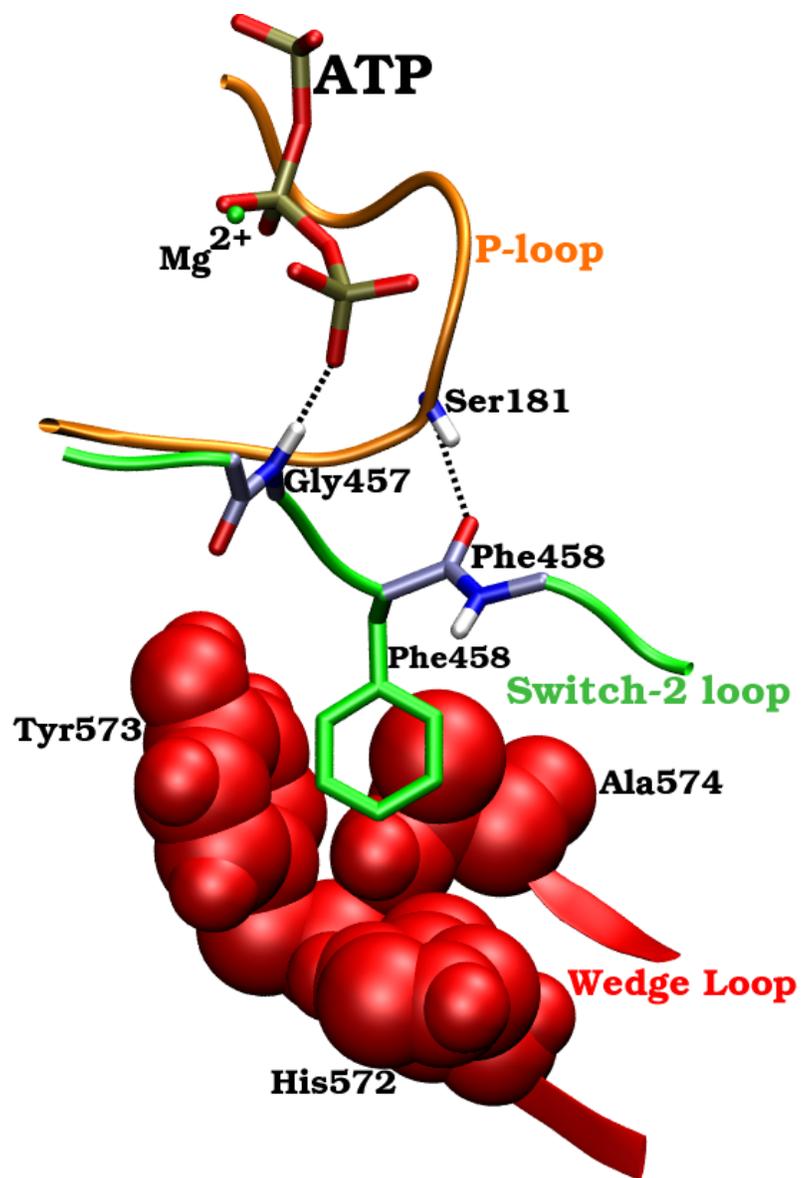


Figure 3.7: The hydrophobic cradle of Phe458.

a productive contraction cycle. This coupling is controlled by small structural changes near the ATP binding site which are amplified into a large domain rotation and vice-versa. The present pathway analysis indicates that the structural coupling between ATPase activation and the recovery stroke occurs in 2 phases. The first, See-saw phase is initiated by the pulling of Gly457 toward the γ -phosphate of ATP, which leads to a see-saw motion of the relay helix and a partial rotation of the converter domain by $\sim 30^\circ$. A molecular fulcrum anchors the relay helix on the main body of the myosin head (at Phe652) and provides a pivoting point for the see-saw motion. The second, Crank-shaft phase, is characterized by the wedging of a loop in-between the N-terminal end of the SH1-helix and the Switch-2 loop. The Wedge loop is a well-conserved structure in the myosin family.^{1,15,37,40} The wedge motion is brought about by movement of Switch-2 loop towards ATP. The strong interaction between the Switch-2 and the Wedge loop pulls the Wedge loop towards the ATP and thus longitudinally translates the SH1-helix relative to the relay helix, leading to further rotation of the converter domain.

The proposed coupling mechanism is consistent with the following experimental point mutational studies:

(i) The mutation of the conserved Ser456 to a Leucine reduces the step size of myosin walking along the actin filament.⁹⁰ The present mechanism suggests that the larger Leucine side chain in this mutant would hinder the movement of the Gly457/Ser456 peptide group toward γ -phosphate in the S456L mutation during the See-saw phase (Figure 3.2A \rightleftharpoons 3.2B), leading to a reduced pull on Asn-475 and thus to a reduced upswing of the relay helix. This hindrance would cause a smaller rotation of the converter domain and hence a smaller step size.

(ii) Two studies involving the mutation of conserved Gly680 to Valine or alanine have shown lower basal ATPase activity and significantly lower in-vitro mobilities.^{91,92} The Wedge loop is just over 3\AA away from the C_α atom of Gly680. The present mechanism suggests that a larger, Valine or alanine residue would prevent the Wedge loop from moving between the SH1 helix and the Switch-2 in the G680V or G680A mutant. The G680V or G680A mutants might also prevent the longitudinal sliding of SH1-helix relative to the relay helix due to wedging, thus preventing the mechanism from proceeding beyond the See-saw phase. The mutation would thus prevent the complete closure of the ATP binding site required for efficient ATP hydrolysis.

The coupling model proposed here could be further tested experimentally because it

predicts the phenotype of several point mutations. For instance, one element of the mechanism involves a pull on the Asn475 side chain by the formation of the γ -phosphate:Gly457 hydrogen bond into a pull on the relay helix (Figure 3.2, right panel). If Asn475 were mutated to a non-polar residue, like Glycine, the hydrogen bond between Asn475 and the Gly457/Ser456 peptide group would be absent. Consequently, this would lessen the pull on the relay helix and might uncouple ATP hydrolysis from converter domain motion during the see-saw phase while the Crank-shaft phase would remain theoretically still active. This mutation might also reduce the ATPase activity in the post-recovery state, due to poorer positioning of the catalytically-important Gly457 amide in the absence of the back hydrogen bond to the Asn475 side chain. In the N475G mutation only the See-saw phase would be abolished which might allow determination of whether the initial see-saw phase is required for the Crank-shaft phase since the Crank-shaft phase is caused by the movement of Phe458 towards the P-loop which should not be affected in this model. If the see-saw phase has no effect on the Crank-shaft phase, the in-vitro mobility rates will not be significantly smaller.

Another point mutation that could give insight into the structural mechanism is mutation of Phe458 to a hydrophilic residue, which would decouple the Switch-2 from the Wedge loop movements. This would essentially block the Crank-shaft phase while the see-saw phase would still be possible, and should thus lead to lower mobility. Consequently, it might be possible to separate and study the two phases during the coupling between Switch-2 movement towards ATP and the rotation of the converter domain. The flow of structural change has been described above in the direction of ATPase activation to lever arm rotation. However, the coupling mechanism is also valid in the reverse direction, i.e., a motion in the converter domain can lead to the corresponding modifications near ATP. The coupling simply ensures that whenever the lever arm is in post-recovery-stroke orientation, the ATPase function is switched on, and when it is in the pre-recovery-stroke position, the ATPase function is switched off. It is plausible that some elements of the coupling mechanism described here might be active during the power stroke, although this is not to imply micro-reversibility, because the power stroke occurs in a different, actin bound, conformation. In the light of the recovery stroke mechanism, one could speculate on the power stroke mechanism. ATP hydrolysis is believed to open Switch-1 and twist the central- β -sheet and the P-loop.^{33,34,45} The twisting of the central β -sheet facilitates actin-cleft closure enabling strong actin binding while the P-loop twist might break the hydrogen bond between Phe458 carbonyl and

Ser181 amide group and allow the Wedge loop, along with the relay helix and switch-2 loop to move away from the ATP binding site. This could allow the back translation of the SH1-helix, leading to straightening of the relay helix and the facilitate back rotation of the converter domain leading to the power stroke and subsequent release of ADP & Pi. At the end of the power stroke (in State I) myosin is found with open Switch-1, Switch-2 in a new conformation denoted by C' with twisted central- β -sheet and P-loop.^{33,34}

In summary, MEP between the two end of the recovery stroke transition has provided insights into key interactions that control the recovery stroke. The Switch-2 and the Wedge loop motions are likely to play key roles during the Lymn-Taylor cycle.

3.5 METHODS

3.5.1 PROTEIN MODELING

The crystal structure 1MMD⁴ complexed with Mg·ADP·BeF₃ (a non-hydrolyzing ATP analog) was used as the “reactant end-state” for the pre-recovery conformation. For the post-recovery conformation, “product end-state”, a structure with Mg·ADP·BeF₃ bound that is very similar to PDB entry 1VOM⁵ was used, as it provides the coordinates for the relay loop whereas 1VOM does not. In both structures, the ATP was modeled by replacing the BeF₃ with a phosphate group. A missing segment in 1MMD (residues 501 to 507) was modeled based on the 2MYS structure⁹³ as previously described.³⁶ 1MMD also lacks coordinates for the non-essential residues 16-35, which were thus left out of all calculations for consistency. 31 crystal water molecules resolved in most myosin structures to date were also included.

3.5.2 PREPARATION OF END STATE STRUCTURES

The reactant and product states were thoroughly energy minimized to relieve bad contacts using CHARMM,⁸⁰ version 29a2 with force-field parameter set 19 for non-aromatic residues and parameter set 22 for aromatic residues.⁹⁴ Unless stated otherwise, all non-bonded interactions were truncated with a switch function between 8 Å and 12 Å⁹⁵ and the non-bonded lists were

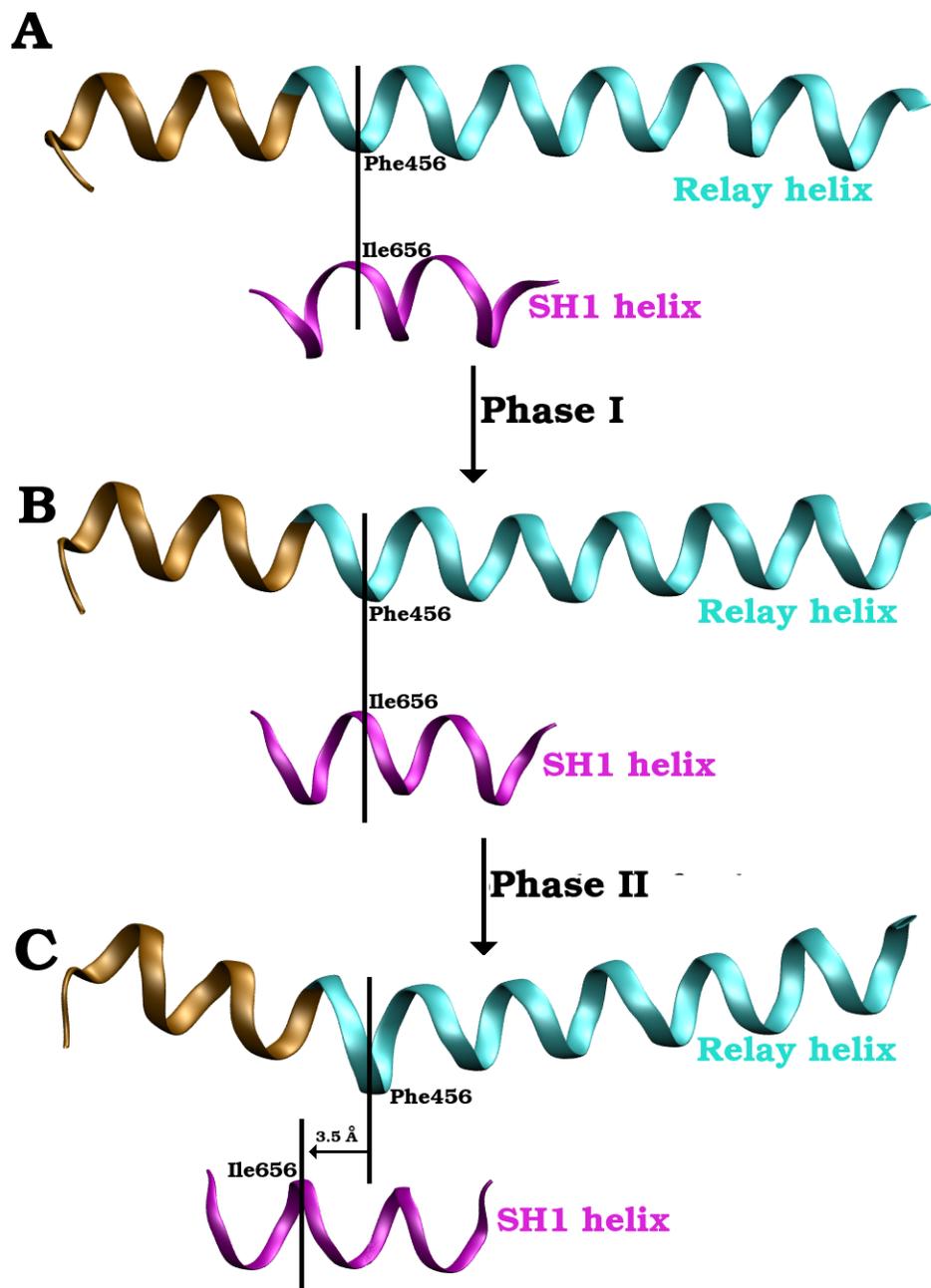


Figure 3.8: Longitudinal translation of the SH1-helix (in purple) relative to the relay helix (in cyan).

updated heuristically. To include electrostatic interactions, we used a variant of GB models called the Analytic Continuum Electrostatics (ACE) model^{68,96} version 2. The Born radius defines the extent of solvation of the charge while calculating the electrostatic solvation energy for a collection of charges. The smaller the Born radius, the higher the solvation.^{57,58,68} Previous studies have indicated that the accuracy of Born radius is the limiting factor for accurate calculation of electrostatic solvation energies.⁷³ Default ACE parameters were optimized here by benchmarking the interaction energies for each pair of charges in the ACE model with the Poisson-Boltzmann (PBEQ) model.⁶¹ In ACE, we use a higher MXBSOLV value of 20 Å (default = 14 Å) which is a parameter that defines the maximum born radii allowed during the electrostatic energy evaluation. This constraint was included to prevent singularities during the evaluation of self energy for each charge and was optimized for small peptides.⁶⁸

A direct minimization often traps the molecule in a high-energy local minimum on a multi-dimensional conformational landscape. Here, Simulated annealing^{97,98} molecular dynamics was used to obtain a conformation that is similar to the end-state crystal structure in terms of RMSD (Root Mean Square Coordinate Deviation) and yet closer to the bottom of the energy basin associated with that particular end state. In performing simulated annealing the energy minimized reactant and product structures were heated to 300 K over 50 ps in steps of 0.6 K per 0.1 ps with harmonic constraints of 1.0 kcal/mol.Å² on all myosin atoms at constant volume (NVE ensemble). During equilibration, the harmonic constraints were reduced to 0.5 kcal/mol.Å² on the back-bone atoms and no constraints were applied to the side-chain atoms. The structures were equilibrated using Nosé-Hoover^{99,100} dynamics at a temperature of 300 K for 750 ps. A production run of 500 ps followed equilibration while maintaining the same constraints on the back-bone and side-chains atoms. The temperature of the system was then slowly reduced to 0 K over 1.5 ns in steps of 0.2 K per 1 ps with harmonic constraints of 0.1 kcal/mol.Å².

Non-relevant structural differences between the annealed end states can lead to undesirable transitions during the CPR³ calculation. These differences can arise from different atom numbering of equivalent atoms, ambiguity of atom positioning in the crystallographic coordinate files or from rotameric transitions during the molecular dynamics simulations. For instance, a CO-NH₃⁺ group can have its hydrogens labeled clockwise as H1, H2, H3 or as H1, H3, H2: arrangements that cannot be mapped by a simple rotation around the C-N bond, but rather would

lead to the exchange of the positions of H2 and H3 during the path, which is associated with a very high energy barrier or a symmetric aromatic ring that has similar orientations in the end states may have its ring atoms named in such a way that it would turn by nearly 180° during the transition. The above undesirable transitions were avoided here by removing all classes of differences listed in Table 3.2. Rotameric differences in non-symmetric groups may also lead to irrelevant barriers that unnecessarily complicate the path. Therefore, all side-chain torsion angles whose differences between the end states exceeded 90° were identified. Each such torsional angle was rotated in the reactant so as to match the corresponding angle in the product and vice versa. One of these two changes was accepted if the rotation did not lead to steric collisions or a poorer hydrogen bond network. All the structural checks mentioned above have been automated with the program called JANUS.¹⁰¹ Both the annealed reactant and product structures were then further energy minimized after checking for non-relevant structural changes using JANUS.¹⁰¹

Table 3.2: Symmetric chemical moieties

Symmetry	Examples
$C_2(sp^2)$	X-CO ₂
$C_3(sp^3)$	X-CH ₃ , X-NH ₃
$C_s(sp^3)$	X-CH ₂ -Y
C_2 rings	Phe, Tyr

The RMS coordinate deviations (RMSD) from the corresponding crystallographic reactant and product end states and the annealed, checked and energy minimized end-states were 2.3 Å and 2.0 Å respectively.

3.5.3 MOLECULAR KINEMATICS

In order to compute the minimum energy pathway (MEP) between the pre-recovery conformation (Figure 3.2A) and the post-recovery conformation (Figure 3.2C) we use CPR, a heuristic method for refining an initial guess path into an MEP.³ Its basic idea of CPR is to identify those points along the path where the energy is highest (the peaks) and to move these points closer to the MEP by a controlled conjugate-gradient minimization. The main advantage of CPR is that it automatically handles paths with many saddle points, each of which is found to any desired accuracy.

The algorithm does not evaluate second derivatives but uses only the energy (which must be continuous) and its gradient. The initial guess path for CPR was generated such that the backbone atoms are interpolated in Cartesian coordinates so as to preserve the backbone fold, and then the side chain atoms are built onto the interpolated backbone, using internal coordinate values. This “combined interpolation” produces an initial path that lies in a reaction channel.⁸⁶ Refining these interpolated states using CPR produces the required MEP between the reactant and the product structures.^{3,36,59,86} Here, the CPR method was used as implemented in the TREK module of CHARMM,⁸⁰ leaving all 7986 atoms free to move independently.

In the previous analysis,³⁶ the solvent was modeled implicitly using a simple distance-dependent dielectric to approximate solvent screening.³⁶ The high polarizability of bulk water, which is reflected by its high dielectric constant, plays a major role in protein conformation and function in physiological conditions.¹⁰² The simple distance dependent dielectric model used in the previous work fails to reproduce the solvent effects accurately.⁵⁹ A straight forward way of more accurately evaluating solvent effects would be to model the bulk solvent with explicit water molecules. The explicit dipole fields of all water molecules then sum to yield the appropriate reaction field for the protein.⁵² However, the explicit solvent molecules increase the size of the system here by a factor of 10, resulting in a large amount of computational time spent on calculating the solvent, rather than the solute of primary interest. Further, explicit solvent “freezes” the solute when minimization-based calculations are performed such as normal mode analysis, ligand docking, or computing minimum-energy pathways between two given conformations of a protein as performed here. An alternative approach is to model the solvent as a implicit polarizable dielectric continuum. The electrostatic potential can then be described by the Poisson-Boltzmann (PB) equation. This second-order differential equation has no analytical solution for any given geometry of a protein and must be solved numerically.⁶⁰ However, a faster approach is to employ analytical approximations to the Poisson-Boltzmann model⁵⁸ such as the Generalized Born (GB) models.⁵⁷ Here, we use a variant of GB models called the Analytic Continuum Electrostatics (ACE) model.^{68,96} The presently-evaluated pathway and that from the previous work indicate closely similar pathways, with differences only in minor details.

SIMULATIONS OF THE MYOSIN II MOTOR REVEAL A NUCLEOTIDE-STATE SENSING ELEMENT THAT CONTROLS THE RECOVERY STROKE

4.1 SUMMARY

During the recovery stroke, the myosin motor is primed for the next power stroke by a 60° rotation of its lever arm. This reversible motion is coupled to the activation of the ATPase function of myosin through conformational changes along the relay helix, which runs from the Switch-2 loop near the ATP to the converter domain carrying the lever arm. Via a hydrogen bond between the side-chain of Asn475 on the relay helix and the Gly457/Ser456 peptide group on the Switch-2, the rotation of the converter domain is coupled to the formation of a hydrogen bond between Gly457 and γ -phosphate that is essential for ATP hydrolysis. Here, molecular dynamics simulations of *Dictyostelium discoideum* myosin II in the two end conformations of the recovery stroke with different nucleotide states (ATP, ADP·Pi, ADP) reveal that the side-chain of Asn475 switches away from Switch-2 upon ATP hydrolysis to make a hydrogen bond with Tyr573. This sensing of the nucleotide state is achieved by a small displacement of the cleaved γ -phosphate towards Gly457 which in turn pushes Asn475 away. The sensing plays a dual role by (i) preventing the wasteful reversal of the recovery stroke while the nucleotide is in the ADP·Pi state, and (ii) decoupling the relay helix from Switch-2, thus allowing the power stroke to start upon initial binding to actin while Gly457 of Switch-2 keeps interacting with the Pi (known to be released only later after tight actin binding). A catalytically important salt bridge between Arg238 (on Switch-1) and Glu459

(on Switch-2), which covers the hydrolysis site, is seen to form rapidly when ATP is added to the pre-recovery stroke conformer and remains stable after the recovery stroke, indicating that it has a role in shaping the ATP binding site by induced fit.

4.2 INTRODUCTION

Molecular motor proteins such as myosins, kinesins and dyneins, use the energy of ATP hydrolysis to generate force and move along actin filaments or microtubules.^{12,13} In myosin, movement is generated by the myosin cross bridges that undergo conformational changes and cyclically interact with actin filaments, as described in the Lymn-Taylor cycle¹⁹ (Figure 4.1, solid arrows). In the rigor conformation, labeled here State I, myosin binds strongly to actin in the absence of ATP. Upon ATP binding to myosin, myosin dissociates from actin and changes into the pre-recovery conformation (also called post-rigor conformation), labeled here State II. Myosin then reversibly undergoes the “recovery stroke” transition to reach the post-recovery conformation (also called pre-power-stroke conformation), labeled here State III. In this conformation, ATP hydrolysis can occur to yield an intermediary State III’, before myosin rebinds to actin (State IV) and performs the “power stroke” to return back to the rigor state while releasing the products of ATP hydrolysis. The Lymn-Taylor cycle is made functional by several coupling mechanisms that link structural changes in different parts of myosin. For example, by sensing the small structural difference between having ATP/Mg²⁺ versus ADP·Pi/Mg²⁺ bound to the nucleotide binding site, myosin controls the conformation of its actin binding region that determines whether the binding affinity for actin is low in step I → II or high in step III’ → IV.^{13,33,34,45} Another example is the activation of the catalytic ATPase function when the lever arm undergoes the recovery stroke in step II → III.^{2,12,35,36} The aim of the present work is, by using molecular dynamics simulations, to understand at atomic detail the coupling mechanisms involved during the recovery stroke and after ATP hydrolysis.

The head of the *Dictyostelium discoideum* Myosin II motor (henceforth ‘myosin’) has been crystallized in the absence of actin with different ATP analogues, (Mg·ATPγS,³⁹ Mg·AMP·PNP,³⁹ Mg·ADP·BeF_x,⁴ Mg·ADP·AlF₄⁻,⁴ Mg·ADP·V_i,⁵) under various conditions. The converter domain/lever arm is found in either of two orientations, between which it is ro-

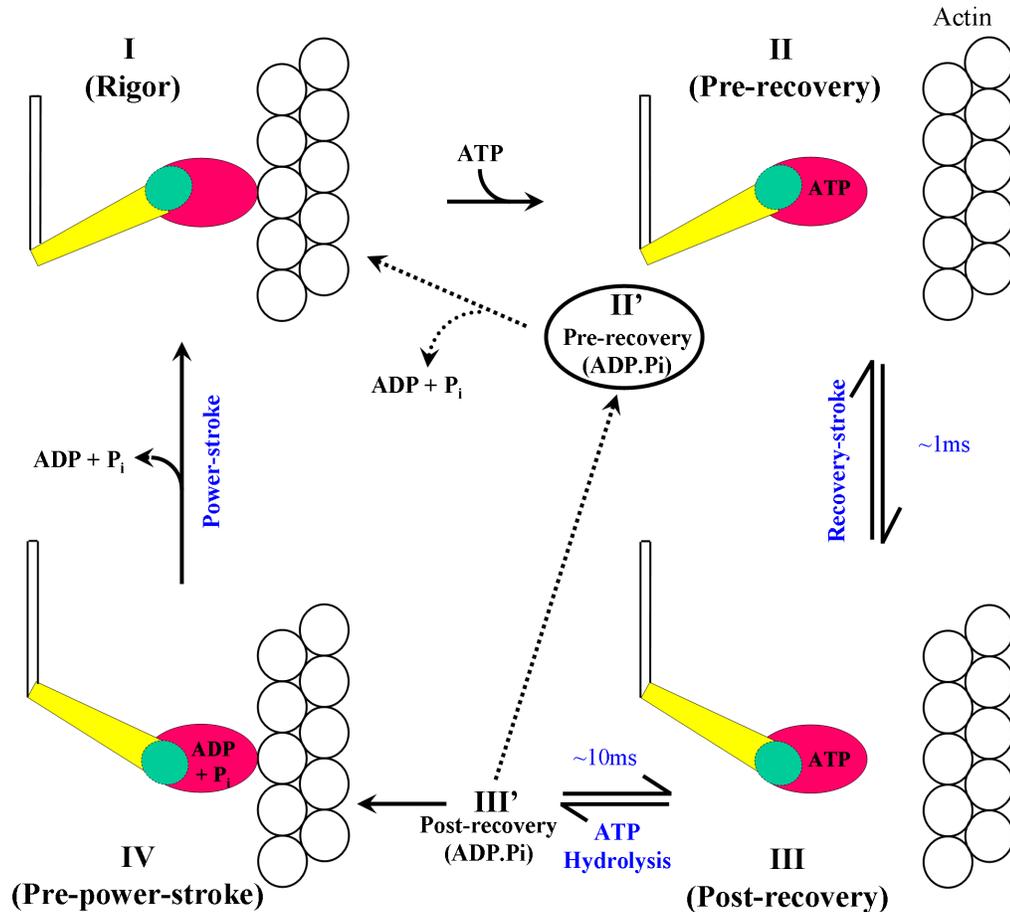


Figure 4.1: Lymn-Taylor Cycle.

The structural domains of myosin are: Lever arm (yellow); myosin head (red); converter domain (green). The actin filament is shown as white spheres. Solid arrows: Standard Lymn-Taylor cycle. Dotted arrows: non-force-generating release of hydrolysis products.

tated by about 60^0 (Figures 4.2A & 4.2D). These two conformations were assigned to States II and III in the Lymn-Taylor cycle.^{2,4,5} The orientation of the converter domain (residues 692-748) is controlled by the position of the “Relay helix” (which spans from the converter domain to the ATP binding site, residues 466-498)³⁶ and by the lengthwise translation of the SH1-helix (to which the converter domain is attached covalently, residues 681-691) relative to the relay helix (Figures 4.2B & 4.2E). The nucleotide-binding site is surrounded by three loop structures, that are conserved among motor proteins (Table 4.1), and also among G-proteins in general.

The p-loop (conserved in many nucleotide binding proteins¹⁰³), the Switch-1 loop, and

Table 4.1: Consensus sequences.

Loop	Consensus ^a	Dictyostelium discoideum	residues ^b
P-loop	GESGAGKT	GESGAGKT	179-186
Switch-1	NxN SSR	NNN SSR	233-238
Switch-2	DxSGFE	DISGFE	454-459

^a Conserved in at least 80 out of 82 myosins¹

^b Residue numbers of *Dictyostelium discoideum* myosin II are used here throughout.

the Switch-2 loop together form the so-called “phosphate tube” around the phosphate moieties of ATP (Figures 4.2C & 4.2F). The p-loop is a Gly-rich element that binds to the phosphate-end of the nucleotide. The two switch loops received their names from the observation that they can each adopt two different conformations, thus possibly serving as switches for information transduction. The combination of Switch-1 and Switch-2 conformations are thought to pertain to different conformational states of the Lymn-Taylor cycle: The Switch-1 closed/Switch-2 open (C/O) conformation corresponds to State II, and Switch-1 closed/Switch-2 closed (C/C) corresponds to State III.^{13,33,34,41}

Significant differences exist between the end states of the recovery stroke. In particular, in State II (pre-recovery) the open Switch-2 is ~ 6.0 Å from the γ -phosphate (Figure 4.2C) while in State III (post-recovery) the closed Switch-2 allows its Gly457/Ser456 peptide group to make a hydrogen bond with the γ -phosphate of ATP (Figure 4.2F). In the conformation of State III, and compared to State II, the relay helix is partially unwound at half length and shifted relative to the SH1-helix, so that the converter domain points the lever arm upwards (in Figure 4.1).

It is known that the ability to catalyze ATP hydrolysis is “off” in the pre-recovery state and is only turned “on” after myosin has undergone the recovery stroke.^{33,34,41,104} This activation of the ATPase function is believed to be due to the closing of Switch-2 and the formation of the hydrogen bond between Gly457 and the γ -phosphate of ATP (Figure 4.2F).^{2,41,46} The coupling of Switch-2 closing and the converter domain/lever arm rotation during the recovery stroke prevents the wasteful hydrolysis of ATP while the lever arm is still in the orientation of State II. Recently, we have proposed a structural model for this coupling mechanism based on the computation of a minimum-energy pathway between the two end-states of the recovery stroke.³⁶ In this model the

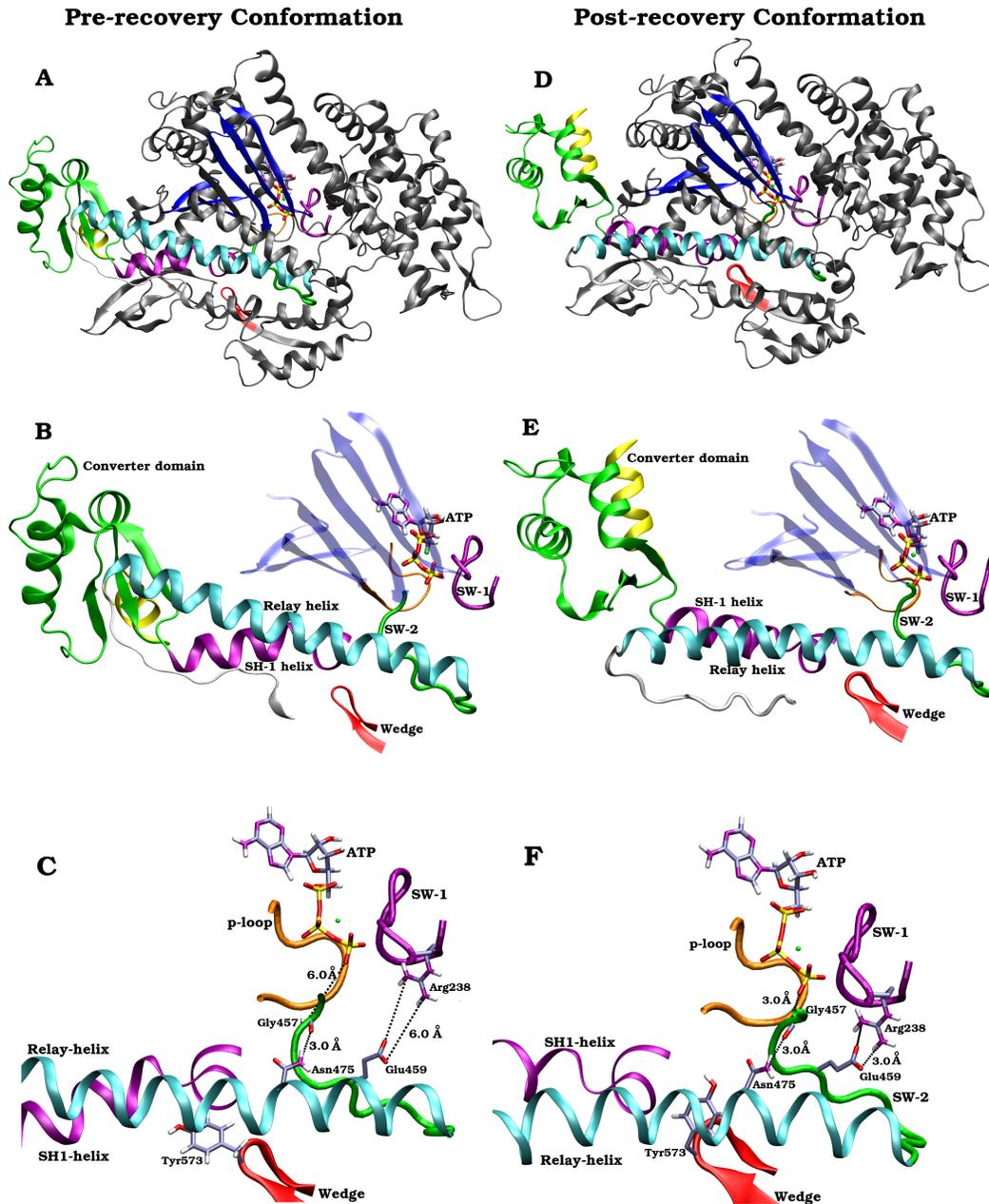


Figure 4.2: End-states of the recovery stroke.

A,B,C) Zoomed views of the Pre-recovery conformation (State II in Figure 4.1). D,E,F) Corresponding views of the Post-recovery conformation (State III in Figure 4.1). Structural elements: Relay helix in cyan; SH-1 helix in purple; converter domain in green; lever arm in yellow 7-stranded β -sheet in blue; P-loop in orange; Switch-1 in purple; Switch-2 in green; Wedge loop in red. Phosphorus atoms of ATP in yellow and Magnesium in green. The pre-recovery structure is PDB entry 1MMD⁴ and the post-recovery conformation is essentially identical to 1VOM⁵ (see Methods). Pictures created using Visual Molecular Dynamics (VMD)⁶ and POVRAY⁷

movement of Gly457 on Switch-2 towards the γ -phosphate is transmitted as a pull on the relay helix, through a hydrogen bond between the Gly457/Ser456 peptide group and the side chain of Asn475 located on the N-terminal half of the relay helix (this hydrogen bond is present in both the end-states, compare Figure 4.2C & 4.2F). The pull on the N-terminal half of the relay helix initiates a “see-saw” motion of the relay helix, translating its C-terminal end (where it is connected to the converter domain) thus pulling on the converter domain, which reacts by a partial rotation (see reference³⁶ for details).

Fluorescence experiments⁴⁷ show that when there is no nucleotide bound or when ADP is bound, myosin predominantly exists in the pre-recovery conformation, which is favored by a factor ~ 10 over the post-recovery conformation. This is confirmed by crystal structures of apo-myosin (PDB IDs: 1FMV⁴³ & 1Q5G³⁴) and of myosin with ADP bound (PDB IDs: 1G8X,¹⁰⁵ 1MMA³⁹), which are all in the pre-recovery conformation. On binding ATP, myosin is found in both the pre-recovery and the post-recovery conformations.⁴⁷ Experiments using fluorescence probes have shown that the conformational equilibrium of myosin with a bound ATP analogue can be made to shift towards either the pre-recovery or the post-recovery conformation by pressure and temperature perturbations.⁴⁷ Myosin has been crystallized in both the pre-recovery and post-recovery conformations under similar experimental conditions in presence of a non-hydrolyzing ATP analogue ($\text{Mg}\cdot\text{ADP}\cdot\text{BeF}_x$) in the catalytic site.^{4,106} The above information indicates that States II and III of the Lymn-Taylor cycle can exchange reversibly when ATP is bound (solid arrows in Figure 4.1). However, this reversibility must somehow be abolished while ADP·Pi is bound to the post-recovery stroke conformation, i.e., Step III' \rightarrow II' (dotted arrows in Figure 4.1) must not be allowed. Otherwise, returning the lever arm to the pre-recovery conformation with the nucleotide in the ADP·Pi state could be followed by actin rebinding (i.e., II' \rightarrow I, dotted arrows in Figure 4.1) since actin can induce a conformational change in the myosin head when ADP·Pi is bound,⁴⁵ leading to actin binding and release of hydrolysis products without force generation. In other words, reversibility of the recovery stroke must occur via states III' \rightarrow III \rightarrow II, after reversal of hydrolysis (indeed step III \leftrightarrow III' is known to be essentially isoenergetic and reversible⁴⁵). This raises a number of questions about the mechanisms that couple the recovery stroke, the activation of the ATPase function and the controlled release of hydrolysis products: (1) How do the interactions with ATP in State II contribute to the coupling between ATPase activation

and the swinging of the lever arm during the recovery stroke ? (2) How does ATP hydrolysis in the post-recovery conformation affect the structure of the active site ? (3) How do these changes prevent the reversal of the recovery stroke while the nucleotide is in the ADP·Pi state ? To address these questions, we performed molecular dynamics simulations of myosin on the two end-states of the recovery stroke and examined the effect of having ATP or ADP bound to the pre-recovery conformation and having either ATP (State III) or ADP·Pi (State III') bound to the post-recovery conformation.

Overall, we find that the p-loop, the Switch-1 and the Switch-2 elements are relatively rigid when ATP is bound to the catalytic site but are generally more flexible with bound ADP or ADP·Pi or in the absence of the nucleotide. The simulations show in which way the key residues around the ATP binding site dynamically react to changes in the nucleotide state. The most striking result is the behavior of the Asn475 side-chain in response to the state of the bound nucleotide. Asn475 makes a stable hydrogen bond to the Gly457/Ser456 peptide group when ATP is bound (States II and III), but when ADP·Pi is bound (post-hydrolysis, State III') this hydrogen bond breaks spontaneously and the Asn475 side-chain switches to form a hydrogen bond with Tyr573. Because in the post-recovery conformation Tyr573 wedges against the end of the SH1-helix (Figure 4.2F), this aromatic residue prevents the length-wise translation of the SH1-helix relative to the relay helix. Thus, formation of the Asn475-Tyr573 hydrogen bond in State III' locks the relay helix and the SH1-helix (and hence the converter domain) in the post-recovery conformation. This locking prevents the reversal of the recovery stroke while the nucleotide is in the ADP·Pi state. As mentioned above, rebinding to actin after such a reversal would be wasteful since it could lead to the release of the hydrolysis products without generating force.

The present simulations give new insights into how changes in the state of the nucleotide affect the structure around the ATP binding site. These effects are instrumental in initiating the coupling mechanisms that make the Lymn-Taylor cycle functional.

4.3 RESULTS

4.3.1 FLEXIBILITY OF THE CATALYTIC SITE

The flexibility of individual structural elements forming the nucleotide binding site (i.e., the p-loop, Switch-1 and the Switch-2) is first analyzed. The time-averaged RMS coordinate fluctuation of each element is plotted in Figure 4.3. In the pre-recovery conformation all the three elements are most flexible in the nucleotide-free state. In absence of nucleotide, the most flexible element is Switch-2 (~ 1.3 Å RMS-fluctuation), followed by the Switch-1 (~ 1.15 Å) and the p-loop (~ 0.75 Å). All these elements become less flexible in presence of ADP and least flexible when ATP is bound. This is indicative of a certain amount of induced fit upon binding the ATP substrate, as might be expected.

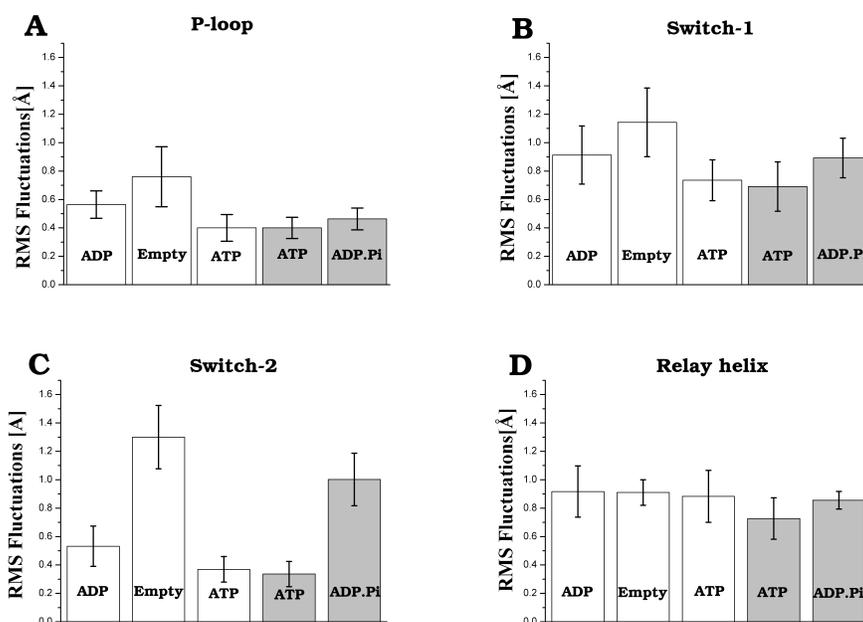


Figure 4.3: Root-Mean-Square coordinate fluctuations of structural elements. A) P-loop; B) Switch-1; C) Switch-2; D) N-terminal half of the relay helix (residues 466-476). White bars: pre-recovery conformation of myosin; grey bars: post-recovery conformation. The state of the nucleotide is indicated in each bar.

Once ATP is bound, going from the pre-recovery to the post-recovery conformation does not significantly modify the flexibility of either of the three elements, which remain rigid.

However, hydrolysis of the ATP results in a clear increase in the flexibility of the Switch-1 and Switch-2 loops, as is apparent from the increase in the coordinate fluctuations in State III' in relative to State III (Figure 4.3B and 4.3C). In contrast, the flexibility of the p-loop is less affected by the hydrolysis of ATP (Figure 4.3A). This suggests that Switch-1 and Switch-2 are the elements that are instrumental in sensing the change of the nucleotide state upon hydrolysis. Thus, changes in the flexibilities of Switch-1 and Switch-2 are likely to be responsible for the required change in the properties of the protein when it is in State III', which are: (i) becoming amenable to the conformational change (induced by actin) that increases actin affinity in State III' relative to State II and (ii) locking of the lever arm into the post-recovery conformation (compared with the freely swinging lever arm in pre-hydrolysis State III). The flexibility of Switch-1 is increased slightly going from the pre-hydrolysis State III (RMS of ~ 0.7 Å) to the post-hydrolysis State III' (RMS ~ 0.9 Å), see Figure 4.3B. This is consistent with the destabilization of Switch-1 that is required for the conformational change of Switch-1 that is believed to accompany the closure of the actin-binding cleft upon actin binding, when going from State III' to State IV.^{33,34,45} The flexibility response to ATP hydrolysis is particularly striking for the Switch-2 (Figure 4.3C). The Switch-2 element is very rigid (time averaged RMS fluctuations of only ~ 0.4 Å) when ATP is bound to the post-recovery conformation (State III), whereas with ADP·Pi bound (State III') the flexibility of Switch-2 increases to an average RMS fluctuation of 1.0 Å. This increase after hydrolysis is due to the decoupling of the Switch-2 from the relay helix, which is caused by the loss of the hydrogen bond between Asn475 and the Gly457/Ser456 peptide group which is triggered by the motion of the cleaved Pi towards Switch-2 (described in detail below).

4.3.2 THE HYDROGEN BONDING OF ASN475

In the crystal structures of both the pre-recovery (State II) and post-recovery (State III) conformations the side chain NH₂ of Asn475 belonging to the relay helix forms a hydrogen bond with the Gly457/Ser456 peptide group belonging to the Switch-2 loop (Figure 4.2C and 4.2F). In previous work, this hydrogen bond was implicated as one of the important interactions coupling the closing of Switch-2 to the rotation of the converter domain during the recovery stroke.³⁶ Therefore, we analyzed the behavior of this hydrogen bond during the MD simulations. In the pre-recovery conformation (State II) the hydrogen bond is found to be very sensitive to the presence of a γ -

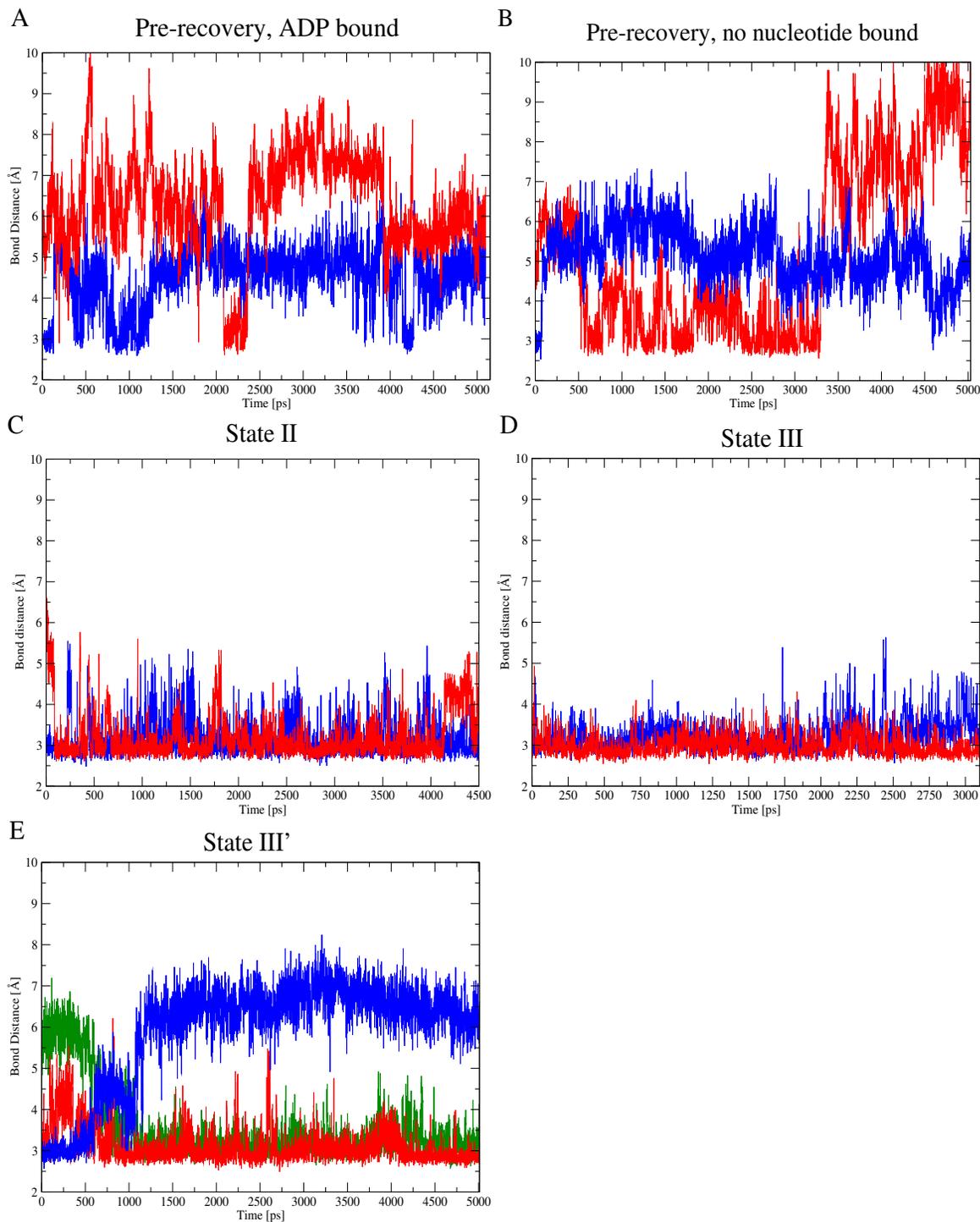


Figure 4.4: Protein-protein interactions in response to the nucleotide state. Hydrogen bond distances: Asn475:ND-Gly457/Ser456:O_{peptide} (Blue), Arg238:N_Z-Glu459:O_E (Red), Asn475:ND-Tyr573:OH (Green). A) Pre-recovery with ADP. B) Pre-recovery without nucleotide (apo). C) Pre-recovery with ATP. D) Post-recovery with ATP. E) Post-recovery with ADP·Pi.

phosphate in the catalytic site, i.e., the bond breaks spontaneously when there is no γ -phosphate. For example with ADP bound (Figure 4.4A, plotted in blue), this hydrogen bond readily breaks in the first 100 ps, reforming only briefly after about 450 ps and 750 ps and after 1.25 ns the hydrogen bond is lost for the rest of the 5 ns simulation. In the simulation of the apo-state (i.e. no nucleotide bound), the interaction breaks already in the first 50 ps (Figure 4.4B, in blue). In contrast, when ATP is bound, this hydrogen bond interaction is quite stable for the total length of the simulation (Figure 4.4C, in blue). Once myosin adopts the post-recovery conformation (State III), the hydrogen bond becomes even more stable (Figure 4.4D, in blue). However, after ATP hydrolysis, i.e., when ADP·Pi is bound in the catalytic site (State III'), the hydrogen bond breaks after only 500 ps, reforms briefly after about 1 ns, and then is lost permanently (Figure 4.4E, in blue). After breaking its interaction with Asn475 side chain, the Ser456 carbonyl oxygen interacts with an additional water molecule coming from a nearby water-pocket. Thus, there is a clear response of this hydrogen bond to the hydrolysis of ATP, going from State III to State III'.

4.3.3 BEHAVIOR OF ASN475 UPON HYDROLYSIS

The reason for the breaking of the hydrogen bond between Asn475 and the Gly457/Ser456 peptide group when going from bound ATP to ADP·Pi in the post-recovery conformation was investigated. In the course of the 5 ns simulation of myosin with bound ADP·Pi, Gly457 on the Switch-2 loop moves away from the nucleotide by ~ 1 Å (Figure 4.7). Even though the Gly457/Ser456 peptide group remains tightly hydrogen bonded to the γ -phosphate, this motion is made possible by a ~ 1.5 Å displacement of the cleaved γ -phosphate away from the β -phosphate (Figure 4.7). As a result, the Switch-2 loop presses against the Asn475 side-chain, which reacts by rotating away (around its torsion angle χ_1), thereby breaking its hydrogen bond with the Gly457/Ser456 peptide group and making a new hydrogen bond with the hydroxyl group of Tyr573 (Figure 4.5). Tyr573 is located on a β -hairpin (residues 571 to 575), called here the “wedge”, which moves by ~ 4 Å during the recovery stroke and wedges against the N-terminus of the SH1-helix (compare Figure 4.2B & 4.2E). In other words, Asn475 switches its hydrogen bond from the Switch-2 loop to the “wedge” in response to ATP hydrolysis.

Upon ATP hydrolysis, there is a charge transfer from the γ -phosphate to the β -phosphate: As ATP^{4-} is converted into $\text{ADP}^{3-} \cdot \text{Pi}^{1-}$, the formal charge on γ -phosphate changes

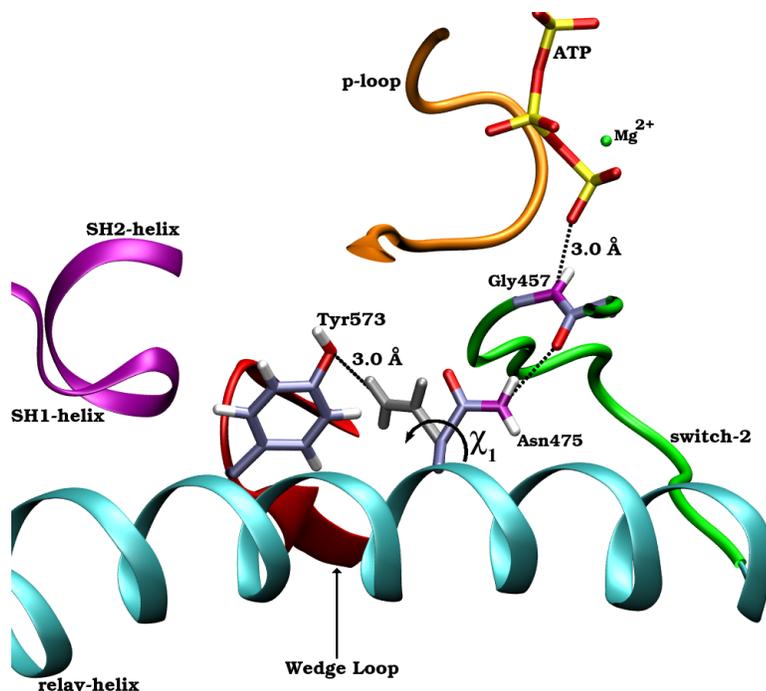


Figure 4.5: Nucleotide sensing by Asn475 in the post-recovery-stroke conformation. Rotation (black arrow) of Asn475 around χ_1 , switching from a hydrogen bond with the Gly457/Ser456 peptide group on Switch-2 to a hydrogen bond with Tyr573 on the wedge loop (residues 571 to 575).

from -2 to -1 and on the β -phosphate from -1 to -2. The question thus arises whether the Asn475 side-chain rotates away due to steric repulsion by Switch-2 or due to the changed electrostatic interaction with the nearby nucleotide. The answer is that the swinging away of the Asn475 side-chain is a steric response to the motion of Switch-2 away from the nucleotide site and is not due to the direct electrostatic interactions between Asn475 and the nucleotide. This becomes evident from the following computational experiment: During a simulation of the post-recovery conformation with ATP bound (State III), the electrostatic interactions of the Gly457/Ser456 peptide group and the nucleotide were modified selectively so that the charge distribution of the ATP seen by this peptide group corresponded to the charge distribution found normally on ADP·Pi, i.e., the peptide group interacts with a -1 charge (instead of the normal -2 charge) on the γ -phosphate and with a -2 charge (instead of -1) on the β -phosphate. All other interactions were kept as they would be in presence of a normal ATP (See Methods). After only 200 ps after this perturbation was introduced to the simulation, the Switch-2 loop moves away from the ATP due to the weakening of

the hydrogen bond between the Gly457/Ser456 peptide group and the γ -phosphate. This motion pushes the side-chain of Asn475, which swings away from Gly457/Ser456 peptide group towards Tyr573 just as described above for simulation with ADP·Pi, (see Figure 4.5).

This can be seen in Figure 4.6, where the electrostatic perturbation is introduced at 350 ps and the hydrogen bond between Asn475 and the Gly457/Ser456 peptide group breaks 200 ps later, concomitantly with the formation of the hydrogen bond between Asn475 and Tyr573. This switching of hydrogen bond partners by the Asn475 side-chain correlates with the movement of the Switch-2 loop away from the γ -phosphate, as seen from the increase in the distance between the Gly457 nitrogen and γ -phosphate (Figure 4.6). During the perturbation, Asn475 interacts with the nucleotide as with a normal ATP⁴⁻, yet perturbing only the interaction of the nucleotide with the Gly457/Ser456 peptide group is sufficient to trigger the motion of the Asn475 side-chain. Thus, while the charge shift on the nucleotide upon hydrolysis may contribute, it is mainly the steric interaction with Switch-2 that is responsible for the switching of hydrogen bond partners by Asn475. When the perturbation is removed (at 715 ps in Figure 4.6), Switch-2 moves back towards the nucleotide and the Gly457 nitrogen restores the initial hydrogen bond with γ -phosphate.

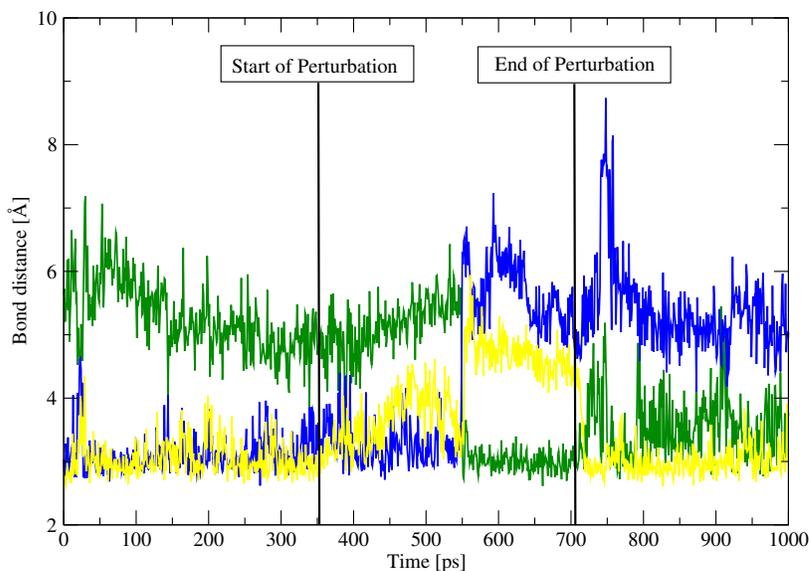


Figure 4.6: Effect of shifting charges on the nucleotide, in the post-recovery conformation. The start and end of the electrostatic perturbations are indicated (modified electrostatics interactions of the Gly457/Ser456 peptide group with the β and γ -Phosphate groups, see text). Hydrogen bond distances: Asn475:N-Gly457/Ser456:O_{peptide} (blue), Asn475:N-Tyr573:O (green) and Gly457- γ -phosphate (yellow).

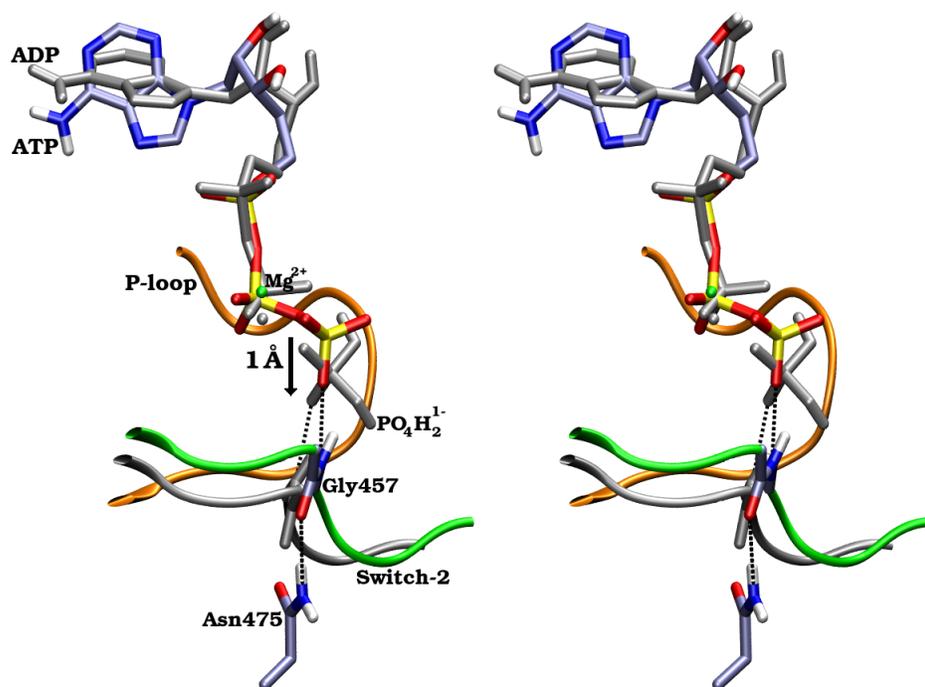


Figure 4.7: Movement of Switch-2 upon hydrolysis in the post-recovery conformation. The post-hydrolysis state III (Crystal structure) is shown in color: P-loop in orange; Switch-1 in purple; Switch-2 in green. The post-hydrolysis state III' (after 5 ns of MD) is shown in grey: ADP, Pi and Switch-2 only. The P-loop and ADP were used for the best-fit overlap of these two conformers. The two panels are for wall-eyed stereo-viewing.

4.3.4 ARG238-GLU459 SALT BRIDGE

Arg238 in Switch-1 and Glu459 in Switch-2 form a salt bridge (see Figure 4.2F) in all post-recovery crystal structures. This salt-bridge has been implicated as having a role in the communication between different functional regions of the myosin head.^{35,46} We analyzed the behavior of the salt bridge in each of the simulations (Figure 4.4A-4.4E, plotted in red). In the pre-recovery conformation, the salt bridge is unstable in the absence of a nucleotide or in the presence of ADP in the catalytic site. With ADP bound, the salt bridge briefly forms for only ~ 250 ps after 2 ns and then remains broken (Figure 4.4A). In the absence of any nucleotide, the salt bridge forms after ~ 500 ps then breaks after 3.25 ns and remains broken for the rest of the 5 ns simulation. In stark contrast, whenever γ -phosphate is present (i.e., when ATP or ADP·Pi are bound), the salt bridge is mostly stable (Figure 4.4C-4.4E). In particular, even though the salt bridge is not present in the

crystal structure that served as the starting point for the simulation of the pre-recovery conformation (PDB-entry : 1MMD⁴), the partners being separated by a distance of ~ 6 Å (see Figure 4.2C), it forms spontaneously (~ 3 Å) after only 20 ps when ATP is added to the active site and remains stable for the rest of 4.5 ns simulation (Figure 4.4C). In the post-recovery conformation, the salt bridge is found to be very stable with ATP bound (Figure 4.4D) and remains mostly stable after ATP hydrolysis (i.e., with ADP·Pi bound, Figure 4.4E).

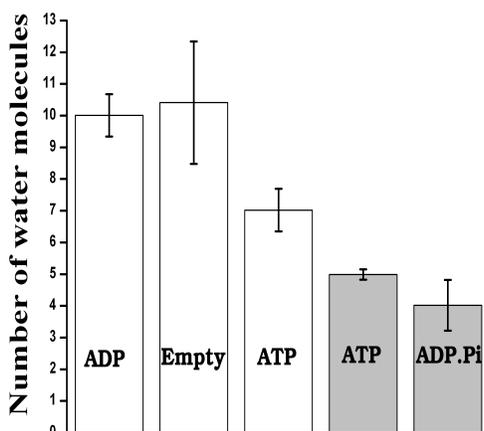


Figure 4.8: Number of water molecules in the catalytic site. Water found within 6.0 Å radius from the geometric center defined over the $C\alpha$ atoms of Ser181 (P-loop), Ser237 (Switch-1) and Gly457 (Switch-2). Bars are labeled as in Figure 4.3.

We find that water content around the ATPase site depends on the bound nucleotide as seen in a similar molecular dynamics study of myosin.¹⁰⁷ Conformations with ATP or with ADP·Pi bound have fewer water molecules near the ATPase site than when ADP or no nucleotide is bound, independent of the myosin conformation (pre-recovery or post-recovery conformations). On average, >10 water molecules are seen around 6.0 Å of ATP site when either ADP or no nucleotide is bound (Figure 4.8). In these states, the salt bridge is not stable (see Figure 4.4A & 4.4B, plotted in red). In the case of bound ATP or ADP·Pi, the number of water molecules is < 7 , while the salt bridge is stable (Figure 4.4C-E). In the absence of the γ -phosphate (or Pi) there is more space available for water in the catalytic site.

4.4 DISCUSSION

In myosin, events at the ATP binding site are tightly coupled to changes in other functional sites of the myosin molecule and vice-versa. Small local structural rearrangements determined by the nucleotide binding state modulate actin binding affinity and the recovery-stroke.^{33,34} Results from fluorescence experiments,⁴⁷ electron-density maps⁴³ and cryo-electron micrographs⁴¹ all suggest that the recovery stroke dominantly occurs when ATP is bound. This shift of the conformational equilibrium away from the pre-recovery state is most likely due to the interactions of the Switch-2 loop with the γ -phosphate of ATP. In particular, the ATPase function is activated when the peptide group of Gly457/Ser456, belonging to the Switch-2 loop, makes a hydrogen bond with the γ -phosphate.⁴⁶ The result is a coupling between activation of the ATPase function and the recovery stroke which ensures that ATP is hydrolyzed only when the converter domain/lever arm is in the post-recovery orientation (i.e., ready to perform the power stroke). The recovery stroke is reversible when ATP is bound.⁴⁷ From the ADP·Pi state III' this reversibility can proceed via state III (i.e., after prior reversal of ATP hydrolysis), but while the nucleotide is in the ADP·Pi state the lever arm is locked in the post-recovery conformation as long as myosin has not bound to actin. This avoids the wasteful release of the hydrolysis products. The present results highlight the role of Asn475 as a sensor to the state of the bound nucleotide during the recovery stroke and after hydrolysis. This residue is involved in the coupling mechanism between ATPase activation and the recovery stroke and is also involved in the locking mechanism of the lever arm after hydrolysis. The following sequence of structural events is proposed for going through State II \rightarrow III \rightarrow III' of the Lymn-Taylor cycle.

In State II, i.e., pre-recovery conformation with ATP bound, the presence of ATP gives rigidity to the Switch-2 loop (as seen from the larger fluctuations of Switch-2 in the absence of a nucleotide, Figure 4.3C). This allows the side-chain of Asn475 on the relay-helix to engage in a hydrogen bond with the Gly457/Ser456 peptide group on the Switch-2 loop (Figure 4.2C). This connects Switch-2 to the N-terminal half of the relay helix, so that when Switch-2 moves towards the γ -phosphate of ATP during the recovery-stroke (States II \rightarrow III), the N-terminal half of the relay helix moves along. This motion of the relay helix is coupled to the rotation of the converter domain (and hence the lever arm) in two ways: (i) By a see-saw like pivoting of the whole relay

helix (as described previously³⁶), whose C-terminal end is connected to the converter domain, and (ii) by a motion of a loop (residues 571 to 575) that is packed against the N-terminal half of the relay helix (Figure 4.2B). This loop wedges against the N-terminal end of the SH1-helix (Figure 4.2E) and pushes the SH1-helix longitudinally relative to the relay helix. This results in further rotation of the converter domain, because it is covalently attached to the C-terminal end of the SH1-helix. We have dubbed the 571-575 loop as the “Wedge-loop” and its behavior during the recovery-stroke will be described in more detail elsewhere. Thus, the hydrogen bond between Asn475 and the Gly457/Ser456 peptide group is essential for the coupling of the activation of the ATPase function to the rotation of the converter domain during the recovery-stroke.

In State III', with ADP·Pi bound, the hydrolyzed γ -phosphate moves away from the β -phosphate and pushes Switch-2 into the Asn475 side-chain (Figure 4.7), which responds by rotating away so as to make a hydrogen bond with the side-chain of Tyr573 on the Wedge Loop (Figure 4.5). As a result, the Wedge loop is now locked in its post-recovery position and prevents the reversal of the longitudinal translation of the SH1-helix relative to the relay helix which in turn prevents reversal of the converter domain rotation. Thus, the Asn475-Tyr573 hydrogen bond is essential in preventing reversal of the recovery-stroke after ATP hydrolysis and before actin binding. The proposed mechanism can be tested by mutating Tyr573, for example into Phe. Such a mutated myosin would be expected to occasionally follow a non-force-generating pathway (dotted arrows in Figure 4.1), thus displaying a reduced efficiency of the motor cycle.

The Wedge Loop (residues 571 to 575), in particular the tip of the loop (composed of amino acids His572, Tyr573 Ala574, see Table 4.2) is a well-conserved structure among 16 classes of myosin motor proteins.^{1,15} In myosin motors that are thought to be processive (Myosin V, VI & VII) the tyrosine at residue 573 is replaced by a phenylalanine (His-Phe-Ala).^{40,108} This precludes the formation of the above described hydrogen bond between the side-chains of Asn475 and residue 573, which thus cannot participate in blocking the Wedge-loop in the post-recovery conformation. However, because processive myosin motors have two heads and both heads are attached to the actin fibril for most of the contractile cycle (even in presence of ATP),^{109–111} reversal of the recovery-stroke is an unlikely event and thus the need for a mechanism for blocking an individual head in the post-recovery conformation after hydrolysis is not needed in these classes of myosins.

Table 4.2: Tip of the “Wedge” loop^a.

Myosin Class	Consensus ^b
I	HYA
II	HYA
III	HYT
IV	HYA
V	HFA
VI	HFA
VII	HFA
VIII ^c	HYA
IX ^c	HYA
X	HYA
X1	HYA
XII	HYA
XIII ^c	HYA
XIV	HTV
XV	HYA
XVI	HYA

^a Residues 572 to 574 in *Dictyostelium discoideum*.

^b Sequences for 143 myosins from public databases.^{1,40}

^c Plant myosins.

The behavior of the salt bridge between Arg238 (on Switch-1) and Glu459 (on Switch-2) shows that this interaction is stable when ATP or ADP·Pi are bound, but is unstable in the absence of the γ -phosphate moiety such as in the case of ADP bound or without a bound nucleotide. Salt bridge formation is seen to correlate with the amount of water in the ATPase site: in the absence of the γ -phosphate moiety more water molecules are present in the active site and hence can better shield the charges on the salt bridge. In contrast, when the γ -phosphate moiety is present (in the case of ATP or ADP·Pi), the space for water is reduced and the salt bridge is more stable. The Arg238-Glu459 salt bridge is not formed in most pre-recovery crystal structures binding an ATP analogue^{4,39} (State II). However, during the present simulations with ATP, this salt-bridge forms spontaneously. The fact that the salt-bridge is broken in the crystal structures could be due to the high salt concentration in the crystallization medium^{4,39} which screens charges and hence might weaken the interaction between Arg238 and Glu459. Alternatively, it could be due to the fact that the ATP analogs (Mg·ADP·BeF₃,⁴ Mg·AMP·PNP,³⁹ etc.,) do not behave exactly as an ATP.

In summary, molecular dynamics simulations of myosin on the conformations at the ends of the recovery stroke with different nucleotides and in the absence of actin have provided insight into key interactions that control the recovery-stroke. These interactions are sensitive to the presence and state of the nucleotides. In particular, the Asn475-Gly457/Ser456 peptide and the Asn475-Tyr573 interactions are likely to play key roles during the Lymn-Taylor cycle.

4.5 METHODS

4.5.1 PROTEIN MODELING

The following protein/nucleotide complexes were prepared: (a) State II (pre-recovery conformation) with $\text{ADP}^{3-}/\text{Mg}^{2+}$, (b) State II (pre-recovery) without nucleotide (apo state), (c) State II (pre-recovery) with $\text{ATP}^{4-}/\text{Mg}^{2+}$, (d) State III (post-recovery) with $\text{ATP}^{4-}/\text{Mg}^{2+}$ and (e) State III (post-recovery) with $\text{ADP}^{3-}\cdot\text{Pi}^{1-}/\text{Mg}^{2+}$. The crystal structure 1MMD⁴ complexed with $\text{Mg}\cdot\text{ADP}\cdot\text{BeF}_3$ (a non-hydrolyzing ATP analog) was used as a starting point for the simulations of the pre-recovery conformation. For the post-recovery conformation, a structure with $\text{Mg}\cdot\text{ADP}\cdot\text{BeF}_3$ bound that is very similar to PDB entry 1VOM⁵ was used, as it provides the coordinates for the relay loop whereas 1VOM does not. In both structures, the ATP was modeled by replacing the BeF_3 with a phosphate group. A missing segment in 1MMD (residues 501 to 507) was modeled based on the 2MYS structure⁹³ as previously described.³⁶ 1MMD also lacks coordinates for the non-essential residues 16-35, which were thus left out of all simulations for consistency. The apo state crystal structure 1FMV⁴³ and 1MMD⁴ are very similar (the backbone RMSD difference is less than 1.0 Å). Therefore, to reduce effects from differences other than the nucleotide state, the apo state of the myosin pre-recovery conformation was taken as the 1MMD structure from which the nucleotide was deleted. The volume occupied by ATP in the binding pocket was calculated to be 450 Å³. This corresponds to about 14 water molecules, which were added into the ATP binding site of the apo-state. The post-hydrolysis state of the post-recovery conformation was modeled by replacing ATP with ADP and Pi ($\text{PO}_4\text{H}_2^{1-}$) in the active site. The Pi and the protons on the Pi were placed as described in the recent QM/MM study investigating the chemo-mechanical coupling of the ATP hydrolysis mechanism.⁴⁵

For each simulation, the protein was placed in an orthogonal box (a=125Å; b=90Å;

$c=75\text{\AA}$) containing water molecules previously equilibrated at 300 K. Waters within 1.75\AA from the protein atoms were deleted, resulting in 27,000 remaining water molecules. The box was replicated with periodic boundary conditions. At least 10\AA separate the protein from each edge of the box. To preserve solution electro-neutrality, sodium ions were placed randomly at distances $>8\text{\AA}$ from the myosin molecule (whose net charge -1 in the present simulations). Three sodium ions were thus added to the complex with ATP or ADP·Pi, two sodium ions to the complex with ADP and one sodium ion to the apo-state.

4.5.2 MOLECULAR DYNAMICS (MD)

MD simulations and energy minimizations were performed using CHARMM,⁸⁰ version 29b2 with force-field parameter set 19⁹⁴ for non-aromatic residues and parameter set 22¹¹² for aromatic residues. Water molecules were modeled with the modified TIP3P potential.¹¹² Non-bonded interactions were truncated with a cubic switching function between 8\AA and 12\AA .⁹⁵ The integration step for MD was 1 fs. Bad initial interactions with water molecules were removed by energy minimization using steepest-descent, followed by a short molecular dynamics (20 ps) simulation at 300 K, keeping the coordinates of the protein fixed. The water was then again energy minimized with the steepest-descent and conjugate gradient methods to a gradient of 0.01 kcal/mol/\AA . The whole system was then energy minimized to a gradient $< 0.01\text{ kcal/mol/\AA}$. The all-atom RMS coordinate deviations with respect to the 1MMD myosin crystal structure of the resulting pre-recovery structure with bound ATP, ADP and no nucleotide were 0.59\AA , 0.68\AA and 0.61\AA , respectively. The all-atom RMS coordinate deviations with respect to the post-recovery crystal structure with bound ATP or ADP·Pi were 0.43\AA and 0.68\AA , respectively. The myosin-in-water system was then heated to 300 K over 20 ps with small harmonic constraints of $0.1\text{ kcal/mol/\AA}^2$ on the positions of the protein atoms, at constant volume. After heating, the harmonic constraints on the protein atoms were reduced slowly to zero (in two steps of $0.05\text{ kcal/mol/\AA}^2$ every 10 ps). The system was then equilibrated in a Nosé-Hoover heat bath^{99,100} at constant pressure of 1 bar and temperature 300 K for 550 ps (NPT conditions). The production MD runs were 3.1 ns and 4.5 ns long for ATP bound to the post-recovery and the pre-recovery conformation, respectively. Production runs of 5.0 ns were performed for the systems of the pre-recovery conformation with bound ADP and no nucleotide and the post-recovery conformation with bound ADP·Pi. The total energy and

temperature were stable. For example, in the pre-recovery conformation with ATP bound, the total energy converges within the first 100 ps of the 550 ps equilibration period and the temperature remains stable (300 ± 1 K), and the backbone RMS deviation between the average structure from 4.5 ns simulation and the 1MMD crystal structure is 2.20 Å.

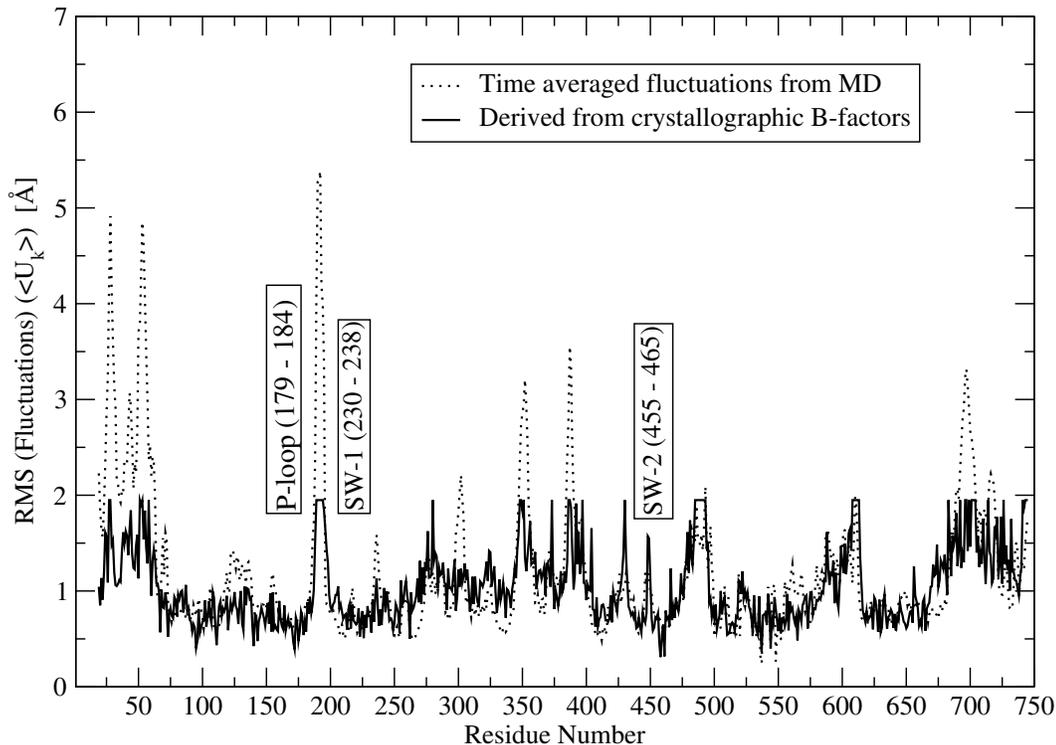


Figure 4.9: Root-Mean-Square fluctuations of C_{α} atoms. Calculated from either the MD trajectory or the experimental B-factors (See Methods).

The flexibility of the protein during MD was compared to the temperature factors determined crystallographically. The time-averaged RMS fluctuations of the C_{α} atoms in the crystal structure, $\langle U_k^2 \rangle$, were derived from the crystallographic B-factors (B_k) using the relation: $\langle U_k^2 \rangle = \frac{3}{8\pi^2} B_k$.¹¹³ The MD fluctuations correlate well with the crystal fluctuations, as shown in Figure 4.9 for the RMS fluctuations of the C_{α} atoms in the pre-recovery conformation with ATP bound (correlation coefficient of 0.77). Note that for the non-resolved residues in the crystal, the temperature factor in the PDB-files is set arbitrarily to 100 \AA^2 , which corresponds to a 2 Å RMS-fluctuation. As expected, during molecular dynamics simulations, non-resolved residues often have higher RMS-fluctuations.

4.5.3 PERTURBATION OF THE ELECTROSTATIC INTERACTIONS WITH ATP.

To perform the computational experiment investigating the cause of the hydrogen bond partner switching of the Asn475 side-chain upon ATP hydrolysis (from the Gly457/Ser456 peptide group to Tyr573), we selectively modified the electrostatic interactions of the Gly457/Ser456 peptide group (atoms C_α , N, H, C & O, whose partial atomic charges sum up to zero) with ATP^{4-} . The electrostatic interaction between the peptide group atoms and the γ -phosphate atoms (P_γ and the three terminal O_γ bonded to P_γ , which together have a net charge of -2) were halved, while the electrostatic interactions with the β -phosphate atoms (P_β and the oxygens bonded to P_β , which together have a net charge of -1) were doubled. This modification mimics the charge changes upon hydrolysis from the initial P_γ^{2-} and P_β^{1-} to P_γ^{1-} and P_β^{2-} . As a result, the Gly457/Ser456 peptide group sees a charge distribution on the ATP that mimics the charge distribution of $ADP^{3-} \cdot Pi^{1-}$, while all other groups in the protein still interact electrostatically with a normal ATP^{4-} and all steric interactions of the nucleotide (including those with the Gly457/Ser456 peptide group) are still those of a normal ATP. After 715 ps, the simulation was switched back to normal interactions between ATP and Gly457/Ser456 peptide group atoms. This perturbation of the electrostatic interactions was done using the BLOCK feature in CHARMM.⁸⁰

4.5.4 ACKNOWLEDGMENTS

We thank S. Schwarzl for helpful discussions and N. Calimet for help with the figures. S. Koppole was supported by a grant from the Deutsche Forschungsgemeinschaft (DFG). All computations were performed on the HELICS supercomputer at the Interdisziplinäres Zentrum für Wissenschaftliches Rechnen (IWR) in Heidelberg. The total CPU time utilized is ~ 250 days on 200 processors i.e., ~ 137 CPU-years (Athlon AMD at 1.4 GHz).

OUTLOOK

The present work described an attempt to understand the conformational mechanism by which Myosin II couples ATPase activation to the $\sim 60^\circ$ rotation of the converter domain during the recovery-stroke in the acto-myosin motor cycle with computational techniques. MEP methods essentially explore the potential energy surface with no entropic component included in the computation and essentially describe the probable mechanism at 0° Kelvin, while MD simulations performed on the two-ends of the recovery-stroke are too short (in the order of a few nano-seconds) to really probe these mechanisms, comprehensively. Thus, the mechanism described in the thesis is a qualitative mechanism and a quantitative procedure still needs to be evolved to understand such complex conformational transitions. Thus, a number of possible projects can be conceived.

5.1 DEVELOPMENT OF ACCURATE CONTINUUM ELECTROSTATICS METHODS.

Generalized Born methods are a huge step forward in evaluating interaction energies of a charge distribution in a protein solvated in a medium of high dielectric. These methods are fast and give reasonably correct solvation free energies (ΔG_{elec}) for systems of small size. When these methods are applied to a large protein-systems like Myosins, the methods are not stable and invariably give large errors while quantitatively describing solvation. It has been shown that the accuracy of a GB method is dependent on the accuracy of the Born radii calculated for each atom in the protein.⁷³ Most GB methods assume the Coulomb-field approximation (described in Chapter 2) which essentially ignores reaction-field effects. Effort has to be done to include some-kind of an approximation

for these effects. Another area that requires refinement is accurately defining the molecular surface as fast as possible which is the speed-determinant in these methods. Considerable progress has been made in the recent years with the publication of Generalized Born Molecular Volume (GBMV)¹¹⁴ and Generalized Born model with a simple smoothing function (GBSW).¹¹⁵ Nevertheless, these models numerically evaluate the Generalized Born equation^{57,58,68} resulting in a discontinuous energy function, thus not useful to work with MEP methods like CPR³ which requires a continuous energy function for gradient evaluation.

5.2 FREE ENERGY CALCULATIONS.

One way to get an estimate for free-energy is by calculating minimum energy pathways (MEP) using methods like CPR³ or Nudged Elastic Band^{116,117} between the reactant and product conformations. This will give a putative mechanism for the conformational change. While using the MEP as the reaction coordinate, one can compute free energies using methods like umbrella sampling to drive the system from the reactant conformation to the product conformation. This gives an estimate for the free-energy along the transition between the initial conformation and the final conformation. Successful application of umbrella sampling methods depends on the choice of the reaction coordinate. MEP would be one of the best estimates as a reaction coordinate for an umbrella sampling method.

5.3 TRANSITION PATHWAY SAMPLING.

One can formulate the conformational changes between two given end-states in protein systems as a “Traveling-Salesman Problem”. The idea is to build a network of points using the “Transition Path Sampling” approach.¹¹⁸ Here, one can use the Minimum Energy pathway (MEP) between the end-states instead of a MD/MC (molecular dynamics or Monte Carlo) trajectory and use points along MEP to shoot trajectories. One can then build a network of points and find the shortest distance between the initial conformation and the final conformation. The distance between any two given points is defined in terms of energy barrier between them, which is computed using a MEP method. It is probable that the protein does not take, just the MEP between the end states but a collection of paths (whose energy barrier is below some threshold) and might pass through

some key transition points (which can be called nodes). If one can find these nodes, it is already interesting to understand function.

5.4 MODELING STATE IV IN THE LYMN-TAYLOR CYCLE

One of the main bottlenecks for computing the conformational change during the “Power-stroke” is the non-availability of the Myosin bound to actin in the pre-power-stroke conformation (State IV, of the Lymn-Taylor Cycle, Figure 1.4). This state has a very small life-time and its in a non-equilibrium conformation, which spontaneously undergoes conformational change to perform the power-stroke.^{12,20,31,33,34} One can model the state IV by performing MD simulations while constraining the global conformation to agree with the 7 or 8Å resolution structural data from cryo-electro microscopy and NMR.^{28,119} Molecular Dynamics simulations could generate a putative structure which can be used to do MEP between State IV and State I and understand the Power-stroke conformational transition in myosin molecular motor.

5.5 MYOSIN II RIGOR CONFORMATION TO MYOSIN II PRE-RECOVERY CONFORMATION

Determination of a possible reaction path corresponding to the O/O to C/O structures of *Dicystelium discoideum* Myosin II, for which crystal structures are available.^{4,34} This can be done with molecular mechanical path calculations using the CPR³ or NEB¹¹⁷ algorithm in analogy to the determination of the reaction path for the return stroke.

5.6 NORMAL MODE ANALYSIS

Once the Normal modes are determined for the Myosin-ATP complex in the C/C and O/C conformations on the potential energy surface, they can then be projected onto the previously determined reaction coordinate using MEP methods. The modes that overlap significantly with the reaction coordinate will couple to the reaction coordinate and thus promote or inhibit the conformational changes in question. The identification of such modes will contribute to understand how the global and collective properties (that emerge only when considering the protein as a single complex entity

rather than as an assembly of somewhat connected atoms) of the myosin protein influence specific functionalities such as the catalysis of ATP hydrolysis.

REFERENCES

1. M. J. T. V. Cope, J. Whisstock, I. Rayment, and J. K. Jones. Conservation within the myosin motor domain: implications for structure and function. *Structure*, 4:969–987, 1996.
2. M. A. Geeves and K. C. Holmes. Structural Mechanism of Muscle Contraction. *Annu. Rev. Biochem.*, 68:687–728, 1999.
3. S. Fischer and M. Karplus. Conjugate Peak Refinement : An algorithm for finding reaction paths and accurate transition states in systems with many degrees of freedom. *Chemical Physics Letters*, 194:252–261, 1992.
4. A. J. Fisher, C. A. Smith, J. B. Thoden, R. Smith, K. Sutoh, H. M. Holden, and I. Rayment. X-ray Structures of the Myosin Motor Domain of Dictyostelium discoideum Complexed with MgADPBeF_x, and MgADPAIF₄⁻. *Biochemistry*, 34(28):8960–8972, 1995.
5. C. A. Smith and I Rayment. X-ray Structure of the Magnesium(II)·ADP·Vanadate Complex of the Dictyostelium discoideum Myosin Motor Domain to 1.9 Å resolution. *Biochemistry*, 35:5404–5417, 1996.
6. W.D. Humphrey. A., and Schulten, K. *J. Mol. Graph*, 14:33–38, 1996.
7. P.O.V.R. Team. Persistence of Vision Ray Tracer (POV-Ray), 1993.
8. K.C. Holmes. Muscle proteins their actions and interactions. *Curr. Opin. Struct. Biol*, 6:781–789, 1996.
9. C. Bustamante, Y. R. Chemla, N. R. Forde, and D. Izhaky. Mechanical Processes in Biochemistry. *Annual Review of Biochemistry*, 73(1):705–748, 2004.
10. M. Schliwa, editor. *Molecular Motors*. Wiley-VCH, Weinheim, 2003.
11. M. Schliwa and W. Woehlke. Molecular Motors. *Nature*, 422:759–765, 2003.
12. K. C. Holmes and M. A. Geeves. The structural basis of muscle contraction. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 355:419–431, 2000.

13. M. A. Geeves and K. C. Holmes. The molecular mechanism of muscle contraction. *Advances in Protein Chemistry*, 71:161–193, 2005.
14. Mermall, V. and Post, PL and Mooseker, MS. Unconventional myosins in membrane traffic, cell movement and signal transduction. *Science*, 279:527–533, 1998.
15. J. R. Sellers. Myosins: A diverse superfamily. *Biochimica et Biophysica Acta*, 1496:3–22, 2000.
16. W. J Wilbur and D. J. Lipman. Rapid Similarity Searches of Nucleic Acid and Protein Data Banks. *Proceedings of the National Academy of Sciences*, 80(3):726–730, 1983.
17. E. W. Myers and W. Miller. Optimal alignments in linear space. *Bioinformatics*, 4(1):11–17, 1988.
18. TP Hodge. Role of the COOH-terminal nonhelical tailpiece in the assembly of a vertebrate nonmuscle myosin rod. *The Journal of Cell Biology*, 118(5):1085–1095, 1992.
19. R. W. Lymn and E. W. Taylor. Mechanism of Adenosine Triphosphate Hydrolysis by Actomyosin. *Biochemistry*, 10(25):4617–4624, 1971.
20. K.C. Holmes and M.A. Geeves. The Structural Basis of Muscle Contraction. *Philosophical Transactions: Biological Sciences*, 355(1396):419–431, 2000.
21. I. Rayment, WR Rypniewski, K. Schmidt-Base, R. Smith, DR Tomchick, MM Benning, DA Winkelmann, G. Wesenberg, and HM Holden. Three-dimensional structure of myosin subfragment-1: A Molecular Motor. *Science*, 261(5117):50–58, 1993.
22. C. A. Smith and I. Rayment. Active site comparisons highlight structural similarities between myosin and other P-loop proteins. *Biophysical Journal*, 70(4):1590–1602, 1996.
23. K. C. Holmes, D. Popp, W. Gebhard, and W. Kabsch. Atomic model of the actin filament. *Nature*, 347:44–49, 1990.
24. I. Rayment, HM Holden, M. Whittaker, CB Yohn, M. Lorenz, KC Holmes, and RA Milligan. Structure of the actin-myosin complex and its implications for muscle contraction. *Science*, 261(5117):58, 1993.
25. M. Lorenz, D. Popp, and KC Holmes. Refinement of the F-actin model against X-ray fiber diffraction data by the use of a directed mutation algorithm. *J Mol Biol*, 234(3):826–36, 1993.
26. W. Kabsch, H.G. Mannherz, D. Suck, E.F. Pai, and K.C. Holmes. Atomic structure of the actin: DNase I complex. *Nature*, 347:37–44, 1990.

27. A. T. Brunger, P. D. Adams, G. M. Clore, W. L. DeLano, P. Gros, R. W. Grosse-Kunstleve, J. S. Jiang, J. Kuszewski, N. Nilges, N. S. Pannu, et al. Crystallography and NMR system (CNS): a new software system for macromolecular structure determination. *Acta Crystallogr. D*, 54:905–921, 1998.
28. R.R. Schröder, D.J. Manstein, W. Jahn, H. Holden, I. Rayment, K.C. Holmes, and J.A. Spudich. Three-dimensional atomic model of F-actin decorated with Dictyostelium myosin S1. *Nature*, 364:171–174, 1993.
29. P. Janmey. Cell biology. A slice of the actin. *Nature*, 364(6439):685–92, 1993.
30. C. E. Schutt, J. C. Myslik, M. D. Rozycki, N. C. W. Goonesekere, and U. Lindberg. The structure of crystalline profilin–bold beta-actin. *Nature*, 365:810–816, 1993.
31. J. E. Molloy, C. Veigel, A. Yildiz, J. N. Forkey, S. A. McKinney, T. Ha, Y. E. Goldman, and P. R. Selvin. Myosin Motors Walk the Walk. *Science*, 300(5628):2045–2046, 2003.
32. K.C. Holmes. The swinging lever-arm hypothesis of muscle contraction. *Curr. Biol*, 7:112–118, 1997.
33. P. D. Coureux, H. L. Sweeney, and A. Houdusse. Three Myosin V structures delineate essential features of chemo-mechanical transduction. *The EMBO Journal*, 23(23):4527–4537, 2004.
34. T. F. Reubold, S. Eschenburg, A. Becker, F. J. Kull, and D. J. Manstein. A structural model for actin-induced nucleotide release in myosin. *Nature Structural Biology*, 10(10):826–830, 2003.
35. G. Li and Q. Cui. Mechanochemical Coupling in Myosin: A Theoretical Analysis with Molecular Dynamics and Combined QM/MM Reaction Path Calculations. *Journal of Physical Chemistry B*, 108:3342–3357, 2004.
36. S. Fischer, B. Windshügel, D. Horak, K. C. Holmes, and J. C. Smith. Structural mechanism of the recovery stroke in the Myosin molecular motor. *Proc. Natl. Acad. Sci. USA*, 102(19):6873–6878, 2005.
37. S. Koppole, J. C. Smith, and S. Fischer. Simulations of the myosin II motor reveal a nucleotide-state sensing element that controls the recovery stroke. *Journal of Molecular Biology*, 361:604–616, 2006.
38. S. Shu, X. Liu, and E. D. Korn. Dictyostelium and Acanthamoeba myosin II assembly domains go to the cleavage furrow of Dictyostelium myosin II-null cells. *Proceedings of the National Academy of Sciences*, 100(11):6499–6504, 2003.

39. A. M Gulick, C. B. Bauer, J. B. Thoden, and I Rayment. X-ray structures of the Mg·ADP, Mg·ATP· γ S, and Mg·AMP·PNP complexes of the Dictyostelium discoideum myosin motor domain. *Biochemistry*, 36:11618–11619, 1997.
40. T. Hodge and M. J. T. V. Cope. Myosin Motor Domain Sequence Alignment. INTERNET: <http://www.mrc-lmb.cam.ac.uk/myosin/trees/colour.html>, 2000.
41. K. C. Holmes, R. R. Schröder, S. Eschenburg, H. L. Sweeney, and A. Houdusse. The structure of the rigor complex and its implications for the power-stroke. *Philosophical Transactions: Biological Sciences*, 359:1819–1828, 2004.
42. S. Park, K. Ajtai, and TP Burghardt. Inhibition of myosin ATPase by metal fluoride complexes. *Biochim Biophys Acta*, 1430(1):127–40, 1999.
43. C. B. Bauer, H. M. Holden, J. B. Thoden, R. Smith, and I. Rayment. X-Ray Structures of the APO and Mg-ATP-Bound States of Dictyostelium Discoideum Myosin Motor Domain. *The Journal of Biological Chemistry*, 275(49):38494–38499, 2000.
44. P.D. Coureux, A.L. Wells, J. Ménétrey, C.M. Yengo, and C.A. Morris. A structural state of the myosin V motor without bound nucleotide. *Nature*, 425:419–423, 2003.
45. S. M Schwarzl, J. C. Smith, and S. Fischer. Insights into the Chemomechanical Coupling of the Myosin Motor from Simulation of Its ATP Hydrolysis Mechanism. *Biochemistry*, 45:5830–5847, 2006.
46. N. Sasaki and K. Sutoh. Structure-mutation analysis of the ATPase site of Dictyostelium discoideum myosin II. *Advances in Biophysics*, 35:1–24, 1998.
47. A. Málnási-Csizmadia, D. S. Pearson, M. Kovács, R. J. Woolley, M. A. Geeves, and C. R. Bagshaw. Kinetic Resolution of a Conformational Transition and the ATP Hydrolysis Step Using Relaxation Methods with a Dictyostelium Myosin II Mutant Containing a Single Tryptophan Residue. *Biochemistry*, 40(42):12727 – 12737, 2001.
48. M. Born and R. Oppenheimer. Zur Quantentheorie der Molekeln. *Ann. Phys.(Leipzig)*, 84(20):457, 1927.
49. M. Stone. Born-Oppenheimer approximation and the origin of Wess-Zumino terms: Some quantum-mechanical examples. *Physical Review D*, 33(4):1191–1194, 1986.
50. M. P. Allen and D. J. Tildesley. Computer simulation of liquids, 385 p, 1987.
51. D.M. York, T.A. Darden, and L.G. Pedersen. The effect of long-range electrostatic interactions in simulations of macromolecular crystals: A comparison of the Ewald and truncated list methods. *Journal of Chemical Physics*, 99(10):8345–8348, 2006.

52. C. Sagui and T. A. Darden. Molecular Dynamics Simulations of Biomolecules: Long-Range Electrostatic Effects. *Annual Review of Biophysics and Biomolecular Structure*, 28:155–179, 1999.
53. R. H. Stote, D. J. States, and M. Karplus. On the treatment of electrostatic interactions in biomolecular simulation. *Journal de Chimie Physique*, 88:2419–2433, 1991.
54. K.A. Sharp and B. Honi. INTERACTIONS IN MACROMOLECULES: Theory and Applications. *Annu. Rev. Biophys. Biophys. Chem*, 19(301):32, 1990.
55. M.K. Gilson. Theory of electrostatic interactions in macromolecules. *Curr. Opin. Struct. Biol*, 5:216–223, 1995.
56. S.A. Hassan and E.L. Mehler. A critical analysis of continuum electrostatics: The screened Coulomb potential-implicit solvent model and the study of the alanine dipeptide and discrimination of misfolded structures of proteins. *Proteins Structure Function and Genetics*, 47(1):45–61, 2002.
57. W. C. Still, A. Tempczyk, R. C. Hawley, and T. Hendrickson. Semi-analytical Treatment of Solvation for Molecular Mechanics and Dynamics. *Journal of American Chemical Society*, 112:6127–6129, 1990.
58. D Bashford and D. A. Case. Generalized Born Models of Macromolecular Solvation Effects. *Annual Reviews in Physical Chemistry*, 51:129–152, 2000.
59. Schwarzl, S. M. And Huang, D. And Smith, J. C. And Fischer, S. Non-uniform Charge Scaling (NUCS): A Practical Approximation of Solvent Electrostatic Screening in Proteins. *Journal of Computational Chemistry*, 26:1359–1371, 2005.
60. W. Im, D. Beglov, and B. Roux. Continuum solvation model: Computation of electrostatic forces from numerical solutions to the Poisson-Boltzmann equation. *Computer Physics Communications*, 111(1):59–75, 1998.
61. W. Im, D. Beglov, and B. Roux. Continuum solvation model: Electrostatic forces from numerical solutions to the Poisson-Boltzmann equation. *Computer Physics Communications*, 111:59–75, 1998.
62. N. Baker, M. Holst, and F. Wang. Adaptive multilevel finite element solution of the Poisson-Boltzmann equation II. Refinement at solvent-accessible surfaces in biomolecular systems. *Journal of Computational Chemistry*, 21(15):1343–1352, 2000.
63. R.E. Bruccoleri, J. Novotny, M.E. Davis, and K.A. Sharp. Finite difference Poisson-Boltzmann electrostatic calculations: Increased accuracy achieved by harmonic dielectric

- smoothing and charge antialiasing. *Journal of Computational Chemistry*, 18(2):268–276, 1997.
64. C.M. Cortis and R.A. Friesner. An Automatic Three-Dimensional Finite Element Mesh Generation System for the Poisson Boltzmann Equation. *Journal of Computational Chemistry*, 18(13):1570–1590, 1997.
65. A. Nicholls and B. Honig. A rapid finite difference algorithm, utilizing successive over-relaxation to solve the Poisson-Boltzmann equation. *Journal of Computational Chemistry*, 12(4):435–445, 1991.
66. M. Born. The mobility of the electrolytic ions. *Z. Physik*, 1:221–249, 1920.
67. L. Onsager. Electric Moments of Molecules in Liquids. *Journal of the American Chemical Society*, 58(8):1486–1493, 1936.
68. M. Schaefer and M. Karplus. A Comprehensive Analytical Treatment of Continuum Electrostatics. *Journal of Physical Chemistry B*, 100:1578–1599, 1996.
69. J. D Jackson. *Classical Electrodynamics* (ed.). Wiley, New York Section, 9:413, 1975.
70. JD Jackson and R.F. Fox. *Classical Electrodynamics*. *American Journal of Physics*, 67:841, 1999.
71. M. Schaefer and C. Froemmel. A precise analytical method for calculating the electrostatic energy of macromolecules in aqueous solution. *J Mol Biol*, 216(4):1045–66, 1990.
72. H. Sklenar, F. Eisenhaber, M. Poncin, and R. Lavery. Including solvent and counterion effects in the forcefields of macromolecular mechanics: The field integrated electrostatic approach (FIESTA). *Theoretical biochemistry & molecular biophysics*. Adenine Press, New York, pages 317–336, 1990.
73. A. Onufriev, D. A. Case, and D. Bashford. Effective Born Radii in the Generalized Born Approximation: The Importance of Being Perfect. *Journal of Computational Chemistry*, 23(14):1297–1304, 2002.
74. U. Matlab. The Mathworks Inc. Natick, MA, 2003.
75. D. McQuarrie. *Statistical Mechanics*. Harper & Row, Harper & Row, New York, 1976.
76. G. Sutmann. *Classical Molecular Dynamics*, pages 211–235. In: *Quantum Simulations of Complex Many-Body Systems: From Theory to Algorithms*, NIC-Directors, 2002.

77. W. G. Hoover. Canonical dynamics: Equilibrium phase-space distributions. *Physical Reviews A*, 31:1695–1697, 1985.
78. M. E. Tuckerman and G. J. Martyna. Understanding modern molecular dynamics: techniques and applications. *Journal of Physical Chemistry B*, 104:159–178, 2000.
79. M.S. Bazaraa and CM Shetty. *Nonlinear programming: theory and algorithms*. Wiley, 1993.
80. B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *Journal of Computational Chemistry*, 4:87–217, 1983.
81. H. M. Warrick and J. A. Spudich. Myosin structure and function in cell motility. *Annual review of cell biology.*, 3:379–421, 1987.
82. S. Itakura, H. Yamakawa, Y. Y. Toyoshima, A. Ishijima, T. Kojima, Y. Harada, T. Yanagida, T. Wakabayashi, and K. Sutoh. Force-generating domain of myosin motor. *Biochemical and Biophysical Research Communications*, 196:1504–1510, 1993.
83. R. Dutzler, T. Schirmer, M. Karplus, and S. Fischer. Translocation mechanism of long sugar chains across the maltoporin membrane channel. *Structure*, 10(9):1273–1284, 2002.
84. A.N. Bondar, M. Elstner, S. Suhai, J.C. Smith, and S. Fischer. Mechanism of Primary Proton Transfer in Bacteriorhodopsin. *Structure*, 12(7):1281–1288, 2004.
85. J. Sopkova-De Oliveira Santos, S. Fischer, C. Guilbert, A. Lewit-Bentley, and J.C. Smith. Pathway for large-scale conformational change in annexin V. *Biochemistry*, 39(46):14065–14074, 2000.
86. F. Noé, F. Ille, J. C. Smith, and S. Fischer. Automated Computation of Low-Energy Pathways for Complex Rearrangements in Proteins: Application to the Conformational Switch of Ras p21. *Proteins: Structure, Function, and Bioinformatics*, 59:534–544, 2004.
87. J Weber and A. E. Senior. ATP synthase: what we know about ATP hydrolysis and what we do not know about ATP synthesis. *Biochimica et biophysica acta*, 1458:300–309, 2000.
88. J. A. Spudich. The Myosin swinging cross-bridge model. *Nature reviews. Molecular cell biology.*, 2:387–392, 2001.
89. S. Karamanou, E. Vrontou, G. Sianidis, C. Baud, T. Roos, A. Kuhn, A. S. Politou, and A. Economou. A molecular switch in SecA protein couples ATP hydrolysis to protein translocation. *Molecular Microbiology*, 34(5):1133–1145, 1999.

90. C. T. Murphy, R. S. Rock, and J. A. Spudich. A Myosin II mutation uncouples ATPase activity from motility and shortens step size. *Nature Cell Biology*, 33:311–315, 2001.
91. B. Patterson, K. M. Ruppel, Y. Wu, and J. A. Spudich. Cold-sensitive mutants G680V and G691C of Dictyostelium Myosin II confer dramatically different biochemical defects. *The Journal of Biological Chemistry*, 272(44):27612–27617, 1997.
92. R. Batra, M. A. Geeves, and D. J. Mamstein. Kinetic analysis of Dictyostelium discoideum myosin motor domains with Glycine-to-Alanine mutations in the reactive thiol region. *Biochemistry*, 38:6126–6134, 1999.
93. L. Rayment, W. R. Rypniewski, K. Schmidt-Base, R. Smith, D. R. Tomchick, M. M. Benning, D. A. Winkelmann, G. Wesenberg, and H. M. Holden. Three-dimensional structure of myosin subfragment-1: a molecular motor. *Science*, 261:80–58, 1993.
94. E. Neria, S. Fischer, and M. Karplus. Simulation of activation free energies in molecular systems. *Journal of Chemical Physics*, 105(5):1902–1921, 1996.
95. J. Norberg and L. Nilsson. On the Truncation of Long-Range Electrostatic Interactions in DNA. *Biophysical Journal*, 79:1537–1553, 2000.
96. M. Schaefer, C. Bartels, and M. Karplus. Solution conformations and thermodynamics of structured peptides: Molecular dynamics simulation with an implicit solvation model. *Journal of Molecular Biology*, 284:835–847, 1998.
97. S Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by Simulated Annealing. *Science*, 220(4598):671–680, 1983.
98. J. Kuriyan, A. T. Brunger, M. Karplus, and W. A. Hendrickson. X-ray refinement of protein structures by simulated annealing: test of the method on myohemerythrin. *Acta Crystallographica Section A, Crystal physics, diffraction, theoretical and general crystallography.*, 45:396–409, 1989.
99. S. Nosé. A unified formulation of the constant temperature molecular dynamics methods. *Journal of Chemical Physics*, 81(1):511–519, 1984.
100. W. G. Hoover. Canonical dynamics: Equilibrium phase-space distributions. *Physical Reviews A*, 31(2):1695–1697, 1985.
101. T Splettstoesser. Filtering Non-relevant Changes from Conformational Transitions in Proteins. Diplomarbeit, Faculty of Biology, University of Heidelberg, 2004.

102. B Honig and A Nicholls. Classical electrostatics in biology and chemistry. *Science*, 268:1144–1148, 1995.
103. A. C. Smith and I. Rayment. Active site comparisons highlight structural similarities between myosin and other P-loop proteins1. *Biophysical Journal*, 70:1590–1602, 1996.
104. S. Park, K. Ajtai, and T.P. Burghardt. Mechanism for coupling free energy in ATPase to the myosin active site. *Biochemistry*, 36(11):3368–3372, 1997.
105. W. Kliche, S. Fujita-Becker, M. Kollmar, D. J. Manstein, and F. J. Kull. Structure of a Genetically Engineered Molecular Motor. *EMBO Journal*, 20:40–46, 2001.
106. F. J. Kull. Private communication. 2000.
107. K. Yamanaka, N. Okimoto, S. Neya, M. Hata, and T. Hoshino. Behavior of water molecules in ATPase pocket of myosin. "Journal of Molecular Structure: THEOCHEM", 758:97–105, 2006.
108. T. Hodge and M. J. T. V. Cope. A Myosin Family Tree. *Journal of Cell Science*, 113:3353–3354, 2000.
109. De La Cruz, E. M. and Wells, A. L. and Rosenfeld, S. S. and Ostap, M.E. and Sweeney, H. L. The kinetic mechanism of Myosin V. *Proc. Natl. Acad. Sci. USA*, 96(24):13726–13731, 1999.
110. De La Cruz, E. M. and Ostap, M.E. and Sweeney, H. L. Kinetic Mechanism and Regulation of Myosin VI. *Journal of Biological Chemistry*, 276(34):32373–32381, 2001.
111. Baker, J. E. and Kremontsova, E. B. and Kennedy, G.G. and Armstrong, A. and Trybus, K. M. and Warshaw, D. M. Myosin V processivity: Multiple kinetic pathways for head-to-head coordination. *Proc. Natl. Acad. Sci. USA*, 101(15):5542–5546, 2004.
112. A. D. MacKerell, Jr., D. Bashford, M. Bellott, R. L. Dunbrack Jr., J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J.C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorcikiewicz-Kuczera, D. Yin, and M. Karplus. All-atom empirical potential for molecular modeling and dynamics Studies of proteins. *Journal of Physical Chemistry B*, 102:3586–3616, 1998.
113. L. Meinhold and J. C. Smith. Fluctuations and Correlations in Crystalline Protein Dynamics: A Simulation Analysis of Staphylococcal Nuclease. *Biophysical Journal*, 88:2554–2563, 2005.

-
114. M. S. Lee, F. R. Jr. Salsbury, and C. L. III Brooks. Novel Generalized Born Methods. *Journal of Chemical Physics*, 24:10606–10614, 2002.
 115. W. Im, M. S. Lee, and C. L. III Brooks. Generalized Born Model with a Simple Smoothing Function. *Journal of Computational Chemistry*, 24:1691–1702, 2003.
 116. H. Jonsson, G. Mills, and K.W. Jacobsen. Nudged elastic band method for finding minimum energy paths of transitions. *Classical and Quantum Dynamics in Condensed Phase Simulations*, pages 385–404, 1998.
 117. Henkelman, G. and Uberuaga, B.P. and Jónsson, H. A climbing image nudged elastic band method for finding saddle points and minimum energy paths. *Journal of Chemical Physics*, 113(22):9901–9904, 2006.
 118. P.G. Bolhuis, D. Chandler, C. Dellago, and P.L. Geissler. TRANSITION PATH SAMPLING: Throwing Ropes Over Rough Mountain Passes, in the Dark. *Annual Review of Physical Chemistry*, 53(1):291–318, 2002.
 119. I. Angert, F.J. Kull, W. Jahn, and R.R. Schröder. Electron cryo-microscopy shows how strong binding of myosin to actin releases nucleotide. *Nature*, 425:423–427, 2003.

Erklärung gemäß §8 (3) b) und c) der Promotionsordnung:

- (a) Ich erkläre hiermit, dass ich die vorgelegte Dissertation selbst verfasst und mich keiner anderen als der von mir ausdrücklich bezeichneten Quellen und Hilfen bedient habe,

- (b) Ich erkläre hiermit, dass ich an keiner anderen Stelle ein Prüfungsverfahren beantragt bzw. die Dissertation in dieser oder anderer Form bereits anderweitig als Prüfungsarbeit verwendet oder einer anderen Fakultät als Dissertation vorgelegt habe.

Sampath Kumar Koppole
November 11, 2006
Heidelberg.

