Inaugural-Dissertation

zur
Erlangung der Doktorwürde
der
Naturwissenschaftlich-Mathematischen
Gesamtfakultät
der
Ruprecht-Karls-Universität
Heidelberg

Vorgelegt von

Dipl.-Chem. Sonja M. Schwarzl
aus Birkenfeld/Nahe

Tag der mündlichen Prüfung: 10. Februar 2006

Understanding the ATP Hydrolysis Mechanism in Myosin Using Computer Simulation Techniques

Gutachter: **Prof. Dr. Peter Comba**
Prof. Dr. Jeremy C. Smith

Bei mir ist wohl immer noch die Mischung von Mystik und Mathematik dominant, die ihre Hauptresultate in der Physik findet. Aber auch die Nachbargebiete wie Parapsychologie und Biologie interessieren mich in zunehmendem Maße.

Wolfgang Pauli (1, S. 182)

Summary

Molecular motors are proteins that convert energy from nucleoside triphosphate hydrolysis into mechanical work. A prominent example is myosin which drives muscle contraction and a large number of additional cellular transport phenomena in all living organisms. While hydrolyzing ATP, myosin translocates along an actin filament. The catalytic cycle for ATP hydrolysis and the mechanical motor cycle are closely coupled. Although a large number of studies have been devoted to understanding the functioning of myosin since its isolation in the 19th century, the details of the chemical mechanism underlying ATP hydrolysis and its coupling to the necessary conformational changes of myosin are poorly understood.

In this thesis, theoretical methods are developed and used to gain a detailed understanding of the mechanism of ATP hydrolysis in myosin and of mechanical events that immediately follow hydrolysis. Three different possible reaction routes are investigated using combined quantum mechanical and molecular mechanical (QM/MM) reaction path simulations. To include solvent screening effects in the calculations, a new approximate method "Non-Uniform Charge Scaling" (NUCS) was developed which scales the partial atomic charges on the molecular mechanical atoms so as to optimally reproduce electrostatic interaction energies between groups of protein atoms and the QM region as determined from an initial continuum solvent analysis with a simple Coulomb potential and scaled charges. NUCS is a generally-applicable method that is particularly useful in cases where an explicit treatment of water molecules is not feasible and interfaces to implicit solvent models are lacking, as is the case for current QM/MM calculations.

Path optimizations were done using Hartree-Fock calculations with 3-21G(d) and 6-31G(d,p) basis sets, followed by point energy calls using density-functional theory B3LYP/6-31+G(d,p). Despite the inaccuracies inherent in this method, the present calculations currently represent the most accurate QM/MM theoretical investigation of an enzyme-catalyzed phosphoanhydride hydrolysis reaction. Possible methodological improvements for future investigations are discussed.

The three pathways studied are isoenergetic within error and are thus equally likely to be populated. The 6-31G(d,p) basis set proved to be reliable in describing the geometries during the phosphate hydrolysis reactions, whereas the 3-21G(d) basis set was found to be too inaccurate. Although the energies were not sufficiently accurate, a number of structural conclusions on the mechanism of ATP hydrolysis can be drawn and related to experimental findings from isotope exchange and mutation studies.

All three paths investigated follow a single-step associative-like mechanism (see movies in

the accompaying CD) and show very similar heavy-atom positions in the transition states regardless of the positions of the protons. In the product states, the coordination bond between Mg^{2+} and Ser237 (and thus the switch-1 loop) is broken. This indicates that product release is likely to occur *via* an exit route that opens by complete opening of the switch-1 loop ("trap door" mechanism). Moreover, the coordination distance between Mg^{2+} and inorganic phosphate (P_i) is extended. This indicates that after hydrolysis this bond may be completely cleaved as an early event necessary for phosphate exit.

Inspired by the simulation results, a Network Hypothesis on the mechanism of ATP hydrolysis in myosin is put forward that combines previous mechanistic proposals and that is consistent with experimental data available from mutational and isotope exchange studies. Moreover, a mechanism is suggested to explain how the catalytic cycle is coupled to the motor activity of myosin.

Zusammenfassung

Molekulare Motoren sind Proteine, die chemische Energie aus der Hydrolyse von Nukleosidtriphosphaten in mechanische Arbeit umsetzen. Ein prominentes Beispiel ist Myosin, das Muskelkontraktion und eine Vielzahl weiterer zellulärer Transportphänome in allen Lebewesen antreibt. Während Myosin ATP hydrolysiert, wandert es an einem Aktinfilament entlang. Der Katalysezyklus und der Motorzyklus des Proteins sind dabei eng miteinander gekoppelt. Obwohl eine große Zahl wissenschaftlicher Arbeiten seit der Entdeckung von Myosin im 19. Jahrhundert der Aufklärung der Funktionsweise von Myosin gewidmet waren, sind die Einzelheiten des chemischen Mechanismus der ATP Hydrolyse und deren Kopplung an den Motorzyklus bislang wenig verstanden.

In der vorliegenden Dissertation werden theoretische Methoden entwickelt und angewendet, um ein detailliertes Verständnis des ATP Hydrolysemechanismus in Myosin und der mechanischen Ereignisse, die direkt auf die Hydrolyse folgen, zu erlangen. Drei verschiedene Reaktionswege werden mit kombinierten quantenmechanische/molekularmechanische (QM/MM) Simulationstechniken berechent. Um Lösungsmitteleffekte in den Rechnungen zu berücksichtigen, wurde eine neue approximative Methode "Non-Uniform Charge Scaling" (NUCS) entwickelt. NUCS skaliert die Partialladungen der MM-Atome so, daß die elektrostatische Wechselwirkungsenergie zwischen MM:MM und QM:MM-Atomgruppen nach einem Poisson-Boltzmann Modell mit einem einfachen Coulomb-Potential optimal reproduziert werden. NUCS ist eine allgemeine Methode, die besonders in Fällen anwendbar ist, in denen eine explizite Modellierung der Wassermoleküle nicht möglich ist und Schnittstellen zu impliziten Lösungsmittelmodellen fehlen, wie dies bei QM/MM Rechnungen der Fall ist.

Reaktionswegoptimierungen wurden mit Hartree-Fock Rechnungen unter Verwendung eines 3-21G(d) oder 6-31G(d,p) Basissatzes durchgeführt. Entlang der optimierten Wege wurden Einzelpunktenergien mit dem B3LYP Dichtefunktional und einem 6-31+G(d,p) Basissatz gerechnet. Trotz der dieser Methode inhärenten Ungenaugkeit stellen die vorgelegten Rechnungen die derzeit genauesten QM/MM Rechnungen einer enzymkatalysierten Phosphoanhydrid-Hydrolyse dar. Mögliche methodische Verbesserungen zum Erreichen höherer Genauigkeit in zukünftigen Studien werden diskutiert.

Die drei untersuchten Reaktionwege sind innerhalb des Fehlers der Methode isoenergetisch und somit gleichermaßen populiert. Der 6-31G(d,p) Basissatz lieferte verläßliche Geometrien zur Beschreibung der Phosphoanhydrid-Hydrolyse, während sich der 3-21G(d) Ba-

sissatz als zu ungenau herausstellte. Obwohl die Energien kein ausreichende Genauigkeit erreichten, lassen sich wichtige strukturelle Schlußfolgerungen bzgl. des Reaktionsmechanismus der ATP Hydrolyse in Myosin ableiten und auf experimentelle Ergebnisse aus Isotopenaustausch- und Mutationsstudien beziehen.

In allen drei untersuchten Wegen vollzieht sich die Reaktion in einem einzigen assoziativen Schritt (siehe Movies auf der Begleit-CD), wobei die Positionen der schweren Atome im Übergangszustand fast identisch und unabhängig von der Position der Protonen sind. In den Produktzuständen ist die Koordinationsbindung zwischen Mg^{2+} und Ser237 (und damit der Switch-1 Schleife) gebrochen und die Koordinationsbindung zwischen Mg^{2+} und dem anorganischen Phosphat (P_i) gedehnt. Dies legt nahe, daß diese Bindung im Anschluß an die Hydrolyse bricht, was ein notwendiger erster Schritt im Zuge der Phosphat-Freisetzung ist.

Angeregt durch die Simulationsergebnisse wird eine Netzwerkhypothese zur Erklärung des ATP Hydrolysemechanismus in Myosin aufgestellt, die die bislang vorgeschlagenen Mechanismen vereinigt und mit experimentellen Ergebnissen konsistent ist. Darüber hinaus wird ein neuer Mechanismus vorgeschlagen, wie der Katalysezyklus und der Motorzyklus in Myosin gekoppelt sind.

Contents

Chapter 1

Preface

Reactions involving phosphates are among the most important (if not the most important) reactions in biochemistry (2). Phosphates serve not only as building blocks of DNA and RNA but also as intermediates in cellular metabolism. In particular, the nucleotides adenosine monophosphate (AMP), adenosine diphosphate (ADP), and adenosine triphosphate (ATP) play a critical role in maintaining an organism's energy balance. ATP is a high-energy molecule that serves as an energy carrier. In humans, ATP is synthesized by the molecular motor F_0F_1-ATPase in daily quantities that exceeds the body weight. During the hydrolysis reaction

$$ATP + H_2O \longrightarrow ADP + P_i \tag{1.1}$$

in which ATP is decomposed into ADP and inorganic phosphate (P_i), the energy can be regained. Usually ATP hydrolysis is catalyzed by proteins that utilize the energy stored in ATP in order to achieve their tasks. Prominent examples of ATP-consuming proteins are molecular motors such as kinesins, dyneins, and myosins that transform the chemical energy of ATP into motion.

The objective of the present work is to develop a generally-applicable method to treat solvent effects in computer simulations of enzyme-catalyzed reactions, to use it for computational investigations of ATP hydrolysis in the molecular motor myosin, and to thereby contribute to a detailed understanding of the functioning of this motor protein.

This thesis is organized in four parts. In the introductory part, the current status of research on ATP hydrolysis in myosin is reviewed and the questions addressed in the present work are formulated. The second part describes the methods used in the present work. In this part, the newly-developed method is described in a self-contained chapter. The third part describes the results of the reaction path simulations on ATP hydrolysis in myosin, followed

by a detailed assessment of the accuracy of the calculations and discussion of mechanistic aspects of ATP hydrolysis in myosin, and early post-hydrolytic chemo-mechanical coupling events in part four. In the Appendix, supplementary material is given.

Part I

Introduction

Opening: Introduction

In this introductory part, the current status of research on ATP hydrolysis is reviewed. Molecular motors in general are introduced in Chapter 2. Chapter 3 gives a detailed view of the specific features of the motor protein myosin. Many studies have been devoted to elucidating the details of myosin function since myosin was discovered in the 19th century. Due to the complexity of this motor protein and its abundance in all living organisms, it has received enormous scientific interest which is reflected in the vast number of publications available. Nevertheless, the details of the chemical mechanism of ATP hydrolysis and its coupling to the motor cycle of myosin are not well understood. To gain deeper insight into these processes requires not only knowledge of the functioning of the motor protein but also an understanding of the mechanism of ATP hydrolysis in aqueous solution, *i.e.*, the uncatalyzed reaction. Therefore, possible mechanisms of phosphate hydrolysis reactions in general are reviewed in Chapter 4, followed by a compilation of available studies on kinetics and thermodynamics of ATP hydrolysis in myosin. The available experimental data are compiled into a free energy diagram of the enzymatic cycle of myosin. The focus is on myosin II from the slime mold *Dictyostelium discoideum* as this myosin is among the best-characterized myosins and has been used throughout this thesis. At the end of the introductory part a number of questions are put forward that are addressed in the present work (Chapter 5).

Chapter 2

Molecular motors

This chapter provides background information on molecular motors in general (Section 2.1), the myosin family of molecular motors (Section 2.2), and movement strategies used by different myosins (Section 2.3). It thus introduces the scientific context into which the current thesis is embedded.

2.1 Motor proteins

Proteins that utilize energy from the hydrolysis of a nucleoside triphosphate and convert it into mechanical work are called molecular motors. Molecular motors are involved in a large variety of cellular tasks. As a group, the motor proteins enable cells to move, contract, change shape, secrete, endocytose, and organize the cytoplasm (3).

There are three families of molecular motors that are involved in motion along filaments: Myosins, kinesins, and dyneins. Myosins move along actin filaments, whereas kinesins and dyneins translocate along microtubules.[1] A compilation of motor protein organization, function, and regulation is given in (5; 6). Motors involved in intracellular transport and trafficking are reviewed in Ref. (7).

Apart from the motor proteins that move along predefined tracks there are molecular motors that couple proton motive force to hydrolysis of nucleoside triphosphate. Examples for this type of motors are the bacterial flagellar motor (5) or the F_1 motor of ATP synthase (8; 9; 10; 11; 12; 13). In addition, molecular motors are involved in RNA and DNA polymerization (polymerases) and unwinding of nucleic acids (helicases).

[1] The cytoskeleton is formed by three types of filaments: Actin filaments are homopolymers consisting of actin monomers, microtubules are formed by tubulin heterodimers, and intermediate filaments are polymeric structures build from various proteins with an α-helical coiled-coil structure (4).

2.2 Myosin classes

The different motors belonging to the myosin superfamily are classified into 18 different myosin classes (14; 15). An unrooted phylogenetic tree visualizing the relationships between the different myosins as judged by sequence similarity is shown in Fig. 2.1. All myosins share the common feature that they consist of a globular motor domain (the head domain), an α-helical neck domain, that binds two light chains, and a long α-helical tail. Small structural changes at the catalytic site in the motor domain are converted into a large swing of the light-chain binding domain that thus serves as a lever arm.

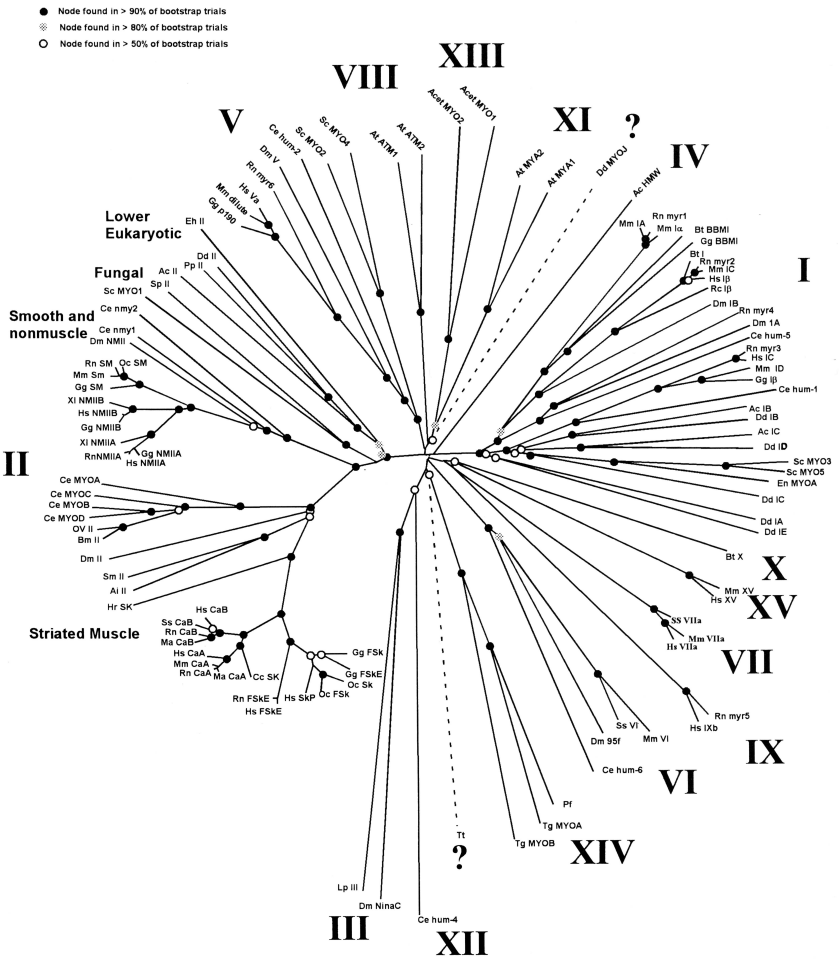

Figure 2.1: Unrooted phylogenetic tree obtained from the myosin motor domain sequences. The figure is taken from Ref. (14)

The best characterized myosins are the class II myosins (also called conventional myosins).

The most prominent among the class II myosins are the muscle myosins that form the thick filaments in a sarcomere of the muscle cell. Muscle contraction is facilitated by the rhythmic interaction between myosin heads and actin filaments (thin filaments) that result in relative movement of the thick and thin filaments. An early model to describe this is the sliding filament model. Investigations of muscle myosins date back to the 1880s, when the skeletal muscle protein was still known as 'muscle globulin' (16) or 'paramyosinogen' (17). The term 'myosin' has been in use at least since 1919 (18).

Unconventional myosins and their receptors are involved in diverse tasks such as in organelle translocation and cytoskeletal reorganization (myosin I[2] (20; 21; 22; 23; 24)), cytokinesis (nonmuscle myosin II), maintenance of cell integrity and structure and signaling (myosin III (25)), vesicle transport and membrane trafficking (myosin V[3] (27; 28)), particle transport and anchoring (myosin VI (29)), cell adhesion, hearing, and maintenance of balance (myosin VII), signaling (myosin IX), filopod extension (myosin X), and gliding motility (myosin XIV). A review of myosin functions and diseases due to myosin dysfunction in given in Ref. (30).

2.3 Movement strategies

Different motors can adopt different strategies on how to move along their filamentous tracks. Which strategy is used depends on the function of the motor.

Motor proteins can move along their tracks either processively or nonprocessivly. Processivity is defined as the average number of steps taken per diffusional encounter between a motor and its filament track (31). Another key feature characteristic of a specific motor is its directionality, *i.e.*, in which direction it moves along its track (32; 33).

Myosin V motors for instance are involved in intracellular trafficking having vesicles as cargo to be transported along the actin cytoskeleton. It has been shown that the myosin V dimer utilizes a mechanism in which one head of the myosin dimer remains attached to the track at all times. Two scenarios are consistent with this: a hand-over-hand mechanism in which the two heads alternate in the lead or an inchworm mechanism in which one head always leads. Using high-resolution fluorescence imaging techniques it was shown that myosin V walks hand-over-hand (34).

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 actin filaments. The "cargo" in this case is the myosin filament itself.

Given the large number of different motors and different movement strategies it is surprising

[2]Myosin I has been shown to produce its working stroke in two steps (19).
[3]The mechanochemical coupling in myosin V has very recently been studied (26).

that at least myosins and kinesins[4] seem to use the same hydrolysis mechanism in their active site to gain energy to be transformed in mechanical energy. This is suggested by crystal structure analysis of different motors that revealed that the motor domain of kinesin is structurally similar to the myosin motor domain (37).

[4]Reviews of the function of kinesins are given in Refs. (35; 36).

Chapter 3

The molecular motor myosin II

In this chapter, details on myosin motors with particular emphasis on conventional myosins are given. The sliding filament model as an early model of actomyosin interaction is introduced in Section 3.1, followed by a description of the structural features of myosin II (Section 3.2). Similarities between the active sites of myosins and other protein are outlined in Section 3.3, followed by descriptions of different conformational states of myosin (Section 3.4) and possible shortcomings of the sliding filament model (Section 3.5). Finally, transmission mechanisms between different sites in the myosin motor domain and the role of myosin II in the slime mold *Dictyostelium discoideum* are discussed in Sections 3.6 and 3.7.

3.1 The sliding filament model

All members of the myosin superfamily interact cyclically with actin, thereby moving past the actin filaments. In muscle, this leads to a sliding of the thick filaments (formed by myosin II) and thin filaments (formed by actin) relative to each other. Motion is initiated by major structural rearrangements in myosin resulting in a hinge-like bending of the myosin head. The necessary energy is provided by ATP hydrolysis. An early model to describe the cycle of events is the sliding filament model that assumes four different structural states of the actomyosin interaction (reviewed for example in Refs. (38; 39; 40)). A schematic representation of this model is shown in Fig. 3.1.

Myosin heads (also called cross-bridges) are detached from actin in the post-hydrolysis state (on the upper left in the figure) in which the reaction products ADP and P_i are bound. Weak actin binding results in formation of the pre-power-stroke state (on the upper right in the figure) and triggers product release. Along with product release, the power-stroke occurs in which myosin moves past actin. After the power-stroke the post-power-stroke

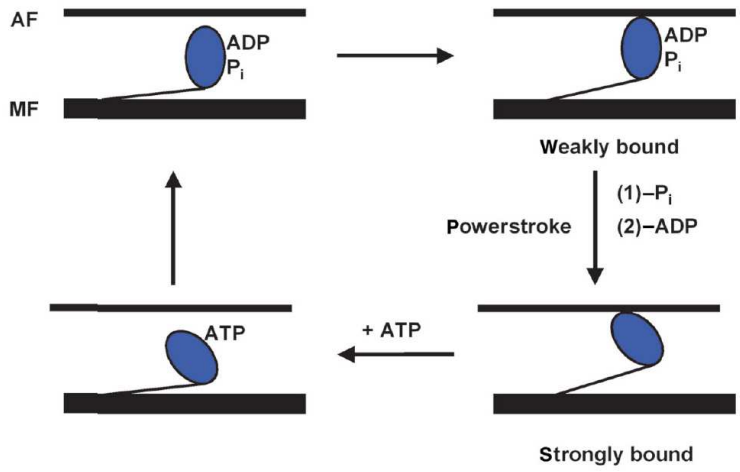

Figure 3.1: The sliding filament model of muscle contraction. The figure was taken from Ref. (41).

state is formed in which myosin strongly binds actin (on the lower right in the figure). This state is also called the rigor state. ATP binding dissociates the actomyosin complex, leading to the pre-recovery-stroke state (on the lower left). A large conformational change called the recovery-stroke reverses the conformational change of the power-stroke while myosin remains detached from actin. The recovery-stroke is coupled to the hydrolysis of ATP. Thus, the cycle is completed by reaching the post-hydrolysis state again.

3.2 Structural features

3.2.1 Global organization

A myosin II monomer consists of one heavy chain and two light chains. The non-muscle myosin II heavy chain of *Dictyostelium discoideum* consists of 2116 amino acids with a molecular weight of 243871 Da (42; 43).

Each myosin heavy chain can be split into one light meromyosin (LMM) and one heavy meromyosin (HMM). HMM can further be split into an N-terminal globular subfragment (S1, also called myosin head fragment MHF) and a rod-shaped subfragment (S2). The rodlike tail sequence is highly repetitive, showing cycles of a 28-residue repeat pattern composed of 4 heptapeptides, characteristic for alpha-helical coiled coils (44). S1 consists of a globular domain at the N-terminus from which a α-helical "tail" extends. The beginning of the tail is stabilized by two calmodulin-like light chains, namely the essential light chain (ELC) and the regulatory light chain (RLC).

The C-terminal α-helices of two heavy chains form a coiled-coil structure thus creating a homodimer. The homodimers self-assemble into filaments. By limited proteolysis, the N-terminal globular S1 domain can be broken into three fragments that are named after their apparent molecular masses: 25 kDa (N-terminal), 50 kDa (middle), and 20 kDa (C-terminal) (45).

The globular S1 domain hosts both the catalytic site and the actin binding region. It is therefore referred to as the catalytic or motor domain. A myosin construct consisting of the first 759 amino acids, *i.e.*, the motor domain alone, has been shown to be able to hydrolyze ATP and to displace along an actin filament (46).

3.2.2 Subdomain organization of the motor domain

The globular fragment of myosin II is visualized in Fig. 3.2. The N-terminal domain (25 kDa, residues 1 to 200) is flanked by the 50KDa domain that is split into the upper (residues 201 to 475) and lower (residues 476 to 613) 50 kDa domains. Between the upper and lower 50 kDa domains is a cleft that is supposed to be closed upon actin binding. At the tip of the cleft are the cardiomyopathy loop (HCM loop, belonging to the upper 50 kDa domain, Ile398-Val405) and the loop-2 (G519-G525 which may be extended to cover a broad surface domain from Gly519 to Lys546, including the conserved residues Phe535 and Pro536, belonging to the lower 50 kDa domain) that are known to play major roles in actin binding (47; 48; 49). In addition, a third loop termed the strut loop (Asp590-Gln593), connects the upper and lower 50 kDa domains. Loop-1 (belonging to the upper 50 kDa domain) is situated at the entrance to the nucleotide binding pocket which is located in a pocket between the upper 50 kDa domain and the N-terminal domain. The C-terminal 20 kDa fragment (residues 614 to 761) consists of a long α helix (residues 648 to 689), the SH1 helix (Val681 to Lys690), and a globular domain called the converter domain (residues 711 to 781) The α-helical neck region (also called light-chain binding domain or LCBD) provides the binding sites to which the essential (ELC) and regulatory (RLC) light chains are bound. The α-helical tail that is responsible for filament formation is not shown.

The rigid light-chain binding domain is believed to serve as a lever arm that amplifies the slight structural changes at the nucleotide binding pocket into large movements. It has been shown that the native LCBD can be replaced by an artificial domain of similar rigidity and dimensions without loss of functionality (50).

Conformational changes in the nucleotide binding pocket are tranduced to the reactive thiol region that got its name from two cysteine residues that can be chemically cross-linked. These are SH1 (Cys707 in chicken skeletal and Thr688 in*Dictyostelium discoideum* myosin II) and SH2 (Cys697 in chicken skeletal and Cys678 in *Dictyostelium discoideum* myosin II). Chemical modification of SH1 or SH2 results in significant alterations of the ATPase activity and actin binding affinity. The reactivity of the thiol groups can also be used to introduce spin labels as has been done to measure EPR spectra (51; 52).

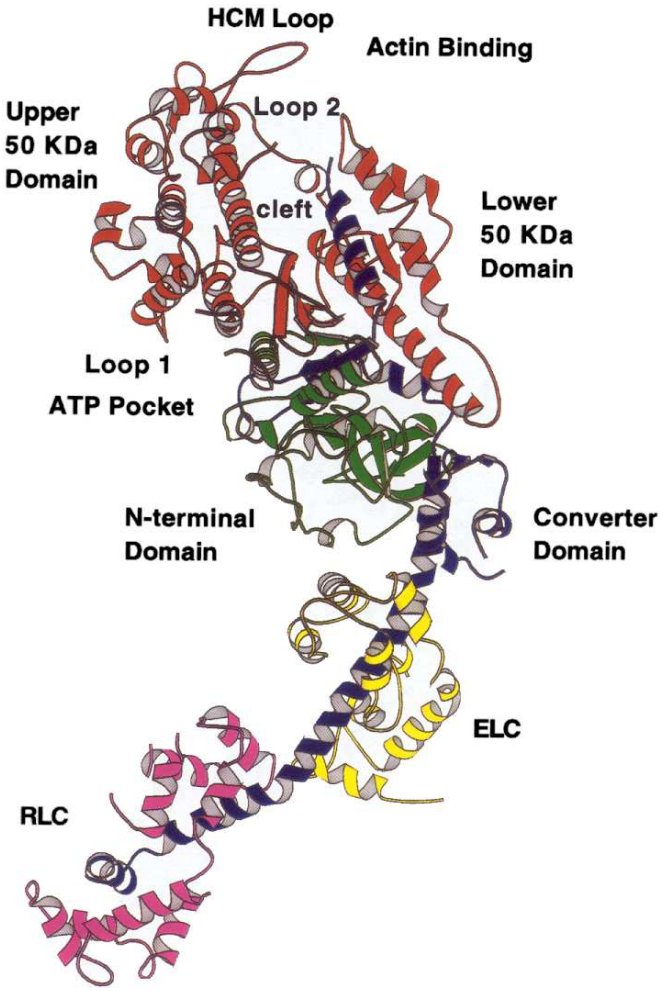

Figure 3.2: Three-dimensional structure of chicken fast skeletal muscle myosin S1. The figure is taken from Ref. (14)

3.2.3 Nucleotide binding site

The nucleotide binding site is located at the interface between the 50 kDa and the N-terminal subdomains (Fig. 3.3). 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 (53), the switch-1 loop and the switch-2 loop. Together they form the so-called phosphate tube. 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 conformation, thus serving as a switch for information transduction by changing their conformation.

The positions of the switch-1 and switch-2 loops can be used to classify different conformational state of myosin. Conformations corresponding to switch-1 closed/switch-2 closed (C/C), switch-1 closed/switch-2 open (C/O), and switch-1 open/switch-2 open (O/O) have been identified by crystallography (see Section 3.4). Binding of ATP induces a conformational change towards the C/C state of the nucleotide binding pocket. This state is 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 (54).

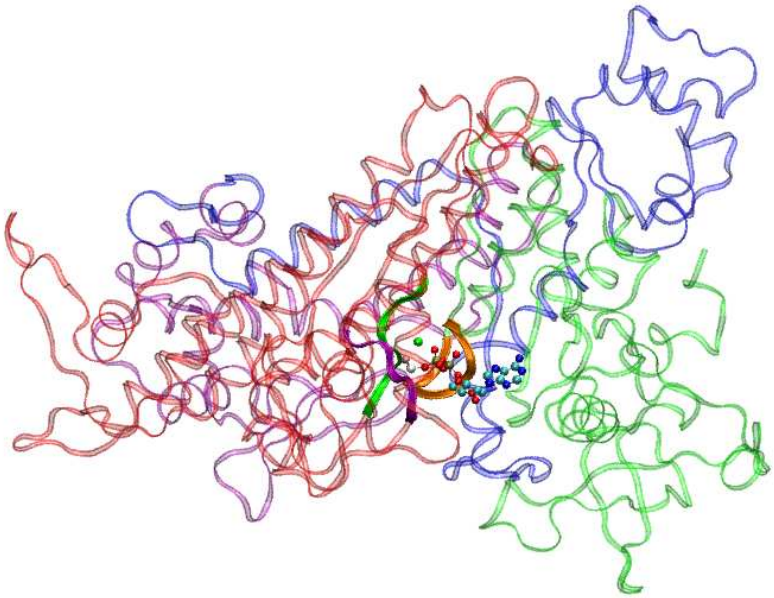

Figure 3.3: Localization of the nucleotide binding pocket in the catalytic domain of *Dictyostelium discoideum* myosin II. The figure was prepared from the ref1 structure (55). The N-terminal 25 kDa domain (1-200) is shown in light green, the upper 50 kDa (201-475) domain in light red, the lower 50 kDa domain (476-613) in light purple, the C-terminal 20 kDa domain (614-759, including the converter domain) in light blue, the P-loop in orange, the switch-1 loop in purple, the switch-2 loop in green, and Mg.ADP.BeF_3 in CPK representation.

The consensus sequences of the three loops in the myosin superfamily and their specific sequences and corresponding residue numbers in *Dictyostelium discoideum* myosin II are shown in Table 3.1.

A detailed view of the nucleotide binding pocket is shown in Fig. 3.4. As can be seen in

loop	consensus	*Dictyostelium*	residues
P-loop	GESGAGKT	GESGAGKT	179-186
switch-1	NxNSSR	NNNSSR	233-238
switch-2	DxSGFE	DISGFE	454-459

Table 3.1: Consensus sequences conserved in at least 80 out of 82 myosins (56) and *Dictyostelium discoideum* myosin II sequences including residue numbers for the phosphate-tube forming loops.

the figure, there is a salt bridge between Arg238 (switch-1) and Glu459 (switch-2). This salt bridge is formed if the nucleotide pocket is in its C/C conformation, in which both switch loops are located close to each other and in close spatial proximity to the P-loop. If the switch-2 loop swings away, the nucleotide pocket adopts its C/O conformation. Fig. 3.5 shows an overlap of the C/O and C/C nucleotide pockets as have been observed in crystallographic studies (55; 57). Both crystals have been prepared from Mg.ADP.BeF_x complexed to *Dictyostelium discoideum* myosin II. It is obvious from the figure that the positions of the P-loop, switch-1 loop, nucleotide and Mg^{2+} (including its coordination sphere) are identical in both structures, whereas the switch-2 loop has moved by about 4 Å thus breaking the salt bridge in the C/O structure.

Nucleotide binding naturally occurs from the solvent-exposed side of the nucleotide binding pocket which is therefore termed the front-door. As long as a nucleotide remains bound this front door is locked. It has been shown, however, that after hydrolysis phosphate release occurs prior to ADP release (58; 59). Thus, it was postulated that P_i release occurs *via* a different pathway than nucleotide binding and that the switch-1 and switch-2 loops must change their conformations thereby opening a “back” door or a “trap” door (see Section 3.6).

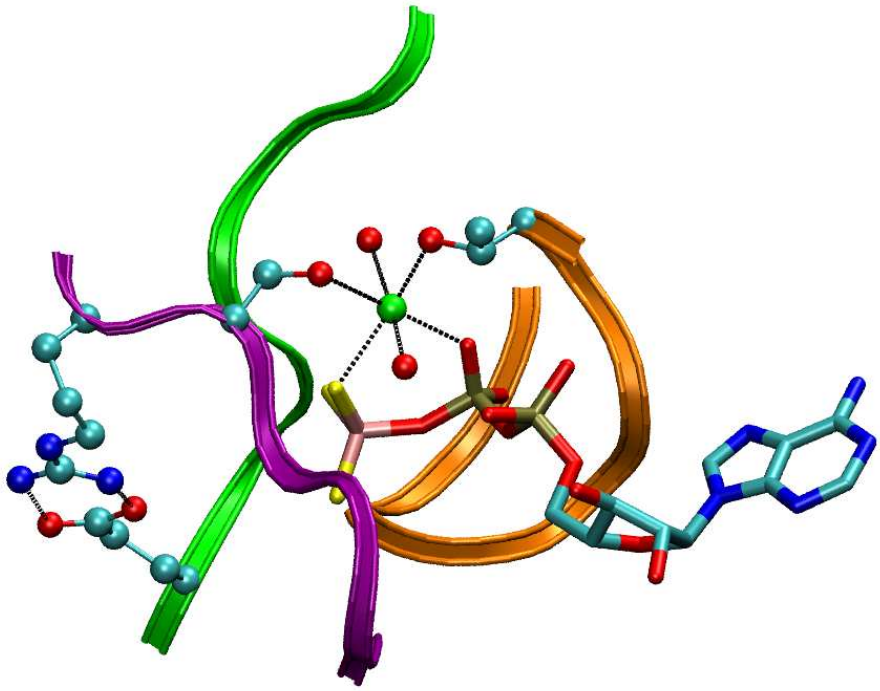

Figure 3.4: Detailed view of the nucleotide binding pocket of *Dictyostelium discoideum* myosin II. The figure was prepared from the ref1 structure (55). The P-loop is shown in orange, the switch-1 loop in purple, the switch-2 loop in green, ADP.BeF_3 in bonds representation, Mg^{2+} and its coordination sphere consisting of an oxygen from ADP, a fluorine atom from BeF_3, two crystallographic water molecules and the side chain oxygens from Thr186 (P-loop) and Ser237(switch-1) in CPK representation, and the salt-bridge between Arg238 (switch-1) and Glu459 (switch-2) in light CPK representation

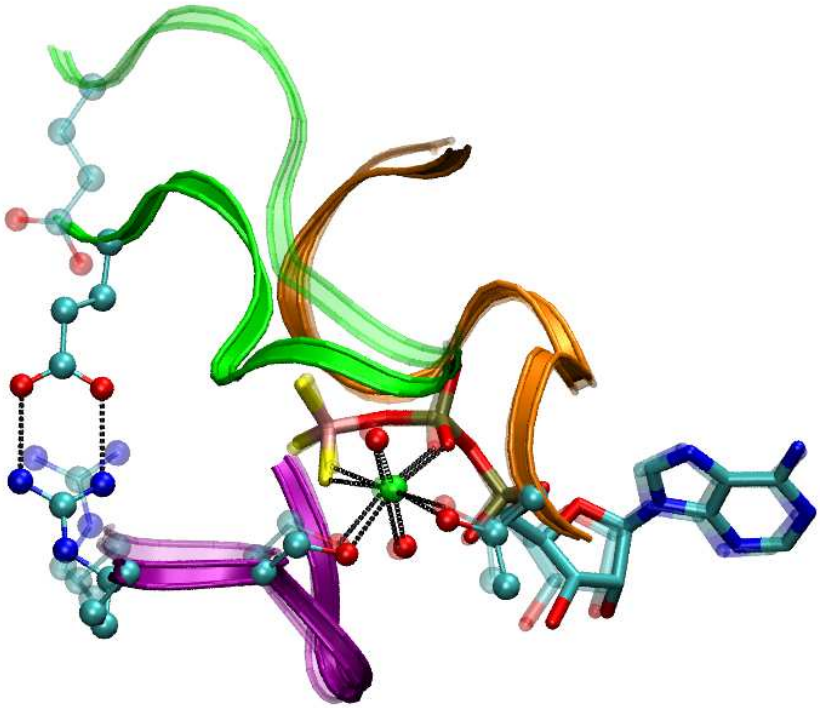

Figure 3.5: Overlap of the closed (opaque colors, ref1 (55)) and open (transparent colors, prepared from PDB code 1MMD (57)) forms of the nucleotide binding pocket of *Dictyostelium discoideum* myosin II. The P-loop is shown in orange, the switch-1 loop in purple, the switch-2 loop in green, ADP.BeF_3 in bonds representation, Mg^{2+} and its coordination sphere consisting of an oxygen from ADP, a fluorine atom from BeF_3, two crystallographic water molecules and the side chain oxygens from Thr186 (P-loop) and Ser237(switch-1) in CPK representation, and the salt-bridge between Arg238 (switch-1) and Glu459 (switch-2) in CPK representation

3.3 Nucleotide binding pocket similarities between myosin and other proteins

Myosin shares its structural elements P-loop, switch-1 and switch-2 with a number of other ATP or GTP hydrolyzing proteins. Apart from other motor proteins such as the kinesins or F_1-ATPase these include the members of the G-protein family (such as Ras, EF-Tu, or transducin-α) (60). Because of the apparent close relationship between these different families of proteins it has been postulated that they evolved from a common ancestor (61).

A comparison of the three-dimensional structure of the active site between myosin and other P-loop proteins such as the GTPase Ras has highlighted close similarities (53). Based on this observation is has been suggested that at least myosin and G-proteins utilize a similar mechanism for nucleotide hydrolysis.

3.4 Myosin conformers

Myosin can adopt different conformational states some of which have been trapped in crystallographic studies. Since the first crystal structure of the motor domain of myosin II has been published in 1993 (62) a large number of X-ray crystallographic structures have been reported. Table 3.2 gives an overview of all currently available crystal structures of myosin II from the organism *Dictyostelium discoideum*. In particular, the switch-1 and switch-2 loops of the nucleotide pocket can be either closed or open. An additional actin-detached structural state has been identified for scallop myosin subfragment 1 in which the SH1 helix is unwound (63; 64; 65).

During the power stroke the converter domain undergoes a conformational change that results in a $\approx 60^\circ$ rotation of the light chain binding domain (LCBD). Motions of the latter have been resolved using EPR spectroscopic techniques with spin-labeled LCBD (66)

Electron cryomicrosopic studies on the actomyosin complex have indicated that none of the crystallographically resolved conformational states of myosin fit into the electron micrograph, thus suggesting the existence of an additional unique actin-bound conformation (67). Such a conformation has been shown to exist by an independent study that in addition suggests that the myosin structure in the actomyosin complex differs in presence and absence of Mg.ADP (68). X-ray fiber diffraction experiments showed a continuous change of the bending angle between the light-chain binding domain and the catalytic domain during the power stroke in intact muscle fibers (69), thereby suggesting the existence of a large number of conformational intermediates during force generation.

Myosin complexed to a specific nucleotide can adopt different conformations as evidenced by the fact that Myosin complexed to ADP.BeF_x has been crystallized in two distinct conformers (55; 57). This is in agreement with previous investigations that have postulated a

PDB ID	resolution [Å]	construct[a]	ligand	state[b]	Ref.
1D0X	2.00	759	m-nitrophenyl aminoethyl.PP_i.BeF_3	C/O	(70)
1D0Y	2.00	759	o-nitrophenyl aminoethyl.PP_i.BeF_3	C/O	(70)
1D0Z	2.00	759	p-nitrophenyl aminoethyl.PP_i.BeF_3	C/O	(70)
1D1A	2.00	759	o,p-dinitrophenyl aminoethyl.PP_i.BeF_3	C/O	(70)
1D1B	2.00	759	o,p-dinitrophenyl aminopropyl.PP_i.BeF_3	C/O	(70)
1D1C	2.30	759	N-methyl-o-nitrophenyl aminoethyl.PP_i.BeF_3	C/O	(70)
1FMV	2.10	759	-	C/O	(71)
1FMW	2.15	759	Mg.ATP	C/O	(71)
1G8X	2.80	761.α-actinin R238E	Mg.ADP	C/O	(72)
1LVK	1.90	759.Q760L, R761P, I762N	Mg.BeF_3.*mant*ADP	C/O	(73)
1MMA	2.10	759.Q760L, R761P, I762N	Mg.ADP	C/O	(74)
1MMD	2.00	759.Q760L, R761P, I762N	Mg.ADP.BeF_3	C/O	(57)
1MMG	1.90	759.Q760L, R761P, I762N	Mg.ATPγS	C/O	(74)
1MMN	2.10	759.Q760L, R761P, I762N	Mg.AMPPNP	C/O	(74)
1MND	2.60	690.Q760L, R761P, I762N	Mg.ADP.AlF_4	C/C	(57)
1MNE	2.70	759.Q760L, R761P, I762N	Mg.PP_i	C/O	(75)
1Q5G	1.90	myosin II fused to dynamin1	-	O/O	(76)
1VOM	1.90	747	Mg.ADP.VO_4	C/C	(77)
REF1		759	Mg.ADP.BeF_3	C/C	(55)

[a]The last resolved amino acid is given. Mutations are indicated.

[b]The conformational states of switch-1/switch-2 are indicated. "O" refers to the open state, whereas "C" refers to the closed state.

Table 3.2: Available crystal structures for *Dictyostelium discoideum* myosin II.

three-step binding process of nucleotides binding to myosin (78; 79). According to these studies, a binding event would be initiated by the formation of a collision complex in which myosin remains in its apo structure, quickly followed by a first conformational change leading to a state that is predominant in M.ADP and terminated by a second conformational change to a state that is predominant in M.AMPPNP and, presumably, in M.ATP (80; 81).

A detailed comparison of the conformations of the nucleotide binding site for 18 different *Dictyostelium discoideum* structures is given in Appendix A.

3.5 Myosin in action

The sliding filament model (see Section 3.1) implies that tiny changes in the myosin head are amplified and result in a large motion of the adjoining neck domain that thus serves as a rigid lever arm. Consequently, the longer the lever arm the larger is the step size taken.

The step-size that muscle myosin II takes is dependent on the load present. In intact muscle fibers the step size has been measured to be between 8 and 13 nm in each interaction of myosin with actin (82).

That enzymatic and mechanical events are indeed tightly coupled in myosin II *in vivo* has been shown by simultaneous measurements of RLC rotations (to monitor the mechanical action) and ADP release (to monitor enzymatic action) in intact muscle fibers (83). It is possible to chemically decouple mechanical and enzymatic events by trinitrophenylation of the reactive Lys84 at the interface between motor and neck domains (84; 85). *In vitro* investigation, however, have suggested that each ATP hydrolysis event may be followed by multiple stepping movements, a scenario that has been described assuming a transiently partially unfolded actomyosin complex (86). Such a partial unfolding mechanism is, however, challenging the classical lever-arm model. Observations that unconventional myosins with short neck domains nevertheless can take large steps also suggest a break-down of the lever-arm theory (for a discussion on this issue see (87)). It is, however, possible that different myosins use different mechanisms with some a employing a lever-arm-type mechanisms and others employing a transient-unfolding mechanism.

3.6 Transmission of information between motor sites

Events at different sites of the myosin motor domain are tightly coupled (reviewed in (39)). Nucleotide binding for instance occurs at the active site in the center of the catalytic domain. However, it induces conformational transitions that are propagated to the actin binding region, resulting in loss of affinity for actin and dissociation of the actomyosin complex. Reversely, posthydrolytic phosphate release is accompanied by an increase in actin affinity. Thus, information is transmitted between different functional units of the

motor domain. In particular, events in the catalytic site must be coupled to events at the actin binding site, thus modulating myosin's affinity for actin, and to events at the converter domain thereby triggering the conformational change leading to the power stroke.

Nucleotide binding. Nucleotide binding has been postulated to occur *via* a front-door mechanism as revealed by a crystal structure in which the nucleotide is partially bound (88). Usually, product release occurs *via* the same route as substrate binding. In myosin, however, P_i release has been shown to occur *prior* to ADP release, with ADP blocking the front door. Thus, phosphate release must occur *via* a different route than substrate binding.

Coupling between nucleotide binding site and converter domain. Nucleotide binding affects the conformation and dynamics of the SH1-SH2 helix (89) that is in close spatial proximity to the converter domain. This influences the distance between the two reactive thiol groups (see (90) and references therein). Based on time-resolved fluorescence measurements on muscle fibers Gly681[1] has been proposed as a pivot for the tertiary structural changes (91).

The available crystal structures indicate that movement of switch-2 triggers a long-range conformational change that leads to a change in the angle between the motor domain and the lever arm. Thus, if switch-2 is in its OPEN state, the lever arm is in its DOWN position and *vice versa*.

A detailed mechanism how the conformational changes at the catalytic site are coupled to conformational changes of the converter domain has recently been proposed based on computational investigations. The the structural elements relay helix (Phe466 to Lys498), relay-loop (Ile499 to Asp509), and SH1-helix (Val681 to Lys690) have been found to play significant roles as mediators (92).

Coupling between nucleotide and actin binding sites. Binding of nucleotide and of actin to myosin is antagonistic: ATP binding to actomyosin causes actin dissociation and actin binding to M.ADP.P_i accelerates P_i and ADP release. It has been shown that the 50 kDa cleft undergoes structural changes upon actin binding (93). A possible rear opening of the phosphate tube which is located at the apex of the 50 kDa cleft has been suggested as a possible "back door" for phosphate release (94). Thus, P_i would be released into the 50 kDa cleft and thereby affect actin affinity. This has been discussed in detail based on molecular dynamics simulations (94).

An alternative view is suggested by a recent nucleotide-free crystal structure of *Dictyostelium discoideum* myosin II in which both the switch-1 and switch-2 loops have moved away from their positions in the CLOSED structure (76). This structure is therefore termed an open/open or O/O structure. The opening of switch-1 suggests that phosphate release can also occur *via* a "trap door" rather than the previously suggested back door. The O/O structure also reveals significant changes in the actin binding region. The 50 kDa cleft is

[1] Gly699 in chicken pectoralis muscle myosin that has been used in the cited experiments.

closed in the O/O structure, thereby suggesting that switch-1 opening and cleft closure are mechanically linked. This linkage has also been suggested based on fitting crystallographic structural information into high-resolution electron micro graphs of actomyosin (95) and by a crystallographic study on myosin V (96).

A schematic summary of the structural states observed in different myosins together with their corresponding positions of switch-1, switch-2, and the lever arm as well as their corresponding actin affinity is shown in Fig. 3.6. How these structural states can be mapped along the ATPase cycle of actomyosin is shown in Fig. 3.7.

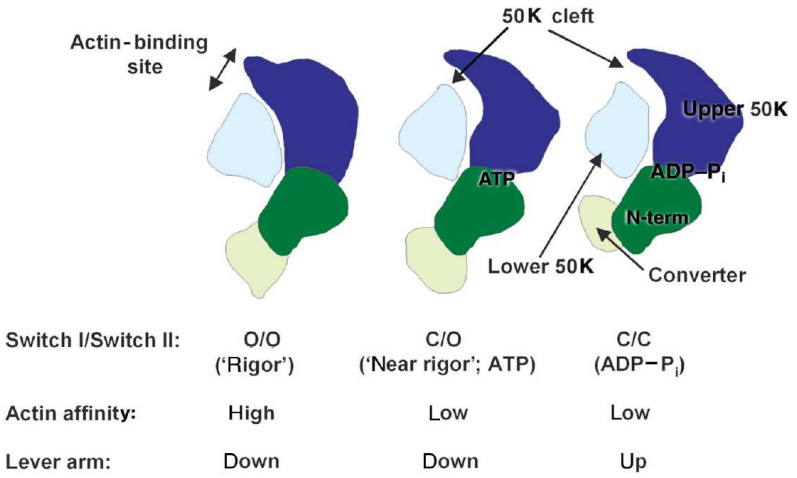

Figure 3.6: Proposed subdomain rearrangements in the myosin motor domain and corresponding switch-1, switch-2, and the lever arm positions as well as actin affinity. The figure was taken from Ref. (41).

3.7 Myosin II in *Dictyostelium discoideum*

Dictyostelium discoideum is a slime mold that grows as a unicellular amoebae. Upon starvation cells interact and by releasing the chemoattractant cAMP and subsequent chemotaxis form a multicellular mound consisting of up to 100 000 cells. Because of its unique features combining unicellular and multicellular characteristics the NIH[2] chose *Dictyostelium discoideum* as a model organism for functional analysis of sequenced genes (97). Specifically, processes such as cytokinesis, motility, phagocytosis, chemotaxis, signal transduction (98), and aspects of development such as cell sorting, pattern formation, and cell-type determination can easily be studied (99). In addition, the localization of specific proteins in the

[2]The NIH (National Institute of Health) was the main organizer of the Human Genome Project and now coordinates the follow-up projects.

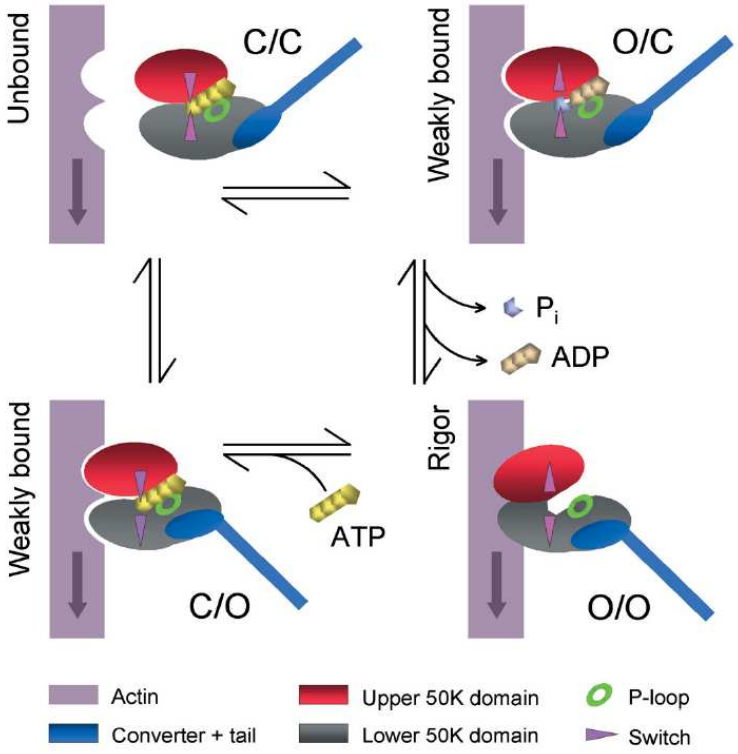

Figure 3.7: Structural model for the actin-activated myosin II ATPase cycle. The figure was taken from Ref. (76).

living cell at different stages off the cell cycle can be monitored using fluorescence confocal microscopy (100; 101).

Both the motion of single *Dictyostelium discoideum* cells during chemotaxis and the subsequent collective motion of the mound require the interaction between a conventional myosin and the actin cytoskeleton (102). Thus, this organism has been used extensively to study the functionality of myosins (103).

Chapter 4

Kinetics and thermodynamics of ATP hydrolysis

A detailed understanding of ATP hydrolysis both in aqueous solution (uncatalyzed reaction) and in myosin (catalyzed reaction) is necessary to explain myosin's enzymatic function. This chapter summarizes the findings of experimental and computational investigations of the thermodynamics and kinetics of ATP and ATP model compound hydrolysis both in solution and in myosin.

4.1 Thermodynamics of ATP hydrolysis in aqueous solution

The energy available to be utilized due to ATP hydrolysis is given by the reaction free energy of ATP hydrolysis in ionic solution under physiological conditions. Already 35 years ago a detailed thermodynamics study has been reported that gives thermodynamic properties of ATP hydrolysis as a function of pH,[1] temperature T, and ionic strength I (104). The usual thermodynamic quantities Gibbs free energy G, enthalpy H, and entropy S as used in chemistry are functions temperature and pressure. To introduce the dependence on pH, pMg,[2] and I into the thermodynamics description as is necessary for a thermodynamic description of *bio*chemical reactions, transformed thermodynamic properties G', H', and S' were introduced (105; 106; 107; 108) and used to describe the hydrolysis of ATP in Ref. (104). This early work has been expanded to yield a theoretical description of all relevant thermodynamic properties of ATP hydrolysis in ionic solution (109; 110) including a study

[1] pH is the negative decadic logarithm of the concentration of H_3O^+ ions.
[2] pMg is the negative decadic logarithm of the concentration of Mg^{2+} ions.

on the change in the binding of hydrogen ions and magnesium ions in the hydrolysis of ATP (111).

A summary of transformed reaction free energies, transformed reaction enthalpies, and transformed reaction entropies at different pH and pMg is given in Table 4.1. Additional thermodynamic data on ATP, ADP, AMP, P_i, adenosine, and adenine can be found in Refs. (112; 113; 114; 115; 116; 117; 118; 119; 120); on NMR measurements on ATP analogues in (121); on NMR measurements of ATP and ADP in (122; 123); on measurements of the MgATP dissociation constant (124; 125; 126); on the acidity of the hydroxyls in the sugar moiety in (127); and on considerations of cation-triphosphate interactions in (128).

property	pMg	pH=7	pH=9
$\Delta_R G'^0$ [J mol^{-1}]([kcal mol^{-1}])	2	-30801 (-7.41)	-41475 (-10.04)
	4	-34998 (-8.37)	-45443 (-10.99)
	6	-36022 (-8.84)	-46684 (-11.23)
$\Delta_R H'^0$ [J mol^{-1}]([kcal mol^{-1}])	2	-26422 (-6.54)	-24410 (-5.97)
	4	-28204 (-6.93)	27508 (-6.69)
	6	-23164 (-5.74)	-21468 (-5.26)
$\Delta_R S'^0$ [J K^{-1} mol^{-1}]([cal K^{-1} mol^{-1}])	2	14.69 (3.51)	57.24 (13.68)
	4	22.79 (5.45)	60.15 (14.38)
	6	43.12 (10.31)	84.58 (20.21)

Table 4.1: Thermodynamic properties for ATP hydrolysis at a temperature of T=298.15K, an ionic strength of I=0.25M, pH values of 7 and 9, and pMg values of 2, 4, and 6. The data are taken from Ref. (109).

The biochemical reaction of ATP hydrolysis is not a chemical reaction in which a single reactant species is transformed into a single product species. Rather, the reactants involved consist of sums of species. In the presence of Mg^{2+} ions ATP can for instance exist as ATP^{4-}, $HATP^{3-}$, $MgATP^{2-}$, or $HMgATP^{-}$. The amount of species present at a given temperature, pH, pMg, and ionic strength depends on the dissociation constants that correspond to the interconversion reactions between species. The necessary equilibrium constants for ATP, ADP, and P_i are listed in Table 4.2. Species are omitted that are not likely to be present if the pH is outside the range 5 to 9 and the pMg is outside the range 2 to 6. The Mg ion binds to the triphosphate moiety, as indicated by ^{1}H NMR experiments (129).[3]

Kinetic measurements on ATP hydrolysis in aqueous solution at 95°C at varying pH gave reaction rates of $3.2 \cdot 10^{-5}$ s^{-1} (pH=8.4, *i.e.*, the predominant species is ATP^{4-}) and $17.5 \cdot 10^{-5}$ s^{-1} (pH=4 to 5, *i.e.*, the predominant species is $HATP^{3-}$) (59; 132). This

[3]This study also indicates that the adenine moiety adopts an *anti* conformation with respect to the ribose ring in aqueous solution.

reaction	pK(I=0)	pK (I=0.25M)
$HADP^{2-} = H^+ + ADP^{3-}$	7.18	6.33
$H_2ADP^- = H^+ + HADP^{2-}$	4.36	3.79
$MgADP^- = Mg^{2+} + ADP^{3-}$	4.65	2.95
$MgHADP = Mg^{2+} + HADP^{2-}$	2.50	1.37
$HATP^{3-a} = H^+ + ATP^{4-}$	7.60	6.47
$H_2ATP^{2-b} = H^+ + HATP^{3-}$	4.68	3.83
$MgATP^{2-} = Mg^{2+} + ATP^{4-}$	6.18	3.91
$MgHATP^- = Mg^{2+} + HATP^{3-}$	3.63	1.93
$Mg_2ATP = Mg^{2+} + MgATP^{2-}$	2.69	1.75
$H_2PO_4^- = H^+ + HPO_4^{2-}$	7.22	6.65
$MgHPO_4 = Mg^{2+} + HPO_4^{2-}$	2.71	1.58

[a]The proton sits on the terminal γ-phosphate group.
[b]The protons sit on the terminal γ-phosphate group and the adenine moiety.

Table 4.2: Equilibrium constants at T=298.15K given as pK values for ATP, ADP, and P_i species that may be present in the pH range of 5 to 9 and the pMg range from 2 to 6. Data are from Refs. (130) and (131).

corresponds to activation free energies of 122.44 kJ/mol = 29.3 kcal/mol and 117.24 kJ/mol = 28.0 kcal/mol (see also Section 4.2.3). At room temperature (25°C) this would correspond to rate constants of $2.1 \cdot 10^{-9} s^{-1}$ and $1.7 \cdot 10^{-8} s^{-1}$ or half lives of about one to ten years.

4.2 Mechanistic considerations on phosphate hydrolysis in aqueous solution

4.2.1 Dissociative *versus* associative mechanism

Definition

Phosphate hydrolysis is initiated by nucleophilic attack of a water molecule or a hydroxide ion. Nucleophilic substitutions at the central phosphorus atom can occur *via* a continuum of mechanisms with two limiting cases (133):

1. **Dissociative mechanism.** A fully dissociative mechanism is characterized by a trigonal metaphosphate moiety as an intermediate. The reaction proceeds in two steps. In the first step, the leaving group dissociates from the phosphate moiety. In the second step the attacking nucleophile reacts with the metaphosphate intermediate

yielding the reaction products. The distances between the phosphorus atom and the leaving and attacking group oxygens in the intermediate are larger than the sum of the van der Waals radii of P and O, *i.e.*, $d(P:O) > r_{vdW}(P) + r_{vdW}(O) = 1.9\mathring{A} + 1.4\mathring{A} = 3.3\mathring{A}$.

2. **Associative mechanism.** A fully associative mechanism is characterized by a pentavalent phosphorane intermediate with trigonal bipyramidal structure. The distances between the phosphorus atom and the leaving and attacking group oxygens in the intermediate are equivalent to the bond length of a P-O single bond, *i.e.*, $d(P:O) = 1.73\mathring{A}$.

Mechanisms that are partially associative (dissociative) have axial P-O bond lengths between 1.73 and 3.3 Å. Dissociative and associative mechanisms differ in the distance between the phosphorus and the incoming oxygen. This distance is therefore a good measure to quantify the degree of associativity. Using Pauling's relationship between bond length $D(n)$ and bond order n (pp. 255-260 in (134)),

$$D(n) = D(1) - 0.60\log(n), \tag{4.1}$$

where $D(1)$ is the single bond distance, the fractional associativity can be defined as being identical to the axial bond order n to the entering group.

A classical S_N2 mechanism is characterized by a transition state in which both entering and leaving groups have bond orders of 0.5. Thus, such a mechanism has a fractional associativity of 50%, and the corresponding bond length is $1.73\mathring{A} - 0.60\log(0.5)\mathring{A} = 1.91\mathring{A}$. Fig. 4.1 shows a summarizing illustration of dissociative, S_N2, and associative mechanisms.

Possibilities to distinguish dissociative and associative mechanisms

Apart from the geometric criteria that have been used to define the associativity of reaction there are several additional possibilities to distinguish dissociative from associative mechanisms:

1. **Linear free energy relationships.** Linear free energy relationships between the pK_a values of different incoming nucleophile and the logarithm of the rate constant (135) can be determined. If the mechanism is dissociative, the rate limiting step (formation of the metaphosphate intermediate) is independent of the pK_a of the attacking nucleophile, whereas if the mechanism is associative, the rate limiting step (formation of the phosphorane intermediate, this requires deprotonation of the nucleophile) does depend on the pK_a of the attacking nucleophile. In contrast, the pK_a of the leaving group influences the rate of transition state formation in a dissociative mechanism

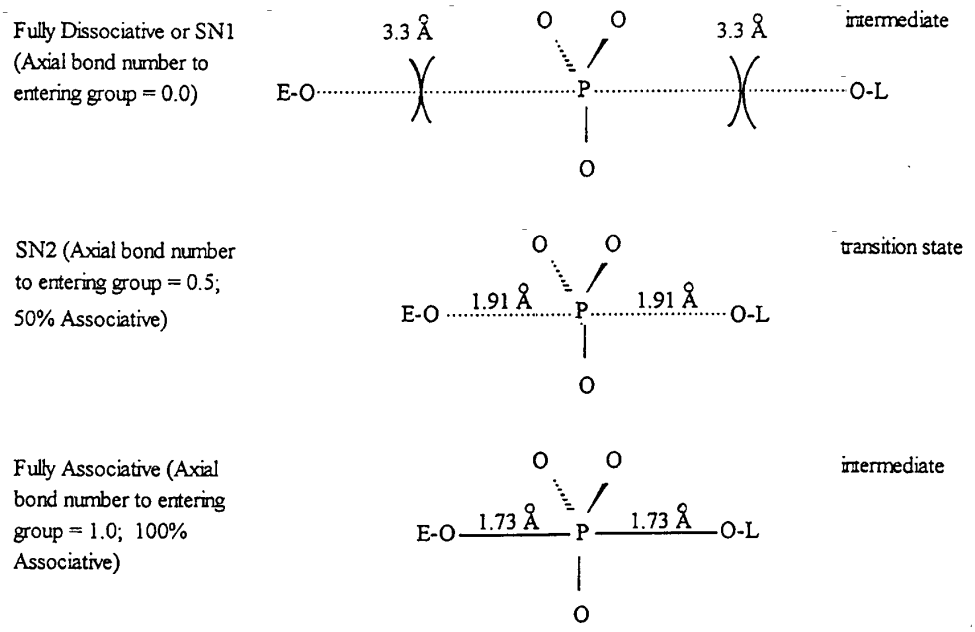

Figure 4.1: Mechanisms of nucleophilic substitution at phosphorus. The distances between the phosphorus atom and the oxygen atom of the entering (E) and the leaving (L) groups in the intermediate (transition) states are given. The figure was taken from Ref. (133).

but not in an associative mechanism. However, the unique interpretability of linear free energy relationships has been questioned (136).

2. **Charge on the non-bridging oxygens.** In a dissociative mechanism, the leaving group dissociates first. If the reacting species is dianionic (as is usually the case in phosphoanhydride cleavage) the leaving group is anionic, leaving only a single negative charge on the metaphosphate intermediate. Thus, the partial atomic charges on the nonbridging oxygens are reduced in magnitude in the transition state as compared to the reactant state. In an associative mechanism, however, a negatively charge nucleophile enters and introduces an additional charge into the pentavalent phosphorane intermediate. Thus, the partial atomic charges on the nonbridging oxygens are increased in magnitude in the transition state as compared to the reactant state.

3. **Charge on the ester or anhydride oxygen.** The charge on the oxygen that connects the leaving group with the phosphorus atom ("bridging oxygen") will become more negative in a dissociative mechanism, in which the anionic leaving group has already left in the transition state, whereas is will remain largely unchanged in an associative mechanism in which the leaving group remains attached in the transition

state.

4. **Isotope labeling.** Experimentally, dissociative and associative mechanisms can be distinguished by stereochemical studies using labeled oxygen species. A dissociative mechanism will lead to a racemization of configuration, whereas an associative mechanism leads to the inversion of configuration (137).

5. **Kinetic isotope effect (KIE).** A KIE is defined as the ratio of reaction rates with unlabeled reactants and reactants labeled with a heavy isotope. A primary KIE refers to labeling of an atom directly involved in bond making or breaking (*i.e.*, the bridge oxygen in phosphate esters), whereas a secondary KIE refers to labeling of neighboring atoms (*i.e.*, the nonbridging oxygens in phosphate esters). KIE arise from the effect of isotopic substitution on the zero point energy levels of the vibrational modes of the reactant as it proceeds from the ground state to the transition state. A decrease in bond order at the transition state results in a normal KIE (>1), whereas an increase in bond order causes an inverse KIE (<1). A normal primary KIE implies that P-O bond cleavage is at least partly rate-limiting and thus suggests a dissociative mechanism. A normal secondary isotope effect implies that the bond order decreases in going form the reactant to the transition state and thus suggests an associative mechanism. An inverse secondary isotope effect, however, implies an increase in bond order as is expected for a dissociative mechanism (137).

Either mechanism involves the reorganization of negative charges on the solute atoms (and thus of the solvent structure surrounding the solute) in going from the reactant state *via* the transition or intermediate state(s) to the product state. Thus, solvation effects are expected to play a significant role in both cases. That solvation indeed plays a major role in ATP hydrolysis in aqueous solution has already been recognized 35 years ago (138).

4.2.2 Acid and base catalysis

The pH of the solution in which the phosphate hydrolysis occurs, affects the protonation states of both the incoming nucleophile and the phosphate ester or anhydride attacked. If the reaction proceeds at high pH it is likely that a hydroxyl ion will attack rather than a water molecule. Because OH^- is a stronger nucleophile than H_2O this may speed up the reaction. However, at high pH values the attacked phosphate may also be deprotonated and negatively charged and thus the resulting charge-charge repulsion slows down the reaction. At low pH values the attacked phosphate is more likely to be neutral and can therefore be more easily attacked by an incoming nucleophile.

Depending on the pK_a values of the nucleophile and the phosphate the pH value at which the reaction is fastest may vary. The hydrolysis of monoesters, for example, is fastest at a pH value of 4 where the major species present is the monoprotonated monoanion (2).

The protonation state may not only influence the magnitude of electrostatic repulsion effects thus altering the reaction rate even if the reaction mechanism remains unchanged but it may influence the reaction mechanism itself. Differently protonated species of the same phosphate may thus be hydrolyzed *via* different reaction routes with correspondingly different reaction barriers and reaction rates.

In biological environments the pH is nearly neutral. However, in an enzymatic environment the protein surrounding may tune the local pK_a value of the phosphate substrate thus creating a protonation pattern and local reaction conditions under which the hydrolysis reaction is favored.

The enzyme thus may tune the reaction rate by using two different strategies: Firstly, the enzyme may create a local protonation pattern that provokes a reaction route with lowest activation barrier and secondly, the enzyme may further lower this activation barrier. To describe the details of a specific enzymatic phosphate hydrolysis reaction it is therefore necessary to know both the protonation states of the nucleophile and the substrate and the mechanistic details of the hydrolysis reaction in the given protonation state as it would occur without the protein environment.

4.2.3 Studies on model compounds

Several studies have been reported that aim at gaining insight into the mechanistic details of phosphate hydrolysis reactions at different protonation states of the attacking nucleophile and the phosphate. This section summarizes such studies.

Hydrolysis of phosphate esters

Phosphate esters serve as model compounds for the hydrolyses of RNA and DNA. As such they have received considerable interest. Here, only investigations on methylphosphate hydrolysis are discussed since methylphosphate also serves as a model compound for phosphoanhydride hydrolysis. Literature is available for other phosphate esters such as ethylene phosphates (139; 140; 141; 142; 143; 144) or for phosphate aryl ester hydrolysis (145; 146; 147; 148; 149). However, because of the limited possibility to transfer mechanistic insight from hydrolyses of such phosphate esters to phosphate anhydride hydrolysis the corresponding studies are not further considered here.

Methylphosphate hydrolysis The maximum rate for methyl phosphate hydrolysis has measured to be $8 \cdot 10^{-6} s^{-1}$ at 373 K and pH = 4.2 (150). This corresponds to an apparent free energy activation barrier of 128.45 kJ/mol or 30.7 kcal/mol. At pH$\approx$ 4 the predominant species is anionic methylphosphate ($MeHPO_4^-$). In another study, the enthalpic barrier was found to be 37.2 kcal/mol and the activation entropy 3.7 cal/mol/K (151).

Dissociative mechanism. A computational investigation of a dissociative mechanism for anionic methylphosphate hydrolysis employing HF/6-31+G(d), B3LYP/6-31+G(d), and MP2/6-31+G(d) calculations has been reported (145). Only the formation of the metaphosphate intermediate was studied. A unimolecular dissociation involving an intramolecular proton transfer from the OH group to the bridging oxygen proceeded with a free energy activation barrier of 46.50, 29.76, and 31.82 kcal/mol, respectively, at the three different levels of theory. Introduction of a water molecule that mediates the proton transfer lowers these barriers to 39.82, 20.10, and 19.74 kcal/mol, respectively.

Different authors report a B3LYP/6-31G(d) free energy activation barrier of 30.7 kcal/mol in gas phase and 34.5 kcal/mol in solution (with a PCM solvation free energy correction) for the unimolecular dissociative mechanism for anionic methylphosphate (152). The reaction of the metaphosphate intermediate with a water molecule leading to dihydrogenphosphate is calculated to proceed *via* a free energy barrier of $\approx$ 25 kcal/mol, in gas phase and 17 kcal/mol in solution. Including one (two) water molecule(s) into the calculations leads to a reduction of the free energy barriers to 26.7 (26.4) kcal/mol (gas phase) and 23.8 (24.3) kcal/mol (solution) for the formation of the metaphosphate intermediate and 19.4 (13.2) kcal/mol (gas phase) and 6.6 (1.5) kcal/mol (solution) for the product formation reaction.

Associative mechanism. A B3LYP/6-31+G(d) study on water attack on anionic methylphosphate reported a free energy activation barrier of 43.8 kcal/mol in gas phase and 36.1 kcal/mol in solution (with a PCM solvation free energy correction) (152). These barriers are reduced to 39.2 kcal/mol (gas phase) and 36.5 kcal/mol (solution) in the presence of an additional water molecule. Bond lengths of 2.28 (only the attacking water molecule is present) and 2.33 Å (both attacking and additional water molecules are present) indicate an associative mechanism with dissociative character. However, the barriers are higher than for the dissociative mechanism, indicating that hydrolysis proceeds *via* a dissociative mechanism. This result was confirmed by B3LYP/cc-PVTZ+//B3LYP/6-31+G(d,p) calculations using the self-consistent reaction field (SCRF) approximation to account for solvation effects (153)

A computational investigation at the MP2/6-31+G(d,p)//HF/6-31G(d) level of theory of OH^- attack to neutral methylphosphate found a barrier of $\approx$ 12 kcal/mol using the Langevin dipole approximation to account for solvent effects and about 17 kcal/mol with the polarizable continuum model (PCM)[4] to account for solvent effects (155). The authors state that OH^- formation by proton transfer from H_2O to monoanionic methylphosphate can be determined from the difference in pK_a values of water (pK_a = 15.7) and methylphosphate(pK_a = 1.65) to occur with an equilibrium constant of $K \approx 10^{-14}$. Thus, the apparent reaction rate $k_1 = K \cdot k_2$ can be used to give an estimate for the rate for OH^- attack on neutral methylphosphate of $\approx 10^7$ $M^{-1}s^{-1}$. The corresponding barrier of $\approx$

[4]The authors of this study state that PCM involves a systematic overestimation of activation barriers originating from overestimation of the hydration free energy of OH^- (-121 kcal/mol calculated *versus* -106.4 kcal/mol experimental (154))

10 kcal/mol agrees well with their computed barrier. Thus, it was shown that even under acidic reaction conditions in which the monoanionic species predominates the hydrolysis reaction may proceed *via* an OH^- attack to the neutral species. The bond length in the transition state of 2.3 Å indicated an associative mechanism with dissociative character.

Dimethylphosphate hydrolysis The experimental barrier for dimethylphosphate ester hydrolysis at 25°C and an ionic strength of 1 has been estimated to be about 28.2 kcal/mol (156). Another study reports a rate of $3.3{\cdot}10^{-6}s^{-1}$ at 100°C, corresponding to an apparent barrier of 131.19 kJ/mol or 31.4 kcal/mol for the neutral species and $1.0{\cdot}10^{-9}s^{-1}$ at 100°C, corresponding to an apparent barrier of 156.32 kJ/mol or 37.4 kcal/mol for the anionic species (157). In the same study a pK_a value of 0.50 was determined for $(CH_3)_2HPO_4$. A detailed experimental investigation yielded an activation enthalpy of 25.9 kcal/mol and an activation entropy of -34 cal/mol/K (151). The conformational flexibility of dimethylphosphate in aqueous solution has been investigated by MP2/6-31+G(d,p)//HF/6-31G(d) computations (158). A maximum barrier of 3 kcal/mol for interconversion between different conformers was found, thus indicating that several conformations may contribute to the reaction in a given protonation state.

An early computational study of the dianionic transition state of OH^- attack to anionic dimethylphosphate in vacuum at the HF/STO-3G, HF/3-21G(d), and HF/3-21+G(d) levels of theory found several pentacovalent intermediate and transition state structures that differed only little in their relative energies (159) with a barrier height of ≈ 90 kcal/mol.

A computational study of anionic dimethylphosphate hydrolysis at the B3LYP/6-31+G(d,p) level of theory including solvent effects showed that the reaction proceeds *via* a phosphorane intermediate with both OH^- and H_2O as nucleophile (160). In vacuum, the highest barriers found are corresponding to the first transition state that connects the reactant with the intermediate. They are 91 kcal/mol for the OH^- attack and 27 kcal/mol for H_2O attack. In the latter, a proton is transferred to one of the phosphoryl oxygen atoms concerted with the formation of the P-OH bond. In water, both barriers are 41 kcal/mol. The P-O bond lengths to the incoming nucleophile in the phosphorane intermediate are all between 1.8 and 1.9 Å, indicating an almost perfect S_N2 reaction.

Trimethylphosphate hydrolysis The experimental rate constant for OH^- attack on trimethylphosphate was reported as $1.6{\cdot}10^{-4}M^{-1}s^{-1}$ at 25°C (161), corresponding to an apparent reaction free energy barrier of 94.64 kJ/mol or 22.6 kcal/mol.

A computational investigation at the MP2/6-31+G(d,p)//HF/6-31G(d) level of theory of OH^- attack to trimethylphosphate found a barrier of ≈ 25 kcal/mol using the Langevin dipole approximation to account for solvent effects and about 32 kcal/mol with the polarizable continuum model (PCM) to account for solvent effects (155). The bond length in the transition state of 2.4 Å indicated an associative mechanism with dissociative character.

Hydrolysis of phosphoanhydrides

Phosphoanhydrides are similar to phosphate monoesters in that they possess a single phosphoryl substituent. Thus, the mechanisms of P-O-P bond cleavage may be similar to the mechanisms of P-O-C bond cleavage. Despite the prevalence of phosphoanhydride hydrolysis reaction in biologic reactions no systematic experimental study on phosphoanhydride hydrolysis reactions had been reported until 1995, when linear free energy relationships between the pK_a values of different incoming alcoholic nucleophiles (and different leaving groups) and the logarithm of the rate constant were measured (135). Only little dependence of the reaction rate on the pK_a of the leaving group was found, which is indicative of a dissociative mechanism. Rate constants of $7.0{\cdot}10^{-7}s^{-1}$ and $2.17{\cdot}10^{-6}s^{-1}$ were measured at 60°C for ATP hydrolysis in absence and presence of Mg^{2+}. This corresponds to apparent free energy barriers of 121.10 kJ/mol (28.9 kcal/mol) and 117.97 kJ/mol (28.2 kcal/mol), respectively. At 95°C and in the absence of Mg^{2+} rate constants of $4.0{\cdot}10^{-5}s^{-1}$ (apparent barrier 121.76 kJ/mol or 29.1 kcal/mol) and $6.67{\cdot}10^{-6}s^{-1}$ (apparent barrier 127.24 kJ/mol or 30.4 kcal/mol) were measured for ATP and pyrophosphate hydrolysis, respectively Thus, the barriers found correspond to earlier measurements of the rate of ATP hydrolysis (see Section 4.1).

Pyrophosphate (PP_i). The first computational study reported on pyrophosphate hydrolysis in gas phase at the MP2/DZP level of theory found a dissociative mechanism for uncatalyzed, acid catalyzed, and Mg^{2+} catalyzed reactions (162; 163). The unimolecular dissociation reaction of $H_2P_2O_7^{2-}$ up to the metaphosphate intermediate proceeds *via* a free energy barrier of 23.1 kcal/mol, that is lowered to 9.3 kcal/mol in the presence of Mg^{2+}. Mg^{2+} is found to act by activating one P-O_{bridge} bond as evidenced by bond elongation. Similar results were found for $H_3P_2O_7^-$ and $HP_2O_7^{3-}$. The latter reactions, however, were not studied in detail.

Another computational study investigated the stable structures and intermediates for a dissociative mechanism of Mg-PP_i considering different protonation states at the MP2/6-31+G(d,p)//HF/6-31+G(d,p) level of theory (164). The calculated reaction free energies in aqueous solution of -48.5, -40.1, and -26.4 kJ/mol for $Mg.H_2P_2O_7$, $Mg.P_2O_7^-$, and $Mg.P_2O_7^{2-}$ are in the expected range as compared to experimental values of pyrophosphate hydrolysis.[5] In the intermediate states the Mg^{2+} was found to bridge the orthophosphate and metaphosphate moieties that therefore did not have any direct contact. A charge analysis of the different species involved revealed significant charge transfer from the negatively charged pyrophosphate to the positively charged Mg^{2+} ion, reducing the charge on Mg^{2+} from formally +2 to 0.5 to 1.0 (1.5 to 1.7) as determined from Mullikan population (CHELP) analyses. In a previous study it was reported that the two P-O_{bridge} bonds in PP_i have the same length in gas phase, whereas an asymmetry is introduced by Mg^{2+}

[5]The experimental reaction free energies (in kcal/mol) as measured calorimetrically for pyrophosphate hydrolysis are -9.5 ($H_4P_2O_7$), -7.5 ($H_3P_2O_7^-$), -7.7 ($H_2P_2O_7^{2-}$), -7.1 ($HP_2O_7^{3-}$), and -10.4 ($P_2O_7^{4-}$)(138).

coordination (165). In this study, the gas-phase free energy barriers for the isomerization reactions leading from the reactant structure to an intermediate structure in which the Mg cation bridges the phosphate and metaphosphate moieties was calculated as 5.6, 10.4, and 13.5 kcal/mol for the neutral ($MgH_2P_2O_7$), anionic ($MgHP_2O_7^-$) and dianionic ($MgP_2O_7^{2-}$) species, respectively. The corresponding calculated reaction free energies were -16.8, -12.7, and 1.8 kcal/mol, respectively.

No computational study on Mg-PP_i hydrolysis with hexacoordinated Mg^{2+} (as is expected in aqueous solution and in enzymatic environment) has been reported to date.

Methyltriphosphate. Methyltriphosphate is a model compound for both ATP and GTP in which the sugar and base are replaced by a methyl group. Its properties are expected to be similar to the properties of triphosphoric acid, $H_5P_3O_{10}$.[6] In gas phase, the dianion $H_3P_3O_{10}^{2-}$(167) and the neutral acid (as well as the triphosphate moiety of neutral ATP and neutral GTP) (168) adopt a cyclic conformation that is stabilized by intramolecular hydrogen bonds between the α- and γ-phosphate moieties. Mg^{2+} coordination increases the P-O_{bridge} bond lengths from 1.59 and 1.65 Å to 1.60 - 1.70 Å and opens the P_γ-O_{bridge}-P_β angle from 119.3° to 125.7° (168). In aqueous solution, the bond lengths are unchanged and the P-O-P angle opens further to 128.3°. The Mg^{2+} is hexacoordinated both in gas phase and in solution.

No systematic computational study exploring different conformers at different protonation states of methyltriphosphate (let alone ATP) neither in gas phase nor in aqueous solution has been reported to date.

The first computational investigations of methyltriphosphate hydrolysis in aqueous solution were reported in 2003. Car-Parrinello molecular dynamics (CPMD) in a cubic box containing the methyltriphosphate molecule, a Mg^{2+} ion and 54 water molecules were used to study both dissociative and associative mechanisms up to the formation of the transition states (168). For the dissociative mechanism a free energy barrier of 36 kcal/mol was found with the water environment playing a passive role throughout the reaction. For the associative mechanism a free energy of 39.1 kcal/mol was found, in which the pentavalent transition state is formed after a proton transfer from the attacking water molecule to the γ phosphate moiety. The P-O distance to the incoming water oxygen at the transition state is 1.9 Å, which indicates an almost perfect S_N2 mechanism. Usage of a mixed reaction coordinate, in which cleavage of the P-O_{bridge} bond is followed by nucleophilic attack of a water molecule leads to a reaction free energy barrier of 35.1 kcal/mol. The authors suggest that the Mg^{2+} cation may have an active role in catalyzing the P-O-P bond cleavage by electrophilic attack onto the bridge oxygen, thereby weakening the bridge bond and initiating the bond breaking reaction. Due to the short simulation length and small simulation system, the metaphosphate intermediate may be artificially stabilized, an effect to which

[6]The pK_a values for triphosphoric acid $H_5P_3O_{10}$ at 25°C and an ionic strength of 1.0M are p$K_{a1} < 1$, p$K_{a2} = 1.05$, p$K_{a3} = 2.11$, p$K_{a4} = 6.00$, and p$K_{a5} = 8.90$ (166). Thus, at neutral pH the predominant species is $HP_3O_{10}^{4-}$.

the authors attribute 10 kcal/mol. Thus, they estimated an effective barrier of about 25 kcal/mol for a dissociative mechanism, which is therefore energetically favored over an associative mechanism. However, the barrier height in an associative mechanism with H_2O attack may be lowered by using OH^- as the incoming nucleophile. Such a possibility was not explored.

The hydrolysis of protonated methyltriphosphate was investigated at the B3LYP/cc-PVTZ+ //B3LYP/6-31+G(d,p) level of theory using the SCRF approximation to account for solvation effects (153). In presence of one explicit water molecule two different associative paths with free energy barriers of 36.4 (47.2) and 34.3 (54.5) kcal/mol in aqueous solution (gas phase) were found. Both possibilities proceeded in a single step reaction *via* pentacovalent phosphorane transition states that differed in their geometry: distances of 2.12 and 1.84 Å between the incoming water oxygen and P_γ indicate that the two pathways differ in their degree of associativity. In presence of two explicit water molecules three out of several possible associative pathways were studied, all of which proceeded stepwise. However, the first and rate-limiting step was found to be identical in all three cases. Its free energy barrier was 33.9 (45.6) kcal/mol in aqueous solution (gas phase) and thus isoenergetic with the result found in the presence of one explicit water molecule. The water oxygen-P_γ distance was 2.11 Å. Several dissociative mechanisms were also explored with rate-limiting barriers of 38.5 (41.3) and 30.5 (42.0) kcal/mol in aqueous solution (gas phase)in presence of one explicit water molecule and 37.3 (38.4) and 30.4 (39.2) kcal/mol in aqueous solution (gas phase) in presence of two explicit water molecules. The authors conclude that in aqueous solution the hydrolysis of methyltriphosphate can proceed *via* either associative or dissociative mechanisms. However, because the reaction barrier of the associative pathways is more sensitive to electrostatic influences by the environment than the barrier for the dissociative pathway (as indicated by larger solvation energies for the associative transition states than for the dissociative transition states), the authors suggest that "the associative pathway may possibly be more susceptible to stabilization by the protein environment than the dissociative pathway".

Nucleotides. No mechanistic study on the hydrolysis of a nucleotide such as AMP, ADP, ATP, GMP, GDP, or GTP is available.

Summary

The computational investigations reviewed above indicate a number of important points:

1. The choice of quantum mechanical calculation method strongly influences the magnitude of the calculated activation barrier even if the basis set is kept constant. HF barriers may be up to 20 kcal/mol higher than DFT or MP2 barriers.

2. The predominant species at a given pH value need not be the reactive species.

3. The barrier for proton transfer may be reduced by 5 to 10 kcal/mol when mediated by one or two water molecules.

4. The presence of an additional water molecule may increase the dissociative character of an associative hydrolysis mechanism.

5. Both dissociative and associative mechanisms can explain experimental reaction rates of phosphate ester hydrolysis. Enzymes can therefore select either mechanism, as has been pointed out earlier (169; 170).

6. Even for a reaction as simple as methylphosphate hydrolysis in water, the reaction mechanism cannot unambiguously be determined. Thus, considerable variability can be expected for enzymatic reactions.

7. Phosphoanhydride hydrolysis reactions of biologically relevant molecules such as PP_i or nucleotides are mechanistically poorly understood.

8. Whether hydrolysis reactions of phosphoanhydrides proceed *via* a dissociative or an associative mechanism and in a single step or multi step fashion in aqueous solution is unresolved. Thus, any possibility seems open for enzyme-catalyzed nucleotide hydrolysis reactions.

4.3 Mechanistic considerations on phosphate anhydride hydrolysis in proteins

A number of studies have investigated enzyme-catalyzed GTP and ATP hydrolysis. Because of the structural similarities in the nucleotide binding pocket (see Section 3.3) these proteins may use similar strategies in catalyzing nucleotide hydrolysis. Among the best characterized proteins is the G-protein Ras, for which a number of both experimental and theoretical studies have been reported. Because of the expected similarities to myosin these studies shall be reviewed here. Kinetic and mechanistic studies on myosin itself are discussed in Sections 4.4 and 4.6.

Studies on phosphoryl transfer proteins such as kinases or phosphatases (171; 172; 173; 174) are not considered or only mentioned if appropriate.

4.3.1 Enzyme strategies for catalysis

A catalyst speeds up a chemical reaction without being changed itself. This is achieved by a transient interaction between the catalyst and the reactants such that reaction intermediate and transition states may be altered in the catalyzed reaction as compared to in the

uncatalyzed reaction. Within the explanatory framework of the transition state theory the speed-up is explained by a lowering of the activation barrier of the reaction (175). Enzymes may adopt several strategies to achieve this:

1. **Decreased substrate flexibility.** Restriction of the conformational flexibility of the substrate by capturing the substrate in a conformation that is optimal for hydrolysis increases the population of the conformation that is competent for hydrolysis.

2. **Ground state destabilization.** Upon binding conformational strain can be introduced into the substrate, forcing its atoms to dislocate along the reaction coordinate *before* the reaction actually starts.

3. **Transition state stabilization.** Favorable interactions between the protein and the substrate in its transition state can lower the energy of the transition state.

4. **Positioning of the attacking water.** The microenvironment of the enzyme active site may position the attacking water molecules for optimal in-line attack (as required for an associative S_N2 reaction mechanism).

5. **Nucleophile activation.** The nucleophilic of the incoming nucleophile may be increased by abstraction of a proton from the attacking water by a general base, thus creating an OH^- ion that attacks. This, however, affects only associative mechanisms, because in dissociative mechanisms the formation of the transition or intermediate state is independent of the attacking nucleophile.

6. **Change in substrate protonation.** The protonation pattern of the substrate may be changed upon binding, thus creating a more reactive species than is predominant in solution.

4.3.2 The GTPase Ras

Ras proteins are involved in cellular signaling leading to cell growth and differentiation (176; 177; 178). They are active when GTP is bound and switched off by the hydrolysis of GTP. In the presence of a GAP (GTPase Activating Protein) the hydrolysis reaction is greatly enhanced. The rate constant for the Ras catalyzed GTPase reaction is $4.7{\cdot}10^{-4}s^{-1}$, corresponding to an effective barrier of 22.4 kcal/mol (179; 180). The interaction with GAP speeds up the hydrolysis resulting in a hydrolysis rate of 19 s^{-1} (181) corresponding to an effective barrier of 16 kcal/mol.

Gln61 in Ras p21 has been proposed to act as a general base that activates the attacking water molecule (182; 183). This, however, has been questioned (184): Based on considerations of the pK_a values of protonated glutamine which is a strong acid and deprotonated water, which is a base, the existence of an ion pair (as would be present after a proton

transfer has happened) was found to be unlikely. The proposed alternative involves catalysis *via* a dissociative mechanism[7] in which the negative charge on the $\beta\gamma$-bridge oxygen in the transition state is stabilized by positive charges in the protein environment, especially by a hydrogen bond from the backbone amide group of Gly13 (which is the fourth residue in the P-loop). In such a scenario, GAP activation can be explained by additional stabilization of the transition state *via* interactions of the positively-charged Arg-finger motif of GAP with the $\beta\gamma$-bridge oxygen. The proposed dissociative mechanism has been confirmed by time-resolved Fourier transform infrared (FTIR) difference spectroscopic studies that revealed that Ras forces GTP into a specific conformation and induces significant charge shift to the β oxygens of GTP upon binding (187). GAP interaction with Ras increases the charge shift (188). Measurements of kinetic isotope effects showed normal primary and secondary KIE, which the authors interpreted as evidence for a loose transition state and a more dissociative-like mechanism (189). However, a normal primary KIE could indicate that bond cleavage is rate limiting, and a normal secondary KIE indicates a more associative transition state (see Section 5). Thus, the interpretation in favor of a dissociative mechanism seems questionable.

A computational free energy perturbation (FEP) study using the empirical valence bond (EBB) description in combination with Langevin dipoles (LD) to model the solvent found that GAP reduces the barrier for hydrolysis by 7 to 10 kcal/mol independent of whether an associative or a dissociative mechanism was used (180). By comparison with barriers obtained for methylphosphate hydrolysis in water in which the barrier for the dissociative mechanism is computed to be higher than for the associative mechanism the authors conclude that Ras takes an associative route. However, no absolute barriers or geometric details on the valence bond states used to describe the transition or intermediate states are given.

A very recent computational study employed both QM-only calculation on a cluster model of the active site at the B3LYP/6-31G(d,p)//B3LYP/4-31G(d)/3-21G level of theory and QM/MM calculation employing the effective fragment potential method at the MP2/6-31+G(d)//HF6-31G level of theory (190). The QM part included the triphosphate moiety of GTP, Mg, the lytic water molecule, and the sidechains of Gln61 of Ras and Arg789 from GAP. An active role for Gln61 as a proton acceptor and for Lys16 as a proton donor was ruled out by the calculations. An associative mechanism in which proton transfer from the lytic water to γ-phosphate occurs simultaneously with bond making between the incoming water oxygen and P_γ was found to proceed *via* a one-step mechanism without intermediate. At the transition state the distances r_1 (P_γ:incoming water oxygen) and r_2 (P_γ:$\beta\gamma$-bridge oxygen) were 1.59 and 1.98 Å, respectively. Since the barrier of 39 kcal/mol for the QM/MM calculations was found to be equally unsatisfying as "non-promising results" for the cluster model the authors performed a two-dimensional grid search on the cluster model, varying

[7]Linear free energy relationship measurements for alkaline phosphates suggests a dissociative mechanism for phosphoryl transfer reactions in proteins (185; 186). Note, however, that the mechanistic interpretation of linear free energy relationship is debated.

both r_1 and r_2. In this way another transition state was located at distances $r_1 = 2.12$ Å and $r_2 = 2.20$ Å in which the lytic water still carried both protons. The calculated barrier was 20.7 kcal/mol. Decay of the transition state structure lead to an intermediate state with an energy of 14.5 kcal/mol above the reactants in which the leaving group was completely dissociated but the lytic water was still doubly protonated. The necessary second step of the reaction corresponding to rearrangement of $(H_2O{\cdot}PO_3$ to H_2OPO_4 was not explicitly determined since the barrier is expected to be significantly lower than 20 kcal/mol. The final product state has an energy of 9.8 kcal/mol with respect to the reactant state. Charge analyses indicated that considerable negative charge is shifted to the bridge oxygen when moving from the reactant to the transition state. Thus, the authors conclude that GTP hydrolysis in Ras occurs *via* a dissociative mechanism as favored by experiment.

4.4 Kinetics of ATP hydrolysis in myosin

Kinetic studies on the contractile cycle of myosin interacting with actin are being conducted for more than 40 years. A landmark was the compilation of available kinetic data into a consistent scheme now known as the Lymn-Taylor cycle (191). Figure 4.2 illustrates this cycle. In Step 1, the actomyosin complex $A.M$ binds ATP, followed by unbinding of actin in Step 2. ATP hydrolysis occurs in Step 3. In Step 4 actin rebinds and triggers product release (Step 5), leading back to the starting point of the contractile cycle.

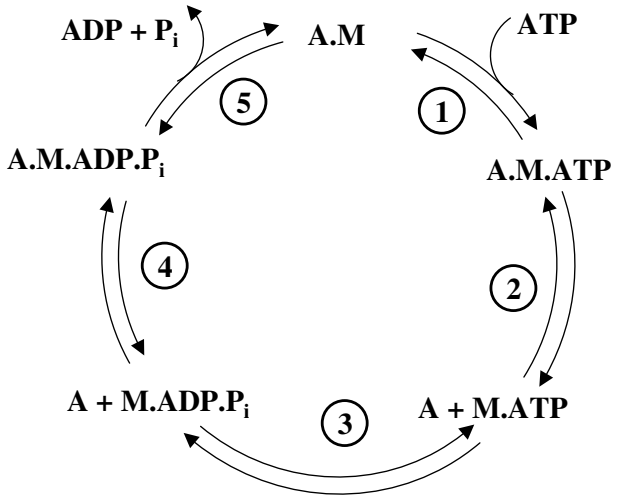

Figure 4.2: Illustration of the actomyosin cycle known as Lymn-Taylor cycle (191).

The Lymn-Taylor scheme forms the basis for the interpretation of a vast number of kinetic measurements on different myosins and from different organisms, both interacting and non-interacting with actin.

Since it was observed that substrate binding and hydrolysis by myosin induces spectral changes that are observable by measuring ultraviolet absorption difference spectra (192) this change in intrinsic protein fluorescence has been widely utilized to relate structural transition of the protein to its enzymatic function (see (193; 194; 195; 196) and references below). Protein intrinsic fluorescence stems mainly from the fluorescent properties of the sidechains of tryptophane residues. If the chemical environment of such a sidechain changes, the fluorescence properties of the protein change correspondingly. In addition, substrates labeled with fluorophores are used to investigate the kinetics of myosin ATP binding, hydrolysis, and product release. Examples are *mant*ATP (2'(3')-*O*-(*N*-methylanthraniloyl)-adenosine-triphosphate), *mant*ADP (2'(3')-*O*-(*N*-methylanthraniloyl)-adenosine-diphosphate), or coumarin-labeled ADP (197). Moreover, kinetic measurements utilizing radioactively labeled phosphate compounds (198) or light scattering techniques (199; 200) were used.

Such measurements established that myosin II proteins from different organisms or tissues differ in their observable response to ATP binding and hydrolysis (201; 202). For example, rabbit skeletal muscle myosin II exhibits an increase in fluorescence intensity upon ATP binding that increases further with hydrolysis (81), whereas *Dictyostelium discoideum* myosin II exhibits a quench in fluorescence intensity, followed by an increase in intensity (81). Although several Trp residues can be found in the primary sequence of myosin II in *all* organisms the change in fluorescence intensity has been shown to stem mainly from the conserved Trp501 (numbering from *Dictyostelium discoideum*) which can therefore be considered a sensor for hydrolysis (203; 204; 205; 206; 207).

It is necessary to adapt the Lymn-Taylor scheme depending on the class of myosins studied and on the organism the protein is extracted from. Such an adapted scheme then may also account for a number of intermediate steps that can be kinetically resolved using modern techniques. An adapted kinetic cycle based on the myosin adenosine triphosphatase reaction cycle in absence of actin as proposed by Bagshaw and Trentham in 1974 (208) is shown in Figure 4.3 for *Dictyostelium discoideum* myosin II, that has been engineered so as to contain no Trp residues other than Trp501 (209). In Step 1 in this scheme ATP binds to apo-myosin (M state) thus forming a collision complex (M.ATP). This followed by a conformational change of myosin as monitored by a quench in the fluorescence intensity (Step 2, $M^{\dagger}$.ATP). Upon ATP hydrolysis in Step 3a fluorescence enhancement is observed (M^{*}.ATP). This step has been further dissected by means of pressure jump and temperature jump methods (209). In the same study, the M state was identified with the *apo* structure of myosin (for which no crystal structure is available), the $M^{\dagger}$ state corresponding to the crystallographic C/O state, and the M^{*} state corresponding to the crystallographic C/C state. Thus, the chemical ATP hydrolysis step, in the C/C conformation of myosin, corresponds to Step 3b. Hydrolysis is followed by a conformational change indicated by a quench in fluorescence (Step 4, M^{*}.ADP.P_i) and subsequent phosphate release (Step 5, M^{*}.ADP). Another conformational change occurs, restoring the level fluorescence intensity of the *apo* structure (Step 6, M.ADP) that precedes ADP release (Step 7).

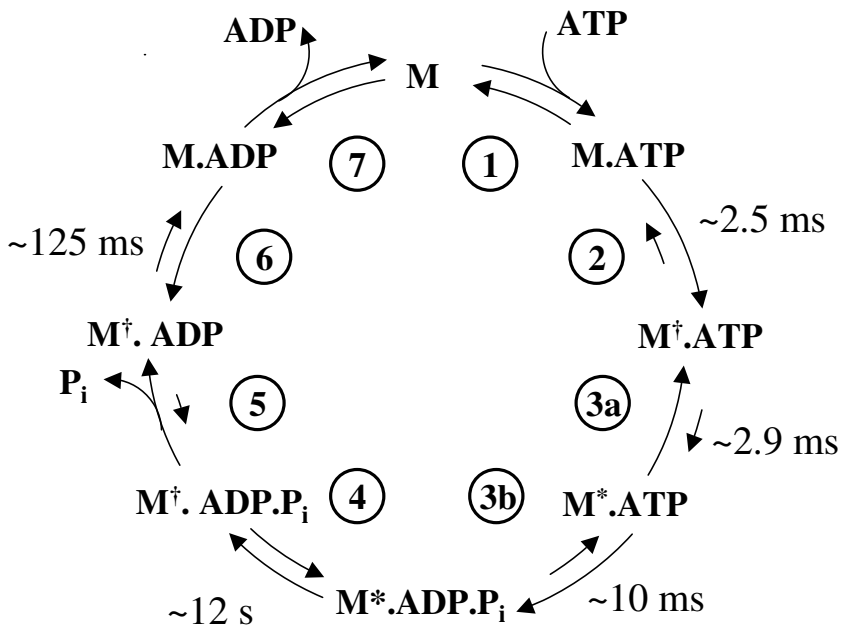

Figure 4.3: Kinetic cycle for ATP hydrolysis by *Dictyostelium discoideum* myosin II in absence of actin as derived from kinetic fits to fluorescence measurements using a protein construct termed W501+ that contained only one fluorescent tryptophan W501 (209). In Step 1, ATP binds to apo-myosin (M), thus forming a collision complex (M.ATP), presumably with both switch-1 and switch-2 loops open (O/O state). This is followed by a conformational change of myosin as monitored by a quench in the fluorescence intensity (Step 2), leading to $M^{\dagger}$.ATP that presumably corresponds to the closing of Switch-1 (C/O state). In Step 3a, fluorescence enhancement is observed, leading to M^{*}.ATP, presumably corresponding to closing of Switch-2 (C/C state), coupled with the recovery stroke of the lever-arm. The catalysis of ATP hydrolysis (step 3b) in the C/C conformation of myosin. Hydrolysis is followed by a conformational change indicated by a quench in fluorescence (Step 4, $M^{\dagger}$.ADP.P_i, presumably the start of the power stroke) and subsequent phosphate release (Step 5, $M^{\dagger}$.ADP). Another conformational change occurs, restoring the fluorescence intensity of the *apo* structure (Step 6, M.ADP) that precedes ADP release (Step 7). The time constants (*i.e.*, the reciprocal of the measured rate constants (81; 196; 209), see also Ref. (210)) for forward reaction steps are indicated.

4.4.1 Experimental kinetic data

Myosin binds and hydrolyses ATP only in the presence of a divalent cation (211; 212). The basal ATPase in presence of Mg^{2+} is 0.12 s^{-1} for the M761-2R construct. Replacing Mg by Ca leads to an increase to a rate of 1.0 s^{-1} (203). The production of force, however, is decreased when calcium is used to replace magnesium (213). Mn^{2+} can also be used (214; 215). In the present work the focus is on the Mg^{2+}-ATPase activity of myosin which is therefore discussed in detail. In presence of actin the basal ATPase rate is increased approximately 1000-fold due to a catalysis of the rate-limiting product release steps. The actin-activated rate of ATP hydrolysis depends strongly on the ionic strength (reducing the ionic strength from 0.1 to 0.04 results in a 20-fold rate increase) (216). The chemical hydrolysis event occurs in the myosin head when myosin is detached from actin. Thus, the following discussion focuses on the myosin ATPase cycle in absence of actin.

Since the possibility of expressing the motor domain of myosin II in the slime mold *Dictyostelium discoideum* was reported (217) this organism has been widely used for detailed characterization of the properties of myosin. This is mainly due to (1) the simplicity with which myosin can be expressed in *Dictyostelium discoideum* compared to bacterial expression systems and (2) the relative ease to prepare large amounts of different myosin constructs (218). Kinetic measurements have been made for a large number of different *Dictyostelium discoideum* myosin II constructs. The nomenclature used to distinguish different constructs and mutants is as follows: A number n following the letter M denotes a myosin head fragment consisting of the N-terminal n wild-type amino acids (219). Mutants are indicated using the one letter code for the wild-type residue followed by the residue number and the one letter code of the mutated amino acid. Thus, M761(G457A) refers to a myosin construct consisting of the first 671 N-terminal amino acids in which Gly457 has been mutated to Ala. The steady-state basal ATPase activities of different motor domain constructs of *Dictyostelium discoideum* myosin II are collected in Table 4.3.

Transient kinetic measurements have been used to measure the kinetics of the different steps associated with the ATPase cycle. The experimental conditions used for the different measurements are collected in Table 4.4 and the resulting values in Table 4.5. In Section 4.5 these data are used to assemble a thermodynamic free energy profile for the ATPase cycle.

The fluorescence enhancement associated with Step 3 does not only occur upon the addition of ATP but also upon the addition of the non-hydrolyzable ATP analog AMP-PNP. This indicates that the conformational transition corresponding to the recovery-stroke is an event separate from the chemical hydrolysis step. Moreover, nucleotide binding is associated with a fluorescence quench in *Dictyostelium discoideum* when myosin adopts the C/O conformation. Upon addition of ATP, however, no change in fluorescence prior to the onset of hydrolysis is observed at 20°C, indicating that the M^*.ATP state exists transiently. This supports the view that Step 3 in the original Bagshaw-Trentham scheme must be split into

construct	k_{cat} $[s^{-1}]$	conditions	ref.
M761	0.07 ± 0.01	40mM NaCl, 20mM TES, 1mM $MgCl_2$, pH=7.5, T=20°	(81)
M761 W501+	0.06 ± 0.01	40mM NaCl, 20mM TES, 1mM $MgCl_2$, pH=7.5, T=20°	(81)
M761-2R	0.12	Mg, T=30 °, rest as in (216) (40mM KCl, 5mM $MgCl_2$, pH=8.0, T=20°)	(203)
M761-2R	1.0	high-salt Ca, T=30°, rest as in (216)	(203)
MHF (=M864)	0.04	25 mM HEPES, 0.1M KCl, 5mM $MgCl_2$, pH=7.0, T=20°	(220)
M754	0.04	25 mM HEPES, 0.1M KCl, 5mM $MgCl_2$, pH=7.0, T=20°	(220)
D1dC (=M759)	0.08	15 mM MOPS, 4mM $MgCl_2$, 1mM DTT, 1mM ATP, 5μg/ml S1dC, T=25°	(46)
D1dC (=M759)	0.028	40mM NaCl, 20mM TES, 1mM $MgCl_2$, pH=7.5, T=20°	(218)

Table 4.3: Steady-state ATPase rates for different *Dictyostelium discoideum* myosin II constructs. Experimental conditions are also given. W501+ is a mutant in which all W except W501 are mutated to F. M761-2R is the M761 fragment fused to two actinin repeats (50). The recombinant protein MHF is the myosin head fragment prepared as described in (217).

construct	conditions	ref.
M864	25mM HEPES, 0.1M KCl, 5mM $MgCl_2$, pH=7.0, T=20°	(221)
M754	25mM HEPES, 0.1M KCl, 5mM $MgCl_2$, pH=7.0, T=20°	(220)
M761	25mM HEPES, 100mM KCl, 5mM $MgCl_2$, pH=7, T=20°	(219)
M781	25mM HEPES, 100mM KCl, 5mM $MgCl_2$, pH=7, T=20°	(219)
M759	40mM NaCl, 20mM TES, 1mM $MgCl_2$, pH=7.5, T=20°	(218)
M761(501+)	40mM NaCl, 20mM TES, 1mM $MgCl_2$, pH=7.5, T=20°	(81)

Table 4.4: Experimental conditions for transient kinetics studies.

property	M864[a]	M754[b]	M761[c]	M781[c]	M759[d]	M761(501+)[e]
K_1k_2 [$M^{-1}s^{-1}$]	$0.94{\cdot}10^6$	$5.7{\cdot}10^4$	$6.06{\cdot}10^5$	$5.5{\cdot}10^5$	$1.7{\cdot}10^5$	$6.6{\cdot}10^5$
k_2 [s^{-1}]	n.d.	n.d.	n.d.	n.d.	n.d.	≈ 400
k_{-2} [s^{-1}]	n.d.	n.d.	n.d.	n.d.	n.d.	≈ 0.02[f]
$k_3 + k_{-3}$ [s^{-1}]	24	80	160	37	17	30
k_6 [s^{-1}]	1.5	0.06	n.d.	n.d.	1.1	7.9
$k_{-6}\frac{1}{K_7}$ [Ms^{-1}]	$9{\cdot}10^5$	$1.25{\cdot}10^5$	n.d.	n.d.	$2.4{\cdot}10^5$	$1.4{\cdot}10^6$
K_1[g]	$2.35{\cdot}10^3$	$1.43{\cdot}10^2$	$1.52{\cdot}10^3$	$1.38{\cdot}10^3$	$4.25{\cdot}10^2$	$1.65{\cdot}10^3$
k_{-6}[h] [s^{-1}]	383	874	n.d.	n.d.	565	848

[a]From Ref. (221). K_1k_2 and $k_3 + k_{-3}$ are measured with ATP, k_6 and $k_{-6}\frac{1}{K_7}$ with *mant*ADP.

[b]From Ref. (220). K_1k_2 and $k_3 + k_{-3}$ are measured with ATP, k_6 and $k_{-6}\frac{1}{K_7}$ with ADP.

[c]From Ref. (219).

[d]From Ref. (218)

[e]From Ref. (81). A myosin construct with only one Trp (Trp501) was used. In addition to the data collected here, the relations $k_{-1} > k_{-2}$ and $k_{-2} << k_3$ were established. Additional information on Step 3 see Table 4.6

[f]Measured for the ATP analog AMP.PNP. The kinetics for ATP may be significantly different.

[g]From $K_1 = \frac{K_1k_2}{k_2}$. From the M761.W501+ construct estimate $k_2 \approx 400s^{-1}$ is used.

[h]From $k_{-6} = k_{-6}\frac{1}{K_7} \cdot K_7$ and assuming $K_7 = \frac{1}{K_1}$, *i.e.*, the collision complex formation constant is independent of the nucleotide.

Table 4.5: Kinetic data as determined by transient kinetics studies for *Dictyostelium discoideum* myosin constructs. Values that have not been determined are marked with n.d.

two separate events, namely the recovery-stroke transition in Step 3a and the chemical hydrolysis event in Step 3b.

Temperature-jump and pressure-jump transient kinetics measurements can been used to decouple Steps 3a and 3b. Such experiments have been reported for rabbit and chicken skeletal S1 (222; 223) and for *Dictyostelium discoideum* M761 (209). These experiments established unambiguously that the recovery-stroke equilibrium (Step 3a) is both temperature and pressure sensitive, whereas the hydrolysis event (Step 3b) is almost independent of temperature and pressure. Estimates for the rates and equilibrium constants associated with Steps 3a and 3b at 20°C for *Dictyostelium discoideum* are collected in Table 4.6.

property	value[a]	estimate[b]
K_{3a}	0.4[c]	≈ 0.2
k_{3a} [s^{-1}]	350[d]	
k_{-3a} [s^{-1}]	870[d]	
K_{3b}	13[d]	≥ 10
K_{3b}	79[e]	
k_{3b} [s^{-1}]	90-110[f]	
k_{-3b} [s^{-1}]	1.1-8.5[g]	

[a]From Ref. (209).
[b]From Ref.(81)
[c]From stopped-flow fluorescence transients and pressure jump analysis.
[d]From pressure jump kinetics.
[e]From steady-state fluorescence.
[f]From stopped-flow transients. The range stems from using the two estimates for K_{3b}.
[g]From $k_{-3b} = \frac{k_{3b}}{K_{3b}}$

Table 4.6: Estimated values for the rates and equilibrium constants at 20°C associated with Steps 3a (recovery-stroke transition) and 3b (hydrolysis) in *Dictyostelium discoideum* myosin II.

The rate of phosphate release (Steps 4 and 5) is limited by the isomerization in Step 4, for which $k_4 = 0.05\ s^{-1}$ was measured (196) for for *Dictyostelium discoideum* myosin II. For rabbit skeletal myosin II is was estimated to be $k_5 = 85\ s^{-1}$ at saturating actin concentration and 20° (224). With actin increasing the rate of phosphate release by approximately 1000-fold, the rate of phosphate release in absence of actin can therefore be estimated as $k_4 = 8.5 \cdot 10^{-2}\ s^{-1}$, which agrees well with the value for *Dictyostelium discoideum* myosin II.

4.4.2 Mutational analyses

A number of different mutational studies on *Dictyostelium discoideum* myosin II motor domain fragments have been reported. These include mutations of the three conserved glycine residues (G680A, G684A, G691A) in the reactive thiol region (90), glutamates close to the nucleotide binding pocket (E459V and E476K) (225), all residues of the switch-1 loop mutated to alanine (226), all residues of the switch-2 loop mutated to alanine (227), mutations of the salt bridge residues in the switch-1 and switch-2 loops (E459R, R238E, and double mutant[8]) (230)

Mutational experiments in the P-loop have only been reported for smooth muscle myosin (Ser181, Lys185, Asn235, Ser236, Arg238) (231).

For the present study, only mutations of the residues of the phosphate tube are relevant. The basal ATPase rates for the mutants of the switch-1, switch-2, and P-loop residues are collected in Table 4.7.

[8]The identical mutations were also carried out for chicken gizzard smooth muscle heavy meromyosin (228). Additional mutations in chicken myosin correspond to D454A, I455M, and G457A (229).

mutant	rate [s^{-1}]	activity [%]	Ref.
HMM(WT)[a]	0.059	100	(231)
HMM(Ser181A)[a]	0.028	47.5	(231)
HMM(Ser181T)[a]	0.017	28.8	(231)
HMM(K185Q)[a]	0	0	(231)
WT[b]	0.062	100	(226)
N233A[b]	0.026	41.9	(226)
N235A[b]	0.046	74.2	(226)
S236A[b]	0.042	67.7	(226)
S237A[b]	0.018	29.0	(226)
R238A[b]	0.010	16.1	(226)
R238C[b]	0.007	11.3	(226)
R238H[b]	0.015	24.2	(226)
HMM(N235I)[a]	0.092	155.9	(231)
HMM(S236A)[a]	0.015	25.4	(231)
HMM(S236T)[a]	0.048	81.4	(231)
HMM(R238I)[a]	0	0	(231)
HMM(R238K)[a]	0.060	101.7	(231)
M765(WT)[c]	0.08	100	(230)
M765(R238E)[c]	$4{\cdot}10^{-4}$	0.5	(230)
M761(WT)[b]	∼0.85	100	(227)
M761(D454A)[b]	∼0.03	∼35	(227)
M761(I455A)[b]	∼0.25	∼300	(227)
M761(S456A)[b]	∼0.19	∼220	(227)
M761(G457A)[b]	∼0.025	∼30	(227)
M761(F458A)[b]	∼0.235	∼275	(227)
M761(E459A)[b]	∼0.02	∼25	(227)
S1(WT)[d]	0.1	100	(233)
S1(S456L)[d]	1.0	1000	(233)
M761-2R(WT)[e]	$1.5{\cdot}10^{-2}$	100	(225)
M761-2R(E459V)[e]	$5.9{\cdot}10^{-5}$	0.4	(225)
M765(E459R)[c]	0.015	18.8	(230)
M765(R238/E459R)[c]	0.06	75	(230)

[a]Measured for gizzard smooth muscle myosin at 25° as described in Ref. (232).
[b]From phosphate release rates at 25°.
[c]Measured using a linked enzyme assay.
[d]From release of labeled P_i at 30°.
[e]From single turnover experiments at 20° using *mant*ATP.

Table 4.7: Basal ATPase rates for *Dictyostelium discoideum* myosin II mutants.

4.5 Thermodynamics of the ATPase cycle in myosin

Reaction rates k can be converted into free energy barriers $\Delta G^{\ddagger}$ with Eyring's formula,

$$k = \frac{k_B T}{h} e^{-\frac{\Delta G^{\ddagger}}{RT}}, \tag{4.2}$$

where k_B denotes the Boltzmann constant, T the absolute temperature, h the Planck constant, and R the ideal gas constant. The values of the constants are given in Appendix B. Equilibrium constants K are related with reaction free energies $\Delta_R G$ by

$$\Delta_R G = -RT \ln K \tag{4.3}$$

Using Eqs. 4.2 and 4.3 the kinetic data given in Section 4.4.1 can be converted into free energies of the different reactant, transition, intermediate, and product states during the ATPase cycle. The resulting energetic data can then be compiled into a consistent free energy profile describing the myosin ATPase cycle.[9]

The free energy difference between the reactant (M + ATP) and product (M + ADP + P_i) states must be equivalent to the reaction free energy of ATP hydrolysis in aqueous solution under the given experimental conditions (compare Section 4.1). The kinetic measurements ATP hydrolysis in *Dictyostelium discoideum* myosin II have been made at pH between 7.0 and 8.0, temperatures between 20° and 25°, Mg^{2+} concentrations of 1mM to 5mM (corresponding to pMg from 3 to 2.3) and ionic strengths varying from 40mM to 120mM (compare Table 4.3). Corresponding to these reaction conditions, the reaction free energy for the overall reaction is $\Delta_r G'^0 \approx \Delta_r G'^0 (T = 298.15K, I = 0.25M, pMg = 2, pH = 7) = -7.4$ kcal/mol.[10] The predominant ATP species under these conditions is $MgATP^{2-}$ (see also Figs. 1.6 and 1.7 in Ref. (131)).

The overall free energy barrier for ATP hydrolysis by myosin is listed in Table 4.8 for the different *Dictyostelium discoideum* myosin II constructs. The free energy barriers range from 17.7 to 19.2 kcal/mol depending on the construct and experimental conditions.

[9] In principle it is possible to determine the enthalpic and entropic contributions to the free energy from the temperature dependence of the rates of the different steps. However, because the rate is also influenced by other factors such as the ionic strength and Mg^{2+} concentration and experiments have been conducted at different conditions, such an analysis cannot easily be done and is therefore omitted here. Measurements of enthalpic changes for the four principle steps of the myosin hydrolysis cycle have been reported for rabbit skeletal muscle myosin S1 (234). However, without knowing to which specific transitions between structural states this information corresponds to, it cannot be used for comparison with theoretical work that is made on a specific structural state.

[10] For comparison: A measurement of the reaction enthalpy for GTP hydrolysis by rabbit myosin yielded a value of $\Delta_r H = -5.3$ kcal/mol (T=25°, pH=8-9, 0.6M KCl, 1mM $CaCl_2$) (235).

construct	$\Delta G^{\ddagger\,a}$ [kJ/mol]	$\Delta G^{\ddagger}$ [kcal/mol]	I [mM]^b	ref.
M761	78.20	18.7	43	(81)
M761 W501+	78.57	18.8	43	(81)
M761-2R	76.89	18.4	≈55	(203)
M761-2R	74.25	17.7	≈55	(203)
MHF (=M864)	79.56	19.0	115	(220)
M754	79.56	19.0	115	(220)
D1dC (=M759)	79.24	18.9	n.d.	(46)
D1dC (=M759)	80.43	19.2	43	(218)

[a]From $k_{cat} = \frac{k_B T}{h} e^{-\frac{\Delta G^{\ddagger}}{RT}} \Leftrightarrow \Delta G^{\ddagger} = -RT \ln \frac{h k_{cat}}{k_B T}$
[b]Calculated from $I = \frac{1}{2} \sum_i c_i Z_i^2$. The following salts were considered: NaCl, KCl, $MgCl_2$

Table 4.8: Free energy barriers $\Delta G^{\ddagger}$ (in kJ/mol and in kcal/mol) and ionic strengths I for ATP hydrolysis by different *Dictyostelium discoideum* myosin II constructs. I is determined from the experimental conditions given in Table 4.3.

The free energies associated with the various steps of the ATPase cycle as derived for the different wild-type *Dictyostelium discoideum* myosin II constructs and the M761(W501+) mutant are listed in Table 4.9. Table 4.10 gives the free energies associated with the recovery-stroke transition and the hydrolysis step for M761(W501+).

property	M864	M754	M761	M781	M759	M761(501+)
$\Delta_R G_1{}^a$	-18.91 (-4.5)	-12.08 (-2.9)	-17.84 (-4.3)	-17.60 (-4.2)	-14.74 (-3.5)	-18.05 (-4.3)
$\Delta G_2^{\ddagger\,b}$	n.d.	n.d.	n.d.	n.d.	n.d.	≈ 57 (13.7)
$\Delta G_{-2}^{\ddagger\,c}$	n.d.	n.d.	n.d.	n.d.	n.d.	≈ 81 (19.4)
$\Delta G_6^{\ddagger}$	70.73 (16.9)	78.57 (18.8)	n.d.	n.d.	71.49 (17.1)	66.68 (15.9)
$\Delta G_{-6}^{\ddagger\,d}$	57.23 (13.7)	55.22 (13.2)	n.d.	n.d.	56.28 (13.5)	55.29 (13.2)

[a]From $\Delta_R G = -RT \ln K$ and
[b]From k_2.
[c]From k_{-2}.
[d]From the estimated k_{-6}

Table 4.9: Barrier heights in kJ/mol (kcal/mol) determined from kinetic data given in Table 4.5 Values that have not been determined are marked with n.d.

The overall reaction free energy together with the energetics of the different steps of the ATPase cycle can be compiled into a consistent free energy profile tabulated in Table 4.11 and shown in Fig. 4.4 for the M761(501+) construct, for which the available kinetic data is almost complete. The values used are $K_1 k_2 = 6.6 \cdot 10^5 \mathrm{M}^{-1}\mathrm{s}^{-1}$ and $k_2 \approx 400\mathrm{s}^{-1}$ from

property	value	estimate
$\Delta_R G_{3a}$	2.23 (0.5)	$\approx$ 3.9 (0.9)
$\Delta G^{\ddagger}_{3a}$	57.45 (13.7)	
$\Delta G^{\ddagger}_{-3a}$	55.23 (13.2)	
$\Delta_R G_{3b}$	-6.25 (-1.5)	$\leq$ -5.61 (-1.3)
$\Delta_R G_{3b}$	-10.64 (-2.5)	
$\Delta G^{\ddagger}_{3b}$	60.27 to 60.76 (14.4 to 14.5)	
$\Delta G^{\ddagger}_{-3b}$	66.51 to 71.49 (15.9 to 17.1)	

Table 4.10: Barrier heights and equilibrium free energies associated with the recovery stroke transition (Step 3a) and the hydrolysis step (Step 3b) in kJ/mol (kcal/mol) from kinetic data given in Table 4.6.

ref. (81); $k_{-2} = 2.3 \cdot 10^{-6}$ s^{-1} from ref. (236); and $K_{3a} = 0.4$, $k_{3a} = 350$ s^{-1}, $k_{-3a} = 870$ s^{-1}, $K_{3b} = 13$, $k_{3b} = 110$ s^{-1}, and $k_{-3b} = 8.5$ s^{-1} from ref. (209), $k_4 = 0.05$ s^{-1} from ref. (196), $k_6 = 7.9$ s^{-1} and $k_{-6}/K_7 = 1.4 \cdot 10^6$ Ms^{-1} from ref. (81). The rapid formation of the collision complexes from either the reactant or the product side of the profile was assumed to have a barrier of 2 kcal/mol. The value for k_{-2} has been inferred from measurements of the ATP binding constant using rabbit myosin II ($K_1K_2 = 3.25 \cdot 10^{11}$ M^{-1}, corresponding to a free energy of about $\Delta_R G_{1,2} = -15.4$ kcal/mol) (237; 238) rather than the upper bound for k_{-2} estimated from displacement experiments with the nucleotide analog AMP.PNP in *D. discoideum* ($k_{-2} \approx 0.02$ s^{-1}). Although this value differs from the rabbit value of k_{-2} by four orders of magnitude, the rabbit muscle myosin value is likely to be more appropriate since the differences between ATP binding to muscle (rabbit) and nonmuscle (*D. discoideum*) myosin II are expected to be smaller than the differences between ATP and AMP.PNP binding to nonmuscle myosin II.

The so-constructed profile shows an estimated value of $\Delta_R G_{4,5} = 1.9$ kcal/mol. This is within 0.5 kcal/mol of the experimental value of 1.4 kcal/mol corresponding to $K_{4,5} = 0.085$ M from ref. (196). This confirms the validity of the assumptions made in the construction of the profile and provides a rough estimate of the accuracy of the energy values given.

The key features with regard to the ATP hydrolysis step (Step 3b) extracted from the free energy profile are:

1. The prehydrolysis state M*.ATP has a free energy of $\approx$-14.8 kcal/mol with respect to the reactant state M + ATP.

2. The posthydrolysis state M*.ADP.P_i has a free energy of $\approx$-16.3 kcal/mol with respect to the reactant state M + ATP and is thus $\approx$1.5 kcal/mol lower in energy than the prehydrolysis state.

3. The experimental reaction free energy barrier for the hydrolysis step is $\approx$14.5 kcal/mol.

state	step	$\Delta G^{\ddagger}_{step}$	ΔG
M + ATP		0.0	0.0
(M+ATP)‡	1[a]	2.0	2.0
M.ATP	-1[b]	6.3	-4.3
(M.ATP)‡	2	13.7	9.4
M†.ATP	-2	24.7	-15.3
(M†.ATP)‡	3a	13.7	-1.6
M*.ATP	-3a	13.2	-14.8
(M*.ATP)‡	3b	14.4	-0.4
M*.ADP.P$_i$	-3b	15.9	-16.3
(M*.ADP.P$_i$)‡	4,5	18.9	2.6
M†.ADP + P$_i$	-4,5[c]	17.0	-14.4
(M†.ADP)‡ + P$_i$	6	15.9	1.5
M.ADP + P$_i$	-6	13.2	-11.7
(M.ADP)‡ + P$_i$	7[d]	6.3	-5.4
M + ADP + P$_i$	-7[e]	2.0	-7.4

[a]Low barrier estimated from the fact that the equilibrium of collision complex formation is fast.

[b]Estimated from $\Delta_R G_1 = -4.3$ kcal/mol and $\Delta G^{\ddagger}_1 = 2$ kcal/mol.

[c]Estimated from the energy difference between the energies of (M*.ADP.P$_i$)‡ as derived from the reactant end of the cycle and the energy of M†.ADP + P$_i$ as derived from the product end of the cycle.

[d]Estimated to be identical to $\Delta G^{\ddagger}_{-1}$.

[e]Estimated to be identical to $\Delta G^{\ddagger}_{1}$.

Table 4.11: Free energies in kcal/mol of the different states during the ATPase cycle of M761.W501+ relative to the reactant state. Up to the posthydrolysis state M*.ADP.P$_i$ the profile was built starting from the reactant energy by successively adding and subtracting the measured activation energies $\Delta G^{\ddagger}_{step}$ as collected in Tables 4.9 and 4.10. The second part of the table was built from the product state in the same manner. The energy of the product state relative to the reactant state is given by the reaction free energy of ATP in aqueous solution at the given reaction conditions (here -7.4 kcal/mol, see main text).

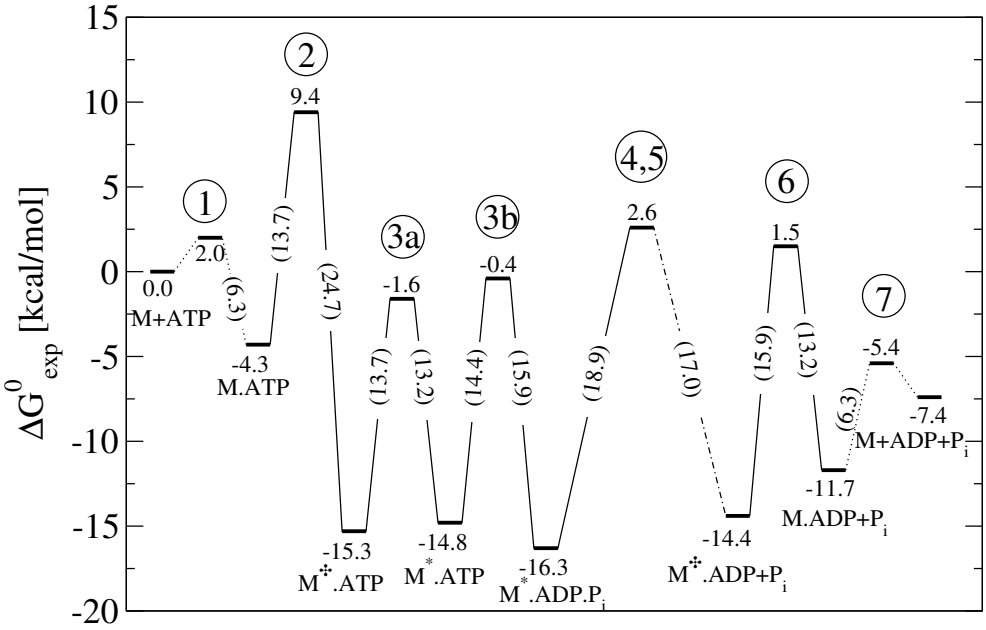

Figure 4.4: Free energy profile corresponding to the enzymatic cycle shown in Fig. 4.3 at standard conditions (T=293K, concentrations of all species 1M) constructed from experimental measurements (see Methods and Table 4.11). Under physiological concentrations (differing from the 1M standard state), under steady-state conditions, and in presence of actin the free energy profile of the contraction cycle is significantly different (236).

4.6 Mechanistic considerations on the ATP hydrolysis step in myosin

4.6.1 Experimental observations

Apart from the extensive kinetic measurements on the ATPase cycle of myosin (as discussed in detail in Section 4.4) a number of experimental studies and one theoretical study have been reported that aim at elucidating details on the mechanism of ATP hydrolysis in myosin.

A number of isotope exchange studies have been made (reviewed in Ref. (239)). Early measurements of isotopic exchange between oxygens from ^{18}O labeled water molecules from the solvent and P_i bound to myosin showed that more than one oxygen from solvent was incorporated into P_i. Thus it was established that hydrolysis must be reversible and that solvent molecules can enter and leave the active site during hydrolysis, thus enabling the observed exchange process (240). Moreover, the results indicate that the active site is flexible enough (or leaves enough space) that the P_i can tumble before rebinding to ADP, such that a different oxygen is released upon the dissociation of a water molecule from P_i than had been incorporated from the attacking water. Alternatively, it must be assumed that water can enter and leave the γ-phosphate moiety from different positions. Assuming a pentavalent intermediate, water attachment and detachment can only occur from the axial positions of the trigonal bipyramid. If, as is observed experimentally, different positions can incorporate ^{18}O a pseudorotation mechanism must be postulated.

Because the temperature dependence of ATP hydrolysis and oxygen exchange reactions were found to be different and because the increased P_i concentration leads to a decrease of the ATPase rate but to an increase of the rate of exchange it was postulated that both reactions happen in structurally distinct but intimately related sites that could be temporarily formed or abolished by conformational fluctuations of myosin (241). This was ruled out based on experiments that clearly separated exchange reactions occurring in the medium and exchange reactions occurring when P_i is strongly bound to myosin (242). It was found that an exchange reaction requires the presence of a divalent cation and a nucleotide and that such conditions are not present in a protein site other than the active site.

Disruption of the hydrolysis reaction of myosin by quenching with $HCLO_4$ containing ^{18}O-labeled water allows to probe the nature of the intermediate or transition state. If the attacking (unlabeled) water enters the reaction upon formation of the transition state quenching will lead to reaction products that do not contain ^{18}O. If, on the other hand, water enters upon decay of the transition or intermediate state ^{18}O must be found in the products after quenching. This would be the case for a phosphorylated myosin intermediate that requires to be hydrolyzed in a second step or for a loose transition state involving

a metaphosphate moiety as found in dissociative hydrolysis mechanisms. However, no ^{18}O was found in the reaction products, thus indicating that myosin is not transiently phosphorylated and that the transition/intermediate state is tight (243), suggesting an associative mechanism.

Isotope exchange must be coupled to ATP hydrolysis *per definitionem* and thus occurs in the conformation of myosin characterized by increased intrinsic protein fluorescence (see Section 4.4). This conformation resembles a tight binding state for both ADP and P_i once the hydrolysis has taken place. It seems unlikely that tightly bound P_i can freely tumble in the active site, thus ruling out a tumbling mechanism to explain isotope exchange. Thus it was proposed that the only mechanism that can explain both isotope exchange results and a tight, associative-like transition state is a pseudo-rotation mechanism (244; 245). Pseudo-rotation in the hydrolysis of phosphate esters has been shown to occur subject to the constraints that equatorial positions are occupied by alkyl groups and a five-membered ring spans one equatorial and one apical position in trigonal bipyramids (147).

In light of the structural rigidity of the myosin binding site a pseudo-rotation mechanism seems, however, unlikely. A pseudo-rotation mechanism implies that a water can come in (and be released) at two different positions. However, the geometry of the binding site suggests that water attacks only in-line with the ADP leaving group and no other incoming or leaving position is visible. This, however would require further studies since the binding site consists of flexible loops that may change their conformation so a to allow water to attack from a 90° angle with respect to the P_γ-O_{bridge} bond.

Exchange can also be explained by rotational flexibility of the γ-phosphate moiety of ATP when bound to myosin. That such a scenario seems likely is suggested by exchange experiments that started from labeled ATP and found that 75% of terminal oxygens had been exchanged 2s after mixing the labeled ATP with myosin in unlabeled water (246). Another indication of rotational freedom is the observed equivalence of all four P_i oxygens to be exchanged (247). Rapid quench flow techniques were used to measure the time-course of oxygen exchange (*i.e.*, number of unlabeled oxygens incorporated into P_i starting from labeled ATP). The resulting distribution of species agreed well with the distribution predicted from a kinetic model of the ATP hydrolysis step that assumed a single step reaction without intermediate (248).

Combining ^{18}O labeling at the γ-nonbridge and the β-γ-bridge position with ^{17}O labeling at the β-nonbridge positions allowed to simultaneously determine rotational freedom of the β- and γ-phosphate moieties of ATP during hydrolysis. No exchange of β-nonbridge oxygens and β-γ-bridge oxygen was observed, indicating that ADP remains tightly bound without undergoing rotation throughout the course of the reaction. The labeled γ-oxygens were, however, exchanged with unlabeled bulk water oxygens, thus confirming significant rotational freedom of either the terminal γ-phosphate moiety or the cleaved off but not yet released P_i (249).

If one terminal oxygen from ATP is replaced by sulfur the resulting ATPγS is slowly

hydrolyzed with the hydrolysis step being rate-limiting. Thus, no reversal of the hydrolysis event is expected when using ATPγS and indeed no isotope exchange was observed (250). This made it possible to use a [$\beta\gamma$-^{18}O;γ-^{18}O$_1$]ATPγS substrate (*i.e.*, an ATP analog with three distinguishable atoms in the γ-nonbridge positions) in ^{17}O-labeled bulk water to determine the stereochemistry of ATP hydrolysis by myosin. Inversion of configuration was found (250), thus confirming the exclusion of a pseudorotation mechanism or of a phosphorylated protein intermediate.

A mechanism with more associative than dissociative character was suggested by Raman difference spectroscopic measurements on the myosin.MgADP.vanadate complex that is considered to be a transition state analog for phosphoryl transfer reactions in enzymes. The experiments showed a decrease in bond order of the equatorial γ-oxygens and an increase in bond order of the axial γ-oxygens of the pentavalent vanadate upon binding (251).

4.6.2 Summary

The experimental findings can be summarized as follows:

1. Hydrolysis proceeds *via* a tight transition state that already includes the attacking water molecule and is more associative than dissociative.
2. Hydrolysis proceeds with inversion of configuration.
3. Hydrolysis proceeds *via* direct in-line attack of the lytic water molecule without a phosphorylated protein intermediate.
4. Hydrolysis is likely to proceed in a single-step reaction without stable intermediate.
5. Either the terminal γ-phosphate moiety in ATP or the already cleaved off but not yet released P_i possesses significant rotational freedom.
6. The ADP moiety including all nonbridging β-oxygens remains positionally constrained throughout the course of the reaction.

4.6.3 Hydrolysis mechanisms proposed

As seen in the available crystal structure of the motor domain of myosin (see Section 3.4) there is no residue that could fulfill the role of a general base in abstracting a proton from the attacking water as has been suggested for Gln61 in Ras. Thus, it was suggested that water might function as the nucleophile with direct transfer of its proton to the γ-phosphate. An alternative suggestion involves a proton relay mechanism in which Ser236

serves as an intermediate in the transfer of one proton from the nucleophilic water to to the γ-phosphate (57; 252).

A semiempirical PM3 molecular orbital study on an active site model of reactant state of myosin revealed that negative charge is transferred from Mg.ATP to the surrounding amino acids (253). However, no residues from the switch-2 loop were included in the model, thus neglecting important interactions. Based on geometric considerations the authors suggest two additional reaction mechanisms. In the first, Ser181[11] acts as an intermediate in the proton transfer from the lytic water to the γ-phosphate in a fashion similar to the Ser236 relay mechanism previously proposed. In the second, Lys185 accepts a proton from the lytic water. The latter, however, requires the sidechain of Lys185 to be deprotonated in the reactant state which seems unlikely due to the basic character of the sidechain of lysine.

A vacuum B3LYP/6-31G(d,p)//HF/6-31G(d,p) study on an active site model consisting of Mg.ATP.$(H_2O)_5$ and the sidechains of Lys185, Arg238, and Glu459 was reported (255). Keeping the sidechains fixed, an activation energy of 41.97 kcal/mol was found for the direct attack mechanism. From their simulations the authors conclude that Lys185 plays a role in maintaining the structure of the phosphate moiety and that the salt-bridge residues Arg238 and Glu459 help to position the lytic water. However, since they don't give any comparison with experiment and do not include any additional sidechains into their model this study is unsuited for comparison between different mechanisms, even if aiming only at a qualitative comparison.

Classical molecular dynamics and Car-Parrinello molecular dynamics (CPMD) studies on Mg.ATP binding to both C/O and C/C structures showed that the C/O structure has a higher affinity for ATP than the C/C structure, as the interactions with the protein in the C/C structure distort ATP towards a transition state (256). This corresponds well with that the $M^{\dagger}$.ATP state is lower in energy than the M^{*}.ATP state (see free energy profile in Fig. 4.4). Ca.ATP binding to the C/O structure was also examined revealing significant differences in Ca^{2+} and Mg^{2+} coordination and, consequently, differences in active site water structure. The higher Ca.ATPase rate than Mg.ATPase rate was attributed to weaker binding of P_i in case of Ca^{2+}. In their calculations the protonated Lys185 forms a stronger hydrogen bond to one of the γ-phosphate oxygens than it does with a β-phosphate oxygen. Thus, the authors suggest that Lys185 may act as a general acid.

The first computational study combining quantum mechanical with molecular mechanical techniques was reported in 2004 (257). The QM region included the triphosphate moiety of ATP, Mg^{2+} coordinated by the sidechains of Thr186 and Ser237 and two water molecules, the attacking water molecule and the sidechain of Ser236. All protein atoms were included in the model as molecular mechanical atoms. An 18 Å water sphere was centered on the active site so as to include solvent effects. Additional solvent screening was modeled by scaling the charges on charged sidechains. QM[B3LYP/6-31+G(d,p)//HF/3-21+G]/MM

[11] Ser181 in the P-loop can be selectively esterified (254). Thus it is possible to introduce a fluorescent probe into the P-loop.

reaction path calculations were performed for the direct attack mechanism and the Ser236 proton relay mechanism in the C/C conformation of myosin and for the direct attack mechanism in the C/O conformation. The Ser236 relay mechanism was found to proceed *via* a concerted proton transfer with a barrier of 22.9 kcal/mol and a O_W-P_γ distance of 1.92 Å towards a pentacovalent intermediate characterized by an energy of 21.2 kcal/mol and a O_W-P_γ distance of 1.79 Å. The intermediate decays to the final product with corresponding energy of -0.4 kcal/mol *via* a second transition state with an energy of 25.9 kcal/mol. The break of the P_γ-$O_{\beta\gamma}$ bond is induced by rotation of the newly formed OH group such that the proton is directed towards ADP in the product structure. Two different direct attack paths were found differing in onto which γ-oxygen the proton is transferred. In the first a transition state with an energy of 38.8 kcal/mol and a O_W-P_γ distance of 2.0 Å decays to a pentacovalent intermediate characterized by an energy of 29.0 kcal/mol and a O_W-P_γ distance that is 0.05 Å shorter than in the first intermediate. A second transition state (energy 32.3 kcal/mol, O_W-P_γ distance 1.65 Å) is passed that again involved rotation of the newly-formed hydroxyl group. In the product state (energy -5.1 kcal/mol) the proton has been transferred to ADP. The second direct attack pathway started from a different reactant state in which the lytic water molecule reoriented so as to hydrogen-bond to the other γ-oxygen. The same intermediate state as for the Ser236 pathway is reached *via* a transition state with an energy of 30.8 kcal/mol and a O_W-P_γ distance of 2.25 Å. The direct attack hydrolysis in the C/O structure of myosin proceeded *via* a barrier of more than 50 kcal/mol, thus clearly indicating that this myosin conformer is not suitable for catalysis.

Compared to the previously reported active site model (255) the activation barriers are much closer to the experimental barrier of about 14.5 kcal/mol. However, they still deviate significantly from experiment. Thus, additional factors not taken into consideration in the reported calculations must influence the hydrolysis reaction.

Based on the available mutational data, an additional mechanism was suggested in analogy to the proposed GTP hydrolysis mechanism in G-proteins. In the "two water hypothesis" (258) a second active-site water molecule held in place by Glu459 and Gly457 is activating the attacking water, taking the role of Gln204 in the GTP hydrolyzing protein $G_{i\alpha}$ (259). The proposed mechanism is dissociative-like, with Asn233 playing the role of Arg178 in $G_{\imath\alpha}$ in stabilizing the negative charge on the bridge oxygen in the transition state.

Chapter 5

Questions addressed in the current work

In the present work, reaction path calculations on the ATP hydrolysis step in myosin are performed using a combined quantum mechanical/molecular mechanical approach. Based on the calculations, insight into the unresolved questions regarding ATP hydrolysis in myosin and its coupling to the mechanical events shall be provided.

The following questions shall be addressed by the simulations:

1. What is the degree of associativity of the myosin-catalyzed ATP hydrolysis reaction (see Eq. 4.1)?
2. Can the findings of the isotopic exchange experiments be rationalized?
3. Can the available mutational data be rationalized?
4. Can different proposals for the mechanism of ATP hydrolysis be verified or ruled out?
5. What is the role of Mg^{2+}?
6. What are the similarities and dissimilarities between the hydrolysis mechanisms of myosin, F_1-ATPase and the GTPase Ras?
7. Which early events of the phosphate exit can be suggested based on the structural details of the chemical hydrolysis of ATP? How can these events be related to the proposed back door or trap door exit mechanisms?

Chapter 6

Summary: Introduction

In this introductory part, the current status of research on ATP hydrolysis in the motor protein myosin was reviewed. Molecular motors in general are proteins that hydrolyse a nucleoside triphosphate molecule and convert the energy into mechanical work. A well-described family of motors is the family of myosins. Among these, conventional myosins as represented by *Discyostelium disoideum* myosin II are the best characterized and are used in the present work.

The sliding filament model was introduced as an early model explaining how the myosin filament moves along the actin filament driven by cyclic hinge-like bending of the myosin head relative to its tail, which is coupled to the attachment and detachment of the myosin heads to and from actin. The myosin head domain only is capable of hydrolysing ATP and displacing along actin. Therefore, many biochemical and crystallographic studies have been conducted using the head domain only. The ATP binding pocket is formed by three loops, the P-loop, switch-1, and switch-2, whose amino acid sequence is highly conserved among all members of the myosin family. These loops are also found in G-proteins, which suggests that myosins and G-proteins share a common hydrolysis mechanism. Both switch-1 and switch-2 can either adopt an open or a closed conformation, thus modulating the angle between the head and tail domains as well as the affinity of myosin for actin. Only the closed/closed state is competent for hydrolysis.

ATP hydrolysis both in aqueous solution and in myosin was reviewed. ATP hydrolysis in aqueous solution is not a single reaction. Rather, it is a sum of reactions starting from different protonation states of ATP. In the presence of Mg^{2+} and under the experimental conditions used to study ATP hydrolysis in myosin the predominant species is $Mg.ATP^{2-}$. The free energy of the hydrolysis reaction is -7.4 kcal/mol and the activation free energy is about 29 kcal/mol. In principle, phosphate hydrolysis can proceed *via* a continuum of mechanisms between a purely dissociative mechanism (with a metaphosphate intermediate) or an associative mechanism (with a pentavalent phosphorane intermediate). Both mech-

anisms can explain the thermodynamics of methylphosphate (the smallest possible model compound for ATP) hydrolysis in aqueous solution, so that enzymes including myosin can in principle select either mechanism. Myosin reduces the activation free energy to about 14.5 kcal/mol and modulates the reaction free energy to about -1.5 kcal/mol. Hydrolysis proceeds with inversion of configuration in the myosin active site. It is likely that a tight associative-like transition state is passed in a single-step reaction, initiated by in-line attack of the lytic water molecule. The terminal γ-oxygens exchange readily with the bulk solvent, indicating rotational freedom of the γ-phosphate moiety. In contrast, the β-oxygens remain positionally constrained. A number of ATP hydrolysis mechanisms have been proposed that are consistent with these experimental findings. These suggestions differ in the activation mechanism of the lytic water molecule that may either transfer its proton directly to ATP or *via* the sidechain of an adjacent amino acid residue. It has been suggested that Ser181, Lys185, Ser236, or a second water molecule participates in the reaction.

The reaction path simulations presented in this thesis give deeper insight into the mechanism of myosin-catalyzed ATP hydrolysis by investigating the degree of associativity, by rationalizing the isotope exchange experiments and mutational studies, by verifying or rejecting previous mechanistic proposals, by clarifying the role of Mg^{2+}, and by comparing the hydrolysis mechanism in myosin with the one in other nucleoside triphosphate hydrolysing proteins. Moreover, details of the chemo-mechanical coupling between the enzymatic and the motor cycles of myosin shall be elucidated.

Part II

Methods

Opening: Methods

In this part the methods used for the present work are described. Here, the newly-developed method to treat solvent effects in computer simulations of enzyme-catalyzed reactions is introduced in a self-contained chapter. This chapter is embedded in the description of the computational methods used to determine possible reaction paths of ATP hydrolysis in myosin. Since a reaction involves bond making and breaking processes, quantum chemical simulation techniques must be used to describe the process. Here, Hartree-Fock and DFT calculations are used. However, the number of atoms present in an enzymatic reactive system far exceeds the size that can be treated quantum mechanically, making it necessary to treat the enzymatic reaction environment with classical molecular mechanics. Thus, quantum mechanical methods must be combined with molecular mechanical techniques. The QM and MM methods and the combination of them are described in Chapter 7. Particular emphasis is given to procedures that can be used to link quantum mechanical and classical mechanical regions across a chemical bond. A generally-applicable link-atom method for this purpose has been implemented and tested. Chapter 8 describes the theory that underlies the determination of minimum-energy pathways that can be used to describe chemical reactions. Enzyme catalysis in general takes place in aqueous solution. Thus, solvent effects must be considered if aiming at an appropriate description of the enzymatic reaction. However, for a variety of reasons, current methods that model the solvent either explicitly or implicitly cannot be used in combination with QM/MM reaction path simulations. The newly-developed method "Non-Uniform Charge Scaling" (NUCS) fills this gap. NUCS is a generally-applicable, approximate method that can be used to model solvent screening effects in macromolecular calculations. Chapter 9 describes NUCS and test applications thereof. The methodological part is completed by a detailed description of the choice and preparation of the myosin model system and of the computational details of the reaction paths calculations of ATP hydrolysis in myosin in Chapter 10.

Chapter 7

Combining quantum mechanical and molecular mechanical calculations

This chapter describes the approach used in this thesis to combine quantum mechanical with molecular mechanical simulation techniques. Section 7.1 outlines the theoretical aspects of QM/MM schemes. Section 7.2 describes in detail how the link between QM and MM regions can be modelled if the boundary crosses a chemical bond.

7.1 Theoretical considerations

7.1.1 Molecular mechanics (MM)

In molecular mechanical (MM) force fields the potential energy of the system under consideration is expressed as an analytical function of the 3N coordinates of the N atoms present (260). The total MM energy, E_{MM}, is written as the sum of bonded, E_b, and nonbonded, E_{nb}, energy terms,

$$E_{MM} = E_b + E_{nb}. \tag{7.1}$$

E_b contains energy terms due to bond stretching, angle bending, and dihedral angle rotation,

$$E_b = \sum_{bonds} k_b(l - l_0)^2 + \sum_{angles} k_a(\alpha - \alpha_0)^2 + \sum_{dihe} \sum_{n} \frac{V_n}{2}[1 \pm \cos(n\tau)], \tag{7.2}$$

where k_b and k_a are the force constants for bond stretching and angle bending, l and θ denote bond lengths and angles, l_0 and θ_0 are the equilibrium bond lengths and equilibrium angles, the integer n is the periodicity of the torsional barrier, V_n is the associated barrier height, and τ is the torsion angle. Force constants, equilibrium values, and barrier heights must be parametrized.

E_{nb} models the nonbonded interactions *via* van der Waals and Coulomb terms,

$$E_{nb} = \sum_{i,j} \left(\frac{-A_{ij}}{r_{ij}^6} + \frac{B_{ij}}{r_{ij}^{12}} \right) + \sum_{i,j} \frac{q_i q_j}{\epsilon r_{ij}}, \tag{7.3}$$

where the sum runs over all unique atom pairs (i, j), A_{ij} and B_{ij} denote the atom-pair-specific attractive and repulsive van der Waals parameters, r_{ij} the distance, q_i and q_j the parametrized partial atomic charges on the atoms, and ϵ the dielectric constant. The van der Waals coefficients are chosen such that an energy minimum V_{ij} occurs at a distance r_{ij}^0, as expressed by

$$A_{ij} = 2(r_{ij}^0)^6 V_{ij} \tag{7.4}$$

and

$$B_{ij} = (r_{ij}^0)^{12} V_{ij}. \tag{7.5}$$

Combination rules are used to obtain V_{ij} and r_{ij}^0 from parametrized atom-specific properties ϵ_i and r_i (where r_i is the van der Waals radius for atom i that is used in the force field):

$$V_{ij} = \sqrt{\epsilon_i \epsilon_j} \tag{7.6}$$

$$r_{ij}^0 = (r_i + r_j). \tag{7.7}$$

Because bonds are modeled with a harmonic potential, bond elongation leads inevitably to an increase in energy and a bond can never be broken. Thus, chemical reactions cannot be modeled with a molecular mechanical force field.

7.1.2 Quantum mechanics (QM)

In quantum mechanical calculations and within the framework of the Born-Oppenheimer approximation the electronic energy for a given position of the nuclei has to be obtained. In *ab initio* wave mechanics, the energy is determined by iterative solution of the eigenvalue problem of the electronic Schrödinger equation,

$$\hat{H}\Psi = E\Psi, \tag{7.8}$$

where $\hat{H}$ is the Hamilton operator, Ψ the electronic wave function, and E the energy. Only the electronic ground-state wave function Ψ_0 and associated energy E_0 are of interest in the present work. The electronic Hamilton operator describes the kinetic and potential energy of the N electrons in the field of the M fixed nuclei. It reads (in atomic units)

$$\hat{H} = -\sum_{i=1}^{N}\frac{1}{2}\nabla_i^2 - \sum_{i=1}^{N}\sum_{A=1}^{M}\frac{Z_A}{r_{iA}} + \sum_{i=1}^{N}\sum_{j>i}^{N}\frac{1}{r_{ij}} + \sum_{A=1}^{M}\sum_{B>A}^{M}\frac{Z_A Z_B}{r_{AB}}, \tag{7.9}$$

where ∇ denotes the Nabla operator, Z_A the nuclear charge on the nucleus A, r_{iA} the distance between electron i and nucleus A, r_{ij} the distance between electrons i and j, and r_{AB} the distance between the nuclei A and B. The fourth term in the above equation is the nuclear repulsion term. It is a constant for a given nuclear configuration and thus does not participate in the solution of the electronic Schrödinger equation.

The N-electron wave function Ψ can be written as a Slater determinant[1] based on N one-electron wave functions ψ, *i.e.*, the molecular orbitals (MO),

$$\Psi = |\psi_1\psi_2...\psi_N> . \tag{7.10}$$

In Hartree-Fock theory, the MOs are varied such that the ground-state energy expectation value

$$E_0 =< \Psi_0|\hat{H}|\Psi_0 > \tag{7.11}$$

is minimized while using a single Slater determinant to approximate the many-electron wave function and subject to the constraint that the MOs remain orthonormal.

[1]For details on the notation see standard textbooks on quantum chemistry, *i.e.*, Refs. (261; 262)

Molecular orbitals ψ can be written as the product of a spin orbital χ and a spatial orbital ϕ. For a closed shell system it suffices to vary the spatial orbitals which can then be occupied by two electrons. This leads to the formulation of the restricted Hartree-Fock equation,

$$\hat{f}_i \phi_j = \epsilon_j \phi_j, \tag{7.12}$$

where $\hat{f}_i$ is the Fock operator for the i-th electron and ϕ_j is the j-th molecular orbital with orbital energy ϵ_j. Because the electrons are indistinguishable, i can be replaced by 1. With this, the Fock operator reads

$$\hat{f}_1 = \hat{h}_1 + \sum_a^{\frac{N}{2}} 2\hat{J}_1^a - \hat{K}_1^a, \tag{7.13}$$

where

$$\hat{h}_1 = -\frac{1}{2}\nabla^2 - \sum_A \frac{Z_A}{r_{1A}} \tag{7.14}$$

is the one-electron part of the Fock operator. $\hat{J}_1^a$ is the Coulomb operator defined by its operation on the spatial one-electron molecular orbital ϕ^b that is occupied by electron 1 (as indicated by the subscript 1),

$$\hat{J}_1^a \phi_1^b = \left[\int dx_2 \phi_2^{a*} \frac{1}{r_{12}} \phi_2^a \right] \phi_1^b \tag{7.15}$$

where the integral is over the full space and x_2 are the coordinates of electron 2 occupying the orbital ϕ^a. The integral equals the expectation value of the Coulomb operator $\frac{1}{r_{12}}$ acting on the molecular orbital ϕ^a that is occupied by electron 2. It thus represents the average Coulomb potential seen by electron 1 due to electron 2. $\hat{K}_1^a$ is the exchange operator defined by its operation on the spatial one-electron molecular orbital ϕ^b that is occupied by electron 1,

$$\hat{J}_1^a \phi_1^b = \left[\int dx_2 \phi_2^{a*} \frac{1}{r_{12}} \phi_2^b \right] \phi_1^a. \tag{7.16}$$

In contrast to the Coulomb integral the exchange integral does not have a classical interpretation.

Spatial molecular orbitals ϕ are written as linear combinations of atomic orbitals η (LCAO),

$$\phi = \sum_{i=1}^{N_\eta} c_i \eta_i \tag{7.17}$$

and can be varied by variation of the coefficients c_i. N_η is the number of atomic orbitals used as basis functions in the expansion.

Once the electronic wave function Ψ is known the values of a number of different observables $\hat{O}$ can be determined by calculation the expectation values $< \Psi|\hat{O}|\Psi >$. In the present work, however, only the ground-state energy is considered (see Eq. 7.11).

Instead of expressing the ground-state energy as an expectation value it can also be determined from the electron density ρ. The existence of a one-to-one relationship between the electron density and the energy of a system has been proven by Hohenberg and Kohn (263). This forms the basis of Density Functional Theory (DFT). The functional $E[\rho]$ that relates the electron density to the energy can, unfortunately, not be mathematically rigorously formulated. It can, however, be approximated. In analogy to Hartree-Fock theory the functional is divided into different energy terms,

$$E[\rho] = T[\rho] + E_{ne}[\rho] + J[\rho] + E_{xc}[\rho], \tag{7.18}$$

where $T[\rho]$ denotes the kinetic energy, $E_{ne}[\rho]$ the electrostatic interaction energy between nuclei and electrons, $J[\rho]$ the Coulomb interaction between electrons, and $E_{xc}[\rho]$ the exchange-correlation functional. Approximating the electron density in terms of one-electron functions ϕ_i (Kohn-Sham orbitals),

$$\rho(r) = \sum_{i=1}^{N} |\phi_i(r)|^2 \tag{7.19}$$

and assuming that the electrons are non-interacting, the kinetic energy can be written as

$$T[\rho] = \sum_{i=1}^{N} < \phi_i| - \frac{1}{2}\nabla|\phi_i > . \tag{7.20}$$

The Coulomb terms $E_{ne}[\rho]$ and $J[\rho]$ are expressed in their classical formulations. The exchange-correlation functional $E_{xc}[\rho]$ is the term that contains all missing parts, *i.e.*, the kinetic correlation energy, the potential correlation energy, and the exchange energy.[2] Since DFT thus includes *all* energy terms necessary to determine the exact energy of the system, it is in principle an exact method. However, the mathematical formulation of the exchange-correlation functional remains elusive. Different DFT methods employ different approximations for $E_{xc}[\rho]$. A frequently-used functional is the B3LYP functional (265; 266; 267) that is employed for all DFT calculations in this thesis. Once a functional is given, the Kohn-Sham orbitals and thus the electron density is determined numerically by a variational approach similar to the one used in Hartree-Fock calculations.

7.1.3 QM/MM calculations

Since it was suggested in 1976 to combine QM and MM calculations (268) to study enzyme-catalyzed chemical reactions a large number of studies have utilized this approach (reviewed in Ref. (269)). In QM/MM calculations a small region of interest is modeled quantum mechanically whereas the major part of the system is treated molecular mechanically as illustrated in Fig. 7.1. Combining QM and MM techniques allows to simulate bond breaking and making processes while taking the full complexity of a protein system in aqueous solution into consideration.

The QM/MM energy is modeled as the sum of the QM energy, the MM energy, and a QM/MM interaction term (270),

$$E_{tot} = E_{QM} + E_{MM} + E_{QM/MM} \tag{7.21}$$

When using a molecular orbital description for the quantum mechanical region (as is done in this thesis) the QM energy can be calculated as outlined in Section 7.1.2. The MM energy is determined as in Section 7.1.1. The QM/MM interaction is readily modeled if the boundary between QM and MM regions does not cross a chemical bond. It is described by the Hamiltonian $\hat{H}_{QM/MM}$

$$\hat{H}_{QM/MM} = -\sum_{i=1}^{N}\sum_{M=1}^{N_M}\frac{q_M}{r_{iM}} + \sum_{A=1}^{N_A}\sum_{M=1}^{N_M}\frac{Z_A q_M}{r_{AM}} + \sum_{A=1}^{N_A}\sum_{M=1}^{N_M}\left[\frac{A_{AM}}{r_{AM}^{12}} - \frac{B_{AM}}{r_{QM}^{6}}\right], \tag{7.22}$$

where N is the number of electrons, N_M the number of MM atoms, N_A the number of QM atoms (nuclei), R_{iM} the distance between electron i and MM atom M, r_{AM} the distance

[2] The correlation energy in wave mechanics is defined as the difference between the exact energy and the corresponding Hartree-Fock value, and the exchange energy is the total electron-electron repulsion minus the Coulomb energy (264, therein p. 180).

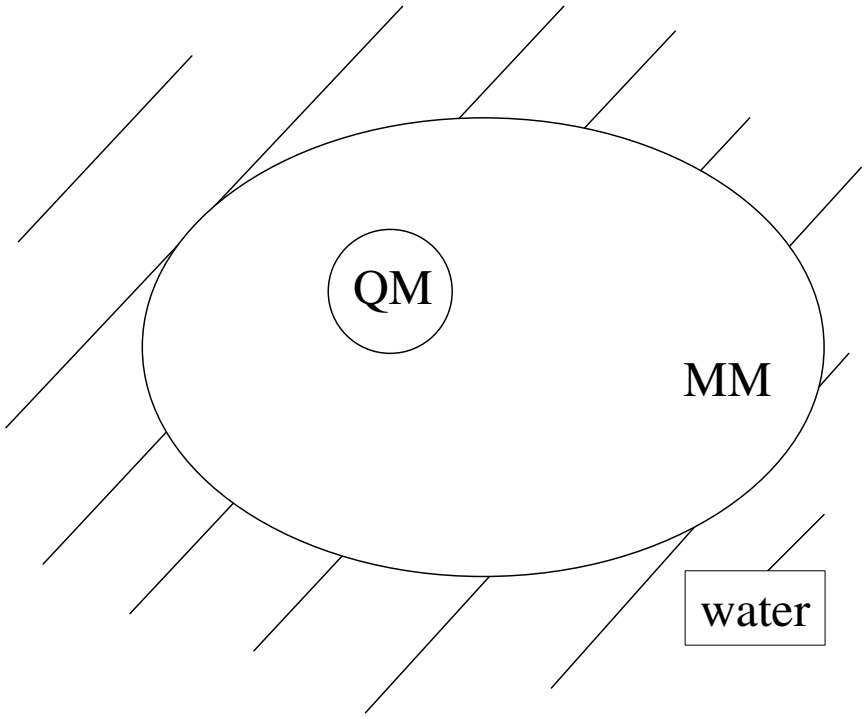

Figure 7.1: Partitioning of a solvated protein simulation system into quantum mechanical (QM) and molecular mechanical (MM) regions.

between QM atom A and MM atom M, Z_A the nuclear charge on QM atom A, q_M the partial atomic charge on MM atom M, and A_{AM} and B_{AM} the parameters describing the van der Waals interactions between QM atom A and MM atom M.

The first term in Eq. 7.22 describes the interaction of the electrons in the QM region with the partial atomic charges of the MM region. It is readily incorporated into the one-electron Hamiltonian (see Eq. 7.14) of the quantum mechanical region. The second term in Eq. 7.22 describes the Coulomb interaction between the QM nuclei and the partial atomic charges on the MM atoms. For a given position of all atoms it is a constant. The third term in Eq. 7.22 introduces a van der Waals attraction/repulsion. This is necessary to prevent QM atoms from intruding the MM region. It can be modeled classically after assigning parametrized values for ϵ_r and r_i onto the QM atoms (see Eqs. 7.4 to 7.7).

If the boundary between the QM and the MM region crosses a chemical bond additional terms must be introduced so as to maintain the correct geometry. How this can be achieved is discussed in detail in Section 7.2.

Apart from the fact that a QM/MM scheme allows to model chemical reaction in an enzymatic environment it can also be used to simulate phenomena that include exited states such as the calculation of redox potentials. Moreover, quantum effects that are

involved in reaction dynamics (*i.e.* tunneling) can in principle be described.

7.2 Linking QM and MM regions across a boundary chemical bond

Computational investigations of enzymatic reactions can be done with combined quantum mechanical/molecular mechanical (QM/MM) calculations in which typically a small part of the system under investigation is treated quantum mechanically, whereas the larger part is treated classically. If the QM/MM boundary cuts a chemical bond, it is necessary to join the quantum and classical regions together. To achieve this a number of methods have been reported. Among these is the link-atom approach in which the QM region is saturated by dummy hydrogens. Here, the link-atom approach is implemented in a series of patches for CHARMM. The usefulness of the approach is approved by showing that the error made by introducing a link atom is smaller than the error intrinsic to treating a large part of the system classically rather than quantum mechanically.[3]

7.2.1 Introduction

Enzymes are remarkable catalysts that speed up chemical reactions by up to 10^{21}-fold (149). Explanation of the efficiency of enzyme catalysis relies on a comparison of the reaction mechanism within the active site of the enzyme with that in solution (272; 273). The reaction mechanism and the role of the enzymatic environment in determining the mechanism cannot easily be determined using only experimental techniques, and computational methods are playing an increasingly important role in elucidating details of enzymatic function.

Computational investigations of enzyme-catalyzed reactions have become possible by combining quantum mechanical (QM) with molecular mechanical (MM) methods (269; 274), an approach that was suggested almost 30 years ago (268). In QM/MM calculations a small part of the overall system, *i.e.*, that in which the reaction proceeds, is modeled quantum mechanically. The remainder of the protein and the surrounding solvent is modeled using molecular mechanics, thus allowing the full complexity of environment to be considered at reasonable computational expense. QM/MM schemes have been reported that model the QM part by means of a molecular orbital (MO) description, as is focused on here, or utilizing valence bond (VB) descriptions (275; 276; 277).

Within the molecular orbital (MO) framework, the total energy of a QM/MM system is determined as the sum of the QM energy of the QM region, the MM energy of the MM

[3]The content of this chapter is in preparation for publication (271).

region, and a QM/MM interaction term. Nonbonded QM/MM interactions can be readily calculated. For example, van der Waals interactions can be modeled by assigning a Lennard-Jones potential with standard MM parameters onto the QM atoms and electrostatic interactions can be modeled by including the partial atomic charges on the MM atoms as external charges into the one-electron Hamiltonian of the QM part. However, a methodological challenge in QM/MM calculations occurs when the boundary between the QM and MM regions crosses a chemical bond. In this case, the QM and MM regions must be linked such that the QM region can be treated as a closed-shell system while maintaining the overall structural integrity of the system.

Several techniques have been reported for linking the QM and MM regions. An intuitive and widely-used approach is the link-atom approach, in which a dummy hydrogen atom is added to saturate the valency of the QM region so as to form a closed-shell system (270; 278; 279). The link atom is positioned on the bond that is cut by the QM/MM boundary. Classical terms are added so as to hold the QM region in place relative to the MM region.

An alternative to the link-atom method is provided by frozen-orbital approaches in which the closed shell at the boundary QM atom is maintained by using strictly localized orbitals. This class of methods includes the Local Self Consistent Field (LSCF) (280; 281) and the Generalized Hybrid Orbital (GHO) (282; 283; 284) formalisms. In the LSCF method a hybrid orbital on the boundary QM atom pointing along the QM-MM bond is frozen, whereas in the GHO method four atomic orbitals are placed on the boundary MM atom, three of which are frozen.

Implementation of any frozen-orbital approach into available simulation packages requires extensive programming. In contrast, this is not necessary for the link-atom approach for which the implementation is straightforward. However, the link-atom approach suffers from the necessity of introducing an artificial atom into the simulation system. This atom is intended to mimic the MM atom (usually a carbon) to which the QM region is attached but is treated as a hydrogen in the QM calculation. Moreover, the link atom interacts electrostatically with the partial atomic charges on the MM atoms, thus distorting the local electrostatic field. To circumvent this it has been suggested to zero the charge on the MM host (285; 286). An alternative approach is to symmetrize the QM/MM link by adding two link atoms and, in addition, to delocalize the MM charges in the direct vicinity of the boundary (287). However, whichever scheme is chosen the local electrostatic field is altered relative to the pure MM treatment.

The frozen orbital approaches avoid the above problems by merely introducing frozen orbitals along the QM/MM boundary without the necessity of charge adaptation on any MM host atoms. Thus, formally frozen orbital approaches are more appropriate than link-atom approaches. However, a comparison of proton affinities, Mulliken charge distributions and geometries of alkanes determined from the link-atom and LSCF approaches has shown that, if properly used, both approaches yield comparable accuracies (288). Thus, the simple

link-atom approach can, in practice, be efficiently used for QM/MM calculations.

However, any QM/MM scheme introduces the unavoidable error that arises from neglecting quantum effects involving the atoms that are modeled classically. Consider one extreme scenario in which one fully neglects the classically-modeled atoms and performs calculations on a smaller model system, which is then treated fully quantum mechanically. An upper bound for the intrinsic linking error can be estimated by calculating the value of the chemical property of interest in this QM subsystem alone (which thus serves as a model system) and comparing it to the reference value of the same property calculated modeling the entire system quantum mechanically. This upper bound can be compared to the actual error which is given by the difference in the value of the property of interest between the QM/MM model and the full QM treatment of the entire system. If the actual error is smaller than the upper bound then the QM/MM scheme can be considered useful. The present chapter evaluates the link-atom approach according to this criterion of usefulness.

Here, proton affinity, a measure often used in testing QM/MM linking approaches (270; 283; 287; 288), is used as a test property. Treating the full system quantum mechanically yields the reference proton affinity, PA_{ref}. Using a smaller model compound that is treated fully quantum mechanically yields a proton affinity PA_{model}. The upper bound is then

$$\Delta PA_{model} = PA_{model} - PA_{ref}. \tag{7.23}$$

If the full system is subjected to a QM/MM calculation giving the proton affinity $PA_{QM/MM}$, the actual linking error is then given as the difference

$$\Delta PA_{QM/MM} = PA_{QM/MM} - PA_{ref} \tag{7.24}$$

and the criterion of usefulness is

$$|\Delta PA_{QM/MM}| < |\Delta PA_{model}|. \tag{7.25}$$

Proton affinities are determined here for the sidechains of three different amino acids (serine, histidine, and aspartate), in order to test the link-atom approach in a scenario that mimics a typical simulation model, *i.e.*, QM/MM calculations on an enzyme.

7.2.2 Methods

The link-atom approach was implemented here into the macromolecular simulation package CHARMM (289) by modifying the input but without changing the code and thus guaranteeing maximal user control. This has the advantages of flexibility, *i.e.*, of allowing for any

kind of QM/MM boundary, of requiring a minimum of changes to the topology and parameter files, and of allowing the use of an identical linking scheme with different quantum packages interfaced with CHARMM. Currently, CHARMM interfaces exist for the quantum packages GAMESS-US (279), GAMESS-UK (290), MNDO97 (291), and TURBOMOLE (292), each of which uses different linking techniques.

For the present implementation the nomenclature used is as follows (Fig. 7.2). The link atom is termed q_0. The MM and QM boundary atoms are termed m_1 and q_1. MM (QM) atoms bound to m_1 (q_1) are termed m_2 (q_2), and MM (QM) atoms bound to m_2 (q_2) are termed m_3 (q_3).

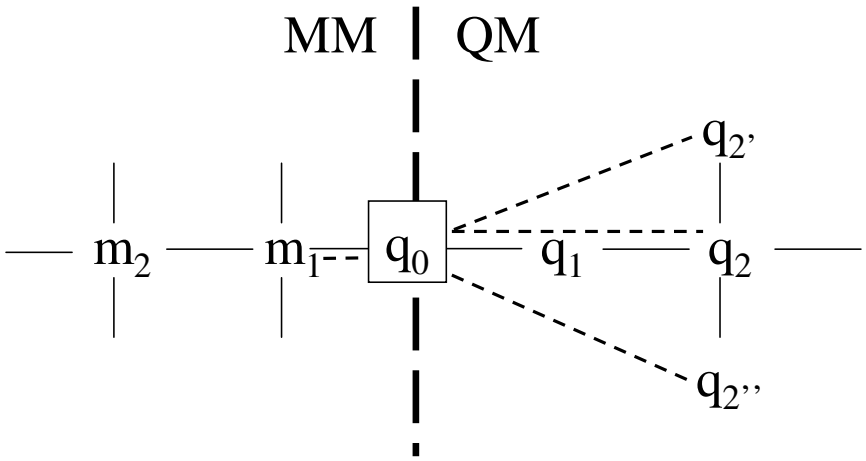

Figure 7.2: Illustration of the nomenclature used in the linking procedure that is described in the text. The link atom q_0 is placed between the MM boundary atom m_1 and the QM boundary atom q_1. Changes from the MM the topology during the linking procedure are: (1) The charge on m_1 is zeroed. (2) The angle m_1-q_0-q_1 is set to be 180°. (3) Dummy bonds (indicated by dashed lines) q_0 -m_1, q_0-q_2, q_0-$q_{2'}$, and q_0-$q_{2''}$ are introduced for van der Waals exclusions. (4) All MM interactions involving QM atoms that are not relevant for the relative positioning of the MM and QM regions are removed (see text).

Classical interactions involving QM atoms are removed, apart from those QM/MM terms that are required for the relative positioning of the QM region within the MM environment. The terms that are kept are the bond m_1-q_1, the angles m_2-m_1-q_1, and the dihedrals m_3-m_2-m_1-q_1 and m_2-m_1-q_1-q_2. An additional angle term m_1-q_0-q_1 keeps the link atom collinear with the m_1-q_1 bond. The charge on the MM boundary atom m_1 is zeroed. Details can be found in Appendix C.

To test the link-atom approach introduced above, calculations were performed of the vacuum proton affinities of the sidechains of serine, histidine, and aspartate. Setting the energy of a proton in vacuum to be zero, the proton affinity is given by the energy difference be-

tween the protonated and the deprotonated species. Here, effects of zero-point energies, vibrational energies, and the loss of translational degrees of freedom are neglected, as they are expected to be of second order. To avoid zwitterionic states of the amino acids, the termini were capped using a C=OCH_3 group for the N-termini and an $NHCH_3$ group for the C-termini.

For each N-acyl-amino-acid-methylamide, three sets of proton affinity calculations were performed, *i.e.*, a QM-only calculation of the entire system yielding the reference proton affinity, PA_{ref}, a QM-only calculation of the sidechain alone yielding the proton affinity of the model compound, PA_{model}, and a QM/MM calculation of the entire system yielding the QM/MM proton affinity, $PA_{QM/MM}$. These three different systems are illustrated in Fig. 7.3 for the case of N-acyl-serine-methylamide. The QM/MM boundary was placed between the C_α and C_β atoms. The proton affinities were calculated from the energies of geometry-optimized structures after optimization to a final root-mean-square gradient of 0.01 kcal/mol/Å using the HF/3-21G*, HF/6-31G**, and B3LYP/6-31G** levels of theory for the QM part and the CHARMM force field for the MM part. A co-compiled version (279) of GAMESS-US (293) interfaced with CHARMM (version 28)(289) was used. The methods and basis sets were chosen appropriately given the present-day computational power available for current QM/MM calculations of complex biological systems. For the present purpose, *i.e.*, testing the link-atom approach, the basis sets chosen here should provide sufficient accuracy. For biological applications it may in certain cases be wise to augment the basis sets with diffuse functions.

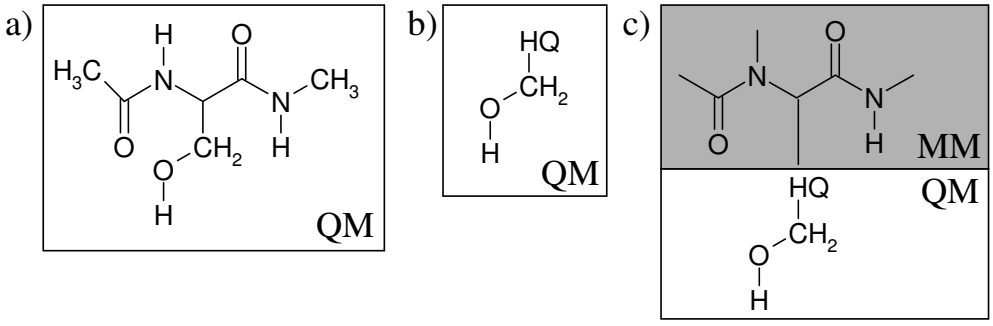

Figure 7.3: Proton affinities are calculated for protonation or deprotonation of the hydroxyl group of the serine sidechain for a) all atoms present and all atoms treated quantum mechanically (yielding the reference proton affinity, PA_{ref}), b) only atoms of the sidechain present and treated quantum mechanically (yielding the proton affinity of the model compound, PA_{model}), and c) all atoms present, sidechain atom treated quantum mechanically and all other atoms treated classically (yielding the QM/MM proton affinity, $PA_{QM/MM}$).

7.2.3 Results and Discussion

To evaluate the link-atom approach according to the criterion of usefulness given in Eq. 7.25 it is necessary to calculate differences in proton affinities. Thus, the calculation procedure chosen must be able to reproduce experimental differences in proton affinities with reasonable accuracy. To test this, the proton affinities of methanol and ethanol were calculated (Table 7.1). Compared to the experimental proton affinities, which are -382 ±1 kcal/mol (294) and -377 to -379 kcal/mol (295; 296; 297; 298), respectively, the calculated absolute affinities are too favorable by 25 to 40 kcal/mol. However, the difference in proton affinities between methanol and ethanol of ΔPA_{exp}= -3 to -5 kcal/mol is captured by the calculations. Therefore, the calculation scheme adequately reproduces differences in proton affinities and is thus suitable for evaluating errors introduced by the link-atom approach.

	PA(MeOH)	PA(EtOH)	ΔPA
HF/3-21G*	-422.45	-418.80	-3.65
HF/6-31G**	-412.38	-409.46	-2.92
B3LYP/6-31G**	-406.16	-402.44	-3.72
Experiment	-382 ±1	-378 ±1	-4 ±2

Table 7.1: Proton affinities of MeOH and EtOH in kcal/mol. Experimental values are from Refs. (294; 295; 296; 297; 298).

Reference proton affinities PA_{ref} were calculated for capped serine, capped histidine, and capped aspartate, treating all atoms quantum mechanically. Proton affinities PA_{model} were determined for the corresponding model compounds methanol, 4-methyl-imidazole, and acetic acid. Finally, the QM/MM proton affinities $PA_{QM/MM}$ were determined for linked capped serine, linked capped histidine, and linked capped aspartate. The resulting proton affinities and the deviations from the reference, ΔPA_{model} and $\Delta PA_{QM/MM}$ and are listed in Tables 7.2, 7.3, and 7.4.

Serine in its neutral form can both accept and release a proton. Both cases are examined here. For the reaction from the anionic to the neutral form the calculated proton affinity is $PA_{ref}^{SerO^-}$ ∼-381 kcal/mol. The proton affinity of the model compound methanolate, $PA_{model}^{SerO^-}$ is ∼-413 kcal/mol. Thus, the error $|\Delta PA_{model}^{SerO^-}|$ is ∼32 kcal/mol. When linking is used, $PA_{QM/MM}^{SerO^-}$ is ∼-410 kcal/mol. Therefore, the error $|\Delta PA_{QM/MM}^{SerO^-}|$ of ∼29 kcal/mol is smaller than $|\Delta PA_{model}^{SerO^-}|$. For the reaction from the neutral to the cationic form the calculated proton affinity is PA_{ref}^{SerOH}∼-227 kcal/mol. The proton affinity of the model compound methanol (PA_{model}^{SerOH}∼-1 96 kcal/mol) is $|\Delta PA_{model}^{SerOH}|$ ∼31 kcal/mol less favorable. Again, linking reduces this error ($PA_{QM/MM}^{SerOH}$ ∼-200 kcal/mol, $|\Delta PA_{QM/MM}^{SerOH}|$ ∼27 kcal/mol).

Histidine in its neutral form is protonated either on $N_{\epsilon 2}$ or on $N_{\delta 1}$. Both forms were used here to determine the proton affinity corresponding to the reaction from neutral to cationic

	PA_{ref}	$PA_{QM/MM}$	$\Delta PA_{QM/MM}$	PA_{model}	ΔPA_{model}
		$SerO^- + H^+ \longrightarrow SerOH$			
HF/3-21G*	-385.21	-420.11	-34.90	-422.45	-37.24
HF/6-31G**	-383.67	-409.38	-25.71	-412.38	-28.71
B3LYP/6-31G**	-375.00	-402.17	-27.17	-406.16	-31.16
		$SerOH + H^+ \longrightarrow SerOH_2^+$			
HF/3-21G*	-231.85	-208.45	23.40	-205.08	26.77
HF/6-31G**	-226.92	-198.37	28.55	-193.91	33.01
B3LYP/6-31G**	-226.60	-195.94	30.66	-190.57	36.03

Table 7.2: Calculated proton affinities for capped serine (PA_{ref}), linked capped serine ($PA_{QM/MM}$), and methanol (PA_{model}) in kcal/mol. The differences $\Delta PA_{QM/MM} = PA_{QM/MM} - PA_{ref}$ and $\Delta PA_{model} = PA_{model} - PA_{ref}$ are also listed. The experimental value is PA_{model}= -382 kcal/mol (294).

	PA_{ref}	$PA_{QM/MM}$	$\Delta PA_{QM/MM}$	PA_{model}	ΔPA_{model}
		His (proton on $N_{\epsilon 2}$) $+H^+ \longrightarrow HisH^+$			
HF/3-21G*	-268.76	-259.56	9.12	-252.82	15.94
HF/6-31G**	-258.09	-253.04	5.05	-245.96	12.13
B3LYP/6-31G**	-257.80	-250.06	7.74	-243.25	14.55
		His (proton on $N_{\delta 1}$) $+H^+ \longrightarrow HisH^+$			
HF/3-21G*	-261.09	-255.28	5.81	-252.56	8.53
HF/6-31G**	-252.02	-248.76	3.26	-246.16	5.86
B3LYP/6-31G**	-251.37	-245.78	5.59	-243.40	7.90

Table 7.3: Calculated proton affinities for capped histidine (PA_{ref}), linked capped histidine ($PA_{QM/MM}$), and 4-methyl-imidazole (PA_{model}) in kcal/mol. The differences $\Delta PA_{QM/MM} = PA_{QM/MM} - PA_{ref}$ and $\Delta PA_{model} = PA_{model} - PA_{ref}$ are also listed. The experimental value is PA_{model}= -227.7 kcal/mol (299).

	PA_{ref}	$PA_{QM/MM}$	$\Delta PA_{QM/MM}$	PA_{model}	ΔPA_{model}
		$Asp^- + H^+ \longrightarrow Asp$ (proton on $O_{\delta 1}$)			
HF/3-21G*	-363.45	-371.17	-7.72	-376.40	-12.95
HF/6-31G**	-360.63	-365.76	-5.13	-371.06	-10.43
B3LYP/6-31G**	-355.98	-363.10	-7.12	-369.71	-13.73
		$Asp^- + H^+ \longrightarrow Asp$ (proton on $O_{\delta 2}$)			
HF/3-21G*	-359.73	-371.48	-11.75	-377.23	-17.50
HF/6-31G**	-356.09	-366.46	-10.37	-371.68	-15.59
B3LYP/6-31G**	-352.17	-363.81	-11.64	-369.70	-17.53

Table 7.4: Calculated proton affinities for capped aspartate (PA_{ref}), linked capped aspartate (PA_{ref}), and acetic acid (PA_{ref}) in kcal/mol. The differences $\Delta PA_{QM/MM} = PA_{QM/MM} - PA_{ref}$ and $\Delta PA_{model} = PA_{model} - PA_{ref}$ are also listed. The experimental value is PA_{model}= -343 to -349 kcal/mol (300; 301; 302; 303).

histidine. In the $N_{\epsilon 2}$ case, a proton affinity of $PA_{ref}^{His_{\epsilon 2}}$∼-261 kcal/mol was found, whereas in the $N_{\delta 1}$ case the proton affinity is $PA_{ref}^{His_{\delta 1}}$∼-254 kcal/mol. In a previous study a full analysis of the conformations accessible to N-formyl-L-histidine-amide (rather than N-acyl-L-histidine-methylamide as used here) in all possible protonation states was made at the HF/6-31G* level of theory (304; 305). The differences in SCF energies between protonated and deprotonated species as calculated from the values tabulated in Refs. (304) and (305) are similar to the proton affinities calculated here, thus providing additional support of the method used here. The proton affinity of the model compound 4-methyl-imidazole is $PA_{model}^{His_{\epsilon 2}} \approx PA_{model}^{His_{\delta 1}}$∼-247 kcal/mol, irrespective of the position of the proton in the neutral form. Thus, the absolute errors are $|\Delta PA_{model}^{His_{\epsilon 2}}|$˜ 14 kcal/mol for the protonated $N_{\epsilon 2}$ tautomer and $|\Delta PA_{model}^{His_{\delta 1}}|$∼7 kcal/mol for the protonated $N_{\delta 1}$ tautomer. In both cases, these errors were again reduced after linking was introduced ($PA_{QM/MM}^{His_{\epsilon 2}}$∼-254 kcal/mol, $|\Delta PA_{QM/MM}^{His_{\epsilon 2}}|$∼7 kcal/mol, $PA_{QM/MM}^{His_{\delta 1}}$∼-249 kcal/mol, $|\Delta PA_{QM/MM}^{His_{\delta 1}}|$∼5 kcal/mol).

In case of aspartate the proton was positioned on either terminal oxygen of the sidechain in the neutral form, allowing investigation of the conformation dependence of the proton affinity. The corresponding two proton affinities are expected to differ due to the asymmetry of the environment provided by the capping groups. Indeed, when the proton is positioned on $O_{\delta 1}$ the proton affinity is $PA_{ref}^{Asp_{\delta 1}}$∼-359 kcal/mol whereas when it is positioned on $O_{\delta 2}$ the proton affinity becomes $PA_{ref}^{Asp_{\delta 2}}$∼-355 kcal/mol, *i.e.*, a difference of ∼4 kcal/mol. As expected, the proton affinity of the symmetric model compound, acetic acid, is $PA_{model}^{Asp_{\delta 1}} = PA_{model}^{Asp_{\delta 2}}$∼-372 kcal/mol independent of the position of the proton. In the QM/MM calculations the proton affinities are identical in the two cases ($PA_{QM/MM}^{Asp_{\delta 1}} = PA_{QM/MM}^{Asp_{\delta 2}}$ ∼-366 kcal/mol), thus reflecting the symmetry of the model compound acetic acid rather than the asymmetry of capped aspartate. However, the errors $|\Delta PA_{QM/MM}^{Asp_{\delta 1}}|$∼7 kcal/mol and $|\Delta PA_{QM/MM}^{Asp_{\delta 2}}|$∼11 kcal/mol are smaller than

$|\Delta PA^{Asp_{\delta 1}}_{model}|$∼13 kcal/mol and $|\Delta PA^{Asp_{\delta 2}}_{model}|$∼17 kcal/mol, respectively.

The above results show that of the error in sidechain proton affinities in the QM/MM treatment is smaller than the error when using a model compound in all cases studied. Thus, the criterion of usefulness given in Eq. 7.25 is satisfied and the suitability of the link-atom approach as investigated here is shown, even though the difference between $|\Delta PA_{QM/MM}|$ and $|\Delta PA_{model}|$, is not great. However, the results do indicate that the link-atom approach can be used for QM/MM calculations without introducing major artifacts as would arise from a violation of the condition given in Eq. 7.25.

The observation that the asymmetry in proton affinities for capped aspartate, found in the full QM treatment, is lost upon QM/MM treatment indicates that, although partial charges in the vicinity of the sidechain are represented in QM/MM calculations, this is not sufficient to capture the full influence of neighboring groups on the proton affinity. Electronic effects such as delocalization of the wave function or mutual polarization may affect the electron density distribution throughout the whole molecule, thereby influencing the proton affinity. However, these effects cannot easily be described within a QM/MM framework. This conclusion can be generalized to any linking protocol. Even though the error in proton affinities of small molecules *in vacuo* may be smaller when using a more elaborate linking protocol (such as a frozen orbital approach) rather than the simple link-atom approach, the error made in protein QM/MM calculations due to not treating neighboring atoms quantum mechanically is inevitable and its magnitude cannot be foreseen. An estimate of the intrinsic error introduced can, in principle, be made only by performing benchmark calculations of proton affinities of sidechains in a protein, treating the whole protein quantum mechanically, and accurately representing solvent electrostatic screening effects. Both full quantum mechanical treatment of proteins with semi-empirical (306; 307) or density functional methods (308) and the treatment of solvent electrostatic screening in QM/MM calculations (309; 310) are areas of active research.

The conformation dependence of proton affinities and pK_a values is an issue that receives growing interest. Recently it was found that the conformation of N-formyl-histidine-amide strongly influences the proton affinity of the sidechain (311). This is in accord with the present results which also indicate that the proton affinities of the amino-acid sidechain depends significantly on the conformation of the capped amino acid. Treating only the sidechain quantum mechanically cannot capture this intra-amino-acid conformation-dependent effect. Thus, to reliably model sidechain protonation events that may play a crucial role in enzymatic catalysis in the framework of QM/MM calculations, it may be necessary to include in the QM region not only the sidechain of the respective amino acid but also adjacent backbone atoms.

7.2.4 Conclusions

In the present work the link-atom approach for QM/MM calculations has been implemented in a set of user-controllable patches for the macromolecular simulation package CHARMM. In all test cases studied the error introduced by this specific linking scheme as measured by $|\Delta PA_{QM/MM}|$ was found to be smaller than $|\Delta PA_{model}|$ which gives an upper bound of the error due to the neglect of quantum effects involving MM atoms in the direct vicinity of the QM atoms (this error is intrinsic to any linking scheme). Thus, the utility of the link-atom procedure has been demonstrated. The linking approach discussed here therefore provides a reliable and flexible tool for QM/MM calculations of reaction paths and energetics of enzyme-catalyzed reactions.

Chapter 8

Determination of reaction paths

In this chapter, the concepts of a minimum-energy path (MEP) to describe chemical reactions (Section 8.1) and of a one-dimenational curvilinear reaction coordinate (Section 8.3) are introduced. Methods to determine a MEP (Section 8.2) and to calculate reaction rates from a MEP are given (Section 8.4).

8.1 Definition of a minimum energy path (MEP)

A chemical reaction involves a change of the relative positions and connectivities of the nuclei involved in the reaction and the associated change in the electronic 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 (312). Mathematically, the PES is a function of the $3N - 6$ internal degrees of freedom of a nonlinear molecule. If the reaction involves electronically excited states as is the case in photochemical reactions, both the ground and excited state surfaces must be available.

If the multidimensional PES is completely and accurately known the reaction dynamics can be determined by solving the equations of motion of the nuclei moving on the PES (313). In such a scenario, the reactant and product of a chemical reaction 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 reactant and the product states.

However, in highly-complex multidimensional systems it is impossible to determine the PES to chemical accuracy since this involves the solution of the electronic Schrödinger equation for all possible nuclear configurations, a task that is computationally not feasible for systems with many degrees of freedom.

Thus, it is often necessary to restrict the investigation to mapping important points on the PES, namely the minima and first-order saddle points connecting the minima. In such a simplified scheme the reactant and product states can be identified with minima that correspond to stable configurations and the transition states accordingly with saddle points.

A continuous path leading from the reactant to the product is called a reaction path. It defines the generalized one-dimensional curvilinear reaction coordinate. If all points on the PES along the reaction path are minima in the $(3N-7)$-dimensional subspace orthogonal to the reaction coordinate the reaction path is called a minimum energy path (MEP). This definition implies that MEP paths always lead *via* n ($n \geq 1$) transition states (first-order saddle points on the PES) and $n-1$ intermediate states (local minima on the PES) such that a steepest descent path (also called intrinsic reaction coordinate) starting on either side from a transition state leads to the neighboring local minimum along the path. A reaction path that is chemically interpretable always starts and ends in a minimum on the PES, *i.e.*, both reactant and product states are stable configurations.

A contour plot of the PES and the MEP for the simplest chemical reaction possible, *i.e.*, a collinear collision reaction $A + BC \rightarrow AB + C$ is shown in Fig. 8.1.

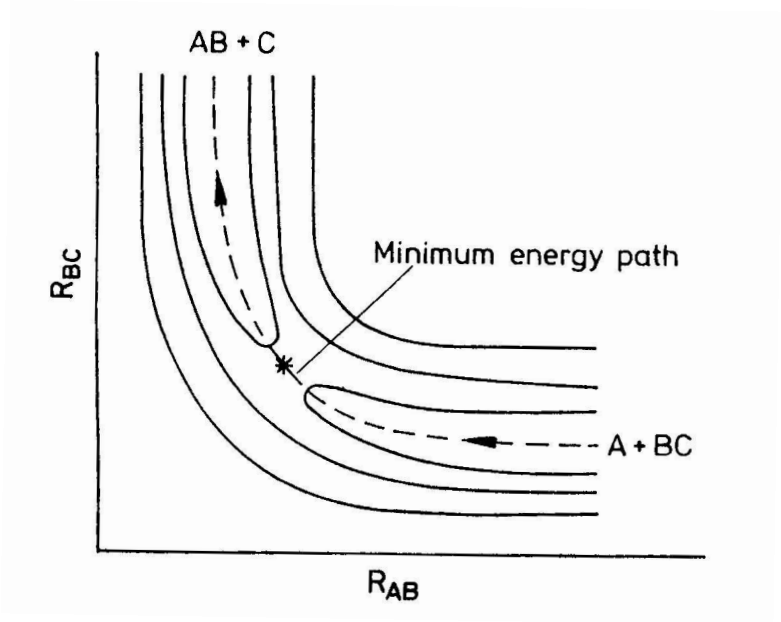

Figure 8.1: Potential energy surface contour plot and minimum energy path for a collinear $A + BC$ collision. The figure was taken from Ref. (312).

8.2 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 for example Ref. (314)). 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.

A heuristic algorithm to achieve this task particularly for the case of conformational transitions in proteins has been suggested (315). This algorithm, Conjugate Peak Refinement (CPR), identifies the highest-energy points along the initial path (the "peaks") and moves these points closer to the MEP by a controlled conjugate gradient minimization. The procedure is illustrated in Fig. 8.2. Starting from an initial path the point of highest energy with coordinates $\vec{r}_{peak}$ is found by maximization along the path. It is then refined by line minimization along the tangent $\vec{s}_0$ to the path at $\vec{r}_{peak}$, resulting in a point $\vec{r}_0$. The minimization is exited if the root-mean-square gradient is lower than a user-defined value g. A new conjugate vector $\vec{s}_1$ with respect to the Hessian is built at $\vec{r}_0$, followed by line minimization. This is repeated until either the numerically built vector $\vec{s}_j$ is no longer conjugate to $\vec{s}_0$ (*i.e.*, the path direction) or until the number of conjugate directions along which the point is a minimum matches the requested number of line minimizations C. If $C = 3N - 7$, *i.e.*, the energy is maximal along the path direction $\vec{s}_0$ and minimal along all other internal degrees of freedom, the point is an exact first-order saddle point. By variation of the minimization gradient exit criterion g and the number of line minimizations C required it is possible to control the accuracy of the path refinement.

CPR has been shown to be a robust algorithm by allowing to find MEPs of very different processes in proteins, such as catalysis of an isomerization reaction (317), proton transfer (318), chloride pumping in halorhodopsin (319), transport across membrane channels (320), and large-scale conformational changes such as the recovery-stroke transition in myosin (92).

8.3 One-dimensional curvilinear reaction coordinate

The MEP can be used to define a normalized, one-dimensional curvilinear reaction coordinate λ. The MEP is described by a discrete set of M points in conformational space, $P = [\vec{r}_0, \vec{r}_1, \ldots, \vec{r}_M]$, where $\vec{r}_0$ corresponds to the reactant structure and $\vec{r}_M$ to the product structure. The progress along the reaction up to a path point $\vec{r}_i$ is measured by summing the root-mean-square coordinate-change along the path up to $\vec{r}_i$,

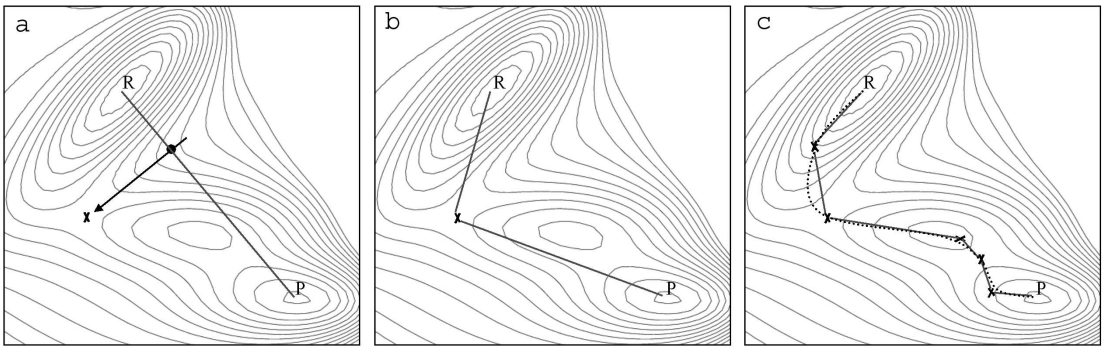

Figure 8.2: Path optimization by CPR (schematic). (a) Starting from an initial guess of the path (here: a linear interpolation between reactant (R) and product(P) states) the point of highest energy is found by maximization along the path (•). This point is moved closer to the MEP by a series of successive line-minimizations along directions conjugate to the path direction at (•). On the present 2D surface this amounts to only one line minimization (shown as →). (b) The optimized point (x) is inserted into the path. (c) This process of maximization/minimization is repeated until all local energy maxima along the path are identified as first-order saddle points. The path thus obtained (x-x-x) is a good approximation to the MEP($\cdot\cdot\cdot\cdot\cdot$). The figure was taken from Ref. (316).

$$\lambda(i) = \frac{\sum_{k=1}^{i} |\vec{r}_k - \vec{r}_{k-1}|}{\sum_{k=1}^{M} |\vec{r}_k - \vec{r}_{k-1}|}, \tag{8.1}$$

where $\sum_{k=1}^{M} |\vec{r}_k - \vec{r}_{k-1}|$ is the normalization factor.

In the following, the term *reactant* refers to the structure at $\lambda = 0$ (with vanishing gradient), *product* to the structure at $\lambda = 1$ (with vanishing gradient), *transition state* to a saddle-point structure (with vanishing gradient), *intermediate* to a minimum-energy structure (with vanishing gradient), and *transient state* to any structure along the reaction coordinate λ.

8.4 Calculating reaction rates from a MEP

Once the reactant, transition, and product states as defined by a MEP are known, reaction rates can be calculated by making use of the transition state theory (TST).[1] The TST is strictly valid only for single-step reactions of the type

[1]The derivation of the rate law given in this Section is mainly based on the derivation given in Ref. (321)

$$R \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} TS^{\ddagger} \overset{k_2}{\rightarrow} P, \tag{8.2}$$

in which a reactant (R) is in rapid equilibrium with the transition state ($TS^{\ddagger}$) that in turn decays irreversibly to the product (P) with a rate k_2. The transition state is identified with an activated complex that is formed from the reactant with the rate k_1 and that decays back to the reactant with rate k_{-1}. Product formation is rate-determining, so that the velocity of the reaction is given by

$$-\frac{d[R]}{dt} = k_2 \cdot [R], \tag{8.3}$$

where $[R]$ is the concentration of the reactant. According to the TST the reaction velocity is determined by the concentration of the transition state, $[TS^{\ddagger}]$, and by the frequency (*i.e.*, probability) $\nu^{\ddagger}$ with which the transition state decays to the product,

$$-\frac{d[R]}{dt} = [TS^{\ddagger}] \cdot \nu^{\ddagger}. \tag{8.4}$$

Combining Eqs. 8.3 and 8.4 and using the law of mass action gives

$$k_2 = \frac{[TS^{\ddagger}]}{[R]} \cdot \nu^{\ddagger} = K^{\ddagger} \cdot \nu^{\ddagger} \tag{8.5}$$

The equilibrium constant $K^{\ddagger}$ can be determined from the partition functions of the transition state ($z^{\ddagger}$) and the reactant state (z^R),

$$K^{\ddagger} = \frac{z^{\ddagger}}{z^R}. \tag{8.6}$$

Assuming separability of the contributing energy terms (*i.e.*, rotational, translational, vibrational, and electronic energies) the partition functions can be written as the product of the rotational, translational, vibrational, and electronic functions. In particular, the partition function of the single vibrational degree of freedom of the transition state that corresponds to the negative eigenvalue ν_0 can be written explicitly,

$$z_{\nu_0} = (1 - e^{-\frac{h\nu_0}{k_B T}})^{-1}, \tag{8.7}$$

where h is Planck's constant, k_B is Boltzmann's constant, and T is the absolute temperature. The vibrational mode corresponding to ν_0 leads to the decay of the activated complex. It is thus possible to equate the decay frequency $\nu^\ddagger$ with ν_0. At the temperature at which the reaction proceeds this mode must always be excited, therefore $h\nu^\ddagger \ll k_BT$ or $h\nu^\ddagger/k_BT \ll 1$. The exponential in Eq. 8.7 can thus be expanded

$$e^{-\frac{h\nu_0}{k_BT}} = 1 - \frac{h\nu_0}{k_BT} + \frac{1}{2}\left(\frac{h\nu_0}{k_BT}\right)^2 - \cdots \tag{8.8}$$

Truncation of the expansion after the linear term and insertion into Eq. 8.7 yields

$$z_{\nu^\ddagger} = \frac{k_BT}{h\nu^\ddagger}. \tag{8.9}$$

The equilibrium constant $K^\ddagger$ can now be written as

$$K^\ddagger = \frac{k_BT}{h\nu^\ddagger}K'^\ddagger, \tag{8.10}$$

where $K'^\ddagger$ is the equilibrium constant of formation of the activated complex determined under neglection of the vibrational decay mode. Insertion of Eq. 8.10 into Eq. 8.5 yields

$$k_2 = \frac{k_BT}{h}K'^\ddagger. \tag{8.11}$$

The free energy of formation of the activated complex, $\Delta G^{0\ddagger}$, is given by

$$\Delta G^{0\ddagger} = -RT\ln K'^\ddagger, \tag{8.12}$$

where R is the ideal gas constant. The combination of Eqs. 8.11 and 8.12 yields the rate law in a form suitable for interpretation within the framework of phenomenological thermodynamics,

$$k_2 = \frac{k_BT}{h}e^{-\frac{\Delta G^{0\ddagger}}{RT}}. \tag{8.13}$$

The free energy of formation of the activated complex, $\Delta G^{0\ddagger}$, is the difference in free energy of the activated complex and the reactant. As such, it is equated to the activation free

energy of the reaction. Since the reaction must proceed *via* the activated complex that is only occasionally formed (due to its high energy), the image of the system having to climb up an energy hill prior to be able to cross a barrier is frequently used. Within this image, the magnitude of $\Delta G^{0\ddagger}$ is equated to the barrier height.

A different formulation of the rate law is obtained if the partition function is explicitly formulated not only for the decay mode of the activated complex but for all degrees of freedom,

$$z = z_t \cdot z_r \cdot z_v \cdot z_{el}, \tag{8.14}$$

where z_t is the translational, z_r the rotational, z_v the vibrational, and z_{el} the electronic partition function. Assuming that only the electronic ground state with energy ϵ_0 is populated at the reaction temperature yields

$$z_{el} = e^{-\frac{\epsilon_0}{k_B T}} \tag{8.15}$$

With this, the rate law in its microscopic form reads

$$k_2 = \frac{k_B T}{h} \frac{z_t^{\ddagger} z_r^{\ddagger} z_v'^{\ddagger}}{z_t^R z_r^R z_v^R} e^{-\frac{\Delta E_0}{k_B T}}, \tag{8.16}$$

where ΔE_0 is the difference in electronic ground state energies of the transition and reactant states and the prime indicates that the decay mode is omitted in the vibrational partition function of the transition state.

Depending on the energy surface that is used to determine the MEP, the corresponding rate law must be chosen. If the reaction path is calculated on a free energy surface, as is the case in free energy perturbation methods, Eq. 8.13 should be used, whereas if a potential energy surface is used as is usually the case in quantum mechanical calculations, Eq. 8.16 should be taken. Details on the energy surface used for the present thesis are given in Section 10.3.

Because of the assumptions made by TST Eqs. 8.13 and 8.16 are valid only as a limiting threshold. If, for instance, the reaction proceeds in several steps, the reaction rate will be determined by the highest barrier that has to be crossed. This highest barrier is thus rate-limiting and the corresponding step is the rate-limiting step. However, if an effective reaction barrier is calculated from the measured overall reaction rate it will overestimate this rate-limiting barrier because the other barriers will also contribute to the reaction rate. The overestimation will be significant when the reaction proceeds *via* a large number

of barriers (as is the case for major conformational transitions in proteins) giving rise to diffusive-like motions as well as when different barriers of equal magnitude have to be crossed such that no single rate-limiting step exists.

If the reaction can proceed *via* different reaction channels that are isoenergetic, the rate law must be multiplied with a symmetry factor σ, *i.e.*, the number of identical reaction paths (312).

One of the major assumptions of TST is that the reaction barrier is crossed only once. This is, however, not true in general. Recrossing events do take place when the system moves along a trajectory that crosses the barrier several times before ending in the product (if the barrier is crossed an odd number of times) or in the reactant state (if the barrier is crossed an even number of times). Thus, the barrier height calculated according to Eqs. 8.13 and 8.16 will overestimate the actual barrier.

Another assumption in TST is that quantum mechanical tunneling and barrier reflection is neglected. If light particles are involved in the reaction (as is the case in hydrogen transfer reactions) the system may evolve from the reactant to the product state due to tunneling although the energy content does not allow for barrier crossing. In this case, TST will underestimate the barrier height. In contrast, if the system contains more energy than necessary for barrier crossing but the reaction nevertheless does not proceed due to quantum mechanical barrier reflection TST will overestimate the barrier.

The TST is formulated for reactions in gas phase. However, most reactions take place in an environment other than gas phase (*i.e.*, in solution, in an enzyme active site, or at a surface of a heterogeneous catalyst). Thus, the energy of the reacting system may be dissipated into the environment due to frictional effects, thus slowing down the reaction. Concurrently, vibrational modes of the environment may couple to the reaction coordinate of the reacting system in a coherent fashion, thus speeding up the reaction.

The collective effects of recrossing, tunneling, quantum mechanical barrier reflection, and interactions between the reacting system and its environment are captured by introduction of a transmission coefficient into the rate law. Thus, the rate law becomes

$$k_2 = \kappa \cdot \sigma \cdot \frac{k_B T}{h} e^{-\frac{\Delta G^{0\ddagger}}{RT}} \tag{8.17}$$

or

$$k_2 = \kappa \cdot \sigma \cdot \frac{k_B T}{h} \frac{z_t^{\ddagger} z_r^{\ddagger} z_v'^{\ddagger}}{z_t^R z_r^R z_v^R} e^{-\frac{\Delta E_0}{k_B T}}, \tag{8.18}$$

Extensions of the TST that include approaches to calculating the transmission coefficient κ both in equilibrium and nonequilibrium systems are available (reviewed for instance in Refs.

(269; 322; 323; 324)). Throughout this thesis, however, both the transmission coefficient κ and the symmetry factor σ are assumed to be equal to 1.

Chapter 9

Non-Uniform Charge Scaling (NUCS)

In molecular mechanics calculations, electrostatic interactions between chemical groups are usually represented by a Coulomb potential between the partial atomic charges of the groups. In aqueous solution these interactions are modified by the polarizable solvent. While the electrostatic effects of the polarized solvent on the protein are well described by the Poisson-Boltzmann equation, its numerical solution is computationally expensive for large molecules such as proteins. The procedure of Non-Uniform Charge Scaling (NUCS) is a pragmatic approach to implicit solvation that approximates the solvent screening effect by individually scaling the partial charges on the explicit atoms of the macromolecule so as to reproduce electrostatic interaction energies obtained from an initial Poisson-Boltzmann analysis. Once the screening factors have been determined for a protein the scaled charges can be easily used in any molecular mechanics program which implements a Coulomb term. The approach is particularly suitable for minimization-based simulations, such as normal mode analysis, certain conformational reaction path or ligand binding techniques for which bulk solvent cannot be included explicitly, and for combined quantum mechanical/molecular mechanical calculations when the interface to more elaborate continuum solvent models is lacking. The method is illustrated using reaction path calculations of the Tyr35 ring flip in the bovine pancreatic trypsin inhibitor.[1]

[1]The contents of this chapter have been published (310; 325).

9.1 Global NUCS

9.1.1 Introduction

Due to the high polarizability of bulk water, which is reflected by its high dielectric constant, the solvent plays a major role in the electrostatic energy of a protein under physiological conditions. The partial charge on a given protein atom generates an electrostatic field, along which the hydrating water molecules tend to orient. The summed dipole fields of the oriented water molecules generate the so-called reaction field. In response to the electrostatic fields of all partial charges, the overall reaction field is the sum of the reaction fields of individual charges (superposition principle). A partial charge on a given protein atom thus interacts with both its own reaction field (giving rise to the so-called Born self energy) and with the reaction field of all other charges. The latter is referred to as the screening energy, because it usually has the opposite sign than the Coulomb interaction with the other charges. Summing Born and screening energies over all partial charges in the protein gives the electrostatic interaction energy between the protein and the solvent, *i.e.*, the electrostatic contribution to the solvation free energy.

Several approaches for modeling solvation in macromolecular simulations are available. However, as new simulation techniques evolve, such as combined quantum mechanical/ molecular mechanical (QM/MM) methods to describe enzyme reactions, new approaches for appropriately treating solvent effects must be developed (309). A straightforward way of accurately treating solvent effects in molecular dynamics simulations is to model the bulk solvent with explicit water molecules (326). The explicit dipole fields of all water molecules then add up to yield the appropriate reaction field. However, the explicit solvent molecules increase the number of atoms, resulting in a large amount of computation time spent on the calculation of the solvent, rather than the solute interactions. Moreover, explicit solvent "freezes" when minimization-based calculations are performed such as normal modes of vibration analysis, ligand docking, or computing minimum-energy pathways. An alternative approach is to model the solvent implicitly as a polarizable dielectric continuum. The electrostatic potential is then described by the Poisson-Boltzmann equation (327). However, this second-order differential equation must be solved numerically, usually via a finite difference scheme (328; 329; 330; 331), a procedure that is computationally demanding and thus not practical when a large number of energy evaluations are required. Analytical approximations to the Poisson-Boltzmann model have been suggested (332), among them the Generalized Born (GB) model (333) and a variant, the Analytic Continuum Electrostatics (ACE) model (334; 335). Both these methods have been shown to produce results in reasonable agreement with the results of numerical solution of the Poisson-Boltzmann equation for test sets of small molecules, peptides, and small proteins (336; 337; 338). Another analytical method, in which the low-dielectric region is replaced by concentric spheres for the electrostatics of globular proteins has been proposed (339).

When applied to large proteins, GB models have been shown to reliably reproduce Poisson-Boltzmann energies (340). However, the ability of these approximate analytical methods to achieve structural stability in molecular mechanics applications of large proteins has been questioned (341; 342).

An even simpler - and already quite old - method introduces a distance-dependent dielectric in the Coulomb potential ($\epsilon = r$) (289). An extension of the distance-dependent dielectric approach including a self-energy term has been reported and applied to molecular dynamics simulations of peptides/small proteins (343). Since the distance-dependent method does not distinguish between atom pairs at the surface of a protein, where the solvent screening effect is large, and atom pairs in the core of a protein, where the screening is small, it oversimplifies solvent effects (344), due partly to the non-uniform nature of these effects. In a variant on the principle of distance-dependence, this had been addressed by scaling the partial atomic charges depending on their distance from the center of the protein (345). However, distance dependence does have the advantage of speed and simplicity over other implicit solvent approaches. Thus, there is a gap between crude but fast methods and more accurate but slower methods. In the present work we aim to fill this gap with a procedure for approximating solvent screening that is as fast and easy to implement as the distance-dependent dielectric approach but is more accurate and versatile in its possible applications.

The present method, Non-Uniform Charge Scaling (NUCS), consists in scaling the partial atomic charges of the protein in such a way that all possible interactions between groups of the protein computed with the standard Coulomb potential optimally reproduce the electrostatic interaction energies in solution. The NUCS procedure is, in principle, comparable to other methods that approximate solvent screening by charge scaling (309; 317; 346). These methods, however, compute scaling factors so as to reproduce the electrostatic potential at a specific site of interest only, whereas NUCS aims at adequately representing the electrostatic potential globally, i.e., at the position of every group in the protein. This is achieved by introducing non-uniform scaling factors, which are used to scale the partial atomic charges while conserving the local electrostatic multipoles of small groups of atoms. The solvated interaction energies that serve as a reference to derive the scaling factors are obtained from an initial solution of the Poisson-Boltzmann equation. The NUCS procedure thus allows solvent screening effects to be included while preserving the speed of the energy evaluations. The scaling factors need to be evaluated only once and can then in principle be used for any type of simulation requiring an implicit solvent representation, independent of the simulation package used. Moreover, there is no need to program an interface to other simulation tools, such as QM/MM, since NUCS generates only an array of modified partial atomic charges, which can be read easily into most available simulation packages. To illustrate the method, it is applied here to the calculation of a typical reaction path, the ring flip of Tyr35 in BPTI. In addition, the effect of NUCS on the structural stability of a protein is tested with molecular dynamics simulations.

9.1.2 Methods

Theory

The NUCS procedure accounts for the solvent screening effect on the interaction energies. As in the distance-dependent dielectric approach, the contributions of the solvation Born energies are neglected. Solvent screening weakens the net electrostatic interaction between two partial charges q_i and q_j of a globular protein in solution as compared to *in vacuo*, while leaving the sign of the interaction unchanged. This can be represented by an atom:atom pairwise screening factor, ϵ_{ij} (where r_{ij} is the inter-atomic distance):

$$E_{ij}^{screen} = \frac{q_i q_j}{\epsilon_{ij} r_{ij}}. \tag{9.1}$$

To evaluate Eq. 9.1 for a protein with N atoms one would have to store the $N \times N$ matrix of the ϵ_{ij} and to modify the Coulomb term in the simulation code so as to use the information stored in the matrix. To avoid this, an empirical combination rule is introduced that approximates the pairwise screening factor ϵ_{ij} in terms of atomic screening factors ϵ_i and ϵ_j:

$$\epsilon_{ij} = \sqrt{\epsilon_i \epsilon_j}, \tag{9.2}$$

where ϵ_i represents the average screening between an atom and the rest of the system. This is equivalent to performing a non-uniform scaling of the charges:

$$q_i \rightarrow q_i' = \frac{q_i}{\lambda_i}, \tag{9.3}$$

where the scaling factors are $\lambda_i = \sqrt{\epsilon_i}$. This then allows to obtain the screened interaction energy by simply using Coulomb's standard law with the scaled charges, *i.e.*, $E_{ij}^{screen} = \frac{q_i' q_j'}{r_{ij}}$. However, assigning different scaling factors λ_i to each partial charge would modify the first non-zero multipole moment of chemical groups. A net charge would appear on a neutral group with a dipole moment (e.g. a bond dipole), *i.e.*, the sum of the partial charges on that group would not add to zero after charge scaling. To avoid this, the scaling factor λ_i of all atoms belonging to a same chemical group I are assigned the same value, thus defining a group screening factor ϵ_I,

$$\lambda_i = \lambda_I \equiv \sqrt{\epsilon_I}. \tag{9.4}$$

This is meaningful because the average screening of an atom depends mostly on its degree of buriedness so that the atomic screening factors ϵ_i of neighboring atoms have similar values. The groups into which the protein is divided here are shown in Fig. 9.1. Each sidechain forms one group. The backbone is partitioned into peptide groups, which thus straddle over consecutive residues. For proline and glycine only a single group is defined comprising the sidechain and the preceding peptide moiety. Molecules other than the protein amino acid chain, such as structural water, substrates, cofactors, or ions, are treated each as a separate group. Large groups or molecules, such as a heme or ATP, are partitioned into smaller groups. For a given group I, the screening factor is obtained here from the ratio

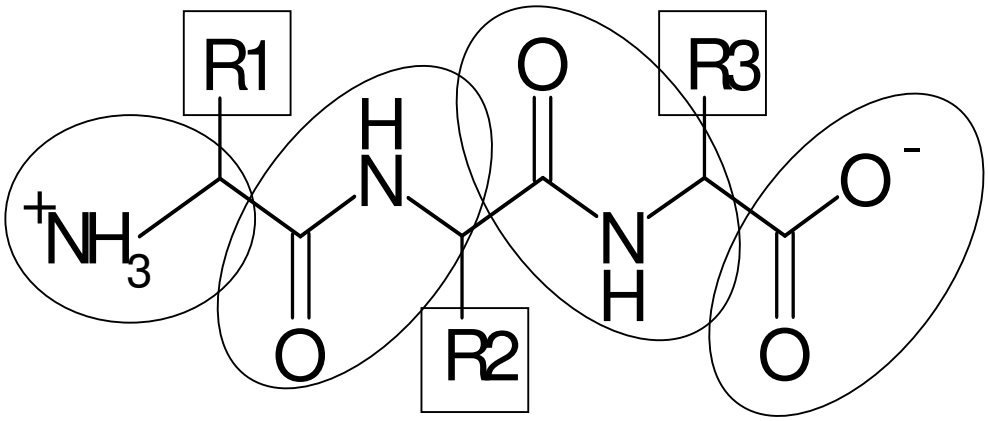

Figure 9.1: Definition of the groups for which different NUCS factors are determined. Each sidechain forms one group (squares). Backbone groups extend over two consecutive amino acids so as to preserve the local peptide dipole (circles).

$$\epsilon_I = \frac{E_I^{vac}}{E_I^{solv}}, \tag{9.5}$$

where E_I^{vac} is the "vacuum" ($\epsilon_{ij} = 1$) Coulomb interaction energy between the unscaled charges belonging to group I and the unscaled charges belonging to all other groups, and E_I^{solv} is the corresponding interaction energy in aqueous solvent calculated by solution of the Poisson-Boltzmann equation. In some cases E_I^{vac} and E_I^{solv} have different signs, which would yield imaginary scaling factors λ_I (from Eqs. 9.4 and 9.5). This happened because both positive and negative partial charges can be present within a group I, so that the interaction energy of group I (E_I^{vac} or E_I^{solv}) is a sum over atom:atom interaction energies of which some are negative (attractive) and some are positive (repulsive), leading to partial cancellation. When the net group interaction energy is close to zero (often the case for E_I^{solv}), the partial cancellation can change the net interaction from positive to negative or *vice versa*. Consequently, to ensure that the scaling factors are real, E_I^{vac} and E_I^{solv} are

first expanded in terms of group:group pairwise interaction energies, E_{IJ}^{vac} and E_{IJ}^{solv}, which yields the corresponding group:group screening factors ϵ_{IJ},

$$\epsilon_{IJ} = \frac{E_{IJ}^{vac}}{E_{IJ}^{solv}}. \tag{9.6}$$

ϵ_I is then computed as a weighted average over the ϵ_{IJ} which have the same sign for E_{IJ}^{vac} and E_{IJ}^{solv}, *i.e.*, for which $\epsilon_{IJ} > 0$. This can be written as follows:

$$\epsilon_I = \sum_{J,\epsilon_{IJ}>0} w_{IJ}\epsilon_{IJ}, \tag{9.7}$$

where w_{IJ} is a weighting factor ($0 \leq w_{IJ} \leq 1$, $\sum_{IJ} w_{ij} = 1$), chosen to give more weight to the strong interactions, which dominate the electrostatics:

$$w_{IJ} = \frac{|E_{IJ}^{solv}|}{\sum_{J,\epsilon_{IJ}>0} |E_{IJ}^{solv}|}. \tag{9.8}$$

Eq. 9.7 can thus be rewritten as

$$\epsilon_I = \frac{\sum_{J,\epsilon_{IJ}>0} |E_{IJ}^{vac}|}{\sum_{J,\epsilon_{IJ}>0} |E_{IJ}^{solv}|}. \tag{9.9}$$

Calculation of NUCS factors

To calculate the group-specific scaling factors, λ_I according to Eqs. 9.4 and 9.9 it is necessary to evaluate the group:group interaction energies in vacuum (E_{IJ}^{vac}) and in solution (E_{IJ}^{solv}). The vacuum interaction energy is given by Coulomb's law,

$$E_{IJ}^{vac} = \frac{1}{2} \sum_{i \in I} \sum_{j \in J} \frac{q_i q_j}{\epsilon r_{ij}} \cdot S(r_{ij}), \tag{9.10}$$

where the sum extends over all atoms i, j within groups I, J, ϵ is the dielectric permittivity, and $S(r_{ij})$ is some truncation function of the long-range nonbonded interaction (see below). The interaction energies in solution are

$$E_{IJ}^{solv} = \sum_{j \in J} q_j \Phi^I(\vec{r}_j), \tag{9.11}$$

where $\Phi^I(\vec{r}_j)$ is the electrostatic potential (at position $\vec{r}_j$ of atom j) due to the solvated charges of group I. $\Phi^I(\vec{r}_j)$ is computed by numerically solving the linearized Poisson-Boltzmann equation (LPBE) that relates the electrostatic potential $\Phi(\vec{r})$ to the charge density $\rho(\vec{r})$,

$$\nabla \cdot \epsilon(\vec{r}) \nabla \Phi(\vec{r}) - \kappa' \Phi(\vec{r}) = -4\pi\rho(\vec{r}), \tag{9.12}$$

where κ' is the Debye-Hückel screening constant related to the ionic strength (here 145 mM). $\Phi^I(\vec{r})$ is obtained by setting all charges other than the charges of group I to zero. The LPBE was solved here using the finite-difference scheme of the PBEQ module in CHARMM (version 28) (289), using a focusing approach with a final grid spacing of 1.0 Å. The dielectric boundary was defined as the van der Waals surface of the protein after increasing all atomic radii by 0.7 Å from the CHARMM radii. The transition between the regions of high ($\epsilon(\vec{r}) = 80$) and low ($\epsilon(\vec{r}) = 1$) dielectric was smoothed over 1.5 Å on each side of this boundary. The optimal value for the interior dielectric constant of a protein is debatable. Here, $\epsilon = 1$ was chosen to be consistent with the procedure used to parameterize the partial charges of CHARMM that implicitly contain the intra-protein electronic polarizability (347).

Truncation of nonbonded interactions

In macromolecular simulations the long-range electrostatic interactions are often neglected to reduce the number of atom pairs that are considered, thus saving computation time. This is usually achieved by multiplying the Coulomb law with some function $S(r)$ which smoothly decays to zero beyond a defined cut-off distance (260), for example a cubic switching function,

$$S(r) = \begin{cases} 1, & r < a \\ \frac{(b^2 - r^2)(b^2 + 2r^2 - 3a^2)}{(b^2 - a^2)^3} & a \le r \le b \\ 0, & r > b \end{cases}, \tag{9.13}$$

where a is the cut-on distance and b is the cut-off distance, or a shift function

$$S(r) = \begin{cases} \left[1 - \frac{r}{b}\right]^2 & r \le b \\ 0 & r > b \end{cases}. \tag{9.14}$$

The effects of these functions in combination with both constant and distance-dependent dielectric approaches have been extensively studied (348; 349; 350). The nonbonded truncation function constitutes an implicit screening of the middle and long-range electrostatic interactions. To avoid over-screening, once by truncation and again by the present charge scaling, it is preferable to calculate the scaling factors under consideration of the nonbonded truncation scheme that will be used in the subsequent simulations. This is done by applying the same truncation function $S(r)$ in Eq. 9.10 when computing E_{IJ}^{vac}. Moreover, to avoid that the many weak very-long range interactions E_{IJ}^{solv} overwhelm the more significant localized interactions in the computation of the weighting factor w_{IJ}, only those interaction E_{IJ}^{solv} are kept in Eq. 9.8 that are within some cutoff distance. It is convenient to use the same cutoff distance b that is used in the truncation function $S(r)$. This is achieved by using $\epsilon_{IJ} > 0$ (as opposed to $\epsilon_{IJ} \geq 0$) in the criterion in the sum in Eq. 9.8 (since from Eqs. 9.6 and 9.10 $\epsilon_{IJ} = 0$ when groups I and J are separated by more than the cutoff distance). This facilitates the computation of Eq. 9.9, where the same criterion can be used in the sums of the numerator and the denominator.

Counter-scaling

The approximations introduced by the combination rule (Eq. 9.2), the neglect of the group-pairs for which $\epsilon_{IJ} \leq 0$ (Eq. 9.9), and the truncation function can introduce a systematic deviation between the reference solvated interaction E_I^{solv},

$$E_I^{solv} = \sum_{J, r_{IJ} < b} E_{IJ}^{solv} \tag{9.15}$$

and the screened interaction of group I computed with scaled charges,

$$E_I^{screen} = \sum_{J, r_{IJ} < b} E_{IJ}^{vac}, \tag{9.16}$$

where E_{IJ}^{vac} is computed in Eq. 9.10 with charges q_i' scaled as in Eq. 9.3. Ideally, the E_I^{screen} would match the corresponding E_I^{solv} for all groups I. The systematic deviation is removed by dividing all scaling factors λ_i by a counterscaling factor γ, $\lambda_i' = \frac{\lambda_i}{\gamma}$, where γ is taken as the square root of the slope of the least-square linear fit between E_I^{screen} and E_I^{solv}. In the results presented here, the charges were scaled by the λ_i', one counterscaling factor being determined for the sidechains and another for the backbone groups. Using this counterscaling in Eq. 9.16 yields E_I^{NUCS}.

Testing of the NUCS factors

To examine the ability of the NUCS approach to reproduce the solvent screening effect, the interaction energies E_{IJ}^{NUCS} computed with the NUCS procedure described above were compared to the corresponding solvated reference interactions E_{IJ}^{solv} obtained from the Poisson-Boltzmann equation. This comparison is made for individual interactions between group:group pairs as well as for the interaction between each group I and the rest of the protein by comparing E_I^{NUCS} and E_I^{solv}. The screened interaction energies of the NUCS approach are contrasted to the interaction energies in absence of screening, E_I^{cdiel} (obtained as E_I^{screen}, but using unscaled charges and $\epsilon = 1$ in Eq. 9.10, as well as compared to the screened interactions in the standard distance-dependent dielectric approach, E_I^{rdiel} (obtained by using unscaled charges and $\epsilon = r$ in Eq. 9.10). The individual group:group interaction energies E_{IJ}^{cdiel} and E_{IJ}^{rdiel} are also directly compared to the reference values E_{IJ}^{solv}. These comparisons were performed for four different proteins, selected to span a wide range of protein sizes. These are the bovine pancreatic trypsin inhibitor (BPTI, 58 residues, PDB code 1BPI (351)), lysozyme (129 residues, PDB code 193L (352)), Staphylococcal nuclease (SNase, 136 residues, PDB code 1STN (353)), and myosin II subfragment S1 (745 residues (55)). Surface water molecules in the crystallographic structures were removed, but internal water was kept.

The two most widely used cutoff functions $S(r)$, the switch and shift functions, are often used in conjunction with a distance-dependent dielectric permittivity ($\epsilon = r$). Therefore, NUCS factors were determined and tested for four different functional forms:

I. Constant ϵ ($\epsilon = 1$) in combination with a cubic function switching from 6 to 12 Å.

II. Constant ϵ ($\epsilon = 1$) in combination with a shift function cutting at 12 Å.

III. Distance-dependent ϵ ($\epsilon = r$) in combination with a cubic function switching from 6 to 12 Å.

IV. Distance-dependent ϵ ($\epsilon = r$) in combination with a shift function cutting at 12 Å.

For each of these four functionals, the corresponding ϵ and $S(r)$ was used to compute the Coulomb interaction energies E_{IJ}^{vac} in Eq. 9.10, both during the determination of the scaling factors and their subsequent testing.

Application to Tyr35 ring flip in BPTI

To illustrate the effect of protein solvation on a simple conformational transition, the minimum-energy path (MEP) of flipping of the phenyl ring of the Tyr35 sidechain in BPTI was calculated using Poisson-Boltzmann continuum electrostatics (referred to hereafter as

"PB path"). This path is compared to the MEP obtained using a vacuum Coulomb potential with unscaled charges ("vacuum path"), the MEP obtained using a distance-dependent dielectric constant with unscaled charges ("rdiel path"), and to the MEP obtained with NUCS ("NUCS path").

The MEP was determined using the conjugate peak refinement (CPR) method (315). CPR is a minimization-based algorithm designed for the determination of reaction paths in highly complex systems without requiring a pre-defined reaction coordinate. It finds the MEP and its transition state(s) by optimizing a crude initial guess of the path, usually some interpolation between the reactant and product states. The default CPR settings in the TReK module of the program CHARMM were used, without ultra-optimization of the saddle-points. The end-states of the 180° ring-flip of Tyr35 were generated by exchanging the coordinates of atoms of the Tyr35 phenyl ring atoms: $C_{\delta 1} \leftrightarrow C_{\delta 2}$, $C_{\epsilon 1} \leftrightarrow C_{\epsilon 2}$, $H_{\delta 1} \leftrightarrow H_{\delta 2}$, and $H_{\epsilon 1} \leftrightarrow H_{\epsilon 2}$. This means that the end states have exactly the same energy. Because Cartesian interpolation between these end states generates structures with singular energies, the initial path was generated via a series of constrained minimizations, in which the torsion angle $C_\alpha - C_\beta - C\gamma - C_{\delta 1}$ of Tyr35 was constrained at successive values. During all path calculations the distant half of BPTI (residues 1 to 6, 23 to 31, and 48 to 58) was kept fixed.

The PB-path was obtained using the PBEQ module of CHARMM to solve the LPBE for the electrostatic energy. For this, the finite-difference grid (grid spacing 0.5 Å) extended at least 15 Å beyond the protein surface. The boundary between the high ($\epsilon_{solvent} = 80$) and low ($\epsilon_{protein} = 1$) dielectric regions was defined using atomic radii that have been optimized to reproduce experimental solvation energies (354). The nonpolar contribution to the solvation energy was taken as proportional to the protein surface area, multiplying it by a surface tension coefficient of 25 cal/mol/Å^2 (355; 356). The end-states used for the CPR calculation were energy minimized to a final root-mean-square gradient of 0.01 kcal/mol/Å. For the vacuum, rdiel, and NUCS paths, the end states were further minimized to a final RMS gradient of 0.001 kcal/mol/Å, using a cubic switching function (function form I) with a constant ($\epsilon = 1$, vacuum and NUCS paths) and a distance-dependent dielectric permittivity ($\epsilon = r$, rdiel path). To favor paths close to the reference PB-path, the saddle points of the PB-path were included in the initial guess paths for CPR.

Molecular dynamics simulations with NUCS

To test the stability of the protein structure when NUCS is used, three molecular dynamics simulation (MD) trajectories were generated for SNase, using unscaled charges ($\epsilon = 1$), a distance-dependent dielectric permittivity ($\epsilon = r$), and charges scaled according to the NUCS scheme (function form I, $\epsilon = 1$). The simulations were performed with the CHARMM param22 (347) all-atom force field. After geometry optimization to a final root-mean square gradient of 0.001 kcal/mol/Å the system was heated to 300 K with ve-

locity scaling corresponding to temperature increments of 1 K every 0.1 ps using a time step of 1 fs. At 300 K, 100 ps equilibration with velocity rescaling every 0.1 ps, another 100 ps equilibration with velocity rescaling every 1.0 ps, and 200 ps equilibration without velocity rescaling were followed by 2 ns production runs. The root-mean-square coordinate deviation (RMSD) of the backbone heavy atoms were determined and compared to the RMSD from explicit water reference MD simulations (357). For comparison with a GB method, the MD simulations were repeated using ACE2 (358) with the CHARMM param19 united-atom force field (359).

9.1.3 Results

The NUCS factors

A statistical overview of the NUCS factors λ_I and counterscaling factors γ obtained for the different proteins and functional forms is given in Table 9.1. There is no clear correlation between the scaling factors and the size of the protein. The cutoff method has minor influence on the NUCS factors, the factors determined with a switch function being somewhat larger than those obtained with a shift function. For example, the mean counterscaled NUCS factor averaged over the four proteins is 1.49 ($\epsilon = 1$) for switching whereas it is only 1.39 ($\epsilon = 1$) for shifting. This is consistent with there being less implicit screening by using a switch function (which reduces interactions only in the 6-12 Å range) than with a shift function (which reduces interactions over the whole 0-12 Å range). In contrast, the use of $\epsilon = 1$ *versus* $\epsilon = r$ has a large effect on the NUCS factors, yielding significantly smaller scaling factors when $\epsilon = r$ is used: the average NUCS factor is 1.44 for $\epsilon = 1$ and 1.08 for $\epsilon = r$. Again, this reflects the large implicit screening of the $\epsilon = r$ functional, which thus needs less additional screening from the charge scaling. However, this is not to say that $\epsilon = r$ and NUCS are equivalent, since NUCS better accounts for the spatial dependence of screening. Fig. 9.2 illustrates this by showing how the value of scaling factors varies within a protein. As expected, the scaling factors for residues in the core of the protein are smaller (closer to 1, *i.e.*, no solvent screening) than those found for surface residues.

Counterscaling factors were calculated separately for the backbone (γ_{BACK}) and sidechain groups (γ_{SIDE}). For the functional forms I and II ($\epsilon = 1$) the counterscaling factors γ are < 1, indicating that without counterscaling Eq. 9.9 would slightly overestimate the screening. For the functional forms III and IV ($\epsilon = r$) the counterscaling factors are around 1, indicating that counterscaling is less important and can be dispensed with when using $\epsilon = r$. Counterscaling factors were also determined using unscaled charges in Eqs. 9.10 and 9.16 and with either CDIEL ($\epsilon = 1$) or RDIEL ($\epsilon = r$). The results are listed at the bottom of Table 9.1. This shows that on average electrostatic interaction energies are overestimated when using CDIEL, whereas this is not the case when using RDIEL.

function form		BPTI	Lysozyme	SNase	myosin
$\epsilon = 1$, switch	min	0.81	0.74	0.79	0.80
(form I)	max	5.46	5.95	5.83	7.40
	ave $\pm$ sd	1.48 $\pm$ 0.82	1.41 $\pm$ 0.70	1.51 $\pm$ 0.86	1.55 $\pm$ 0.90
	γ_{BACK}	0.71	0.68	0.67	0.70
	γ_{SIDE}	0.82	0.75	0.79	0.90
$\epsilon = 1$, shift	min	0.85	0.83	0.84	0.87
(form II)	max	4.43	3.20	4.79	5.97
	ave $\pm$ sd	1.40 $\pm$ 0.62	1.28 $\pm$ 0.41	1.43 $\pm$ 0.66	1.45 $\pm$ 0.64
	γ_{BACK}	0.83	0.84	0.82	0.86
	γ_{SIDE}	0.93	0.90	0.91	0.99
$\epsilon = r$, switch	min	0.74	0.71	0.73	0.64
(form III)	max	3.06	3.89	2.97	3.57
	ave $\pm$ sd	1.14 $\pm$ 0.39	1.09 $\pm$ 0.37	1.11 $\pm$ 0.34	1.05 $\pm$ 0.35
	γ_{BACK}	0.93	0.93	0.95	0.92
	γ_{SIDE}	1.12	0.99	1.03	1.07
$\epsilon = r$, shift	min	0.72	0.72	0.72	0.62
(form IV)	max	2.42	3.63	2.32	2.83
	ave $\pm$ sd	1.09 $\pm$ 0.30	1.09 $\pm$ 0.33	1.06 $\pm$ 0.25	1.01 $\pm$ 0.26
	γ_{BACK}	0.98	0.98	1.01	0.98
	γ_{SIDE}	1.17	1.08	1.09	1.13
CDIEL[a]	γ	1.29	1.28	1.47	1.23
RDIEL[b]	γ	1.08	1.01	1.06	0.82

[a]Coulomb potential with $\epsilon = 1$ and a switching cutoff function, unscaled charges.
[b]Coulomb potential with $\epsilon = r$ and a switching cutoff function, unscaled charges.

Table 9.1: Statistical overview of NUCS factors λ_I (Obtained from Eqs. 9.4 and 9.9, with counterscaling.) and counterscaling factors γ. Different counterscaling factors were used for sidechain groups and backbone groups.

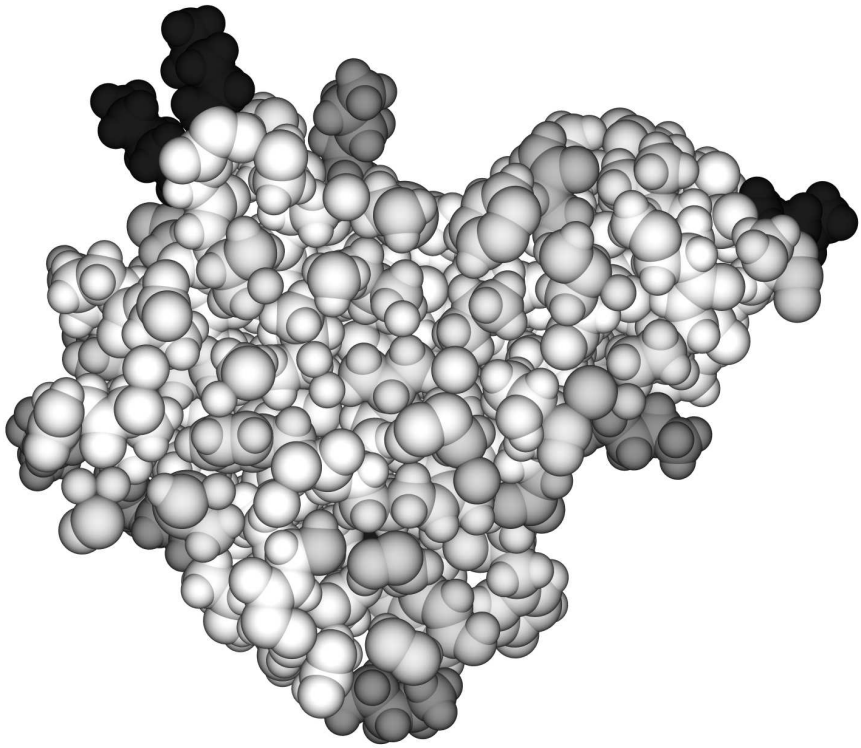

Figure 9.2: Cut through the middle of Staphylococcal nuclease. Groups are shaded proportionally to the value of their NUCS factor (for functional form I). White corresponds to a factor of about 1 (*i.e.*, no scaling) and dark gray to a factor of about 6.

Testing the NUCS approach

To evaluate how the NUCS approach accounts for the solvent screening, the NUCS electrostatic interaction energy between each group and the rest of the protein (E_I^{NUCS} from Eq. 9.16) is plotted against the reference solvated interaction (E_I^{solv}, Eq. 9.15) for Staphylococcal nuclease (Fig. 9.3A). For comparison, the corresponding interactions obtained with the Coulomb potential without charge scaling, either in vacuum ($\epsilon = 1$, E_I^{cdiel}) or with a distance-dependent dielectric ($\epsilon = r$, E_I^{rdiel}) are also plotted. Fig. 9.3A shows that treatment of the solvent screening is significantly improved with the NUCS approach. E_I^{cdiel} shows very large deviations from E_I^{solv} and poor correlation (correlation coefficient c.c.=0.67). The introduction of a distance-dependent dielectric permittivity improves the correlation (c.c.=0.89). Using NUCS, a similar correlation is observed (c.c.=0.92), however the deviations from E_I^{solv} are larger for E_I^{rdiel} than for E_I^{NUCS} for large energies. Because the large interaction energies contribute most to the total electrostatic energy, it is of par-

ticular importance to adequately represent them. This is reflected in Table 9.2 which lists the RMS energy deviations (RMSD) relative to E_I^{solv} of E_I^{cdiel}, of E_I^{rdiel}, of E_I^{NUCS}, and of interactions calculated after scaling the charges according to the distance-from-center scaling approach of Ref. (345). The deviations obtained with unscaled charges or charges scaled according to their distance from the protein center are significantly larger than with NUCS. For example in the case of myosin the RMSD is 29.7 kcal/mol when using unmodified charges with $\epsilon = 1$. The use of a distance-dependent dielectric permittivity reduces this RMSD to 13.4 kcal/mol, respectively to 15.57 kcal/mol with distance-from-center charge scaling. With NUCS, the RMSD is much smaller, ranging between 3.3 kcal/mol (functional form II) and 4.5 kcal/mol (functional form I). For the smaller proteins the RMSD with NUCS is nearly half the RMSD with unscaled charges and a distance-dependent dielectric permittivity, for example 1.9 kcal/mol for Staphylococcal nuclease (functional form IV) *versus* 5.83 kcal/mol. Table 9.2 indicates that in general, the combination of NUCS with a shift cutoff function leads to a smaller deviation than with a switch cutoff function. Combining NUCS with a distance-dependent dielectric permittivity ($\epsilon = r$) rather than using a constant dielectric ($\epsilon = 1$), further reduces the deviation from the solvated interaction energies.

The analysis of the deviation from the reference solvated interactions was repeated for the individual group:group interaction energies E_{IJ} that sum up to E_I in Eq. 9.16. The corresponding RMS deviations are listed in Table 9.3. The RMSD(E_{IJ}^{cdiel}) for myosin is 4.2 kcal/mol. Introducing a distance-dependent dielectric permittivity reduces this RMSD(E_{IJ}^{rdiel}) to 1.6 kcal/mol. Using NUCS the RMSD(E_{IJ}^{NUCS}) values for myosin range from 0.5 kcal/mol (function form II) to 0.8 kcal/mol (function form I). The improvement of NUCS over $\epsilon = r$ is apparent in the plot of E_{IJ}^{NUCS} and E_{IJ}^{rdiel} versus E_{IJ}^{solv} shown in Figure 9.3B. The smallest RMSD, 0.3 kcal/mol, is observed for Staphylococcal nuclease (function form IV). Again, the smallest RMSD values are found when combining NUCS with a distance-dependent dielectric permittivity ($\epsilon = r$) and a shift cutoff function. To estimate how the NUCS interaction energies compare with those of a GB method, the group:group interaction energies have been recalculated for myosin using ACE (334; 335) with parameters optimized for large proteins (360). The resulting RMSD(E_{IJ}^{ACE}) is 0.6 kcal/mol and thus within the range of RMSD values obtained with NUCS.

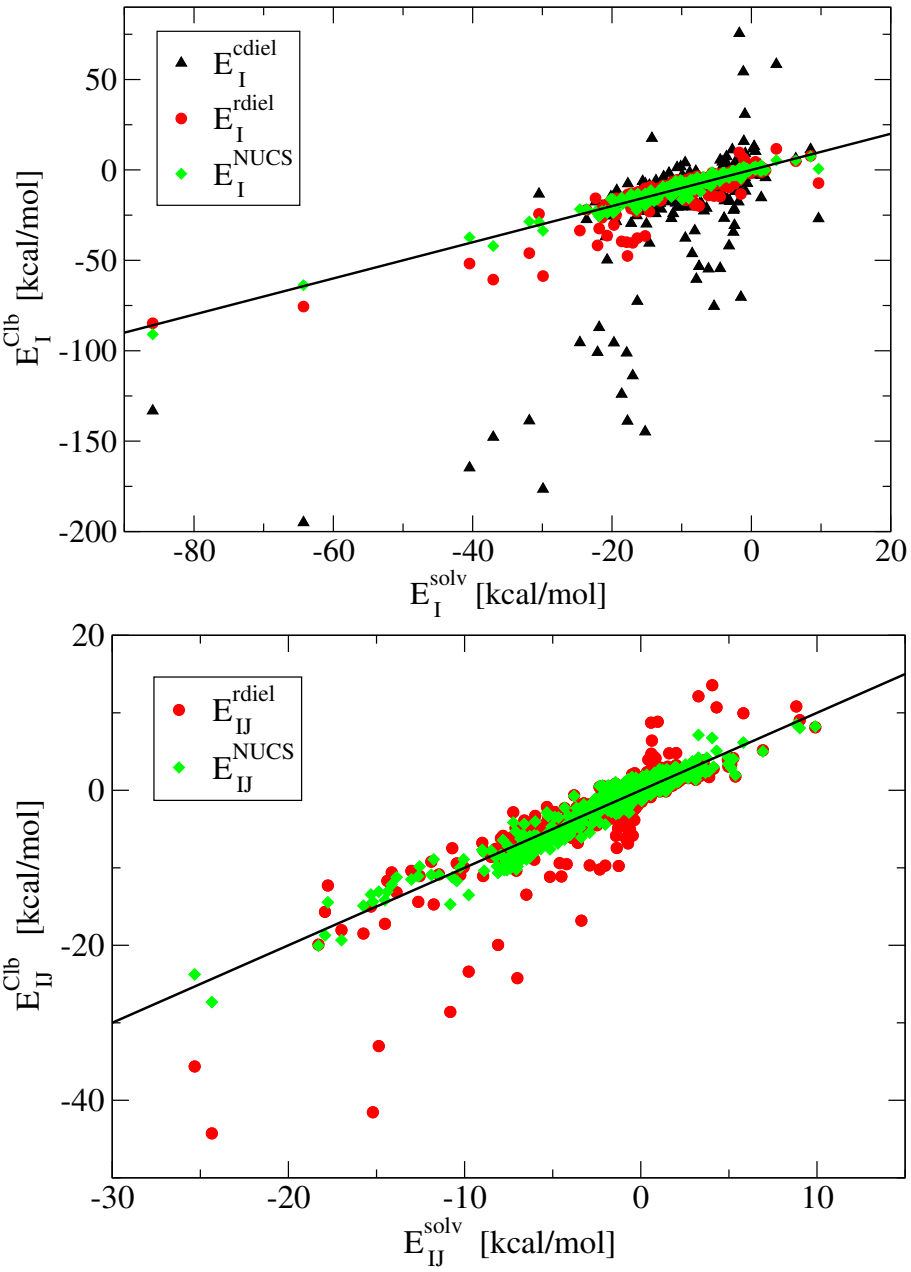

Figure 9.3: A) Comparing Coulombic interaction energies (Eq. 9.16) to the reference solvated interactions E_I^{solv}(Eq. 9.15). The interactions between each group and the rest of the protein, E_I^{cdiel}, E_I^{rdiel}, and E_I^{NUCS}, are for Staphylococcal nuclease and functional form III. B) Comparing Coulombic pair-wise interactions to the reference solvated interactions E_{IJ}^{solv}. The group:group interaction energies, E_{IJ}^{rdiel} and E_{IJ}^{NUCS}, are for Staphylococcal nuclease and functional form III. The diagonal line indicates ideal agreement.

	unscaled charges		NUCS				
function form[a]	CDIEL	RDIEL	d_{center}[b]	I ($\epsilon = 1$)	II ($\epsilon = 1$)	III ($\epsilon = r$)	IV ($\epsilon = r$)
BPTI	16.73	5.11	5.29	3.49	3.09	2.86	2.83
Lysozyme	18.84	5.60	8.77	4.45	3.41	3.67	3.48
SNase	28.78	5.83	9.25	3.93	2.84	2.05	1.90
Myosin	29.66	13.37	15.57	4.54	3.26	3.52	3.97

[a]CDIEL: switched cutoff and $\epsilon = 1$, RDIEL: switched cutoff and $\epsilon = r$.

[b]d_{center}: charges scaled according to the distance-from-center approach described in Ref. (345), using a switched cutoff and $\epsilon = 1$.

Table 9.2: Root mean square energy deviations (in kcal/mol). Deviations are calculated between the reference interaction energy E_I^{solv} (Eq. 9.15) and the Coulombic interaction energy (Eq. 9.16), for interactions taken between individual groups I and the rest of the protein.

	unscaled charges		NUCS			
function form[a]	CDIEL	RDIEL	I ($\epsilon = 1$)	II ($\epsilon = 1$)	III ($\epsilon = r$)	IV ($\epsilon = r$)
BPTI	2.65	0.58	0.55	0.39	0.33	0.31
Lysozyme	2.22	0.60	0.70	0.47	0.42	0.41
SNase	3.69	0.68	0.66	0.42	0.29	0.28
Myosin	4.20	1.57	0.84	0.51	0.63	0.80

[a]CDIEL: switched cutoff and $\epsilon = 1$, RDIEL: switched cutoff and $\epsilon = r$.

Table 9.3: Root mean square energy deviations (in kcal/mol). Deviations are calculated between the reference interaction energy E_{IJ}^{solv} (Eq. 9.11) and the Coulombic interaction energy (Eq. 9.10), for interactions taken between pairs of groups I and J.

Transferability of NUCS factors

The amount of solvent screening experienced by a partial charge in a protein depends mostly on how buried the charge is. Therefore, the scaling factors that represent the screening are mostly geometry dependent and are expected to be determined more by the overall protein conformation than by the precise value of the partial charges. This is convenient, because it means that scaling factors determined once for a given protein structure can be transferred between different force-fields which have different charge parameterizations. To test the transferability of the NUCS factors between different force fields, the factors were recalculated for Staphylococcal nuclease using partial atomic charges from the AMBER (361; 362) force field (function form I). Fig. 9.4A shows a plot of the NUCS factors thus determined versus factors determined for CHARMM charges. It shows a high correlation (correlation coefficient = 0.88) between the scaling factors obtained for the two force-fields. This indicates the degree of robustness of the NUCS factors with respect to charge variation (see Fig. 9.4B). Therefore, the scaling factors calculated using the partial atomic charges of one force field can be usefully transferred to another force field, after a rescaling by the average of factor ratios that can be determined once for each pair of force-fields (see caption of Fig. 9.4A).

The NUCS factors are dependent on the functional form (e.g., I-IV), as discussed above. However, Fig. 9.5 shows a nearly linear dependence between factors calculated using different functional forms. These linear relationships exist for all the proteins used in this study (not shown). Thus, in practice is not necessary to calculate the NUCS factors for all functional forms but only for a single functional form and then derive the factors for another functional form by simple rescaling (see caption of Fig. 9.5).

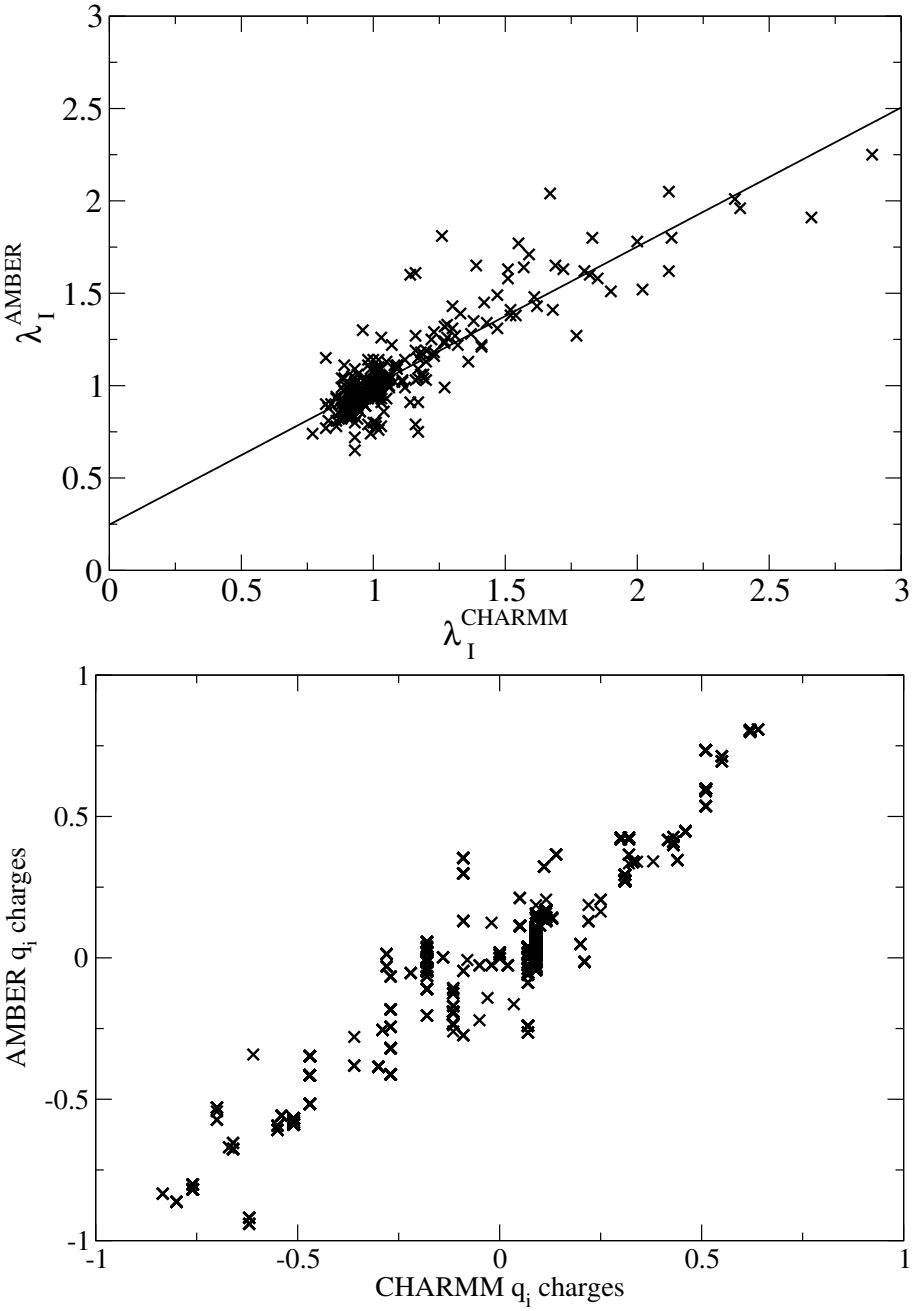

Figure 9.4: A) Dependence of NUCS factors on charges (AMBER *versus* CHARMM) for Staphylococcal nuclease (using function form III). A linear regression through (1,1) yields a slope of 0.75. The average ratio $\frac{\lambda_I^{AMBER}}{\lambda_I^{CHARMM}}$ is 0.98. The correlation coefficient is 0.88. B) Comparison of the partial atomic charges from the AMBER and CHARMM force-fields (for Staphylococcal nuclease).

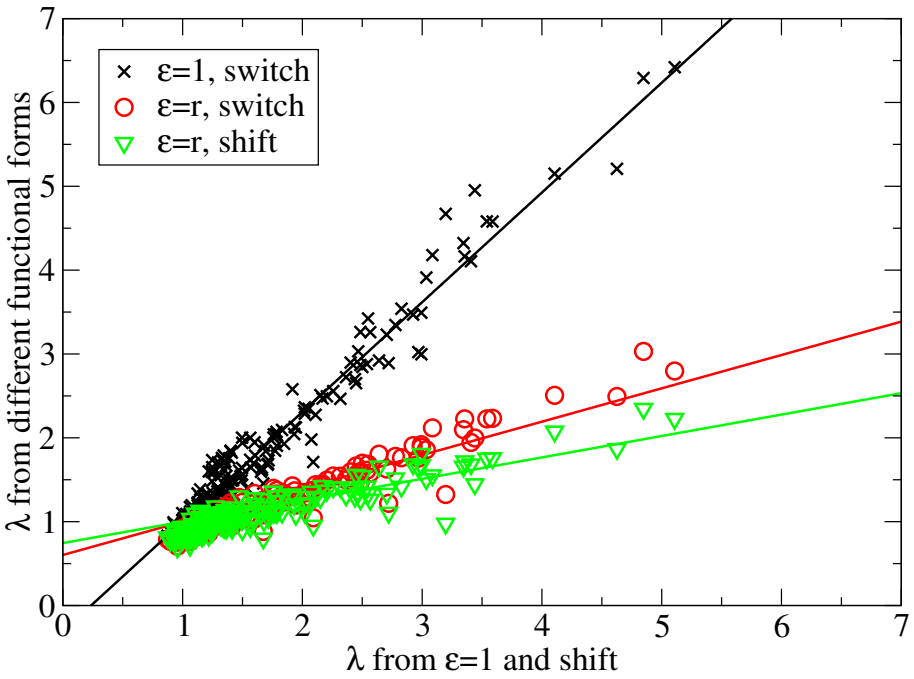

Figure 9.5: NUCS factors derived for Staphylococcal nuclease using different functional forms plotted against the factors obtained with functional form II ($\epsilon = 1$, shift). Linear regressions through (1,1) yield slopes of 1.31 ($\epsilon = 1$, switch), 0.40 ($\epsilon = r$, switch), and 0.26 ($\epsilon = r$, shift). The average ratios between factors from functional forms I, III, and IV and factors from functional form II are 1.05, 0.76, and 0.74, respectively.

Tyr35 ring flip in BPTI

The usefulness of the NUCS approach is illustrated on the minimum-energy path (MEP) for a simple conformational transition: the Tyr35 ring flip in BPTI. In the reactant state, the hydroxyl group of the Tyr35 sidechain is in the ring plane and points towards the solvent (Fig. 9.6A). To serve as a reference, the MEP was first optimized with Poisson-Boltzmann (PB) solvation, using CPR. The energy profile along that path is shown in Fig. 9.7A (black). A first transition state is found at a value of the normalized curvilinear reaction coordinate (RC) of 0.2 (appears as a shoulder along the curve), with a barrier of 13.9 kcal/mol, in which the hydroxyl group reorients to point towards the protein and is orthogonal to the ring plane (not shown). Throughout the whole path the Tyr35 hydroxyl group rarely forms hydrogen bonds with the protein, pointing mainly directly into the solvent and only occasionally forming contacts to the backbone carbonyl oxygens of Gly37 and Cys38. At RC = 0.45 the phenyl ring starts rotating. During this rotation, the Tyr35 hydroxyl group remains in the ring plane and follows its rotation. The rate-limiting transition state is reached at RC = 0.58 (shown in Fig. 9.6B), with a barrier of 32.5 kcal/mol. This is in excellent agreement with the experimental enthalpy barrier $\Delta H^{\ddagger} = 33$ kcal/mol (363; 364), although this agreement may be coincidental. At RC = 0.85 the ring rotation is complete and is followed by structural rearrangements of the protein environment.

Also shown in Fig. 9.7A are profiles along the PB path where the energy was re-computed (no geometry optimization) with a Coulomb potential (functional form I), using either unscaled charges, a distance-dependent dielectric constant, or charges scaled according to the NUCS procedure. The vacuum Coulomb potential (unscaled charges, $\epsilon = 1$) yields an energy profile that strongly differs from that calculated with the PB method. Firstly, the lowest-energy structure is now at RC = 0.17 (-9.33 kcal/mol). In this structure, the NH_3^+ group on the solvent-exposed Lys46 sidechain points towards the neighboring solvent-exposed negatively-charged sidechain of Asp50. The favorable interaction between these two groups is screened by the solvent in the PB solvation model. Secondly, the energy at the PB transition state (RC = 0.58), 10.5 kcal/mol, is considerably underestimated. The shape of the energy profile determined with a distance-dependent dielectric constant (unscaled charges, $\epsilon = r$) follows the shape of the vacuum energy profile, although the deviations from the PB profile are slightly milder than in the vacuum case. In contrast, when using NUCS, the general form of the profile more closely resembles that of the PB reference. The lowest-energy structure along the PB path remains the lowest in energy with NUCS. The energy at the transition state is around 23 kcal/mol with NUCS. This is much closer to the PB barrier height than the value calculated using unscaled charges.

We now examine how the shape of the PB path changes when subjected to re-optimization with another electrostatic method. The two transition state structures of the PB path (at RC 0.2 and 0.58) were taken as intermediates in the initial guess path for further CPR pathway optimizations, using either unscaled charges, a distance-dependent dielectric constant, or NUCS. The energy profiles of the resulting paths are shown in Fig. 9.7B. In

the unscaled case, the lack of screening of the electrostatic interaction between the solvent-exposed sidechains of Lys46 and Asp50 results in high energy peaks (the first and third peaks, 29 kcal/mol and 17 kcal/mol) corresponding to local structural rearrangements at the protein surface that are irrelevant for the actual ring flip. The ring flip takes place between RC = 0.33 and RC = 0.48 with a saddle point at RC = 0.42 ($\Delta H^{\ddagger}$ = 30.5 kcal/mol). In this path segment, the Tyr35 hydroxyl group rotation proceeds without a barrier (in contrast to the PB path, where this event formed a saddle point) and is directly followed by the barrier-causing event, which is the ring flipping accompanied by back-rotation of the hydroxyl group. The events along this path differ significantly from the events in the PB path, even though the PB path transition states had been used to seed the CPR. When using a distance-dependent dielectric constant, a large number of structural rearrangements in the protein environment take place between RC = 0 and RC = 0.29, at which point the barrierless rotation of the sidechain hydroxyl group of Tyr35 begins. At RC = 0.35 this rotation is completed. The minimum-energy structure along the rdiel path at RC = 0.54 (ΔE = -9.2 kcal/mol) corresponds to the starting structure for the ring rotation that proceeds via a saddle point at RC = 0.83 ($\Delta E^{\ddagger}$ = 6.1 kcal/mol) up to RC = 0.80. The ring flip is directly followed by back-rotation of the hydroxyl group up to RC = 0.85 with a barrier at RC = 0.83 ($\Delta E^{\ddagger}$ = -3.1 kcal/mol). Thus, as was the case for the unscaled path, the events along the rdiel path differ significantly from the events in the PB path. In contrast, the nature and sequence of events are adequately conserved in the path optimized using NUCS (Fig. 9.7B). Moreover, the overall shape of the energy profile is very similar to the PB reference path, although the energy barriers of 6.6 kcal/mol at RC = 0.26 (Tyr35 hydroxyl group rotation) and 17.2 kcal/mol at RC = 0.48 (ring flip accompanied by hydroxyl group back-rotation) are lower than the PB barriers.

MD trajectories of SNase

Fig. 9.8 shows the RMSD of the backbone heavy atoms of SNase with respect to the crystal structure during MD simulations generated with either the GB method ACE2, unscaled charges, a distance-dependent dielectric permittivity, or charges scaled according to the NUCS scheme. The average RMSD values are ∼3.3 Å, ∼2.9 Å, ∼2.2 Å, and ∼1.7 Å for the ACE2, unscaled, rdiel, and NUCS trajectories, respectively. The RMSD for the explicit water simulation is ∼1 Å (357). All simulations based on a Coulomb potential exhibit significantly smaller fluctuations in the RMSD than the ACE2 simulation, the latter displaying structural fluctuations that correspond better to those of the protein in explicit water than the Coulomb-based methods. Nevertheless, the average structural deviation along the NUCS trajectory is smaller than that obtained in the ACE2 and rdiel trajectories. Moreover, the structure of the protein remains stable throughout the NUCS simulation, indicating that the charge scaling did not introduce large artefactual strain into the protein.

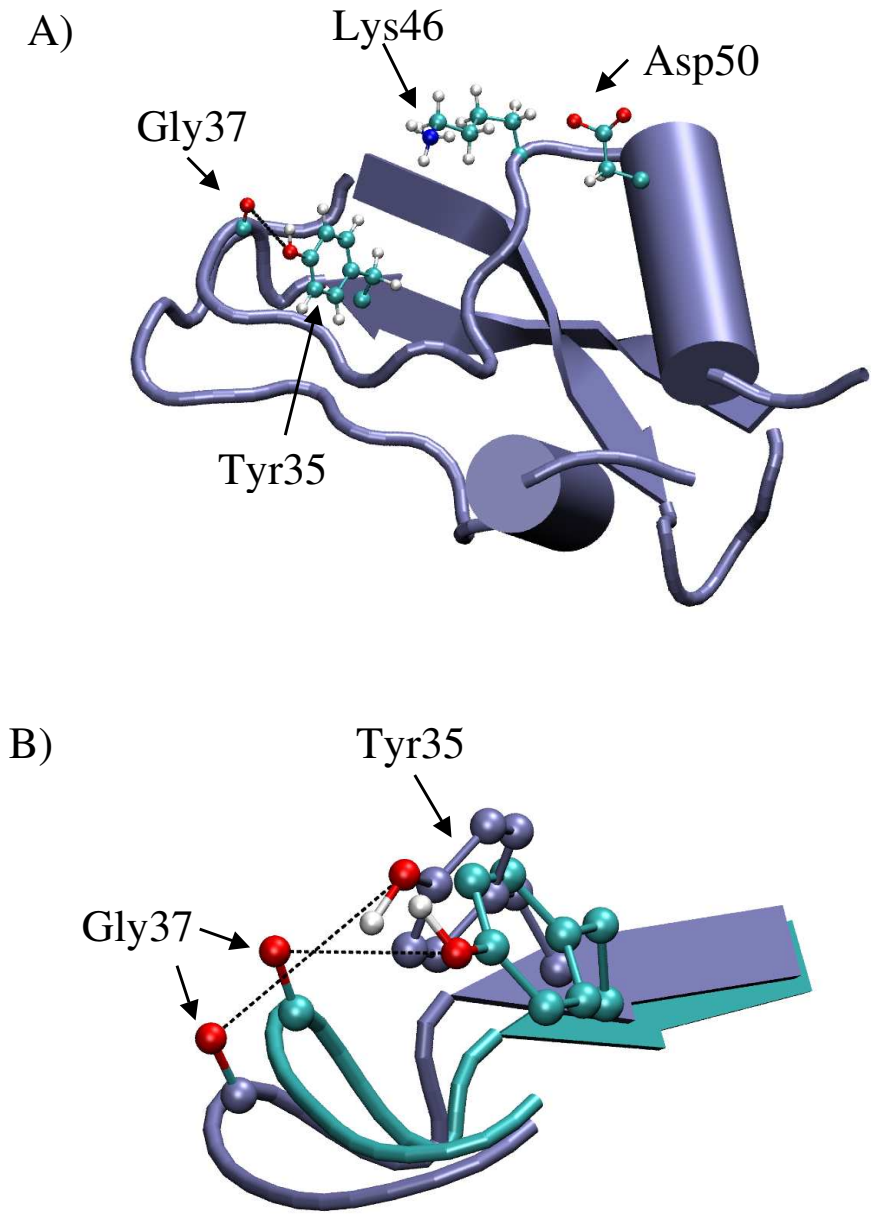

Figure 9.6: Tyr35 ring flip in BPTI, optimized with the Poisson-Boltzmann solvation model. A) BPTI structure in the reactant state. B) Overlap of the reactant state (green) and transition state (RC = 0.58 in Fig. 9.7A, in blue).

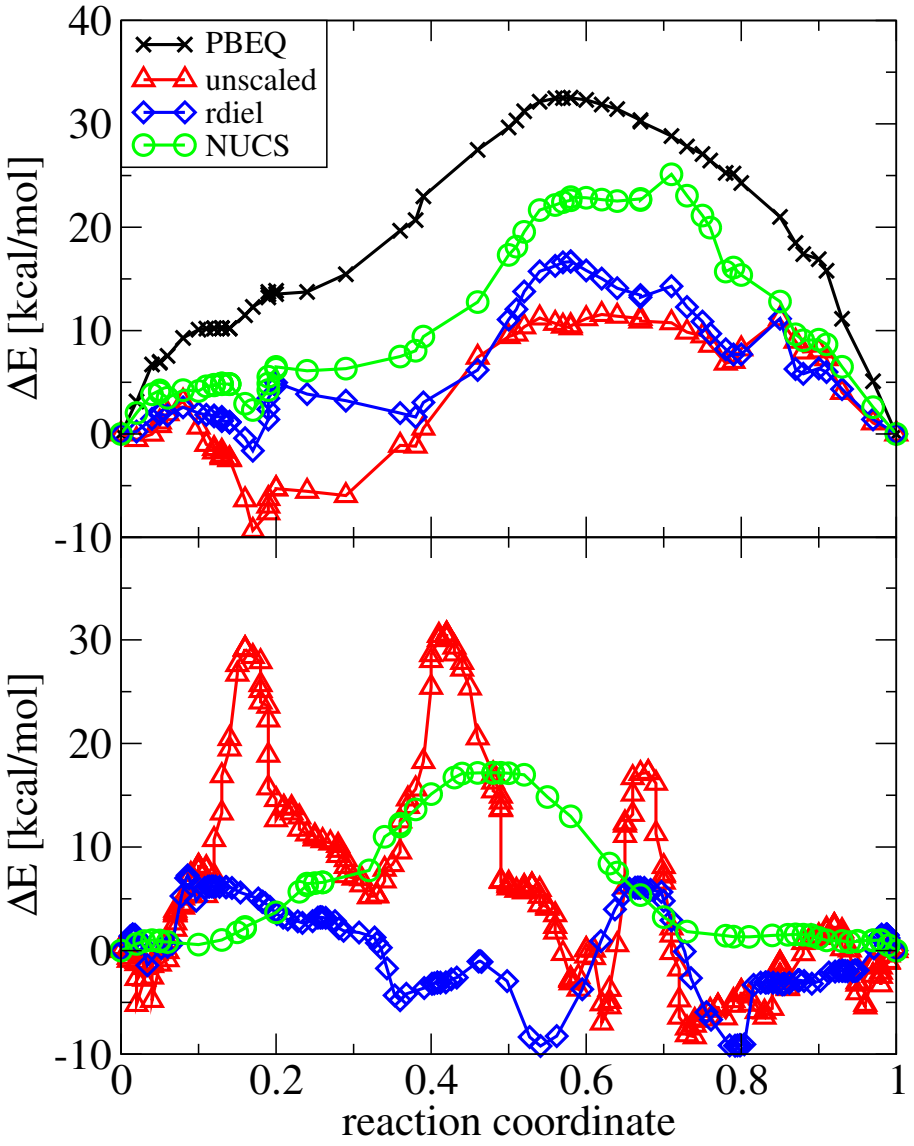

Figure 9.7: Energy profiles for the ring flip of Tyr35 in BPTI. A) Minimum-energy path calculated with Poisson-Boltzmann continuum electrostatics (black crosses), and point energy calls along that path using a Coulomb potential (switch function) with either unscaled charges (red triangles), a distance-dependent dielectric constant (blue diamonds), or NUCS (green circles). B) Minimum-energy path optimised by CPR, using either unscaled charges (red triangles), a distance-dependent dielectric constant (blue diamonds), or NUCS (green circles), taking the saddle points from the Poisson-Boltzmann path (Panel A) as intermediates in the initial path for CPR. The reaction coordinate (RC) is the curvilinear distance along the path (normalised), measured as RMS change in all Cartesian coordinates of the protein.

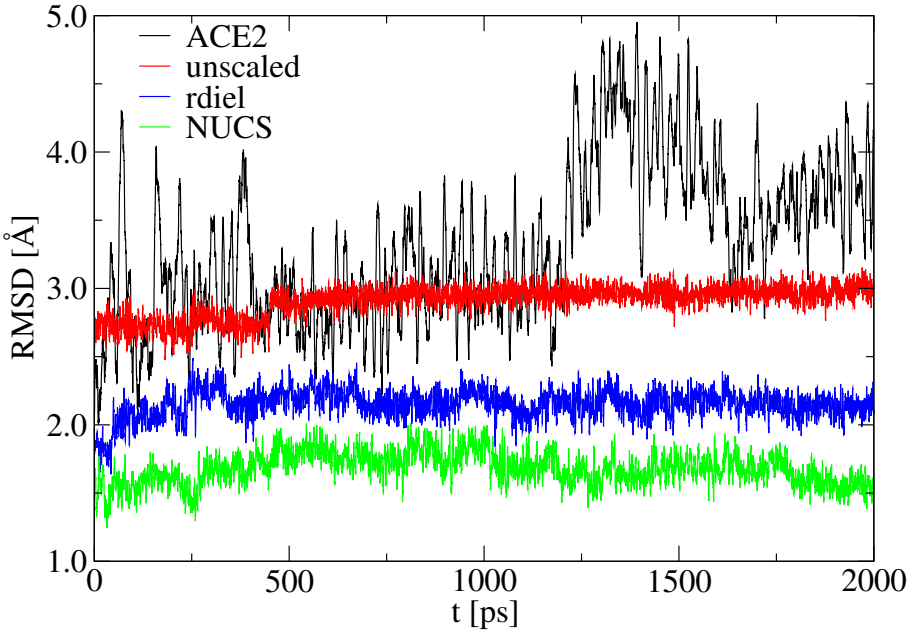

Figure 9.8: RMSD of the backbone heavy atoms of SNase with respect to the crystal structure for MD trajectories generated with ACE2 (black, top curve), unscaled charges (red, upper middle curve), a distance-dependent dielectric constant (blue, lower middle curve), or NUCS (green, bottom curve). The reference RMSD from an explicit water simulation is ∼1 Å (357).

Conformation dependence of NUCS factors

The conformation dependence of the NUCS factors is investigated by comparing the results obtained from the C/O (PDB code 1MMD) (57) and C/C (ref1) (55) conformations of myosin. A statistical overview of the NUCS factors λ_I and counterscaling factors γ obtained for both structures is given in Table 9.4.

	C/O	C/C
λ_{min}	1.07	1.08
λ_{max}	7.62	8.22
λ_{ave}	1.93	1.93
γ_{BACK}	0.71	0.70
γ_{SIDE}	0.87	0.90

Table 9.4: Statistical overview of NUCS factors λ_I (before counterscaling) and counterscaling factors γ obtained for myosin in the C/O and C/C conformations.

The raw NUCS factors (before counterscaling) range from 1.07 (C/O) to 8.22 (C/C) with an average of 1.93 for both structures. Counterscaling factors were calculated separately for backbone (γ_{BACK}) and sidechain groups (γ_{SIDE}). They are all smaller than 1 indicating that the scaling procedure without counterscaling slightly overestimates screening.

The dependence of NUCS factors on conformation can be investigated by comparing the resulting factors for myosin II subfragment S1 for the two conformations, C/O and C/C. Although this is a major conformational change an obvious correlation is observed (Fig. 9.9). In this figure, surface groups are indicated by the use of a filled symbol (a group is defined as a surface group if at least one of its atoms has an atomic solvent accessible surface > 16 Å^2). Solvent accessibility was calculated using the built-in surface module in CHARMM. The figure shows that all residues with high scaling factors are at the protein surface.

Fig. 9.10 shows a plot of interaction energies between each group I and the rest of the protein determined for the for the C/C conformation calculated using a Coulomb potential with unscaled charges and a uniform dielectric ($\epsilon = 1$, E^{vac}), with unscaled charges and a distance-dependent dielectric permittivity ($\epsilon = r$, E^{rdiel}), and with scaled charges and a uniform dielectric ($\epsilon = 1$, E^{shield}) *versus* the reference Poisson-Boltzmann energies (E^{solv}). As was the case for Staphylococcal nuclease, there is very poor correlation between E^{vac} and E^{solv} that is improved when introducing a distance-dependent dielectric permittivity. Using NUCS, however, improves the correlation even further. This is also reflected in the RMSD values that are collected in Table 9.5.

When simulating the conformational transitions a protein undergoes one might wish to use the average NUCS factors obtained for the two end state structures. Therefore, the average of the counterscaled NUCS factors obtained for the C/O and C/C structures of

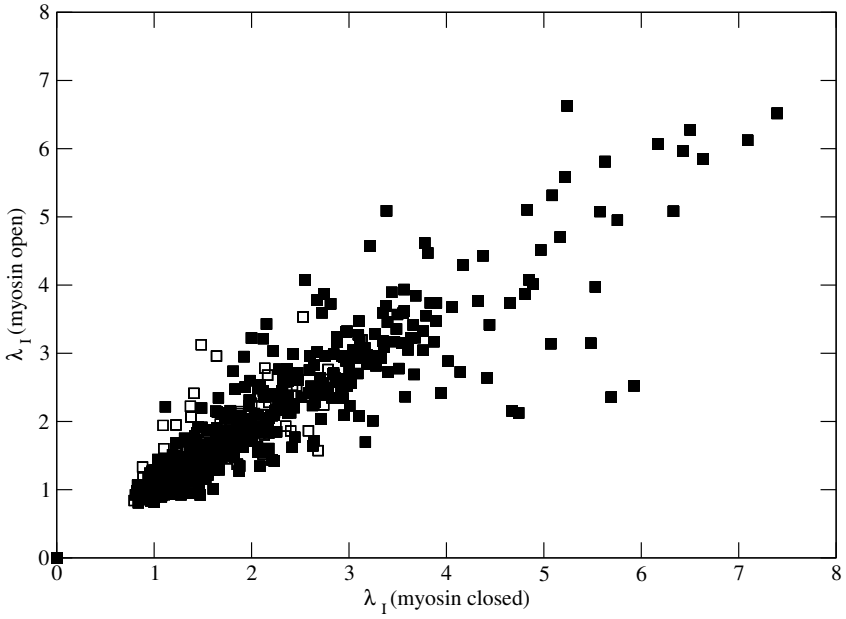

Figure 9.9: Counterscaled NUCS factors λ_I calculated for groups I using the myosin C/O structure versus factors for the identical groups I obtained using the myosin C/C structure. Filled symbols indicate surface residues.

	E^{vac}	E^{rdiel}	E^{shield}
C/O	34.1	9.9	5.5
C/C	29.7	13.4	4.5

Table 9.5: Root mean square energy deviations of the interaction energies between each group and those atoms of the protein within the cutoff distance from the reference values obtained from Poisson-Boltzmann interaction energies. Units are kcal/mol.

myosin were determined and used to recalculated the electrostatic interaction energies, E^{shield}, for the C/C structure. Fig. 9.11 shows that the use of these average factors leads to a correlation between E^{shield} and E^{solv} similar to the correlation observed when using the original factors, indicating the usefulness of the averaged NUCS factors in the context of the study of conformational transitions.

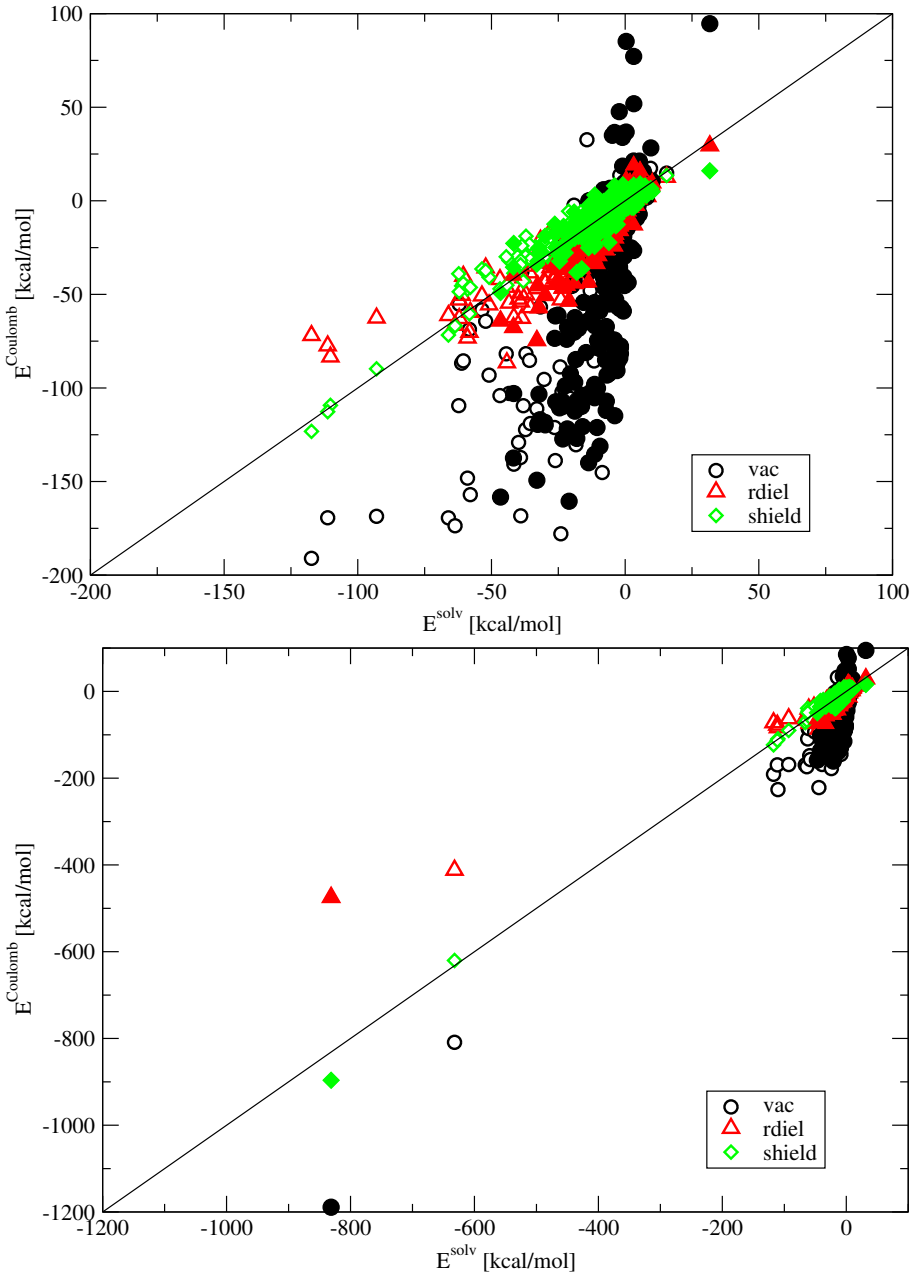

Figure 9.10: Interaction energies between each group I and the rest of the protein for the C/C conformation of *Dictyostelium discoideum* myosin II calculated using a Coulomb potential with unscaled charges and a uniform dielectric ($\epsilon = 1$, E^{vac}), with unscaled charges and a distance-dependent dielectric permittivity ($\epsilon = r$, E^{rdiel}), and with scaled charges and a uniform dielectric ($\epsilon = 1$, E^{shield}). Reference values are the Poisson-Boltzmann energies (E^{solv}). All energies are given in kcal/mol, filled symbols indicate surface residues. The two plots only differ in the scale shown.

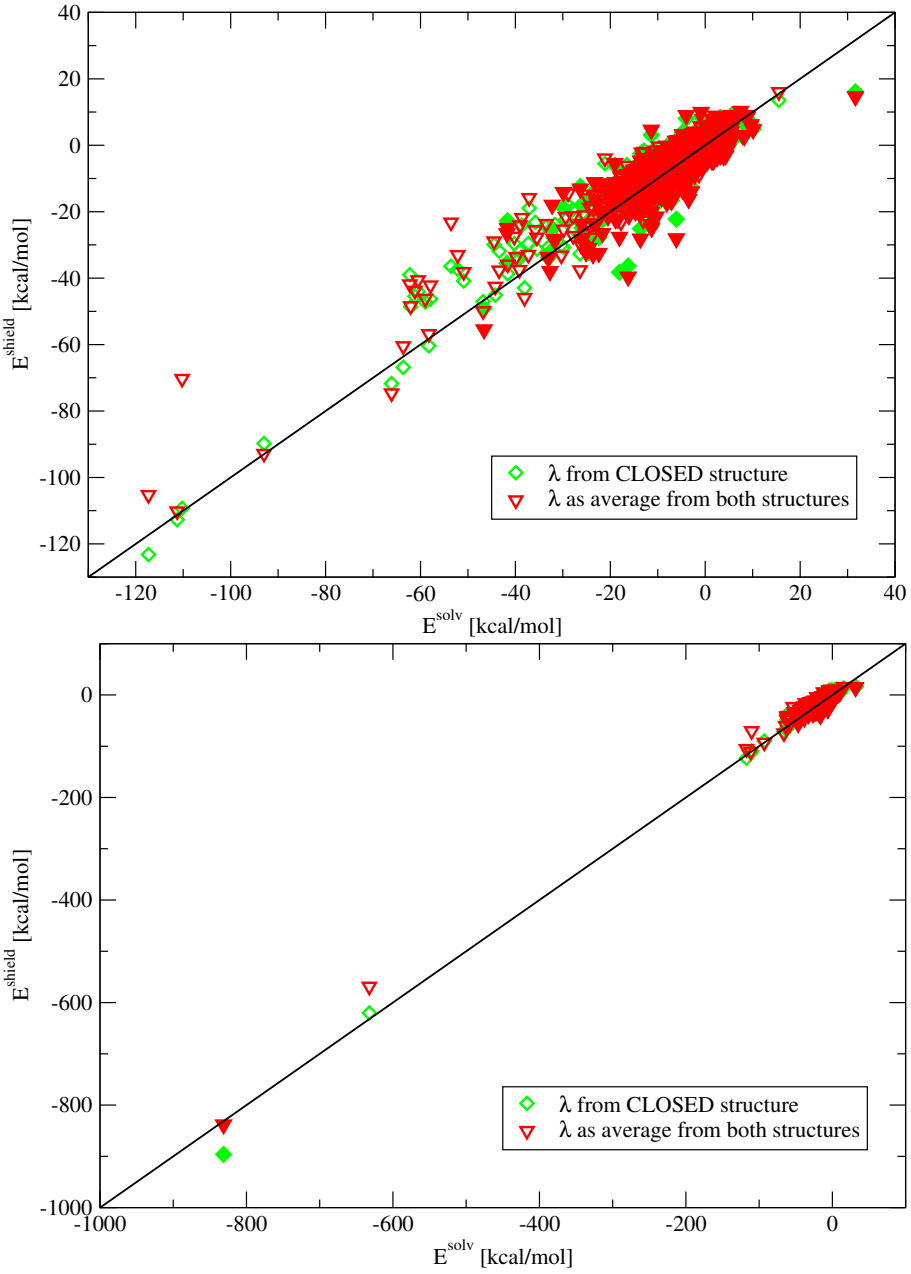

Figure 9.11: Interaction energies between each group and the rest of the protein for the C/C conformation of *Dictyostelium discoideum* myosin II calculated with scaled charges obtained from NUCS factors determined for the C/C conformation and from average factors determined for the C/O and the C/C conformations. Reference values are the Poisson-Boltzmann energies (E^{solv}). All energies are given in kcal/mol, filled symbols indicate surface residues. The two plots only differ in the scale shown.

9.1.4 Discussion

NUCS is a procedure for implicitly including solvent screening effects in protein calculations while maintaining the computational simplicity of a vacuum Coulombic intraprotein model. This is achieved by introduction of non-uniform scaling factors that reduce the effective partial atomic charges of protein atoms relative to the vacuum case. NUCS models solvent screening effects with higher accuracy than assigning a single uniform value or distance-dependent value ($\epsilon = r$) to the dielectric constant. Focusing on screening as a major contribution of electrostatic solvation, NUCS does not consider other solvent effects, such as the electrostatic Born energy contributions, hydrophobic interactions, or dynamical effects such as friction. In cases where the simulation method used and available computation time allow the inclusion of explicit water, this is preferable to a continuum description of the solvent, since an explicit-water representation captures frictional effects as well as solvent screening effects. However, in specific cases such as the refinement of minimum-energy paths, treating solvent explicitly may not be advisable. In these cases, a continuum treatment of the solvent may be adequate, with a Poisson-Boltzmann treatment being the most accurate continuum method available. Still, PB or PB-derived methods such as GB cannot be employed in certain application areas. In such cases, NUCS provides a useful alternative. Some of these areas of application of the NUCS scheme are illustrated by previous studies that employed site-specific charge-scaling schemes similar to NUCS. An example study investigated ligand binding to the immunosuppressent receptor FKBP (365). Other minimization-based computational techniques such as normal mode calculations should also benefit from NUCS. Dynamics-based free energy perturbation studies simulations employing a charge-scaling protocol similar to NUCS have been reported (366; 367). Due to the neglect of the self-energy term in the determination of the NUCS factors, the method presented here is limited to applications that do not require accurate treatment of desolvation energies. Some examples in which NUCS should not be applied are the determination of pKa values or processes involving the desolvation of charged ligands or charged surface residues (such as salt-bridge formation). For such applications, PB or GB methods should be preferred over NUCS. An anonymous reviewer of the NUCS paper (310) has suggested in analogy to NUCS, for which the scaling factors are calculated once and then used for subsequent simulations, GB methods could in principle be approximated and sped up by performing the time-consuming determination of Born radii only once, using PB to obtain "perfect" Born radii (368) from the self-energies of each atom and taking the resulting radii without modification throughout the subsequent simulation. This would require one PB calculation per atom, only slightly more than the one PB calculation per group necessary for NUCS.

The applicability of non-uniform charge scaling to conformational transitions that do not significantly modify the shape of the protein has been demonstrated here for the minimum-energy path calculations on the ring flip of Tyr35 in BPTI. A reference path calculated with continuum electrostatics exhibited an enthalpic barrier in excellent agreement with the

experimental enthalpy barrier of $\Delta H^{\ddagger} = 33$ kcal/mol (363; 364). Previous simulations had given enthalpic barriers for this process as low as of 10 kcal/mol (369). This discrepancy was attributed to the possibility that different reaction paths may be followed due to entropic effects (370). Here, it is shown that a minimum-energy path calculated using Poisson-Boltzmann electrostatics is able to reproduce the experimental enthalpy barrier, suggesting that the earlier discrepancies might rather be due to a neglect of the solvation effects than of the entropic effects. The fact that the path obtained with NUCS is able to preserve the structural character of the PB path shows that the NUCS approach was able to capture part of these solvation effects. The speed advantage gained in using NUCS is remarkable: The refinement of the Poisson-Boltzmann ring flip path presented here required a few weeks CPU time on a single 2.6 GHz processor, as compared to about one hour for a path refinement using the NUCS potential. In addition, example MD simulations of SNase have shown that structures generated from vacuum MD simulations stay closer to the structures generated from explicit water MD simulations when using NUCS than when using a distance-dependent dielectric permittivity or the GB method ACE2. This suggests that the NUCS scheme is useful in the context of standard MD applications.

Large-scale conformational transitions are a limiting case for the application of NUCS because the screening factors might change during the transition along with the change in protein/solvent boundary and thus might have to be recalculated. An example of such an application is annexin V, a system where ligand binding is accompanied by significant conformational changes (371), which was studied using CPR (315). The conformation-dependence of NUCS factors has been investigated by determining NUCS factors for myosin, that undergoes a major conformational change involving the 60° rotation of one subdomain (38). The factors found for the two conformations indeed differed. However, using factors averaged between the two conformations did not lead to a significant increase in the RMSD between E_I^{NUCS} and E_I^{solv}. Thus, after individual validation, NUCS could in principle be applicable for studying conformational transitions as large as observed in the myosin case.

Another domain of application is when using specialized code currently not interfaced with the more elaborate implicit solvent models, as is the case in combined quantum mechanical/molecular mechanical (QM/MM) calculations. In this case, non-uniform charge scaling as presented here would be a useful alternative. A charge scaling procedure for QM/MM calculations has been reported recently (309). However, the scaling factors are calculated differently than in the present NUCS procedure. The approach presented in Ref. (309), which is based on earlier work (346), introduces scaling factors only for exposed ionizable groups, not for all groups as in the present work. Ref. (309) scales partial charges so as to accurately reproduce the electrostatic potential at a specified region of interest, such as an enzyme active site. A scaling factor is applied that screens the electrostatic interactions between the charges of a group I outside the region of interest and the charges on the atoms within the region of interest. Only the charges on group I are scaled, the charges on the region of interest remaining unscaled. Thus, the full screening effect is carried by one partner in this interaction, namely the group I. This approach had been used earlier

to determine the catalytic mechanism of a prolyl-isomerase (317). NUCS, in contrast, aims at describing the electrostatic potential equally well throughout the whole protein. As described in Section 9.1.2, this is achieved by determining group-specific screening constants designed to optimally reproduce the screening between the group I and the rest of the protein. The screening between any two given groups, I and J, is contained in both group screening constants ϵ_I and ϵ_J and is distributed over both groups via application of the square root in Eq. 9.4. Thus, the resulting NUCS factors will in general be smaller in magnitude than the scaling factors of Ref. (309).

In QM/MM simulations, a widely-used computational setup involves a shell model, in which the QM region is surrounded by a flexible MM region, keeping the rest of the protein fixed (269; 274; 372; 373). With this type of setup it is important to accurately represent the electrostatic effect of the protein environment on the QM region while at the same time maintaining the integrity of the electrostatic interactions within the flexible molecular mechanical region. Use of the scheme of Ref. (309) should represent the potential at the QM region more accurately than NUCS can. However, two charged groups that are within the flexible MM region interact not only with the QM region but also with each other. Scaling them according to the scheme in Ref. (309) may lead to problems since the screening between the two groups will be overestimated. The NUCS scheme avoids this problem by taking all interactions into consideration when determining a group scaling factor, including interactions between groups within the flexible region. A useful approach for QM/MM calculations may be to combine the present method with that of Ref. (309). Groups in the flexible region would be scaled according to the NUCS scheme, whereas charged groups in the fixed region that influence only the QM region *via* long-range electrostatic interactions without changing their positions, could be scaled according to the scheme from Ref. (309). This approach has already proven useful in determining the translocation mechanism of a poly-sugar chain across a trans-membrane pore (320), where interaction within the long sugar-chain were treated as in NUCS while the interactions between sugar and protein were treated with a method similar to Ref. (309). It is investigated in detail in Section 9.3

Due to the definition of the NUCS factors (Eqs. 9.4 and 9.9) their squares, $\lambda_I^2 = \epsilon_I$, serve as effective dielectric screening constants that represent the degree to which the surrounding solvent screens the electrostatic interactions in a given protein part. However, λ_I^2 should not be confused with the protein dielectric constant that arises from internal charge fluctuations. Estimates of the latter based on molecular dynamics simulations have provided values ranging from 1-2 in the protein core to up to 40 in the protein surface regions (374; 375; 376; 377). Interestingly, this range is similar to that spanned by the λ_I^2, which range from ~ 1 to ~ 50 (compare Fig. 9.2). This suggests that solvent electrostatic screening and internal protein charge fluctuations might be of approximately equal magnitude in determining effective electrostatic interactions between atoms in proteins.

9.2 Targeted NUCS

Enzymatic processes in proteins are local events. When studying such processes it is therefore necessary to accurately describe the electrostatic potential *at the active site*, whereas the electrostatic potential in regions far from the active site is of minor importance. Thus, the protein can be split into two regions, namely the region of interest (referred to as *target region* hereafter) and the rest of the protein.

The electrostatic potential in the target region shall be optimally described using a simple Coulomb potential. This is achieved by scaling the partial atomic charges on all atoms *outside* the target region such that the Coulomb interaction energy determined with scaled charges optimally reproduces the interaction energy in solution.

Charges are scaled in a non-uniform fashion using scaling factors that are determined for groups outside the target region. The same group definitions as for the global NUCS procedure are used. Scaling factors are determined according to

$$\lambda_I = \frac{|E^{PB}_{I,target}|}{|E^{vac}_{I,target}|}, \tag{9.17}$$

where $E^{PB}_{I,target}$ is the Poisson-Boltzmann interaction energy between the charges of the group I and all charges within the target region, and $E^{vac}_{I,target}$ is the corresponding Coulomb interaction energy calculated using unscaled charges on the atoms of group I and scaled charges on the atoms within the target region.

This definition of scaling factors is comparable to the definition of the group:group screening factors ϵ_{IJ} in the global charge scaling routines. However, here, the whole screening is assigned to the group I.

Eq. 9.17 uses absolute values for the interaction energies in order to compensate for the change in sign which frequently occurs if the interaction energies in solution and/or in vacuum are small. It also occurs that $|E^{vac}_{I,target}|$ is smaller than $|E^{PB}_{I,target}|$, leading to scaling factors larger than 1. Thus, charges would be increased instead of decreased, which is undesired. Since this usually occurs only when the interaction energies are small, the scaling factors are set to 1.0 in such cases.

9.3 Mixed NUCS

In many QM/MM calculations, the QM region is surrounded by a flexible MM region, which in turn is surrounded by a region in which protein atoms are kept fixed. In such simulations it is necessary to adequately represent the electrostatic potential at the QM and

flexible regions, whereas it is not necessary to correctly represent the electrostatic potential at the fixed region. Thus, the following points should be satisfied:

1. The interaction energy between the QM region and any given protein group calculated with scaled charges should be close to the Poisson-Boltzmann interaction energy. This ensures that the electrostatic potential at the positions of the QM atoms is adequately represented and thus the electrostatic QM/MM interactions are appropriately captured.

2. The interaction energies between two group within the flexible region calculated with scaled charges should be close to the Poisson-Boltzmann interaction energies. This ensures that the electrostatic interaction between any two mobile MM atoms is captured.

3. The interaction energy between a group in the flexible region and *all* other groups calculated with scaled charges should be close to the Poisson-Boltzmann interaction energy. This ensures that the overall electrostatic potential at the position of a mobile MM atom is adequately captured.

This is achieved by deriving scaling factors in a three-step procedure:

1. Determination of global scaling factors for groups in the target region (*i.e.*, the flexible and QM regions).

2. Determination of targeted scaling factors for groups in the fixed region. In the calculation of $E^{vac}_{I,target}$ according to Eq. 9.17 the partial atomic charges on the atoms within the target region are scaled using the previously-determined global NUCS factors.

3. Validation of the scaling factors according to the three aspects given above.

Because this procedure involves a "mixing" of the global and targeted NUCS schemes, it is referred to as *mixed NUCS* hereafter.

Both global NUCS and mixed NUCS have been implemented in a set of CHARMM input scripts. The procedures for determining NUCS factors with CHARMM are described in Appendices D and E.

Chapter 10

System setup and computational details

In this chaper, the choice and preparation of the crystal structure used for the calculations on myosin-catalyzed ATP hydrolysis is detailed in Sections 10.1 and 10.2. A description of the potential energy function used in given in Section 10.3. Computational details on the optimizations of the end states and the paths are described in Sections 10.4 and 10.5. The assumption made are summarized in Section 10.6. Finally, the path analysis methods used are given in Section 10.7.

10.1 Choice of the structure

Only the C/C structure of myosin (see Section 3.2.3) is competent for hydrolysis. Three crystal structures in the closed state are available for *Dictyostelium discoideum*, namely a Mg.ADP.VO_4 (PDB code 1VOM (77)), a Mg.ADP.AlF_4 (PDB code 1MND (57)) and a Mg.ADP.BeF_3 (ref1 (55)) structure. The resolution of the AlF_4 structure is only 2.6 Å and thus this structure is not further considered. Both the vanadate structure and the beryllium fluoride structure, however, seem equally well suited as a starting point for the calculation at first glance. The pentavalent VO_4 moiety may serve as a transition state analog corresponding to an associative hydrolysis mechanism (see Section 4.2). The tetravalent BeF_3, however, may serve as a template for the ATP-bound reactant state. Both structures are in principle equally suitable for the purpose of mechanistic studies (compare also Appendix A).

To decide which structure should be taken as a starting point for the calculations, the structures of the nucleotide binding pocket are compared in detail. As shown in Fig. 10.1, the structures of the P-loop, switch-1, switch-2, and nucleotide are very similar, including

the Mg^{2+} coordination sphere and orientation of the sidechains. Differences are found in the sidechains of Glu180 in the P-loop; a slight shift by a maximum of 0.5 Å of the backbone structure of the switch-1 loop; and a small shift of approximately 0.5 Å in C_α of Gly457 and O_γ of Ser456 of the switch-2 loop.

The V-O (1.64 to 1.67 Å for the equatorial bonds and 2.08 and 2.28 Å for the axial bonds) bonds are naturally longer than the Be-F (1.76 to 1.83 Å) bonds. The bond lengths between α- or β-phosphorus to the terminal oxygens vary between 1.44 and 1.51 Å, whereas the bond lengths of the P-O bridging bonds are in the range of 1.58 to 1.63 Å. Thus, the Be-F bond lengths are closer to the P-O bond lengths than the V-O lengths. Therefore, the BeF_3 structure (ref1) seems more suitable as a starting structure for computation and is used in the present work.

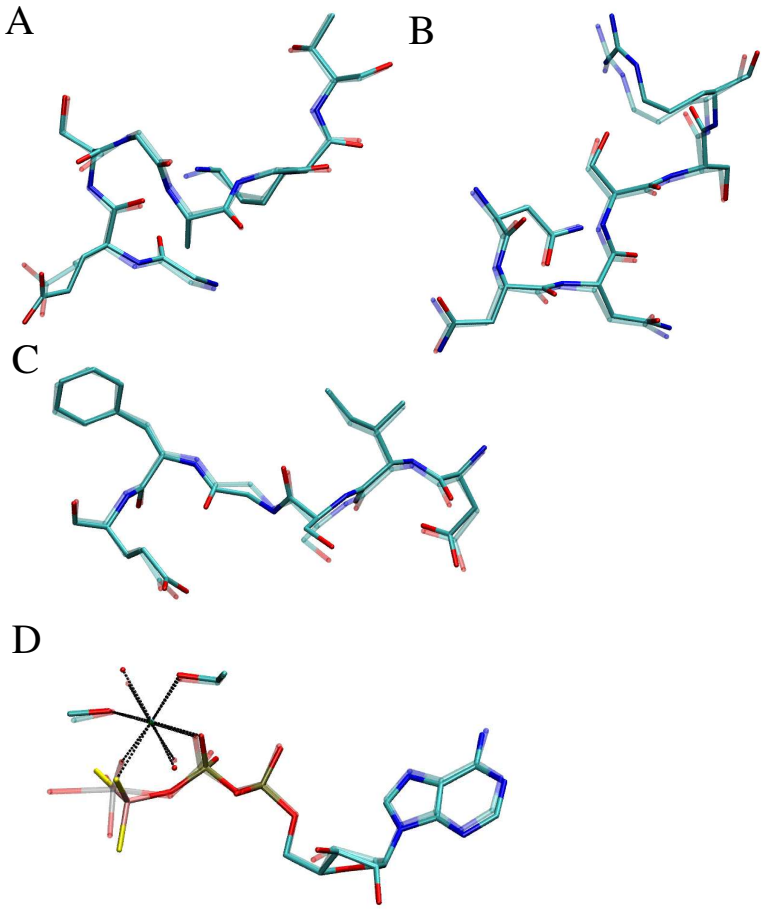

Figure 10.1: Overlap of the reactant-like (opaque colors, ref1 (55)) and transition-state-like (transparent colors, prepared from PDB code 1VOM (77)) structures of *Dictyostelium discoideum* myosin II. Panel A shows the P-loop, panel B the switch-1 loop, panel C the switch-2 loop, and panel D the nucleotide.

10.2 Modeling the myosin:ATP complex

To prepare the myosin.Mg.ADP.BeF_x crystal structure (55) for the simulations the following steps were made:

1. **Atom replacement.** The beryllium atom was replaced by a phosphorus atom (*i.e.*, the P_γ). The fluorine atoms were replaced by oxygen atoms.

2. **Pronotation state of ATP and titratable groups.** At neutral pH the predominant species of Mg.ATP is fully deprotonated (131). Thus, ATP was assumed to have a total charge of -4e in the myosin active site. Standard protonation states were assumed for all titratable groups of the protein.

3. **Crystal water.** Crystal water molecules at the surface of the protein were removed whereas internal water molecules were kept.

4. **Hydrogens.** No hydrogens are resolved in the crystal structure. Hydrogens were added to the polar groups with the HBUILD routine of CHARMM.

5. **Water structure.** The ref1 structure does not show any resolved water molecules in the cavity formed by the BeF_3 moiety and the switch-1 and switch-2 loops that could serve as the attacking water molecule. A cavity analysis revealed a large cavity in the attacking position region (see Fig. 10.2). It is likely that this cavity is filled with flexible water molecules that due to their flexibility do not have well-defined electron densities and whose positions can therefore not be determined by X-ray crystallography.[1] A water molecules was placed inside the cavity (see Fig. 10.3) in a position suitable for inline attack onto the P_γ. This water is termed *attacking water* hereafter.

The final model contained 7986 atoms, out of which 7849 atoms are protein atoms, 90 are water atoms (corresponding to a total of 30 explicit water molecules), and 47 Mg.ATP atoms.

[1] Apart from the two Mg^{2+}-coordinating crystallographic water molecules, there is only one (water 679) water molecule within 5 Å of the vanadate atom in the 1VOM structure. This water molecule has a temperature factor of more than 50 $Å^2$, indicating that its position could be determined only approximately.

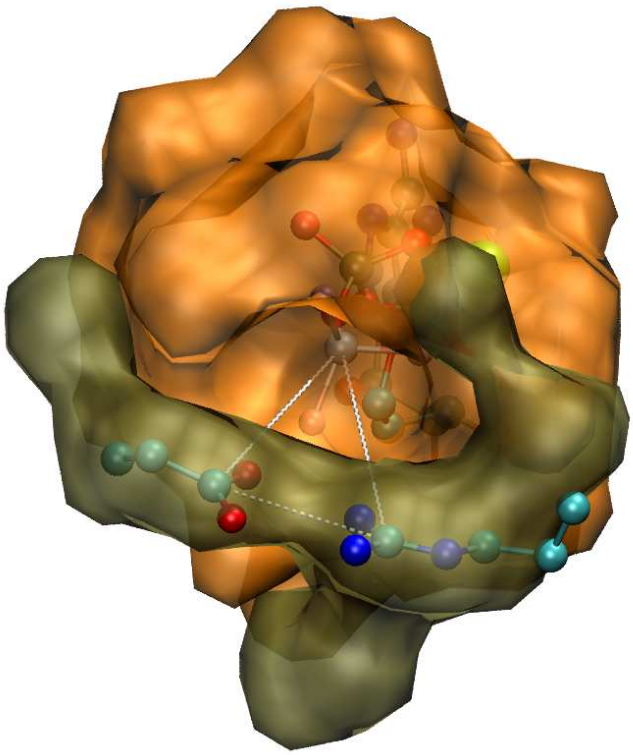

Figure 10.2: Cavity between the beryllium fluoride moiety and the salt bridge between switch-1 and switch-2 in the ref1 structure. Mg.ADP.BeF_3 and the sidechains of Arg238 and Glu459 are shown in CPK representation, the surface enveloping all atoms within 6 Å of the beryllium atom in transparent orange, and the surface enveloping all atoms with a distance between 6 and 8 Å from the beryllium atom and within 6 Å of C_ζ of Arg238 or C_δ of Glu459 in transparent tan. The plane defined by the beryllium atom, C_ζ of Arg238, and C_δ of Glu459 is indicated with dotted white lines.

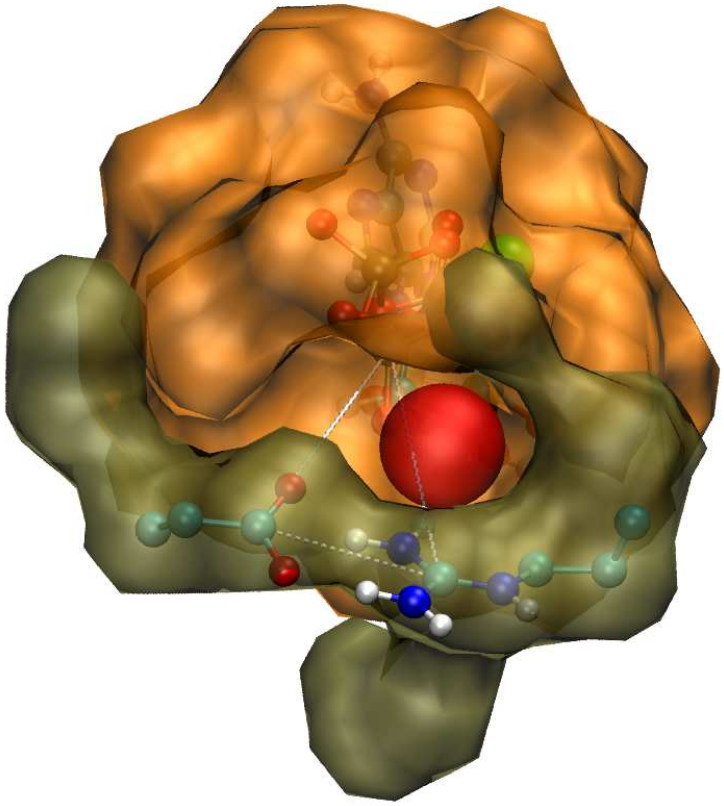

Figure 10.3: The cavity between the beryllium fluoride moiety and the salt bridge between switch-1 and switch-2 is filled with a water molecules (*i.e.*, the attacking water, shown as van der Waals sphere) in the structure as prepared for the simulations. Mg.ATP and the sidechains of Arg238 and Glu459 are shown in CPK representation, the surface enveloping all atoms within 6 Å of the P_γ atom in transparent orange, and the surface enveloping all atoms with a distance between 6 and 8 Å from the P_γ atom and within 6 Å of C_ζ of Arg238 or C_δ of Glu459 (except the water molecule) in transparent tan. The plane defined by the beryllium atom, C_ζ of Arg238, and C_δ of Glu459 is indicated with dotted white lines.

10.2.1 Partitioning of the system

The protein is partitioned into QM and MM regions using a four-layer "onion" model (see Fig. 10.4). The inner quantum mechanical region (termed QM) is surrounded by a first molecular mechanical layer in which all atoms are fully flexible (termed flexible) and a second molecular mechanical region in which the atoms are positionally constrained to their positions in the crystal structure (termed constrained). The remaining atoms are kept fixed (termed fixed).

The quantum mechanical part consists of the triphosphate moiety of ATP, the Mg^{2+} ion, the two Mg-coordinating crystal water molecules, the Mg-coordinating sidechains of Thr186 and Ser237, the attacking water molecule, the helper water molecule, and the sidechains of Ser181 and Ser236. The valency at the boundary atoms was saturated by adding a total of 15 dummy hydrogen atoms, 5 of which are hydrogen link atoms. The total number of QM atoms is 57, with 28 non-hydrogen atoms. The QM and MM regions were linked together using the default linking procedure as implemented in the CHARMM/GAMESS-US interface (279; 378).[2]

The flexible MM region consists of all residues of the P-loop, the switch-1 loop, and the switch-2 loop augmented by all atoms within 12 Å of P_γ that are not included in the QM region (475 atoms). The constrained MM region consists of all atoms with a distance between 12 and 20 Å from P_γ (1441 atoms). The remaining 6028 atoms are kept fixed.

10.3 Energy function

10.3.1 Description of the energy surface used

The energy function used for all calculations is the sum of the quantum mechanical and the molecular mechanical energy terms,

$$E_{tot} = E_{QM} + E_{MM}. \tag{10.1}$$

E_{QM} contains both the quantum mechanical energy of the QM region and the electrostatic interaction energy between the QM and MM regions. Geometry optimization were done using Hartree-Fock calculations (261) for the QM atoms with a 3-21G(d) basis set (379;

[2]Alternatively, it was tried to use the user-controllable linking procedure as described and tested in Section 7.2. However, although this procedure had been found to be stable for the molecules used for testing, stable geometry optimizations could not be achieved for the myosin QM/MM system. Since no obvious reasons for the observed structural instabilities (that sometimes even lead to disruptions of the chemical structure of the triphosphate moiety) could be found in the user-controllable linking procedure it was replaced by the default linking procedure that was found to be stable.

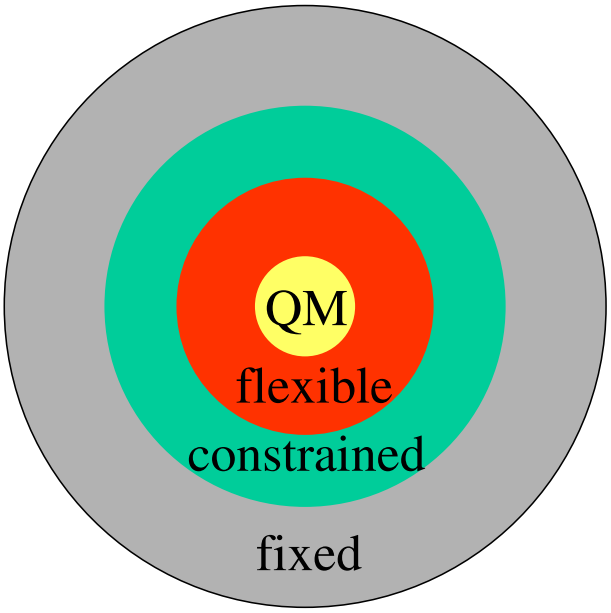

Figure 10.4: Onion model for QM/MM calculations. The inner quantum mechanical region (**QM**) is surrounded by a first molecular mechanical layer in which all atoms are fully flexible (**flexible**) and a second molecular mechanical region in which the atoms are positionally constrained to their positions in the crystal structure (**constrained**). The remaining atoms are kept fixed (**fixed**).

380; 381) and subsequently a 6-31G(d,p) (382; 383) basis set. Point-energy calls were made on the structures optimized with both basis sets using density-functional theory with the B3LYP functional (265; 266; 267) and 6-31+G(d,p), a basis set that includes diffuse functions (384).

The choice of Hartree-Fock calculations as the QM method implies large energetic inaccuracies. The lack of electron correlation effects leads to an error on the order of 10 kcal/mol in the computation of barrier heights, even if large basis sets are used (385). The energy variation between quantum chemical methods of as much as ∼15 kcal/mol were found using various quantum methods and basis sets for the reaction energy of a water exchange reaction of the much smaller $Sc(OH_2)_6^{3+}$ system (386). Use of density functional theory improves the accuracies, however, common B3LYP calculations show an average (maximum) error of ∼3 (∼20) kcal/mol (387) in the calculation of heats of formation of a set of 148 molecules. Improvement of accuracy would be expected if, during geometry optimization, diffuse functions were included in the basis set (388; 389) and a quantum method including electron correlation effects were employed (390). This, however, exceeds currently available computational resources when combined with the refinement of complete reaction paths in a protein, which requires significantly more energy evaluations than a simple geometry

optimization. This is reflected in the CPU time required by the calculations presented in this thesis: on eight 2.6 GHz processors, the QM/MM geometry optimization of a single protein structure at the HF/6-31G(d,p) level requires ∼2 days, while the refinement of a path (composed of 10 to 30 structures) requires 2 to 4 months. Nevertheless, the structural information obtained is expected to be reasonably accurate. Moreover, qualitative structural results common to the levels of theory used are likely to represent stable and reliable findings. The analyses in the present work thus focus mostly on these method-invariant aspects.

E_{MM} is the sum of bonded, nonbonded, and constraint energy terms arising from the flexible and constrained MM regions. It also contains the van der Waals interaction energy between the QM and MM regions. The CHARMM (289) polar hydrogen force field (param 19/22) (359) is used. Nonbonded interactions are truncated with a cubic switching function in combination with a uniform dielectric ($\epsilon = 1$) in a switch region between 99 and 100 Å. This large cutoff distance equals using an infinite cutoff, because due to the definition of the regions any two moving atoms are at a maximum distance of 40 Å. To approximate solvent screening effects, the charges on all MM atoms are scaled as described in Section 9.3. The target region is defined as the QM and flexible MM regions, augmented by atoms from the constrained region so as to set the boundary between target and non-target regions at the boundaries of the groups for which NUCS factors are calculated, resulting in a total of 628 atoms. The non-target region contains the remaining parts of the proteins.

Atoms in the constrained regions are harmonically constrained to their positions $\vec{r}_0$ in the reference structure,

$$E_{cons} = k_{cons}(\vec{r} - \vec{r}_0)^2. \tag{10.2}$$

The constraint force constants, k_{cons}, are determined from the crystallographic temperature factors, B, according to

$$k = \frac{4\pi^2 k_B T}{B}. \tag{10.3}$$

The notation QM[B3LYP/6-31+G(d,p)//HF/6-31G(d,p)]/MM/NUCS used hereafter refers to an energy surface that combines QM calculations with MM calculations while solvent screening effects are approximated by the NUCS procedure. The quantum methods for geometry optimizations and subsequent single point energy evaluations are included in square brackets.

10.3.2 Some remarks on the nature of the energy surface

A large number of approximations have been made in the definition of the energy surface. Thus, the question arises whether the resulting surface corresponds to a microscopic potential energy surface or, alternatively, approximates a free energy surface.

The free energy ΔG of a solvated protein:substrate complex in a given state as defined by a subspace S in conformational space can be written as

$$\Delta G = \sum_i p_i \cdot \Delta G_i, \tag{10.4}$$

where the sum extends over all conformations $i \in S$, p_i is the probability of finding the system in the conformation i ($\sum_i p_i = 1$), and ΔG_i is the free energy of conformation i. The probability p_i can be determined as the Boltzmann factor,

$$p_i = \frac{e^{-\frac{\Delta G_i}{RT}}}{\sum_j e^{-\frac{\Delta G_j}{RT}}}. \tag{10.5}$$

Combining a quantum mechanical description of the substrate with a molecular mechanical description of the protein and a continuum description of the ionic solution that solvates the protein:substrate system, ΔG_i can be expanded as (see Ref. (391))

$$\begin{aligned} \Delta G \;=\; & E_{bonded}(MM) + E_{coul}(MM) + E_{vdW}(MM) + E_{vdW}(QM:MM) \\ & + E_{QM}(QM) + G_{solv}(QM/MM) + G_{ideal}(QM/MM), \end{aligned} \tag{10.6}$$

where $E_{bonded}(MM)$ is the MM energy due to bonded interactions in the MM region MM, $E_{coul}(MM)$ and $E_{vdW}(MM)$ are the electrostatic Coulomb and the van der Waals energy terms within MM, $E_{vdW}(QM:MM)$ is the van der Waals interaction term between MM and the QM part QM, $E_{QM}(QM)$ is the QM energy of the QM part(including the electrostatic interaction energy between the QM region and the MM regions), $G_{solv}(QM/MM)$ is the solvation free energy of the entire system, and $G_{ideal}(QM/MM)$ enthalpic and entropic contributions from translational, rotational, and vibrational degrees of freedom of the entire system, as determined from an ideal-gas description that is assumed to be equally appropriate for infinitely dilute solvated systems.

The first four terms in Eq. 10.6 are contained in E_{MM} (Eq. 10.1) in the present work. In addition, E_{MM} approximately accounts for the electrostatic contribution to the solvation

free energy, G_{solv}, because scaled charges are used. The nonpolar contribution to G_{solv}, G_{solv}^{np} can be assumed to be proportional to the solvent accessible surface A, $G_{solv}^{np} = \gamma \cdot A$, where γ is the surface tension coefficient. In the present work, G_{solv}^{np} can be neglected because the protein surface does not change during the reaction and the resulting constant G_{solv}^{np} consequently does not affect the energy difference between the reactant state and any state along the reaction path. $E_{QM}(L_i)$ in Eq. 10.6 is identical to E_{QM} in Eq. 10.1.

The translational and rotational free energy contributions do not change along the reaction coordinate either, because they depend on the total mass and the principle moments of inertia of the QM:MM system, neither of which changes significantly during the reaction. The vibrational contribution, however, may change along the reaction coordinate. It is dependent on the vibrational frequencies as determined from normal mode analyses at stationary points on the energy surface. Necessarily, the modes change throughout the course of the reaction. However, it is currently not possible to perform normal mode analyses on *ab initio* QM/MM energy surfaces since the implementation of required routines into the program packages is lacking. Thus, vibrational contributions must be neglected.

Apart from the neglection of the vibrational free energy contributions, the energy function used in the present thesis (Eq. 10.1) very closely matches the free energy function given in Eq. 10.6. It can therefore be viewed as an approximate free energy function. Accordingly, the rate law is used in its phenomenological form given in Eq. 8.17 rather than its microscopic form throughout this thesis.

However, the neglect of vibrational contributions and the only very approximate treatment of solvation effects suggests that equating the energy surface used with a free energy surface is daring. In addition, the very severe approximation of working with only a single conformation for each state of interest is made, thus circumventing the necessity of averaging (Eq. 10.4). To indicate this, the symbol E is used to denote energies hereafter, rather than using the symbol G.

10.4 Preparation of the reactant and product structures

10.4.1 Reactant structure

The structure prepared as described in Section 10.2 was subjected to QM[HF/3-21G(d)]/MM/NUCS energy minimization to a final root mean square gradient of 0.01 kcal/mol/Å. A combination of conjugate gradient and adopted basis newton raphson minimizations was used while slowly releasing initial positional constraints (force constant 10 kcal/mol/Å^2) on the QM atoms. The resulting structure was further minimized on the QM[HF/6-31G(d,p)]/MM/NUCS surface. The geometry-optimized structure serves as the reactant

structure (R) for all subsequent simulations.

10.4.2 Product structures

To generate a product structure corresponding to the direct pathway the energy surface was explored by adiabatic mapping along the predefined reaction coordinate λ_{ini}

$$\lambda_{ini} = d(P_\gamma : O_{\beta\gamma}) - d(P_\gamma : O_W), \tag{10.7}$$

where $d(P_\gamma : O_{\beta\gamma})$ is the distance between the γ phosphorus and the $\beta\gamma$-bridge oxygen (*i.e.*, the bond broken) and $d(P_\gamma : O_W)$ is the distance between the γ phosphorus and the oxygen of the attacking water (*i.e.*, the bond formed).

By restraining λ_{ini} to different values, the system was forced from the reactant geometry into a product-like geometry. Due to the definition of λ_{ini} this restraint operated only on the bond distances of the P-O bonds broken and formed, neglecting the proton transfer from the attacking water onto the γ-phosphate moiety. This transfer did not occur spontaneously, resulting in a product-like structure in which the ADP and P_i were clearly separated from each other but in which one oxygen of P_i carried two protons. By manually transferring one of the two protons onto the neighboring oxygen, a product structure was created that was further energy-minimized to a final RMS gradient of 0.01 kcal/mol/Å without restraints. The resulting geometry-optimized structure served as the product structure for the direct pathway (P_{direct}).

The product structures for the Ser181 and Ser236 pathways, P_{S181} and P_{S236}, were generated from P_{direct} by manually shifting the protons so as to make their positions compatible with the proton transfer event and subsequent QM[HF/3-21G(d)]/MM/NUCS geometry optimization to a RMS gradient of 0.01 kcal/mol/Å.

The QM[HF/3-21G(d)]/MM/NUCS-optimized product structures were further minimized on the QM[HF/6-31(d,p)]/MM/NUCS energy surface to a final root-mean-square gradient of 0.01 kcal/mol/Å.

10.5 Path refinement

Reaction path calculations were performed using the CPR algorithm as implemented in TReK module of CHARMM. Initial seed paths corresponding to the direct path, the Ser181 path, and the Ser236 path were prepared manually. The seed paths were refined until all energy peaks along the path were minima (RMS gradient less than 0.05 kcal/mol/Å) along

the first 44 directions[3] of the conjugate subspace.

Initial paths consisted of the reactant and the product structures augmented by manually-generated transient structures that have been included when necessary to prevent steric clashes due to the Cartesian interpolation between adjacent path points as done by CPR. The initial paths were first refined on the QM[HF/3-21G(d)]/MM/NUCS surface. The transition states and two adjacent path points in either direction along the reaction coordinate were then used to seed further path refinements on the QM[HF/6-31G(d,p)]/MM/NUCS surface.

10.6 Summary: assumptions

A number of assumptions have been introduced in the above sections. They can be summarized as follows:

1. ATP is fully deprotonated and complexed to Mg^{2+} (*i.e.*, Mg.ATP^{2-}) when bound to myosin. This is in agreement with that the predominant species in solution under the experimental conditions is Mg.ATP^{2-} (compare Section 4.5). The pK_a values of the titratable groups of ATP may change on binding. The resulting microscopic pK_a values can in principle be calculated (119; 392; 393). However, such calculations are beyond the scope of the present work.

2. The ref1 structure of Mg.BeF_3.ADP complexed to myosin II provides a suitable structural model for myosin in its catalytically active form.

3. A single water molecules fills the cavity formed by the γ-phosphate moiety and the rear end of the phosphate tube.

4. The chemical cleavage step takes place without major conformational rearrangements of the protein environment. Allowing a 12 Å sphere around P_γ to be flexible (embedded in a 12 to 20 Å buffer zone) provides sufficient flexibility of the protein to accommodate local conformational changes.

5. Solvent effects can be taken into account by non-uniformly scaling the partial atomic charges on the MM atoms.

6. QM and MM regions can be effectively joined by introduction of a link atom.

[3] A heuristic rule states that if a peak is a minimum in the first $\sqrt{N}$ directions of the conjugate subspace (N number of moving atoms) it is likely to be a real saddle-point. The number of moving atoms in the present work is 1958, and thus $\sqrt{1958} \approx 44$ line-minimizations must be made.

7. The accuracy of HF/3-21G(d) and HF/6-31G(d,p) calculations is sufficient to refine the structures along the ATP hydrolysis reaction. B3LYP/6-31+G(d,p) point energy calls on these structures provide a reliable energy profile along the reaction coordinate.

8. Neglection of vibrational energy contributions does not introduce major artifacts.

9. Single stationary points on the multidimensional energy landscape sufficiently describe the reactant, product, transition, and intermediate states that describe the myosin-catalyzed ATP hydrolysis reaction.

10.7 Path analysis

The refined paths are analyzed with the following analysis methods:

1. **Geometry.** Bond distances and angles reveal possible distortions of ATP in the reactant state and the degree of associativity as judged from the transition or intermediate state geometry.

2. **Energy decomposition.** The total energy ΔE_{tot} is decomposed into its contributing terms,

$$\Delta E_{tot} = \Delta E_{QM} + \Delta E_{bonded} + \Delta E_{vdW} + \Delta E_{elec} + \Delta E_{cons}, \tag{10.8}$$

   where ΔE_{QM} denotes the QM energy including the QM/MM electrostatic interactions, ΔE_{bonded} the bonded energy terms, *i.e.*, the sum of bond, angle, dihedral, Urey-bond, and improper terms of the force field, ΔE_{vdW} the van der Waals interactions within the MM region and between MM and QM regions, ΔE_{elec} the electrostatic interactions within the MM region, and ΔE_{cons} the energy terms arising from the positional constraints in the bfactor region. This decomposition reveals the origin of energy peaks along the reaction path.

3. **Decomposition of the QM energy.** The quantum mechanical contribution to the total energy, ΔE_{QM} comprises both the energy of the quantum region, ΔE_{QM}^{QM}, and the electrostatic interaction energy between the quantum atoms and the remaining proteins atom, $\Delta E_{elec}^{QM/MM}$, which also includes polarization effects. ΔE_{QM}^{QM} is given by the quantum mechanical energy contribution to the total energy from QM/MM single-point energy evaluations along the refined path after setting all partial atomic charges on MM atoms to zero. $\Delta E_{elec}^{QM/MM}$ is then calculated as the difference

$$\Delta E_{QM/MM}^{elec} = \Delta E_{QM} - \Delta E_{QM}^{QM}. \tag{10.9}$$

4. **Perturbation analysis.** The electrostatic contribution from individual residues of the switch-1, the switch-2, and the P-loops was analyzed by recomputing the energy difference between stationary points of the paths, ΔE_{tot}, after perturbing the charges on a given residue X by setting them to zero, $\Delta E_{tot}^{q(X)=0}$. The difference

$$\Delta E_{elec}^{X} = \Delta E_{tot} - \Delta E_{tot}^{q(X)=0} \tag{10.10}$$

gives the electrostatic contribution to the barrier height or to the reaction energy due to the charges on residue X (394).

5. **Charge Distribution.** Mulliken charge analyses (395) are performed on the stationary points of the paths. This reveals how charge shifts occur along the reaction.

Chapter 11

Summary: Methods

This part summarized all methods that are used in the present thesis, including the newly-developed method of Non-Uniform Charge Scaling (NUCS). First, an outline of the mathematical formulations of molecular mechanical (MM) force fields, the quantum mechanical (QM) Hartree-Fock and density functional theories, and the basic formulation of combining MM and QM calculations were given. Moreover, the implementation and testing of a specific method to link MM and QM regions across a chemical bond, *i.e.* the link-atom method, was described. A general criterion of usefulnes to evaluate linking schemes was formulated. A linking scheme is considered useful if the error of some chemical property of interest determined using the QM/MM approach rather than modelling the whole system quantum mechanically is smaller than the error made by treating a small model system only instead of treating the whole system quantum mechanically. According to this criterion, the link-atom scheme implemented here was found to be useful.

Subsequently, the concept of minimum-energy paths (MEP) was introduced. A MEP connects a minimum on the potential energy landscape corresponding to the reactant state to another minimum corresponding to the product state such that each path point is a minimum in the subspace orthogonal to the reaction coordinate. Maxima along a continuous MEP correspond to first-order saddle points on the energy landscape and thus to transition states of the reaction. The progress of a reaction can be measured by a normalized curvilinear reaction coordinate that changes from 0 (reactant) to 1 (product). Transition state theory can then be used to derive reaction rates from the energy barrier at the saddle point with highest energy along the MEP.

The development and testing of the new method Non-Uniform Charge Scaling (NUCS) was described in a self-contained chapter. NUCS aims to represent the electrostatic potential in solution as determined from an initial Poisson-Boltzmann analysis with a simple Coulomb potential. This is achieved by the introduction of non-uniform scaling factors that reduce the partial atomic charges on the protein atoms. To determine the scaling factors, the

electrostatic interaction energies between all possible pairs of groups of the protein are calculated both using the Poisson-Boltzmann approach and a Coulomb potential. The ratio between the two energy values gives a pairwise screening constant. These pairwise screening constants are averaged to arrive at group-specific scaling factors. NUCS is particularly useful for simulations that rely on minimization such as the determination of MEPs, ligand binding, or normal mode analyses for which an explicit treatment of water molecules is not possible. Moreover, it can be used in cases where interfaces to implicit solvent models are lacking, as is the case in current QM/MM calculations. Because the scaling factors depend on the overall conformation of the protein, large-scale conformational changes would be a limiting case for the application of NUCS. The method was illustrated using reaction path calculations of the Tyr35 ring flip in the bovine pancreatic trypsin inhibitor. The NUCS-derived path captured the qualitative features of a reference Poisson-Boltzmann path unlike paths that were determined using unscaled charges, even though the energy barrier was underestimated by roughly 15 kcal/mol.

In an extension of the NUCS method, it is possible to consider only pairwise interactions for a subregion of the protein and to scale the partial atomic charges outside this subregion such that their effect on the potential in the subregion is optimally captured. This "mixed NUCS" scheme reproduces the solution potential in the subregion more accurately than outside the subregion and is therefore particularly useful to study events in an active site of a protein. It has been used for the determination of reaction paths for ATP hydrolysis in myosin in the present thesis.

Details of the setup of the reaction path simulations on ATP hydrolysis in myosin were described. A structure of *Dictystelium discoideum* myosin II complexed to the Mg.ATP analog Mg.Be.F_x.ADP is used for the calculations. In this structure, the lytic water molecule is not resolved. In was therefore placed into the corresponding cavity. The QM region comprises the triphosphate moiety of ATP, Mg^{2+} and its coordination sphere, the sidechains of Ser181 and Ser236, the attacking water molecule and an additional water molecule. All protein atoms within 12 Å of P_γ are allowed to move, whereas atoms between 12 and 20 Å from P_γ are positionally constrained to their crystal structure position. All other atoms are kept fixed. The crystal structure was then minimized using HF/3-21G(d) and HF/6-31G(d,p) for the QM atoms, the CHARMM united atom force field for the MM atoms, and NUCS to treat solvent screening. Product structures were generated from the resulting reactant structure by adiabatic mapping along a predefined reaction coordinate defined from the distances between P_γ and the incoming/leaving oxygens. Path refinements are made with the same energy functions as during minimizations using the Conjugate Peak Refinement (CPR) algorithm. The resulting paths are analyzed in geometric terms and by decomposition of the total energies into their contributing terms. Moreover, the electrostatic contributions of the binding-pocket-forming residues to the barrier heights and reaction energies are calculated and Mulliken analyses are performed to analyze charge shifts during the reaction.

Part III

Myosin: results

Opening: Results

In this part all results from the myosin calculations presented in this thesis are described. Chapter 12 gives the results of the NUCS factors determined according to the mixed NUCS scheme described in Section 9.3. This chapter shows that the mixed NUCS scheme can be applied successfully to myosin. Chapter 13 gives the results of the hydrolysis reaction path calculations and analyses thereof. The sequence of events is outlined for each of the three paths as refined both with HF/3-21G(d) and HF/6-31G(d,p) for the QM atoms in Section 13.1. In the same section, the stationary points are described in geometric terms. Subsequently, the energy profiles along the paths as determined with both Hartree-Fock and DFT are given (Section 13.2). The total energy is decomposed into its contributing terms in Section 13.3, thus revealing which energy terms contribute most to the overall shape of the energy profile. The results from the perturbation analyses shown in Section 13.4 give insight into how specific amino acids that form the ATP binding pocket influence the energy barrier height and reaction energy. Finally, the Mulliken analyses given in Section 13.5 show that only very small charge shifts occur during the reaction.

Chapter 12

Mixed NUCS factors for myosin

Mixed NUCS factors were determined for the model structure of ATP complexed to myosin II as prepared from the crystal structure (Chapter 10). The target region contained the QM and flexible MM atoms, augmented by the atoms of the constrained region necessary to complete the NUCS groups.

12.1 NUCS factors in the target region

Global NUCS factors (see Section 9.1) are determined for all groups in the target region. The two panels in Fig. 12.1 show the Coulomb electrostatic interaction energy between a group in the target region and all other groups in the target region calculated with unscaled and with scaled charges *versus* the corresponding reference Poisson-Boltzmann interaction energies for backbone and for sidechain groups.

The root-mean square energy deviations (RMSD) are calculated separately for backbone (BACK) and sidechain (SIDE) groups. They are RMSD(BACK, vac) = 11.7 kcal/mol, RMSD(BACK, shield) = 9.7 kcal/mol, RMSD(SIDE, vac) = 118.0 kcal/mol and RMSD(SIDE, shield) = 58.7 kcal/mol. Thus, the introduction of scaling factors does not significantly improve the backbone interactions, but it does improve the sidechain interactions. Counterscaling does not lead to any further improvement. Therefore, the global scaling factors in the target region as originally calculated are used for subsequent calculations without counterscaling.

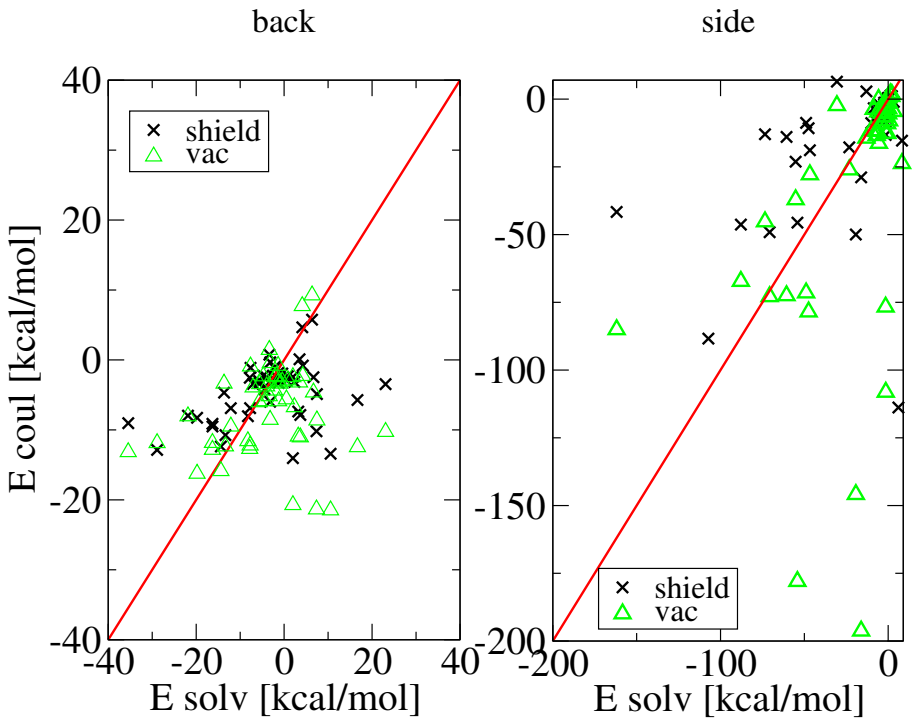

Figure 12.1: Interaction energies in kcal/mol between groups in the target region and all other groups in the target region. Coulomb energies determined from unscaled (triangles) and from scaled (crosses) charges are plotted *versus* the reference Poisson-Boltzmann energies.

12.2 NUCS factors in the non-target region

The necessity of scaling the charges in he non-target region can be shown by computing the interaction energies E_{vac} and E_{solv} of a group I outside the target region with the target region *versus* the distance of the center of mass of group I from P_γ. The resulting values are plotted in Fig. 12.2. The vacuum Coulomb interaction energies fall into three groups, corresponding to their overall charge (neutral, negatively, or positively charged). Charged groups as distant from the active site as 50 Å or more still exhibit absolute interaction energies of approximately 15 kcal/mol. In contrast, when interaction energies are determined using Poisson-Boltzmann electrostatics, the interaction energy of all groups further than 20 Å from P_γ (*i.e.*, all groups in the constrained and fixed regions) are close to zero. Thus, the necessity for scaling is obviously given.

Scaling factors for the NUCS groups in the non-target region have therefore been determined as described in Section 9.3. The Coulomb interaction energies determined from unscaled and scaled charges are plotted *versus* the Poisson-Boltzmann energies in Fig.

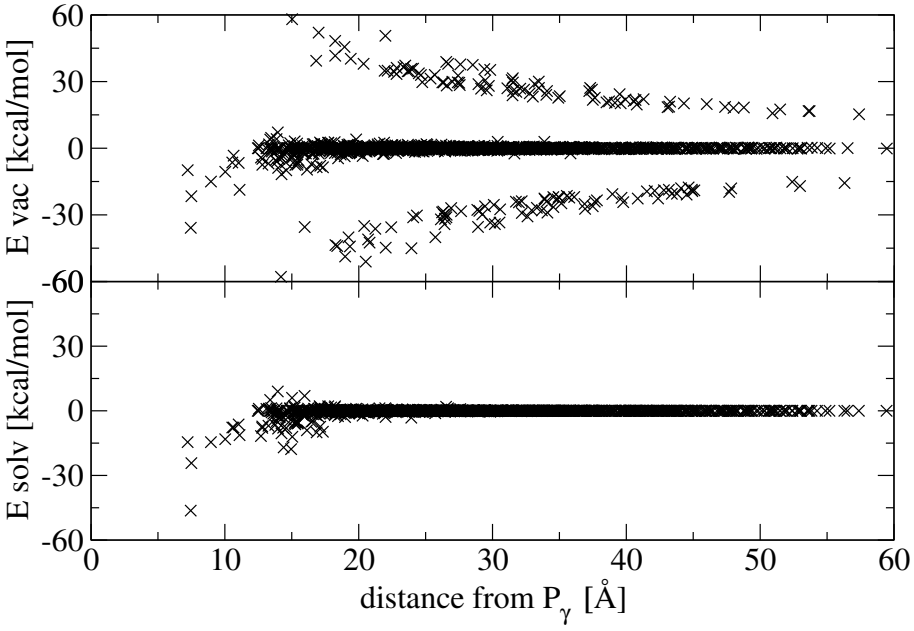

Figure 12.2: Interaction energies between the target region and groups in the non-target region as a function of the distance from P_γ.

12.3. As expected, scaling shifts all interaction energies under consideration close to the diagonal.

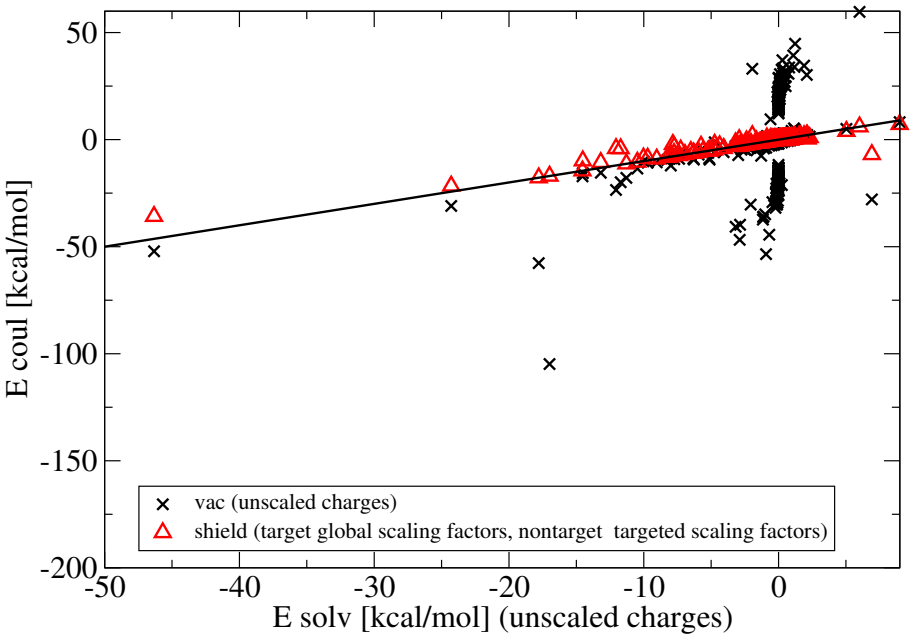

Figure 12.3: Interaction energies in kcal/mol between groups outside the target region and the target region.

12.3 Usefulness for QM/MM calculations

To validate that the NUCS factors are useful for QM/MM calculations, the three criteria given in Section 9.3 were evaluated.

1. **Potential in QM region.** The Coulomb interaction energies with unscaled (E_{vac}) and with scaled charges (E_{shield}) *versus* the PB interaction energies for each group I (irrespective of its position in space) interacting with the QM region are compared in Fig. 12.4. It is evident that the the E_{shield} are much closer to the diagonal than the E_{vac} and thus that the mixed scaling procedure can be used to derive scaling factors that reproduce the electrostatic potential in solution at the position of the QM atoms.

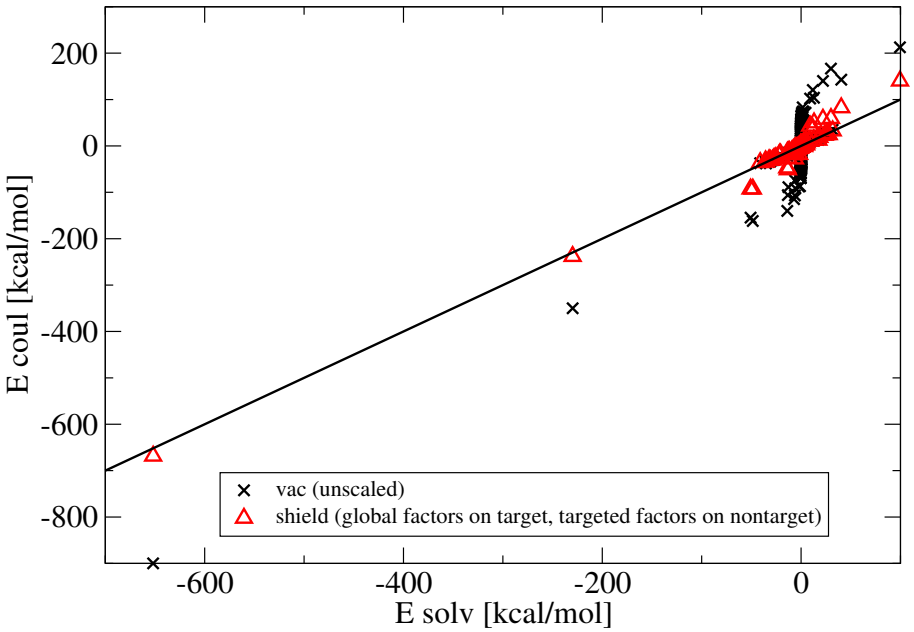

Figure 12.4: Interaction energies between the QM region and groups both in and outside of the target region.

2. **Interactions within target region.** The Coulomb interaction energies with both unscaled and scaled charges are plotted *versus* the PB interaction energies between two groups within the target region in Fig. 12.5. In contrast to the energies determined with unscaled charges, the energies determined with scaled charges are close to the diagonal. Thus, usage of scaled charges adequately approximates the Poisson-Boltzmann interaction energies as desired.

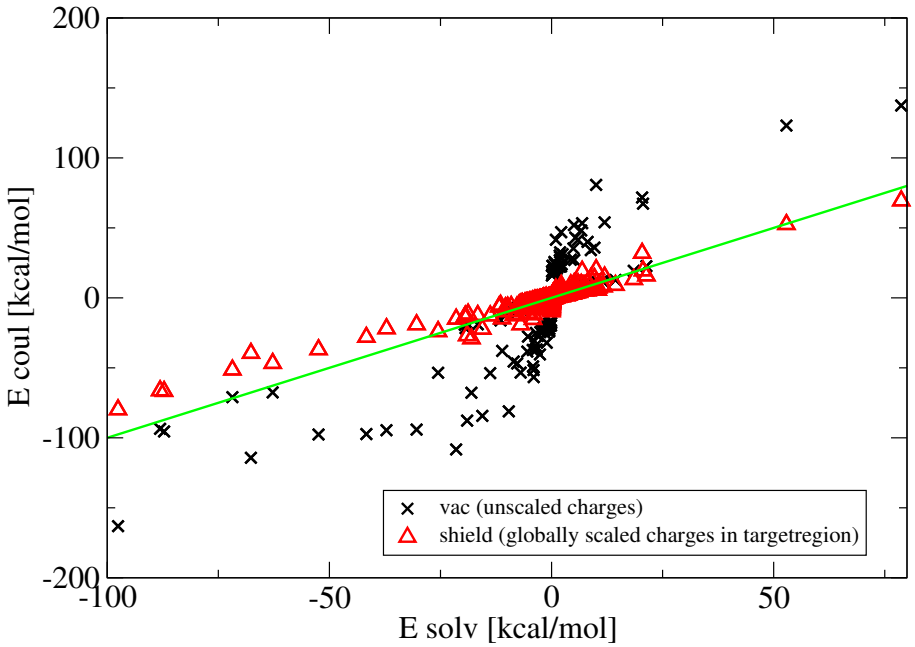

Figure 12.5: Group:group interaction energies between the groups in the targetregion.

3. **Group:nongroup interactions.** The Coulomb interaction energies with both unscaled and scaled charges *versus* the PB interaction energies between a group within the target region and all non-group atoms are compared in Fig. 12.6. It shows that mixed scaling improves the electrostatic potential at the position of a group in the target region due to the charges of all other groups.

Thus, the usefulness of the mixed NUCS scheme for QM/MM calculations is given, as shown by the satisfaction of all three criteria. For all subsequent simulations the NUCS factors described in this Section have been used.

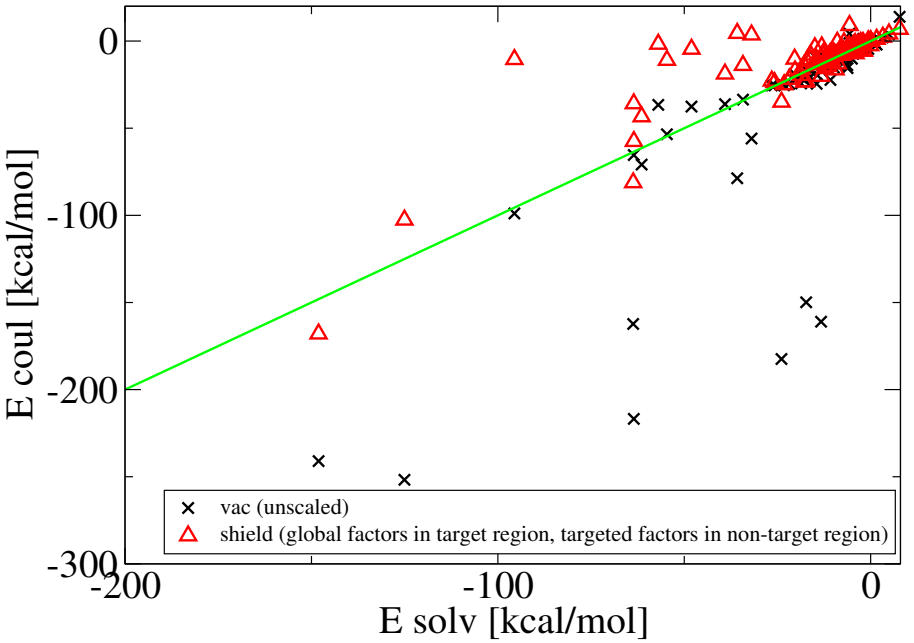

Figure 12.6: Interaction energies between a group in the target region and all nongroup atoms.

Chapter 13

Reaction paths for ATP hydrolysis in myosin

Three different reaction routes for ATP hydrolysis in myosin were studied. Fig. 13.1 shows a sketch of the reaction site (*i.e.*, the QM region) along with the nomenclature used hereafter to uniquely denote the different atoms involved. In the figure, the three water activation paths are indicated by colored arrows. The purple arrow corresponds to the direct path, in which the proton from the attacking water molecular is directly transferred to the γ-phosphate moiety of ATP. In the Ser181 path (red arrows) the proton transfer is mediated by the helper water molecule and the Ser181 sidechain. Similarly, the proton transfer is mediated by the sidechain of Ser236 in the Ser236 path (green arrows). The attack of O^a onto P_γ common to all three activation mechanisms is indicated by a dashed black arrow.

Fig. 13.2 shows the energy profiles along the normalized, one-dimensional, curvilinear reaction coordinates for the three paths as refined on the QM[HF/3-21G(d)]/MM/NUCS and QM[HF/6-31G(d,p)]/MM/NUCS potential energy surfaces. Remarkably, the barrier heights of the three mechanisms, when comparing the values obtained using a given QM method, are very close to each other. Thus, the three activation mechanisms are, to within error, likely to be equally populated.

Detailed descriptions of the geometries of the stationary points and the sequences of events along the paths are given in Section 13.1. Energetic analyses are described in detail Sections 13.2 to 13.4. The results of the charge shift analysis are given in Section 13.5.

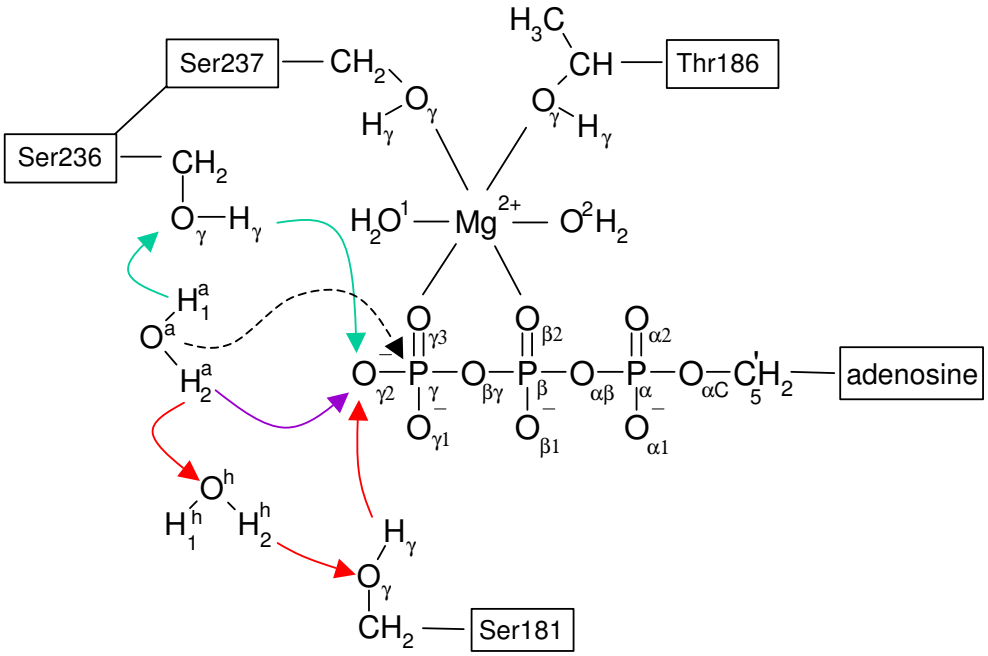

Figure 13.1: Sketch of the reaction site of ATP hydrolysis in myosin. All QM atoms are shown. The three different water activation paths studied are indicated with arrows: The purple arrow corresponds to the direct path, red arrows to the Ser181 path, and green arrows to the Ser236 path.

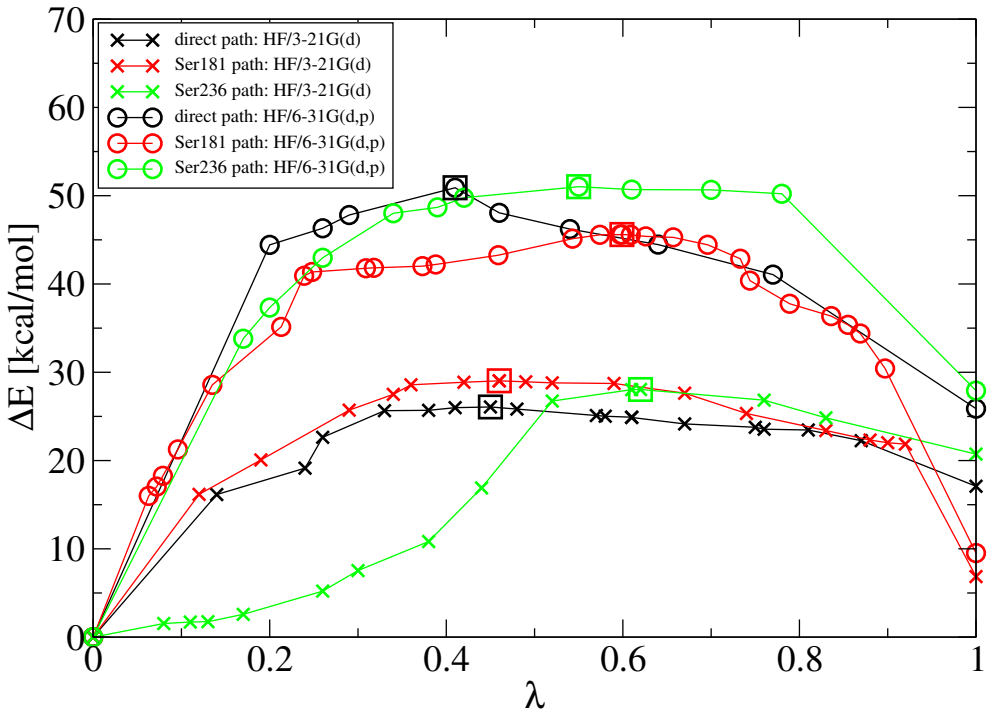

Figure 13.2: Energy profiles along the normalized curvilinear reaction coordinate λ for the direct (black), Ser181 (red), and Ser236 (green) paths as determined on the QM[HF/3-21G(d)]/MM/NUCS (crosses) and QM[HF/6-31G(d,p)]/MM/NUCS (circles) potential energy surfaces. Saddle points are indicated by squares.

13.1 Geometric description of the paths

The three different paths that were considered, *i.e.*, the direct attack pathway, the Ser181 pathway and the Ser236 pathway all start from the identical reactant state. Thus, in this Section, the geometry of the reactant state is described first, followed by a geometric description of the three different reaction pathways.

13.1.1 Reactant

The reactant state as minimized on the QM[HF/3-21G(d)]/MM/NUCS potential energy surface is visualized in Fig. 13.3.

Geometric distortions may occur at the boundary between the QM and MM regions that were linked together using the default link-atom method as is built in the CHARMM/GAMESS-US interface (279). The link atom should in principle be positioned along the bond between the QM and MM boundary atoms. As shown in Fig. 13.4 this could, however, not be achieved. The angle m_1:q_0:q_1[1] varies between 131° and 167° indicating significant deviation from linearity (see Table 13.1). In addition, the distances between MM and QM boundary atoms are larger than their respective values as parametrized in the CHARMM force field. For ATP, $d(C_5'^{ATP} : C_4'^{ATP})$ is 1.52 Å, as opposed to an MM distance of 1.51 Å.[2] In case of Thr[3] and Ser[4] the distances $d(C_\alpha : C_\beta)$ are 1.59 Å and 1.63 to 1.65 Å, respectively, as opposed to MM distances of 1.53 and 1.52 Å. However, the maximum distance deviation is < 0.15 Å and thus the resulting inconsistency of the geometry of the QM region is supposedly negligible. For the angles m_1:q_1:q_2 deviations from the parametrized values are also found. Instead of $C_4'^{ATP} : C_5'^{ATP} : O_{\alpha C}^{ATP} = 108.4°$[5], $C_\alpha^{Ser} : C_\beta^{Ser} : O_\gamma^{Ser} = 111.0°$[6], and $C_\alpha^{Thr} : C_\beta^{Thr} : O_\gamma^{Thr} = 104.5°$[7] as parametrized in the CHARMM force field, values of 113.0° (ATP), 105.5° to 106.8° (Ser), and 103.5° (Thr) are found. However, no angle deviates by more than ∼5°, as is in the tolerated range. In addition, the Ser angles are larger than the Thr angle, as is the case in the force field. Thus, the geometries at the QM/MM boundary are sufficiently well described.

Table 13.2 summarizes the distances and angles that characterize the reactant state. The geometry of the triphosphate moiety of ATP remains largely conserved as revealed by comparing the crystal structure, the QM[HF/3-21G(d)]/MM/NUCS-minimized structure, and the QM[HF/6-31G(d,p)]/MM/NUCS-minimized with the solution structure for Mg^{2+}-

[1] For the nomenclature see Section 7.2.2.
[2] C_4' is of type CN7 and C_5' of type CN8B.
[3] Both C_α^{Thr} and C_β^{Thr} are of type CH1E.
[4] C_α^{Ser} is of type CH1E and C_β^{Ser} of type CH2E.
[5] C_4' is of type CN7, C_5' of type CN8B, and $O_{\alpha C}$ of type ON2
[6] C_α^{Ser} is of type CH1E, C_β^{Ser} of type CH2E, and O_γ^{Ser} of type OH1.
[7] C_α^{Thr} is of type CH1E, C_β^{Thr} of type CH1E, and O_γ^{Thr} of type OH1.

coordinated methyl triphosphate (Mg.MTP$_{aq}$) (168). The distance between a phosphorus and a terminal oxygen varies between 1.48 and 1.54 Å an is thus close to the range observed for Mg.MTP$_{aq}$ (1.51 to 1.55 Å). In the crystal structure, the P:O$_{bridge}$ bonds are modelled to be of equal length (1.63 Å). Minimization introduces an asymmetry by elongating the $P_\gamma : O_{\beta\gamma}$ bond to 1.65 (HF/3-21G(d)) and 1.68 Å(HF/6-31G(d,p)) and shortening the $P_\beta : O_{\beta\gamma}$ to 1.60 (both HF/3-21G(d) and HF/6-31G(d,p)). These distances are slightly shorter than the equivalent distances in Mg.MTP$_{aq}$ which are 1.70 and 1.62 Å, respectively. Nevertheless, the asymmetry of the P:O$_{bridge}$ bond lengths indicates that Mg^{2+} activates the P$_\gamma$:O$_{\beta\gamma}$ bond for cleavage. The angle a($P_\gamma : O_{\beta\gamma} : P_\beta$) is 126.2° and 126.6° on the HF/3-21G(d) and HF/6-31G(d,p) surfaces, respectively and thus similar to the value of 128.3° for Mg.MTP$_{aq}$. The angles a($O_{\beta\gamma} : P_\gamma : O_\gamma$) vary between 103.2° and 106.7° and are thus somewhat smaller than the angle expected for tetrahedral arrangement (109.5°). Thus, the enzymatic environment slightly distorts the γ-phosphate moiety of ATP into the direction of a trigonal transition state configuration in which the angle a($O_{\beta\gamma} : P_\gamma : O_\gamma$) would be 90°.

Overlaps of the reactive regions of the reactant state as minimized on the QM[HF/3-21G(d)]/MM/NUCS energy surface and the crystal structure (Fig. 13.5) and of the reactant as minimized on the QM[HF/6-31G(d,p)]/MM/NUCS and QM[HF/3-21G(d)]/MM/ NUCS energy surfaces (Fig. 13.6) indicate that significant changes only occurred in the position and relative orientation of the attacking and the helper water molecules. Especially the triphosphate moiety of ATP exhibits only minor geometric changes with the geometry varying only little between the two different basis sets used.

Figs. 13.7 and 13.8 show the corresponding overlaps of the Mg^{2+} coordination sphere. In all three structures, an almost perfect octahedron is observed which is in accordance with the Mg.MTP$_{aq}$ structure. Thus, the enzyme environment does not introduce changes in the Mg^{2+} coordination pattern as compared to in aqueous environment.

The attacking water (H_2O^a in Fig. 13.1) forms hydrogen bonds to the carbonyl oxygen of Ser237 (distance d($O^a : O^{S237}$) = 2.81 Å) and to O$_{\gamma 2}$ of the γ-phosphate moiety of ATP (d($O^a : O_{\gamma 2}$) = 3.06 Å). The sidechains of both Ser181 and Ser236 also form hydrogen bonds to O$_{\gamma 2}$ with distances d($O^{S181} : O_{\gamma 2}$) = 2.77 Å and d($O^{S236} : O_{\gamma 2}$) = 2.86 Å. The helper water is hydrogen bonded to the sidechains of Ser181 (d($O^h : O^{S181}$) = 2.96 Å) and Ser236 (d($O^h : O^{S236}$) = 3.13 Å). Altogether, this forms a hydrogen-bond network around the γ-phosphate which favors multiple pathways for proton transfer and positions the proton donors corresponding to the three paths for efficient transfer of a proton onto O$_{\gamma 2}$. These paths are described in the following sections and can best be understood by watching the accompanying molecular movies.

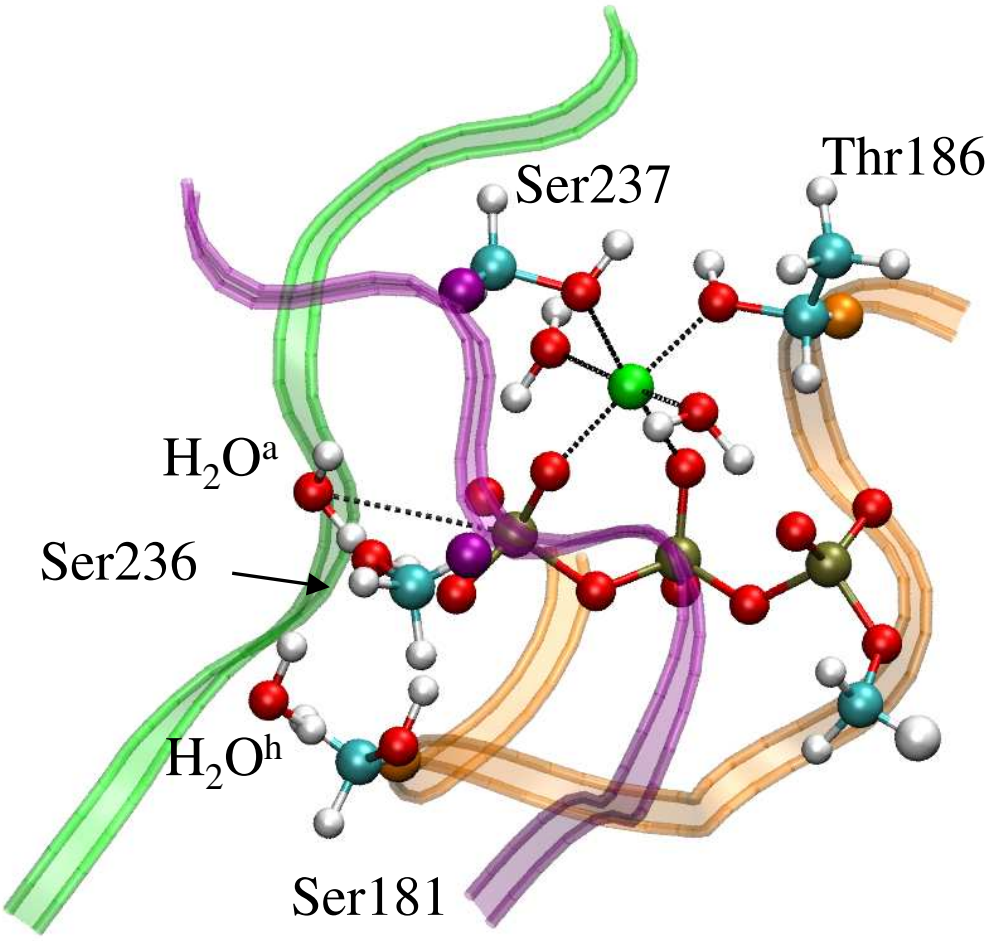

Figure 13.3: eactant state as minimized on the QM[HF/3-21G(d)]/MM/NUCS potential energy surface. Only atoms treated quantum mechanically are explicitly shown, *i.e.*, the triphosphate moiety of ATP, Mg^{2+}, the attacking (H_2O^a), helper (H_2O^h), and Mg^{2+}-coordinating water molecules, and the sidechains of Ser181, Thr186, Ser236, and Ser237. The P-loop is drawn in orange, the switch-1 loop in purple, and the switch-2 loop in green. The link atoms (shown as spheres with larger radius than usual hydrogen atoms) are colored according to the loop to which the residue is attached.

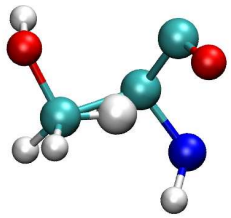

Figure 13.4: Visualization of Ser181 in the reactant state as minimized on the QM[HF/3-21G(d)]/MM/NUCS potential energy surface. The link atom is shown as a white sphere with larger radius than usual hydrogen atoms.

	HF/3-21G(d)	HF/6-31G(d,p)
$\mathrm{d}(C_5'^{ATP} : QQH^{ATP})$	1.092	1.095
$\mathrm{d}(C_5'^{ATP} : C_4'^{ATP})$	1.522	1.522
$\mathrm{a}(C_5'^{ATP} : QQH^{ATP} : C_4'^{ATP})$	131.553	130.958
$\mathrm{a}(C_4'^{ATP} : C_5'^{ATP} : O_{\alpha C}^{ATP})$	113.190	112.961
$\mathrm{d}(C_\beta^{S181} : QQH^{S181})$	1.085	1.088
$\mathrm{d}(C_\beta^{S181} : C_\alpha^{S181})$	1.632	1.633
$\mathrm{a}(C_\beta^{S181} : QQH^{S181} : C_\alpha^{S181})$	139.486	139.926
$\mathrm{a}(C_\alpha^{S181} : C_\beta^{S181} : O_\gamma^{S181})$	101.186	101.604
$\mathrm{d}(C_\beta^{T186} : QQH^{T186})$	1.079	1.083
$\mathrm{d}(C_\beta^{T186} : C_\alpha^{T186})$	1.590	1.592
$\mathrm{a}(C_\beta^{T186} : QQH^{T186} : C_\alpha^{T186})$	143.930	143.289
$\mathrm{a}(C_\alpha^{T186} : C_\beta^{T186} : O_\gamma^{T186})$	102.333	103.490
$\mathrm{d}(C_\beta^{S236} : QQH^{S236})$	1.083	1.087
$\mathrm{d}(C_\beta^{S236} : C_\alpha^{S236})$	1.629	1.630
$\mathrm{a}(C_\beta^{S236} : QQH^{S236} : C_\alpha^{S236})$	139.850	139.917
$\mathrm{a}(C_\alpha^{S236} : C_\beta^{S236} : O_\gamma^{S236})$	101.568	101.801
$\mathrm{d}(C_\beta^{S237} : QQH^{S237})$	1.072	1.078
$\mathrm{d}(C_\beta^{S237} : C_\alpha^{S237})$	1.650	1.650
$\mathrm{a}(C_\beta^{S237} : QQH^{S237} : C_\alpha^{S237})$	166.601	165.784
$\mathrm{a}(C_\alpha^{S237} : C_\beta^{S237} : O_\gamma^{S237})$	105.115	105.476

Table 13.1: Distances d (in Å) and angles a (in °) characteristic of the QM/MM linking sites in the reactant state.

	crystal	HF/3-21G(d)	HF/6-31G(d,p)	Mg.MTP$_{aq}$ [a]
d($P_\gamma : O^a$)	4.266	3.353	3.556	-
d($P_\gamma : O_{\gamma 1}$)	1.516	1.498	1.497	1.51 - 1.55
d($P_\gamma : O_{\gamma 2}$)	1.514	1.515	1.503	1.51 - 1.55
d($P_\gamma : O_{\gamma 3}$)	1.528	1.513	1.516	1.51 - 1.55
d($P_\gamma : O_{\beta\gamma}$)	1.634	1.653	1.676	1.70
d($P_\beta : O_{\beta\gamma}$)	1.635	1.598	1.599	1.62
d($P_\beta : O_{\beta 1}$)	1.526	1.481	1.485	1.51 - 1.55
d($P_\beta : O_{\beta 2}$)	1.541	1.499	1.495	1.51 - 1.55
a($P_\gamma : O_{\beta\gamma} : P_\beta$)	133.547	126.214	126.646	128.3
a($O_{\beta\gamma} : P_\gamma : O_{\gamma 1}$)	101.199	105.962	106.696	-
a($O_{\beta\gamma} : P_\gamma : O_{\gamma 2}$)	103.742	104.834	103.797	-
a($O_{\beta\gamma} : P_\gamma : O_{\gamma 3}$)	103.490	103.203	103.235	-
a($O_{\beta\gamma} : P_\gamma : O^a$)	166.717	156.258	157.148	-
a($O^a : P_\gamma : O_{\gamma 1}$)	80.948	72.442	72.899	-
a($O^a : P_\gamma : O_{\gamma 2}$)	64.038	58.162	58.910	-
a($O^a : P_\gamma : O_{\gamma 3}$)	87.128	98.973	97.683	-
a($P_\gamma : O_{\gamma 1} : O_{\gamma 2} : O_{\gamma 3}$)	24.724	28.837	28.739	-
d($O^a : H_2^a$)	1.002	0.966	0.947	-
d($H_2^a : O_{\gamma 2}$)	3.456	1.949	2.141	-
d($H_2^a : O^h$)	3.852	3.543	3.847	-
d($O^h : H_2^h$)	0.964	0.973	0.948	-
d($H_2^h : O_\gamma^{S181}$)	1.869	1.850	2.045	-
d($O_\gamma^{S181} : H_\gamma^{S181}$)	0.978	0.988	0.956	-
d($H_\gamma^{S181} : O_{\gamma 2}$)	3.279	1.652	1.810	-
d($O^a : H_1^a$)	0.959	0.968	0.948	-
d($H_1^a : O_\gamma^{S236}$)	3.375	2.564	2.894	-
d($O_\gamma^{S236} : H_\gamma^{S236}$)	0.977	0.987	0.956	-
d($H_\gamma^{S236} : O_{\gamma 2}$)	1.860	1.707	1.901	-
d($Mg : O_{\gamma 3}$)	1.738	1.965	1.968	2.15
d($Mg : O_{\beta 2}$)	1.788	2.038	2.049	2.15
d($Mg : O^{w1}$)	1.753	2.042	2.103	-
d($Mg : O^{w2}$)	1.750	2.052	2.104	-
d($Mg : O_{\gamma 1}^{T186}$)	1.905	2.019	2.080	-
d($Mg : O_\gamma^{S237}$)	1.874	2.140	2.233	-
a($O_{\gamma 3} : Mg : O_{\beta 2}$)	94.279	90.279	91.101	90.0

[a]Geometric data for Mg.MTP (methyl triphosphate) in aqueous solution as determined from Car-Parrinello molecular dynamics (168).

Table 13.2: Distances d (in Å) and angles a (in °) characteristic of the QM region in the reactant state.

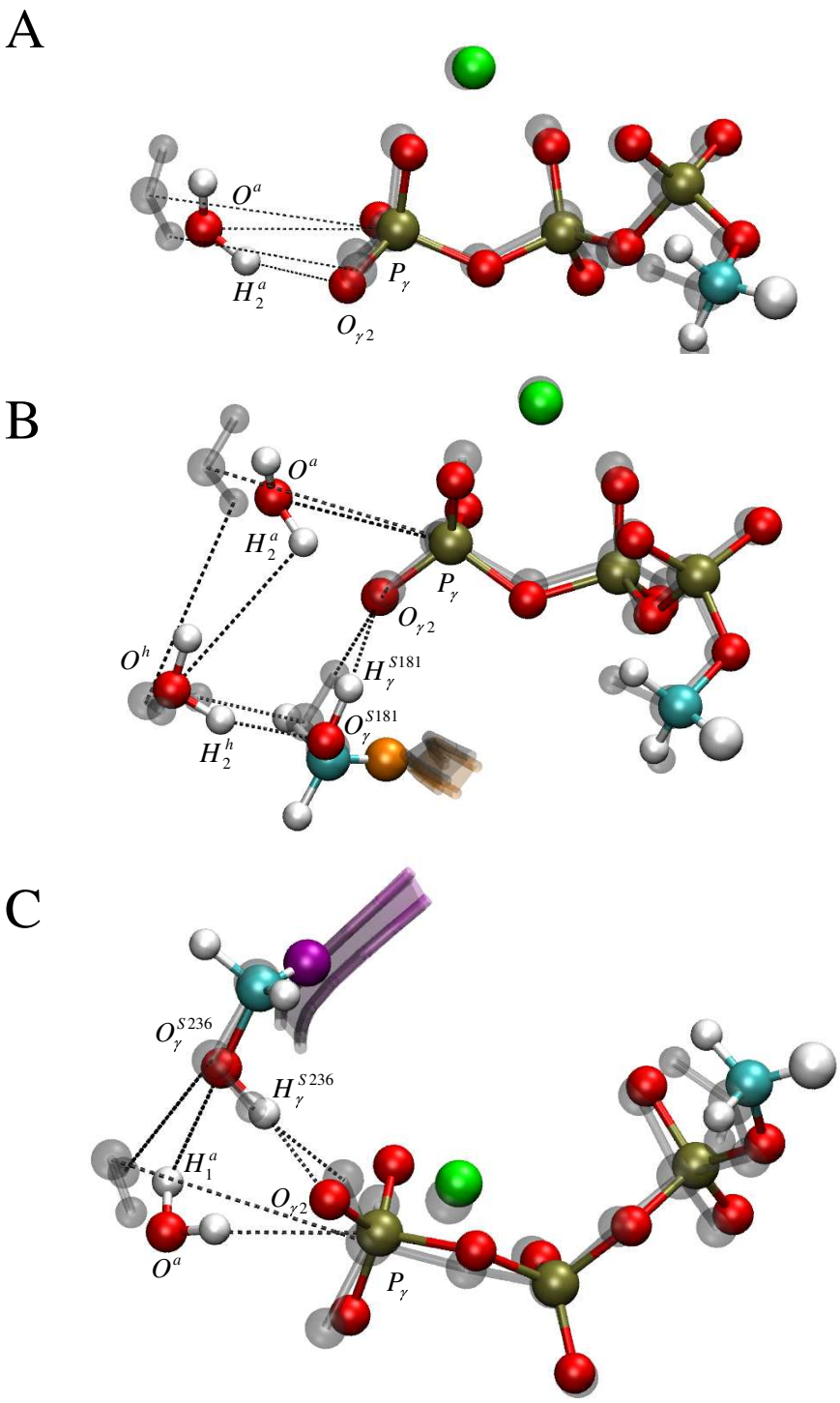

Figure 13.5: Visualization of the reactant state for the direct pathway (panel A), the Ser181 pathway (panel B), and the Ser236 pathway (panel C). Both the minimized structure (on the QM[HF/3-21G(d)]/MM/NUCS potential energy surface, in color) and the starting structure for minimization (as derived from the crystal structure (55), in gray) are shown. The link atoms are shown as spheres with larger radius than usual hydrogen atoms.

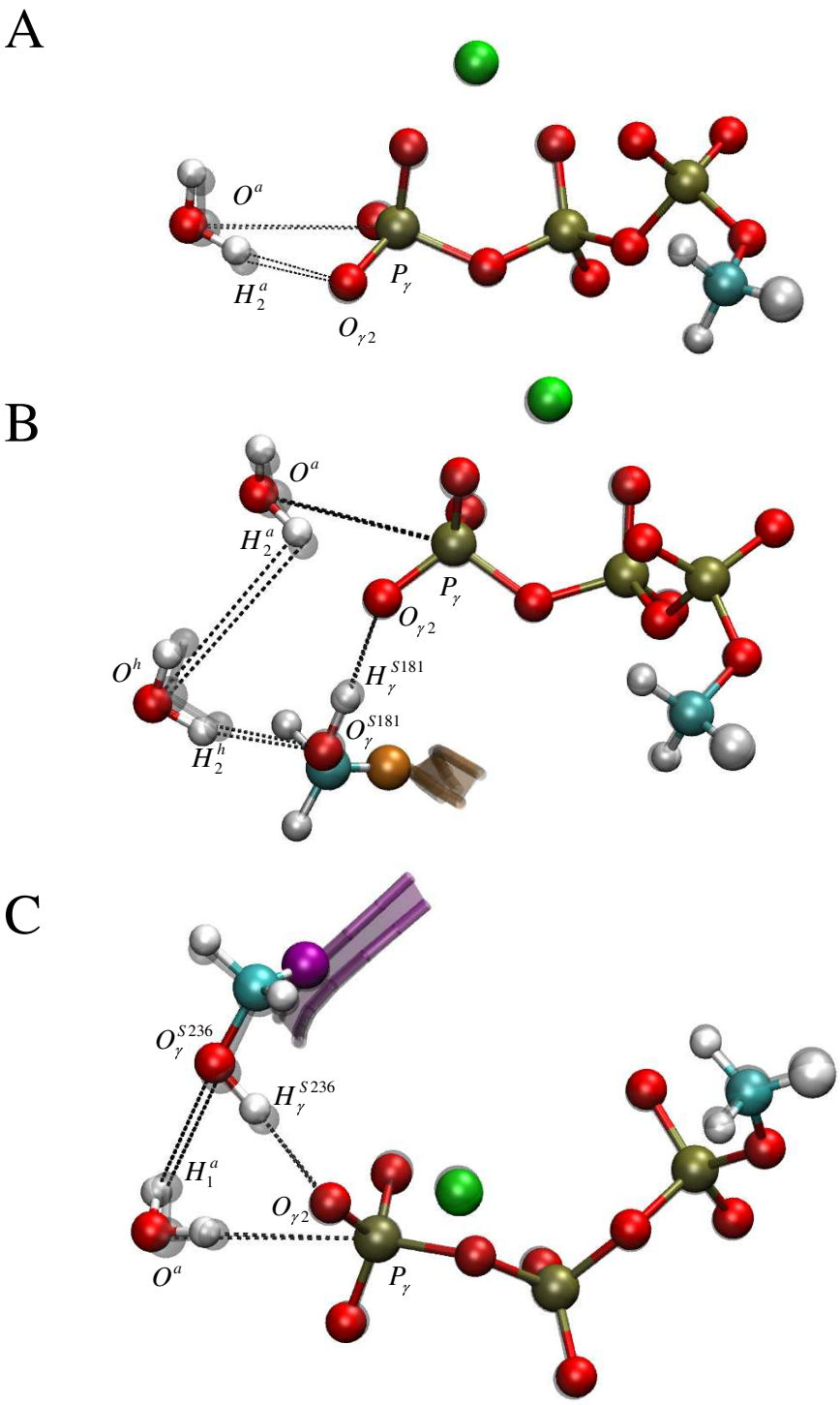

Figure 13.6: Visualization of the reactant state for the direct pathway (panel A), the Ser181 pathway (panel B), and the Ser236 pathway (panel C). The minimized structures on the QM[HF/6-31G(d,p)]/MM/NUCS potential energy surface (in color) and on the QM[HF/3-21G(d)]/MM/NUCS potential energy surface (in gray) are shown. The link atoms are shown as spheres with larger radius than usual hydrogen atoms.

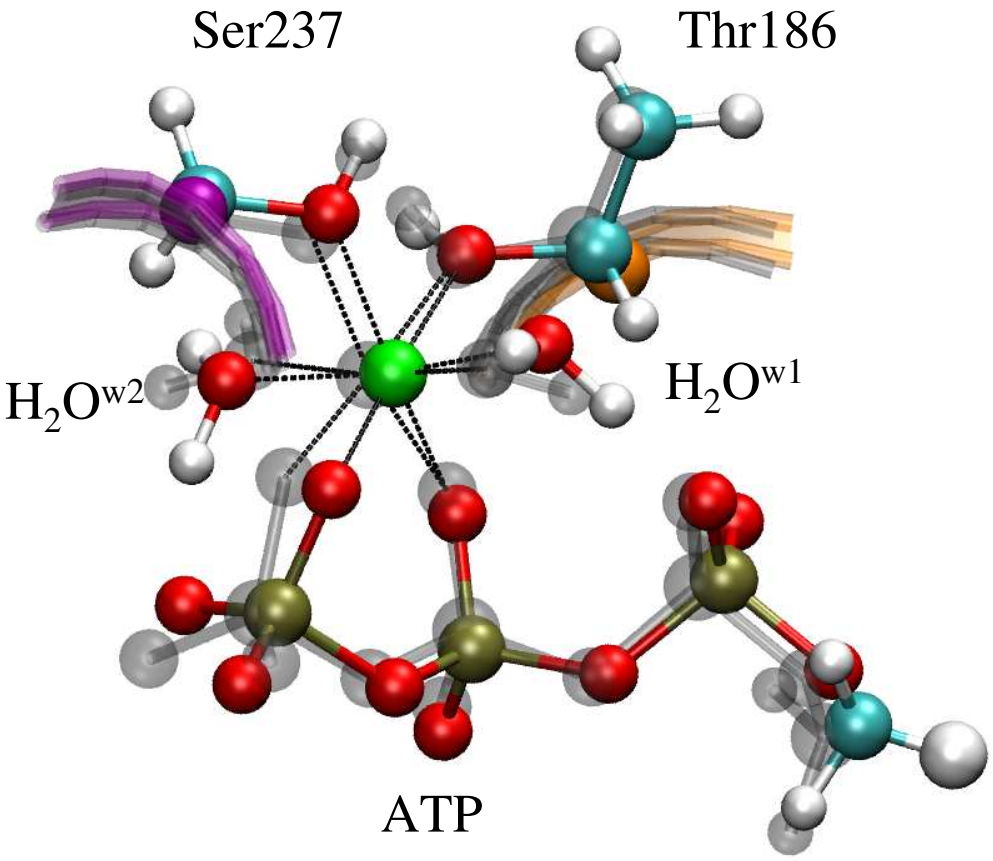

Figure 13.7: Visualization of the Mg^{2+} coordination sphere in the reactant state. Both the minimized structure (on the QM[HF/3-21G(d)]/MM/NUCS potential energy surface, in color) and the starting structure for minimization (as derived from the crystal structure (55), in gray) are shown. The link atoms are shown as spheres with larger radius than usual hydrogen atoms.

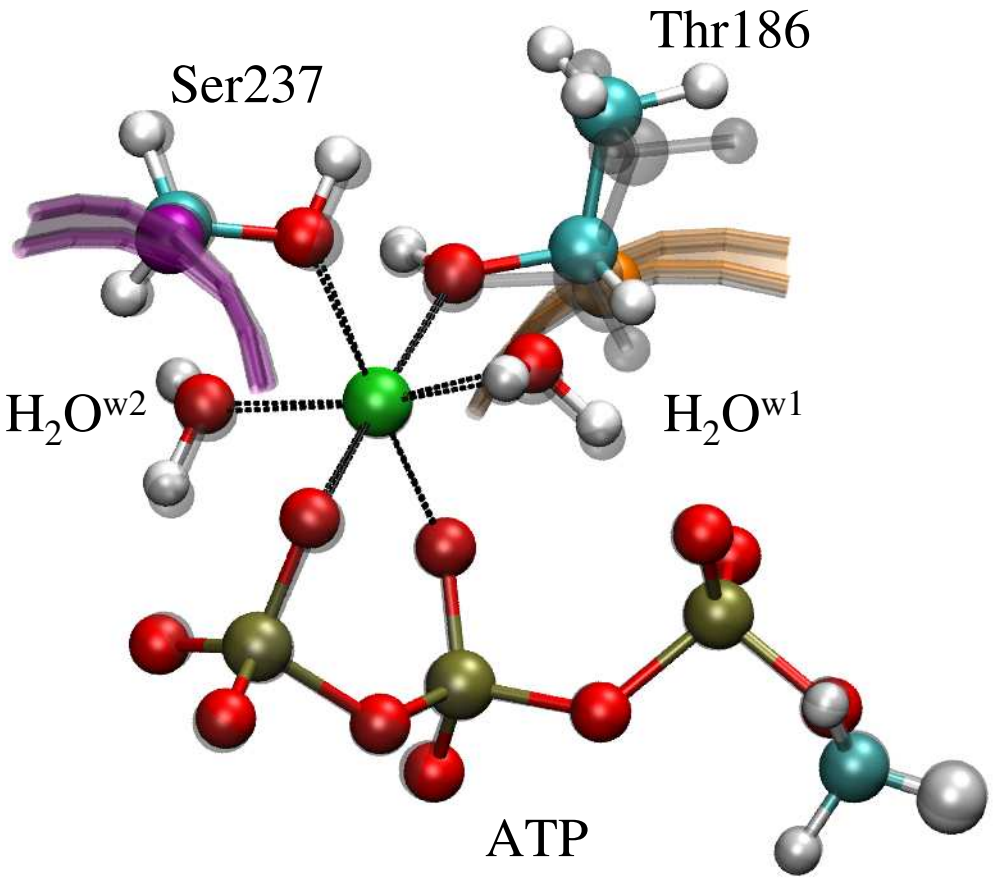

Figure 13.8: Visualization of the Mg^{2+} coordination sphere in the reactant state. Both the minimized structure on the QM[HF/6-31G(d,p)]/MM/NUCS potential energy surface (in color) and on the QM[HF/3-21G(d)]/MM/NUCS potential energy surface (in gray) are shown. The link atoms are shown as spheres with larger radius than usual hydrogen atoms.

13.1.2 Direct path

In the direct path, one of the protons from the attacking water (namely H_2^a) is directly transferred to the $O_{\gamma 2}$-oxygen of the γ-phosphate moiety of ATP. Tables 13.3 and 13.4 show the geometries of the linking sites in the transition and product states as refined on both the QM[HF/3-21G(d)]/MM/NUCS and QM[HF/6-31G(d,p)]/MM/NUCS potential energy surfaces. As in the reactant state, the angle m_1:q_0:q_1 deviates from linearity and the distances between MM and QM boundary atoms are larger than their respective distances as parametrized in the CHARMM force field. Tables 13.5 and 13.6 summarize the distances and angles that characterize the transition and product states for the direct path. The paths as refined on the two potential energy surfaces are described below.

QM[HF/3-21G(d)]/MM/NUCS. Starting from the reactant structure (reaction coordinate $\lambda = 0.00$) the first event in the direct path is the activation of the attacking water by proton transfer of the H_2^a proton from the attacking water to $O_{\gamma 2}$ of ATP, which is completed at $\lambda = 0.38$. Almost concertedly, Ser236 transfers its hydroxyl proton H_γ^{S236} onto the attacking water. This proton transfer event is completed at $\lambda = 0.41$. At $\lambda = 0.45$ the transition state is reached. As shown in Fig. 13.9 the transition-state structure of ATP is close to its reactant-state structure. At $\lambda = 0.67$ a transient state with trigonal bipyramidal geometry is reached that is characterized by an improper angle a($P_\gamma : O_{\gamma 1} : O_{\gamma 2} : O_{\gamma 3}$) of -1.4° and distances d($P_\gamma : O^a$) = 1.82 Å and d($P_\gamma : O_{\beta\gamma}$) = 1.91 Å. In this state, the H_γ^{S236} is backtransferred onto the sidechain of Ser236. Up to the transient state at $\lambda = 0.67$ the angle a($O_{\gamma 3} : Mg : O_{\beta 2}$) remains close to 90° and the distance d($Mg : O_\gamma^{S237}$) remains smaller than 2.2 Å. Upon decay of the trigonal bipyramidal transient state, the angle a($O_{\gamma 3} : Mg : O_{\beta 2}$) widens to its final value of 117.3° and the distance d($Mg : O_\gamma^{S237}$) increases to 3.3 Å (Fig. 13.10). The increase of the $Mg : O_\gamma^{S237}$ distance is due to both Mg moving towards the ADP.P_i moieties and O_γ^{S237} moving away from its original position.

QM[HF/6-31G(d,p)]/MM/NUCS. Starting from the reactant structure (reaction coordinate $\lambda = 0.00$) the first event in the direct path is the activation of the attacking water by proton transfer of the H_2^a proton from the attacking water to $O_{\gamma 2}$ of ATP, which is completed at $\lambda = 0.29$. The resulting HO^{a-} hydroxide then attacks P_γ. At $\lambda = 0.41$ the transition state is reached. Unlike the transition state on the QM[HF/3-21G(d)]/MM/NUCS surface, the transition state on the QM[HF/6-31G(d,p)]/MM/NUCS surface is almost trigonal bipyramidal (Fig. 13.11). A transient state with clear trigonal bipyramidal structure as characterized by an improper angle a($P_\gamma : O_{\gamma 1} : O_{\gamma 2} : O_{\gamma 3}$) of -0.04° and distances d($P_\gamma : O^a$) = 1.81 Å (to the attacking oxygen) and d($P_\gamma : O_{\beta\gamma}$) = 1.84 Å (to the leaving oxygen) is reached at $\lambda = 0.46$ which decays to the product state (Fig. 13.12). In contrast to the path on the QM[HF/3-21G(d)]/MM/NUCS surface, no spontaneous proton transfer from the sidechain of Ser236 to the attacking water is observed. Up to the transition state at $\lambda = 0.41$ the angle a($O_{\gamma 3} : Mg : O_{\beta 2}$) remains close to 90° and the distance d($Mg : O_\gamma^{S237}$) stays at 2.2 Å. Upon decay of the transition state, the angle a($O_{\gamma 3} : Mg : O_{\beta 2}$) widens

to its final value of 125.1°, pulling the Mg^{2+} between the β-phosphate and the leaving γ-phosphate and provoking an increase of the distance d($Mg : O_{\gamma}^{S237}$) to 3.6 Å. As in the HF/3-21G(d)-refined path, this increase is due to Mg^{2+} and O_{γ}^{S237} moving in opposite directions. This is very important, because it breaks the coordination bond between Mg^{2+} and the switch-1 loop in the product state (Fig. 13.12), thereby weakening one of the strong interactions that maintain switch-1 closed over the nucleotide.

The reaction paths as refined on the two different potential energy surfaces share common characteristics. As shown in Fig. 13.13 both paths proceed *via* a transient state in which P_{γ}, $O_{\gamma 1}$, $O_{\gamma 2}$, and $O_{\gamma 3}$ lie in a plane (as expected for both an associative and a dissociative mechanism). However, on the QM[HF/3-21G(d)]/MM/NUCS surface the transition state is much earlier than the trigonal bipyramidal state. Both paths exhibit linear dependence of the distance d($Mg : O_{\gamma}^{S237}$) *versus* the angle a($O_{\gamma 3} : Mg : O_{\beta 2}$) as shown in Fig. 13.14. In addition, both paths follow the same route when plotting the distance d($P_{\gamma} : O^{a}$) *versus* the distance d($P_{\gamma} : O_{\beta\gamma}$) (Fig. 13.15). This figure reveals that the distance between the P_{γ} and the incoming oxygen is reduced *prior* to an increase of the distance between P_{γ} and the $\beta\gamma$-bridge oxygen. This indicates that the direct path follows an associative mechanism.

	HF/3-21G(d)	HF/6-31G(d,p)
$\mathrm{d}(C_5'^{ATP} : QQH^{ATP})$	1.091	1.096
$\mathrm{d}(C_5'^{ATP} : C_4'^{ATP})$	1.522	1.522
$\mathrm{a}(C_5'^{ATP} : QQH^{ATP} : C_4'^{ATP})$	131.420	130.828
$\mathrm{a}(C_4'^{ATP} : C_5'^{ATP} : O_{\alpha C}^{ATP})$	113.325	112.918
$\mathrm{d}(C_\beta^{S181} : QQH^{S181})$	1.085	1.086
$\mathrm{d}(C_\beta^{S181} : C_\alpha^{S181})$	1.633	1.636
$\mathrm{a}(C_\beta^{S181} : QQH^{S181} : C_\alpha^{S181})$	139.745	143.168
$\mathrm{a}(C_\alpha^{S181} : C_\beta^{S181} : O_\gamma^{S181})$	101.284	103.020
$\mathrm{d}(C_\beta^{T186} : QQH^{T186})$	1.078	1.083
$\mathrm{d}(C_\beta^{T186} : C_\alpha^{T186})$	1.591	1.593
$\mathrm{a}(C_\beta^{T186} : QQH^{T186} : C_\alpha^{T186})$	143.882	143.575
$\mathrm{a}(C_\alpha^{T186} : C_\beta^{T186} : O_\gamma^{T186})$	102.313	103.106
$\mathrm{d}(C_\beta^{S236} : QQH^{S236})$	1.090	1.085
$\mathrm{d}(C_\beta^{S236} : C_\alpha^{S236})$	1.634	1.637
$\mathrm{a}(C_\beta^{S236} : QQH^{S236} : C_\alpha^{S236})$	140.633	147.418
$\mathrm{a}(C_\alpha^{S236} : C_\beta^{S236} : O_\gamma^{S236})$	103.856	105.442
$\mathrm{d}(C_\beta^{S237} : QQH^{S237})$	1.071	1.077
$\mathrm{d}(C_\beta^{S237} : C_\alpha^{S237})$	1.648	1.650
$\mathrm{a}(C_\beta^{S237} : QQH^{S237} : C_\alpha^{S237})$	167.007	167.112
$\mathrm{a}(C_\alpha^{S237} : C_\beta^{S237} : O_\gamma^{S237})$	105.033	105.464

Table 13.3: Distances d (in Å) and angles a (in °) characteristic of the QM/MM linking sites in the transition state (direct path).

	HF/3-21G(d)	HF/6-31G(d,p)
d($C_5'^{ATP} : QQH^{ATP}$)	1.090	1.095
d($C_5'^{ATP} : C_4'^{ATP}$)	1.521	1.521
a($C_5'^{ATP} : QQH^{ATP} : C_4'^{ATP}$)	133.109	131.807
a($C_4'^{ATP} : C_5'^{ATP} : O_{\alpha C}^{ATP}$)	112.519	112.243
d($C_\beta^{S181} : QQH^{S181}$)	1.082	1.087
d($C_\beta^{S181} : C_\alpha^{S181}$)	1.630	1.629
a($C_\beta^{S181} : QQH^{S181} : C_\alpha^{S181}$)	140.877	139.910
a($C_\alpha^{S181} : C_\beta^{S181} : O_\gamma^{S181}$)	101.672	101.606
d($C_\beta^{T186} : QQH^{T186}$)	1.077	1.081
d($C_\beta^{T186} : C_\alpha^{T186}$)	1.602	1.604
a($C_\beta^{T186} : QQH^{T186} : C_\alpha^{T186}$)	145.311	145.344
a($C_\alpha^{T186} : C_\beta^{T186} : O_\gamma^{T186}$)	103.613	103.943
d($C_\beta^{S236} : QQH^{S236}$)	1.079	1.084
d($C_\beta^{S236} : C_\alpha^{S236}$)	1.641	1.637
a($C_\beta^{S236} : QQH^{S236} : C_\alpha^{S236}$)	151.845	145.233
a($C_\alpha^{S236} : C_\beta^{S236} : O_\gamma^{S236}$)	106.853	104.278
d($C_\beta^{S237} : QQH^{S237}$)	1.074	1.078
d($C_\beta^{S237} : C_\alpha^{S237}$)	1.656	1.653
a($C_\beta^{S237} : QQH^{S237} : C_\alpha^{S237}$)	164.878	165.054
a($C_\alpha^{S237} : C_\beta^{S237} : O_\gamma^{S237}$)	103.805	104.100

Table 13.4: Distances d (in Å) and angles a (in °) characteristic of the QM/MM linking sites in the product state (direct path).

	HF/3-21G(d)	HF/6-31G(d,p)
$\mathrm{d}(P_\gamma : O^a)$	2.750	2.180
$\mathrm{d}(P_\gamma : O_{\gamma 1})$	1.490	1.491
$\mathrm{d}(P_\gamma : O_{\gamma 2})$	1.568	1.604
$\mathrm{d}(P_\gamma : O_{\gamma 3})$	1.501	1.505
$\mathrm{d}(P_\gamma : O_{\beta\gamma})$	1.636	1.728
$\mathrm{d}(P_\beta : O_{\beta\gamma})$	1.606	1.580
$\mathrm{d}(P_\beta : O_{\beta 1})$	1.479	1.488
$\mathrm{d}(P_\beta : O_{\beta 2})$	1.498	1.499
$\mathrm{a}(P_\gamma : O_{\beta\gamma} : P_\beta)$	125.118	128.156
$\mathrm{a}(O_{\beta\gamma} : P_\gamma : O_{\gamma 1})$	107.334	101.903
$\mathrm{a}(O_{\beta\gamma} : P_\gamma : O_{\gamma 2})$	98.380	87.912
$\mathrm{a}(O_{\beta\gamma} : P_\gamma : O_{\gamma 3})$	103.480	98.226
$\mathrm{a}(O_{\beta\gamma} : P_\gamma : O^a)$	159.507	161.208
$\mathrm{a}(O^a : P_\gamma : O_{\gamma 1}$	88.218	90.456
$\mathrm{a}(O^a : P_\gamma : O_{\gamma 2}$	62.429	73.663
$\mathrm{a}(O^a : P_\gamma : O_{\gamma 3}$	80.253	87.556
$\mathrm{a}(P_\gamma : O_{\gamma 1} : O_{\gamma 2} : O_{\gamma 3})$	24.457	11.893
$\mathrm{d}(O^a : H_2^a)$	1.682	1.658
$\mathrm{d}(H_2^a : O_{\gamma 2})$	1.006	0.962
$\mathrm{d}(H_2^a : O^h)$	2.202	3.453
$\mathrm{d}(O^h : H_2^h)$	0.965	0.946
$\mathrm{d}(H_2^h : O_\gamma^{S181})$	2.016	2.130
$\mathrm{d}(O_\gamma^{S181} : H_\gamma^{S181})$	0.975	0.947
$\mathrm{d}(H_\gamma^{S181} : O_{\gamma 2})$	1.800	1.942
$\mathrm{d}(O^a : H_1^a)$	0.966	0.942
$\mathrm{d}(H_1^a : O_\gamma^{S236})$	2.925	3.121
$\mathrm{d}(O_\gamma^{S236} : H_\gamma^{S236})$	1.393	0.979
$\mathrm{d}(H_\gamma^{S236} : O_{\gamma 2})$	2.745	2.908
$\mathrm{d}(Mg : O_{\gamma 3})$	1.991	1.990
$\mathrm{d}(Mg : O_{\beta 2})$	2.048	2.042
$\mathrm{d}(Mg : O^{w1})$	2.025	2.091
$\mathrm{d}(Mg : O^{w2})$	2.048	2.103
$\mathrm{d}(Mg : O_{\gamma 1}^{T186})$	2.026	2.089
$\mathrm{d}(Mg : O_\gamma^{S237})$	2.122	2.239
$\mathrm{a}(O_{\gamma 3} : Mg : O_{\beta 2})$	88.309	90.676

Table 13.5: Distances d (in Å) and angles a (in °) characteristic of the QM region in the transition state (direct path).

	HF/3-21G(d)	HF/6-31G(d,p)
d($P_\gamma : O^a$)	1.632	1.612
d($P_\gamma : O_{\gamma 1}$)	1.491	1.477
d($P_\gamma : O_{\gamma 2}$)	1.562	1.593
d($P_\gamma : O_{\gamma 3}$)	1.501	1.498
d($P_\gamma : O_{\beta\gamma}$)	2.987	3.300
d($P_\beta : O_{\beta\gamma}$)	1.497	1.497
d($P_\beta : O_{\beta 1}$)	1.496	1.500
d($P_\beta : O_{\beta 2}$)	1.541	1.538
a($P_\gamma : O_{\beta\gamma} : P_\beta$)	127.030	124.769
a($O_{\beta\gamma} : P_\gamma : O_{\gamma 1}$)	86.773	83.422
a($O_{\beta\gamma} : P_\gamma : O_{\gamma 2}$)	73.213	69.797
a($O_{\beta\gamma} : P_\gamma : O_{\gamma 3}$)	71.474	71.216
a($O_{\beta\gamma} : P_\gamma : O^a$)	163.381	163.095
a($O^a : P_\gamma : O_{\gamma 1}$	109.085	110.990
a($O^a : P_\gamma : O_{\gamma 2}$	94.932	95.519
a($O^a : P_\gamma : O_{\gamma 3}$	104.677	108.082
a($P_\gamma : O_{\gamma 1} : O_{\gamma 2} : O_{\gamma 3}$)	-24.463	-28.386
d($O^a : H_2^a$)	2.362	2.340
d($H_2^a : O_{\gamma 2}$)	1.019	0.955
d($H_2^a : O^h$)	1.531	2.053
d($O^h : H_2^h$)	0.977	0.950
d($H_2^h : O_\gamma^{S181}$)	1.852	2.033
d($O_\gamma^{S181} : H_\gamma^{S181}$)	0.977	0.949
d($H_\gamma^{S181} : O_{\gamma 2}$)	1.840	2.073
d($O^a : H_1^a$)	0.976	0.956
d($H_1^a : O_\gamma^{S236}$)	3.212	3.334
d($O_\gamma^{S236} : H_\gamma^{S236}$)	0.980	0.945
d($H_\gamma^{S236} : O_{\gamma 2}$)	2.699	2.386
d($Mg : O_{\gamma 3}$)	1.992	2.033
d($Mg : O_{\beta 2}$)	1.884	1.906
d($Mg : O^{w1}$)	1.965	2.027
d($Mg : O^{w2}$)	2.062	2.119
d($Mg : O_{\gamma 1}^{T186}$)	2.097	2.201
d($Mg : O_\gamma^{S237}$)	3.271	3.583
a($O_{\gamma 3} : Mg : O_{\beta 2}$)	117.279	125.173

Table 13.6: Distances d (in Å) and angles a (in °) characteristic of the QM region in the product state (direct path).

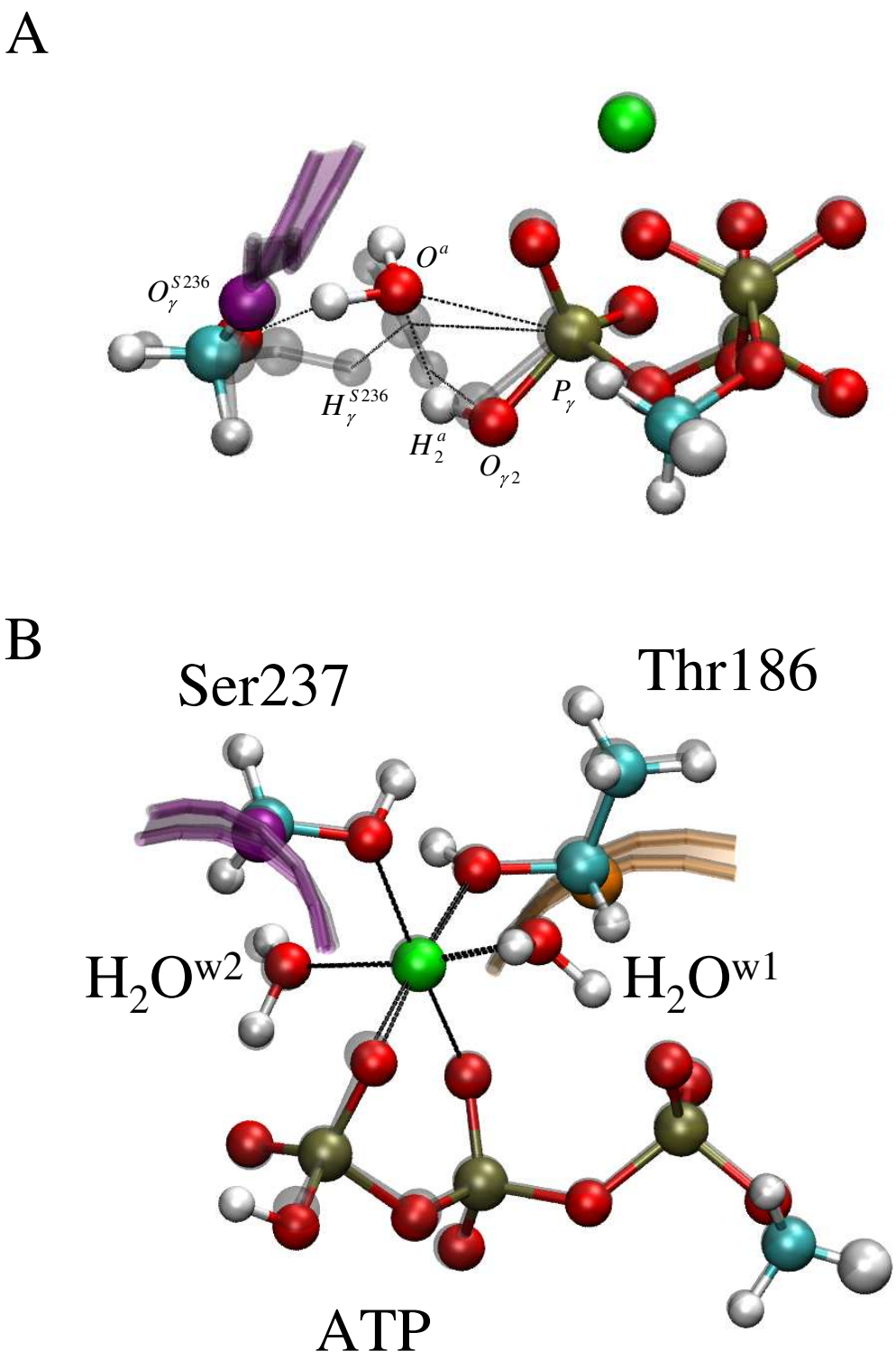

Figure 13.9: Overlap of the transition state (in color) of the direct path and the reactant state (in gray) as refined on the QM[HF/3-21G(d)]/MM/NUCS potential energy surface. Both the reaction site (panel A) and the Mg^{2+} coordination sphere (panel B) are shown. The link atoms are shown as spheres with larger radius than usual hydrogen atoms.

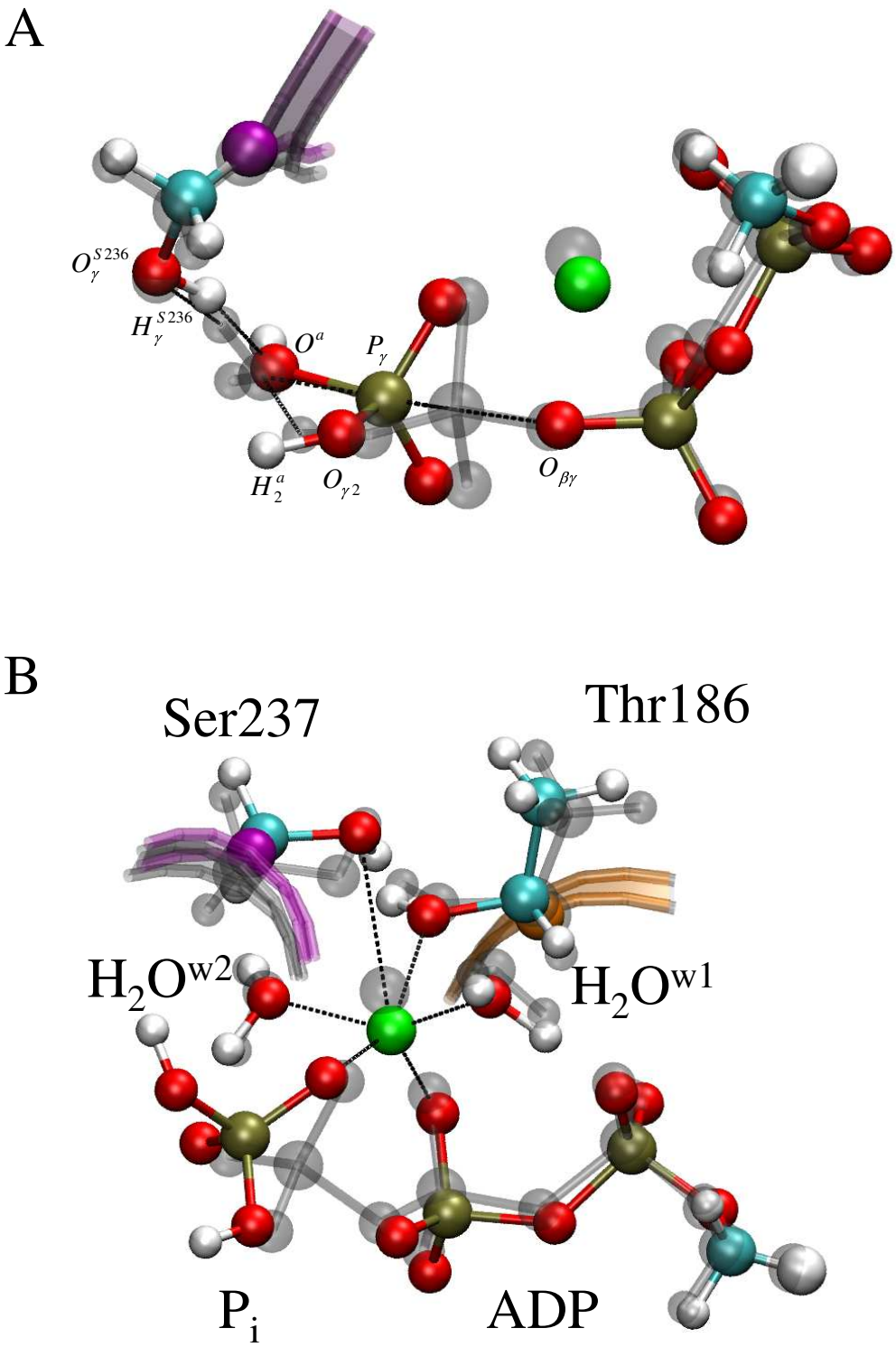

Figure 13.10: Overlap of the product state (in color) and the transition state (in gray) of the direct path as refined on the QM[HF/3-21G(d)]/MM/NUCS potential energy surface. Both the reaction site (panel A) and the Mg^{2+} coordination sphere (panel B) are shown. The link atoms are shown as spheres with larger radius than usual hydrogen atoms.

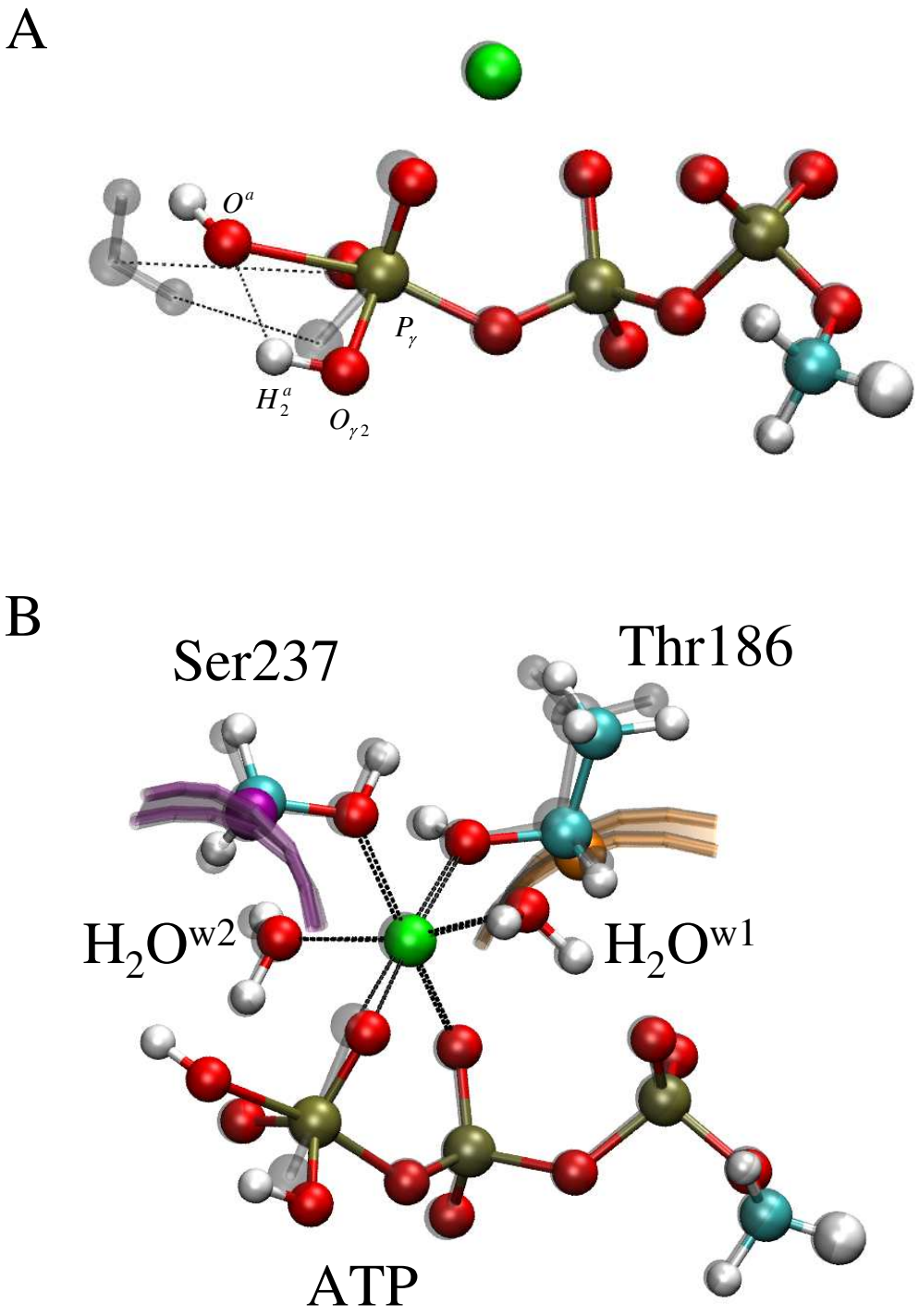

Figure 13.11: Overlap of the transition state (in color) of the direct path and the reactant state (in gray) as refined on the QM[HF/6-31G(d,p)]/MM/NUCS potential energy surface. Both the reaction site (panel A) and the Mg^{2+} coordination sphere (panel B) are shown. The link atoms are shown as spheres with larger radius than usual hydrogen atoms.

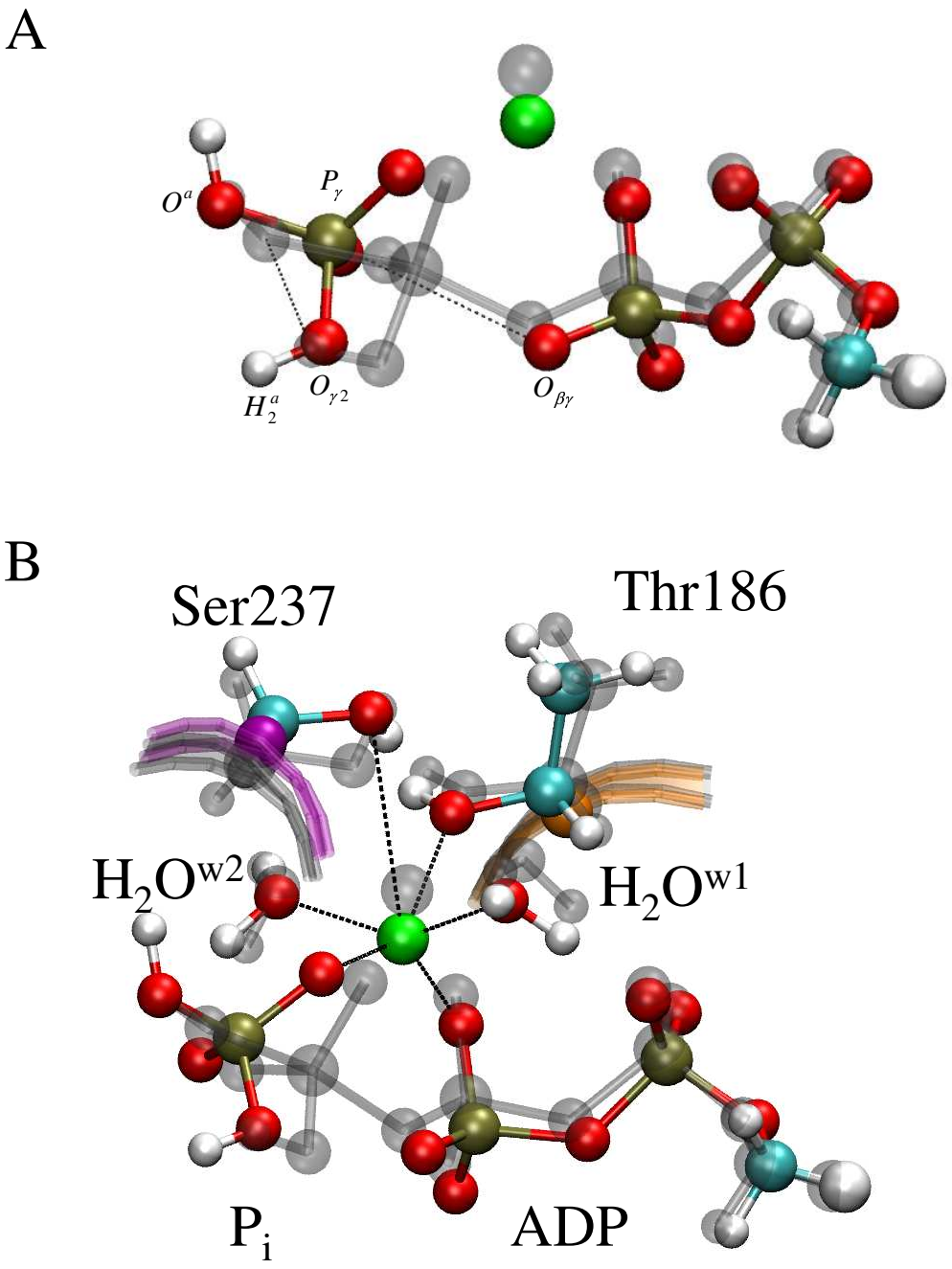

Figure 13.12: Overlap of the product state (in color) and the transition state (in gray) of the direct path as refined on the QM[HF/6-31G(d,p)]/MM/NUCS potential energy surface. Both the reaction site (panel A) and the Mg^{2+} coordination sphere (panel B) are shown. The link atoms are shown as spheres with larger radius than usual hydrogen atoms.

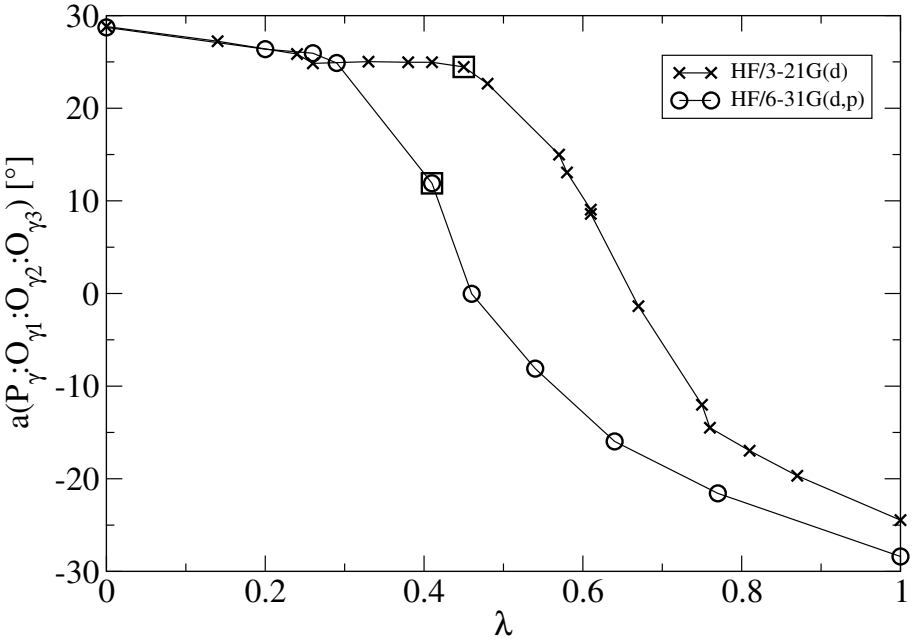

Figure 13.13: Improper angle a($P_\gamma : O_{\gamma 1} : O_{\gamma 2} : O_{\gamma 3}$) *versus* the reaction coordinate λ for the direct pathway.

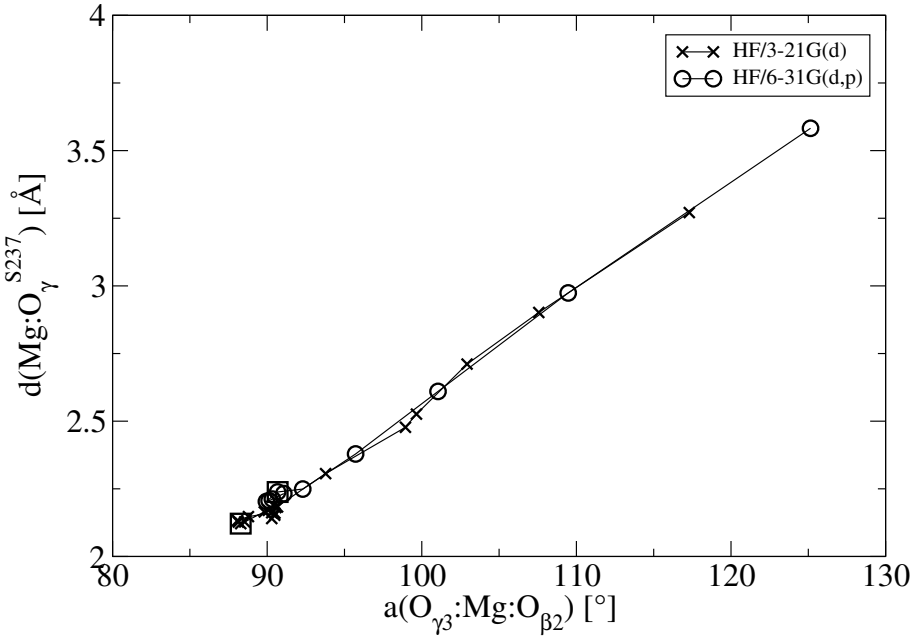

Figure 13.14: Angle a($O_{\gamma 3} : Mg : O_{\beta 2}$) *versus* the distance d($Mg : O_\gamma^{S237}$) for the direct pathway.

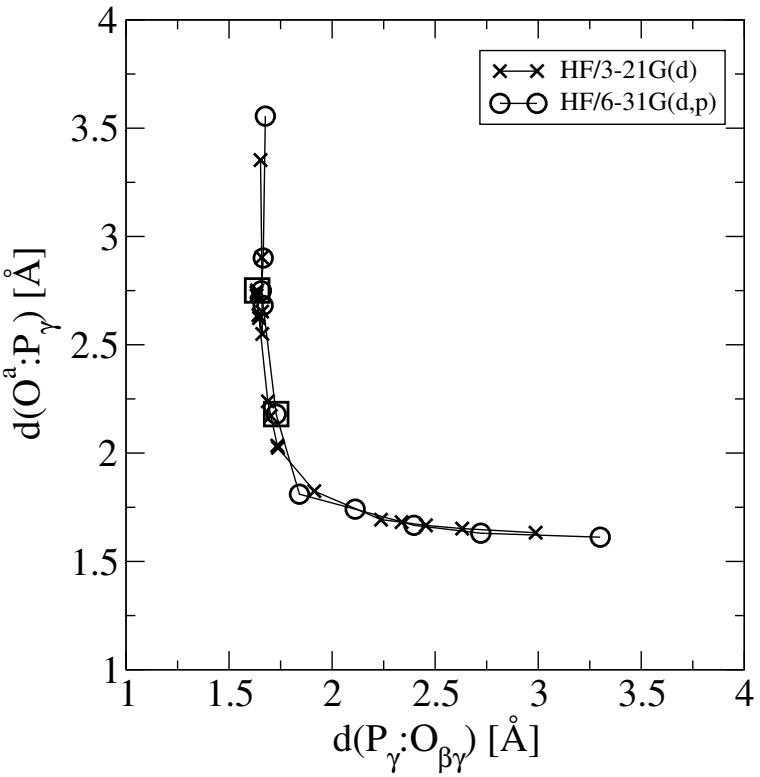

Figure 13.15: Distance between P_γ and the incoming oxygen $d(P_\gamma : O^a)$ *versus* the distance between P_γ and the β-γ-bridge oxygen $d(P_\gamma : O_{\beta\gamma})$ for the direct pathway.

13.1.3 Ser181 path

In the Ser181 path, one of the protons from the attacking water (H_2^a) is transferred to the oxygen of the helper water (O^h) that in turn transfers one of its hydrogens (H_2^h) to the sidechain oxygen of Ser181 (O_γ^{S181}). In a third proton transfer, the hydroxyl hydrogen of Ser181 (H_γ^{S181}) is transferred to the $O_{\gamma 2}$-oxygen of the γ-phosphate moiety of ATP. Tables 13.7 and 13.8 show the geometries of the linking sites in the transition and product states as refined on both the QM[HF/3-21G(d)]/MM/NUCS and QM[HF/6-31G(d,p)]/MM/NUCS potential energy surfaces. As in the reactant state and the stationary points on the direct path, the angle m_1:q_0:q_1 deviates from linearity and the distances between MM and QM boundary atoms are slightly larger than their respective values as parametrized in the CHARMM force field. Tables 13.9 and 13.10 summarize the distances and angles that characterize the transition and product states for the Ser181 path. The paths as refined on the two potential energy surfaces are described below.

QM[HF/3-21G(d)]/MM/NUCS. In the beginning of the path up to $\lambda = 0.29$, the attacking water reorients so as to be able to transfer its H_2^a proton onto the helper water. This proton transfer event is completed at $\lambda = 0.36$. Concertedly, the two other proton transfers, *i.e.*, the transfer of the helper water H_2^h proton onto the sidechain of Ser181 and the transfer of the H_γ^{S181} from the sidechain of Ser181 onto the $O_{\gamma 2}$ oxygen of ATP, occur between $\lambda = 0.19$ and $\lambda = 0.36$. An additional proton transfer occurs between $\lambda = 0.34$ and $\lambda = 0.46$, where the H_γ^{S236} proton is transferred from the sidechain of Ser236 onto the attacking water in a fashion similar to the analogous proton transfer event observed in the direct pathway on the QM[HF/3-21G(d)]/MM/NUCS surface.

At $\lambda = 0.46$ the transition state is reached. As shown in Fig. 13.16 the transition-state structure of ATP is close to its reactant-state structure. This situation is identical to that observed for the direct pathway on the QM[HF/3-21G(d)]/MM/NUCS surface.

Between $\lambda = 0.59$ and $\lambda = 0.83$ the H_2^a proton is transiently transferred from the helper water to the sidechain of Ser236, the H_2^h proton is backtransferred from the sidechain of Ser181 onto the helper water, and the H_γ^{S181} proton is backtransferred from the $O_{\gamma 2}$ oxygen of ATP onto the sidechain of Ser181. Subsequently, the H_γ^{S236} is backtransferred from the attacking water onto the sidechain of Ser236 between $\lambda = 0.83$ and $\lambda = 0.92$. Thus, at $\lambda = 0.90$ ($\lambda = 0.92$), the sidechain of Ser236 is doubly protonated with O:H bond distances of 1.14 (1.09) and 1.09 (1.19) Å.

The final proton transfers occur at the end of the path between $\lambda = 0.90$ and $\lambda = 1.0$. In this path region, the H_2^a proton is backtransferred from the sidechain of Ser236 onto the helper water, the H_2^h proton is transferred from the helper water onto the sidechain of Ser181, and the H_γ^{S181} proton is transferred from the sidechain of Ser181 onto the $O_{\gamma 2}$ oxygen of ATP.

At $\lambda = 0.83$ a transient state with distorted trigonal bipyramidal geometry is reached that is characterized by an improper angle a($P_\gamma : O_{\gamma 1} : O_{\gamma 2} : O_{\gamma 3}$) of -6.2° and distances

d($P_\gamma : O^a$) = 1.86 Å and d($P_\gamma : O_{\beta\gamma}$) = 2.3 Å. Up to $\lambda = 0.52$ the angle a($O_{\gamma 3} : Mg : O_{\beta 2}$) remains between 90° and 93° and the distance d($Mg : O_\gamma^{S237}$) stays at 2.1 Å. Between $\lambda = 0.59$ and $\lambda = 1.0$ the angle a($O_{\gamma 3} : Mg : O_{\beta 2}$) widens to its final value of 123.5° and the distance d($Mg : O_\gamma^{S237}$) increases to 3.3 Å (Fig. 13.17). As is the case in the direct path, the increase of the $Mg : O_\gamma^{S237}$ distance is due to the motions of both Mg^{2+} and O_γ^{S237}.

QM[HF/6-31G(d,p)]/MM/NUCS. In the beginning of the path up to $\lambda = 0.21$, the attacking water reorients so as to be able to transfer its H_2^a proton onto the helper water. In the same path segment ($\lambda = 0.10$ – $\lambda = 0.21$), Ser181 transfers its H_γ^{S181} proton onto the the $O_{\gamma 2}$ oxygen of ATP between $\lambda = 0.10$ and $\lambda = 0.21$. Directly after the H_γ^{S181} proton has been transferred it points towards the sidechain of Ser181, as expected. The second proton transfer event, in which the helper water transfers its H_2^h onto the sidechain of Ser181, occurs between $\lambda = 0.21$ and $\lambda = 0.25$. This proton remains oriented towards the oxygen, O^h, from which it originates, throughout the rest of the path. A third proton transfer occurs between $\lambda = 0.37$ and $\lambda = 0.46$. In this path segment the attacking water transfers its H_2^a onto the helper water oxygen, O^h.

Meanwhile, the $O_{\gamma 2}H_\gamma^{S181}$ group rotates such that the proton points towards the attacking water oxygen at $\lambda = 0.39$ and then towards the helper water oxygen, O_h, at $\lambda = 0.60$. Up to $\lambda \approx 0.70$, H_2^a keeps its orientation towards O^a, whereas the helper water reorients in the final part of the path such that in the product state H_2^a mediates a hydrogen bond between the helper water and the sidechain of Ser236. Between $\lambda = 0.66$ and $\lambda = 1.0$ the $O_{\gamma 2}H_\gamma^{S181}$ group rotates back in an almost 180° rotation, so that in the product state it mediates a hydrogen bond between $O_{\gamma 2}$ of the P_i moiety and $O_{\beta\gamma}$ of the ADP (See Fig. 13.19B). This hydrogen bond to the anionic ADP is more favorable than the hydrogen bond made by P_i in the product states of the direct path or of the Ser236 path (to the helper water, see Fig.13.12B and 13.26B). This is the reason why the product of the Ser181 path is lower in energy (Fig. 13.2).

The transition state at $\lambda = 0.60$ exhibits near trigonal bipyramidal geometry (Fig. 13.18). At $\lambda = 0.66$ the trigonal bipyramidal geometry is reached (improper dihedral angle a($P_\gamma : O_{\gamma 1} : O_{\gamma 2} : O_{\gamma 3}$) of zero with distances to the incoming and leaving oxygens of d($P_\gamma : O^a$) = 1.92 Å and d($P_\gamma : O_{\beta\gamma}$) = 1.88 Å. As observed in the direct path, up to the transition state at $\lambda = 0.60$, the angle a($O_{\gamma 3} : Mg : O_{\beta 2}$) remains below 93° and the distance d($Mg : O_\gamma^{S237}$) stays around 2.2 Å. During decay to the product (after $\lambda = 0.66$) the angle a($O_{\gamma 3} : Mg : O_{\beta 2}$) widens (final value 120°) and the distance d($Mg : O_\gamma^{S237}$) increases to 3.4 Å (Fig. 13.19), due to the simultaneous motion of Mg^{2+} and O_γ^{S237}. As in the direct path, this increase of the d($Mg : O_\gamma^{S237}$) distance has the effect of weakening the link between the Switch-1 loop and the Mg^{2+}-coordination sphere.

In contrast to the Ser181 path as refined on the QM[HF/3-21G(d)]/MM/NUCS surface, the three proton transfer event are stepwise in the QM[HF/6-31G(d,p)]/MM/NUCS-optimized path. Additional qualitative differences between the paths refined on the two surfaces

are that in the QM[HF/6-31G(d,p)]/MM/NUCS-optimized path Ser236 is not involved in proton transfers (as is the case for the QM[HF/3-21G(d)]/MM/NUCS-optimized path) and that no transient reversal of the proton transfers is observed in the QM[HF/6-31G(d,p)]/MM/NUCS-optimized path. As in the HF/6-31G(d,p)-optimized direct path, the attacking water ist activated before the resulting O^aH^- nucleophile attacks the γ-phosphate moiety of ATP, thus forming the rate-limiting transition state. The activation relay involving the helper water and Ser181 was described here as being sequential, as is indeed found in the CPR-optimized path. However, this implies that intermediate states separated by barriers corresponding to each proton transfer event should have been located as minima on the potential energy surface, which is not the case. Proton transfers here occur from a neutral species to an anionic species, thereby leading to an effective transport of a hydroxide anion through the myosin active site. The mechanism of this transport may be similar to the transport of hydrated hydroxide in bulk water that has been shown to be sequential *via* energy barriers of approximately 3 kcal/mol (396). Those small barriers may be unresolved in the present energy profile and therefore invisible.

As was the case for the direct paths, the Ser181 paths as refined on the two different potential energy surfaces share common characteristics. As shown in Fig. 13.20 both paths proceed *via* a transient state in which P_γ, $O_{\gamma 1}$, $O_{\gamma 2}$, and $O_{\gamma 3}$ lie in a plane. As in the direct paths, the transition state is earlier on the QM[HF/3-21G(d)]/MM/NUCS surface than on the the QM[HF/6-31G(d,p)]/MM/NUCS surface which in turn in earlier than the trigonal bipyramidal state. In both paths the distance d($Mg : O_\gamma^{S237}$) depends on the angle a($O_{\gamma 3} : Mg : O_{\beta 2}$) as shown in Fig. 13.21. The dependence is, however, not as clearly linear as was the case in the direct paths. As in the direct paths, both Ser181 paths follow the same associative route when plotting the distance d($P_\gamma : O^a$) *versus* the distance d($P_\gamma : O_{\beta\gamma}$) (Fig. 13.22). Remarkably, the transition state of the Ser181 path on the QM[HF/6-31G(d,p)]/MM/NUCS surface is characterized by the *identical* distances as found at the transition state of the direct path on the same potential energy surface (compare Figs. 13.22 and 13.15 and Tables 13.9 and 13.5), namely d($P_\gamma : O^a$) = 2.18 Å and d($P_\gamma : O_{\beta\gamma}$) = 1.73 Å.

	HF/3-21G(d)	HF/6-31G(d,p)
d($C_5'^{ATP} : QQH^{ATP}$)	1.092	1.095
d($C_5'^{ATP} : C_4'^{ATP}$)	1.521	1.522
a($C_5'^{ATP} : QQH^{ATP} : C_4'^{ATP}$)	132.218	131.385
a($C_4'^{ATP} : C_5'^{ATP} : O_{\alpha C}^{ATP}$)	112.769	112.714
d($C_\beta^{S181} : QQH^{S181}$)	1.079	1.088
d($C_\beta^{S181} : C_\alpha^{S181}$)	1.633	1.633
a($C_\beta^{S181} : QQH^{S181} : C_\alpha^{S181}$)	144.319	143.032
a($C_\alpha^{S181} : C_\beta^{S181} : O_\gamma^{S181}$)	100.014	103.152
d($C_\beta^{T186} : QQH^{T186}$)	1.079	1.083
d($C_\beta^{T186} : C_\alpha^{T186}$)	1.593	1.594
a($C_\beta^{T186} : QQH^{T186} : C_\alpha^{T186}$)	144.471	143.750
a($C_\alpha^{T186} : C_\beta^{T186} : O_\gamma^{T186}$)	102.616	103.306
d($C_\beta^{S236} : QQH^{S236}$)	1.091	1.085
d($C_\beta^{S236} : C_\alpha^{S236}$)	1.633	1.634
a($C_\beta^{S236} : QQH^{S236} : C_\alpha^{S236}$)	138.963	143.865
a($C_\alpha^{S236} : C_\beta^{S236} : O_\gamma^{S236}$)	103.962	103.913
d($C_\beta^{S237} : QQH^{S237}$)	1.072	1.077
d($C_\beta^{S237} : C_\alpha^{S237}$)	1.648	1.649
a($C_\beta^{S237} : QQH^{S237} : C_\alpha^{S237}$)	166.406	164.796
a($C_\alpha^{S237} : C_\beta^{S237} : O_\gamma^{S237}$)	105.229	105.000

Table 13.7: Distances d (in Å) and angles a (in °) characteristic of the QM/MM linking sites in the transition state (Ser181 path).

	HF/3-21G(d)	HF/6-31G(d,p)
d($C_5'^{ATP} : QQH^{ATP}$)	1.090	1.095
d($C_5'^{ATP} : C_4'^{ATP}$)	1.521	1.520
a($C_5'^{ATP} : QQH^{ATP} : C_4'^{ATP}$)	133.325	131.427
a($C_4'^{ATP} : C_5'^{ATP} : O_{\alpha C}^{ATP}$)	112.299	112.346
d($C_\beta^{S181} : QQH^{S181}$)	1.079	1.083
d($C_\beta^{S181} : C_\alpha^{S181}$)	1.631	1.635
a($C_\beta^{S181} : QQH^{S181} : C_\alpha^{S181}$)	142.800	145.435
a($C_\alpha^{S181} : C_\beta^{S181} : O_\gamma^{S181}$)	98.928	101.492
d($C_\beta^{T186} : QQH^{T186}$)	1.077	1.082
d($C_\beta^{T186} : C_\alpha^{T186}$)	1.604	1.602
a($C_\beta^{T186} : QQH^{T186} : C_\alpha^{T186}$)	145.655	144.510
a($C_\alpha^{T186} : C_\beta^{T186} : O_\gamma^{T186}$)	104.495	103.190
d($C_\beta^{S236} : QQH^{S236}$)	1.079	1.085
d($C_\beta^{S236} : C_\alpha^{S236}$)	1.648	1.636
a($C_\beta^{S236} : QQH^{S236} : C_\alpha^{S236}$)	154.448	143.273
a($C_\alpha^{S236} : C_\beta^{S236} : O_\gamma^{S236}$)	107.612	103.123
d($C_\beta^{S237} : QQH^{S237}$)	1.073	1.077
d($C_\beta^{S237} : C_\alpha^{S237}$)	1.658	1.655
a($C_\beta^{S237} : QQH^{S237} : C_\alpha^{S237}$)	165.722	166.946
a($C_\alpha^{S237} : C_\beta^{S237} : O_\gamma^{S237}$)	104.409	104.402

Table 13.8: Distances d (in Å) and angles a (in °) characteristic of the QM/MM linking sites in the product state (Ser181 path).

	HF/3-21G(d)	HF/6-31G(d,p)
d($P_\gamma : O^a$)	2.890	2.177
d($P_\gamma : O_{\gamma 1}$)	1.493	1.492
d($P_\gamma : O_{\gamma 2}$)	1.531	1.592
d($P_\gamma : O_{\gamma 3}$)	1.502	1.509
d($P_\gamma : O_{\beta\gamma}$)	1.654	1.731
d($P_\beta : O_{\beta\gamma}$)	1.577	1.569
d($P_\beta : O_{\beta 1}$)	1.482	1.488
d($P_\beta : O_{\beta 2}$)	1.500	1.501
a($P_\gamma : O_{\beta\gamma} : P_\beta$)	131.453	130.480
a($O_{\beta\gamma} : P_\gamma : O_{\gamma 1}$)	104.617	101.180
a($O_{\beta\gamma} : P_\gamma : O_{\gamma 2}$)	96.511	85.784
a($O_{\beta\gamma} : P_\gamma : O_{\gamma 3}$)	103.471	98.569
a($O_{\beta\gamma} : P_\gamma : O^a$)	166.505	163.236
a($O^a : P_\gamma : O_{\gamma 1}$	81.479	88.954
a($O^a : P_\gamma : O_{\gamma 2}$	70.139	77.530
a($O^a : P_\gamma : O_{\gamma 3}$	84.071	88.046
a($P_\gamma : O_{\gamma 1} : O_{\gamma 2} : O_{\gamma 3}$)	23.711	10.527
d($O^a : H_2^a$)	2.133	1.892
d($H_2^a : O_{\gamma 2}$)	3.215	2.875
d($H_2^a : O^h$)	0.982	0.959
d($O^h : H_2^h$)	1.501	1.916
d($H_2^h : O_\gamma^{S181}$)	1.012	0.950
d($O_\gamma^{S181} : H_\gamma^{S181}$)	1.457	2.894
d($H_\gamma^{S181} : O_{\gamma 2}$)	1.021	0.943
d($O^a : H_1^a$)	0.971	0.944
d($H_1^a : O_\gamma^{S236}$)	2.863	3.077
d($O_\gamma^{S236} : H_\gamma^{S236}$)	1.432	0.976
d($H_\gamma^{S236} : O_{\gamma 2}$)	2.851	3.024
d($Mg : O_{\gamma 3}$)	1.972	1.995
d($Mg : O_{\beta 2}$)	2.036	2.029
d($Mg : O^{w1}$)	2.035	2.092
d($Mg : O^{w2}$)	2.060	2.106
d($Mg : O_{\gamma 1}^{T186}$)	2.035	2.099
d($Mg : O_\gamma^{S237}$)	2.110	2.230
a($O_{\gamma 3} : Mg : O_{\beta 2}$)	92.576	93.254

Table 13.9: Distances d (in Å) and angles a (in °) characteristic of the QM region in the transition state (Ser181 path).

	HF/3-21G(d)	HF/6-31G(d,p)
d($P_\gamma : O^a$)	1.639	1.597
d($P_\gamma : O_{\gamma 1}$)	1.493	1.481
d($P_\gamma : O_{\gamma 2}$)	1.541	1.575
d($P_\gamma : O_{\gamma 3}$)	1.503	1.508
d($P_\gamma : O_{\beta\gamma}$)	2.863	3.244
d($P_\beta : O_{\beta\gamma}$)	1.495	1.506
d($P_\beta : O_{\beta 1}$)	1.497	1.497
d($P_\beta : O_{\beta 2}$)	1.539	1.530
a($P_\gamma : O_{\beta\gamma} : P_\beta$)	131.984	116.727
a($O_{\beta\gamma} : P_\gamma : O_{\gamma 1}$)	82.120	86.755
a($O_{\beta\gamma} : P_\gamma : O_{\gamma 2}$)	71.899	53.002
a($O_{\beta\gamma} : P_\gamma : O_{\gamma 3}$)	77.954	82.213
a($O_{\beta\gamma} : P_\gamma : O^a$)	167.667	151.273
a($O^a : P_\gamma : O_{\gamma 1}$	107.359	111.083
a($O^a : P_\gamma : O_{\gamma 2}$	96.277	98.435
a($O^a : P_\gamma : O_{\gamma 3}$	104.100	108.654
a($P_\gamma : O_{\gamma 1} : O_{\gamma 2} : O_{\gamma 3}$)	-25.444	-30.627
d($O^a : H_2^a$)	3.056	3.805
d($H_2^a : O_{\gamma 2}$)	3.338	3.585
d($H_2^a : O^h$)	0.993	0.949
d($O^h : H_2^h$)	1.577	1.975
d($H_2^h : O_\gamma^{S181}$)	0.999	0.946
d($O_\gamma^{S181} : H_\gamma^{S181}$)	1.522	3.299
d($H_\gamma^{S181} : O_{\gamma 2}$)	1.007	0.976
d($O^a : H_1^a$)	0.975	0.956
d($H_1^a : O_\gamma^{S236}$)	3.159	3.173
d($O_\gamma^{S236} : H_\gamma^{S236}$)	0.988	0.951
d($H_\gamma^{S236} : O_{\gamma 2}$)	2.442	2.104
d($Mg : O_{\gamma 3}$)	1.952	1.990
d($Mg : O_{\beta 2}$)	1.890	1.908
d($Mg : O^{w1}$)	1.983	2.050
d($Mg : O^{w2}$)	2.076	2.120
d($Mg : O_{\gamma 1}^{T186}$)	2.075	2.127
d($Mg : O_\gamma^{S237}$)	3.322	3.447
a($O_{\gamma 3} : Mg : O_{\beta 2}$)	123.469	119.764

Table 13.10: Distances d (in Å) and angles a (in °) characteristic of the QM region in the product state (Ser181 path).

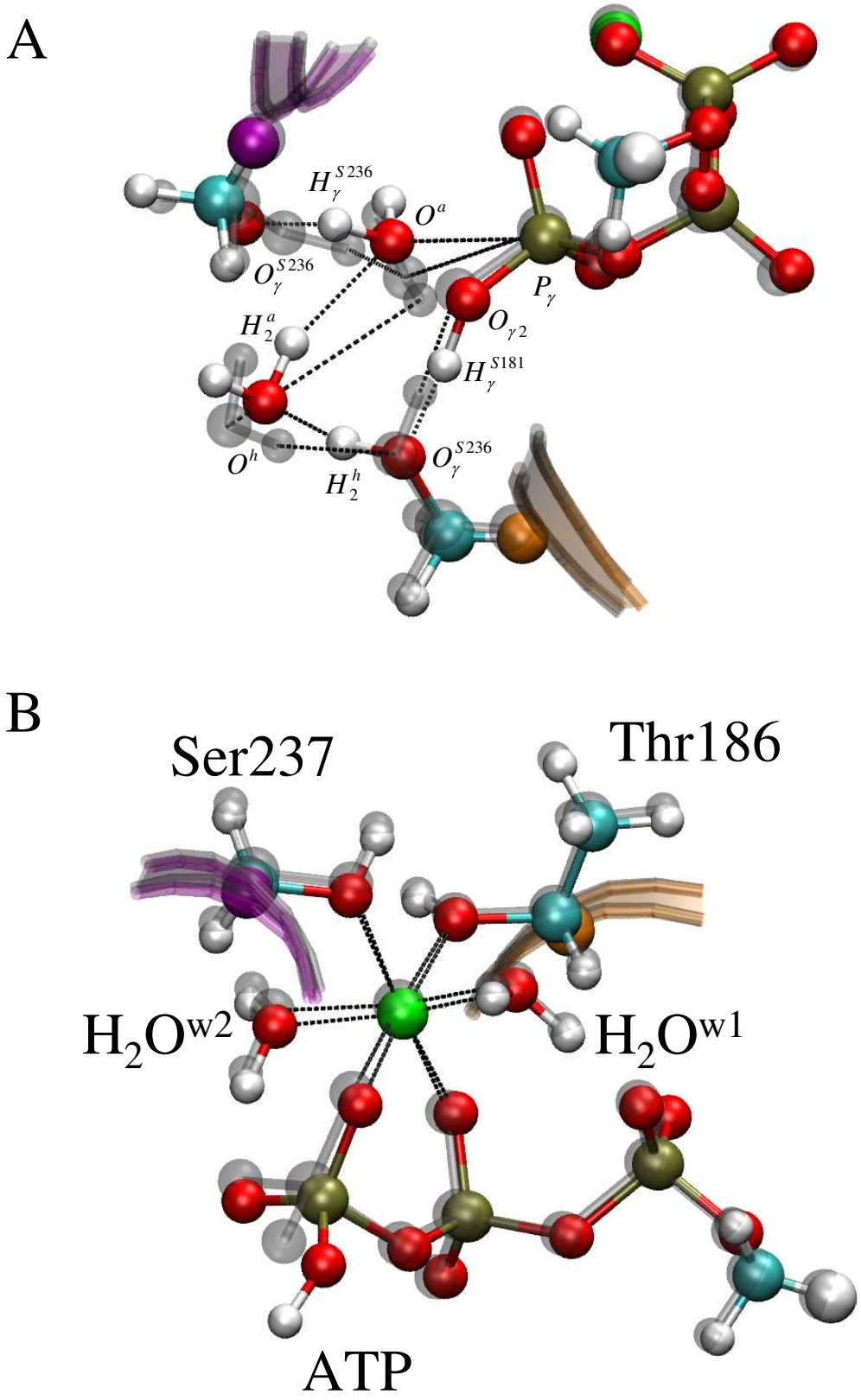

Figure 13.16: Overlap of the transition state (in color) of the Ser181 path and the reactant state (in gray) as refined on the QM[HF/3-21G(d)]/MM/NUCS potential energy surface. Both the reaction site (panel A) and the Mg^{2+} coordination sphere (panel B) are shown. The link atoms are shown as spheres with larger radius than usual hydrogen atoms.

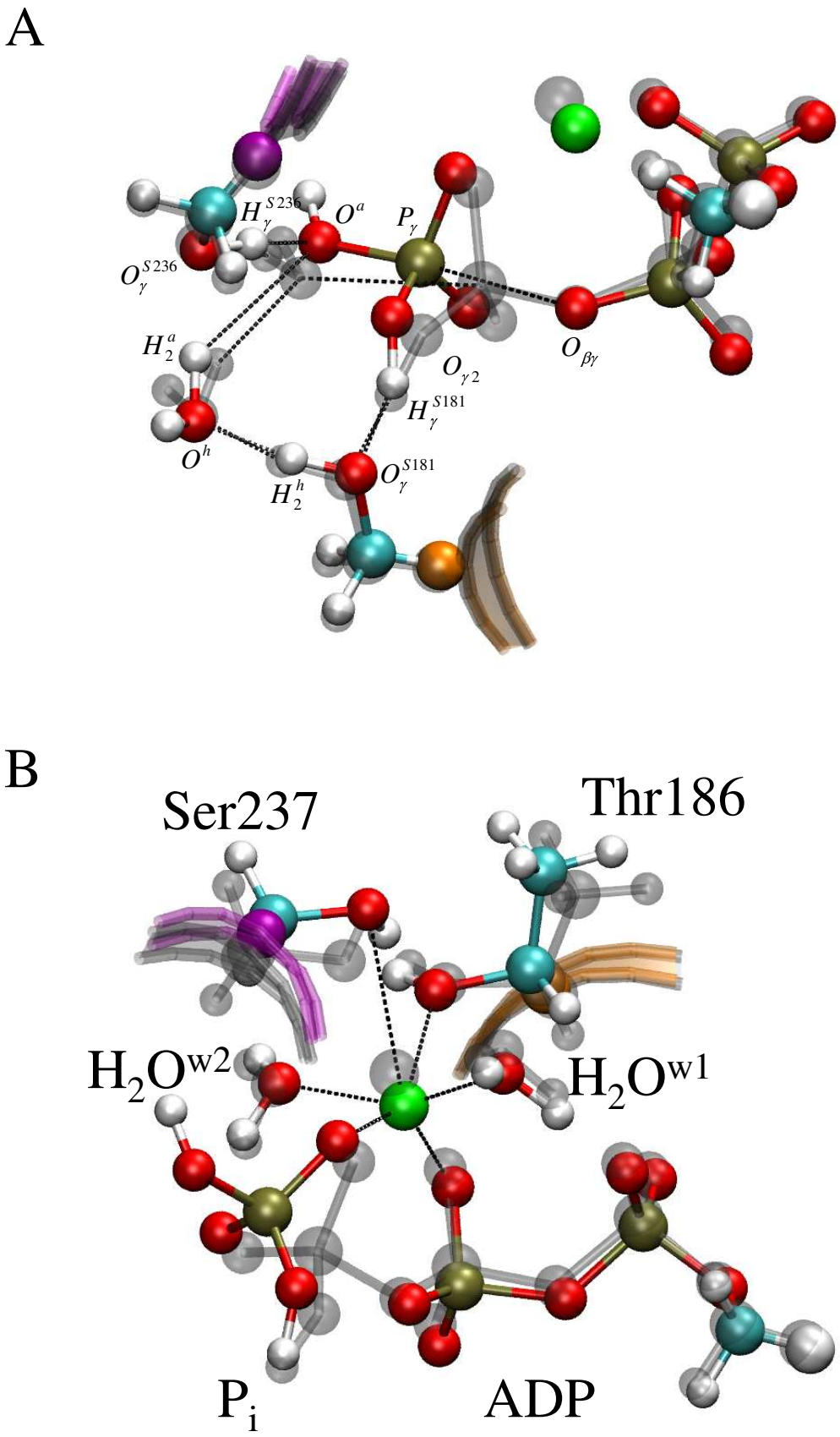

Figure 13.17: Overlap of the product state (in color) and the transition state (in gray) of the Ser181 path as refined on the QM[HF/3-21G(d)]/MM/NUCS potential energy surface. Both the reaction site (panel A) and the Mg^{2+} coordination sphere (panel B) are shown. The link atoms are shown as spheres with larger radius than usual hydrogen atoms.

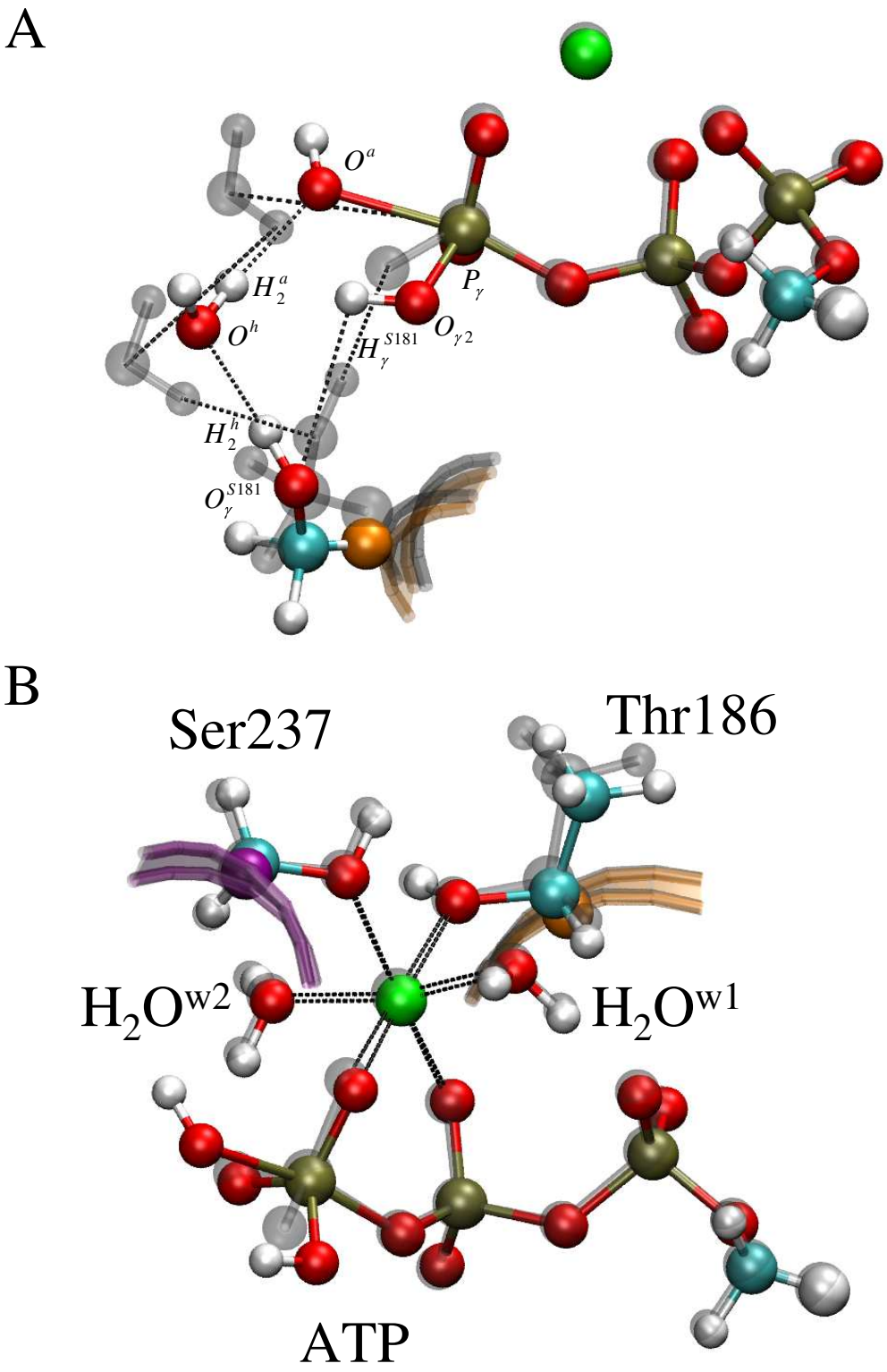

Figure 13.18: Overlap of the transition state (in color) of the Ser181 path and the reactant state (in gray) as refined on the QM[HF/6-31G(d,p)]/MM/NUCS potential energy surface. Both the reaction site (panel A) and the Mg^{2+} coordination sphere (panel B) are shown. The link atoms are shown as spheres with larger radius than usual hydrogen atoms.

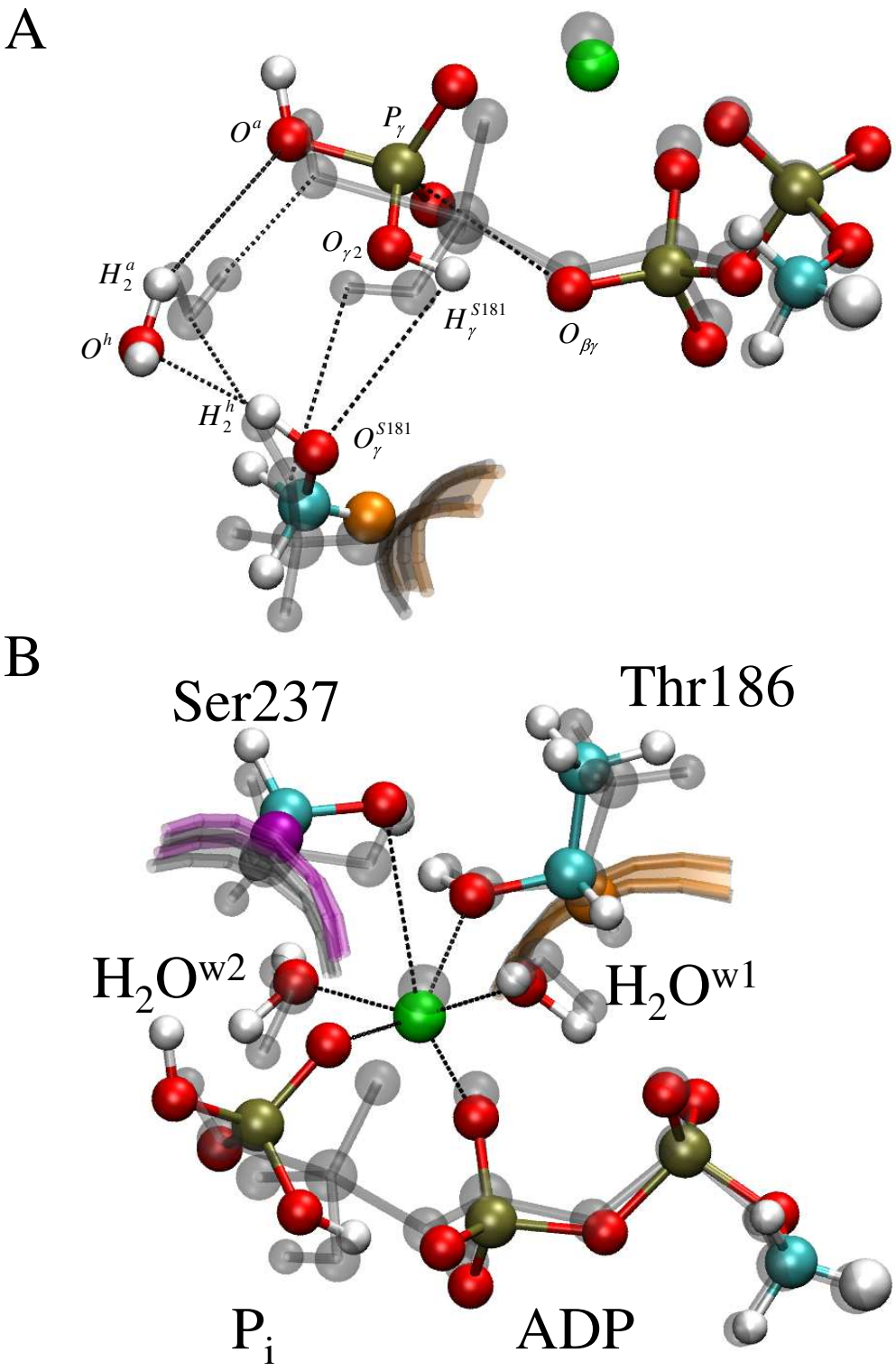

Figure 13.19: Overlap of the product state (in color) and the transition state (in gray) of the Ser181 path as refined on the QM[HF/6-31G(d,p)]/MM/NUCS potential energy surface. Both the reaction site (panel A) and the Mg^{2+} coordination sphere (panel B) are shown. The link atoms are shown as spheres with larger radius than usual hydrogen atoms.

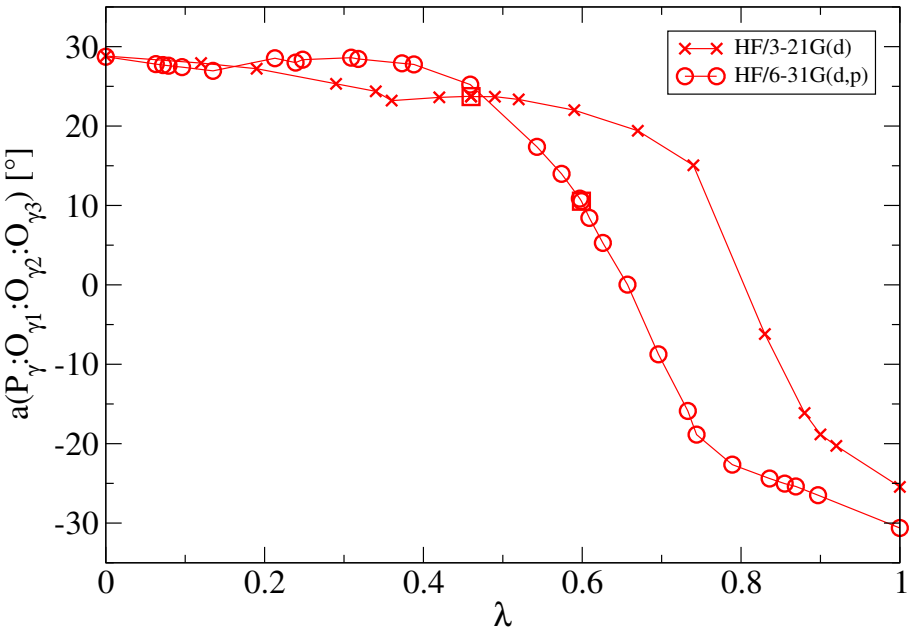

Figure 13.20: Improper angle a($P_\gamma : O_{\gamma 1} : O_{\gamma 2} : O_{\gamma 3}$) *versus* the reaction coordinate λ for the Ser181 pathway.

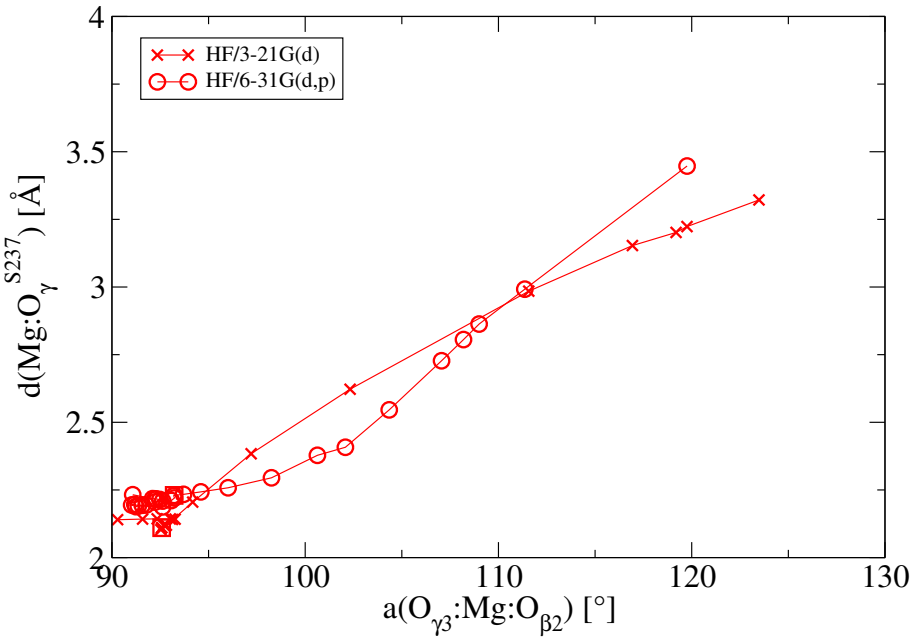

Figure 13.21: Angle a($O_{\gamma 3} : Mg : O_{\beta 2}$) *versus* the distance d($Mg : O_\gamma^{S237}$) for the Ser181 pathway.

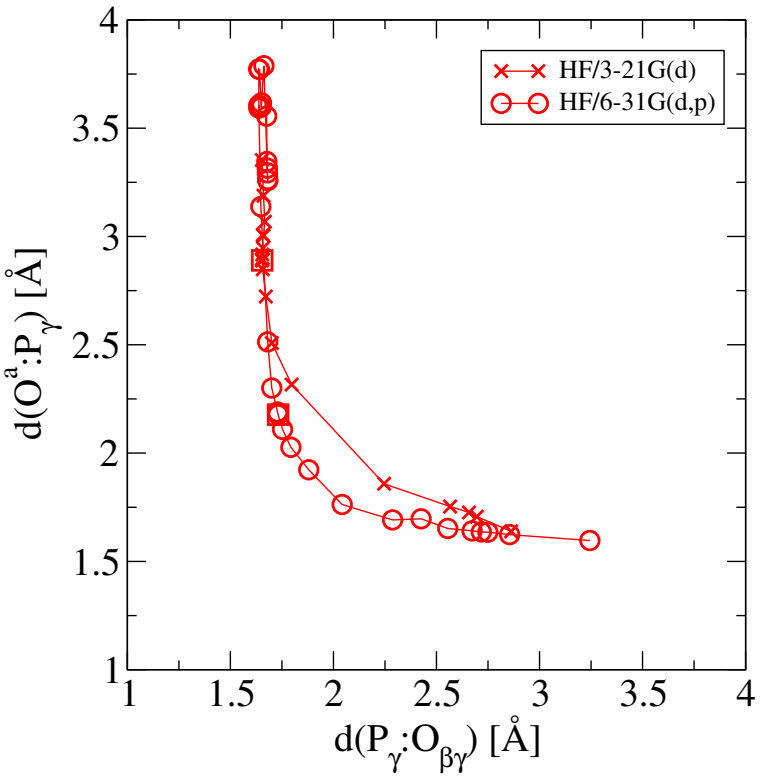

Figure 13.22: Distance between P_γ and the incoming oxygen $d(P_\gamma : O^a)$ *versus* the distance between P_γ and the β-γ-bridge oxygen $d(P_\gamma : O_{\beta\gamma})$ for the Ser181 pathway.

13.1.4 Ser236 path

In the Ser236 path, one of the protons from the attacking water (namely H_1^a) is transferred to the sidechain-oxygen of Ser236 (O_γ^{S236}). The Ser236 hydroxyl hydrogen H_γ^{S236} is transferred to the $O_{\gamma 2}$-oxygen from the γ-phosphate moiety of ATP. Tables 13.11 and 13.12 show the geometries of the linking sites in the transition and product states as refined on both the QM[HF/3-21G(d)]/MM/NUCS and QM[HF/6-31G(d,p)]/MM/NUCS potential energy surfaces. As in the reactant state and the stationary points on the direct and Ser181 paths, the angle m_1:q_0:q_1 deviates from linearity and the distances between MM and QM boundary atoms are larger than their respective distances as parametrized in the CHARMM force field. Tables 13.13 and 13.14 summarize the distances and angles that characterize the transition and product states for the Ser236 path. The paths as refined on the two potential energy surfaces are described below.

QM[HF/3-21G(d)]/MM/NUCS. Between $\lambda = 0.30$ and $\lambda = 0.52$ the H_γ^{S236} proton is transferred from the sidechain of Ser236 onto the $O_{\gamma 2}$ oxygen of ATP. At $\lambda = 0.62$ the transition state is reached. As shown in Fig. 13.23 the transition-state structure of ATP is close to its reactant-state structure as was also the case for the direct and Ser181 pathways on the QM[HF/3-21G(d)]/MM/NUCS surface. In the transition state, the attacking water carries both its protons. Thus, all three paths show an attacking water molecule in the transition state on the QM[HF/3-21G(d)]/MM/NUCS surface rather than an attacking negatively charged hydroxide ion. Between $\lambda = 0.62$ and $\lambda = 0.83$ the H_2^a proton is transferred from the attacking water onto the sidechain of Ser236. Thus, the two proton transfers occur stepwise in the Ser236 pathway on the QM[HF/3-21G(d)]/MM/NUCS surface unlike in the Ser181 pathway where several proton transfer events are concerted.

Only very close to the end of the path (at $\lambda = 0.83$) a transient state with trigonal bipyramidal geometry is reached that is characterized by an improper angle a($P_\gamma : O_{\gamma 1} : O_{\gamma 2} : O_{\gamma 3}$) of -2.2° and distances d($P_\gamma : O^a$) = 1.76 Å and d($P_\gamma : O_{\beta\gamma}$) = 1.93 Å. Up to this transient state the angle a($O_{\gamma 3} : Mg : O_{\beta 2}$) remains close to 90° and the distance d($Mg : O_\gamma^{S237}$) remains smaller than 2.2 Å. Upon decay of the trigonal bipyramidal transient state, the angle a($O_{\gamma 3} : Mg : O_{\beta 2}$) widens to its final value of 105.7° and the distance d($Mg : O_\gamma^{S237}$) increases to 2.5 Å (Fig. 13.24), with the increase of the $Mg : O_\gamma^{S237}$ distance resulting from the motions of both Mg and O_γ^{S237}, as was the case for the direct and Ser181 paths.

QM[HF/6-31G(d,p)]/MM/NUCS. The first event in the Ser236 path, between $\lambda = 0.00$ and $\lambda = 0.34$, is the proton transfer of the H_γ^{S236} proton from the sidechain of Ser236 onto the $O_{\gamma 2}$ oxygen of ATP. Simultaneously, the attacking water reorients into a position optimal for proton transfer of its H_1^a proton onto the sidechain of Ser236. This second proton transfer occurs between $\lambda = 0.39$ and $\lambda = 0.55$, where the transition state is reached. Thus, on the QM[HF/6-31G(d,p)]/MM/NUCS surface, the second proton transfer occurs *prior* to transition state formation, unlike in the Ser236 path as refined on the QM[HF/3-

21G(d,)]/MM/NUCS surface, where the attacking water still has both its protons bound in the transition state.

Unlike the transition state on the QM[HF/3-21G(d)]/MM/NUCS surface, the transition state on the QM[HF/6-31G(d,p)]/MM/NUCS surface is almost trigonal bipyramidal (Fig. 13.25). The trigonal bipyramidal structure between $\lambda = 0.61$ and $\lambda = 0.70$ has an improper angle a($P_\gamma : O_{\gamma 1} : O_{\gamma 2} : O_{\gamma 3}$) going from 5.9° to -5.4°, while the distances to the attacking oxygen d($P_\gamma : O^a$) decreases from 2.0 to 1.8 Å and the distance to the leaving oxygen d($P_\gamma : O_{\beta\gamma}$) increases from 1.8 to 2.1 Å, respectively. As was the case for the direct and Ser181 paths, the angle a($O_{\gamma 3} : Mg : O_{\beta 2}$) remains close to 90° and the distance d($Mg : O_\gamma^{S237}$) stays at 2.2 Å up to the transition state at $\lambda = 0.55$. Upon decay of the transition state, the angle a($O_{\gamma 3} : Mg : O_{\beta 2}$) widens to its final value of 122.1° and the distance d($Mg : O_\gamma^{S237}$) increases to 3.5 Å (Fig. 13.26), again breaking the link between switch-1 and the Mg^{2+}.

The Ser236 reaction paths as refined on the two different potential energy surfaces share common characteristics, as was the case for the Ser181 paths and the direct paths. As shown in Fig. 13.27 both paths proceed *via* a transient state in which P_γ, $O_{\gamma 1}$, $O_{\gamma 2}$, and $O_{\gamma 3}$ lie in a plane (as expected for both an associative and a dissociative mechanism). However, on both potential energy surfaces the transition state is earlier than the trigonal bipyramidal state, although already on the path segment corresponding to the configurational flip at the γ-phosphorus. Both paths exhibit linear dependence of the distance d($Mg : O_\gamma^{S237}$) *versus* the angle a($O_{\gamma 3} : Mg : O_{\beta 2}$) as shown in Fig. 13.28. This was also the case for the direct path with the slope found for the Ser236 path on the QM[HF/6-31G(d,p)]/MM/NUCS surface being identical to the slopes found for the direct path on both energy surfaces. In addition, both Ser236 paths follow the same route when plotting the distance d($P_\gamma : O^a$) *versus* the distance d($P_\gamma : O_{\beta\gamma}$) (Fig. 13.29). This route is identical to the route followed by the direct and Ser181 paths, *i.e.*, indicative of an associative mechanism. Remarkably, the transition state of the Ser236 path on the QM[HF/6-31G(d,p)]/MM/NUCS surface is characterized by the *identical* distances as found at the transition states of the direct and Ser181 paths on the same potential energy surface (compare Figs. 13.29, 13.22 and 13.15; and Tables 13.13, 13.9 and 13.5), namely d($P_\gamma : O^a$) = 2.18 Å and d($P_\gamma : O_{\beta\gamma}$) = 1.73 Å.

	HF/3-21G(d)	HF/6-31G(d,p)
$\mathrm{d}(C_5'^{ATP} : QQH^{ATP})$	1.092	1.095
$\mathrm{d}(C_5'^{ATP} : C_4'^{ATP})$	1.522	1.522
$\mathrm{a}(C_5'^{ATP} : QQH^{ATP} : C_4'^{ATP})$	131.070	131.141
$\mathrm{a}(C_4'^{ATP} : C_5'^{ATP} : O_{\alpha C}^{ATP})$	113.433	112.935
$\mathrm{d}(C_\beta^{S181} : QQH^{S181})$	1.084	1.087
$\mathrm{d}(C_\beta^{S181} : C_\alpha^{S181})$	1.632	1.637
$\mathrm{a}(C_\beta^{S181} : QQH^{S181} : C_\alpha^{S181})$	139.600	143.166
$\mathrm{a}(C_\alpha^{S181} : C_\beta^{S181} : O_\gamma^{S181})$	101.269	103.026
$\mathrm{d}(C_\beta^{T186} : QQH^{T186})$	1.078	1.082
$\mathrm{d}(C_\beta^{T186} : C_\alpha^{T186})$	1.590	1.592
$\mathrm{a}(C_\beta^{T186} : QQH^{T186} : C_\alpha^{T186})$	143.682	143.819
$\mathrm{a}(C_\alpha^{T186} : C_\beta^{T186} : O_\gamma^{T186})$	102.097	103.030
$\mathrm{d}(C_\beta^{S236} : QQH^{S236})$	1.089	1.086
$\mathrm{d}(C_\beta^{S236} : C_\alpha^{S236})$	1.633	1.638
$\mathrm{a}(C_\beta^{S236} : QQH^{S236} : C_\alpha^{S236})$	140.415	147.345
$\mathrm{a}(C_\alpha^{S236} : C_\beta^{S236} : O_\gamma^{S236})$	103.571	105.319
$\mathrm{d}(C_\beta^{S237} : QQH^{S237})$	1.071	1.077
$\mathrm{d}(C_\beta^{S237} : C_\alpha^{S237})$	1.650	1.650
$\mathrm{a}(C_\beta^{S237} : QQH^{S237} : C_\alpha^{S237})$	167.343	166.737
$\mathrm{a}(C_\alpha^{S237} : C_\beta^{S237} : O_\gamma^{S237})$	104.919	105.441

Table 13.11: Distances d (in Å) and angles a (in °) characteristic of the QM/MM linking sites in the transition state (Ser236 path).

	HF/3-21G(d)	HF/6-31G(d,p)
$\mathrm{d}(C_5'^{ATP} : QQH^{ATP})$	1.090	1.095
$\mathrm{d}(C_5'^{ATP} : C_4'^{ATP})$	1.523	1.521
$\mathrm{a}(C_5'^{ATP} : QQH^{ATP} : C_4'^{ATP})$	133.650	131.842
$\mathrm{a}(C_4'^{ATP} : C_5'^{ATP} : O_{\alpha C}^{ATP})$	112.578	112.260
$\mathrm{d}(C_\beta^{S181} : QQH^{S181})$	1.082	1.086
$\mathrm{d}(C_\beta^{S181} : C_\alpha^{S181})$	1.630	1.628
$\mathrm{a}(C_\beta^{S181} : QQH^{S181} : C_\alpha^{S181})$	141.107	140.007
$\mathrm{a}(C_\alpha^{S181} : C_\beta^{S181} : O_\gamma^{S181})$	101.812	101.702
$\mathrm{d}(C_\beta^{T186} : QQH^{T186})$	1.078	1.081
$\mathrm{d}(C_\beta^{T186} : C_\alpha^{T186})$	1.595	1.602
$\mathrm{a}(C_\beta^{T186} : QQH^{T186} : C_\alpha^{T186})$	145.143	144.392
$\mathrm{a}(C_\alpha^{T186} : C_\beta^{T186} : O_\gamma^{T186})$	102.485	103.447
$\mathrm{d}(C_\beta^{S236} : QQH^{S236})$	1.079	1.084
$\mathrm{d}(C_\beta^{S236} : C_\alpha^{S236})$	1.639	1.638
$\mathrm{a}(C_\beta^{S236} : QQH^{S236} : C_\alpha^{S236})$	149.803	145.593
$\mathrm{a}(C_\alpha^{S236} : C_\beta^{S236} : O_\gamma^{S236})$	106.045	104.411
$\mathrm{d}(C_\beta^{S237} : QQH^{S237})$	1.073	1.077
$\mathrm{d}(C_\beta^{S237} : C_\alpha^{S237})$	1.655	1.650
$\mathrm{a}(C_\beta^{S237} : QQH^{S237} : C_\alpha^{S237})$	166.250	164.227
$\mathrm{a}(C_\alpha^{S237} : C_\beta^{S237} : O_\gamma^{S237})$	103.197	103.058

Table 13.12: Distances d (in Å) and angles a (in °) characteristic of the QM/MM linking sites in the product state (Ser236 path).

	HF/3-21G(d)	HF/6-31G(d,p)
$\mathrm{d}(P_\gamma : O^a)$	2.686	2.178
$\mathrm{d}(P_\gamma : O_{\gamma 1})$	1.487	1.491
$\mathrm{d}(P_\gamma : O_{\gamma 2})$	1.576	1.604
$\mathrm{d}(P_\gamma : O_{\gamma 3})$	1.500	1.506
$\mathrm{d}(P_\gamma : O_{\beta\gamma})$	1.644	1.729
$\mathrm{d}(P_\beta : O_{\beta\gamma})$	1.604	1.580
$\mathrm{d}(P_\beta : O_{\beta 1})$	1.481	1.488
$\mathrm{d}(P_\beta : O_{\beta 2})$	1.500	1.499
$\mathrm{a}(P_\gamma : O_{\beta\gamma} : P_\beta)$	125.350	128.184
$\mathrm{a}(O_{\beta\gamma} : P_\gamma : O_{\gamma 1})$	105.999	101.831
$\mathrm{a}(O_{\beta\gamma} : P_\gamma : O_{\gamma 2})$	96.557	87.968
$\mathrm{a}(O_{\beta\gamma} : P_\gamma : O_{\gamma 3})$	102.945	98.177
$\mathrm{a}(O_{\beta\gamma} : P_\gamma : O^a)$	166.703	161.238
$\mathrm{a}(O^a : P_\gamma : O_{\gamma 1}$	83.093	90.540
$\mathrm{a}(O^a : P_\gamma : O_{\gamma 2}$	70.436	73.654
$\mathrm{a}(O^a : P_\gamma : O_{\gamma 3}$	80.615	87.541
$\mathrm{a}(P_\gamma : O_{\gamma 1} : O_{\gamma 2} : O_{\gamma 3})$	23.042	11.854
$\mathrm{d}(O^a : H_2^a)$	0.970	0.942
$\mathrm{d}(H_2^a : O_{\gamma 2})$	3.556	3.159
$\mathrm{d}(H_2^a : O^h)$	4.275	4.701
$\mathrm{d}(O^h : H_2^h)$	0.965	0.947
$\mathrm{d}(H_2^h : O_\gamma^{S181})$	2.040	2.127
$\mathrm{d}(O_\gamma^{S181} : H_\gamma^{S181})$	0.973	0.947
$\mathrm{d}(H_\gamma^{S181} : O_{\gamma 2})$	1.839	1.940
$\mathrm{d}(O^a : H_1^a)$	1.025	1.644
$\mathrm{d}(H_1^a : O_\gamma^{S236})$	1.470	0.978
$\mathrm{d}(O_\gamma^{S236} : H_\gamma^{S236})$	2.236	2.828
$\mathrm{d}(H_\gamma^{S236} : O_{\gamma 2})$	0.982	0.963
$\mathrm{d}(Mg : O_{\gamma 3})$	1.979	1.987
$\mathrm{d}(Mg : O_{\beta 2})$	2.038	2.042
$\mathrm{d}(Mg : O^{w1})$	2.026	2.093
$\mathrm{d}(Mg : O^{w2})$	2.052	2.102
$\mathrm{d}(Mg : O_{\gamma 1}^{T186})$	2.024	2.088
$\mathrm{d}(Mg : O_\gamma^{S237})$	2.134	2.238
$\mathrm{a}(O_{\gamma 3} : Mg : O_{\beta 2})$	88.901	90.710

Table 13.13: Distances d (in Å) and angles a (in °) characteristic of the QM region in the transition state (Ser236 path).

	HF/3-21G(d)	HF/6-31G(d,p)
d($P_\gamma : O^a$)	1.642	1.616
d($P_\gamma : O_{\gamma 1}$)	1.490	1.479
d($P_\gamma : O_{\gamma 2}$)	1.569	1.591
d($P_\gamma : O_{\gamma 3}$)	1.499	1.497
d($P_\gamma : O_{\beta\gamma}$)	2.846	3.279
d($P_\beta : O_{\beta\gamma}$)	1.497	1.495
d($P_\beta : O_{\beta 1}$)	1.497	1.501
d($P_\beta : O_{\beta 2}$)	1.538	1.536
a($P_\gamma : O_{\beta\gamma} : P_\beta$)	124.232	124.183
a($O_{\beta\gamma} : P_\gamma : O_{\gamma 1}$)	86.776	84.511
a($O_{\beta\gamma} : P_\gamma : O_{\gamma 2}$)	74.047	69.319
a($O_{\beta\gamma} : P_\gamma : O_{\gamma 3}$)	73.578	71.429
a($O_{\beta\gamma} : P_\gamma : O^a$)	164.208	162.004
a($O^a : P_\gamma : O_{\gamma 1}$	107.730	110.867
a($O^a : P_\gamma : O_{\gamma 2}$	93.643	95.102
a($O^a : P_\gamma : O_{\gamma 3}$	104.151	107.886
a($P_\gamma : O_{\gamma 1} : O_{\gamma 2} : O_{\gamma 3}$)	-22.521	-27.901
d($O^a : H_2^a$)	0.975	0.955
d($H_2^a : O_{\gamma 2}$)	3.286	3.260
d($H_2^a : O^h$)	4.120	4.537
d($O^h : H_2^h$)	0.975	0.949
d($H_2^h : O_\gamma^{S181}$)	1.849	2.045
d($O_\gamma^{S181} : H_\gamma^{S181}$)	0.978	0.949
d($H_\gamma^{S181} : O_{\gamma 2}$)	1.822	2.069
d($O^a : H_1^a$)	1.722	2.068
d($H_1^a : O_\gamma^{S236}$)	0.983	0.946
d($O_\gamma^{S236} : H_\gamma^{S236}$)	2.822	2.960
d($H_\gamma^{S236} : O_{\gamma 2}$)	1.013	0.955
d($Mg : O_{\gamma 3}$)	2.001	2.038
d($Mg : O_{\beta 2}$)	1.925	1.912
d($Mg : O^{w1}$)	1.981	2.027
d($Mg : O^{w2}$)	2.066	2.112
d($Mg : O_{\gamma 1}^{T186}$)	2.096	2.182
d($Mg : O_\gamma^{S237}$)	2.503	3.502
a($O_{\gamma 3} : Mg : O_{\beta 2}$)	105.666	122.112

Table 13.14: Distances d (in Å) and angles a (in °) characteristic of the QM region in the product state (Ser236 path).

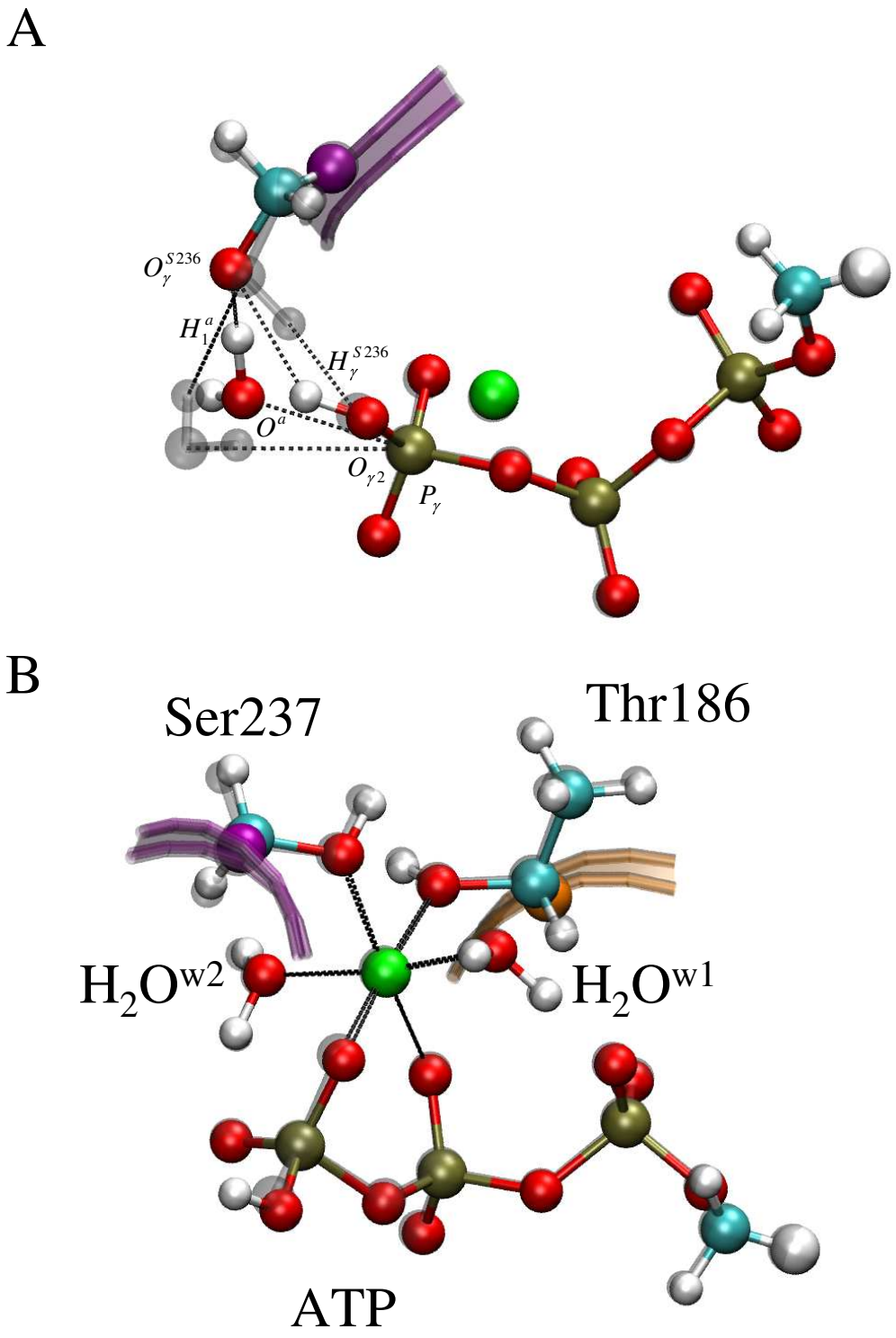

Figure 13.23: Overlap of the transition state (in color) of the Ser236 path and the reactant state (in gray) as refined on the QM[HF/3-21G(d)]/MM/NUCS potential energy surface. Both the reaction site (panel A) and the Mg^{2+} coordination sphere (panel B) are shown. The link atoms are shown as spheres with larger radius than usual hydrogen atoms.

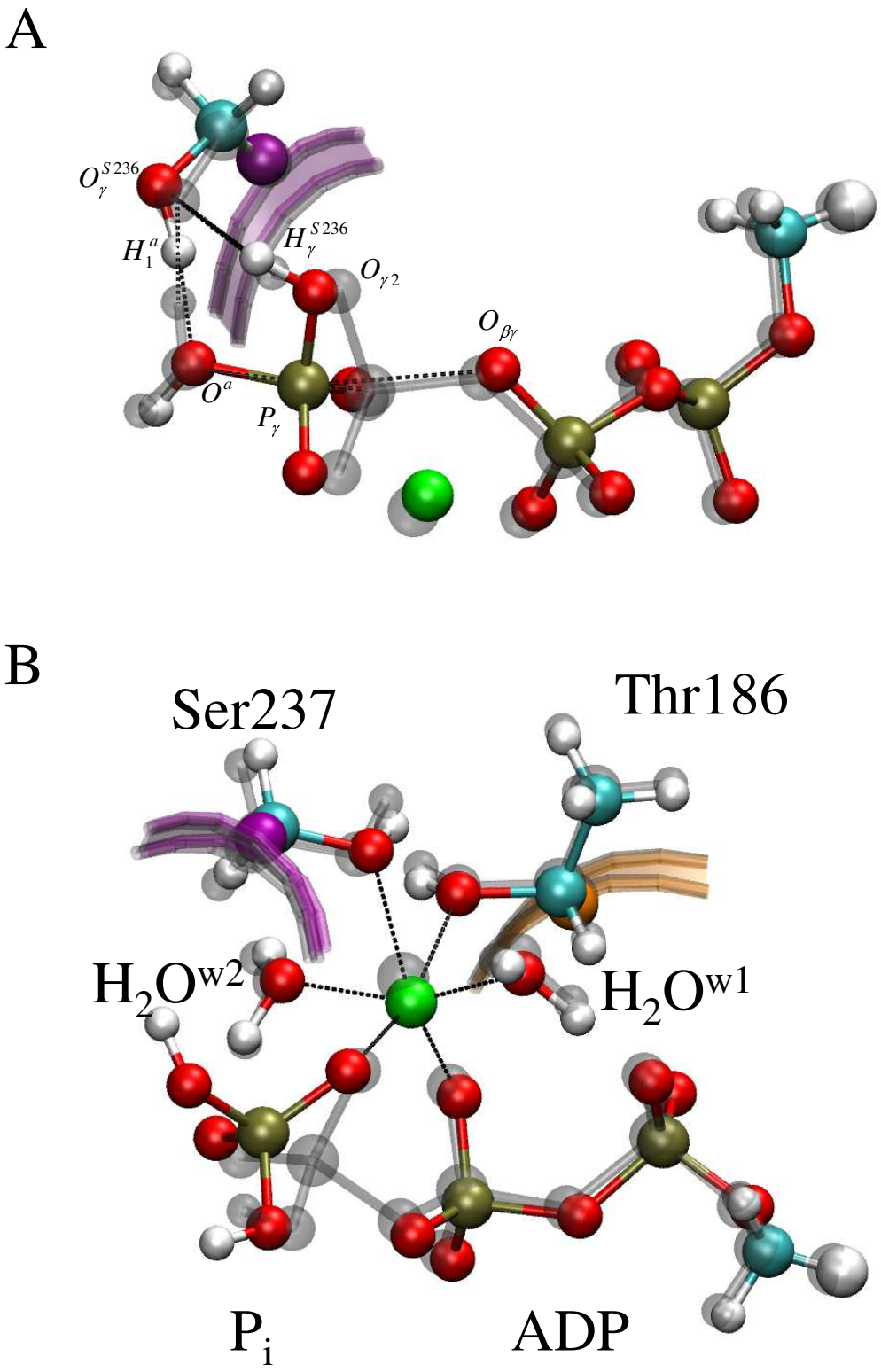

Figure 13.24: Overlap of the product state (in color) and the transition state (in gray) of the Ser236 path as refined on the QM[HF/3-21G(d)]/MM/NUCS potential energy surface. Both the reaction site (panel A) and the Mg^{2+} coordination sphere (panel B) are shown. The link atoms are shown as spheres with larger radius than usual hydrogen atoms.

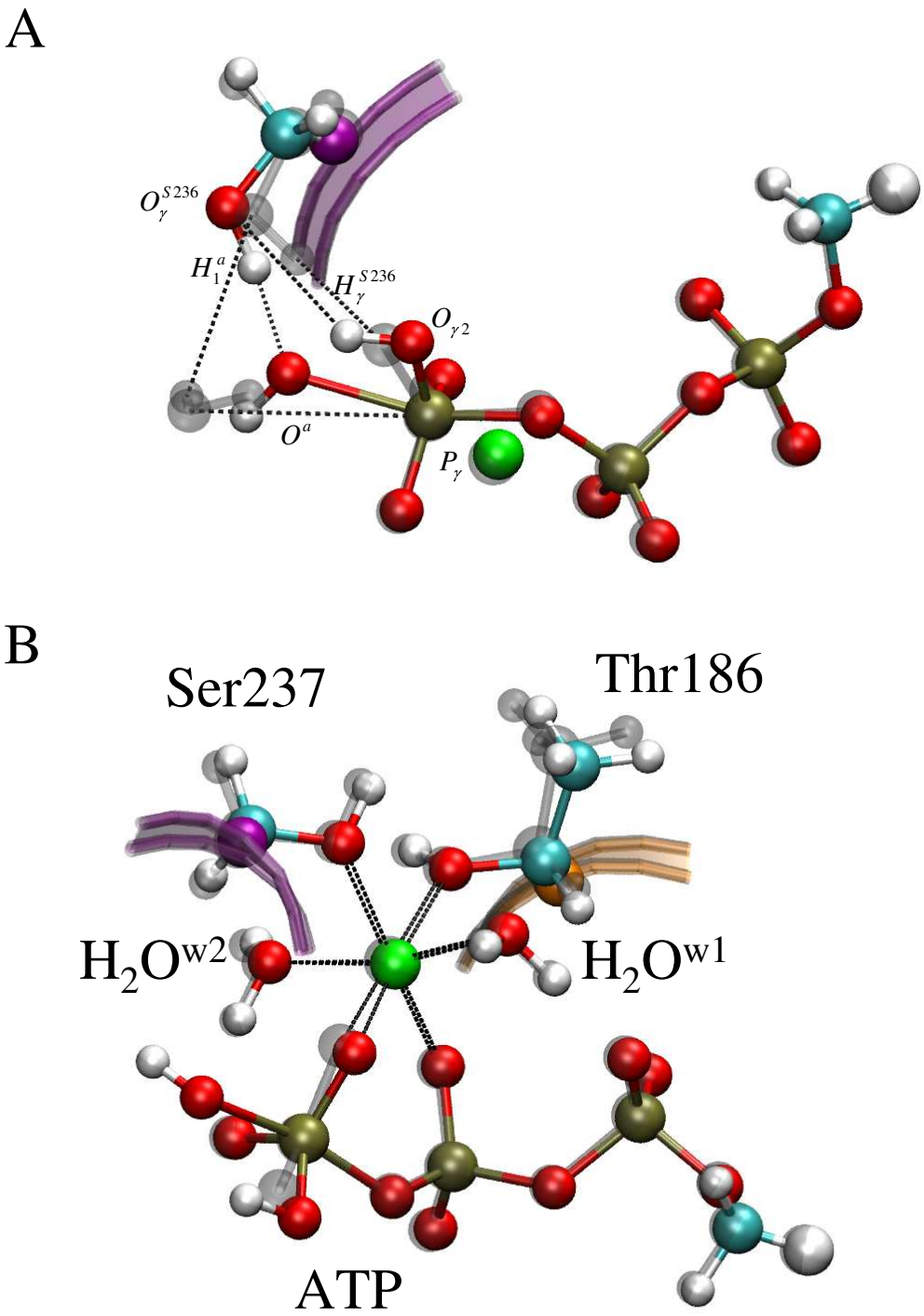

Figure 13.25: Overlap of the transition state (in color) of the Ser236 path and the reactant state (in gray) as refined on the QM[HF/6-31G(d,p)]/MM/NUCS potential energy surface. Both the reaction site (panel A) and the Mg^{2+} coordination sphere (panel B) are shown. The link atoms are shown as spheres with larger radius than usual hydrogen atoms.

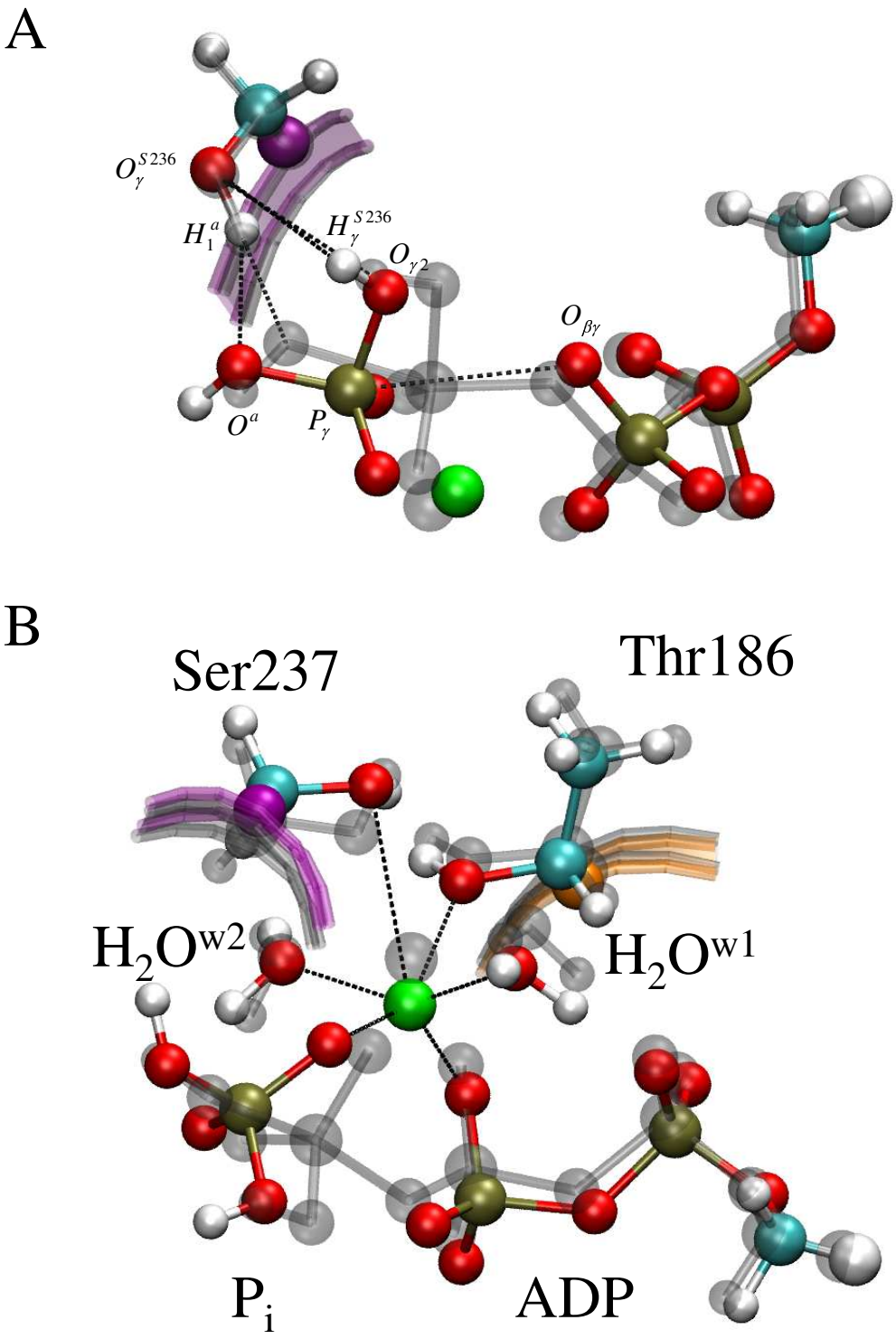

Figure 13.26: Overlap of the product state (in color) and the transition state (in gray) of the Ser236 path as refined on the QM[HF/6-31G(d,p)]/MM/NUCS potential energy surface. Both the reaction site (panel A) and the Mg^{2+} coordination sphere (panel B) are shown. The link atoms are shown as spheres with larger radius than usual hydrogen atoms.

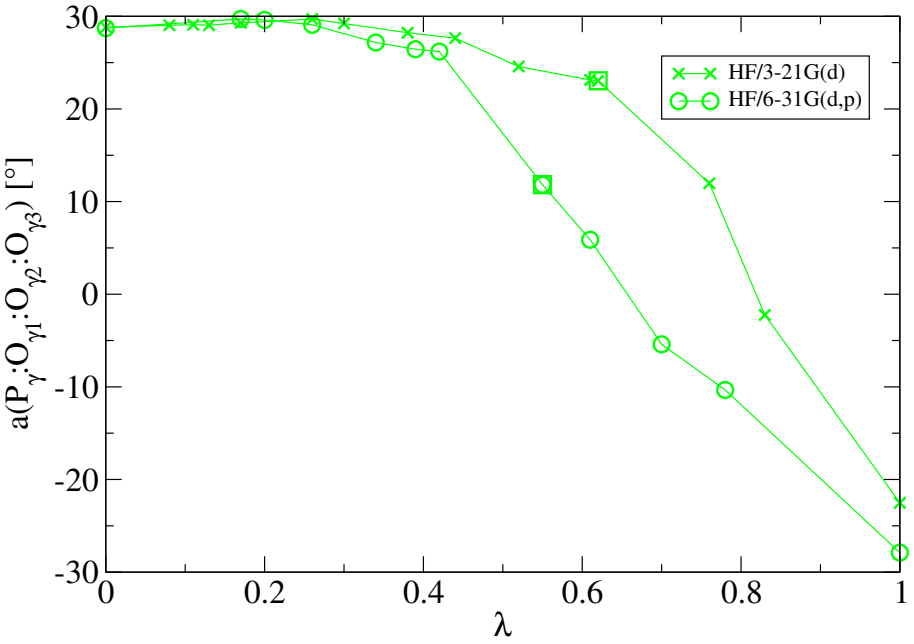

Figure 13.27: Improper angle a($P_\gamma : O_{\gamma1} : O_{\gamma2} : O_{\gamma3}$) *versus* the reaction coordinate λ for the Ser236 pathway.

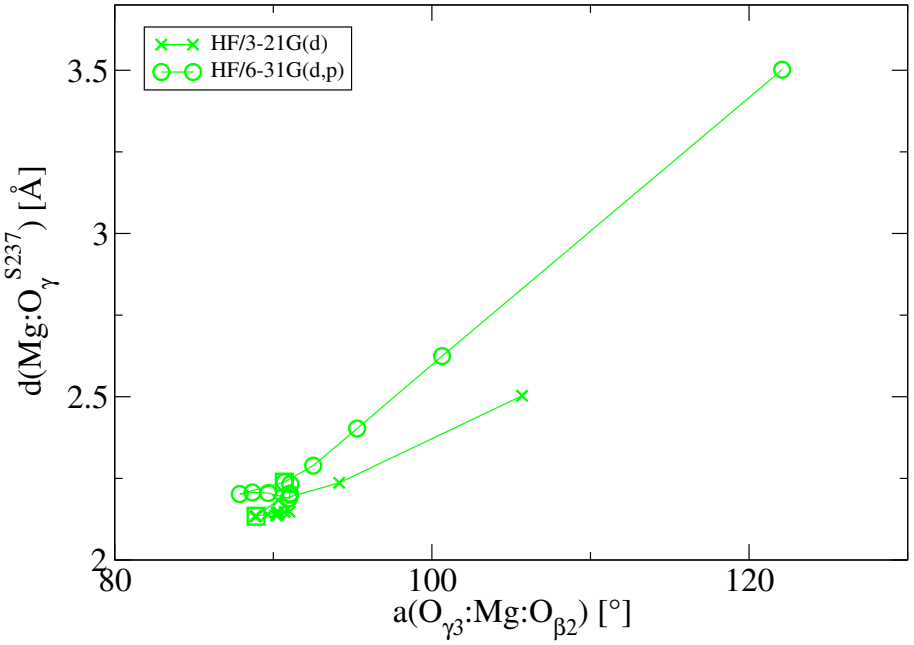

Figure 13.28: Angle a($O_{\gamma3} : Mg : O_{\beta2}$) *versus* the distance d($Mg : O_\gamma^{S237}$) for the Ser236 pathway.

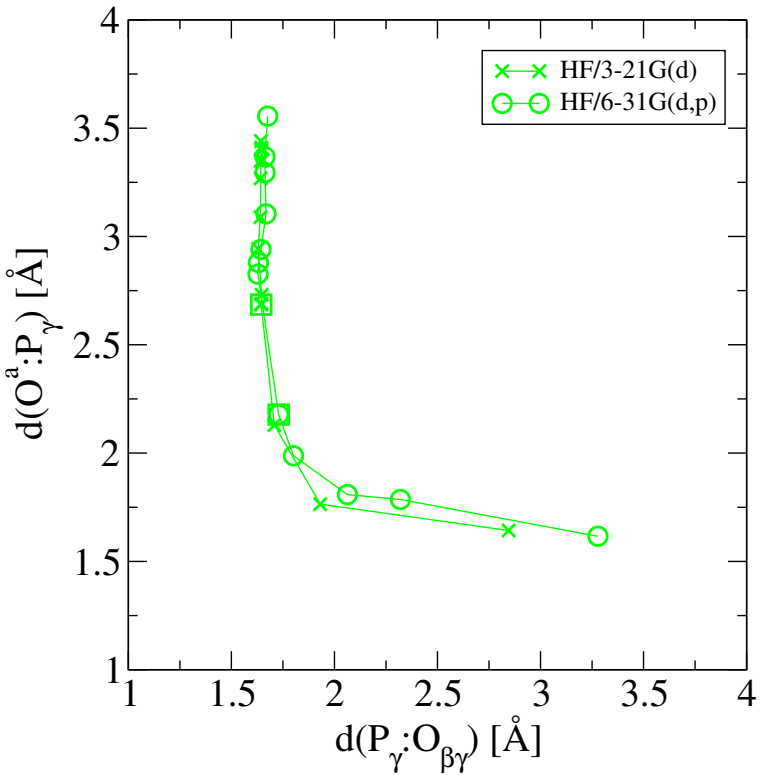

Figure 13.29: Distance between P_γ and the incoming oxygen $d(P_\gamma : O^a)$ *versus* the distance between P_γ and the β-γ-bridge oxygen $d(P_\gamma : O_{\beta\gamma})$ for the Ser236 pathway.

13.2 Energy profiles along the reaction paths

The path refinements have been made with Hartree-Fock calculations, thus no electron correlation effects were considered during reactant and product geometry optimizations and during path refinements. In addition, the basis sets used did not include diffuse function. It is therefore likely that the energetics along the paths is only approximately captured at this level of theory. To estimate the error in energies, single point energy calls were performed along the reaction coordinates of the three paths using density functional theory with the B3LYP exchange-correlation functional (265; 266; 267) and a 6-31+G(d,p) basis set. The numerical values of the reaction barrier heights and reaction energies for the three paths are collected in Table 13.15.

	direct	Ser181	Ser236
	barrier heights		
HF/3-21G(d)//HF3-21G(d)	26.07	29.03	28.05
B3LYP/6-31+G(d,p)//HF/3-21G(d)	31.10	31.41	31.34
HF/6-31G(d,p)//HF/6-31G(d,p)	50.90	45.60	51.02
B3LYP/6-31+G(d,p)//HF/6-31G(d,p)	41.08	38.04	41.28
	reaction energies		
HF/3-21G(d)//HF3-21G(d)	17.09	6.86	20.73
B3LYP/6-31+G(d,p)//HF/3-21G(d)	20.90	11.20	25.86
HF/6-31G(d,p)//HF/6-31G(d,p)	25.88	9.52	27.92
B3LYP/6-31+G(d,p)//HF/6-31G(d,p)	27.64	10.82	29.70

Table 13.15: Barrier heights and reaction energies in kcal/mol for the direct path, the Ser181 path, and the Ser236 path.

Remarkably, the barrier heights of the three paths as calculated on the identical energy surface are very close to each other, independent of the energy surface used. Thus, the three paths are isoenergetic and are therefore equally likely to be populated. However, the QM method greatly influences the barrier height. The highest energy barriers of 45.6 (Ser181 path) to 51.0 kcal/mol (Ser236 path) are found on the HF/6-31G(d,p)//HF/6-31G(d,p) surface, followed by 38.0 (Ser181 path) to 41.3 kcal/mol (Ser236 path) on the B3LYP/6-31+G(d,p)//HF/6-31G(d,p) and 31.1 (direct path) to 31.4 kcal/mol (Ser181 path) on the B3LYP/6-31+G(d,p)//HF/3-21G(d) surface. The lowest barrier heights of 26.1 (direct path) to 29.0 kcal/mol (Ser181 path) are found on the HF/3-21G(d)//HF/3-21G(d) surface. Thus, the barrier height varies by ∼25 kcal/mol depending on the calculation method chosen. Given the low accuracy of the quantum methods chosen, this is not surprising. However, it indicates strongly that higher-level quantum methods are required to obtain reliable energies. Compared to the experimental barrier of 14.5 kcal/mol (see Section 4.5) all calculated energy barriers are too high.

The complete energy profiles as obtained on the four different energy surfaces are shown in Figs. 13.30, 13.31, and 13.32 for the direct, the Ser181, and the Ser236 pathways, respectively. In all three paths, the shapes of the energy profiles on the HF/3-21G(d)//HF/3-21G(d) and B3LYP/6-31+G(d,p)//HF/3-21G(d) differ significantly, with the B3LYP/6-31+G(d,p)// HF/3-21G(d) energies being higher than the HF/3-21G(d)//HF/3-21G(d) throughout the whole paths. In particular, the saddle points on the HF/3-21G(d)//HF/3-21G(d) surface are no longer the highest-energy points on the B3LYP/6-31+G(d,p)//HF/3-21G(d) surface. In addition, for the Ser181 path a single energy maximum is found on the HF/3-21G(d)//HF/3-21G(d) surface, whereas two energy maxima are found on the B3LYP/6-31+G(d,p)//HF/3-21G(d) surface. This indicates that the QM[HF/3-21G(d)]/MM/NUCS geometries are different from geometries as would be found when using more reliable higher-level methods. Thus, the HF/3-21G(d) method is not sufficiently accurate to describe the geometries along the ATP hydrolysis reaction pathway in myosin.

This situation is different when comparing the shapes of the HF/6-31G(d,p)//HF/6-31G(d,p) and B3LYP/6-31+G(d,p)//HF/6-31G(d) energy profiles. Here, the B3LYP/6-31+G(d,p)// HF/6-31G(d) energies are in general smaller than the HF/6-31G(d,p)//HF/6-31G(d,p) energies. The shapes of the paths are reasonably conserved. In addition, on the B3LYP/6-31+G(d,p)//HF/6-31G(d) surface, the energy of the saddle point deviates by no more than 5 kcal/mol from the highest-energy point. Thus, the saddle points remain the highest-energy points within the error of the method. This indicates that the HF/6-31G(d,p)-refined geometries are more reliable than the HF/3-21G(d) geometries.

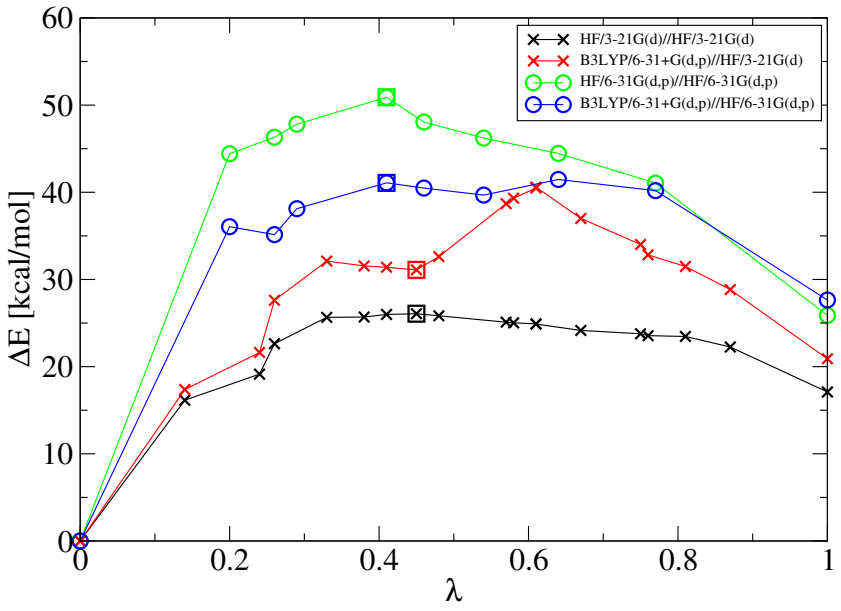

Figure 13.30: Total energies in kcal/mol along the reaction coordinate for the direct pathway. The transition states are indicated by squares.

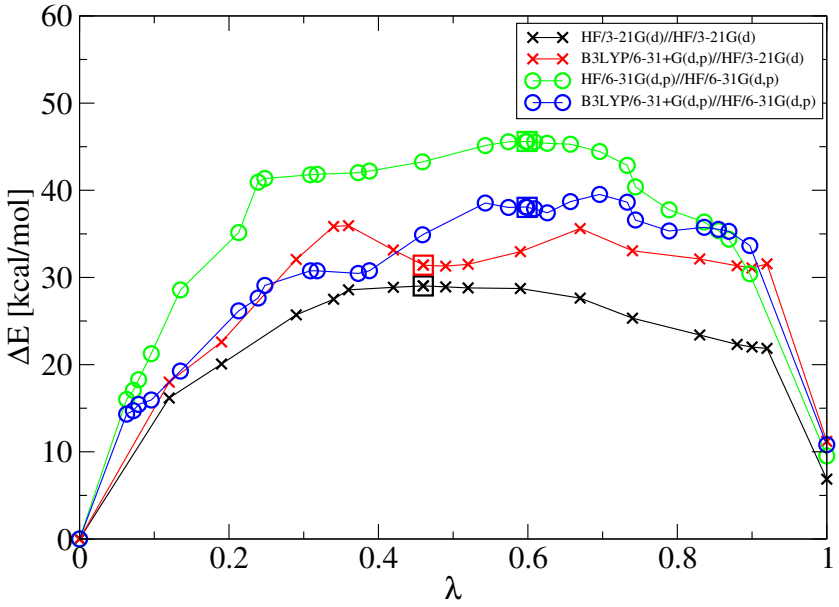

Figure 13.31: Total energies in kcal/mol along the reaction coordinate for the Ser181 pathway. The transition states are indicated by squares.

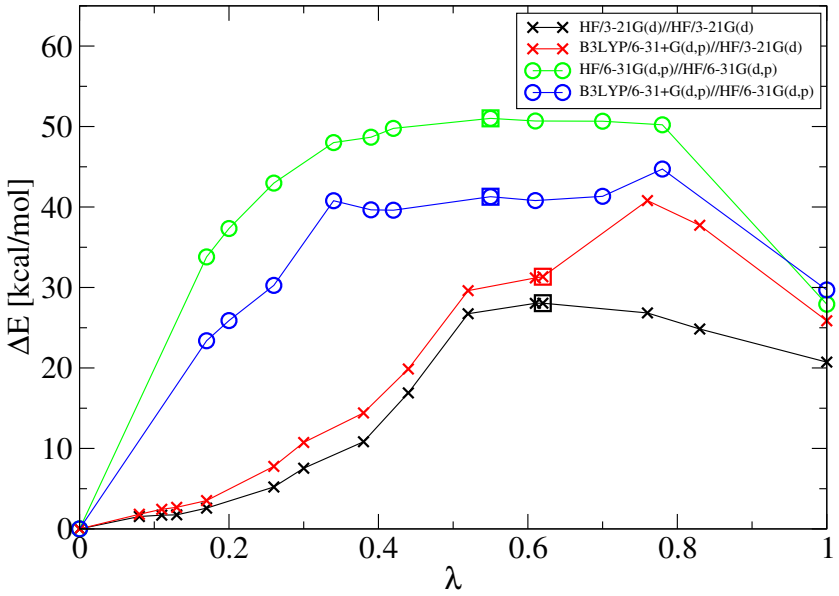

Figure 13.32: Total energies in kcal/mol along the reaction coordinate for the Ser236 pathway. The transition states are indicated by squares.

13.3 Energy decomposition along the reaction paths

Further insight into the energetics along the reaction paths is gained by dissecting the total energies as discussed in Section 13.2 into their contributing terms. These are the bonded energy terms, the van der Waals energy terms, the MM electrostatic interaction terms, the constraint energy terms, and the quantum energy terms. The QM energy includes both the quantum mechanical energy of the QM atoms and the QM/MM electrostatic interactions. Figs. 13.33, 13.34, and 13.35 show the resulting dissections for the three paths as refined on the QM[HF/3-21G(d)]/MM/NUCS energy surface. The corresponding plots for the paths as refined on the QM[HF/6-31(d,p)]/MM/NUCS energy surface are shown in Figs. 13.36, 13.37, and 13.38.

In all cases, the shape of the total energy profile is dominated by the shape of the QM energy profile. This energy term also reflects the insufficient accuracy of the QM[HF/3-21G(d)]/MM/NUCS potential energy surface in that the highest-energy points on the HF/3-21G(d)//HF/3-21G(d) and the B3LYP/6-31+G(d,p)//HF/3-21G(d) surfaces do not coincide. In contrast, the shapes of the QM energy profiles on the HF/6-31G(d,p)//HF/6-31G(d,p) and B3LYP/6-31+G(d,p)//HF/6-31G(d,p) surfaces are conserved, thus confirming the finding that the QM[HF/6-31G(d,p)]/MM/NUCS-optimized geometries are likely to be reliable.

The total energy of the transition states is clearly dominated by the QM energy in all cases. For the paths as refined on the QM[HF/3-21G(d)]/MM/NUCS surface, none of the bonded, van der Waals, MM electrostatics, and constraint terms contributes more than 3 kcal/mol (absolute energies) to the total energies of the transition states, as compared to 22 to 31 kcal/mol that are contributed by the QM energies. Thus, the MM energy terms constitute only minor contributions to the energies of the transition states in this case. For the paths as refined on the QM[HF/6-31G(d,p)]/MM/NUCS surface, the QM energies contribute 32 to 43 kcal/mol to the total energies of the transition states. Here, for the direct and Ser236 paths the largest contributions of the MM energy terms comes from the MM electrostatics that contributes 5.6 to 6.1 kcal/mol to the total energies of the transition states. For the Ser181 path, the largest MM contribution to the transition state energy comes from the van der Waals term, which contributes 3.3 kcal/mol.

However, at the product states the total energies are sums of terms of comparable magnitude. For the direct path, the QM energies are favorable on the HF/3-21G(d)//HF/3-21G(d) (-4.3 kcal/mol) and B3LYP/6-31+G(d,p)//HF/3-21G(d) surfaces (-0.5 kcal/mol), whereas they become slightly unfavorable on the HF/6-31G(d,p)//HF/6-31G(d,p) (0.3 kcal/mol) and B3LYP/6-31+G(d,p)//HF/6-31G(d) surfaces (2.0 kcal/mol). The largest contributions come from the MM electrostatics and van der Waals terms that are unfavorable by 9.3 and 7.2 kcal/mol (13.9 and 8.1 kcal/mol) for the QM[HF/3-21G(d)]/MM/NUCS-optimized (QM[HF/6-31G(d,p)]/MM/NUCS-optimized) product state, respectively. Thus, the major energy terms responsible for the large unfavorable reaction energy of the direct

path (*i.e.*, 17.1 to 27.6 kcal/mol, depending on the energy surface, see also Table 13.15) are found in the MM terms rather than the QM terms. A similar situation is observed for the Ser236 path, in which the reaction energies are unfavorable by 20.7 to 29.7 kcal/mol (see Table 13.15). The value of the reaction energies is the sum of the QM energies that are unfavorable by 2.8 to 7.9 kcal/mol and the MM electrostatics and van der Waals terms that are unfavorable by 6.2 and 6.5 kcal/mol (14.4 and 7.5 kcal/mol) for the QM[HF/3-21G(d)]/MM/NUCS-optimized (QM[HF/6-31G(d,p)]/MM/NUCS-optimized) product structure. In both the direct and Ser236 paths, the bonded and constraint energy terms are also unfavorable, however, they contribute less than 3 kcal/mol each.

For the Ser181 path, the reaction energies of 6.9 to 11.2 kcal/mol (see Table 13.15) are not as unfavorable as in the direct and Ser236 paths. This is due to almost all energy terms becoming less unfavorable or even favorable. In particular, the QM energies in the product states are clearly favorable by -8.8 to -14.3 kcal/mol, due to formation of the hydrogen bond between P_i and ADP that is not formed in the product states of the direct and the Ser236 paths. The MM electrostatics and van der Waals terms remain unfavorable by 5.8 and 7.8 kcal/mol (7.8 and 7.8 kcal/mol) for the QM[HF/3-21G(d)]/MM/NUCS-optimized (QM[HF/6-31G(d,p)]/MM/NUCS-optimized) product structure.

The constraint energy terms stay close to zero throughout the whole paths in all cases. This indicates that the constrained region is not significantly distorted throughout the course of the reaction. Thus, all protein motions that are necessary to allow the reaction to proceed are captured in the flexible region. This shows that the size of flexible region has been chosen large enough to accommodate local structural rearrangements that accompany ATP hydrolysis in myosin. The highest constraint energy is found for the product state of the Ser181 path (4.1 kcal/mol, in the QM[HF/3-21G(d)]/MM/NUCS-optimized structure), *i.e.*, for the state of lowest QM energy. This suggests that conformational changes in the constrained region may be necessary in later stages of the hydrolysis reaction to accommodate a product geometry that may be characterized by a larger distance between ADP and P_i and by larger distortions of the Mg^{2+} coordination sphere than observed in the present product states.

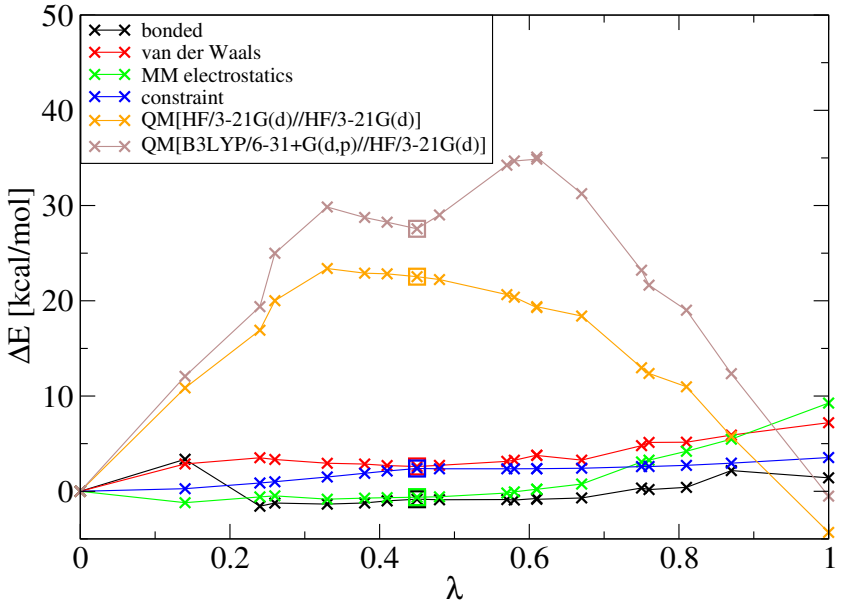

Figure 13.33: Dissection of the total energy into its contributing terms (in kcal/mol) for the direct pathway as refined on the QM[HF/3-21G(d)]/MM/NUCS energy surface.

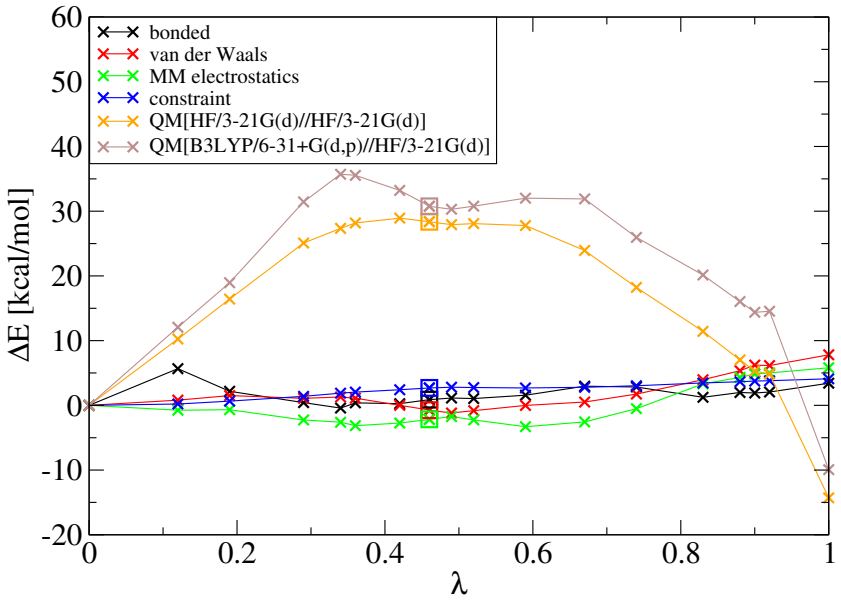

Figure 13.34: Dissection of the total energy into its contributing terms (in kcal/mol) for the Ser181 pathway as refined on the QM[HF/3-21G(d)]/MM/NUCS energy surface.

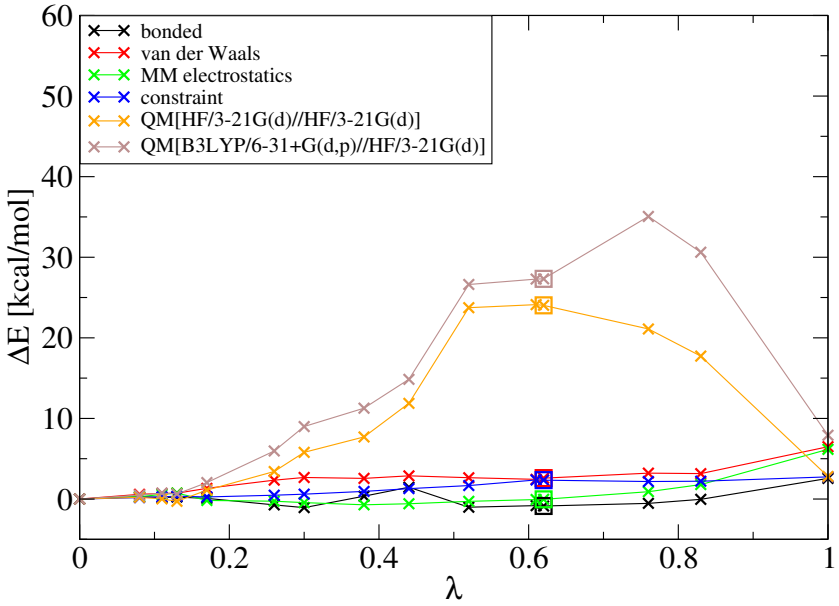

Figure 13.35: Dissection of the total energy into its contributing terms (in kcal/mol) for the Ser236 pathway as refined on the QM[HF/3-21G(d)]/MM/NUCS energy surface.

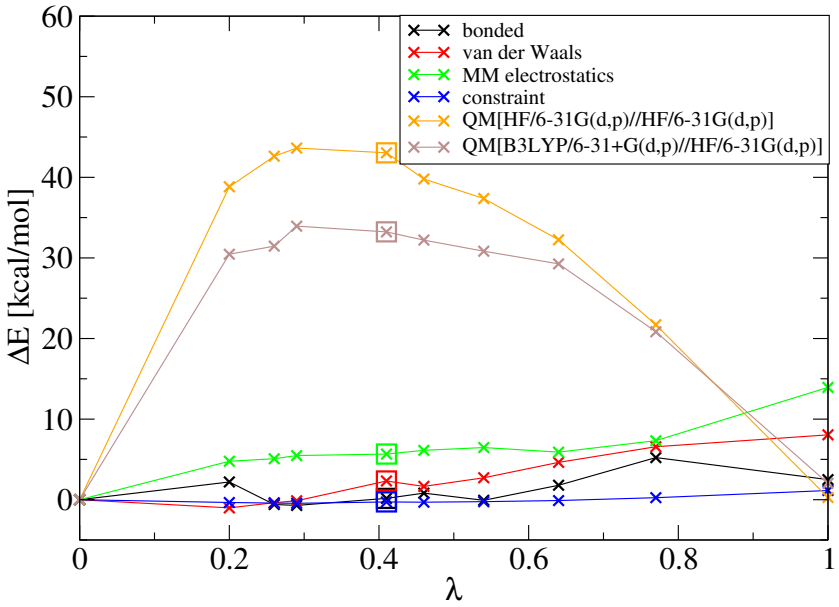

Figure 13.36: Dissection of the total energy into its contributing terms (in kcal/mol) for the direct pathway as refined on the QM[HF/6-31G(d,p)]/MM/NUCS energy surface.

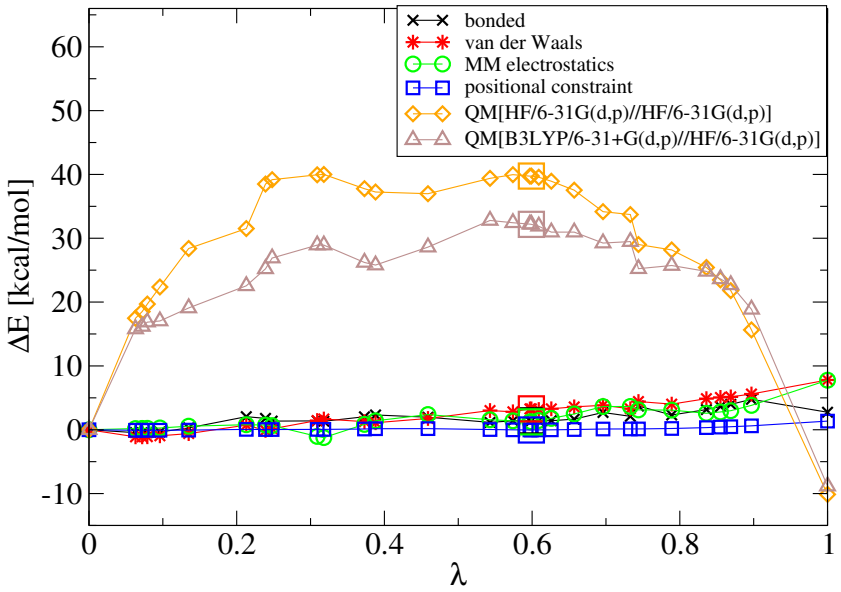

Figure 13.37: Dissection of the total energy into its contributing terms (in kcal/mol) for the Ser181 pathway as refined on the QM[HF/6-31G(d,p)]/MM/NUCS energy surface.

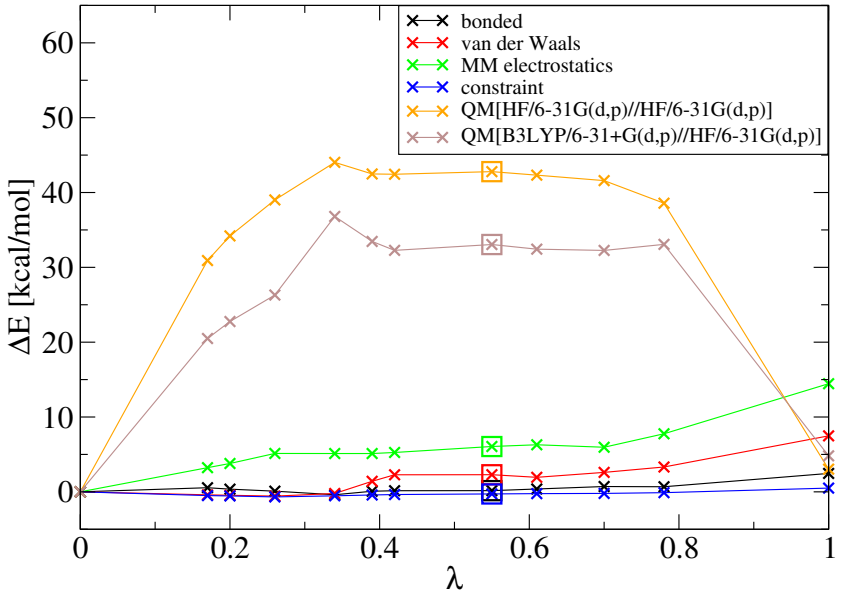

Figure 13.38: Dissection of the total energy into its contributing terms (in kcal/mol) for the Ser236 pathway as refined on the QM[HF/6-31G(d,p)]/MM/NUCS energy surface.

To further dissect the QM energy ΔE_{QM} into the quantum energy of the quantum region, ΔE_{QM}^{QM} and the QM/MM electrostatic interaction energy, $\Delta E_{QM/MM}^{elec}$, the single point energy calls were repeated after setting the partial atomic charges on all MM atoms to zero. The resulting total energy, $\Delta E_{tot}^{q=0}$, contains all energy contributions except for the MM electrostatics and the QM/MM electrostatic interactions. The corresponding total energy profiles along the reaction coordinates are shown in Figs. 13.39, 13.40, and 13.41 for the direct path, the Ser181 path, and the Ser236 path. In all three cases, $\Delta E_{tot}^{q=0}$ *increases* almost throughout the paths. At the transition states, $\Delta E_{tot}^{q=0}$ is in no case an energy maximum. This indicates that the electrostatic interactions (as described by the MM electrostatic terms and the QM/MM electrostatic interaction energy, $\Delta E_{QM/MM}^{elec}$) do play a major role in determining the stationary points that describe the hydrolysis reaction.

On the QM[HF/3-21G(d)//HF/3-21G(d)]/MM/NUCS and QM[B3LYP/6-31+G(d,p)//HF/3-21G(d)]/MM/NUCS surfaces, $\Delta E_{tot}^{q=0}$ increases nearly monotonically for the direct and Ser236 paths, reaching a maximum close to 60 kcal/mol at the product states. For the Ser181 path, $\Delta E_{tot}^{q=0}$ increases on these surfaces up to $\lambda = 0.92$, where an energy value of more than 80 kcal/mol is reached. In the product state, $\Delta E_{tot}^{q=0}$ drops to 68.44 and 66.7 kcal/mol on the QM[HF/3-21G(d)//HF/3-21G(d)]/MM/NUCS and QM[B3LYP/6-31+G(d,p)//HF/3-21G(d)]/MM/NUCS surfaces, respectively.

On the QM[HF/6-31G(d,p)//HF/6-31G(d,p)]/MM/NUCS and QM[B3LYP/6-31+G(d,p)//HF/6-31G(d,p)]/MM/NUCS surfaces, $\Delta E_{tot}^{q=0}$ increases nearly monotonically for all paths up to $\lambda \approx 0.8$, where maxima of around 70 kcal/mol are reached. Towards the product state $\Delta E_{tot}^{q=0}$ then drops to ∼40 kcal/mol for the Ser181 path and ∼55 kcal/mol for the direct and Ser236 paths. Thus, $\Delta E_{tot}^{q=0}$ of the product state of the Ser181 path is ∼15 kcal/mol less unfavorable than $\Delta E_{tot}^{q=0}$ of the product states of the direct and Ser236 paths. Therefore, $\Delta E_{tot}^{q=0}$ is the energy term that represents the main contribution to the reaction energy difference observed for the three paths (see Table 13.15).

The QM/MM electrostatic interaction energies, $\Delta E_{QM/MM}^{elec}$ (compare Section 10.7) as determined on the four different potential energy surfaces and the QM energies *without* the QM/MM electrostatic interaction energies, ΔE_{QM}^{QM}, are plotted in Figs. 13.42, 13.43, and 13.44 for the direct path, the Ser181 path, and the Ser236 path, respectively.

In all three paths, $\Delta E_{QM/MM}^{elec}$ as determined on the QM[B3LYP/6-31+G(d,p)//HF/3-21(d)]/MM/NUCS surface closely follows $\Delta E_{QM/MM}^{elec}$ as determined on the QM[HF/3-21(d)//HF/3-21(d)]/MM/NUCS surface. This indicates that the QM/MM electrostatic interactions are well captured by the Hartree-Fock method. Throughout the whole paths, the deviations in $\Delta E_{QM/MM}^{elec}$ between energies calculated on the DFT surfaces and the Hartree-Fock surface increase with the reaction coordinate λ. In all three paths, $\Delta E_{QM/MM}^{elec}$ (QM[B3LYP/6-31+G(d,p)//HF/3-21(d)]/MM/NUCS) is less favorable than $\Delta E_{QM/MM}^{elec}$ (QM[HF/3-21G(d)//HF/3-21(d)]/MM/NUCS), reaching a maximum deviation of about 6 kcal/mol at $\lambda = 1$ for all three paths. On the Hartree-Fock and DFT surfaces using the

HF/3-21G(d)-optimized coordinates and in the direct path, $\Delta E^{elec}_{QM/MM}$ decreases from 0 kcal/mol to -7.2 and -5.4 kcal/mol between $\lambda = 0$ and $\lambda = 0.26$, followed by an increase to -3.3 and -1.1 kcal/mol at $\lambda = 0.38$. Between $\lambda = 0.38$ and $\lambda = 1$, $\Delta E^{elec}_{QM/MM}$ decreases to its final values of -55.7 and -49.8 kcal/mol. In the Ser181 path, $\Delta E^{elec}_{QM/MM}$ decreases monotonically to its final values of -67.4 and -61.3 kcal/mol. A small bump at $\lambda = 0.36$ and a major drop between $\lambda = 0.74$ and $\lambda = 0.83$ are characteristic of the curve. In the Ser236 path, $\Delta E^{elec}_{QM/MM}$ increases to 4.7 and 6.6 kcal/mol up to $\lambda = 0.26$, followed by a decrease to -9.3 and -6.5 kcal/mol up to $\lambda = 0.44$ and a second increase to -6.6 and -4.4 kcal/mol at $\lambda = 0.52$. In the second half of the path, $\Delta E^{elec}_{QM/MM}$ decreases to its final values of -51.4 and -45.3 kcal/mol.

As was the case for the HF/3-21G(d)-optimized paths, $\Delta E^{elec}_{QM/MM}$ as determined on the QM[B3LYP/ 6-31+G(d,p)//HF/6-31(d,p)]/MM/NUCS surface closely follows $\Delta E^{elec}_{QM/MM}$ as determined on the QM[HF/6-31(d,p)//HF/6-31(d,p)]/MM/NUCS surface for all three paths. Similarly, in all three paths, $\Delta E^{elec}_{QM/MM}$ (QM[B3LYP/6-31+G(d,p)//HF/6-31(d,p)]/MM/NUCS) is less favorable than $\Delta E^{elec}_{QM/MM}$(QM[HF/6-31G(d,p)//HF/6-31(d,p)]/MM/NUCS), reaching a maximum deviation of about 3 kcal/mol at $\lambda = 1$ for all three paths. Thus, the value of the maximum deviation $\Delta\Delta E^{elec}_{QM/MM}$ for the HF/6-31G(d,p)-optimized paths is only half the value for the HF/3-21G(d)-optimized paths. Unlike the behavior of $\Delta E^{elec}_{QM/MM}$ in the HF/3-21G(d)-optimized paths, $\Delta E^{elec}_{QM/MM}$ increases initially. In the direct path, it increases to values of 4.9 and 5.5 kcal/mol at $\lambda = 0.29$ on the Hartree-Fock and DFT surfaces, respectively. Subsequently, $\Delta E^{elec}_{QM/MM}$ decreases monotonically to its final values of -44.8 and -41.9 kcal/mol. In the Ser181 path it decreases from 0 kcal/mol to -7.35 and -7.59 kcal/mol at $\lambda = 0.21$, followed by an increase to 1.26 and 0.90 kcal/mol at $\lambda = 0.39$ and a subsequent decrease to its final values of -41.97 and -39.78 kcal/mol. In the Ser236 path, the initial increase continues up to $\lambda = 0.26$ with values of $\Delta E^{elec}_{QM/MM}$ of 10.8 and 11.7 kcal/mol on the Hartree-Fock and DFT surfaces, respectively. In the remaining part of the path $\Delta E^{elec}_{QM/MM}$ decreases almost linearly to its final values of -43.5 and -40.5 kcal/mol.

ΔE^{QM}_{QM} as determined on the QM[B3LYP/6-31+G(d,p)//HF/3-21(d)]/MM/NUCS surface is not as clearly related to ΔE^{QM}_{QM} as determined on the corresponding Hartree-Fock surface as had been the case for $\Delta E^{elec}_{QM/MM}$ determined on the two surfaces, although in some path segments the two curves do agree to each other remarkably well. In the HF/3-21G(d)-optimized direct path (Fig. 13.42), ΔE^{QM}_{QM} increases monotonically to its final values of 51.4 and 49.3 kcal/mol on the Hartree-Fock and DFT surfaces, respectively. Throughout the whole path except for the product state, ΔE^{QM}_{QM} (QM[B3LYP/6-31+G(d,p)//HF/3-21(d)]/MM/NUCS) is more unfavorable than ΔE^{QM}_{QM} (QM[HF/3-21(d)//HF/3-21(d)]/MM/NUCS). Up to $\lambda = 0.24$ the two curves are nearly identical. Between $\lambda = 0.26$ and $\lambda = 0.87$ the curves deviate with a maximum deviation of 12.8 kcal/mol at $\lambda = 0.61$. Thus, $\Delta\Delta E^{QM}_{QM}$ contributes much more to the deviation between ΔE_{QM} as calculated on the two different

surfaces (which is $\Delta\Delta E_{QM} = 15.7$ kcal/mol at $\lambda = 0.61$ for the direct path, compare Fig. 13.33) than $\Delta E^{elec}_{QM/MM}$, which contributes 2.9 kcal/mol.

In the HF/3-21G(d)-optimized Ser181 path (Fig. 13.43), ΔE^{QM}_{QM} increases up to $\lambda = 0.92$ to values of 71.8 and 75.9 kcal/mol on the Hartree-Fock and DFT surfaces and subsequently decreases to 53.1 and 51.4 kcal/mol at the product state. As is the case for the direct path, ΔE^{QM}_{QM} (QM[B3LYP/6-31+G(d,p)//HF/3-21(d)]/MM/NUCS) is more unfavorable than ΔE^{QM}_{QM} (QM[HF/3-21(d)//HF/3-21(d)]/MM/NUCS) throughout the whole path except for the product state. $\Delta\Delta E^{QM}_{QM}$ between the Hartree-Fock and DFT surfaces is close to zero up to $\lambda = 0.19$ and between $\lambda = 0.46$ and $\lambda = 0.52$. The maximum deviation of $\Delta\Delta E^{QM}_{QM} = 6.4$ kcal/mol is observed at $\lambda = 0.34$. At this path point, $\Delta\Delta E_{QM}$ is 8.4 kcal/mol (Fig. 13.34) and $\Delta\Delta E^{elec}_{QM/MM}$ is 2.0 kcal/mol. Thus, at this path point $\Delta\Delta E^{QM}_{QM}$ contributes more to $\Delta\Delta E_{QM}$ than $\Delta\Delta E^{elec}_{QM/MM}$ as was also the case in the direct pathway. This situation changes at $\lambda = 0.90$, where $\Delta\Delta E_{QM}$ reaches its maximum of 9.1 kcal/mol, out of which 3.6 kcal/mol come from $\Delta\Delta E^{QM}_{QM}$ and 5.5 kcal/mol come from $\Delta\Delta E^{elec}_{QM/MM}$. Thus, in contrast to the direct path, in the Ser181 path the deviations in the QM energy between the DFT and Hartree-Fock surfaces (on the HF/3-21G(d)-optimized coordinates) are equally due to the fluctuations in the QM energy of the QM part and the fluctuations of the QM/MM electrostatic interaction energy.

In the Ser236 path (Fig. 13.44) as refined on the QM[HF/3-21G(d)//HF/3-21G(d)]/MM/NUCS surface, ΔE^{QM}_{QM} increases throughout the whole path to its final values of 54.2 and 53.3 kcal/mol on the Hartree-Fock and DFT surfaces, respectively. Again, as is the case for the direct and Ser181 paths, ΔE^{QM}_{QM} (QM[B3LYP/ 6-31+G(d,p)//HF/3-21(d)]/MM/NUCS) is more unfavorable than ΔE^{QM}_{QM} (QM[HF/3-21(d)//HF/3-21(d)]/MM/NUCS) throughout the whole path except for the product state. $\Delta\Delta E^{QM}_{QM}$ remains close to zero for a significant part of the path, up to $\lambda = 0.62$. The largest deviation is observed at $\lambda = 0.76$, where $\Delta\Delta E^{QM}_{QM}$ is 11.6 kcal/mol. At the same path point, $\Delta\Delta E_{QM}$ reaches its maximum of 14.0 kcal/mol (Fig. 13.35). Thus, as in the direct path but unlike in the Ser181 path, $\Delta\Delta E^{QM}_{QM}$ contributes more to $\Delta\Delta E_{QM}$ than $\Delta\Delta E^{elec}_{QM/MM}$ (which is 2.3 kcal/mol at $\lambda = 0.76$).

ΔE^{QM}_{QM} as determined for the on HF/6-31G(d,p)-optimized paths exhibit a different behavior from the corresponding values determined for the HF/3-21G(d)-optimized paths on both the Hartree-Fock and DFT surfaces. In particular, the curves pass through a maximum in the final parts of the path, which is a qualitative behavior that is not observed for the HF/3-21G(d)-optimized paths. In addition, ΔE^{QM}_{QM}(QM[B3LYP/6-31+G(d,p)//HF/6-31(d,p)]/MM/NUCS) is less unfavorable than ΔE^{QM}_{QM}(QM[HF/6-31(d,p)//HF/6-31(d,p)]/MM/NUCS) throughout the whole paths, which is opposite to the situation observed for the HF/3-21G(d)-optimized paths. Moreover, $\Delta\Delta E^{QM}_{QM}$ on the Hartree-Fock and DFT surfaces for the HF/6-31G(d,p)-optimized paths differs from zero except for the reactant and the product states of the paths, unlike the corresponding $\Delta\Delta E^{QM}_{QM}$ on the Hartree-Fock

and DFT surfaces for the HF/3-21(d)-optimized paths that have been close to zero for a significant number of path segments.

In the HF/6-31G(d,p)-optimized direct path (Fig. 13.42), ΔE_{QM}^{QM} reaches its maximum at $\lambda = 0.64$ with values of 58.5 and 53.0 kcal/mol on the QM[HF/6-31G(d,p)//HF/6-31G(d,p)]/MM/NUCS and QM[B3LYP/6-31+G(d,p)//HF/6-31G(d,p)]/MM/NUCS surfaces, respectively. The final values at the product state are 45.1 and 43.9 kcal/mol, respectively. The maximum deviation $\Delta\Delta E_{QM}^{QM}$ of -11.8 kcal/mol is observed at $\lambda = 0.26$. This is the major contribution to $\Delta\Delta E_{QM}$ that also reaches its maximum at $\lambda = 0.26$ with a value of -11.2 kcal/mol (Fig. 13.36). The contribution of the QM/MM electrostatic interaction energy is as small as $\Delta\Delta E_{QM/MM}^{elec} = 0.6$ kcal/mol.[8] Thus, $\Delta\Delta E_{QM}^{QM}$ clearly dominates the difference in QM energies $\Delta\Delta E_{QM}$ between the DFT and Hartree-Fock energy surfaces.

In the HF/6-31G(d,p)-optimized Ser181 path (Fig. 13.43), ΔE_{QM}^{QM} increases up to $\lambda = 0.73$ to values of 68.4 and 61.3 kcal/mol on the Hartree-Fock and DFT surfaces and subsequently decreases to 31.8 and 31.0 kcal/mol at the product state. The maximum deviation $\Delta\Delta E_{QM}^{QM} = -12.8$ kcal/mol is observed at $\lambda = 0.24$. At this path point, $\Delta\Delta E_{QM}$ also reaches its maximum with a value of -13.3 kcal/mol (Fig. 13.37). $\Delta\Delta E_{QM/MM}^{elec}$ is -0.5 kcal/mol. Thus, at $\lambda = 0.24$ $\Delta\Delta E_{QM}^{QM}$ clearly dominates $\Delta\Delta E_{QM}$ as was also the case in the direct pathway. This behavior is observed throughout the whole path which is in contrast to the behavior of Ser181 path as refined on the QM[HF/3-21G(d)//HF/3-21G(d)]/MM/NUCS surface but which is comparable to the situation in all other paths.

In the HF/6-31G(d,p)-optimized Ser236 path (Fig. 13.44), ΔE_{QM}^{QM} reaches its maximum at $\lambda = 0.78$ with values of 63.0 and 54.9 kcal/mol on the QM[HF/6-31G(d,p)//HF/6-31G(d,p)]/MM/NUCS and QM[B3LYP/6-31+G(d,p)//HF/6-31G(d,p)]/MM/NUCS surfaces, respectively. The final values at the product state are 46.5 and 45.3 kcal/mol, respectively. The maximum deviation $\Delta\Delta E_{QM}^{QM}$ of -13.6 kcal/mol is observed at the same value of the reaction coordinate as in the case of the direct path, *i.e.*, $\lambda = 0.26$. As in the direct path, this is the major contribution to $\Delta\Delta E_{QM}$ that also reaches its maximum at $\lambda = 0.26$ with a value of -12.7 kcal/mol (Fig. 13.38). Again, the contribution of the QM/MM electrostatic interaction energy of $\Delta\Delta E_{QM/MM}^{elec} = 0.9$ kcal/mol is small. Thus, as before, ΔE_{QM}^{QM} clearly dominates the difference in QM energies ΔE_{QM} between the DFT and Hartree-Fock energy surfaces.

The energy difference between the product state of the Ser181 path and the product states of the direct and Ser236 paths is due to differences in the quantum energy of the QM atoms, ΔE_{QM}^{QM}. This reflects the favorable hydrogen bond between P_i and ADP that is

[8] The difference in sign between $\Delta\Delta E_{QM}^{QM}$ and $\Delta\Delta E_{QM/MM}^{elec}$ stems from the fact that ΔE_{QM}^{QM} is less unfavorable (or more favorable) on the DFT surface than on the Hartree-Fock surface, leading to a negative difference. In contrast, $\Delta E_{QM/MM}^{elec}$ is less favorable (or more unfavorable) on the DFT surface than on the Hartree-Fock surface, leading to a positive difference.

formed only in the Ser181 path. In contrast, the QM/MM electrostatic interactions with the protein $\Delta\Delta E^{elec}_{QM/MM}$ stabilize the product states of the substrate by an equal amount in all three paths.

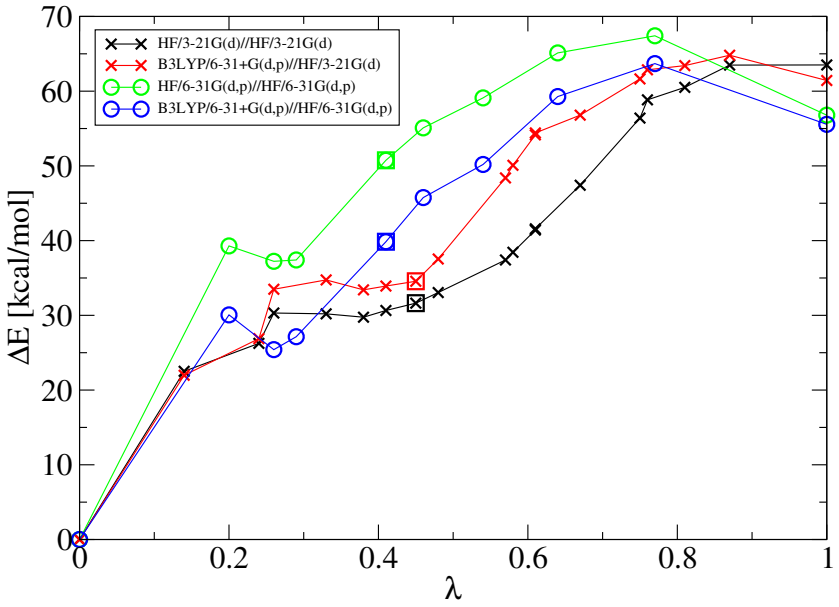

Figure 13.39: Total energies ($DeltaE^{q=0}_{tot}$) in kcal/mol along the reaction coordinate for the direct pathway. All MM partial atomic charges have been set to zero. The transition states are indicated by squares.

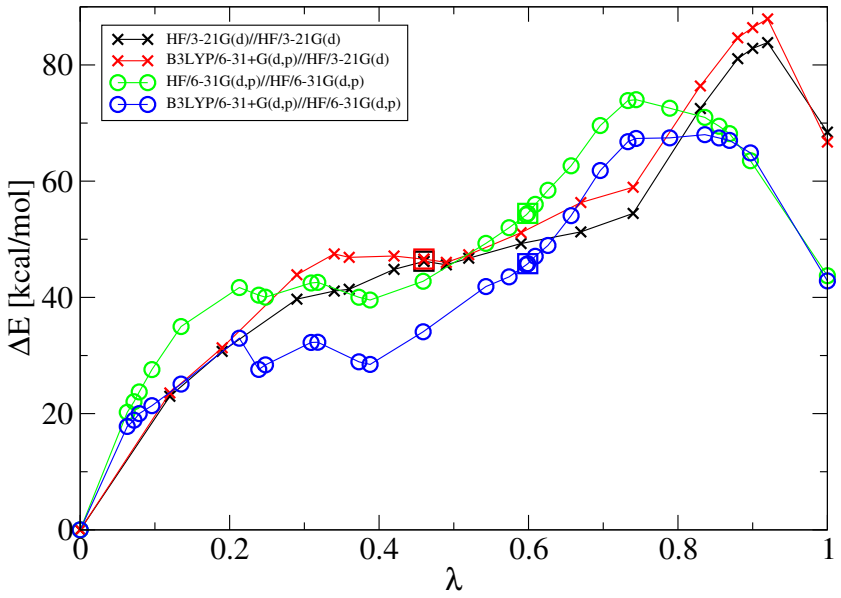

Figure 13.40: Total energies ($\Delta E_{tot}^{q=0}$) in kcal/mol along the reaction coordinate for the Ser181 pathway. All MM partial atomic charges have been set to zero. The transition states are indicated by squares.

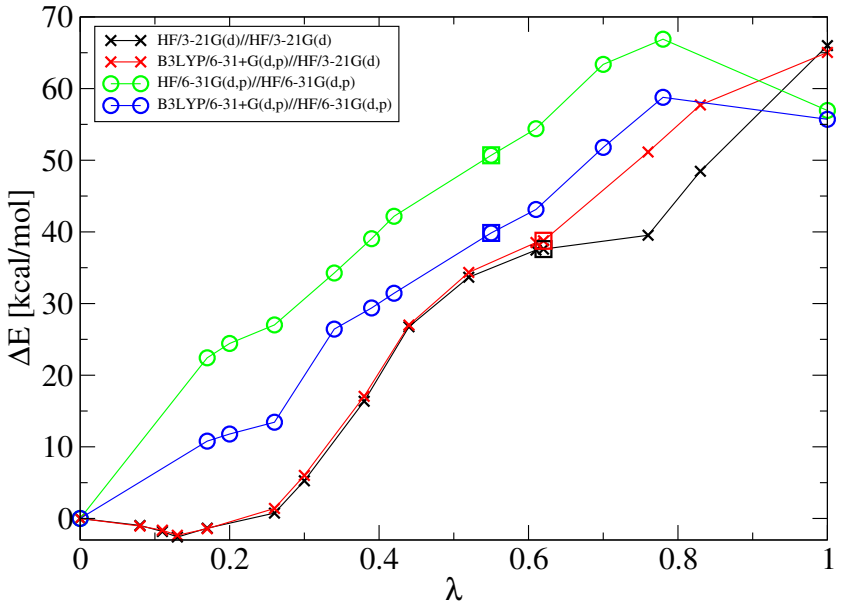

Figure 13.41: Total energies ($\Delta E_{tot}^{q=0}$) in kcal/mol along the reaction coordinate for the Ser236 pathway. All MM partial atomic charges have been set to zero. The transition states are indicated by squares.

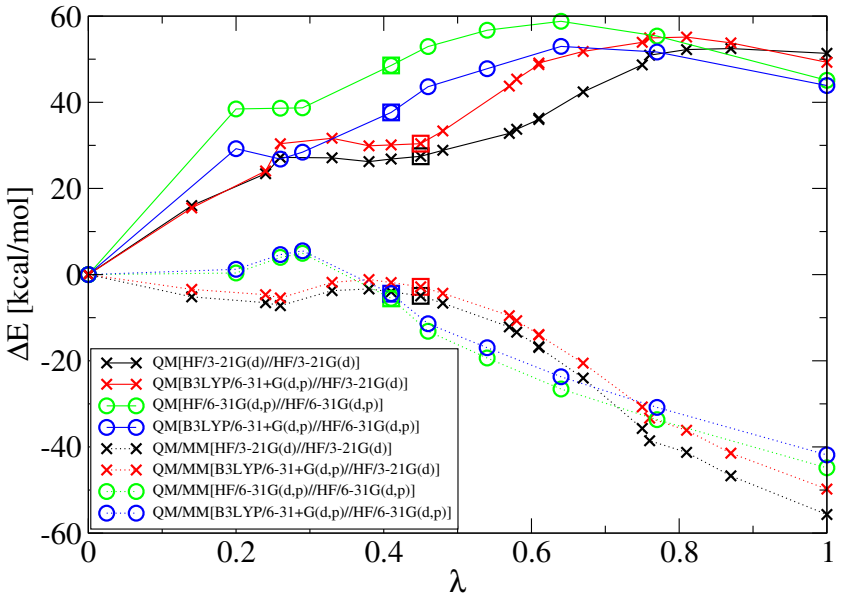

Figure 13.42: Dissection of the QM energy ΔE_{QM} into the quantum energy of the quantum region (ΔE_{QM}^{QM}) and the QM/MM electrostatic interaction energy ($\Delta E_{QM/MM}^{elec}$) for the direct path. The transition states are indicated by squares.

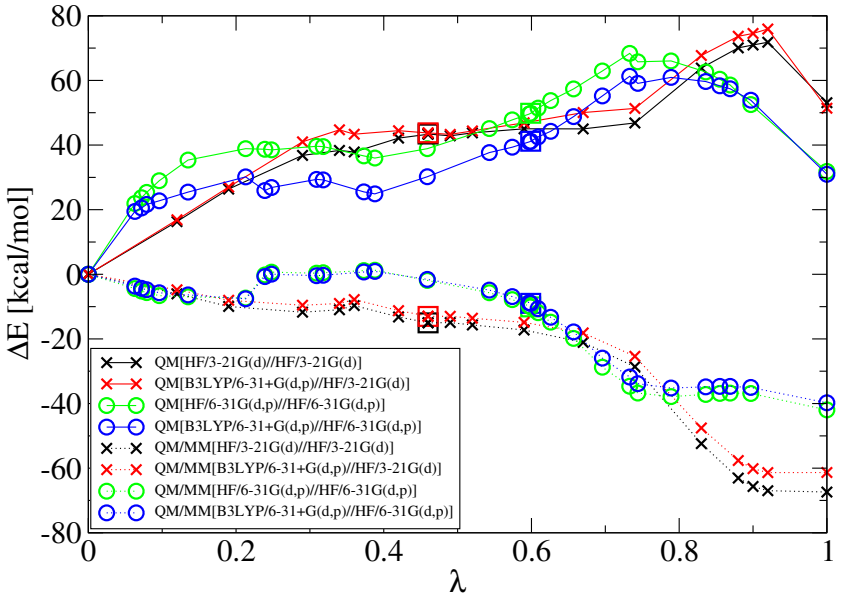

Figure 13.43: Dissection of the QM energy ΔE_{QM} into the quantum energy of the quantum region (ΔE_{QM}^{QM}) and the QM/MM electrostatic interaction energy ($\Delta E_{QM/MM}^{elec}$) for the Ser181 path. The transition states are indicated by squares.

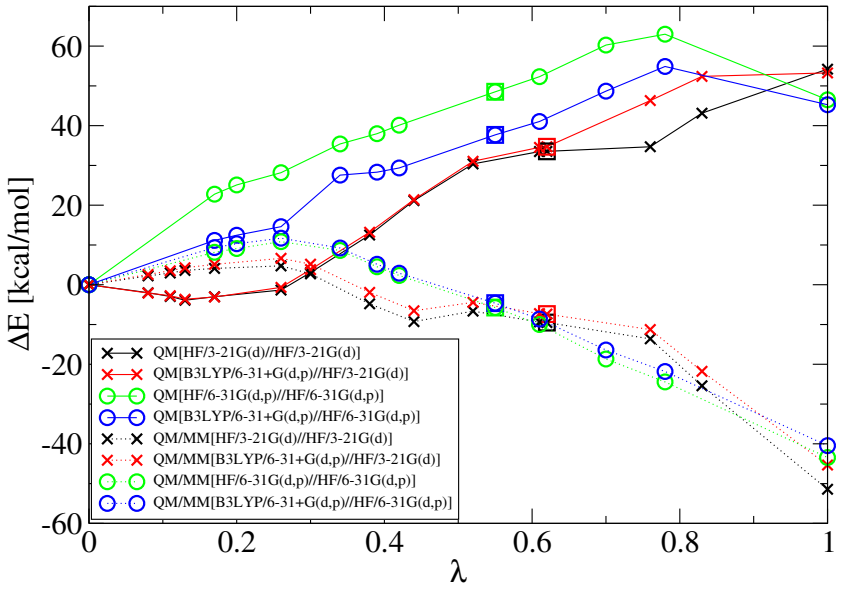

Figure 13.44: Dissection of the QM energy ΔE_{QM} into the quantum energy of the quantum region (ΔE_{QM}^{QM}) and the QM/MM electrostatic interaction energy ($\Delta E_{QM/MM}^{elec}$) for the Ser236 path. The transition states are indicated by squares.

13.4 Perturbation analysis

The total energy difference ΔE_{tot} between any two states can be split into the MM and QM/MM electrostatic interactions involving a specific residue X, ΔE^{X}_{elec}, and the remaining energy, $\Delta E^{q(X)=0}_{tot}$ that is determined as the total energy after perturbing the system by setting the charges on the residue X to zero (compare Section 10.7).

This perturbation analysis has been done to elucidate the contributions of the P-loop, switch-1, and switch-2 residues to the energy barriers and the reaction energies of all three paths. The resulting residue contributions are visualized in the subsequent figures (for an assignment of figure numbers and contributions see Table 13.16).

	direct	Ser181	Ser236
residue contributions to barrier heights			
P-loop	13.45	13.46	13.47
switch-1	13.48	13.49	13.50
switch-2	13.51	13.52	13.53
residue contributions to reaction energies			
P-loop	13.54	13.55	13.56
switch-1	13.57	13.58	13.59
switch-2	13.60	13.61	13.62

Table 13.16: Assignment of figure numbers to residue contributions for the three paths.

The residue contributions as determined for the transition state structures of the three paths as optimized on the QM[HF/3-21G(d)]/MM/NUCS and QM[HF/6-31G(d,p)]/MM/NUCS surfaces differ qualitatively from each other. A striking example is the contribution of Lys185 in the transition state of the Ser181 path (Fig. 13.46) that is unfavorable by ∼3 kcal/mol as calculated with HF/3-21G(d)//HF/3-21G(d) and B3LYP/6-31+G(d,p)//HF/3-21G(d) whereas it is favorable by ∼-3 kcal/mol as calculated with HF/6-31G(d,p)//HF/6-31G(d,p) and B3LYP/6-31+G(d,p)//HF/6-31G(d,p). Another example is the contribution of Arg238 in the transition state of the Ser236 path (Fig. 13.50) that is favorable by ∼-3.5 kcal/mol as calculated with HF/3-21G(d)//HF/3-21G(d) and B3LYP/6-31+G(d,p)//HF/3-21G(d) whereas it is unfavorable by ∼2 kcal/mol as calculated with HF/6-31G(d,p)//HF/6-31G(d,p) and B3LYP/6-31+G(d,p)//HF/6-31G(d,p). These discrepancies reflect the different geometries of the transition states as refined on the HF/3-21G(d) and HF/6-31G(d,p) surfaces with the former being more reactant-like (*i.e.*, with tetrahedral configuration at P_γ) than the latter that are trigonal bipyramids. Since the latter more adequately reflect the transition state (see Section 13.2), only the residue contributions as determined for the HF/6-31G(d,p)-optimized geometries are likely to be significant.

In the product states the qualitative nature of the residue contribution is independent of the energy surface used. This reflects that here the HF/3-21G(d) and HF/6-31G(d,p)-optimized geometries are similar to each other. Nevertheless, significant variation in the magnitude of the energy contributions occurs, as is the case for Lys185 in the product states of the Ser181 and Ser236 pathways (Figs. 13.55 and 13.56), that is favorable by ∼-3 to ∼-10 kcal/mol.

The contributions of the P-loop residues to the barrier heights are smaller than 1.5 kcal/mol for all paths (as refined on the HF/6-31G(d,p) surface, Figs. 13.45, 13.46, and 13.47). The only exception is Lys185 in the Ser181 pathway that reduces the barrier height by ∼-3 kcal/mol. In the direct and Ser236 pathways this residue does not contribute to the barrier height. The contribution of Gly179 is slightly unfavorable in all three paths, whereas the contributions of Gly182 and Gly184 are always slightly favorable. Glu180 contributes unfavorably to the barrier height in the Ser181 pathway, whereas its contribution is close to zero in the direct and Ser236 pathways. In contrast, the contribution of Ser181 (backbone atoms only) is unfavorable in the direct and Ser236 paths, whereas it is favorable in the Ser181 path. The contributions of Ala183 and Thr186 (backbone atoms only) are insignificant.

The contributions of the switch-1 residues to the barrier heights vary between ∼-2 and ∼6 kcal/mol (all paths as refined on the HF/6-31G(d,p) surface, Figs. 13.48, 13.49, and 13.50). The contribution of Asn233 is slightly unfavorable in the direct and Ser236 paths, whereas is reduces the barrier height by ∼-2 kcal/mol in the Ser181 path. Asn234 contributes always slightly favorable, whereas the contributions of Asn235 ans Ser236 (backbone atoms only) are close to zero. The contribution of Ser237 (backbone atoms only) are unfavorable by 1 to 1.5 kcal/mol in all three paths. Arg238 also contributes unfavorably to the barrier height. The magnitude of its contribution is ∼2 kcal/mol in the direct and Ser236 paths, whereas it amounts to ∼6 kcal/mol in the Ser181 path.

The contributions of the switch-2 residues to the barrier heights vary between ∼-6 and ∼1 kcal/mol (all paths as refined on the HF/6-31G(d,p) surface, Figs. 13.51, 13.52, and 13.53) and are thus somewhat antagonistic to the switch-1 residues. The contribution of Asp452 is unfavorable in the Ser181 path but close to zero in the direct and Ser236 paths. In contrast, Phe458 does not contribute in the Ser181 path, but contributes favorably in the direct and Ser236 paths. Ile453 does not contribute to the barrier height in any of the three paths. The contribution of Ser456 is unfavorably by ∼1 kcal/mol in all three paths, which is compensated for by the favorable contribution of -1.5 to -2 kcal/mol of Gly457 that hydrogen-bonds to the γ-phosphate. Glu459 reduces the barrier height by ∼-4 kcal/mol in the direct and Ser236 paths and by ∼-6 kcal/mol in the Ser181 path.

Of particular interest is the combined contribution of the salt bridge between Arg238 and Glu459 that connects the Switch-1 and Switch-2 loops and has been shown to be essential (230). In all three paths, interactions with Glu459 lower the barrier, whereas interactions with Arg238 raise the barrier by a similar amount. Thus, the contribution of Arg238 cancels

at least partially the favorable contribution of Glu459, and the total contribution of the salt-bridge residues to the barrier height is negligible. This indicates that the likely role of this essential salt-bridge is to lock the active site, rather than to participate directly in the catalysis.

The contributions of the P-loop residues to the reaction energies are mainly favorable (all paths as refined on the HF/6-31G(d,p) surface, Figs. 13.54, 13.55, and 13.56). Only Gly179, Glu180, and Thr186 (backbone atoms only) contribute unfavorably. However, the magnitude of their energy contribution is less than 1 kcal/mol. This is clearly outweighed by the favorable contributions of Gly182 (-3 to -5 kcal/mol), Ala183 ($\sim$-3 kcal/mol), Gly184 (-1.5 to -2 kcal/mol), and Lys185 (-3 to -5 kcal/mol). The contribution of Ser181 (backbone atoms only) is negligible.

The contributions of the switch-1 residues to the reaction energies vary between $\sim$-6 and 11 kcal/mol (all paths as refined on the HF/6-31G(d,p) surface, Figs. 13.57, 13.58, and 13.59). The favorable contribution of Asn233 has a magnitude of $\sim$-6 kcal/mol in the direct and Ser236 paths and $\sim$-4 kcal/mol in the Ser181 path. Asn234 and Asn235 also contribute favorably, however, by the much smaller magnitude of -0.5 to -2 kcal/mol. The contribution of Ser236 (backbone atoms only) is unfavorable by 1 to 2 kcal/mol, whereas the contribution of Ser237 (backbone atoms only) is favorable by -5 to -6 kcal/mol. As was the case in the transition states, the contribution of Arg238 is clearly unfavorable with magnitudes of 10 to 11 kcal/mol.

The contributions of the switch-2 residues to the reaction energies vary between $\sim$-10 and 5 kcal/mol (all paths as refined on the HF/6-31G(d,p) surface, Figs. 13.60, 13.61, and 13.62) and are thus somewhat antagonistic to the switch-1 residues, as has also been observed in the transition states. The contribution of Asp452 is favorable by -1 to -2 kcal/mol. Ile also contributes slightly favorably. The contribution of Ser456 is, however, unfavorable by 4 to 5 kcal/mol. Gly457 and Phe458 contribute again favorably by $\sim$-3 kcal/mol and $\sim$-1 kcal/mol. Glu459 compensated the unfavorable contribution of its salt-bridge partner Arg238 with its favorable contribution of -9 to -10 kcal/mol. Thus, as in the transition states, the total contribution of the Arg238:Glu459 salt-bridge to the reaction energy is small.

Remarkably, the residue contributions are often similar in the direct and Ser236 paths, whereas they differ in the Ser181 path. This confirms that the Ser181 path is more different from the direct and Ser236 paths than the mutual difference between the direct and Ser236 paths.

The total contribution of all binding-pocket-forming residues to the reaction energy amounts to -21.7 (-19.1) kcal/mol, -18.4 (-16.7) kcal/mol, and -22.5 (-20.2) kcal/mol with HF/6-31G(d,p)//HF/6-31G(d,p) (or B3LYP/6-31+G(d,p)//HF/6-31G(d,p)) for the direct, Ser181, and Ser236 product state, respectively. These values include the combined MM and QM/MM electrostatic contributions of the residues of the binding pocket, but are dominated by the QM/MM electrostatics. They contribute about half of the total QM/MM

electrostatic stabilization energy that stabilizes the anionic charge-pair P_i^-/ADP^{3-} in the product state.

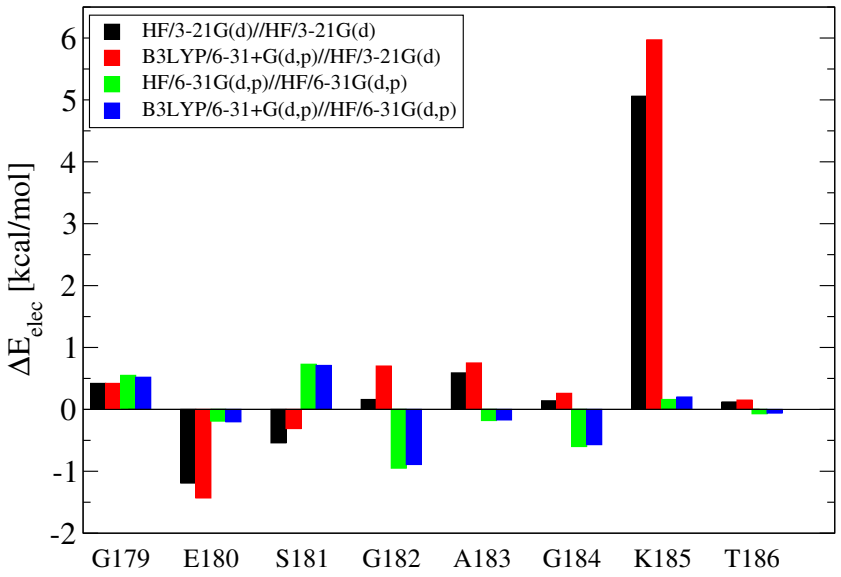

Figure 13.45: Contributions of the P-loop residues to the barrier height of the direct pathway.

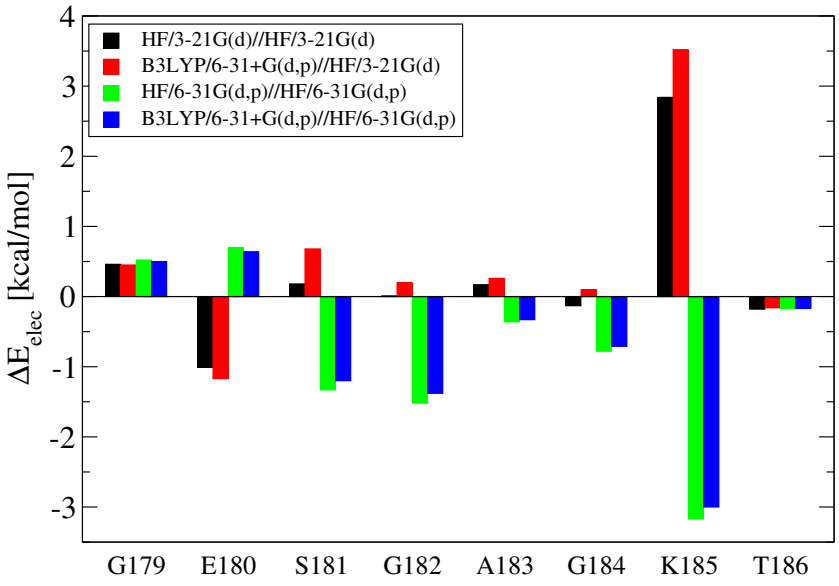

Figure 13.46: Contributions of the P-loop residues to the barrier height of the Ser181 pathway.

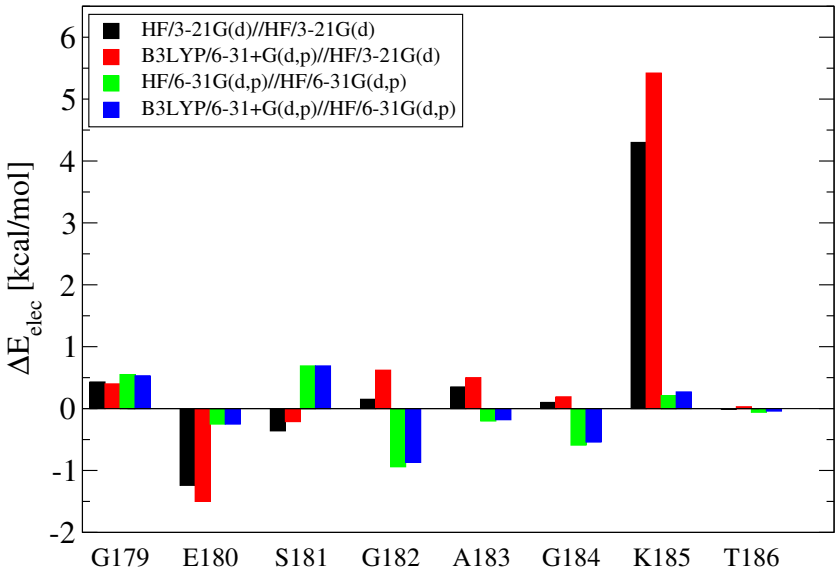

Figure 13.47: Contributions of the P-loop residues to the barrier height of the Ser236 pathway.

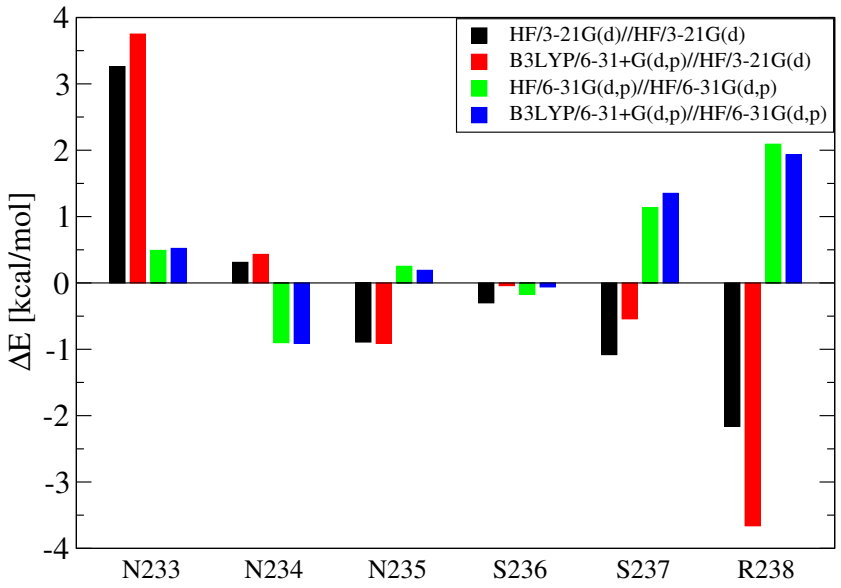

Figure 13.48: Contributions of the switch-1 residues to the barrier height of the direct pathway.

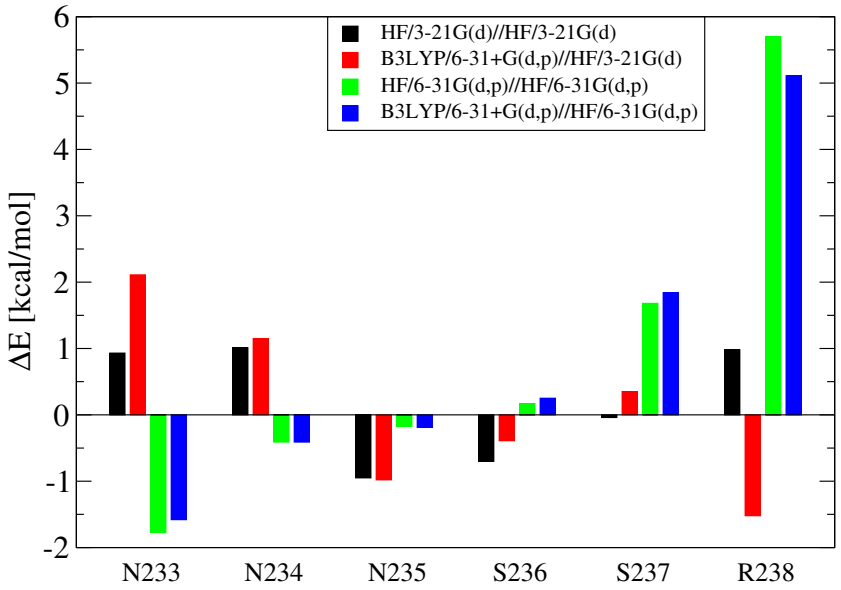

Figure 13.49: Contributions of the switch-1 residues to the barrier height of the Ser181 pathway.

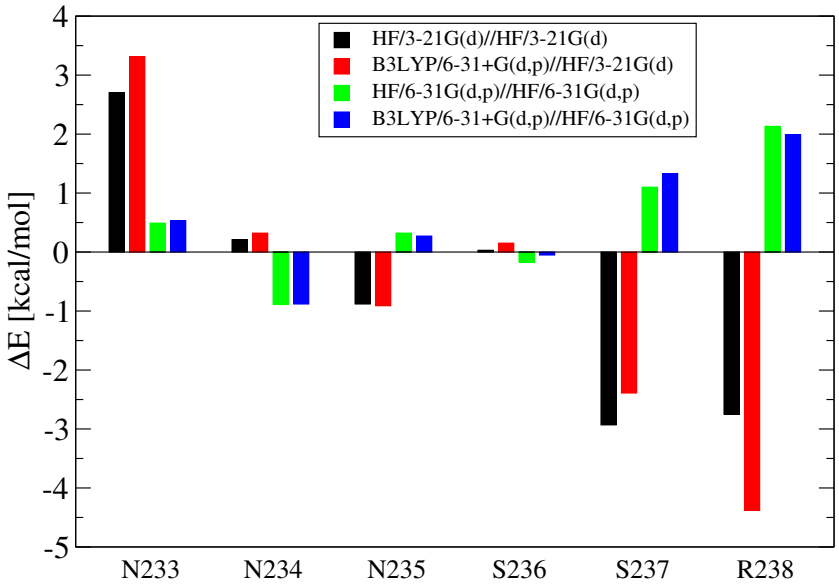

Figure 13.50: Contributions of the switch-1 residues to the barrier height of the Ser236 pathway.

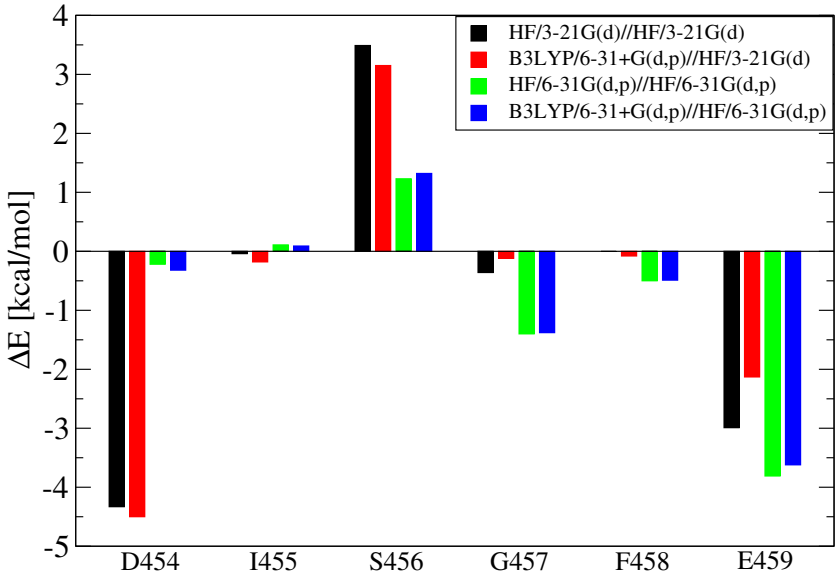

Figure 13.51: Contributions of the switch-2 residues to the barrier height of the direct pathway.

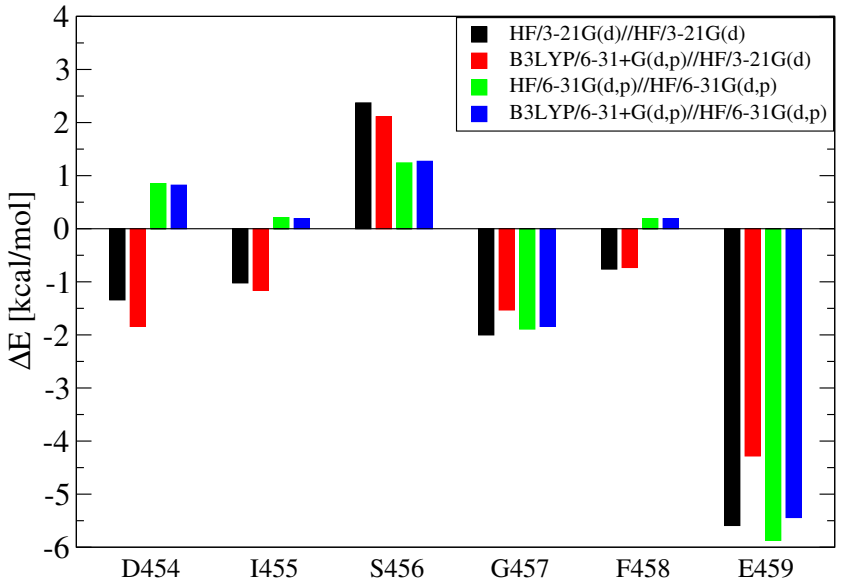

Figure 13.52: Contributions of the switch-2 residues to the barrier height of the Ser181 pathway.

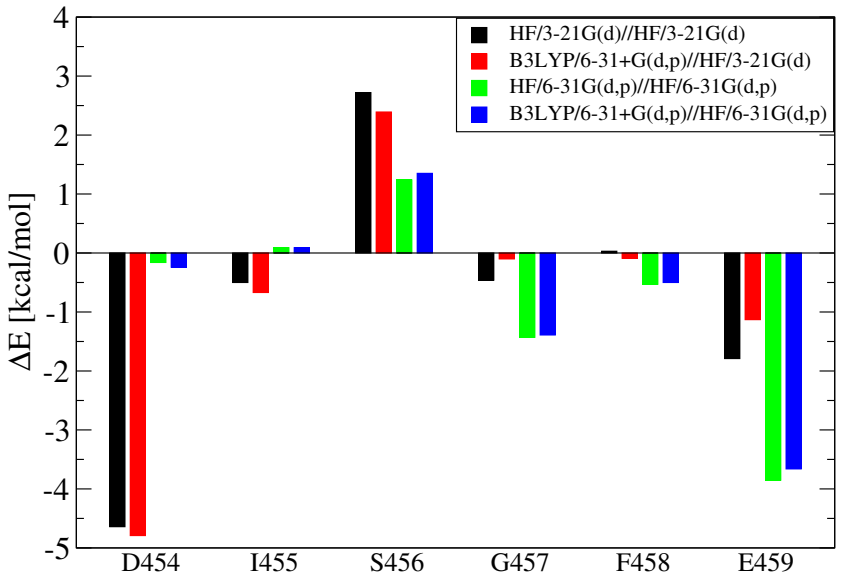

Figure 13.53: Contributions of the switch-2 residues to the barrier height of the Ser236 pathway.

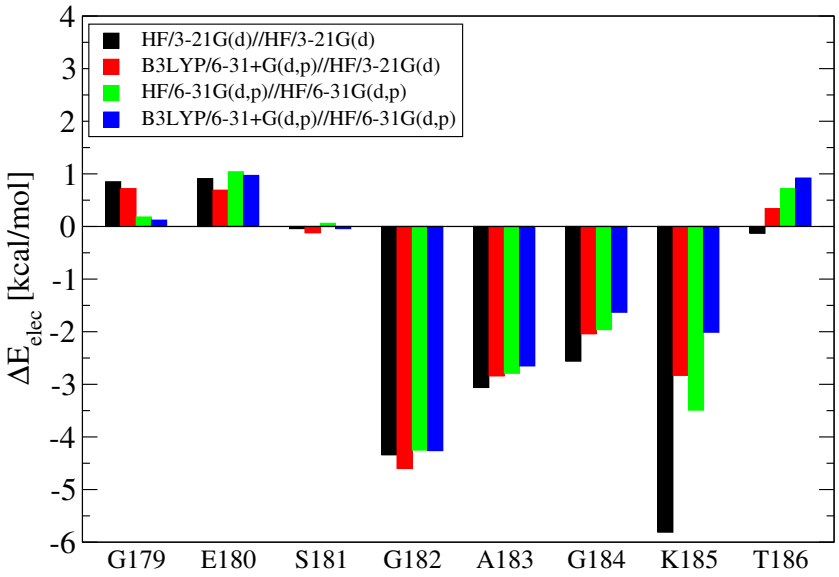

Figure 13.54: Contributions of the P-loop residues to the reaction energy of the direct pathway.

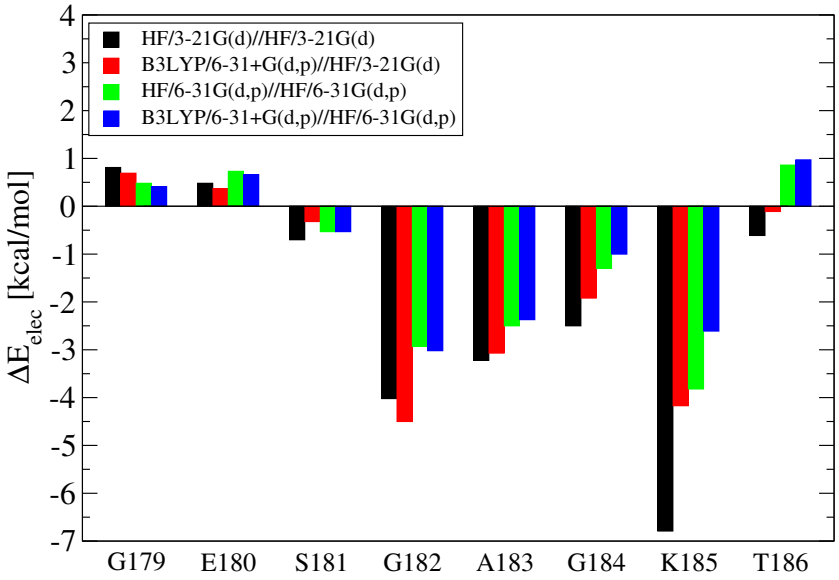

Figure 13.55: Contributions of the P-loop residues to the reaction energy of the Ser181 pathway.

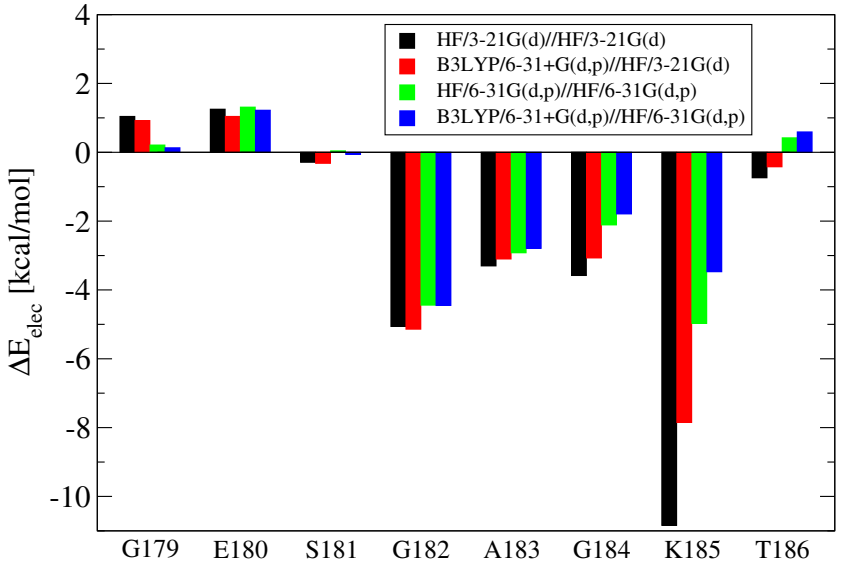

Figure 13.56: Contributions of the P-loop residues to the reaction energy of the Ser236 pathway.

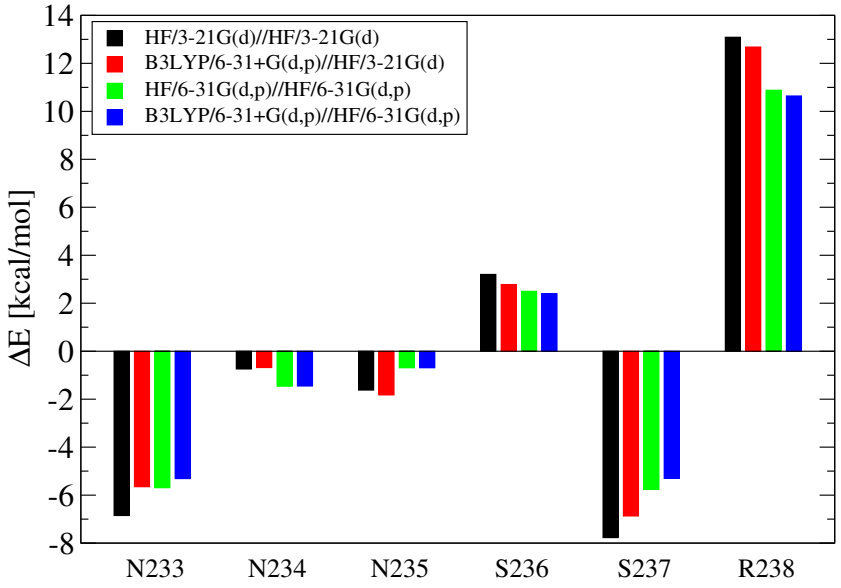

Figure 13.57: Contributions of the switch-1 residues to the reaction energy of the direct pathway.

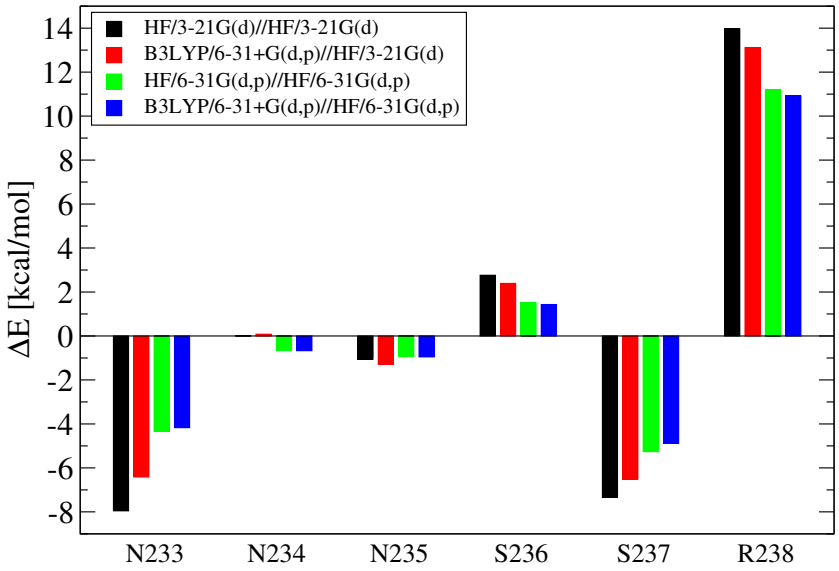

Figure 13.58: Contributions of the switch-1 residues to the reaction energy of the Ser181 pathway.

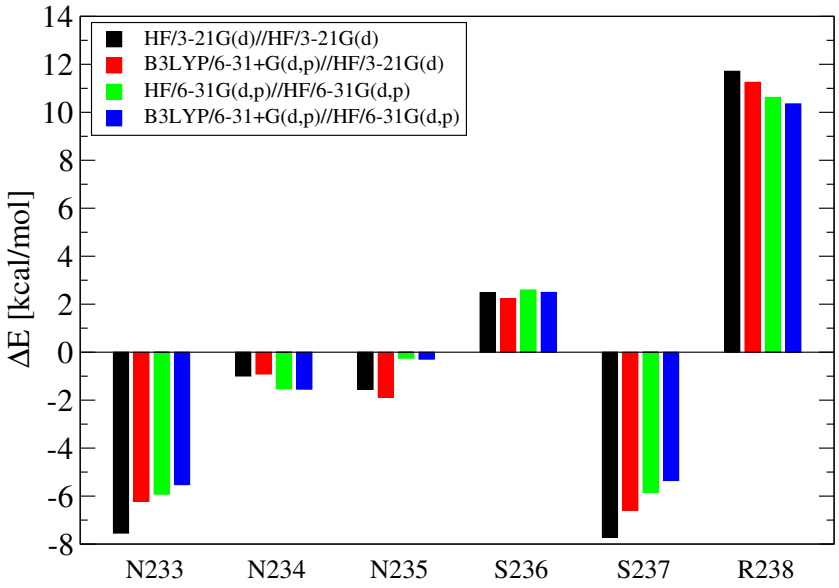

Figure 13.59: Contributions of the switch-1 residues to the reaction energy of the Ser236 pathway.

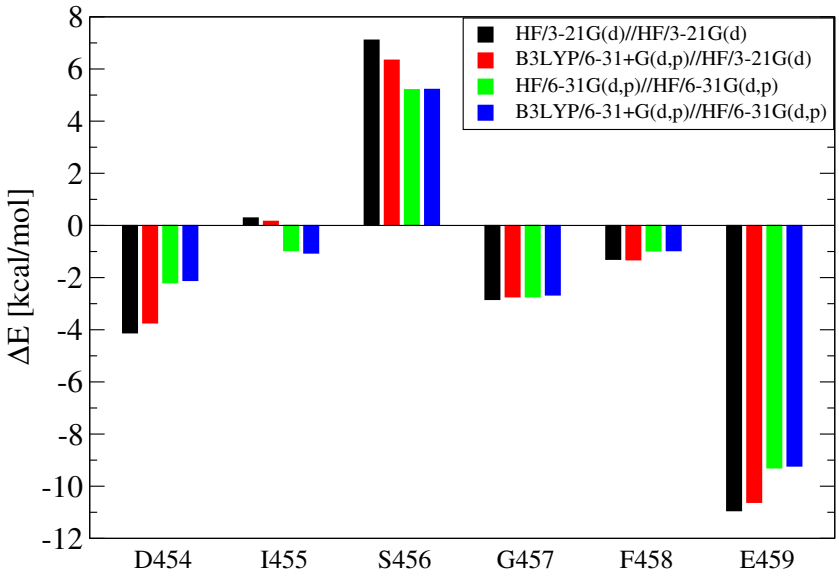

Figure 13.60: Contributions of the switch-2 residues to the reaction energy of the direct pathway.

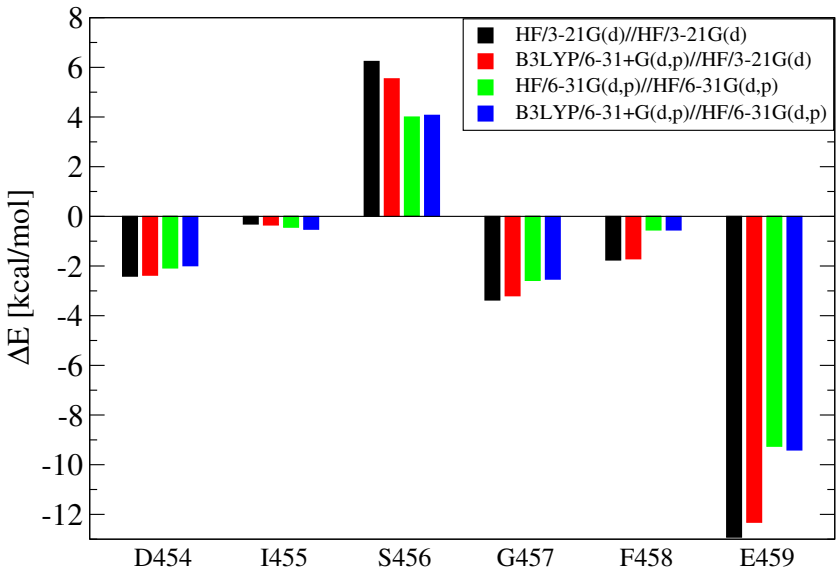

Figure 13.61: Contributions of the switch-2 residues to the reaction energy of the Ser181 pathway.

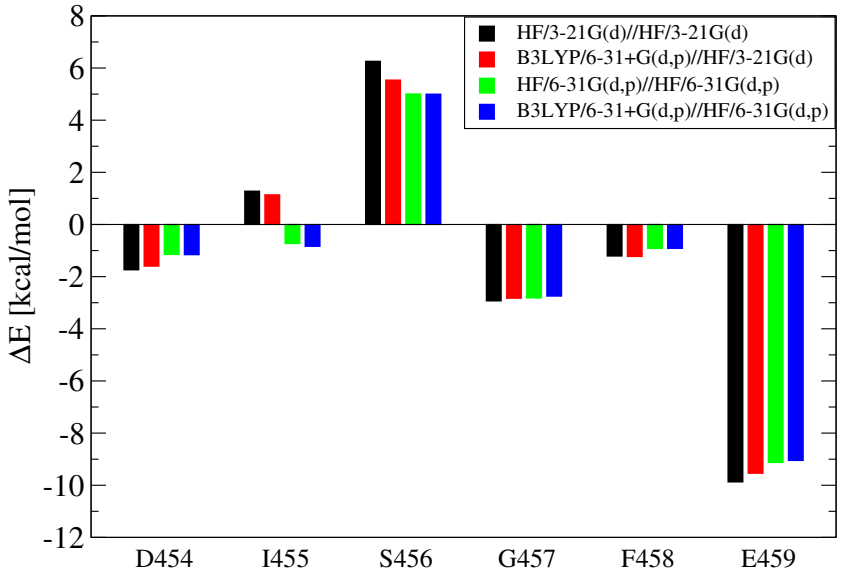

Figure 13.62: Contributions of the switch-2 residues to the reaction energy of the Ser236 pathway.

13.5 Charge shifts

Charge shifts were analyzed for the reaction paths as refined on the QM[HF/6-31G(d,p)]/MM/NUCS energy surface by performing a Mulliken charge analysis at the stationary points of the paths. The Mulliken charges q_M determined for the reactant state are collected in Table 13.17. The formal charge of +2 on the Mg cation is reduced to 1.20e (HF/6-31G(d,p)) and 0.21e (B3LYP/6-31+G(d,p)), while the total Mulliken charges on the triphosphate moiety of ATP is -3.37e (HF/6-31G(d,p)) and -3.07e (B3LYP/6-31+G(d,p)). Thus, in the reactant state, a significant amount of negative charge is transferred from ATP^{4-} to Mg^{2+}. The attacking water molecule is neutral (-0.01e and 0.07e), as is the helper water molecule (-0.05e and 0.03e). The ADP moiety of ATP carries a charge of 2.29e and -2.32e and is thus significantly more negative than its formal charge of -2. In addition, the sum of Mulliken charges on the attacking water and the γ-phosphate atoms (*i.e.*, all atoms that form the inorganic phosphate in the product state) is -1.09e and -0.68e. This is identical to the formal charge on $H_2PO_4^-$ rather than the expected formal charge of -2 on the γ-phosphate moiety. This means that most of the electrons pulled towards the Mg^{2+} come from the P_γ group of ATP, preparing it for attack by a nucleophile. Thus, the electrostatic charge distribution of the product state is already preformed in the reactant state.

The sidechains of Ser181 (-0.02e and -0.02e) and Ser236 (-0.02e and -0.06e) are slightly negatively charged. In contrast, the two Mg-coordinating water molecules carry a fractional positive charge (w1: 0.09e and 0.21e; w2: 0.16e and 0.08e). The same is true for the Mg-coordinating sidechains of Thr186 (0.06e and 0.29 e) and Ser237 (0.05 e and 0.18e). Thus, the positive charge on the Mg cation is partially distributed over the Mg-coordinating moieties.

For most atoms in the QM region the Mulliken charges do not change between the reactant and the transition state or product state geometries. The changes in sums of Mulliken charges on the different moieties are collected in Table 13.18. In the transition states of the three paths, the formal charge of +2 on the Mg cation is further reduced by -0.02 to -0.03e (HF/6-31G(d,p) and -0.06 to -0.11e (B3LYP/6-31+G(d,p)), while the ATP.H_2O^a moiety is more positive than in the reactant by 0.04 to 0.06e and 0.05 to 0.10e. Interestingly, the shift to more positive a charge on the ATP.H_2O^a moiety is due to a decrease in negative charge on the phosphate moiety by 0.08 to 0.10e and 0.13 to 0.14e on the HF and DFT surfaces, respectively, while the ADP moiety becomes slightly more negatively charged (-0.04e and -0.05 to -0.07e). The charge shifts on all other moieties are $\leq$ 0.05e. That the Mulliken charges on the moieties are essentially the same at the transition state as in the reactant state explains why the electrostatic contributions of the binding-pocket-forming residues to the barrier heights are small (compare Section 13.4). The geometric changes are small upon going from the reactant to the transition states. Therefore, changes in electrostatic interaction energies could only arise from changes in charge distributions. The Mulliken

analysis shows, however, that only insignificant charge shifts occur between the reactant and transition states, leading to small changes in electrostatic interaction energies between the QM and MM regions.

In the product state the charge on the Mg cation is similar to its charge in the reactant state (Δq_M HF/6-31G(d,p): -0.04 to -0.06e, Δq_M B3LYP/6-31+G(d,p): 0.03 to 0.06e). In contrast, the Mg-coordinating moieties Thr186 and Ser237 are slightly less positively charged than in the reactant (Δq_M HF/6-31G(d,p): 0.00 to -0.07e, Δq_M B3LYP/6-31+G(d,p): -0.02 to -0.15e). On the HF/6-31G(d,p) surface, the charge on ADP is more negative than in the reactant by -0.26 to -0.28e, whereas the charge on the inorganic phosphate is more positive by 0.28 to 0.31e and 0.07 to 0.08e. Thus, as expected, negative charge is shifted from the γ-phosphate moiety to the ADP moiety. However, instead of the full negative charge that is formally shifted only one third of an elementary charge is shifted. Nevertheless, since the charge pattern of the product state is already preformed in the reactant state, this charge shift seems sufficient to reach the final charge pattern of the product state.

For those atoms for which the charges change by $|\Delta q_M| \geq 0.08e$ the changes are collected in Table 13.19. As expected, the attacking water oxygen, O^a, is significantly more negative in the transition state than in the reactant state by -0.24e (HF/6-31G(d,p)) and -0.39 to -0.47e (B3LYP/6-31+G(d,p)), whereas is is only slightly more negative in the product state (HF/6-31G(d,p): 0.00 to -0.04e; B3LYP/6-31+G(d,p): -0.18 to -0.21e). Thus, in the transition state, the attacking moiety has the character of an hydroxide ion. The γ-phosphorus is more positive in the transition state than in the reactant state (HF/6-31G(d,p): 0.14 to 10.15e; B3LYP/6-31+G(d,p): 0.22 to 0.27e), whereas it becomes more negative in the product state on the B3LYP/6-31+G(d,p)//HF/6-31G(d,p) surface (-0.19 to -0.32e). $O_{\gamma 2}^{ATP}$, *i.e.*, the proton-accepting terminal oxygen on ATP, becomes more positive in both transition and product states than in the reactant state (HF/6-31G(d,p): 0.13 to 0.19e; B3LYP/6-31+G(d,p): 0.21 to 0.39 e). In the transition state, the negative charge is taken by the bridge oxygen, $O_{\beta\gamma}^{ATP}$ that is more negative than in the reactant state on the B3LYP/6-31+G(d,p) surface (-0.12 to -0.15e). Surprisingly, this atom is more positive in the product state than in the reactant state on the DFT surface (0.15 to 0.20e), although it transforms from a bridging oxygen in the reactant into a terminal oxygen on the ADP moiety in the product state. The negative charge is carried by the β-phosphorus, that is more negative in the product state than in the reactant state (HF/6-31G(d,p): -0.01 to -0.06e; B3LYP/6-31+G(d,p): -0.08 to -0.17e). The two non-bridging β-oxygens, $O_{\beta 1}^{ATP}$ and $O_{\beta 2}^{ATP}$, keep their charges.

In an associative mechanism the charges on the non-bridging γ-oxygens should be more negative in the transition state than in the reactant state (see Section 4.2.1). This is clearly not the case. In contrast, the proton-accepting $O_{\gamma 2}^{ATP}$ becomes more positive, which, however, may be due to its accepting a positive charge from of a proton. The two other non-bridging γ-oxygens do not show significant charge shifts. Alternatively, a dissociative

mechanism would be characterized by the $\beta\gamma$-bridge oxygen becoming more negative in the transition state. Although this is the case it is not very pronounced, since the charge shift is only observed on the B3LYP/6-31+G(d,p) surface and not on the HF/6-31G(d,p) surface. Thus, no clear argument can be made in favor of either an associative or a dissociative mechanism based on charge shifts.

In summary, the Mulliken charge analysis thus does not give insight into the associative or dissociative nature of the mechanism used by myosin to catalyze ATP hydrolysis. However, it does indicate that myosin acts electrostatically by preforming in the reactant state the charge distribution of the product state and thus minimizing the charge shifts that are actually required during hydrolysis.

atom	HF/6-31G(d,p)	B3LYP/6-31+G(d,p)	atom	HF/6-31G(d,p)	B3LYP/6-31+G(d,p)
C_{β}^{S181}	-0.05	-0.27	H_{1}^{ATP}	0.13	0.17
O_{γ}^{S181}	-0.71	-0.65	H_{2}^{ATP}	0.11	0.13
H_{γ}^{S181}	0.41	0.47	$O_{\alpha C}^{ATP}$	-0.72	-0.68
H_{q}^{S181}	0.11	0.15	P_{α}^{ATP}	1.67	2.37
$H_{\beta 1}^{S181}$	0.09	0.12	$O_{\alpha 1}^{ATP}$	-0.84	-0.93
$H_{\beta 2}^{S181}$	0.13	0.16	$O_{\alpha 2}^{ATP}$	-0.96	-1.23
C_{β}^{T186}	0.12	0.09	$O_{\alpha\beta}^{ATP}$	-0.84	-1.30
$O_{\gamma 1}^{T186}$	-0.74	-0.51	P_{β}^{ATP}	1.65	2.66
$H_{\gamma 1}^{T186}$	0.42	0.46	$O_{\beta 1}^{ATP}$	-0.87	-1.07
H_{q}^{T186}	0.09	0.09	$O_{\beta 2}^{ATP}$	-0.89	-1.05
$C_{\gamma 2}^{T186}$	-0.33	-0.43	$O_{\beta\gamma}^{ATP}$	-0.88	-1.34
H_{β}^{T186}	0.17	0.19	H_{q}^{ATP}	0.26	0.29
$H_{\gamma 1}^{T186}$	0.11	0.14	O^{a}	-0.75	-0.73
$H_{\gamma 2}^{T186}$	0.10	0.13	P_{γ}^{ATP}	1.73	2.50
$H_{\gamma 3}^{T186}$	0.11	0.14	$O_{\gamma 2}^{ATP}$	-0.91	-1.22
C_{β}^{S236}	-0.03	-0.26	$O_{\gamma 3}^{ATP}$	-0.95	-0.95
O_{γ}^{S236}	-0.72	-0.67	$O_{\gamma 1}^{ATP}$	-0.95	-1.08
H_{γ}^{S236}	0.41	0.46	H_{1}^{a}	0.36	0.40
H_{q}^{S236}	0.13	0.16	H_{2}^{a}	0.38	0.41
$H_{\beta 1}^{S236}$	0.11	0.14	O^{w1}	-0.79	-0.73
$H_{\beta 2}^{S236}$	0.09	0.12	H_{1}^{w1}	0.41	0.44
C_{β}^{S237}	0.01	-0.15	H_{2}^{w1}	0.47	0.50
O_{γ}^{S237}	-0.72	-0.54	O^{w2}	-0.75	-0.73
H_{γ}^{S237}	0.41	0.46	H_{1}^{w2}	0.40	0.44
H_{q}^{S237}	0.07	0.08	H_{2}^{w2}	0.43	0.45
$H_{\beta 1}^{S237}$	0.11	0.14	O^{h}	-0.75	-0.78
$H_{\beta 2}^{S237}$	0.16	0.19	H_{1}^{h}	0.34	0.40
Mg	1.20	0.21	H_{2}^{h}	0.37	0.41
$C_{5}^{\prime ATP}$	-0.10	-0.34			

Table 13.17: Mulliken charges q_M determined after single point energy calls using HF/6-31G(d,p) and B3LYP/6-31+G(d,p) in the reactant state geometry as optimized on the QM[HF/6-31G(d,p)]/MM/NUCS surface.

moiety	$\sum q_M$(reactant)	$\Delta \sum q_M$(direct)	$\Delta \sum q_M$(Ser181)	$\Delta \sum q_M$(Ser236)
		transition state		
Ser181	-0.02 (-0.02)	0.02 (0.01)	-0.01 (0.03)	0.02 (0.01)
Thr186	0.06 (0.29)	-0.01 (-0.02)	-0.01 (-0.03)	-0.01 (-0.02)
Ser236	-0.02 (-0.06)	-0.04 (-0.01)	-0.05 (0.02)	-0.04 (-0.01)
Ser237	0.05 (0.18)	0.00 (0.01)	0.00 (0.01)	0.00 (0.01)
Mg	1.20 (0.21)	-0.03 (-0.11)	-0.02 (-0.06)	-0.03 (-0.10)
ATP.H_2O^a	-3.38 (-3.00)	0.04 (0.10)	0.06 (0.05)	0.04 (0.10)
ATP	-3.37 (-3.07)	0.32 (0.47)	0.34 (0.50)	0.32 (0.46)
H_2O^a	-0.01 (0.07)	-0.27 (-0.37)	-0.29 (-0.44)	-0.27 (-0.37)
ADP	-2.29 (-2.32)	-0.04 (-0.05)	-0.04 (-0.07)	-0.04 (-0.05)
P_i	-1.09 (-0.68)	0.08 (0.14)	0.10 (0.13)	0.08 (0.14)
H_2O^{w1}	0.09 (0.21)	0.01 (0.00)	0.00 (0.00)	0.01 (0.00)
H_2O^{w2}	0.07 (0.16)	0.01 (0.02)	0.00 (0.01)	0.01 (0.02)
H_2O^h	-0.05 (0.03)	0.00 (-0.01)	0.02 (-0.03)	0.00 (-0.01)
		product state		
Ser181		0.03 (0.05)	-0.02 (0.04)	0.03 (0.05)
Thr186		0.01 (-0.05)	0.01 (-0.02)	0.00 (-0.06)
Ser236		0.01 (0.04)	0.03 (0.04)	0.02 (0.04)
Ser237		-0.07 (-0.15)	-0.06 (-0.14)	-0.06 (-0.15)
Mg		-0.06 (0.03)	-0.04 (0.06)	-0.05 (0.05)
ATP.H_2O^a		0.03 (0.11)	0.02 (0.08)	0.02 (0.08)
ATP		0.00 (0.20)	-0.10 (0.12)	-0.01 (0.19)
H_2O^a		0.03 (-0.09)	0.12 (-0.04)	0.03 (-0.11)
ADP		-0.27 (0.02)	-0.26 (0.01)	-0.28 (0.01)
P_i		0.31 (0.08)	0.28 (0.07)	0.30 (0.08)
H_2O^{w1}		-0.01 (-0.02)	-0.01 (-0.02)	0.00 (-0.01)
H_2O^{w2}		0.01 (0.00)	0.01 (0.01)	0.01 (0.00)
H_2O^h		0.04 (0.00)	0.05 (-0.04)	0.04 (0.00)

Table 13.18: Sums of Mulliken charges for different moieties in the reactant state, $\sum_{moiety} q_M$, and differences in sums, $\Delta \sum_{moiety} q_M$, between transition and reactant state geometries as well as between product and reactant state geometries as optimized on the QM[HF/6-31G(d,p)]/MM/NUCS surface for the direct path, the Ser181 path, and the Ser236 path. Mulliken population analyses were done after single point energy calls using HF/6-31G(d,p) (B3LYP/6-31+G(d,p)). Protons have been assigned to the heavy atoms to which they are bound in the different states.

atom	direct	Ser181	Ser236
	transition states		
O_{γ}^{S181}	0.02 (0.03)	0.02 (0.10)	0.02 (0.03)
Mg	-0.03 (-0.11)	-0.02 (-0.06)	-0.03 (-0.10)
P_{β}^{ATP}	0.01 (0.11)	0.01 (0.15)	0.01 (0.11)
$O_{\beta 2}^{ATP}$	-0.01 (0.01)	-0.01 (-0.01)	-0.01 (0.01)
$O_{\beta\gamma}^{ATP}$	-0.01 (-0.12)	0.00 (-0.15)	0.01 (-0.12)
O^{a}	-0.24 (-0.39)	-0.24 (-0.47)	-0.24 (-0.39)
P_{γ}^{ATP}	0.14 (0.27)	0.15 (0.22)	0.15 (0.27)
$O_{\gamma 2}^{ATP}$	0.14 (0.21)	0.19 (0.37)	0.14 (0.21)
$O_{\gamma 1}^{ATP}$	0.03 (-0.05)	0.02 (-0.04)	0.03 (-0.05)
H_{1}^{a}	-0.06 (-0.02)	-0.05 (-0.01)	0.06 (0.10)
	product states		
O_{γ}^{S181}	0.02 (0.06)	0.03 (0.11)	0.02 (0.06)
Mg	-0.06 (0.03)	-0.04 (0.06)	-0.05 (0.05)
P_{β}^{ATP}	-0.01 (-0.17)	-0.04 (-0.08)	-0.06 (-0.15)
$O_{\beta 2}^{ATP}$	-0.07 (0.09)	-0.06 (0.06)	-0.07 (0.08)
$O_{\beta\gamma}^{ATP}$	-0.02 (0.20)	-0.06 (0.15)	-0.02 (0.19)
O^{a}	-0.04 (-0.20)	0.00 (-0.18)	-0.04 (-0.21)
P_{γ}^{ATP}	0.04 (-0.20)	-0.04 (-0.32)	0.03 (-0.19)
$O_{\gamma 2}^{ATP}$	0.13 (0.26)	0.13 (0.39)	0.13 (0.25)
$O_{\gamma 1}^{ATP}$	0.08 (0.12)	0.06 (0.08)	0.08 (0.12)
H_{1}^{a}	0.05 (0.07)	0.05 (0.06)	0.03 (0.03)

Table 13.19: Differences in atomic Mulliken charges, Δq_M, between the transition and reactant state geometries as well as between product and reactant state geometries as optimized on the QM[HF/6-31G(d,p)]/MM/NUCS surface for the direct path, the Ser181 path, and the Ser236 path. Mulliken population analyses were done after single point energy calls using HF/6-31G(d,p) (B3LYP/6-31+G(d,p)). Only atoms with $|\Delta q_M| > 0.08e$ for any transition or product state structure are shown.

Chapter 14

Summary: Results

The mixed NUCS scheme developed in this thesis could be applied successfully to myosin, as is evident from the ability to reproduce the solution electrostatic potential at the QM region with a simple Coulomb potential to satisfactory accuracy. The scaling factors obtained were used for all subsequent calculations.

The reactant state for ATP hydrolysis in myosin was obtained by minimizing the crystal structure in which Be.F_x.ADP had been replaced by ATP with HF/3-21G(d) and HF/6-31G(d,p) for the QM atoms. Minimization resulted in a slight distortion of the geometry of ATP in the direction of the transition state geometry. Moreover, the P_γ:$O_{\beta\gamma}$ bond is elongated and thus activated for cleavage. The hydrogen-bonding network around the γ-phosphate favors multiple pathways for proton transfer and positions the proton donors corresponding to the three paths investigated for efficient transfer of a proton to $O_{\gamma 2}$. The three paths differ in their water activation mechanism (compare accompaying movies): In the direct path, the proton from the attacking water is transferred directly to $O_{\gamma 2}$, whereas in the Ser181 and Ser236 paths a proton relay involving the sidechains of the respective sidechains is built up.

Path optimizations with HF/3-21G(d) and HF/6-31G(d,p) revealed that the reaction proceeds in one step *via* a single transition state in all cases. The barrier heights of the three mechanisms, when comparing the values obtained using a given QM method, are very close to each other. Thus, the three activation mechanisms are, to within error, equally likely to be populated. Depending on the QM method used, barrier heights ranging from $\sim$25 to $\sim$50 kcal/mol and reaction energies ranging from $\sim$7 to $\sim$28 kcal/mol are obtained

The HF/3-21G(d)-optimized transition state geometries resemble the reactant state more than the expected trigonal bipyramidal transition state geometry. Moreover, the shape of the energy profile along the reaction coordinate changes significantly when point energy calls with B3LYP/6-31+G(d,p) for the QM atoms are performed on the HF/3-21G(d)-optimized structures. Thus, the 3-21G(d) basis set is too inaccurate to yield reliable struc-

tures. In contrast, the 6-31G(d,p) basis set was found to yield reliable structures. Thus, only the 6-31G(d,p)-optimized paths are considered further. The heavy-atom movements are very similar in all three paths. The paths follow an associative route *via* a trigonal bipyramidal transition state characterized by distances to the incoming and leaving oxygens of d(P_γ:O^a) = 2.18 Å and d(P_γ:O^a) = 1.73 Å, respectively. Up to the transition state, the angle a($O_{\gamma 3}$:Mg:$O_{\beta 2}$) remains close to 90°, indicative of octahedral coordination of the Mg^{2+} cation. Upon decay of the transition state, the angle widens to 120 - 125° and the distance d(Mg:O_γ^{S237}) increases to ∼3.5 Å. This breaks the coordination bond between Mg^{2+} and the switch-1 loop. In the product state, Mg^{2+} is only five-fold coordinated. The reaction energy of the Ser181 path is significanly lower than the reaction energy of the direct or Ser236 paths. This is due to the formation of a hydrogen bond between P_i and ADP in the Ser181 path that is not formed in the other two paths.

The energy decomposition showed that the energy barriers are in all cases clearly dominated by the QM energy that comprises both the quantum mechanical energy of the quantum atoms and the QM/MM electrostatic interaction terms. In the product state, the QM energy is close to zero for the direct and Ser236 paths and favorable for the Ser181 path. The unfavorable reaction energies are mainly due to the MM electrostatics and van der Waals terms. Further decomposition of the QM energy revealed that the quantum energy of the quantum atoms increases along the reaction coordinate with a maximum of ∼70 kcal/mol shortly before reaching the product states, where it drops to 30 - 40 kcal/mol. No energy maximum is observed at the transition states. This large unfavorable energy is compensated by the favorable QM/MM electrostatic interactions of up to -40 kcal/mol in the product states (HF/6-31G(d,p)-optimized paths). The energy variation among QM methods was found to be clearly dominated by the variation of the quantum energy of the quantum atoms.

The perturbation analyses showed that the residues forming binding pocket contribute little to the barrier height. In contrast, they stabilize the anionic ion-pair P_i^-:ADP^{3-} in the product state, contributing approximately half of the overall QM/MM electrostatic stabilization. This indicates that long-range electrostatic interactions contribute significantly to the stabilization of the product state. The combined contribution of Arg238 and Glu459, which form a salt bridge between the switch-1 and switch-2 loops, is small in all three paths, both in the transition and in the product state, indicating that these residues do not actively participate in catalyzing the hydrolysis reaction.

The Mulliken charge analyses revealed that the charge pattern of the product state is already preformed in the reactant state and that, consequently, only small charge shifts occur during the reaction. This explains the findings of the perturbation analyses that did not reveal significant electrostatic stabilization of the transistion state by the binding-pocket-forming residues.

Part IV

Myosin: discussion and outlook

Opening: Discussion and outlook

In this part hosts the assessment of the accuracy and the discussion of the reaction path calculations of ATP hydrolysis in myosin (15). After a section dealing with the accuracy and reliability of the present simulations (Section 15.1), the mechanism of ATP hydrolysis in myosin is discussed in Section 15.2. This involves consideration of the reaction profiles, elaboration on the degree of associativity of the hydrolysis reaction, and relating the simulation results to findings from isotope exchange experiments and mutation studies. Based on these considerations a Network Hypothesis for myosin-catalysis of ATP hydrolysis is put forward that unifies previous mechanistic proposals and in the framework of which experimental findings can be explained. The catalytic role of the Mg^{2+} metal cofactor is discussed, followed by considerations on the enzymatic strategy of myosin. A comparison of the hydrolysis mechanisms of myosin, the G-protein Ras, and F_1-ATPase completes this section. In Section 15.3 early post-hydrolytical chemo-mechanical coupling events are discussed and implications for the myosin motor cycle are outlined. Finally, in an outlook in Chapter 17, a number of future projects are suggested that cover both methodological aspects and simulations that are likely to give deeper insight into the chemo-mechanical coupling mechanisms in the molecular motor myosin.

Chapter 15

Discussion

15.1 Accuracy and reliability of the calculations

The calculations reported here represent the most accurate calculations on enzyme-catalyzed nucleoside triphosphate hydrolysis to date. None the less, computational limitations restrict the level of theory that can currently be employed for QM/MM calculations of the present kind (optimization of one path takes 2 to 4 months on 8 CPUs at 2.6 GHz with HF/6-31G(d,p)). The calculated reaction barrier heights and reaction energies for ATP hydrolysis in myosin of ∼25 to ∼50 kcal/mol and ∼6 to ∼25 kcal/mol, respectively (Table 13.15), show variations of ∼25 kcal/mol, thus suggesting that convergence in energies could not be achieved with the present methods. They also deviate significantly from the experimental values of ∼15 kcal/mol and ∼-2 kcal/mol (Table 4.11).

The variation in energies is dominated by the variation in the quantum mechanical energy (see Section 13.3) and can thus be traced back to the different accuracies of the quantum mechanical methods and basis sets used. A recent study compared the accuracy of different quantum mechanical methods including Hartree-Fock, MP2, and DFT methods by determining the geometries and energies of metal aqua complexes in gas-phase with a SBKJ (397) basis set on the metal ion and a 6-31G(d) basis set for the ligands (386). Although the geometries as optimized with the different methods were similar, the energies varied significantly. For example, the reaction energy of a water exchange reaction of $Sc(OH_2)_6^{3+}$ varied by as much as 62.4 kJ/mol or ∼15 kcal/mol. In the present study, the number of atoms in the quantum region is much larger than the number of atoms in a metal hexa-aqua complex. Thus, the total energies are also larger and it may be expected that the error in energy differences may be larger as well. In this reasoning, the errors due to usage of an effective potential basis set for transition metals may be equated with the errors introduced by treating the three phosphorus atoms present in the quantum region of the present study with a rather low basis set. In addition, the 3-21G(d) basis set used here yields so low an

accuracy that it is not any more considered in present-day quantum chemical studies on small molecules. Thus, the energy variation of ∼25 kcal/mol that is observed in the present work for a given path of ATP hydrolysis in myosin calculated with Hartree-Fock or DFT single point energy calls on either HF/3-21G(d)-optimized or HF/6-31G(d,p)-optimized coordinates does not seem unexpectedly large.

To estimate the error of the present calculations it would be necessary to perform benchmark simulations with a higher-order quantum mechanical method and with a larger basis set. Since "to get reliable results for compounds with first and second row elements, basis sets of at least a split-valence + polarization quality should be used" (386) the minimum basis set for such benchmark calculations should be 6-31+G(d,p). To get to an accuracy that is somewhat close to chemical accuracy it is necessary to use a quantum method that includes electron correlation effects (reviewed in (390)). One should, however, be aware that some methods including electron correlation nevertheless are too inaccurate to reach the chemical accuracy of ∼1 kcal/mol. For example, density functional theory with a B3LYP exchange-correlation functional still shows an average (maximum) error of ∼3 (∼20) kcal/mol (387) in the calculation of heats of formation. Second-order Møller-Plesset perturbation (MP2) theory (398) has an intrinsic error of up to ∼3.5 kcal/mol in the calculation of reaction enthalpies (399), even in the basis set limit. Thus, even more accurate methods would have to be used. However, the computational requirements to perform such high-level quantum chemical calculations for a quantum region of the size used in the present work and in the context of refining minimum-energy paths greatly exceed the currently available computational power.

The lack of diffuse functions in the basis sets used during path refinements will add to the deviation of the calculated from the experimental reaction barriers and energies. Since it is well established that for accurate quantum chemical treatment of anionic systems diffuse functions are required (388; 389) omitting diffuse function represents a major source of error. Augmenting the basis sets used with diffuse functions was not possible in the present work since this introduced convergence problems. This may be due to several reasons. Firstly, the addition of diffuse functions leads to a wave function that extends further from the centers of the quantum atoms than without diffuse functions. Thus, the electron density stretches further into the MM region in the direct neighborhood of the QM region with diffuse functions than without diffuse functions. If charges are present adjacent to the QM region this may lead to convergence problems. This may be the case for myosin in the present work, since the positively-charged sidechain of Lys185 is in intimate vicinity to the QM region, forming hydrogen bonds to oxygens on the γ- and β-phosphate moieties of ATP. Secondly, the choice of the quantum region itself may influence the convergence behavior. Depending on the boundary between QM and MM regions, the energy gap between the HOMO and the LUMO may change, due to different constitutions of the QM region and, accordingly, differences in the MM environment of the QM region. This may influence the convergence of the SCF procedure. Thirdly, the linking protocol used to cut chemical bonds between QM and MM regions may influence the convergence properties. In the

link-atom approach used in the present work some partial atomic charges on MM atoms bound to the boundary MM atom are zeroed, thus distorting the local electrostatic field and thereby also the QM/MM electrostatic interactions. In addition, constraints are used to position the QM atoms relative to the MM environment. This may introduce artificial forces in addition to the MM and QM forces, thus affecting the convergence stability. More elaborate linking protocols that avoid both the distortion of the local electrostatic field and the introduction of artificial constraint forces such as methods based on frozen orbital approaches are expected to yield more stable results (Compare Section 7.2).

The deviation in the total quantum energy (E_{QM}) between the two quantum methods used to evaluate the energetics of the HF/6-31G(d,p)-optimized paths, HF and B3LYP, amounts to about 10 kcal/mol. It is interesting to ask whether this discrepancy comes from the energy of the QM region itself (E_{QM}^{QM}) or from the interaction between the QM region and the protein ($E_{QM/MM}^{elec}$). E_{QM}^{QM} reaches a maximum deviation between HF and B3LYP of also about 10 kcal/mol, whereas the contribution of the QM/MM electrostatic interaction energy is always smaller than 1 kcal/mol. Thus, the difference between the quantum methods stems from the E_{QM}^{QM} term. This means that difference relative to higher-level quantum methods that will be observed in energy calculations of gas-phase ATP hydrolysis will be directly transferable to QM/MM calculations of ATP hydrolysis in myosin.

The energetics of the minimum-energy paths for ATP hydrolysis in myosin as calculated in the present work may deviate from experiment even if a high-level *ab initio* quantum chemical method including electron-correlation effects and a large basis set would have been used. A number of reasons could be responsible for this. Firstly, if the flexible MM region would be chosen too small to adapt to conformational changes in the quantum region, geometric changes would be propagated into the constrained MM region. Thus, atoms in the constrained region would be displaced from their crystal-structure positions to which they are harmonically constrained, resulting in an artificial energy increase that may distort the energy profile along the minimum-energy path. In the present work, this source of error was avoided by choosing the flexible MM region large enough to accommodate conformational changes in the QM region, as evidenced *a posteriori* by constraint energy terms close to zero for all three paths (Section 13.3). Secondly, the protein conformation resulting from geometry-optimization starting from the crystal structure may not be the geometry that is optimal to support the hydrolysis event. Even though the crystal structure used in the present study was solved with the ATP analog ADP.Be.F_3 bound to myosin in the C/C conformation (55) that is supposedly competent for hydrolysis it is possible that some sidechains are in a conformation so as to not optimally allow the hydrolysis to occur. Relaxing the crystal structure by short molecular dynamics simulations *prior* to geometry optimization may overcome this problem.[1] Thirdly, the calculation of minimum-

[1]Such MM-only molecular dynamics simulations were made with both ATP.H_2O and ADP.P_i bound to myosin (results not shown). Several structures from the MD trajectories were quenched by MM/NUCS minimizations, followed by QM[AM1]/MM/NUCS and QM[HF/3-21G(d)]/MM/NUCS minimizations. How-

energy paths intrinsically neglects the effect of the flexibility of the protein environment. Actually, the chemical reaction of ATP hydrolysis in myosin proceeds in the environment of the protein, with the protein undergoing constant conformational fluctuations. Thus, the reactive motions of the nuclei of the quantum region take place in the mean field created by the fluctuating atoms of the protein environment. To capture this effect one would have to replace the potential energy surface used in the present work by a free energy surface as obtained from averaging the potential energy over many conformations of the protein environment for a given conformation of the reactive QM region by means of free-energy perturbation (FEP) techniques. The number of energy evaluations required to perform such averaging, however, greatly exceeds the capacity of currently available computers if *ab initio* or density-functional quantum chemical techniques are used. Using semi-empirical methods such as AM1 (400), AM1-d (401), MNDO (402; 403), MNDO-d (404; 405; 406), or SCC-DFTB (407; 408; 409) may overcome this problem. However, parameters appropriate for treatment of phosphate hydrolysis reactions are not available (or not yet tested) for these methods.[2] In addition, not all of the necessary QM/MM interfaces between the respective QM program packages and CHARMM are currently publicly accessible. Thus, currently it is not possible to reliably treat enzyme-catalyzed phosphate hydrolysis reactions with semi-empirical QM/MM techniques. Nevertheless, the lack of averaging over protein conformers represents a non-negligible source of error. A marked example for this has been demonstrated for aldose reductase, where minimum-energy path simulations using an AM1/CHARMM potential yielded a barrier of more than 30 kcal/mol (412), whereas FEP simulations using an empirical valence bond approach for the quantum region found a free energy barrier of ∼17 kcal/mol (413), that approaches the experimental barrier of ∼15 kcal/mol (414).

An additional source of inaccuracy is the only approximate treatment of solvent effects that have been included in the present study *via* non-uniformly scaling the charges. As discussed in detail in Section 9.1.4 the NUCS procedure allows to capture qualitatively the shape of the energy profile along the reaction coordinate of an example reaction as determined from reference Poisson-Boltzmann calculations. However, the deviation in barrier heights when using scaled charges and a Coulomb potential compared to Poisson-Boltzmann electrostatics was as large as ∼15 kcal/mol in case of the Tyr35 ring flip event in BPTI (Section

ever, the resulting reactant and product structures differed in orientations of the sidechains in the flexible region. Since for each sidechain reorientation event the system must cross a saddle point on the potential energy surface, the number of saddle points that would have been necessary to optimize in order to connect one of the quenched reactant structures with one of the quenched product structures was too large to allow refining minimum-energy paths with CPR on a QM/MM/NUCS potential energy surface. Attempts to generate pairs of reactant and product structures by adiabatic mapping strategies as described in Section 10.4.2 starting from previously quenched structures failed due to large jumps in energy of up to ∼200 kcal/mol, the origin of which could not be elucidated.

[2]Recently, AM1/d parameters for phosphorus were developed aiming at the description of RNA cleavage reactions (410; 411). However, whether the parameters are transferable to other phosphate hydrolysis reactions is unknown.

9.1.3). Thus, the error due to approximate treatment of solvent effects is comparable in magnitude to the error due to the neglect of protein flexibility. Even though the NUCS procedure is not capable to adequately capture the energetics along the reaction coordinate, the geometric changes accompanying the BPTI ring flip event as determined with Poisson-Boltzmann electrostatics could be reproduced when using non-uniformly scaled charges. Thus, the geometries determined when approximating electrostatic solvent effects with the NUCS procedure can be considered plausible.

In summary, the computational approach chosen here to study the ATP hydrolysis reaction in myosin is too inaccurate to reliably describe the reaction energetics. However, due to the lack of more adequate feasible methods it represents the current status of research and available computational resources. Even though the present methodology cannot yield quantitative results it is possible to gain important insight into the qualitative nature of ATP hydrolysis in myosin. In particular, comparisons between different reaction mechanisms are possible and insight into the coupling between the ATPase activity of myosin and the motor cycle can be gained. Such conclusions can safely be drawn if they are supported by the results as obtained on all four energy surfaces used. For example, all three pathways studied are isoenergetic independent of the energy surface (Section 13.2). This can therefore be considered a reliable simulation result.

Moreover, the geometric changes of the heavy atoms along the reaction coordinates are similar in the QM[HF/ 3-21G(d)]/MM/NUCS-refined and QM[HF/6-31G(d,p)]/MM/NUCS-refined paths, even though the location of the transition state differs. Thus it is likely that the successions of geometries as determined by the simulations represent sequences of events that are indeed accessible to the system. This makes it possible to compare the geometric changes as determined for the three different paths studied as well as to relate the local geometric changes in the vicinity of the γ-phosphate moiety to changes in geometry in more distant parts of myosin. Thus, some propositions on how the ATPase activity and the motor cycle of myosin are coupled are possible to be made.

As described in Section 13.1 the geometries of the transition states as refined on the QM[HF/3-21G(d)]/MM/NUCS energy surface closely resemble the geometry of reactant state, whereas the QM[HF/6-31G(d,p)]/MM/NUCS-optimized transition state geometries exhibit a trigonal bipyramidal geometry at the P_γ, as is expected. In addition, the shapes of the energy profiles along the reaction coordinates for the three paths studied change when moving from the HF/3-21G(d)//HF/3-21G(d) surface to the B3LYP/6-31+G(d,p)//HF/3-21G(d) surface, whereas they are conserved when moving from the HF/6-31G(d,p)//HF/6-31G(d,p) to the B3LYP/6-31+G(d,p)//HF/6-31G(d,p) surface (Section 13.2). This indicates that using a 3-21G(d) basis set yields too inaccurate results to reliably describe the geometries at the stationary points. In contrast, usage of a split-valence 6-31G(d,p) basis that includes polarization functions seems to be able to capture the geometries of the stationary points. Thus, the following Sections of the discussion will consider only the three QM[HF/6-31G(d,p)]/MM/NUCS-optimized paths.

The degree of inaccuracy of the 3-21G(d) basis set found here puts the reliability of previous studies on enzyme-catalyzed nucleoside triphosphate hydrolysis reactions into question that have combined *ab initio* quantum chemical methods utilizing a small basis set with molecular mechanics. In particular, it seems questionable whether the conclusion of the recent study on GTP hydrolysis by the Ras-GAP complex in which a dissociative mechanism was found (190) is trustable, since geometry optimizations were performed using a 4-31G(d) basis set on the attacking water and the triphosphate moiety of GTP and only a 3-21G basis set on all other quantum atoms. The previous QM/MM study on ATP hydrolysis in myosin may suffer from similar inaccuracies since geometry optimizations were performed using a 3-21+G basis set (257). Although this basis set includes diffuse functions and is therefore presumably better suited to described the diffusiveness of the electronic wave function due to the negative charges present in the anionic quantum system in the myosin active site than the basis sets used in the present study, it lacks polarization functions and is therefore presumably less appropriate than the 6-31G(d,p) basis set used here to describe proton transfer events. In addition, use of a 6-31 split valence set provides more accurate results than use of a 3-21 split valence basis set since the former includes a larger number of basis functions. Thus, it is to be assumed that the HF/6-31G(d,p)-refined geometries of the present study are more reliable than the HF/3-21+G-refined geometries of the preceding study. The heights of the rate-limiting barriers of 25.9 to 38.8 kcal/mol as calculated on a B3LYP/6-31+G(d,p)//HF/3-21+G surface in this previous study are within the range of barrier heights observed in the current study (∼25 to ∼50 kcal/mol). The two available computational QM/MM studies on ATP hydrolysis in F_1-ATPase used a 6-31G(d) basis set for geometry optimizations (415; 416), thus including polarization functions on heavy atoms but not on hydrogen atoms. The reliability of this basis for description of proton transfer events can be questioned, nevertheless these studies represent the computationally most accurate QM/MM studies on nucleoside triphosphate hydrolyses published to date. Nonetheless, the present study utilizes a larger basis set than any of the previously-published studies and can therefore be considered the most accurate computational study currently available for an enzyme-catalyzed phosphoanhydride hydrolysis reaction.

15.2 Mechanistics of ATP hydrolysis in myosin

In this section, the mechanistic details that can be deduced from the simulation results using the QM[HF/6-31G(d,p)]/MM/NUCS potential energy surface are discussed and related to previous experimental and theoretical work. This contributes to answering the questions raised in Chapter 5.

15.2.1 Reaction profiles

All three water activation mechanisms studied, that is, the direct path, the Ser181 path, and the Ser236 path represent single-step reactions, *i.e.*, the reactant and the product states are connected *via* a single transition state. This is in contrast to the previous QM/MM study on ATP hydrolysis in myosin, in which a pentacoordinated intermediate state was found (257). However, it agrees with the experimental evidence from time-course experiments measuring oxygen exchange that suggested that ATP hydrolysis in myosin occurs without passing through a stable intermediate state (248) (see Section 4.6.1). Since both the current study that can be considered more accurate than the previous one (see above), and experiment indicate that ATP hydrolysis occurs *via* a single-step rather than a multi-step reaction mechanism it can be assumed that the actual enzymatic process indeed proceeds in a one-step fashion.

15.2.2 Degree of associativity (question 1)

The geometric analyses of the three pathways considered in the present study shows clearly that ATP hydrolysis in myosin follows an associative mechanism, in that the attacking oxygen first approaches the P_γ and the bond cleaved is broken afterwards (Figs. 13.15, 13.22, and 13.29 in Section 13.1). However, the mechanism is not purely associative, since a completely associative mechanism would be characterized by a pentacovalent phosphorane intermediate in which the bond distances of the bonds formed and cleaved are equal to the P:O single bond distance. Here, however, no intermediate is found (see above). To determine the degree of associativity as defined by Eq. 4.1 it is necessary to know the P:O single bond distance. Since the calculation methods chosen presumably yield a distance different from 1.73Å (as used in Ref. (133)) but with unknown numerical value, it is not possible to assign a fractional associativity to the pathways. Nevertheless, the bond distances of $d(P_\gamma : O^a) = 2.18$ Å and $d(P_\gamma : O_{\beta\gamma}) = 1.73$ Å found in the transition states of all three pathways studied are obviously shorter than the sum of van der Waals radii of phosphorus and oxygen (3.3 Å, see Section 4.2.1), thus ruling out a dissociative mechanism. That hydrolysis does not occur *via* a dissociative reaction route agrees with the findings from reaction quenching experiments (243). The bond lengths in the transition state are similar to the bond lengths found in the transition state of the uncatalyzed reaction for which values of 2.12 Å and 1.84 Å have been reported (153) (see Section 4.2.3). In addition, the mechanism cannot be characterized as an S_N2 mechanism, because the bond distances of the bonds formed and cleaved differ in the transition state, thereby contradicting the definition of an S_N2 mechanism that implies equal bond distances. Thus, the mechanism will be termed associative hereafter, owing to the fact that the incoming oxygen is bound before the outgoing oxygen is cleaved off but implying that certain features of a dissociative mechanism may nevertheless contribute. This is in accordance with findings from Raman spectroscopic studies that suggested a mechanism with more associative than dissociative

character (251).

A contribution from a dissociative-like mechanism could in principle be provided by stabilization of the charge on the bridge oxygen in the transition state. Such a role has been proposed for Asn233 (258). However, the present calculation do not support this proposal since no significant charge shifts are observed throughout the whole paths (see Section 13.5). Rather, myosin seems to act electrostatically by preforming the charge pattern of the product configuration already in the reactant state. Thus, charge shifts are apparently not used by myosin to support the hydrolysis reaction step itself and can therefore not be used to distinguish an associative mechanism from a dissociative one.

Experiments used to determine linear free energy relationships for a number of uncatalyzed phosphoanhydride hydrolysis reactions show only a small dependence of the hydrolysis rate constant on the pK_a of different incoming groups, but a large dependence on the pK_a of leaving groups (135). This had been interpreted in favor of a dissociative mechanism in aqueous solution. However, the unique interpretability of linear free energy relationships has been questioned (136). Moreover, hydrolysis in aqueous solution of methylphosphate – the smallest possible model system for phosphate hydrolysis – which has been examined using a variety of methods, can be explained indiscriminately using associative or dissociative mechanisms (145; 152; 153; 155). Furthermore, computational investigations on methyltriphosphate hydrolysis in aqueous solution did not give clear evidence in favor of either mechanism (153; 168). Since neither mechanism seems to be inherently more favorable, enzymes (including myosin) can, in principle, select either mechanism (169; 170).

The question of whether enzymes utilize a dissociative or an associative mechanism for phosphoryl transfer reactions thus remains open. Substrates such as vanadates are frequently used to crystallize proteins in a conformational state that is supposed to mimic the transition state (417) of the phosphate hydrolysis reaction. This, however, implies that a pentacovalent state indeed exists in the course of the enzyme-catalyzed reaction. Crystallographic evidence of a pentavalent phosphorus intermediate has been reported in 2003 with distances of 2.0 and 2.1 Å to the axial oxygens and 1.7 Å to the equatorial oxygens (418). This has provoked a new suggestion based on hypervalent phosphorus chemistry, namely that enzyme catalysis may proceed *via* a hexavalent phosphorus transition state in which a protein sidechain donates an additional oxygen, and a pentavalent intermediate (419). Details of such a mechanism, however, have so far not been suggested for any enzyme. The current study suggests that in case of myosin it seems unlikely that such a "hypervalent phosphorus" mechanism would be followed, since no pentacovalent intermediate could be located.

15.2.3 Relating the simulation results to isotope exchange experiments (question 2)

Isotope exchange experiments investigating the myosin catalysis of ATP hydrolysis have established that the terminal γ-oxygens on ATP rapidly exchange with the solvent. All four oxygens on the product P_i are equivalent (247), *i.e.*, the three terminal γ-phosphate oxygens on ATP and the attacking water oxygen. The $\beta\gamma$-bridge oxygen and the terminal β-oxygens do not exchange (249). Since hydrolysis in found in the present work to proceed without passing a stable intermediate state, the exchange of the terminal γ-oxygens can only occur if either the γ-phosphate moiety of ATP in the reactant state or the P_i in the product state possesses rotational freedom. In the reactant state, both the γ- and the β-phosphate moiety donate one oxygen each to the Mg^{2+} cation. Rotation of the γ-phosphate therefore would require breakage of the coordination bond. However, since the hexacoordination of Mg^{2+} represents the preferred coordination number of Mg^{2+}, found in 80% of structures in the Cambridge crystallographic databank (420), it seems unlikely that this bond could be broken spontaneously. Thus, it seems more likely that rotation occurs in the posthydrolysis state.

In the present simulations the $Mg{:}O^{S237}$ coordination bond is broken and Mg^{2+} is only five-fold coordinated after hydrolysis. Although this is an unusual coordination pattern for Mg^{2+}, five- or four-coordinate Mg^{2+} complexes are observed each in $\sim$8% of structures in the Cambridge crystallographic databank (421; 422). The present calculations show that, in the post-hydrolysis state, the $d(Mg^{2+}{:}O_{\gamma 3})$ coordination bond is longer and thus weaker than the $d(Mg^{2+}{:}O_{\beta 2})$ bond (see Tables 13.6, 13.10, and 13.14), indicating that the former is more likely to break than the latter. This makes it possible for the P_i to transiently dissociate from Mg^{2+}, yielding a four-fold coordinated Mg^{2+} cation and allowing the P_i to rotate. In such a scenario, the β-phosphate remains permanently coordinated to Mg^{2+}, hindering its rotation,which explains the experimentally-observed lack of isotope exchange in this group. Thus, the presently found product structure, with its five-fold Mg^{2+} coordination and weakened $d(Mg^{2+}{:}O_{\gamma 3})$ bond is consistent with the experimentally-observed features of the oxygen exchange. Moreover, breaking of the $d(Mg{:}O_{\gamma 3})$ bond as proposed here *must* eventually occur to allow exit of the P_i. The above scenario gives a possible explanation of this crucial early post-hydrolysis event.

To verify the above speculations, further computational studies would be required, such as conformational searches with a four-fold coordinated Mg^{2+} in the posthydrolysis M.ADP.P_i state combined with analyses of rotational barriers for P_i tumbling in the myosin catalytic site.

15.2.4 Relating the simulation results to mutational experiments (question 3)

Extensive kinetic characterization of a large number of active-site mutants of *Dictyostelium discoideum* myosin II has been reported (see Table 4.7), with different mutants having significantly different effects on the basal ATPase rate. Mutants that speed up the ATPase cycle are expected to facilitate the rate-limiting phosphate release step without significantly affecting the rate of the hydrolysis step itself. Mutants that slow down or completely inhibit the ATPase activity may interfere with the ATPase cycle by inhibiting substrate binding, by preventing the return stroke (*i.e.*, the conformational transition from the C/O state to the C/C state of myosin) thus making it impossible for myosin to adopt the conformation competent for hydrolysis, by inhibiting the hydrolysis step itself, or by further slowing down the product release steps. Due to the complicated regulation of the myosin motor activity, it is not always possible to unambiguously assign the inhibitory action of mutations to these four possibilities.

The P-loop residues are supposed to be necessary for nucleoside triphosphate binding, since this loop is found in all ATP or GTP hydrolyzing proteins (53). Apart from the structural role some P-loop residues may nevertheless play a catalytic role. This is proposed to be the case in the present work, in which Ser181 acts as a proton relay in the Ser181 mechanism that is found to be occupied. Indeed, when Ser181 is mutated to Ala, the enzymatic activity is halved (231). Mutating Ser181 to Thr does not recover the enzymatic activity (231) even though it is conceivable that the Thr hydroxyl group may play a similar role as the Ser hydroxyl group. However, due to the active site being closely packed it is possible that a Thr sidechain could not be positioned in a geometry appropriate for proton transfers and that in contrast Thr (that is more bulky than Ser) disrupts the balanced organization of the surrounding heavy atoms that serve as proton donors and acceptors. Thr186 is clearly important as a ligand coordinating the Mg^{2+} ion. However, no mutational data are available for this residue. Since perturbation analysis can only be made for residues not included in the QM region and Thr186 is included in the QM region in the present study, it is not possible to infer any further details on the role of Thr186 from the present simulations. A key role is played by Lys185 that is hydrogen-bonded to both a β- and a γ-oxygen of ATP in a fashion similar to Mg^{2+} (see comparison of crystal structures, Item 3 in Appendix A). When mutated in gizzard muscle myosin II, no enzymatic activity can be observed (231). However, other authors claim that a K185A mutant can catalyze hydrolysis (258). In the present work, Lys185 is found to lower the barrier height only in the Ser181 pathway (Fig. 13.46). Interestingly, it electrostatically stabilizes the product states in all three paths studied (Figs. 13.54, 13.55, and 13.56). Thus, apart of the possibility of its active participation in hydrolysis (as is discussed in Section 15.2.5) is may contribute to hydrolysis by pulling the hydrolysis equilibrium towards the product side.

Both the switch-1 and the switch-2 loops must be in their closed conformation for hydrol-

ysis to proceed. The C/C conformation is maintained by a salt-bridge between Arg238 (switch-1) and Glu459 (switch-2). When mutating Arg238 to Ala (226), Cys (226), His (226), Ile (231), or Glu (230), no or significantly reduced enzymatic activity is observed. Similarly, when mutating Glu459 into Ala (227), Val (225), or Arg (230) enzymatic activity is severely hindered. However, the single mutant R238K (231) and the double mutant R238/E459R (230) maintain their capacity for hydrolysis. This indicates that formation of the salt bridge, but not the detailed chemical nature of the salt-bridge partners is critical for hydrolysis in *Dictyostelium discoideum* myosin II. Since in both the R238K and the double mutant the local hydrogen pattern is necessarily affected, it seems unlikely that Arg238 and Glu459 play an essential role in positioning the attacking water, as has been suggested based on mutational data for rabbit skeletal myosin II (258). The rabbit myosin double mutant R247E/E470R cannot recover the loss of basal ATPase rate (423) as could the corresponding *Dictyostelium discoideum* myosin double mutant R238E/E459R (230). This may be explained by assuming that myosin II motors from different organisms utilize slightly different strategies for ATP hydrolysis, even though the nucleotide binding pocket residues are extremely conserved across species (see Section 3.2.3). In the present work which is based on the crystal structure of *Dictyostelium discoideum* myosin II, the unfavorable electrostatic contribution of Arg238 to both the barrier heights and the reaction energies is compensated by the favorable contribution of Glu459. Thus, the overall contribution of the salt-bridge residues to catalysis is small, making an active participation in catalyzing ATP hydrolysis unlikely.

The first residue in the switch-1 loop, Asn233, does play a role as evidenced by its halving the basal ATPase rate of myosin when mutated to Ala (226). It has been suggested to act by electrostatically stabilizing the negative charge on the bridge oxygen in the transition state (258), assuming dissociative-like mechanism similar to the mechanism suggested for the G-protein Ras. In Ras, this charge stabilization is exerted by an Arg residue (termed the "Arg-finger") from the GTPase activating protein (GAP) whose presence is necessary for hydrolysis to proceed. That indeed an Asn can take a role equivalent to the Arg-finger has recently been shown for Rap1GAP, *i.e.* the GTPase-activating protein for Rap, a G-protein belonging to the Ras superfamily (424). However, the present work puts the assumption into question that a similar mechanism takes place in myosin. A residue can act by stabilizing a negative charge only if significant negative charge is accumulated in the transition state. This is, however, not the case in myosin, as indicated by the Mulliken charge analyses of the present work (Section 13.5). In addition, the perturbation analyses show that Asn233 stabilizes the transition state only in the Ser181 pathway, whereas it is slightly destabilizing in the direct and Ser236 pathways. However, it clearly stabilizes the product states of all three pathways. Thus, Asn233 may act similarly to Lys185 by pulling the hydrolysis equilibrium towards the product side. The two other Asn residues of the switch-1 loop, Asn234 and Asn235, contribute only marginally to both barrier heights and reaction energies in all three pathways studied. This coincides with the finding that mutating Asn235 to Ala alters the basal ATPase rate of myosin to only 75% of the wild-

type activity (226), and mutating Asn235 to Ile even elevates the activity to ∼150% of the wild-type activity (231).

Both Ser residues of the switch-1 loop have been proposed to be involved in catalyzing ATP hydrolysis. Ser236 may serve as an intermediate proton shuttle in a fashion similar to Ser181. Thus, mutations should reduce the ATPase activity, which is indeed the case. However, different experimental studies disagree in the extent of loss of activity when mutating Ser236 to Ala. Using a *Dictyostelium discoideum* myosin II constructs, the basal ATPase rate of myosin is reduced to ∼70% of its wild-type activity (226), whereas when using gizzard muscle myosin it is reduced to ∼25% (231). In the latter case, some activity can be recovered when replacing Ser236 with a Thr with the S236T mutant showing 80% of the wild-type activity. In the present study, Ser236 in contained in the quantum mechanical region, thus making it impossible to perform a perturbation analysis. The fact that the Ser236 pathway is found to be isoenergetic to the Ser181 path in the present study nevertheless coincides with approximately comparable loss of ATPase activity upon mutating Ser236 or Ser181 to Ala. Ser237 coordinates the Mg^{2+} ion. If mutated to Ala this coordination is made impossible. Thus, it seems likely that a S237A mutant would be hampered in closure of the switch-1 loop upon nucleotide binding that is accompanied by formation of the hexacoordinate Mg^{2+} complex, thereby slowing down the basal ATPase activity to the observed ∼30% of the wild-type activity (226). Thus, it is expected that such a mutant would not significantly slow down the rate of the hydrolysis step itself. Since however, Ser237 is also treated quantum mechanically in the present work, thus making it impossible to study its electrostatic contributions by means of perturbation analysis, the impact of Ser237 on the hydrolysis step cannot be investigated here.

Several residues seem to not participate in the hydrolysis step as indicated by an increase in basal ATPase rate upon mutation to Ala or Leu. These are Ile455 (227), Ser456 (227; 233), and Phe458 (227). Indeed, Ile455 and Phe458 do not significantly contribute to either the reaction barrier heights or the reaction energies of any of the studied pathways. Ser456 exhibits an unfavorable contribution to both barrier heights and reaction energies in all pathways. Thus, it is possible that the increased basal ATPase rate can at least partly be explained by an accelerated ATPase step due to removal of the inhibitory effect of Ser456. Asp454 contributes slightly favorable to both reaction barriers and reaction energies in all three pathways, except in the transition state of the Ser181 path, where its contribution is slightly unfavorable. In contrast, in the mutational analysis the activity of the D454A mutant is reduced to ∼35% of the wild-type activity (227). Thus, the loss of activity is almost as pronounced as for the E459A mutant. To explain the discrepancies between the results from the perturbation analysis of the present study and the results from the mutational experiments further investigations would be necessary.

The backbone amide NH group of Gly457 group forms a hydrogen bond to $O_{\gamma 1}$ only in the C/C structure of myosin, an interaction that has been proposed to be crucial for hydrolysis to proceed (38; 258). In the present path calculations, this hydrogen bond is maintained

throughout the whole paths with only insignificant movements of $O_{\gamma 1}$. Since the hydrogen bond is apparently unaffected by the hydrolysis itself, it seems unlikely that formation of this hydrogen bond induces the hydrolysis event as has previously been suggested (77). Nevertheless, its presence seems to be necessary for hydrolysis to occur, as is revealed by the favorable contributions of Gly457 to both barrier heights and reaction energies in all three pathways studied.

15.2.5 Network hypothesis for myosin catalysis of ATP (question 4)

The question arises as to whether it is possible to synthesize the findings of the present study with previous proposals in a new hypothesis on the mechanism of ATP hydrolysis in myosin that is capable of explaining both the experimental observations and the findings of the present study. Several features of the ATP hydrolysis reaction paths in myosin as obtained here seem remarkable. Firstly, all reaction pathways studied here have approximately the same barriers and are thus approximately equally likely to be populated. It is therefore not possible to speak of *the* reaction pathway in myosin since several possible reaction channels could be taken. Secondly, all proton transfer events are found to occur without visible barrier. This may be due to the CPR algorithm used to refine the reaction pathways. CPR locates the highest saddle points that connect pre-defined minima on the potential energy landscape. However, it does not refine transient path points so as to consistently follow the intrinsic reaction coordinate. The reaction paths as refined with CPR can in principle be smoothed by further optimization using the SCM command of the TReK module (315) in CHARMM, thereby revealing further bumps in the energy profile that correspond to minor barriers. Since such a path smoothing is computationally costly, it has been omitted here. Nevertheless, the additional barriers that may show up are *per definitionem* negligibly small compared to the reaction barrier refined with CPR. Indeed, proton transfers in alkaline solutions that lead to the transport of hydrated hydroxide in aqueous solution occurs *via* a free energy barrier of about 3 kcal/mol (396). Thus, even if a barrier must be crossed in a proton transfer event (as is expected), this barrier is low and will not significantly hinder the proton transfer. Thirdly, the geometries of the transition states of the three pathways are nearly identical when considering only the heavy atoms (Fig. 15.1). Exceptions are the positions of the helper water and Ser181 sidechain atoms in the Ser181 pathway. However, the heavy atoms of the triphosphate moiety of ATP and the attacking oxygen overlap remarkably well, as is also indicated by the distances $P_\gamma : O^a$ and $P_\gamma : O_{\beta\gamma}$ being identical in all three paths (see Section 13.1.4).

This inspires the following *Network Hypothesis* on the mechanism of ATP hydrolysis in myosin: Myosin maintains a preorganized, dynamic hydrogen-bonding network in the active

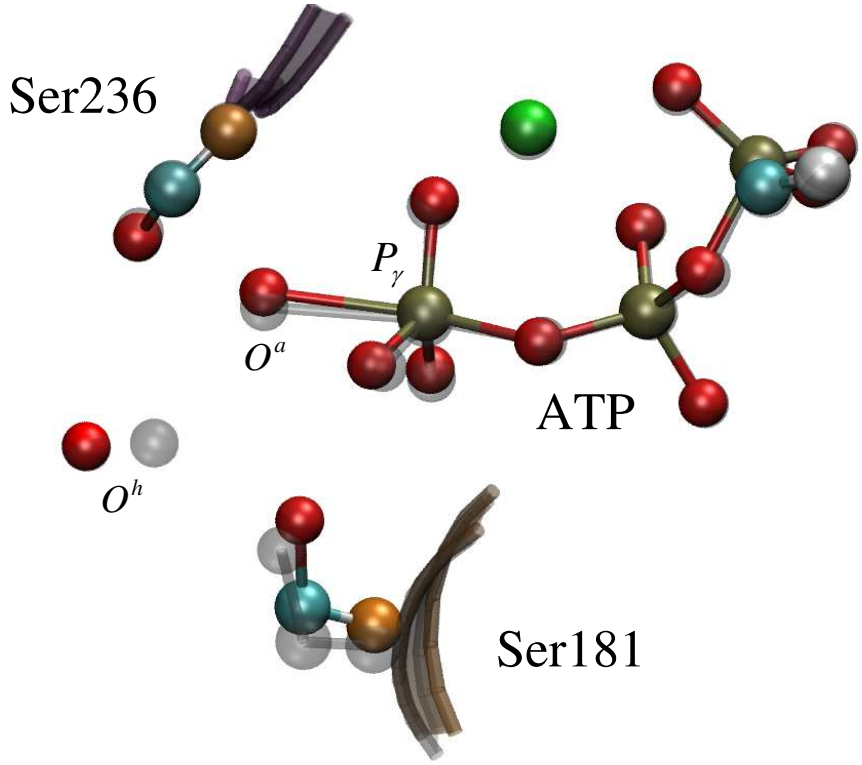

Figure 15.1: Overlap of the QM[HF/6-31G(d,p)]/MM/NUCS-optimized transition state structures of the direct (colored by atom), Ser181 (gray), and Ser236 pathways (tan, almost invisible since the atoms occupy practically identical positions as in the direct path). Only the heavy atoms are shown.

site when it adopts the catalysis-competent C/C conformation. Depending on the specific protonation state and conformation of the binding-pocket-forming residues at a given time, ATP hydrolysis proceeds *via* different reaction channels. Those different reaction paths are equally valid in explaining the details of ATP hydrolysis in myosin.

The dynamic network consists of heavy atoms that can carry one or more protons (such as the water molecules adjacent to the γ-phosphate moiety including the attacking water, the sidechains of Ser181, Lys185, Ser236, Arg238, Ser456, Gly457 (main chain), and Glu459 and the γ-phosphate moiety of ATP) and of as many protons as necessary to balance the electrostatic charges. The protons are rapidly exchanged between the donors and acceptors *via* negligible energy barriers in a fashion similar to the proton transfer events taking place in bulk water. Thus, the somewhat delocalized hydrogens provide a mean field in which the heavy atom motion takes place. The progress of the ATP hydrolysis is determined predominantly by the movements of the heavy atoms involved, in particular the movement of P_γ. In principle, a number of different paths are conceivable, depending on the specific

configuration and protonation pattern of the myosin active site. Each such "macropath" can be realized by a number of different proton transfer pathways. A specific ATP hydrolysis pathway that takes place in a given protonation pattern of the active site, within a single conformation of the protein, and with specific proton transfer pathways can be viewed as one possible realization of the observable hydrolysis event (Fig. 15.2). In the present work, three different such "micropaths" have been studied. However, a number of additional "micropaths" might, in principle, exist. The hitherto-suggested reaction mechanisms for ATP hydrolysis in myosin, *i.e.*, the direct or Ser236 relay mechanisms (57; 252), the Ser181 relay mechanism or Lys185 as a general base (253), Lys185 as a general acid (256), and the two-water-hypothesis (258), can be considered to correspond to different possible "micropaths". The sharp distinction between associative and dissociative mechanisms is somewhat blurred in the network hypothesis, since some "macropaths" may follow a route with more associative character while others may be more on the dissociative side of the continuum of possible reaction mechanisms.

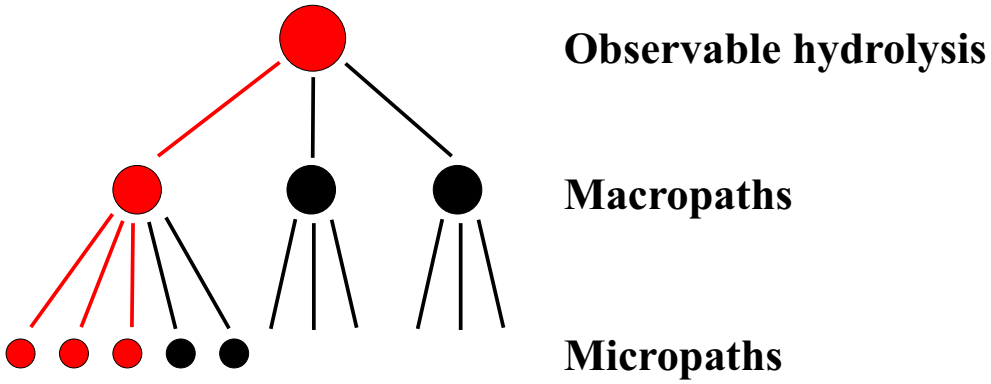

Figure 15.2: Path hierarchy. The experimentally observable ATP hydrolysis can be realized by a number of macropaths characterized by a specific pattern of motion of the heavy atoms involved in the hydrolysis. Each macropath can be realized by a number of micropaths characterized by specific proton transfer pathways. The branch of the resulting hierarchy tree corresponding to the three hydrolysis pathways investigated in this study is marked in red.

If the proposed dynamic network hypothesis holds, then the ATP hydrolysis mechanism in myosin can be expected to proceed in a similar fashion as ATP hydrolysis in bulk water. Since so far no definite details on the mechanism of nucleotide hydrolysis in aqueous solution are known (compare Section 4.2.3) it is, however, not possible to comment on the degree of similarity between the uncatalyzed and the catalyzed processes.

The network hypothesis decouples the ATP hydrolysis from the necessity of myosin to capture a well-defined number of water molecules in the active site during the return

stroke (C/O to C/C transition). In the proposed dynamic network scenario, any number of captured water molecules could possibly be involved in the dynamic hydrogen bonding network and thus contribute to hydrolysis. However, due to steric factors is seems likely that either only one or two water molecules could be captured in addition to the attacking water molecule in the cavity formed between the γ-phosphate moiety of ATP and the switch-1 and switch-2 loops. These water molecules are expected to possess considerable degree of mobility, because they can be viewed as key mediators of proton exchange among the heavy atoms of the dynamic hydrogen bonding network. That the crystallographically-observed water molecules in this region are indeed mobile is indicated experimentally by their high temperature factors or even their absence in the available crystal structures (see Item 5 in Section 10.2). In the present study, both the helper water and the attacking water molecules can reorient within the active site without energetic cost, enabling the system to proceed from a single reactant structure *via* three different water activation routes to three different product structures.

It is currently not possible to test the proposed network hypothesis using computer simulation techniques since such a testing would require determination of all possible reaction paths in all possible protonation patterns, preferably while averaging over many protein conformation in a free-energy perturbation approach. This, however, is computationally too costly to be done with *ab initio* quantum techniques and would thus require the use of semi-empirical quantum techniques that have so far not been shown to be able to yield reliable geometries and energies for phosphoanhydride reactions (see Section 15.1). However, some rationalization could be possible by calculating the pK_a values of the residues involved. Knowledge of the pK_a values would give insight into the equilibria associated with proton transfer events. However, such an argumentation must be made with care, since it is possible that a Ser residue will always carry a single proton on its sidechain even when it participates in a proton hopping chain. Thus, the pK_a value that reflects the likelihood of a protonation or deprotonation event cannot be used as a criterion of likelihood of a simultaneous protonation/deprotonation process under such circumstances.

It is possible to explain the findings of the mutational studies within the framework of the dynamic network hypothesis for myosin catalysis of ATP hydrolysis by assigning functional roles to the residues in the nucleotide binding pocket within the network hypothesis. Some may be responsible to put together and maintain the enzymatic environment in such a way that a dynamic network best suitable to support hydrolysis can emerge (Category I). A second group of residues (Category II) may position ATP in the binding pocket such that ATP sits rather immobilized in the pocket and can be acted upon by the members of the third group of residues (Category III) that form the active participants in the proton exchange processes and are thus the key players in the chemical hydrolysis event. The water molecules (including the attacking water) that are trapped in the cavity adjacent to the γ-phosphate moiety of ATP in the C/C conformation of myosin form the central core of the dynamic proton exchange network. This is due on the one hand to their ability to donate and accept two protons each and on the other hand to their mobility within

the cavity that makes them ideal proton shuttles able to mediate proton transfers between protein residues that participate in the network.

The salt-bridge residues Arg238 and Glu459 are in Category I. If mutated such that salt-bridge formation is made impossible ATPase activity is lost (e.g., R238A (226), R238C (226), R238H (226), R238I (231), R238E (230), E459A (227), E459V (225), and E459R (230)), whereas in mutants that maintain the capacity of salt-bridge formation, ATPase activity is maintained as well (*i.e.*, R238K (231) and the double mutant R238E/E459R (230)). Thus, salt-bridge formation seems to be necessary to firstly sequester the reaction site from other parts of the molecular motor, thereby defining the boundaries of the dynamic network, and, secondly, to trap a number of water molecules within the reactive area. In addition, Arg238 and Glu459 may also be involved in proton exchanges, albeit only at the boundary of the network. Therefore, the active participation of Arg238 and Glu459 in proton exchanges seems to be of minor importance. This agrees with the results of the present perturbation analysis that indicate that the combined effects of Arg238 and Glu459 on the reaction barrier height are small. The P-loop residues are expected to belong to the ATP-positioning residues (Category II) and to thus contribute indirectly to assembling the dynamic network.

Mutation of any of the key residues of the network (Category III), *i.e.*, Ser181, Ser236, or Lys185, into Ala would be expected to reduce the number of possible micropaths, thus leading to a reduced but not abolished ATPase activity, as is observed for Ser181 (231) and Ser236 (226; 231), and is suggested for Lys185 (258). The Ser residues participate in the proton exchange network *via* their sidechain hydroxyl group. Thus, in principle, loss of activity upon mutation to Ala should be reduced when mutating Ser into Thr. This is indeed observed for Ser236, but not for Ser181 (231). This may be explained by the observation that the sidechain of Ser181 can interact with only one water molecule of the water cluster, whereas Ser236 points directly into the cluster and can interact with two water molecules. In position 181, the hydroxyl group of the Thr may be inaccessible to hydrogen donor/acceptor partners of the water cluster, thus blocking the reaction paths accessible with a Ser in position 181. In contrast, in position 236 the structure of the water cluster may simply reorganize, thus facilitating proton transfers to and from the hydroxyl group of Thr.

Residues that seemingly cannot be assigned specifically to one of the three categories introduced above may not be directly involved in the hydrolysis step itself but play key roles in mediating the communication between the different sites of myosin. Examples are Ser237 that may contribute to opening of the trap door for product release (discussed in detail in Section 15.3) or Ser456, that has been shown to decouple the ATPase cycle from the motor cycle when mutated to Ala (233), or Gly457 in the switch-2 loop that may couple the local structural changes during ATP hydrolysis to the movement of the lever arm (92).

An experimental test of the network hypothesis would require to substitute the water molecules central to the network hypothesis. This could in principle be realized by sol-

vating a wild-type construct of the myosin II motor in methanol instead of in water and measuring the ATPase activity. However, it may be difficult to assign the expected observable changes in reaction rates specifically to the ATP hydrolysis step, since a change of solvent will also change the dynamics of the conformational transitions and thus influence all steps of the ATPase cycle. Thus, one would probably have to measure the kinetics along the whole ATPase cycle, as has been done for the W501+ mutant of *Dictyostelium discoideum* (209). In addition, the chemical reaction kinetics of hydrolysis and methanolysis of ATP are presumably significantly different from each other. The extent of dissimilarity is, however, unknown, thus making it difficult to separate changes in the kinetics of the hydrolysis step (once these would be known from experiment) between intrinsic changes of the reaction channels of the uncatalyzed reactions and kinetic changes due to a differences in the dynamic network environment within myosin.

The network hypothesis implies that ATP hydrolysis in myosin is accompanied by a significant number of proton transfer events. Thus, it seems possible that a kinetic isotope effect could be measured, if proton transfers are found to contribute to the rate-limiting step. In addition, it is in principle possible that tunneling effects[3] need to be considered if experimental rates shall be accurately predicted by computational methods.

15.2.6 Role of Mg^{2+} (question 5)

Mg^{2+} may both be catalytically active and help in positioning the rather flexible triphosphate moiety of ATP in a conformation suitable for hydrolysis. The latter has been proposed in the case of 3-phosphoglycerate kinase (426) and it seems likely that Mg^{2+} is indeed required for this positioning, since without Mg^{2+} hydrolysis cannot take place. However, if the role of Mg^{2+} were only a structural one then it should be possible to exchange Mg^{2+} with Ca^{2+} without disrupting the motor activity of myosin. Indeed, when Ca^{2+} is used, the ATPase activity of myosin as measured by the basal ATPase rate is not only not alleviated but even increased (203; 226). Thus, it seems likely that the cation does indeed contribute catalytically to the ATP hydrolysis reaction. However, when Ca^{2+} is used instead of Mg^{2+}, the force production is decreased (213). Thus, the cation seems to also be involved in the communication between different sites of the myosin motor (discussed in Section 15.3).

The formation of the coordination bonds between Mg^{2+} and ATP^{4-} introduces an asymmetry in the distances $d(P_\gamma{:}O_{\beta\gamma})$ and $d(O_{\beta\gamma}{:}P_\beta)$. An equivalent asymmetry has been observed for methyltriphosphate (168). Even in pyrophosphate, being a symmetric molecule, a corresponding activation of a $d(P{:}O_{bridge})$ bond by Mg^{2+} has been observed, if the Mg^{2+} is coordinated by one oxygen of one phosphate moiety and two oxygens of the other phosphate moiety in pyrophosphate (165). This asymmetry of the $d(P : O_{bridge})$ bond lengths indi-

[3]Tunneling effects do play a non-negligible role in determining the free energy barrier of reaction involving proton, hydrogen, or hydride transfers. Tunneling effects occur both in solution and in enzyme environment, thus they probably do not contribute to the catalytic power of enzymes (425).

cates that Mg^{2+} contributes to activate the d($P_\gamma : O_{\beta\gamma}$) bond for cleavage both in myosin and in aqueous solution. Thus, a possible catalytic role of Mg^{2+} can be seen in activating the bond that will be cleaved both in solution and in the enzymatic environment.

Moreover, the present work indicates that the Mg^{2+} cation moves towards the ADP.P_i moieties upon decay of the pentacovalent phosphorane transition state (Sections 13.1.2). This adds to the dissociative character of the paths studied, because when the cation piles in between the γ-phosphate moiety and ADP it contributes to the dissociation by stabilizing the separated state. Another possible catalytic role of the Mg^{2+} ion can therefore be seen in pushing the reaction in the forward direction by triggering the separation of the reaction products. This contributes to lowering the energy of the product state. According to the Hammond postulate, stabilization of the product state shifts the transition state towards the reactant state (427). Indeed, the transition state is seen here to be structurally and electrostatically close to the reactant state. Given that the energy of the unhydrolyzed ATP is much lower than the anion pair in proximity, this is one way for the protein to lower the activation barrier.

The pushing effect of the cation might be more pronounced when using Ca^{2+} instead of Mg^{2+}, because the ionic radius of Ca^{2+} is larger than that of Mg^{2+}, and thus might more efficiently separate the P_i from the ADP. Thus, the Ca-ATPase reaction is expected to be more dissociative in character than the Mg-ATPase reaction. If this is true, more negative charge would accumulate on the $\beta\gamma$-bridge oxygen in the transition state of the Ca-ATPase reaction than in the Mg-ATPase reaction. This negative charge would have to be stabilized by the protein environment. A possible candidate is Asn233 that has been suggested to fulfill exactly this role in the Mg-ATPase cycle (258). Indeed, the Ca-ATPase activity is almost abolished in the N233A mutant, in contrast to the Mg-ATPase activity for the same mutant (226). Thus, it seems possible that different reaction channels would be accessible when using different metal cations as cofactors. If the metal ion cofactor indeed helps to advance the reaction by product separation and if the effect of Ca^{2+} is more pronounced than the effect of Mg^{2+} due to the difference in ionic radii, this may provide an explanation of the increase in ATPase rate upon replacement of Mg^{2+} with Ca^{2+}. This hypothesis could be tested by repeating the reaction simulations of the current work while replacing Mg^{2+} by Ca^{2+}.

Apart from the steric effects discussed above, Mg^{2+} may act catalytically by affecting the charge distribution in the active site. That this is indeed the case is revealed by the Mulliken charge analyses (see Section 13.5) that show that the positive charge of the dication is distributed among the Mg-coordinating ligands. In particular, the charges on ATP and the attacking water are rearranged such that the charge pattern of the product state is already preformed in the reaction state. This charge preorganization is presumably not possible in the absence of a divalent cation, which may explain why hydrolysis does not occur without a metal cofactor (211; 212).

15.2.7 Enzymatic strategy of myosin

In lowering the barrier of ATP hydrolysis from 28.9 – 29.3 kcal/mol (132; 135) in aqueous solution to about 14.5 kcal/mol in the myosin active site, myosin functions as an enzyme. Enzymes can, in principle, achieve catalysis by destabilizing the reactant state and/or stabilizing the rate-limiting transition state. In the present work it is found that the residues in direct proximity to the reaction do not significantly contribute to lowering the barrier. Therefore, no clear stabilization of the transition state is observed. In contrast, factors that may contribute to reactant state destabilization are observed. These are the preformation of the charge pattern of the transition state in the reactant state and the slight geometric deformation of ATP into the direction of the trigonal bipyramidal transition-state geometry. Thus, myosin as an enzyme is likely to act mainly by destabilization of the reactant state.

15.2.8 Comparison of the hydrolysis mechanisms of myosin, Ras, and F_1-ATPase (question 6)

In contrast to Ras, neither myosin nor F_1-ATPase require an additional protein equivalent to GAP to assemble the hydrolysis machinery. However, in all three proteins the hydrolysis machinery is functional only transiently. Thus, certain structural requirements must be met before hydrolysis can efficiently take place. Myosin must adopt the C/C conformation, whereas F_1-ATPase hydrolyses ATP only in the β subunit that adopts the β_{TP} structure. Ras needs the Arg-finger from GAP before it can efficiently hydrolyze GTP (179; 180; 181). For F_1-ATPase, a similar Arg-finger motif is provided by the α-subunit that neighbors the catalytic β-subunit (416). Differences occur, however, in the proposed degree of associativity of the hydrolysis mechanisms used by the three proteins. In myosin it seem likely that the Mg-ATPase reaction proceeds more associative-like than dissociative-like (Section 15.2.2), whereas in Ras a more dissociative-like mechanism appears probable (see Section 4.3.2). It is, however, difficult to distinctly assign the degree of associativity to any of the mechanisms under question. This is evident by the case of myosin, where in the current work a geometrically clearly associative mechanism is found that nevertheless exhibits certain features of a dissociative mechanism (such as the lack of charge shifts or the motion of the Mg^{2+} ion). Thus, it seems probably that a delicate equilibrium is created between a more associative-like and a more dissociative-like mechanism that may be shifted into one or the other direction by slight changes in the environment. It is likely that at the same time other, more robust features of the hydrolysis mechanism are maintained. This would make the discussion on the degree of associativity misleading, and attention should be directed more towards the invariant features of enzyme-catalyzed triphosphate hydrolysis.

Such an invariant feature may be found in the role of the water cluster that has been suggested in the present work to be the central core of a dynamic water network that fosters hydrolysis. A similar water cluster is observed in F_1-ATPase, where different possible

mechanisms involving different water molecules of the water cluster have been studied (415; 416). Thus, it seems possible that indeed a similar dynamic water network mechanism may be functional in F_1-ATPase as has been proposed here for myosin. An additional similarity between the mechanisms in myosin and F_1-ATPase lies in that the Arg-finger in F_1-ATPase stabilizes the hydrolysis product state in a fashion similar to the product stabilization by Asn233 observed in this work (Figs. 13.57, 13.58, and 13.59). Thus, it may be possible that the Arg- or Asn-finger motif is involved in regulation and fine-tuning of the motor activity in the motor proteins myosin and F_1-ATPase by coupling the post-hydrolysis state to subsequent states along the motor cycle rather than in activating the hydrolysis itself as is the case for Ras.

The dynamic water network hypothesis implies that nucleoside-triphosphate-hydrolyzing enzymes do not significantly alter the mechanism of the uncatalyzed reaction. However, to validate this statement detailed comparisons would be required between the mechanisms of triphosphate hydrolysis in water and in enzymatic environment. This can currently not be done due to the lack of knowledge of details of the uncatalyzed hydrolysis reaction.

Nevertheless, the dynamic network hypothesis exhibits certain features that make it a reasonable hypothesis worth of further consideration to also explain features of the ATP hydrolysis events catalyzed by proteins other than myosin. A number of features inherent to the network hypothesis are generally valid for motor proteins. For example, according to the network hypothesisonly a minimum of mechanistic changes are required in going from the uncatalyzed to the catalyzed reaction. In aqueous solution, ATP is surrounded by the dynamic water structure formed by the bulk solvent. By restricting the size of the dynamic water network and optimal positioning of the water cluster with respect to ATP the enzyme may block unproductive attacks while still allowing productive attacks. Thus, the motor protein could be thought of being optimally designed for its motor function with minimum effort to optimize the hydrolysis mechanism for speed-up of the uncatalyzed reaction. Indeed, the efficiency of motor proteins as enzymes is not as high as the efficiency of proteins that exhibit enzymatic function only. Moreover, allowing many possible reaction routes introduces significant redundancy that leads to robustness of the hydrolysis event. Since the nucleoside triphosphate hydrolysis step represents a key event in maintaining the function of motor proteins and thus of a large number of proteins that are responsible to uphold the functioning of the mechanical events in cells and organisms, it seems necessary to minimize the possibility of disruptions of this important reaction. The dynamic network hypothesis represents a possible explanatory framework how the thus required robustness of the hydrolysis event is achieved.

15.3 Implications for the myosin motor cycle

In myosin, the ATPase activity is tightly coupled to its motor cycle by transmission of events at the ATP binding site to events at the other functional sites of myosin. In particular, the small local structural changes associated with ATP hydrolysis in the active site are amplified into the ∼60° rotation of the lever arm and the state of the bound nucleotide modulates the affinity of myosin for actin (see Section 3.6).

Apart from insight into the chemical nature of the ATP hydrolysis mechanism in myosin (Section 15.2.5), the present work indicates the existence of a number of intermediary states along the actomyosin cycle based on which a scenario of the functioning of the actomyosin cycle can be drawn. In particular, speculations on the mechanism of phosphate release are possible. The breaking of the coordination bond between Mg^{2+} and Ser237 implies that the switch-1 loop is no longer immobilized by this strong interaction. Just as formation of this coordination bond accompanies the closure of the switch-1 loop upon ATP binding, breaking of this bond can be considered a prerequisite for the opening of the switch-1 loop after hydrolysis. The present simulations suggest that, in the hitherto unkown (428) O/C conformation of myosin, product release is likely to occur *via* an exit route that opens by movement of switch-1, a mechanism that has been termed the "trap door" mechanism (76), rather than along an exit route in which P_i is released into the 50 kDa cleft by breaking of the Arg238/Glu459 salt-bridge (the "back door" mechanism) (94). Since motion of the switch-1 loop is believed to be coupled to cleft closure (95; 96), and thereby modulation of the actin affinity (41; 93), weakening of the Mg^{2+}:switch-1 interaction, as is observed here, is a key event that couples the hydrolysis step to the conformational changes in the post-hydrolysis steps of the actomyosin cycle. This coupling might occur *via* a cooperative mechanism, in which small movements of switch-1 would facilitate the partial cleft closure that induces initial actin binding (the so-called "weak binding" state of actomyosin) whose strengthening induces a further cleft closing that propagates back to the switch-1 loop, which fully opens to allow the release of P_i after actin is more strongly bound (428, compare Fig. 1 therein).

The five-fold coordination of Mg^{2+} in the posthydrolysis state observed in the present work can be viewed as an intermediate state that is likely to react into a state in which the coordination bond between Mg^{2+} and P_i is broken. Since the breakage of the Mg^{2+}:Ser237 coordination bond is already observed in the present work, the three different structures of the posthydrolysis state corresponding to the product states of the three hydrolysis pathways investigated here provide a first indication that the proposed O/C state indeed exists. Experimentally, it may not be possible to observe the putative intermediary state with four-fold Mg^{2+} coordination, since it is to be expected that this coordination pattern is very short-lived. It is known that the water molecules in the active site can rapidly exchange with the bulk solvent, thus indicating that bulk solvent molecules can easily stream into the active site. Partial opening of the switch-1 loop will further facilitate the streaming of

water into the vicinity of Mg^{2+}. Thus, it is to be expected that the coordination sphere of Mg^{2+} will rapidly be filled to its preferred coordination number of six by coordination of two additional water molecules.

Postulating an intermediary state with four-fold coordinated Mg^{2+} allows to draw the following scenario that provides a possible succession of events along the actomyosin cycle. Beginning at the post-hydrolysis state of the ATPase cycle, the coordination bond between Mg^{2+} and the switch-1 loop is broken, followed by breakage of the coordination bond between Mg^{2+} and P_i. Myosin is still in its C/C conformation and not bound to actin. Switch-1 starts to swing from its closed into its open position. This local conformational transition is transmitted *via* cleft closure to the actin binding region such that the affinity of myosin for actin is increased. Formation of the actomyosin complex induces complete cleft closure that backpropagates to the switch-1 loop that can now adopt its open position. Myosin is now in its O/C conformation and actin is bound.

The complete swinging open of the switch-1 loop opens the exit route for phosphate release through which P_i can easily escape, since it is not coordinated to Mg^{2+} any more. At the same time, the switch-1 movement loosens the clamp provided by the Arg238-Glu459 salt-bridge that has prevented the strained myosin spring (see below) from snapping into its released state. As soon as the clamp is removed, switch-2 swings into its open state, which at this stage of the ATPase cycle is energetically favored over the closed state. Associated with the swinging open of switch-2, the lever arm rotates from its UP into its DOWN position, while actin in bound. After this power stroke, myosin is in its O/O conformation and actin is tightly bound (rigor state). Both the power stroke and the product release events are ultimately triggered by the same event, namely the swinging open of the switch-1 loop. Thus, they occur simultaneously in the actomyosin cycle, although they are not causatively connected with each other.

ATP binding dissociates the actomyosin complex by inducing the switch-1 loop to swing from its open into its closed position, a movement that is coupled to cleft opening and thus to loss of actin binding affinity. The resulting conformational state of myosin is the C/O state. The binding event causes strain in myosin that is compensated for by a favorable binding energy. This agrees with the thermodynamic free energy profile along the ATPase cycle of myosin (Fig. 4.4) that shows that the energy of the system is reduced significantly upon binding, whereas later events do not decrease the energy as markedly. The subsequent isoenergetic swinging of switch-2 from its open into its closed position puts myosin into its C/C conformation competent for hydrolysis. The hydrolysis event itself occurs as soon as it is made possible by the preceding conformational changes of the protein environment and does not distinctly lead to a decrease in free energy of the overall system. After hydrolysis is completed, the system is ready for the next cycle.

Thus, the ATP hydrolysis step itself is not used to provide energy that is *directly* used to perform the mechanical work of the lever-arm rotation, but rather as a chemical clock that coordinates the events along the motor cycle by fine-tuning the succession of conformational

states of the switch-1 and switch-2 loops and thus the opening and closing of the 50kDa cleft and associated actin affinity as well as the position of the lever arm. Energy transduction proceeds *via* intermediate storage of energy in the myosin head domain that thus functions as a spring that is spanned by ATP binding and whose relaxation can be equated with the power stroke.

To ensure efficacy in force production by myosin it is necessary to prevent premature and thus ineffective product release. This makes it necessary to keep P_i and ADP bound after hydrolysis occurred until structural rearrangements in the myosin motor have prepared the system to undergo the power stroke. Since both P_i and ADP carry negative charges and thus repel each other, it is difficult to keep them both bound in the active site of myosin where they are in close contact to each other. This task is achieved by the strong stabilization of the product state by means of the QM/MM electrostatic interactions. Only half of this stabilization energy can be assigned to the residues in direct vicinity of the binding site (see Section 13.4). Thus, long-range electrostatic interactions are crucial in controlling the events along the myosin or actomyosin cycle. Further analysis is required to gain more detailed insight into the nature of the long-range electrostatic stabilization of the product state and the necessary modulation mechanisms that eventually allow the products to be released.

In the above scenario product release is triggered through the movement of switch-1. This supports the trap door mechanism rather than the back door mechanism for the opening of the phosphate exit pathway (question 7, compare Section 3.6). If indeed product release is due only to movement of switch-1 and not due to movement of switch-2, and if indeed the lever arm movement and associated power stroke is due only to movement of switch-2 and not due to movement of switch-1, then it should be possible to decouple the catalytic cycle from the motor cycle by introduction of an appropriate single point mutation. The mutation should be such that ATP hydrolysis still occurs but the motor activity is hindered. Thus, in the mutant formation of the C/C conformation and uninterrupted swinging of switch-1 must be possible, while at the same time hindering the movement of switch-2. Such a mutant indeed exists: when mutating Ser456 into Leu the capability of myosin alone to hydrolyze ATP is unaffected as shown by a basal ATPase rate that is even elevated compared to the wild-type, whereas the catalytic cycle and the motor cycle are decoupled from each other (233).

Within the suggested scenario the differences between the actomyosin cycles with Mg^{2+} and Ca^{2+} as metal cofactors can be explained. The ability of Mg^{2+} to coordinate only four ligands is central to the coupling of the ATPase activity and the motor cycle, because the intermediary four-fold coordination is necessary to allow both the motion of the switch-1 loop and to dissociate P_i from the cation to facilitate phosphate release. This provides a possible explanation why ATPase and motor activities are partially decoupled when replacing Mg^{2+} by Ca^{2+}. Ca^{2+} cannot form stable complexes when four-fold coordinated, thus the putative intermediary state cannot be formed and the coupling mechanism is at

least partially disrupted.

Five- or four-coordinated Mg^{2+} complexes are not unusual, as they are observed in ∼10% and 23% of Mg-containing structures in the Cambridge Crystallographic Databank, compared to 60% with coordination number 6 (429). The hypothesis that intermediary fourfold metal coordination is required, is supported by measurements of intermediate solvent/phosphate isotope exchange in presence of a number of divalent cations (430). No exchange is observed with Ca^{2+}, Sr^{2+}, or Ba^{2+}, all ions that do not allow for the coordination number 4 (429), and thus for intermediate dissociation of the metal:P_i coordination bond. In contrast, exchange is observed in presence of Co^{2+}, Ni^{2+}, or Mn^{2+}, all ions that allow the coordination numbers 6 and 4 (429).

Direct evidence for the existence of a tetra-coordinate metal intermediate in the myosin ATPase cycle might be obtained from time-resolved EPR spectroscopy using Mn^{2+} as a metal cofactor. When a ^{17}O directly coordinates to Mn^{2+} the EPR signal for Mn^{2+} is inhomogeneous broadened. This has been used to investigate the myosin.Mn.ADP complex (431). When using ATP with all terminal γ-oxygens labeled with ^{17}O a broadening of the EPR signal should be observed as long as a terminal γ-oxygen coordinates to Mn^{2+}. If the hypothesis holds that the P_i:metal coordination bond breaks and reforms during the reversible hydrolysis, then an oscillatory signal broadening should be observable in time-resolved EPR experiments in the initial phase of the experiment before solvent/phosphate isotope exchange occurs.

15.4 Concluding Remarks

The present work intended to contribute to a detailed understanding of ATP hydrolysis in the motor protein myosin II by determining and analyzing different possible reaction routes for this reaction that is ubiquitous in all living organisms. A number of qualitative statements could be made that may aid in gaining more insight into the mystery of the actomyosin motor cycle. Moreover, the simulations inspired the formulation of a new type of ATP hydrolysis mechanism (the network hypothesis) that may generally be employed by phosphate-hydrolyzing enzymes. In addition, the present work facilitated putting forward a new suggestion regarding the succession of events along the actomyosin motor cycle, regarding in particular the phosphate release step with associated conformational changes in the myosin motor domain.

Chapter 16

Summary: Discussion

The combined quantum mechanical/classical mechanical reaction path calculations on ATP hydrolysis in myosin presented in this thesis are the most accurate computational investigations of enzyme-catalyzed nucleoside triphosphate hydrolysis reported to date. Nonetheless, limitations in the available computational time (path refinements required 2 to 4 months of CPU time on eight 2.6 GHz Linux processors) did not allow the use of *ab inito* methods that include electron correlation effects, thereby limiting the accuracy that could be reached. This is reflected in the fact that the calculated barrier heights and reaction energies varied by ∼25 kcal/mol with the choice of the QM method. In addition to the energetic uncertainties associated with the QM methods used here, the deviations of the calculated energy barriers and reaction energies from the experimental values may be due to a variety of reasons, including the possibility that the crystal structure used may not describe a conformation that optimally supports hydrolysis, the neglect of the effects of the flexibility of the protein environment that is inherent to the calculation of minimum-energy paths, and the approximate treatment of solvent effects with Non-Uniform Charge Scaling. However, qualitative findings that are method-independent can be safely regarded as reliable results.

All three pathways investigated in the present thesis proceed in a one-step reaction without a stable reaction intermediate, which agrees with previous experimental findings. No obvious stabilization of the transition state was found. In contrast, the charge pattern of the product state was found to be already preformed in the reactant state. Moreover, the geometry of ATP is slightly distorted in the direction of a pentavalent phosphorane state when bound to myosin. This suggests that myosin acts as an enzyme by reactant state destabilization rather than by transition state stabilization. An associative route with some dissociative character is found, which also agrees with experimental findings that suggested a mechanism with more associative than dissociative character. Isotope exchange experiments established that the terminal γ-oxygens exchange equivalently with bulk solvent, whereas the β-oxygens do not exchange. Since hydrolysis occurs without passing through

a stable intermediate, either the γ-phosphate moiety in the prehydrolysis state or the P_i in the posthydrolysis state must therefore possess rotational freedom. In the the product state of present simulations the coordination bond between Mg^{2+} and the switch-1 loop is broken and the coordination bond between Mg^{2+} and P_i is weakened. This suggests that P_i can transiently dissociate from Mg^{2+}, allowing P_i to tumble in the active site and to echange its oxygens equivalently with bulk solvent. The β-phosphate remains permanently coordinated to Mg^{2+}, which prevents rotation and, therefore, exchange. The dissociation of the Mg^{2+}:P_i coordination bond is also a prerequisite for the eventual release of P_i.

The fact that all three water activation pathways investigated here were found to be isoenergetic and that the heavy atoms adopt a very similar geometry in all three transition states has inspired the formulation of a Network Hypothesis for ATP hydrolysis in myosin. Myosin maintains a preorganized, dynamic hydrogen-bonding network in the active site when it adopts the catalytically-active C/C conformation. This dynamical network consists of the γ-phosphate, water molecules adjacent to the γ-phosphate, and sidechains that can carry protons and serve as proton donors and/or acceptors. The protons can rapidly exchange between the donors and acceptors, thereby creating specific hydrogen-bonding patterns in the active site and allowing for a number of different water activation mechanisms. For several such hydrogen-bonding patterns, ATP hydrolysis proceeds with the same movements of the heavy atoms, in particular the movement of P_γ and Mg^{2+}. Thus, there is no unique mechanism of ATP hydrolysis in myosin, but a number of equivalent pathways that the system can take depending on its current protonation state and specific conformation. Hitherto suggested mechanisms correspond to water activation mechanisms that occur *via* different proton transfer routes. In the present work, three possible water activation pathways have been studied and were found to be equivalent. Other nucleoside triphosphate hydrolyzing proteins such as the G-protein Ras or the F_1-ATPase exhibit similar structural features in their active sites to those found in myosin. Thus, it seems likely that the network hypothesis may also be fruitful in explaining the hydrolytic properties of these proteins.

In myosin, both Ser181 and Ser236 participate in the postulated dynamic hydrogen-bonding network. Mutating either into Ala would therefore be expected to reduce the number of possible proton transfer routes for water activation, thus leading to reduced, but not abolished, ATPase activity. This is indeed observed in experimental mutation studies. Mutating the salt-bridge residues Arg238 and Glu459 such that salt-bridge formation is made impossible, results in loss of ATPase activity, whereas in mutants that maintain the capacity of salt-bridge formation, ATPase activity is maintained as well. Thus, salt-bridge formation seems to be necessary to assemble the active site into its active conformation but the active participation of those residues in catalysis seems unlikely. This agrees with the findings of the perturbation analyses in the present work which indicate that the combined effects of the Arg238 and Glu459 on the reaction barrier height and therefore on catalysis are small.

The cationic metal cofactor Mg^{2+} is necessary for hydrolysis to occur. It is likely that, apart from a structural role, Mg^{2+} is also catalytically active. Mg^{2+} here is found to activate the bond to be cleaved by introducing an asymmetry into the triphosphate moiety of ATP. Upon decay of the transition state to the product state, Mg^{2+} moves towards P_i.ADP, thereby contributing to the separation of the hydrolysis products. This separation effect may be more pronounced when replacing Mg^{2+} by Ca^{2+}, thus giving a possible explanation of the experimental finding that the rate of myosin-catalyzed ATP hydrolysis is increased when using Ca^{2+} instead of Mg^{2+}.

The breaking of the coordination bond between Mg^{2+} and Ser237 implies that the switch-1 loop is no longer immobilized by this strong interaction. Just as formation of this coordination bond accompanies the closure of the switch-1 loop upon ATP binding, breaking of this bond can be considered a prerequisite for the opening of the switch-1 loop after hydrolysis. The present simulations suggest that product release is likely to occur *via* an exit route that opens through movement of switch-1, a mechanism that has been termed the "trap door" mechanism. Switch-1 movement is also believed to be coupled to the closure of the cleft between the upper and lower 50 kDa domains, thereby modulating the affinity of myosin for actin. Thus, weakening of the Mg^{2+}:switch-1 interaction, as observed here, is a key event that couples the hydrolysis step to the conformational changes in the posthydrolysis steps of the actomyosin cycle. This coupling might occur *via* a cooperative mechanism, in which small movements of switch-1 would facilitate the partical cleft closure that induces initial weak binding of actin. Strengthening of the actomyosin interaction then induces further cleft closure that propagates back to the switch-1 loop, which fully opens to allow the release of P_i after actin is strongly bound. Thus, the present simulations shed new light onto a crucial step in the actomyosin cycle which underlies motion in all living organisms.

Chapter 17

Outlook

The present work describes the second attempt ever undertaken to understand the reaction mechanism of ATP hydrolysis in myosin and its mechanochemical coupling to the actomyosin motor cycle with computational techniques. Given the complexity of both the ATP hydrolysis event itself and of myosin being a large motor protein, and considering the fact that combined quantum mechanical/molecular mechanical simulations do not yet belong to the standard toolkit of molecular simulations, it must be viewed as an explorative study that opens up a number of possible follow-up investigations. These can be grouped into projects that further develop the methods necessary for quantitative QM/MM simulations and projects that apply existing (or the newly-developed) methods to the myosin catalysis of ATP hydrolysis. A number of possible projects are listed below.

17.1 Method development

1. Parameterization of MNDO-d and/or AM1-d for phosphate anhydride hydrolysis reactions. This can be done by optimizing the necessary parameters such that the properties determined from MNDO-d or AM1-d mirror the reaction properties of pyrophosphate hydrolysis in aqueous solution that are necessary to be known as a reference (see Item 1 in Section 17.2).

2. Development of an efficient algorithm for QM/MM normal mode calculations. Analytical and numerical methods for normal mode calculations are available both for molecular mechanical and for quantum chemical calculations. However, the available methods either do not allow the combination of quantum and classical techniques or are inefficient and unstable. Thus, methods that allow normal mode calculation within the onion model usually used for QM/MM calculations are needed.

3. Development of a continuum solvent method that can be combined with QM/MM calculations. Currently, no continuum solvent methods, algorithms, or implementations are available in combination with QM/MM calculations The NUCS method developed in the present thesis represents a first attempt for an approximate and crude continuum treatment of solvent effects for this purpose. However, since continuum solvent techniques belong to the standard techniques used both in molecular mechanical and in quantum chemical calculations it seems possible and advisable to develop mathematical formulations and implementations thereof to make continuum methods available for QM/MM simulations.

4. Comparison of linking protocols. The performance of different linking protocols to combine quantum mechanical with molecular mechanical calculations should compared in a real-case scenario rather than on model compounds. This could for instance be achieved by determining proton affinities on some protein sidechains in a protein that is small enough to be treated fully quantum mechanically.

17.2 Simulating phosphoanhydride hydrolysis in aqueous solution

1. Mechanistic study of pyrophosphate hydrolysis in aqueous solution. Compared to phosphate monoester hydrolysis reaction, phosphoanhydride reactions have received much less interest, which is probably due to the higher complexity of the smallest possible model system in case of phosphoanhydride cleavage, *i.e.* pyrophosphate hydrolysis compared to methylphosphate hydrolysis. This is in contrast to the importance of phosphoanhydride cleavage reactions in biology. A systematic study that explores different possible reaction routes with different possible protonation states in presence and absence of Mg^{2+} both in gas phase and in aqueous solution using electron-correlated quantum chemical simulation techniques would therefore be extremely valuable and useful.

2. Network simulations of pyrophosphate hydrolysis in aqueous solution. To gain insight into a possible *active* role of the water network surrounding pyrophosphate in aqueous solution by partaking in an dynamic proton transfer network that provides the environment for the reactive changes in heavy atom positions (as has been proposed in the Network Hypothesis in the present thesis), a large number of different possible protonation patterns not only of pyrophosphate itself but of its surrounding water clusters must be modeled. Such an investigation will be made possible once an appropriately-tested semi-empirical quantum chemical method is available (see Section 17.1). To build up a close-to-complete network consisting of vertices corresponding to different configurations and conformations of the reactive system and of

edges corresponding to reaction paths that connect different states, a large number of optimizations will be necessary that can only be accomplished with an appropriate semi-empirical method.

3. Determination of the microscopic pK_a values of ATP, ADP, and the ADP.P_i adduct in aqueous solution in presence and absence of Mg^{2+}. Such calculations can be done by searching for the optimal geometries of the molecules under consideration in all possible protonation states using quantum chemical methods that include electron correlation effects and an adequate model of the solvent environment. This will show the number of protons bound and thus the charge of ATP at different pH and pMg values and at which position ATP will be protonated. Knowing such details is a necessary prerequisite for both a detailed description of the ATP hydrolysis event in aqueous solution and the determination of pK_a values in myosin (see below).

4. Simulations of pyrophosphate hydrolysis reactions in presence of Ca^{2+}, Mn^{2+}, or other divalent cations. Such simulations can be performed in analogy to the simulation proposed in Items 1 and 2. This will give insight into how the size and coordination properties of the divalent metal ion influence phosphoanhydride hydrolysis.

17.3 Simulating details of the myosin ATPase cycle

1. Determination of pK_a values of titratable groups in myosin and calculation of the most probable protonation patterns in the different structural states of myosin. This requires knowledge of the microscopic pK_a values of all titratable groups present, including the triphosphate moiety of ATP. pK_a calculations can be performed with continuum electrostatic techniques, with molecular mechanical free energy perturbations calculations, and combinations thereof. The simulation results will give insight into possible changes in protonation patterns along the myosin motor cycle that may contribute to the coupling of different events along the cycle, justify the adequacy or inadequacy of the assumption of the present work that ATP is fully deprotonated when bound to myosin, and allow preliminary testing of the Network Hypothesis vy considerating likelihoods of proton transfer events.

2. Re-determination of reaction path calculations of ATP hydrolysis in myosin. Once different possible protonation patterns of the active-site residues and the substrate in myosin are known (Item 1) the QM/MM reaction path simulations presented in the current thesis can be repeated within the new protonation pattern. This will make it possible to validate or corroborate the hypotheses on the functioning of myosin both as an enzyme and as a motor made in the present work. In addition, the size of the QM region can be varied, thus exploring whether Lys185 actively contributes to hydrolysis.

3. Free-energy perturbation (FEP) calculations on ATP hydrolysis in myosin. The reaction paths determined in the present work indicate that the reaction coordinate defined by the bond distances of the bonds formed and broken can be used as a pre-defined reaction coordinate without introduction of major inaccuracies. This provides the possibility of performing free-energy perturbation calculations while explicitly modeling the solvent on ATP hydrolysis in myosin using umbrella sampling techniques once a sufficiently fast and accurate quantum chemical method is available (see Item 1 in Section 17.1). The FEP calculations can be complemented by calculations of tunneling effects using the variational transition state theory with multidimensional tunneling (VTST/MT) approach. Such simulations should in principle be capable of reproducing the experimentally-observed energetics of ATP hydrolysis in myosin.

4. Network simulations of ATP hydrolysis in myosin. Such simulations can be performed and analyzed in analogy to the simulations proposed in Item 2 in Section 17.2. Such simulations should provide an alternative to FEP simulations to reproduce the energetics of ATP hydrolysis in myosin that are experimentally observed.

5. Combining reaction path calculations on the ATP hydrolysis in myosin with path calculation of the C/O to C/C transition in myosin (return stroke). This can be done by utilizing different structures along the return stroke as determined previously using the CPR algorithm (92) as starting structures for QM/MM reaction path calculations as performed in the present thesis.

6. Determination of a possible reaction path corresponding to the O/O to C/O structures of *Dictyostelium discoideum* myosin II, for which crystal structures are available. This can be done with molecular mechanical path calculations using the CPR algorithm in analogy to the determination of the reaction path for the return stroke (92).

7. Reaction path calculation on ATP hydrolysis in myosin with Ca^{2+} and Mn^{2+} as metal cofactors. Combined with the results of the simulations proposed in Item 4 in Section 17.2 this will clarify the role of the cation in both the catalytic cycle and the coupling of the catalytic and motor cycles.

8. Simulations on mutant myosins. Once the experimental ATPase activity of the wild-type myosin can be computationally reproduced (Items 3 and 4), explicit mutations of amino acids in the catalytic site can be made. This will possibly yield a detailed explanation of the effect of the mutation and thus of the role of the respective amino acid in catalyzing ATP hydrolysis and in coupling of the catalytic and the motor cycles.

9. Normal mode analyses of the myosin.ATP complex in the C/C conformation using QM/MM techniques. The determination of normal modes on a QM/MM potential

energy surface requires the availability of appropriate methods (2 in Section 17.1). Once the normal modes are determined they can be projected onto the previously-determined reaction coordinate (Items 2 and 4). The modes that overlap significantly with the reaction coordinate will couple to the reaction coordinate and thus promote or inhibit the reaction. 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.

Part V

Appendix

Appendix A

Comparison of the nucleotide binding sites of myosin II crystal structures

To compare the nucleotide binding pocket of the different *Dictyostelium discoideum* myosin II crystal structures all structures except the O/O structure were downloaded from the protein databank. The structures were then matched onto the $Mg.ADP.VO_4$ structure (77) with the program Moloc (432) by a rigid body match so as to optimally fit the C_α carbons of the P-loop.

Only the nucleotide binding pocket was then further analyzed using the visualization program VMD (433).

A.1 Comparison

A number of questions were put forward and answered:

1. Is the ADP moiety in both C/O and C/C conformations in the same position?

   Structures: 1vom, 1mmg, 1lvk, 1mma, 1mmn, 1mnd, ref1, 1fmw, 1g8x

   The N1 of the adenine ring varies about 1 Å perpendicular to the plane of the ring system (Fig. A.1). The positions of the P_β phosphorus varies by 0.6 Å. The positions of the P_α phosphorus varies by 0.5 Å. The maximum distance between the O4 atoms of the ribose is 0.75 Å. Since the resolution of the crystal structures varies between 1.9 Å and 2.6 Å, this does not seem significant. Thus, the conclusion is that the ADP moiety moves only very little.

2. What is the relative position of Mg^{2+} in the ADP structures (C/O) compared to the 1VOM (C/C) structure?

   Structures: 1vom, 1mmg, 1lvk, 1mma, 1mmn, 1mnd, ref1, 1fmw, 1g8x

   The maximum distance between the positions of the Mg^{2+} in *all* structures is 0.7 Å (Fig. A.2). This is within the crystallographic error. Thus, the Mg^{2+} position does *not* change between the C/O and C/C conformations of myosin. In addition, the coordination pattern of the Mg^{2+} coordination sphere does not change.

   In 1mma the Mg^{2+} is coordinated by the oxygens of the sidechains of Thr186 and Ser237, one oxygen on PB and three water molecules (bfactors on water smaller than on O Thr186 and Ser237!). The situation is identical in 1g8x. Here, the bfactors of the coordinated water oxygens are in the same range as the oxygens on the sidechains.

3. Are the sidechain orientations of G179, E180, S181, G182, A183, G184, K185, T186, N233, N234, N235, S236, S237, R238, D454, I455, S456, G457, and E449 identical? What about temperature factors?

   Structures: all

   - G179. The backbone fold is identical in all structures. However, the position of the C_α in the three C/C structures clearly is different from its position in the C/O structures (Fig. A.3). The bfactors are given in Table A.1.
   - E180. The backbone atoms overlap well in all structures. The side chain organization is similar in all C/O structures but differs in the C/C structures resulting in different position of the terminal negatively charged carboxylate group (Fig. A.3). The bfactors are given in Table A.2.
   - S181. Both backbone and sidechain atoms overlap well for all structures (Fig. A.3). The bfactors are given in Table A.3.
   - G182. The backbone fold is identical in all structures (Fig. A.4). The bfactors are given in Table A.4.
   - A183. Backbone and sidechain atoms overlap well in all structures (Fig. A.4). The bfactors are given in Table A.5.
   - G184. All atoms overlap well except for the oxygen of 1mne (Fig. A.4). The bfactors are given in Table A.6.
   - K185. The backbone atoms overlap well. Given the flexibility of the long sidechain, the sidechain atoms overlap remarkably well, too. This results in a position of the NZ that is very much conserved in all structures (only 1mnd shows a slightly different position, Fig. A.5). This suggests, that the positively charged nitrogen serves a similar role as the Mg^{2+} in coordinating the oxygens from the phosphate moiety of ATP or ADP . The bfactors are given in Table A.7.

- T186. The backbone atoms overlap well in all structures. So do the sidechain atoms (Fig. A.5). Only in 1fmv the orientation of the sidechain differs, but in this structure no Mg^{2+} is coordinated. This means that if the Mg^{2+} is around, the sidechain orientation is well-defined. The bfactors are given in Table A.8.
- N233. Only in 1g8x and 1mma the carboxamid group is rotated by 180° with respect to the other structures (Fig. A.6). The backbone atoms overlap well irrespective of the structure being in the C/O or the C/C form. The bfactors are given in Table A.9.
- N234. Only in ref1 the carboxamid group is rotated by 180° with respect to the other structures (Fig. A.6). The sidechain of 1d0z adopts a different orientation than in the remaining structures. However, the backbone atoms overlap well irrespective of the structure being in the C/O or the C/C form. The bfactors are given in Table A.10.
- N235. The sidechain orientations are identical in all structures. The bfactors are given in Table A.11.
- S236. All atoms overlap well. The orientation of the sidechain OH group is identical in all structures. (Fig. A.7). The bfactors are given in Table A.12.
- S237. The backbone atoms overlap well. The orientation of the sidechain OH group is different in the 1mnd, 1fmw, and 1fmv structures from the orientation in the other structures. (Fig. A.7). The bfactors are given in Table A.13.
- R238. The backbone atoms of the C/O structures overlap well, whereas the backbones are different in the C/C structures. However, the three C/C structures exhibit different positions among each other as well. The sidechain orientation differ as well. Among the C/O structures a clear pattern can be observed from which the C/C structures clearly are separated. The maximum distance observed between nitrogens NH2 of different structures is 2.1 Å (Fig. A.7). The bfactors are given in Table A.14.
- D454. The side chain orientation is identical in all structures. The three C/C structures are clearly separated from the C/O structures in both backbone and sidechain position (Fig. A.8). The bfactors are given in Table A.15.
- I455. The backbone atoms of all C/O structures overlap as well as the backbone atoms of all C/C structures. The sidechain orientation differs significantly between the different structures (Fig. A.8). The bfactors are given in Table A.16.
- S456. The complete residue is clearly separated in the C/O and C/C structures. In the C/O structures the orientation of the sidechain is identical in all structures except 1mmd and 1d1b. In the C/C structures the sidechain orientation of 1mnd differs from the orientation of the ref1 and 1vom structures (Fig. A.8). The bfactors are given in Table A.17.

- G457. The "fold" of the backbone is identical in all structures. However, the C/O structures are clearly separated by more than 5 Å from the C/C structures (Fig. A.9). The bfactors are given in Table A.18.
- F458. Backbone and sidechain atoms overlap well for all C/O structures. The C/C structures have a common geometry that differs significantly from that of the C/O structures. The distance between backbone atoms of the C/C structures and of the C/O structures is more than 5 Å (Fig. A.9). The bfactors are given in Table A.19.
- E459. The three C/C structures adopt an almost identical geometry and position. The C/O structures create a bundle of structures with a maximum distance of 1.7 Å between one of the carboxylate oxygens of the different structures. The distance between the C/O and C/C structures is about 5 Å (Fig. A.9). The bfactors are given in Table A.20.

4. Where are the crystal water molecules in the binding pocket?

   Structures: all

   Answer: All crystal water molecules within 6 Å around PG (or the equivalent atom) and PB are considered.

   - 1D0X.
     (a) HOH 1048 (index 5934) distances to other atoms in Å in brackets: Mg (2.13), ASN235:O (2.93), MNQ:OA2 = PA oxygen (2.71).
     (b) HOH 1712 (index 6599) to Mg (2.06), ASP454:OD2 (2.79), HOH1426 (2.86).
     (c) HOH 1371 (index 6258) to HOH1090 (2.76), HOH1426 (2.98), SER456:N (3.49)
     (d) HOH 1437 (index 6324) to MNQ:F1 = PG oxygen not Mg coordinating (2.53), MNQ:F3 = PG oxygen not Mg coordinating (3.35), ARG238:NH2 (3.39), SER237:O (3.45)
     (e) HOH 1110 (index 5997) to HOH1252 (2.47), GLU459:OE2 (2.74), SER456:OG (3.47)
     (f) HOH 1115 (index 6002) to THR230:OG (2.71), SER181:OG (2.76), HOH1146 (2.83), SER236:OG (3.35)
   - 1D0Y.
     (a) HOH 1023 (index 5937) to Mg (2.16), ONP:OA = PA oxygen (2.65), THR186:OG (2.75), ONP:OB = Mg coordinating O on PB (2.80), SER237:OG (2.86), ONP:F2 = Mg coordinating O on PG (3.02), ASN235:O (3.04)
     (b) HOH 1699 (index 6602) to *MG* (1.91), *ASP454:OD2* (2.63), *HOH1332* (2.90), SER237:OG (2.65), THR186:OG1 (2.81), ONP:F2 = Mg coordinating O on PG (2.88), ONP:OB2 = Mg coordinating O on PB (3.02)

(c) HOH 1432 (index 6346) to HOH1504 (3.16), HOH1036 (3.25), SER237:O (3.36), ONP:F1 = not Mg coordinating O on PG (3.43)

(d) HOH 1504 (index 6418) to ONP:F1 not Mg coordinating O on PG (2.73), HOH1432 (3.16), ARG238:NH2 (3.28), ONP:F3 not Mg coordinating O on PG (3.30), HOH1012 (3.41), SER237:O (3.41)

(e) HOH 1012 (index 5926) to HOH1303 (2.50), GLU459:OE2 (2.59), HOH1208 (3.37), HOH1504 (3.41)

(f) HOH 1081 (index 5995) to THR 230:OG (2.63), HOH1116 (2.68), SER181:OG (2.87), ARG238:NH2 (3.24), HOH1208 (3.29), SER236:OG (3.50)

- 1D0Z.

(a) HOH 1453 (index 6347) to *MG* (2.07), *PNQ:OA2* = not Mg coordinating O on PA (2.73), *ASN235:O* (3.03), SER237:OG (2.90), THR186:OG1 (2.92), PNQ:OB2 = MG coordinating O on PB (2.94), PNQ:F3 = Mg coordinating O on PG (3.00)

(b) HOH 1454 (index 6348) to *MG* (1.92), *ASP454:OD2* (2.92), *HOH1201* (3.46), SER237:OG (2.62), PNQ:F3 = Mg coordinating O on PG (2.64), THR186:OG1 (2.76), PNQ:OB2 = Mg coordinating O on PB (2.89)

(c) HOH 1149 (index 6043) to HOH1203 (2.88)

(d) HOH 1371 (index 6265) to PNQ:F1 = not Mg coordinating O on PG (2.78), ARG238:NH2 (3.04), SER237:O (3.32), HOH1118 (3.35), PNQ:F2 = not Mg coordinating O on PG (3.39)

(e) HOH 1118 (index 6012) to HOH1194 (2.63), GLU459:OE2 (2.75), SER456:OG (3.30), HOH1371 (3.35)

(f) HOH 1037 (index 5931) to SER181:OG (2.71), THR230:OG1 (2.71), HOH1373 (3.08), ARG238:NH2 (3.39)

- 1D1A.

(a) HOH 995 (index 5801) to *MG* (2.09), *DAE:OA* = not Mg coordinating O on PA (2.80), *ASN235:O* (2.92), SER237:OG (2.82), THR186:OG1 (2.82), DAE:F2 = Mg coordinating O on PG (2.96), DAE:OB2 = Mg coordinating O on PB (2.98)

(b) HOH 996 (index 5802) to *MG* (2.16), *ASP454:OD2* (2.71), *HOH1151* (3.38), *HOH1313* (3.44), ASP454:OD1 (3.43), THR186:OG1 (2.80), SER237:OG (2.88), DAE:OB2 = Mg coordinating O on PB (2.90), DAE:F2 = Mg coordinating O on PG (3.08)

(c) HOH 1151 (index 5953) to HOH1132 (3.32), HOH996 (3.38), HOH1313 (3.46), SER456:N (3.50)

(d) HOH 1347 (index 6149) to GLU459:OE2 (2.97)

(e) HOH1030 (index 5832) to THR 230:OG1 (2.31), SER181:OG (2.80), SER236:OG (3.43), HOH1125 (3.48)

- 1D1B.
  (a) HOH 1540 (index 6372) to *MG* (2.05), *ASN235:O* (2.95), *DAQ:OA2* = not Mg coordinating O on PA (2.92), SER237:OG (2.72), THR186:OG1 (2.90), DAQ:OB2 = Mg coordinating O on PB (3.04), DAQ:F3 = Mg coordinating O on PG (3.06)
  (b) HOH 1541 (index6373) to *MG* (2.07), *HOH1058* (3.05, *ASP454:OD2* (2.78), THR186:OG1 (2.77), SER237:OG (2.78), DAQ:F3 = Mg coordinating O on PG (2.92), DAQ:OB2 = Mg coordinating O on PB (2.98), ASP454:OD1 (3.31)
  (c) HOH 1245 (index 6077) to HOH1526 (2.14), HOH1246 (2.88), HOH1058 (3.05), PHE239 (3.12)
  (d) HOH 1526 (index 6358) to HOH1245 (2.14), SER237:O (2.27), HOH1531 (2.38), HOH1441 (3.12), DAQ:F1 = not MG coordinating O on PG (3.19)
  (e) HOH 1531 (index 6363) to HOH1256 (2.38), DAQ:F1 (2.78), HOH1123 (3.06), HOH1410 (3.09), HOH1441 (3.13), SER456:OG (3.28), ARG238:NH2 (3.49)
  (f) HOH 1123 (index 5955) to GLU459:OE2 (2.61), SER456:OG (2.70), HOH1519 (2.75), HOH1531 (3.06), HOH1410 (3.15)
  (g) HOH 1061 (index 5893) outside the binding pocket
- 1D1C.
  (a) HOH 1005 (index 5941) to *MG* (2.20), *NMQ:OA2* = not MG coordinating O on PA (3.00), *ASN235:O* (3.18), NMQ:OB2 = Mg coordinating O on PB (2.56), NMQ:F2 = Mg coordinating O on PG (2.63), SER237:OG (2.67, THR186:OG1 (2.93)
  (b) HOH 1458 (index 6394) to *MG* (3.27), HOH1141 (2.88), HOH1357 (2.88), ASP454:O (3.08), PHE239:O (3.28), ASP454:OD2 (3.29), ASP454:OD1 (3.33), SER237:OG (3.45)
  (c) HOH 1357 (index 6293) to HOH1458 (2.86), HOH1222 (3.11), NMQ:F1 = not MG coordinating O on PG (3.23), SER237:O (3.32)
  (d) HOH 1464 (index 6400) to ARG238:NH2 (2.81), HOH1452 (2.93)
  (e) HOH 1452 (index 6388) to HOH1188 (2.35), HOH1464 (2.93), GLU459:OE2 (3.14), SER456:OG (3.16)
  (f) HOH 1135 (index 6071) to THR230:OG1 (2.44), SER181:OG (2.78), HOH1499 (3.30), SER236:OG (3.43), HOH1536 (3.48)
- 1FMV. apo structure. By comparison with ATP position in 1fmw I identify HOH1154 (index 6476) and HOH 1155 (index 6127) as water molecules occupying the positions of PB and PG, respectively.
  (a) HOH 1158 (index 6475) to HOH1167 (2.58), HOH1154 (3.29), GLU187:N (3.30)

(b) HOH 1167 (index 6478) to HOH1166 (2.53), HOH1158 (2,58), HOH1142 (2.82), HOH1531(3.30), ASN127:ND (3.40)

(c) HOH 1400 (index 6473) to SER186:OG1 (2.67), ASP454:OD1 (3.08), ASP454:OD2 (3.41)

(d) HOH 1043 (index 6026) to HOH966 (2.95), PHE239:O (3.37)

(e) HOH 1156 (index 6474) to LYS185:NZ (2.83), HOH1155 (3.12), HOH1058 (3.15), SER181:N (3.43)

(f) HOH 935 (index 5918) to THR230:OG1 (2.37), SER181:OG (3.03), HOH1498 (3.28)

- 1FMW (strange structure!).

(a) HOH 1227 (index 6012) to *MG* (2.18), *ASN235:O* (3.09), *ATP:O1A* (3.16), THR186:OG1 (2.81), ATP:O1G = Mg coordinating (3.03), ATP:O1B = Mg coordinating (3.15), SER237:N (3.45)

(b) HOH 1226 (index 6011) to *MG* (2.49), *ASP454:OD1* (2.63), *HOH1375* (2.81), THR186:OG1 (3.18), LYS241:NZ (3.34)

(c) HOH 1375 (index 6159) to SER237:O (2.33), HOH1372 (2.52), HOH1226 (2.81), ATP:O1G = MG coordinating (3.03), ATP:O3G = not MG coordinating (3.36)

(d) HOH 1372 (index 6156) to HOH1375 (2.52), HOH1214 (2.54), PHE239:O (2.98)

(e) HOH 1157 (index 5942) to ATP:O3G = not MG coordinating (2.84)

(f) HOH 1306 (index 6091) to SER181:OG (2.59), THR230:OG1 (2.60), SER236:OG (3.36)

(g) HOH 1274 (index 6059) to ATP:O1A (2.48), ATP:O3* = sugar (3.11), ASN235:ND2 (3.28)

- 1G8X.

(a) HOH 1201 (index 16278) to *MG* (2.20), *ADP:O2A* (2.75), *ASN235:O* (3.09), THR186:OG1 (2.73), SER237:OG (2.82), ADP:O2B = Mg coordinating (2.92), ADP:O3B = MG coordinating (3.30)

(b) HOH 1202 (index 16279) to *MG* (2.33), *ASP454:OD2* (2.53), *ASP454:OD1* (2.98), THR186:OG1 (3.06), SER237:OG (3.33)

(c) HOH 1203 (index 16280) to *MG* (2.36), *ADP:O3B* (2.78), *HOH1204* (2.68), SER237:N (2.79), SER237:O (3.24), SER237:OG (3.49)

(d) HOH 1204 (index 16281) to HOH1203 (2.68), LYS185:NZ (2.85)

(e) HOH 1205 (index 16282) to ADP:O2A (2.54), ADP:O3* (2.64), ASN235:ND2 (2.94), GLU187:OE1 (3.28)

- 1LVK.

(a) HOH 997 (index 5907) to *MG* (2.19), *MNT:O2A* = mant-ADP (2.62), *ASN:235:O* (2.91), THR186:OG1 (2.91), SER237:OG (2.92), MNT:O2B = Mg coordinating (2.94), BEF:F3 = Mg coordinating (3.00)

(b) HOH 996 (index 5906) to *MG* (2.04), *ASP454:OD1* (2.67), *ASP454:OD2* (3.31), THR186:OG1 (2.84), SER237:OG (2.87), BEF:F3 (2.90), MNT:O2B (2.98)

- 1MMA. (take HOH 9951 index 6135 as substitute for PG)

(a) HOH 9954 (index 6138) to *MG* (2.24), *ASN235:O* (3.00), *ADP:O2A* (3.01), SER237:OG (2.68), THR186:OG1 (2.94), ADP:O2B = Mg coordinating (2.98), HOH 9951 = Mg coordinating (3.20, SER237:N (3.45)

(b) HOH 9953 (index 6137) to *MG* (2.10), *ASP454:OD1* (2.50), *HOH8158* (2.64), HOH8023 (3.34), ASP454:OD2 (3.38), THR186:OG1 (2.98), SER237:OG (3.01), ADP:O2B (3.14), HOH9951 (3.22)

(c) HOH 9951 (index 6135) to *MG* (2.25), *ADP:O3B* = not Mg coordinating (2.18), *HOH9955* (2.69), SER237:N (3.21), HOH9952 (3.36), ADP:O2B = Mg coordinating (2.75), HOH9954 = Mg coordinating (3.20), SER237:OG (3.22), HOH9953 = Mg coordinating (3.22)

(d) 8202 (index 5917) to SER237:OG (2.54), ASN227:OD (2.69), HOH8142 (2.91), HOH8310 (3.13), ASN227:ND2 (3.18)

(e) HOH 8158 (index 5873) to HOH8142 (2.31), HOH9953 (2.64), ASP454:O (2.72), PHE239:O (2.90), HOH8023 (3.12), ASP454:OD1 (3.35)

(f) HOH 8023 (index 5738) to HOH8158 (3.12), HOH9953 (3.34), SER456:N (3.44)

(g) HOH 9952 (index 6136) to LYS185:NZ (2.61), HOH995 (3.36), HOH9955 (3.43), SER181:N (3.49)

(h) HOH 9955 (index 6139) to SER181:OG (2.17), SER236:OG (2.59), HOH9951 (2.69), ADP:O3B = not Mg coordinating (2.72), ASN233:ND2 (3.25), GLY182:N (3.36), HOH9952 (3.43)

(i) HOH 8074 (index 5789) to ADP:O2A (2.53), GLU187:OE2 (2.67), ADP:O3* (2.99), ASN235:ND2 (3.19)

- 1MMD.

(a) HOH 1001 (index 5908) to *MG* (2.15), *ADP:O2A* (2.78), *ASN235:O* (2.83), BEF:F3 (2.85), ADP:O2B (2.90), SER237:OG (3.05), THR:OG1 (3.12)

(b) HOH 1002 (index 5909) to *MG* (2.06), *ASP454:OD2* (2.53), THR186:OG1 (2.86), BEF:F3 (3.00) ADP:O2B (3.07), SER237:OG (3.23)

(c) HOH 1181 (index 6088) to BEF:F1 = not MG coordinating O on PG (2.72), SER237:O (3.15), ARG238:NH2 (3.44)

(d) HOH 1046 (index 5953) to SER181:OG (2.55), THR230:OG1 (2.78)

- 1MMG.
  (a) HOH 9951 (index 6233) to *MG* (2.10), *AGS:O2A* (2.63), *ASN235:O* (2.99), THR186:OG1 (2.81), SER237:OG (2.87), AGS:O2B (2.89), AGS:O3G (2.99)
  (b) HOH 9952 (index 6234) to *MG* (2.06), *ASP454:OD1* (2.73), *HOH8266* (3.22), ASP454:OD2 (3.33), THR186:OG1 (2.78), SER237:OG (2.84), AGS:O2B (2.87), AGS:O3G (3.17)
  (c) HOH 8257 (index 6144) to HOH8103 (3.09), HOH8266 (3.24), PHE239:O (3.43)
  (d) HOH 8194 (index 6081) to THR230:OG1 (2.50), SER181:OG (2.95)
  (e) HOH 8178 (index 6065) to AGS:O2A (2.62), GLU187:OE2 (2.69), ASN235:ND2 (2.96), AGS:O3* = sugar (3.07)
- 1MMN.
  (a) HOH 9951 (index 6320) to *MG* (2.07), *ANP:O2A* (2.87), *ASN235:O* (3.10), THR186:OG1 (2.55), SER237:OG (2.82), ANP:O2B (2.93), ANP:O3G (3.07)
  (b) HOH 9952 (index 6321) to *MG* (2.07), *ASP454:OD1* (2.69), ASP454:OD2 (3.47), SER237:OH (2.94), THR186:OG1 (2.96), ANP:O2B (2.98), ANP:O3G (3.07)
  (c) HOH 8308 (index 6142) to HOH8273 (2.43), HOH8101 (2.54), PHE239:O (3.18)
  (d) HOH 8273 (index 6107) to HOH8308 (2.43), ANP:O1G (2.47)
  (e) HOH 8266 (index 6100) to THR230:OG1 (2.31), SER181:OG (2.79), SER236:OG (3.43), HOH8272 (2.46)
- 1MND.
  (a) HOH 801 (index 5139) to *MG* (1.84), *ADP:O2A* (2.95), *ASN235:O* (3.07), SER237:OG (2.32), ALF:F2 = Mg coordinating (2.51), ADP:O2B = Mg coordinating (2.85), THR186:OG1 (3.13), ALF:F4 = Mg coordinating (3.19), ADP:O3B = not directly Mg coordinating (3.02)
  (b) HOH 799 (index 5137) to GLY457:O (2.77), GLU459:OE2 (2.99), ALF:F3 (2.99), GLY457:N (3.35)
  (c) HOH 739 (index 5077) to SER181:OG (2.68), GLU459:OE2 (2.77), THR230:OG1 (2.85)
  (d) HOH 802 (index 5140) to ADP:O2A (2.40)¡ ADP:O1A) (2.98), HOH800 (3.13)
- 1MNE.
  (a) HOH 947 (index 6026) to *MG* (1.87), *ASN235:O* (3.20), POP:O1 (2.65), THR186:OG1 (2.66), POP:O4 (2.72), SER237:OG (2.80)

(b) HOH 946 (index 6025) to *MG* (1.99), *ASP454:OD2* (2.74), ASP454:OD1 (3.30), THR186:OG1 (2.72), SER237:OG (2.85), POP:O1 (2.95), POP:O4 (3.04)

(c) HOH 821 (index 5900) to POP:O6 (2.59)

(d) 818 (index 5897) to SER181:OG (2.61), THR230:OG1 (3.08)

(e) 803 (index 5882) to GLU187:N (2.86), POP:O3 (3.23), GLY184:O (3.24), HOH800 (3.41), HOH804 (3.44)

(f) HOH 804 (index 5883) to HOH800 (2.73), POP:O3 (3.04), HOH806 (3.27), GLY182:O (3.37), HOH803 (3.44)

- 1VOM.

(a) HOH 695 (index 6477) to *MG* (2.04), *ADP:O2A* (2.73), *ASN235:O* (2.80), SER237:N (3.50), VO:O1 = Mg coordinating (2.77), SER237:OG (2.81), THR186:OG1 (2.94), ADP:O1B = Mg coordinating (2.94)

(b) HOH 696 (index 6479) to *MG* (2.06), *ASP454:OD1* (2.55), *ILE455:O* (2.69), THR186:OG1 (2.72), VO:O1 (2.84), ADP:O1B (3.03), SER237:OG (3.10)

(c) HOH697 (index 6479) to VO:O4 (2.62), GLY457:O (2.65), GLY459:OE1 (2.66), GLY457:N (3.33), ARG238:NH1 (3.35)

(d) HOH 104 (index 5886) to THR 230:OG1 (2.57), SER181:OG (2.64), GLY459:OE1 (3.03)

- REF1.

(a) HOH 1 (index 6112) to *MG* (2.26), *ADP:O2A* (2.43), *ASN235:O* (2.52), ASN233:ND2 (3.36), BEF:F2 = Mg coordinating (2.83), ADP:O2B = Mg coordinating (2.04), SER237:OG (3.14), THR186:OG1 (3.31)

(b) HOH 1 (index 6113) to *MG* (2.35), *ASP454:OD1* (2.51), *ILE:455:O* (2.69), ASP454:OD2 (3.02), THR186:OG1 (2.94), SER237:OG (3.31)

(c) HOH 1 (index 6117) to THR230:OG1 (2.76), SER181:OG (2.66), GLU459:OE1 (2.99)

A.2 Conclusions from the comparison

The nucleotide binding pocket is structurally well conserved. Almost no flexibility is observed for the positions of the Mg^{2+} cation and ADP, irrespective of considering the C/O or the C/C conformations.

The P-loop residues adopt mostly identical geometries in all structures. For G179 a slight difference in position between C/O and C/C structures is observed. The sidechain of E180 in the C/C structures is flexible and thus needs be considered when modeling the $ADP.P_i$ end state of the hydrolysis. K185 exhibits some structural flexibility in the sidechain,

however, the position of the terminal NZ is *very* conserved. Thus, it seems that this nitrogen serves a similar structural role as the Mg^{2+} in preserving the geometry of the binding pocket.

The switch-1 residues also overlap very well. Structural flexibility is observed mainly in R238 and, to some extent, the sidechain orientations of N233 and S237. In switch-2, however, distinct positions are observed between C/C and C/O structures for all residues. Nevertheless, the three C/C structures show very similar patterns. The most disordered side chain in that of I455, which, however, does not contribute to hydrolysis at all due to its hydrophobic nature.

The C/O structures contain more crystal water molecules in the vicinity of the hydrolysis site than the C/C structures. When ADP (or an analog) is bound, a crystal water molecule occupies the position of the O_γ that coordinates Mg^{2+}. Thus, it seems likely that the geometry of the binding pocket including all major interactions is predefined . If some atoms are not present (i.e. γ-phosphate in ADP structures), they are replaced by water molecules with only slight changes in side chain orientations. Thus, it should be possible to replace water molecules with an inorganic phosphate molecule to model the hydrolysis product state.

The most useful structures for modeling the hydrolysis event in myosin are ref1 and 1vom (C/C).

code	min	max
1D0X	13.07 (CA)	17.17 (N)
1D0Y	13.77 (O)	26.66 (C)
1D0Z	5.06 (C)	12.62 (N)
1D1A	14.80 (CA)	28.08 (N)
1D1B	13.49 (O)	19.89 (N)
1D1C	7.60 (CA)	10.32 (O)
1FMV	3.95 (CA)	13.91 (N)
1FMW	6.22 (CA)	17.81 (O)
1G8X	26.45 (CA)	27.75 (C)
1LVK	14.49 (O)	17.53 (N)
1MMA	9.31 (CA)	31.05 (N)
1MMD	8.97 (CA)	14.20 (N)
1MMG	16.66 (C)	26.15 (N)
1MMN	11.68 (CA)	22.27 (N)
1MND	29.40 (O)	76.46 (CA)
1MNE	5.88 (CA)	29.91 (N)
1VOM	10.74 (CA)	22.44 (C)
ref1	26.04 (N)	32.34 (O)

Table A.1: Minimal and maximal temperature factors for atoms of G179.

code	min	max
1D0X	10.05 (CA)	40.95 (OE2)
1D0Y	14.77 (C)	44.42 (OE1)
1D0Z	7.52 (N)	41.62 (OE2)
1D1A	17.53 (C)	97.09 (OE1)
1D1B	8.45 (CA)	36.75 (CD)
1D1C	6.91 (O)	59.46 (OE1)
1FMV	14.73 (CA)	48.12 (OE2)
1FMW	10.17 (CA)	39.78 (OE2)
1G8X	25.00 (C)	49.77 (OE1)
1LVK	12.74 (C)	79.42 (CD)
1MMA	16.57 (CG)	54.76 (CD)
1MMD	4.70 (C)	36.96 (CD)
1MMG	15.46 (C)	59.46 (CD)
1MMN	8.83 (C)	41.95 (OE2)
1MND	21.19 (OE1)	57.76 (OE2)
1MNE	6.24 (CA)	67.37 (OE2)
1VOM	8.02 (CD)	18.27 (OE1)
ref1	30.72 (N)	44.42 (OE1)

Table A.2: Minimal and maximal temperature factors for atoms of E180.

code	min	max
1D0X	9.34 (CA)	22.46 (C)
1D0Y	15.26 (N)	26.09 (C)
1D0Z	7.37 (CA)	21.36 (C)
1D1A	14.46 (CB)	38.71 (C)
1D1B	11.82 (CA)	28.27 (C)
1D1C	1.06 (CA)	20.31 (O)
1FMV	27.64 (O)	100.00 (C)
1FMW	10.96 (N)	47.24 (C)
1G8X	23.76 (CB)	35.61 (OG)
1LVK	10.49 (CA)	20.27 (OG)
1MMA	19.13 (N)	37.00 (C)
1MMD	7.59 (CB)	23.72 (C)
1MMG	18.91 (N)	27.67 (CB)
1MMN	7.90 (CA)	22.72 (O)
1MND	16.74 (CA)	77.07 (O)
1MNE	7.65 (CA)	26.61(O)
1VOM	9.14 (CB)	16.02 (O)
ref1	33.90 (C)	36.23 (OG)

Table A.3: Minimal and maximal temperature factors for atoms of S181.

code	min	max
1D0X	8.74 (CA)	22.15 (N)
1D0Y	13.27 (CA)	30.18 (N)
1D0Z	4.06 (CA)	18.79 (C)
1D1A	19.34 (CA)	29.87 (C)
1D1B	11.24 (CA)	23.12 (N)
1D1C	1.34 (CA)	29.91 (N)
1FMV	20.06 (CA)	36.39 (N)
1FMW	16.09 (C)	33.59 (O)
1G8X	25.39 (CA)	26.71 (O)
1LVK	12.09 (CA)	21.28 (C)
1MMA	13.69 (CA)	31.80 (O)
1MMD	10.46 (CA)	20.99 (C)
1MMG	14.29 (CA)	23.97 (O)
1MMN	9.34 (CA)	27.57 (N)
1MND	26.09 (CA)	41.00 (N)
1MNE	10.98 (C)	27.17 (CA)
1VOM	14.53 (N)	18.90 (C)
ref1	32.54 (CA)	33.99 (C)

Table A.4: Minimal and maximal temperature factors for atoms of G182.

code	min	max
1D0X	8.48 (CA)	15.82 (N)
1D0Y	15.52 (CA)	19.40 (O)
1D0Z	6.33 (CA)	9.48 (CB)
1D1A	15.38 (CB)	29.94 (O)
1D1B	10.43 (C)	21.00 (O)
1D1C	6.14 (C)	21.82 (O)
1FMV	9.08 (CA)	21.45 (N)
1FMW	7.03 (CA)	22.74 (O)
1G8X	22.88 (O)	29.91 (CB)
1LVK	12.34 (C)	18.51 (N)
1MMA	12.38 (C)	27.63 (O)
1MMD	8.72 (CB)	21.58 (C)
1MMG	14.79 (C)	26.62 (O)
1MMN	4.56 (CA)	21.18 (O)
1MND	10.92 (CB)	45.98 (O)
1MNE	24.45 (N)	37.24 (CB)
1VOM	10.68 (CB)	22.33 (C)
ref1	26.81 (O)	33.90 (N)

Table A.5: Minimal and maximal temperature factors for atoms of A183.

code	min	max
1D0X	17.74 (N)	24.46 (C)
1D0Y	14.77 (CA)	24.19 (O)
1D0Z	9.47 (CA)	20.94 (C)
1D1A	15.66 (CA)	35.12 (C)
1D1B	13.88 (N)	33.49 (C)
1D1C	6.14 (CA)	37.59 (O)
1FMV	7.51 (CA)	18.22 (C)
1FMW	4.39 (CA)	22.35 (O)
1G8X	25.18 (CA)	27.29 (O)
1LVK	15.22 (O)	18.68 (C)
1MMA	19.12 (CA)	23.09 (N)
1MMD	8.47 (N)	17.53 (C)
1MMG	19.59 (O)	26.19 (C)
1MMN	13.42 (CA)	24.43 (N)
1MND	15.80 (C)	34.65 (N)
1MNE	1.00 (N)	17.57 (C)
1VOM	9.40 (O)	16.75 (C)
ref1	28.63 (CA)	31.18 (O)

Table A.6: Minimal and maximal temperature factors for atoms of G184.

code	min	max
1D0X	9.77 (CD)	19.94 (C)
1D0Y	14.11 (NZ)	21.47 (CG)
1D0Z	7.97 (CA)	13.64 (C)
1D1A	18.94 (CB)	31.47 (N)
1D1B	9.34 (CG)	24.77 (N)
1D1C	3.05 (CB)	29.25 (N)
1FMV	13.50 (O)	27.75 (CD)
1FMW	7.90 (CG)	34.50 (N)
1G8X	21.81 (CA)	25.65 (N)
1LVK	8.95 (CD)	18.18 (CE)
1MMA	7.76 (CG)	25.28 (CE)
1MMD	5.18 (CG)	20.44 (NZ)
1MMG	11.00 (CG)	23.16 (C)
1MMN	8.91 (CE)	18.26 (C)
1MND	13.35 (C)	69.91 (CB)
1MNE	3.09 (CB)	23.83 (CD)
1VOM	8.33 (CB)	20.87 (CD)
ref1	25.56 (CD)	30.41 (CG)

Table A.7: Minimal and maximal temperature factors for atoms of K185.

code	min	max
1D0X	9.86 (CA)	28.07 (C)
1D0Y	14.96 (N)	39.00 (C)
1D0Z	7.15 (CA)	38.54 (C)
1D1A	17.59 (CA)	39.00 (C)
1D1B	10.15 (CA)	34.21 (C)
1D1C	7.33 (CB)	78.86 (OG1)
1FMV	14.91 (N)	84.97 (CB)
1FMW	8.15 (CA)	61.76 (OG1)
1G8X	19.07 (CG2)	29.39 (O)
1LVK	10.67 (N)	28.69 (C)
1MMA	7.50 (CA)	32.02 (C)
1MMD	4.02 (CA)	14.32 (CB)
1MMG	10.21 (CA)	32.69 (OG1)
1MMN	8.46 (CA)	29.96 (C)
1MND	1.93 (CG2)	95.05 (O)
1MNE	2.41 (CG2)	23.21 (OG1)
1VOM	9.99 (CA)	26.47 (C)
ref1	25.01 (CG2)	29.29 (C)

Table A.8: Minimal and maximal temperature factors for atoms of T186.

code	min	max
1D0X	17.95 (CA)	24.72 (O)
1D0Y	16.09 (CA)	35.34 (O)
1D0Z	3.99 (CB)	26.21 (O)
1D1A	28.07 (O)	42.33 (ND2)
1D1B	17.31 (CB)	28.77 (CG)
1D1C	3.27 (CA)	40.77 (CG)
1FMV	25.38 (C)	100.00 (CG)
1FMW	8.80 (CB)	32.16 (CG)
1G8X	21.79 (ND2)	26.94 (N)
1LVK	11.39 (CA)	29.46 (O)
1MMA	20.61 (N)	34.19 (CG)
1MMD	6.66 (CB)	20.30(OD1)
1MMG	16.25 (C)	23.96 (CG)
1MMN	15.58 (C)	40.78 (CG)
1MND	8.08 (CB)	61.78 (O)
1MNE	16.78 (CA)	47.43 (OD1)
1VOM	9.94 (ND2)	23.51(C)
ref1	27.63 (CB)	31.39(O)

Table A.9: Minimal and maximal temperature factors for atoms of N233.

code	min	max
1D0X	17.65 (N)	95.93 (OD1)
1D0Y	17.70 (C)	100.00 (OD1)
1D0Z	10.95 (N)	100.00 (OD1, CG)
1D1A	24.10 (C)	100.00 (OD1)
1D1B	17.34 (CB)	100.00 (OD1)
1D1C	12.07 (CA)	67.43 (CG)
1FMV	24.30 (N)	93.43 (OD1)
1FMW	21.61 (C)	100.00 (OD1)
1G8X	21.85 (CB)	27.92 (O)
1LVK	19.28 (N)	47.10 (OD1)
1MMA	21.52 (CA)	100.00 (OD1)
1MMD	10.31 (N)	37.75 (OD1)
1MMG	20.27 (N)	83.42 (CG)
1MMN	16.52 (N)	79.09 (OD1)
1MND	11.43 (CG)	49.77 (OD1)
1MNE	20.15 (N)	53.04 (OD1)
1VOM	9.66 (C)	50.66 (OD1)
ref1	27.80 (CB)	35.71 (O)

Table A.10: Minimal and maximal temperature factors for atoms of N234.

code	min	max
1D0X	18.28 (CA)	32.19 (C)
1D0Y	21.69 (CA)	60.39 (C)
1D0Z	13.28 (CA)	31.83 (C)
1D1A	26.55 (C)	45.24 (CA)
1D1B	20.51 (CB)	31.77 (C)
1D1C	12.81 (C)	35.53 (CB)
1FMV	23.55 (CB)	86.33 (C)
1FMW	10.73 (CB)	42.84 (C)
1G8X	24.59 (O)	26.67 (OD1)
1LVK	17.45 (CA)	23.87 (ND2)
1MMA	24.58 (CA)	41.18 (OD1)
1MMD	9.09 (CG)	21.64 (C)
1MMG	21.82 (CA)	40.81 (C)
1MMN	20.04 (N)	42.89 (C)
1MND	30.46 (O)	62.32 (CG)
1MNE	15.10 (CB)	56.20 (OD1)
1VOM	9.80 (C)	24.02 (OD1)
ref1	26.66 (ND2)	30.67 (O)

Table A.11: Minimal and maximal temperature factors for atoms of N235.

code	min	max
1D0X	15.89 (CA)	22.58 (O)
1D0Y	22.55 (N)	29.03 (OG)
1D0Z	11.28 (N)	25.50 (C)
1D1A	20.98 (C)	35.24 (CA)
1D1B	13.57 (C)	27.90 (O)
1D1C	13.84 (CB)	27.20 (O)
1FMV	19.58 (CA)	30.98(C)
1FMW	17.94 (C)	29.77 (CB)
1G8X	23.46 (CB)	26.66 (OG)
1LVK	12.82 (C)	22.90 (CB)
1MMA	17.01 (C)	45.25 (CA)
1MMD	10.85 (CB)	29.90 (O)
1MMG	13.35 (C)	28.79 (CB)
1MMN	11.07 (C)	20.29 (N)
1MND	7.69 (CA)	60.65(CB)
1MNE	22.88 (C)	43.28 (O)
1VOM	11.51 (C)	18.64 (O)
ref1	25.43 (CB)	28.22 (O)

Table A.12: Minimal and maximal temperature factors for atoms of S236.

code	min	max
1D0X	12.87 (CA)	23.06 (N)
1D0Y	12.58 (CA)	21.22 (OG)
1D0Z	7.37 (C)	19.29 (CB)
1D1A	15.44 (CB)	32.62 (N)
1D1B	11.38 (C)	21.72 (OG)
1D1C	9.63 (CA)	30.17 (CB)
1FMV	13.56 (C)	37.74 (CB)
1FMW	11.78 (C)	41.91 (OG)
1G8X	25.06 (N)	27.31 (C)
1LVK	8.84 (C)	19.12 (N)
1MMA	15.33 (CA)	39.94 (OG)
1MMD	1.10 (C)	16.67 (N)
1MMG	13.01 (C)	23.25 (N)
1MMN	6.13 (C)	18.29 (O)
1MND	15.46 (O)	51.53 (C)
1MNE	9.28 (CA)	22.99 (OG)
1VOM	10.13 (C)	12.78 (OG)
ref1	26.16 (CB)	30.32 (OG)

Table A.13: Minimal and maximal temperature factors for atoms of S237.

code	min	max
1D0X	12.41 (CA)	46.40 (CD)
1D0Y	15.28 (CB)	41.37 (CZ)
1D0Z	4.61 (CA)	41.34 (CD)
1D1A	24.50 (CG)	34.98 (NH1)
1D1B	12.22 (N)	84.30 (CD)
1D1C	10.13 (CA)	42.78 (C)
1FMV	9.18 (N)	42.86 (NH1)
1FMW	19.21 (CA)	40.73 (CD)
1G8X (GLU 238)	30.06 (N)	46.64 (OE1)
1LVK	12.44 (N)	58.38 (CD)
1MMA	19.10 (CB)	33.81 (NH1)
1MMD	8.99 (N)	22.03 (C)
1MMG	12.30 (CB)	42.11 (CD)
1MMN	13.75 (N)	36.94 (CD)
1MND	4.14 (CG)	51.03 (O)
1MNE	15.76 (N)	34.06 (C)
1VOM	11.63 (CG)	22.99 (C)
ref1	26.72 (N)	31.27 (CG)

Table A.14: Minimal and maximal temperature factors for atoms of R238.

code	min	max
1D0X	10.95 (CA)	34.47 (CG)
1D0Y	13.37 (CA)	27.85 (CG)
1D0Z	2.65 (CB)	20.58 (CG)
1D1A	16.79 (N)	40.33 (OD1)
1D1B	10.21 (CA)	25.41 (CG)
1D1C	4.42 (CB)	29.64 (O)
1FMV	11.81 (OD1)	22.47 (CG)
1FMW	15.87 (C)	43.36 (CB)
1G8X	37.99 (O)	51.76 (OD1)
1LVK	17.18 (O)	31.36 (CG)
1MMA	16.04 (N)	32.39 (CA)
1MMD	3.76 (C)	12.44 (OD1)
1MMG	14.54 (CA)	27.19 (CG)
1MMN	10.76 (N)	28.13 (C)
1MND	13.97 (CB)	71.85 (OD2)
1MNE	6.04(N)	32.24 (O)
1VOM	12.28 (CG)	18.26 (CA)
ref1	30.84 (O)	42.79 (OD1)

Table A.15: Minimal and maximal temperature factors for atoms of D454.

code	min	max
1D0X	12.69 (C)	39.72 (CG1)
1D0Y	11.12 (CB)	100.00 (CD1)
1D0Z	8.77 (CG2)	17.51 (CD1)
1D1A	17.74 (N)	54.07 (CD1)
1D1B	13.88 (CG2)	29.38 (CD1)
1D1C	1.00 (CA)	100.00 (CD1)
1FMV	6.40(C)	17.56 (O)
1FMW	1.81 (C)	34.36 (CB)
1G8X	32.85 (CG2)	40.18 (CD1)
1LVK	8.74 (C)	100.00 (CD1)
1MMA	7.87 (C)	51.12 (CD1)
1MMD	2.47 (C)	35.13 (CD1)
1MMG	8.48 (C)	34.64 (CD1)
1MMN	9.72 (C)	35.16 (CD1)
1MND	4.65 (CG2)	75.90 (CD1)
1MNE	6.59 (CB)	34.61 (CD1)
1VOM	12.39 (N)	26.43 (C)
ref1	26.98 (CG2)	31.77 (N)

Table A.16: Minimal and maximal temperature factors for atoms of I455.

code	min	max
1D0X	16.25 (N)	52.64 (OG)
1D0Y	15.91 (C)	48.95 (OG)
1D0Z	10.87 (C)	48.68 (OG)
1D1A	19.40 (CA)	86.30 (O)
1D1B	12.86 (N)	63.82 (OG)
1D1C	11.18 (C)	44.75 (OG)
1FMV	3.63 (CA)	24.63 (OG)
1FMW	11.04 (CA)	32.81 (C)
1G8X	34.53 (O)	45.36 (OG)
1LVK	12.18 (N)	37.63 (OG)
1MMA	20.99 (C)	37.73 (CB)
1MMD	10.13 (CA)	80.97 (OG)
1MMG	18.41 (C)	44.14 (OG)
1MMN	14.88 (N)	34.74 (OG)
1MND	24.99 (OG)	71.26 (CB)
1MNE	5.26 (C)	38.65 (OG)
1VOM	10.48 (CA)	17.01 (C)
ref1	30.06 (N)	33.58 (OG)

Table A.17: Minimal and maximal temperature factors for atoms of S456.

code	min	max
1D0X	11.26 (C)	14.84(N)
1D0Y	15.01 (C)	19.88 (O)
1D0Z	5.57 (CA)	12.01 (O)
1D1A	18.62 (N)	23.31 (CA)
1D1B	9.86 (C)	14.36 (O)
1D1C	9.91 (CA)	29.25 (N)
1FMV	4.92 (N)	11.96 (C)
1FMW	11.23 (C)	16.60 (CA)
1G8X	27.22 (O)	34.26 (N)
1LVK	9.59 (CA)	16.88 (O)
1MMA	5.59 (C)	25.48 (CA)
1MMD	1.00 (C)	8.35 (CA)
1MMG	9.25 (CA)	15.99 (O)
1MMN	10.83 (C)	13.81 (O)
1MND	21.97 (C)	45.33 (CA)
1MNE	8.09 (N)	14.11 (C)
1VOM	8.41 (CA)	16.15 (N)
ref1	29.56 (N)	31.88 (O)

Table A.18: Minimal and maximal temperature factors for atoms of G457.

code	min	max
1D0X	14.33 (CB)	28.94 (C)
1D0Y	15.77 (CB)	34.21 (CD1)
1D0Z	4.75 (CB)	29.12 (C)
1D1A	15.87 (CB)	72.22 (CD2)
1D1B	15.51 (N)	31.84 (CG)
1D1C	7.44 (CB)	28.25 (O)
1FMV	7.36 (CE2)	22.05 (C)
1FMW	16.74 (N)	46.82 (O)
1G8X	30.07 (N)	37.19 (CE1)
1LVK	13.90 (CB)	25.93 (O)
1MMA	15.81 (CB)	65.80 (CD1)
1MMD	6.44 (N)	37.07 (C)
1MMG	12.64 (CB)	35.67 (CZ)
1MMN	3.46 (CB)	34.20 (CD1)
1MND	10.35 (CA)	50.69 (CE2)
1MNE	3.41 (CE2)	38.26 (CD1)
1VOM	11.57 (N)	29.72 (C)
ref1	26.54 (CD1)	32.75 (CD2)

Table A.19: Minimal and maximal temperature factors for atoms of D458.

code	min	max
1D0X	19.36 (CB)	31.28 (CD)
1D0Y	21.07 (O)	40.93 (CD)
1D0Z	13.63 (CA)	28.45 (N)
1D1A	27.53 (CB)	69.72 (CD)
1D1B	20.38 (OE2)	35.77 (CD)
1D1C	15.27 (CA)	34.83 (CD)
1FMV	10.91 (O)	26.79 (CD)
1FMW	19.62 (CA)	41.99 (CD)
1G8X	35.10 (N)	52.76 (OE1)
1LVK	19.79 (O)	66.22 (CD)
1MMA	26.78 (C)	100.00 (CD)
1MMD	12.77 (CA)	100.00 (OE2)
1MMG	21.25 (O)	47.54 (CD)
1MMN	15.04 (CB)	100.00 (CD)
1MND	4.52 (CG)	42.26 (O)
1MNE	16.99 (N)	84.85 (OE2)
1VOM	12.91 (CA)	34.13 (CD)
ref1	32.46 (N)	39.39 (OE1)

Table A.20: Minimal and maximal temperature factors for atoms of E459.

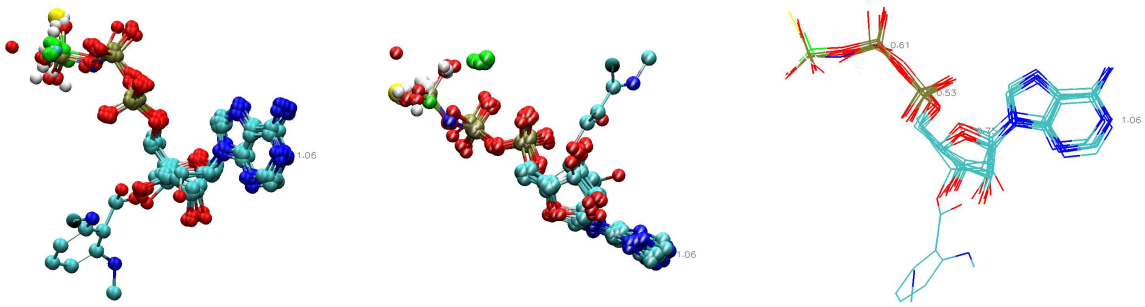

Figure A.1: Overlap of the nucleotide position in 9 structures. Two different viewing angles are shown.

Figure A.2: Positions of the Mg^{2+} ions in structures containing the ADP moiety. The van der Waals sphere of the Mg^{2+} of the 1mmn structure is shown.

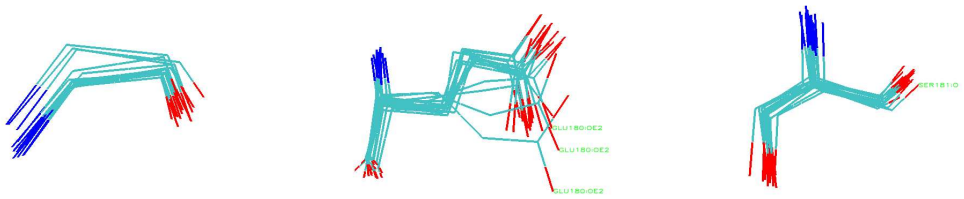

Figure A.3: Overlap of G179, E180, and S181 backbone and sidechain for all structures.

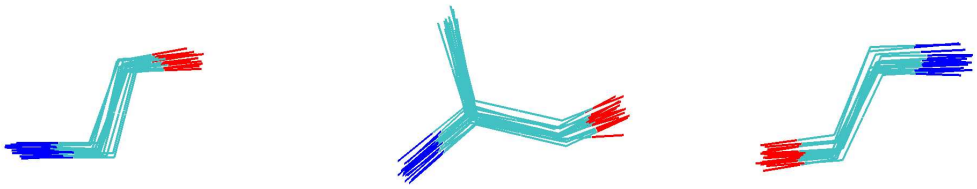

Figure A.4: Overlap of G182, A183, and G184 backbone and sidechain for all structures.

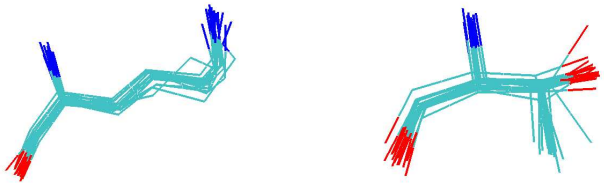

Figure A.5: Overlap of K185 and T186 backbone and sidechain for all structures.

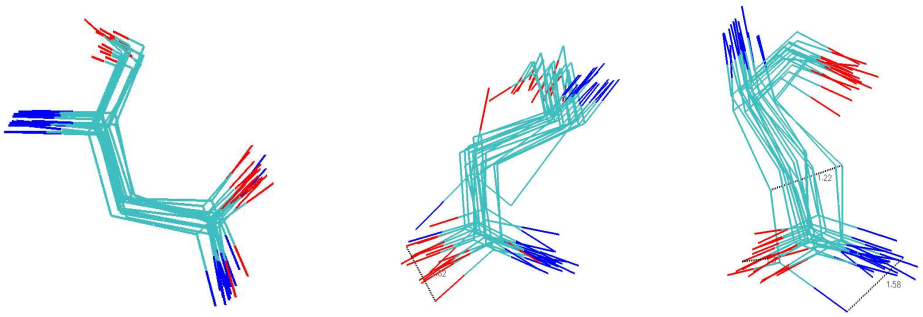

Figure A.6: Overlap of N233, N234, and N235 backbone and sidechain for all structures.

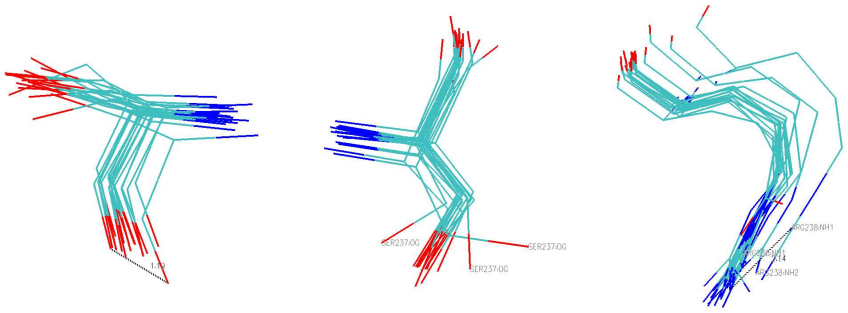

Figure A.7: Overlap of S238, S237, and R238 backbone and sidechain for all structures.

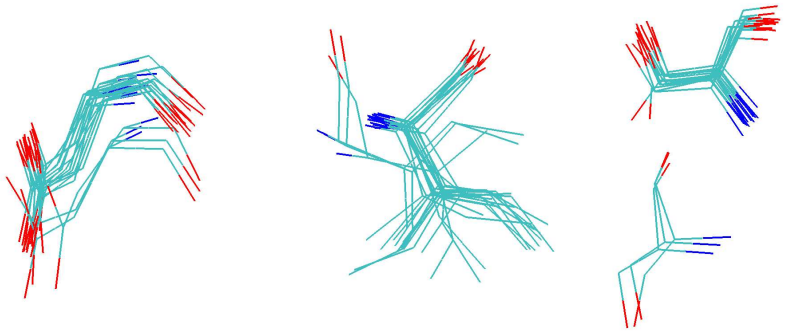

Figure A.8: Overlap of D454, I455, and S456 backbone and sidechain for all structures.

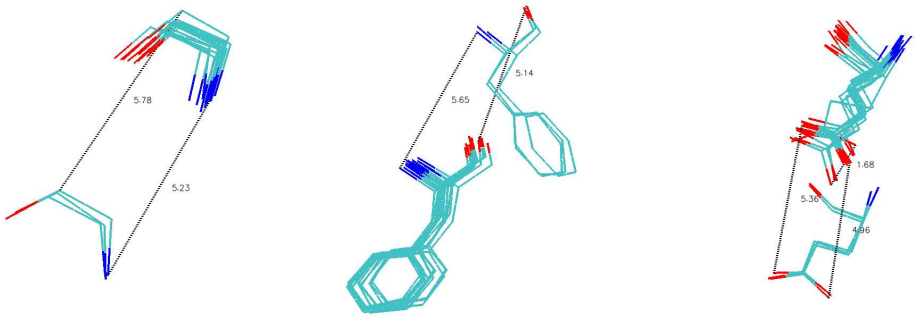

Figure A.9: Overlap of G457, F458, and E459 backbone and sidechain for all structures.

Appendix B

Physical constants used in the thesis

symbol	property	value	unit
k_B	Boltzmann constant	$1.380662 \cdot 10^{-23}$	JK^{-1}
h	Planck constant	$6.626176 \cdot 10^{-34}$	Js
$\kappa = \frac{k_B T}{h}$	prefactor for rate law at T=300K	$6.25094 \cdot 10^{12}$	s^{-1}
N_A	Avogadro constant	$6.022045 \cdot 10^{23}$	mol^{-1}
$R = kN_A$	gas constant	8.31441	$JK^{-1}mol^{-1}$

Table B.1: Physical constants.

Appendix C

Supplementary material for the link-atom procedure

C.1 Implementation

The nomenclature used in this Appendix is as follows (compare Fig. 7.2 in Section 7.2). The link atom is termed q0. The MM and QM boundary atoms are termed m1 and q1. MM (QM) atoms bound to m1 (q1) are termed m2 (q2), and MM (QM) atoms bound to m2 (q2) are termed m3 (q3). Linking is performed in four steps:

1. Generation of the protein according to the CHARMM topology containing the residue X to be split at the bond m1-q1.
2. Addition of the link atom q0 by applying a first patch to the residue X.
3. Declaration the quantum region using the QUANTUM or the GAMESS command in CHARMM.
4. Joining the QM and MM regions by applying a second patch to the residue X.

All classical interactions between atoms in the QM region and some classical interactions involving both QM and MM atoms are automatically removed after invoking the QUANTUM or the GAMESS command. QM/MM classical interactions that are kept are the bond m1-q1, the angles m2-m1-q1 and m1-q1-q2, the dihedrals m3-m2-m1-q1, m2-m1-q1-q2, and, when using the QUANTUM command, m1-q1-q2-q3. In the present link-atom implementation, only those classical interactions which are necessary for positioning the QM region relative to the MM region are kept. The terms m1-q1-q2 and m1-q1-q2-q3, however, are unnecessary for this relative positioning and are therefore deleted in Step 4

using the second patch. To keep the link atom q0 collinear with the m1-q1 bond an angle term with a large force constant is added. Finally, to exclude 1-2 and 1-3 van der Waals interactions, dummy bonds q0-m1 and q0-q2 are introduced with zero force constants. Example patches and necessary additions to the parameter file are given below.

C.2 Topologies

For the CHARMM parameter set 19 (united atoms representation) it is necessary to add the link atom HQ and in addition two dummy hydrogen atoms to the topology. The dummy hydrogens must be positioned to be bound to C_β. The reqired patches are:

- Adding the link atom.

```
PRES ADDQ  0.000 ! Patch for adding a link-hydrogen on any residue
GROUP
ATOM HQ   HQ      0.000
```

- Adding dummy hydrogens.

```
PRES ADD2 0.000 ! add two aliphatic hydrogens on any residue
GROUP
ATOM HDU1  HDU  0.00
ATOM HDU2  HDU  0.00
```

After adding the required dummy atoms, the following CHARMM patches were used for serine, histidine, and aspartate to link the QM and MM regions.

- Serine.

```
PRES SERQ         0.00  ! QM patch for Serine.
                        ! Apply after QUANTUM/GAMESS command
GROUP
ATOM CA  CH1E 0.00 ! changed from 0.1
ATOM H   H    0.30 ! changed from 0.25
ATOM N   NH1 -0.30 ! changed from -0.35
DELETE ANGL CA  CB  OG  ! delete QM/MM terms left by QUANTUM:
!                       ! angles with central atom being QM
DELETE DIHE HG  OG  CB  CA ! dihedrals with 3 QM atoms
!
ANGL CA  CB   HQ        ! keeps link atom colinear with Ca - CB bond
```

```
!                        ! parameter entry needed!
!
BOND HQ  CA              ! dummy bonds for vdW exclusions between CA
BOND HQ  OG              ! and OG
!                        ! parameter entry needed!
BOND HQ HDU1             ! q0 q2'
BOND HQ HDU2             ! q0 q2"
```

- Histidine.

```
PRES HISQ         0.00  ! QM patch for histidine.
                        ! Apply after QUANTUM/GAMESS command
GROUP
ATOM CA  CH1E 0.00 ! changed from 0.1
ATOM H   H    0.30 ! changed from 0.25
ATOM N   NH1 -0.30 ! changed from -0.35
DELETE ANGL CA  CB  CG  ! delete QM/MM terms left by QUANTUM:
!                       ! angles with central atom being QM
DELETE DIHE ND1  CG  CB  CA ! dihedrals with 3 QM atoms
DELETE DIHE CD2  CG  CB  CA
!
ANGL CA  CB   HQ        ! keeps link atom colinear with Ca - CB bond
!                       ! parameter entry needed!
!
BOND HQ  CA             ! dummy bonds for vdW exclusions between CA
BOND HQ  CG             ! and CG
!                       ! parameter entry needed!
BOND HQ HDU1            ! q0 q2'
BOND HQ HDU2            ! q0 q2"
```

- Aspartate.

```
PRES ASPQ  0.00  ! QM patch for aspartate with link on the CA-CB bond
GROUP
ATOM CA  CH1E 0.00 ! changed from 0.1
ATOM H   H    0.30 ! changed from 0.25
ATOM N   NH1 -0.30 ! changed from -0.35
DELETE ANGLE CA CB CG     ! m1 q1 q2
DELETE DIHE CA CB CG OD1  ! m1 q1 q2 q3
DELETE DIHE CA CB CG OD2  ! m1 q1 q2 q3
ANGLE CA CB HQ            ! m1 q1 q0
BOND HQ CA                ! q0 m1
```

```
     BOND HQ CG                ! q0 q2
     BOND HQ HDU1              ! q0 q2'
     BOND HQ HDU2              ! q0 q2"
```

C.3 Parameters

```
BONDS
HQ  CH1E     0.00      1.0   ! dummy bond q0-m1
HQ  HDU      0.00      1.0   ! dummy bond q0-q2' and q0-q2''
HQ  OH1      0.00      1.0   ! dummy bond q0-q2 (Ser)
HQ  C        0.00      1.0   ! dummy bond q0-q2 (His, Asp)

ANGLES
CH1E CH2E  HQ  200.0   0.0   ! angle m1-q0-q1

NONBOND
HQ    0.00   0.00      1.0   ! has no energetic effect
HDU   0.00  -0.022     1.32  ! taken from atom type HA in
                             ! parameter set 22
```

Appendix D

Procedure to determine global NUCS factors with CHARMM

The determination of global non-uniform charge scaling (NUCS) factors (see Section 9.1) has been implemented in a set of CHARMM input scripts that are described in this Section.

The procedure can be summarized as follows: First, a group I is selected, for which the scaling factor is to be determined. The charges on this group are set to their respective values. The charges of all other groups are set to zero. Then the electrostatic potential ϕ_I due to the charges of group I is calculated using a finite difference scheme with a focusing approach. It is then interpolated to all atom positions $\phi_I(\vec{r}_j)$.

Next, the electrostatic interaction energies E_{IJ}^{vac} and E_{IJ}^{solv} between the source group I and all other groups J are calculated. From this the pairwise dielectric constants ϵ_{IJ} are determined. If $\epsilon_{IJ} < 0$, the energies are zeroed, otherwise they are put to absolute values. Finally, the dielectric constant ϵ_I and the scaling factor λ_I are determined.

D.1 Determination of λ_I with CHARMM

This procedure has been set up in a set of CHARMM input files. CHARMM version 28 or higher is necessary, as the scripts make use of the PBEQ module. The structure and file names of the input scripts are visualized in Figure D.1.

The user only has to modify *get_box.inp*, *generate.str*, and *settings.str*.

First, the finite difference box dimensions must be determined (see Section D.2). For the subsequent determination of NUCS factors, the input file to be piped into CHARMM is *scaling.inp*. This file streams *generate.str*, where the user must set up the system (i.e. read in topology and parameter files, sequence information, generate the protein structure file

with all segments and read in the coordinates). After that *settings.str* is streamed, where the user must specify several variables needed for the calculation. In *initialize.str* several variables are initialized. After an energy call to initialize the energy function, *loop.str* is streamed that loops from the first to the last residue, for which scaling factors shall be determined. After definition of the source group I, $calcfactors.str$ is streamed. There, *pbeq.str* is streamed to determine the electrostatic potential $\phi_I(\vec{r}_j)$. The scaling factor λ_I is then calculated in *scaling.str*, which streams *loop2.str* and determines the actual λ_I. *loop2.str* loops over all residues within the cutoff distance around group I and streams *calculation.str*. In this stream file, the pairwise interaction energies in vacuum E_{IJ}^{vac} and in solution E_{IJ}^{solv} are calculated (in the stream files *vacuum.str* and *solution.str*, respectively) and the check for change in sign is performed.

The stream files *terminal.str*, *terminal2.str*, *terminal3.str*, and *terminal4.str* are used in *loop.str* and *loop2.str*, respectively, to ensure adequate group definition at segment borders in CHARMM.

D.2 How to get the correct box dimensions

The CHARMM input file *get_box.inp* can be used to derive the correct box dimensions. It streams *generate.str* and *settings.str*. In *settings.str* only the energy function variables must be correctly set. Via COOR ORIENT and COOR STAT commands the dimensions of the system are printed. ATTENTION: Make sure that you use the CHARMM oriented coordinates in your calculation of scaling factors! The oriented coordinates may be written to a new coordinate file, whose name must be given in *get_box.inp*. The output gives the minimum and maximum X-, Y-, and Z-coordinates of the centers of the atoms. To this maximum distance, a total of 6 Å have to be added (1.5 Å for the atom radius plus 1.5 Å for the water radius on each side) in order to get to an estimate of the solvent accessible surface dimensions.

For the first PBEQ calculation, 25 Å should be added on each side of the system to get the size of the finite difference box. A grid spacing of 4.0 Å is recommended. The number of grid points can then be calculated as $boxsize/gridspacing = (Xmax - Xmin + 6 + 2 \cdot 25)/4.0$. In order not to get any box size problems, the values should always be rounded up. CHARMM only takes odd numbers, if an even number of grid points is specified, it is automatically incremented by one. For the second PBEQ calculation, 10 Å should be added on each side and a grid spacing of 2.0 Å is recommended. For the third calculation it should be 5 Å and 1.0 Å, respectively.

D.2.1 User-defined settings

The user must specify the following variables in *settings.str*:

- General settings. The variable $FIRSTRESIDUE$ should be set to the first residue for which the scaling factor should be determined. If the residue $FIRSTRESIDUE-1$ has a side chain, this scaling factor is also determined. $LASTRESIDUE$ is the last residue, for which λ is determined. $BACK1$ of $LASTRESIDUE+1$ is included.

  The names of the following output files must be given:

  1. Scaling factors. The final scaling factors are written into the WMAIN column of a .crd coordinate file. The name is specified in $OUTPUT$
  2. Interaction energies E_I^{solv}. The reference interaction energies between each group and the rest of the protein calculated from the Poisson-Boltzmann calculation are written into the WMAIN column of a .crd coordinate file. The name is specified in $OUTPUT1$
  3. Pairwise interaction energies. The pairwise interaction energies E_{IJ}^{vac} and $E_I^{solv}J$ that are evaluated during the calculation are printed into an output file specified in $OUTPUT2$.

- Energy function settings. The parameter file version has to be given in $PARAM$ (19 for extended carbon model or 22 for an explicit hydrogen representation), the cutoff type in $ELECTR$ (cdiel for a constant dielectric permeability and rdiel for a distance-dependent dielectric permeability), and cutoff values in the variables $UCTONNB$, $UCTOFNB$, $UCUTNB$, $UWMIN$ (U for user-defined).

- PBEQ settings. These settings are needed for the calculation of E_I^{solv} with the Poisson-Boltzmann module in CHARMM (PBEQ). $offset$ is a value added to the van der Waals radii used for the determination of the dielectric boundary. sw gives the value of the smoothing window to be used for smoothing the transition of the dielectric constant from $EPSWAT$ to $EPSPROT$ at the dielectric boundary between high and low dielectricum. A combination of $sw = 1.5$ and $offset = 0.7$ has been found to reproduce accurate electrostatic calculations done with UHBD. The ion concentration in the solvent is given in CON and the temperature in TEM. The grid spacings and number of grid points in all direction are given in $DCELn$ and $NCLdn$, where $n = 1, 2, 3$ and $d = X, Y, Z$. To determine these values see Section D.2.

- Useratoms. For each group, a single unique atom must be given as a representative. For standard groups such as protein sidechains, protein back bones, or water this is automatically done. The user must, however, provide this definitions for each additional group such as retinal, ions, or nucleotides.

D.2.2 Initialization

The initialization of storage vectors is done in *initialize.str*. The following storage vectors are used:

- SCA1 storage of radii
- SCA2 storage of charges
- SCA3 intermediate storage of pairwise interaction energies in vacuum
- SCA4 intermediate storage of pairwise interaction energies in solution
- SCA5 intermediate storage of pairwise scaling factors
- SCA6 storage vector for scaling factors
- SCA7 intermediate storage of pairwise interaction energies in solution within cutoff distance E_{IJ}^{solv}
- SCA8 storage of interaction energies in solution between each group and the rest of the protein within cutoff distance E_{I}^{solv}
- SCA9 intermediate storage of electrostatic potential

initialize.str also streams *defi19.str* or *defi22.str*, where the definition of backbone and sidechain atoms is done. The backbone groups are defined over two residues: BACK1 includes the atom types C_α, H_α, N, HN; BACK2 includes C and O for param22. For param19 the definitions are accordingly. The termini and capping groups ACE and CBX (param19 only) are treated correctly. All atoms not belonging to BACK1 or BACK2 are supposed to be sidechain atoms. Thus, each protein residue is split into two groups, a backbone group (spanning two amino acids) and a sidechain group.

All non-protein moieties, such as water, ions, or ligands, are treated as protein residues without backbones. A scaling factor is determined for each group. The prolines are treated specially in that only one scaling factor is determined for both backbone and side chain atoms.

Finally, a selection is defined that contains exactly one atom per group.

D.3 Determination of NUCS factors

The NUCS factors are determined automatically once the settings have been correctly made. The input file *scaling.inp* must be piped into CHARMM. The print and warning levels are set so as to suppress most of the output. Otherwise, the output files will be very large.

D.4 Testing and counterscaling of NUCS factors

The NUCS factors as determined by the procedure described above are only raw factors. They will be denoted with λ'_I in this section. Having calculated the λ'_I it is possible to calculate the electrostatic interaction energy between a group I and the rest of the protein in vacuum $E_I^{shield'}$ using the raw scaling factors. The $E_I^{shield'}$ are an estimate for the target E_I^{solv}.

$$E_I^{shield'} = \sum_{i \in I} \sum_{j \in J \neq I}^{groups} \frac{q_i}{\lambda'_I} \frac{q_j}{\lambda'_J} \frac{1}{r_{ij}} \tag{D.1}$$

The λ'_I may be counterscaled to give the NUCS factors λ_I. The E_I^{shield} calculated with the λ_I give a more accurate correlation with the E_I^{solv}. Counterscaling is achieved by dividing the raw scaling factors λ'_I by a factor γ,

$$E_I^{shield} = \sum_{i \in I} \sum_{j \in J \neq I}^{groups} \frac{q_i}{\lambda_I} \frac{q_j}{\lambda_J} \frac{1}{r_{ij}} = \gamma^2 \sum_{i \in I} \sum_{j \in J \neq I}^{groups} \frac{q_i}{\lambda'_I} \frac{q_j}{\lambda'_J} \frac{1}{r_{ij}} = \gamma^2 E^{shield'} \tag{D.2}$$

The E_I^{shield} are an estimate to the E_I^{solv}, so γ^2 can be determined by doing a least square fit of the E_I^{solv} against the $E_I^{shield'}$ according to Equation D.2. For reasons of accuracy, γ may be determined separately for backbone groups (γ_{BACK}) and side chain groups (γ_{SIDE}).

D.4.1 Calculation of E_I^{solv}, $E_I^{shield'}$, and γ

To get the data necessary for the least squares fit, the E_I^{solv} must be extracted from the output file specified in $OUTPUT1$ and the $E_I^{shield'}$ must be calculated. The procedure to do this has been set up in a set of CHARMM input files. The structure and file names are visualized in Figure D.2.

The user only has to modify *generate.str* and *settings.str*.

The input file to be piped into the CHARMM calculation is *test_factors.inp*. This file streams *generate.str*, which should be exactly the same generation file as used before to calculate the scaling factors. After that *settings.str* is streamed. Attention: The settings necessary for the testing of the scaling factors are slightly different from the corresponding settings to calculate the scaling factors. The coordinate files with the raw scaling factors λ'_I in the WMAIN column and the interaction energies in solution E_I^{solv} need to be specified. In *initialize.str* several variables are initialized. After an energy call to initialize the energy

function, *loop.str* is streamed. It loop over all residues for which the scaling factors shall be tested. After definition of the source group I, *loop.str* streams *calenergy.str*, which in turn streams *vacuum.str* to calculate E_I^{vac} with unscaled charges, extracts E_I^{solv}, and streams *shield.str* to calculate $E_I^{shield'}$ with scaled charges using the raw scaling factors. Two output files are generated, in which the calculated values for backbone groups and side chain groups are tabulated. These values must then be used in another program to determine γ_{BACK} and γ_{SIDE} by performing linear regression fits.

This can for instance be done with the programs provided as tools by Frank Noe. An example input line would be

./select_columns -s1 "4,5" < back_cdiel.dat | ./linear-regression

The square root of the slope of the resulting linear regression through the origin then gives γ_{BACK}.

After having determined γ_{BACK} and γ_{SIDE} the E_I^{shield} are determined by setting both BACKFAC and SIDEFAC to their respective values and repeating the calculation.

The root-mean-square deviation can be determined accordingly:

./select_columns -s1 "4,5" < back_cdiel.dat | ./rmsd

D.4.2 User-defined settings

The user must specify the following variables in *settings.str*

- NUCS factors and energies. The variable $INPUTFAC$ must be set to the path and file name of the .crd coordinate file with NUCS factors in the WMAIN column. The variable $INPUTENER$ must be set to the path and file name of the .crd coordinate file with the interaction energies in solution E_I^{solv} in the WMAIN column.
- General settings. The variables $FIRSTRESIDUE$ and $LASTRESIDUE$ should be set to the numbers of the first and last residues, for which the scaling factors should be tested.
- Energy function settings. These settings should be identical to the specifications used for the calculation of the scaling factors (see Section D.2.1). They may also be set to a different energy function, if the interaction energies in vacuum shall be recomputed for comparison.
- Charge scaling settings. The variables $BACKFAC$ and $SIDEFAC$ should be set to 1.0, if E_I^{vac}, E_I^{solv}, and $E_I^{shield'}$ are to be calculated. They should be set to γ_{BACK} and γ_{SIDE}, respectively, if the E_I^{shield} are to be calculated.

  The output file names must be given in the variables $OUTPUTBACK$ and $OUTPUTSIDE$, respectively.

D.4.3 Initialization

initialize.str streams *defi19.str* or *defi22.str*, calculates scaled charges and initializes the following storage vectors:

- SCA1 storage of scaling factors
- SCA2 storage of scaled charges
- SCA3 storage of original partial atomic charges
- SCA4 storage of E_I^{solv}
- SCA5 storage of E_I^{vac} (original charges)
- SCA6 storage of $E_I^{shield'}$ or E_I^{shield} (scaled charges)

D.5 How to use the scaling factors in subsequent calculations

The modified charge scaling factors $\lambda_I = \lambda_I' \cdot \gamma$ can easily be used for subsequent molecular mechanics calculations such as energy minimizations or reaction path simulations. For this, a coordinate file with the raw scaling factors in the WMAIN column must be read. The modification factors $BACKFAC = \gamma_{BACK}$ and $SIDEFAC = \gamma_{SIDE}$ must have been determined before. Then, the CHARGE vector within CHARMM must be modified using the following lines of CHARMM input code:

```
STREAM "defi??.str"   ! ?? = 19, 22

! store scaling factors in SCA1
SCAL SCA1 = WMAIN

! modify scaling factors
SET BACKFAC ??
SET SIDEFAC ??
SCAL SCA1 MULT @BACKFAC SELE BACK1 .OR. BACK2 END
SCAL SCA1 MULT @SIDEFAC SELE SIDE END

! scale charges
SCAL SCA1 RECI
SCAL CHARGE PROD SCA1 SELECT ALL END
```

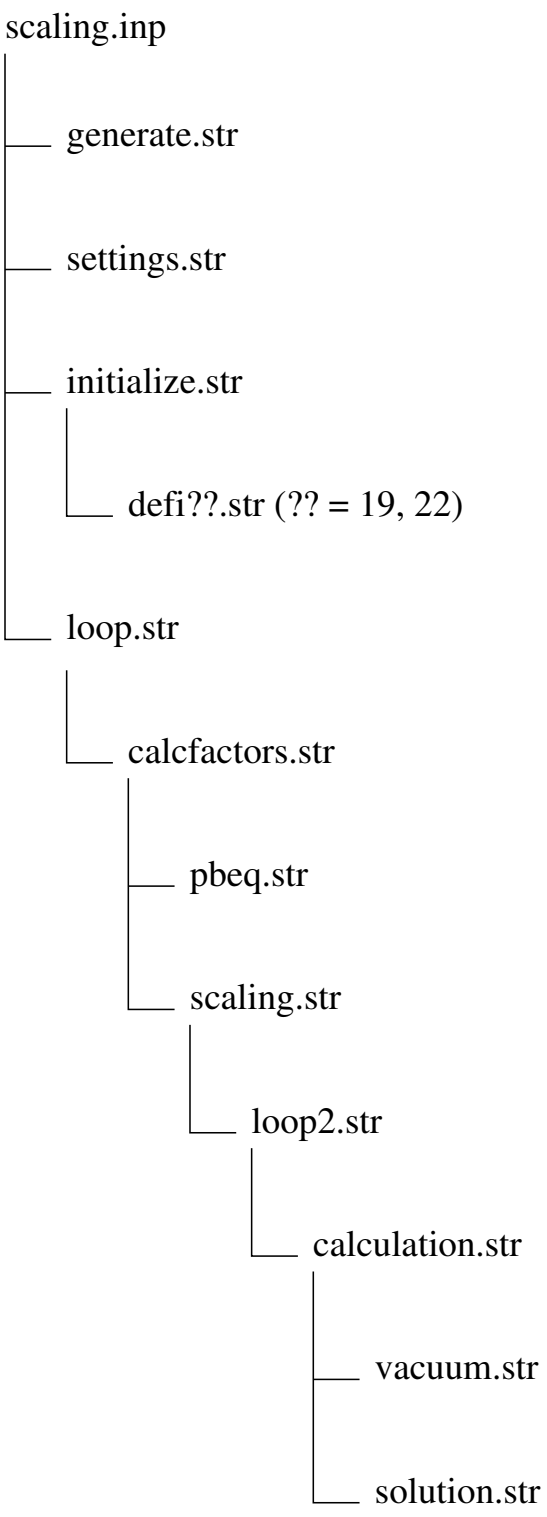

Figure D.1: Structure of input and stream files for the determination of scaling factors.

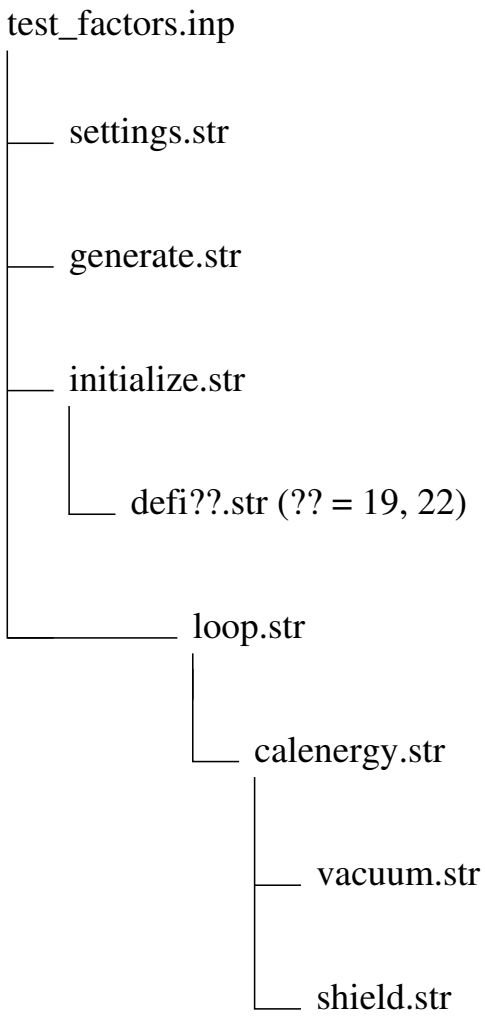

Figure D.2: Structure of input and stream files for testing the previously determined scaling factors.

Appendix E

Procedure how to combine global and targeted charge scaling

E.1 Determination of NUCS factors in the target region

The scripts for the determination of global scaling factors for the groups within the flexible and QM regions are almost identical to the scripts described in Appendix D. The only difference is that the user must define a target region in *settings.str*. In QM/MM applications this target region should comprise both the QM and the flexible regions.

This target region is increased in *target.str* (streamed from *initialize.str*) so as to ensure that the boundary between the target and the nontarget regions is along the group definitions used for the calculation of NUCS factors. All charges outside the target region are set to zero. Thus, scaling factors are calculated only for groups within the target region, and only interactions between groups belonging to the target region are considered.

The scaling factors for groups in the target region must be calculated, counterscaled, and verified as described in NUCS_global_protocol. The counterscaled scaling factors are provided in a CHARMM coordinate file in which the WMAIN column is set to the values of the scaling factors for atoms belonging to the target region and to zero for atoms outside the target region.

E.2 Determination of NUCS factors in the non-target region

The procedure to determine targeted scaling factors in the non-target region as described in Section 9.2 has been implemented in a set of CHARMM input files.

The user only has to modify *generate.str* and *settings.str*. *generate.str* should be identical to the file used in the determination of global scaling factors. The settings provided in *settings.str* should be identical to the settings used in the determination of global scaling factors, apart from the following differences:

- The name of the CHARMM coordinate file with scaling factors for atoms in the target region must be given in $INPUTFAC$.
- The name of the output CHARMM coordinate file which will contain the scaling factors for atoms outside the target region must be given in $OUTPUT$.
- A filename must be provided in $OUTPUT2$, into which the electrostatic interaction energies between the groups outside the target region and the target region calculated with Poisson-Boltzmann and with a Coulomb potential, the distance between the group and the target region, and the scaling factor are given

The input file to be piped into CHARMM is *interaction.inp*. This streams *generate.str*, *settings.str*, and *initialize.str*. In *initialize.str*, *defi19.str* or *defi22.str* and *target.str* are streamed. After an energy call to initialize the energy function, the electrostatic potential due to the atoms of the target region is determined by numerical solution of the Poisson-Boltzmann equation in *pbeq.str*. The vector $SCA9$ contains the resulting potential at the positions of all atoms. *loop_inter.str* then loops over all residues and streams *calculation.str*. In *calculation.str*, *vacuum.str* (returns $E^{vac}_{I,target}$) and *solution.str* (returns $E^{PB}_{I,target}$) are streamed for all groups that are outside the target region. Then, the scaling factor is determined and stored to $SCA6$. Moreover, the distance between the group and the target region is computed. The group identification, $E^{vac}_{I,target}$, $E^{PB}_{I,target}$, the distance, and the scaling factor are written to $OUTPUT2$.

Finally, *get_all_factors.inp* can be used to concatenate the global scaling factors determined for the target region and the targeted scaling factors determined for the non-target region.

E.3 Validation of the NUCS factors

CHARMM input scripts for the validation of the scaling factors according to the three aspects given in Section 9.3 have been prepared. In all sets of scripts only *settings.str* and

generate.str must be modified by the user according to the settings used in the determination of the scaling factors. Additional settings are described below.

In all cases, interaction energies are computed with Poisson-Boltzmann electrostatics to provide reference values. The corresponding interaction energies are then determined using a Coulomb potential with either scaled or unscaled charges, allowing for a comparison.

1. **Potential in QM region.** The input file to be piped into CHARMM is *interactions_with_qm.inp*. This streams *generate.str*, *settings.str*, and *initialize.str*. In *settings.str* the user must specify the QM region, the file name of the CHARMM coordinate file with scaling factors for all groups in WMAIN, and the output file name. The electrostatic potential due to the atoms of the QM region is determined in *pbeq.str* and returned in *SCA*9. *loop_screen.str* loops over *all* groups and streams *calculation.str*. *calculation.str* streams *solution.str* that returns the Poisson-Boltzmann interaction energy, calculates the Coulomb interaction energy with scaled and unscaled charges, determines the distance between the group and the QM region, and writes the output.

2. **Potential in flexible region.** The input file to be piped into CHARMM is *interactions.inp*. This streams *generate.str*, *settings.str*, and *initialize.str*. In *settings.str* the user must specify the target and QM regions, the file name of the CHARMM coordinate file with scaling factors for all groups in WMAIN, and the output file name. *loop.str* loops over all groups and streams *calcener.str* for groups that are in the target region but not in the QM region. The electrostatic potential due to the atoms of the group is determined in *pbeq.str* and returned in *SCA*9. Then, *calculation.str* is streamed that streams *solution.str* that returns the Poisson-Boltzmann interaction energy between the group and all other groups, calculates the Coulomb interaction energy with scaled and unscaled charges and writes the output.

3. **Group:group interactions in flexible region.** The input file to be piped into CHARMM is *interactions.inp*. This streams *generate.str*, *settings.str*, and *initialize.str*. In *settings.str* the user must specify the target and QM regions, the file name of the CHARMM coordinate file with scaling factors for all groups in WMAIN, and the output file name. *loop.str* loops over all groups and streams *calcener.str* for groups that are in the target region but not in the QM region. The electrostatic potential due to the atoms of the group is determined in *pbeq.str* and returned in *SCA*9. Then, *loop2.str* loops over all groups and streams *calculation.str* for groups that are in the target region but not in the QM region. *calculation.str* streams *solution.str* that returns the Poisson-Boltzmann interaction energy between the two groups, calculates the Coulomb interaction energy with scaled and unscaled charges, determines the distance between the two groups and writes the output.

Bibliography

[1] Wolfgang Pauli. *Wissenschaftlicher Briefwechsel mit Bohr, Einstein, Heisenberg u.a. Band IV.* Springer, Heidelberg, berlin, New York, 2001.

[2] F. H. Westheimer. Why Nature chose phosphates. *Science*, 235:1173–1178, 1987.

[3] George S. Bloom. Preface. *Biochim. Biophys. Acta*, 1496:1–2, 2000.

[4] Thomas D. Pollard. The cytoskeleton, cellular motility and the reductionist agenda. *Nature*, 422:741–745, 2003.

[5] Manfred Schliwa, editor. *Molecular motors*. Wiley-VCH, Weinheim, 2003.

[6] Manfred Schliwa and Günther Woehlke. Molecular motors. *Nature*, 422:759–765, 2003.

[7] Ronald D. Vale. The molecular motor toolbox for intracellular transport. *Cell*, 112:467–480, 2003.

[8] George Oster and Hongyun Wang. Reverse engeneering a protein: the mechanochemistry of ATP synthase. *Biochim. Biophys. Acta*, 1458:482–510, 2000.

[9] Manuel Diez, Boris Zimmermann, Michael Börsch, Marcelle König, Enno Schweinberge, Stefan Steigmiller, Rolf Reuter, Suren Felekyan, Volodymyr Kudryavtsev, Claus A. M. Seidel, and Peter Gräber. Proton-powerde subunit rotation in single membrane-bound F_oF_1-ATP synthase. *Nature Struct. Mol. Biol.*, 11:135–141, 2004.

[10] Hiroyasu Itoh, Akira Takahashi, Kengo Adachi, Hiroyuki Noju, Ryohei Yasuda, Masasuke Yoshida, and Kazuhiko Kinosita, Jr. Mechanically driven ATP synthesis by F_1-ATPase. *Nature*, 427:465–458, 2004.

[11] Jianpeng Ma, Terence C. Flynn, Qiang Cui, Andrew G. W. leslie, John E. Walker, and Martin Karplus. A dynamic analysis of the rotation mechanism for conformational change in F_1ATPase. *Structure*, 10:921–931, 2002.

[12] Iris Antes, David Chandler, Hongyun Wang, and George Oster. The unbinding of ATP from F_1-ATPase. *Biophys. J.*, 85:695–706, 2003.

[13] Yi Qin Gao, Wei Yang, Rudolph A. Marcus, and Martin Karplus. A model for the cooperative free energy transduction and kinetics of ATP hydrolysis by F_1-ATPase. *Proc. Natl. Acad. Sci.*, 100:11339–11344, 2003.

[14] James R. Sellers. Myosins: a diverse superfamily. *Biochim. Biophys. Acta*, 1496:3–22, 2000.

[15] Jonathan S. Berg, Bradford C. Powell, and Richard E. Cheney. A millennial myosin census. *Mol. Biol. Cell*, 12:780–794, 2001.

[16] A. Danilewsky. Myosin, seine darstellung, eigenschaften, umwandlung in syntonin und rückbildung aus demnselben. *Z. Physiol. Chem.*, 5:158, 1881.

[17] W. D. Halliburton. On muscle-plasma. *J. Physiol.*, 8:133–202, 1887.

[18] Otto von Fürth. Die Kolloidchemie des Muskels und ihre Beziehungen zu den Problemen der Kontraktion und der Starre. *Ergebn. Physiol.*, 17:363–571, 1919.

[19] Claudia Veigel, Lynne M. Coluccio, James D. Jontes, John C. Sparrow, Ronald A. Milligan, and Justin E. Molloy. The motor protein myosin-I produces its working stroke in two steps. *Nature*, 398:530–533, 1999.

[20] Barbara Barylko, Derk D. Binns, and Joseph P. Albanesi. Regulation of the enzymatic and motor acrivities of myosin I. *Biochim. Biophys. Acta*, 1496:23–35, 2000.

[21] Setsuko Fujia-Becker, Ulrike Dürrwang, Muriel Erent, Richard J. Clark, Michael A. Geeves, and Dietmar J. Manstein. Changes in Mg^{2+}-ion concentration and heavy chain phosphorylation regulate the motor activity of a class-I myosin.

[22] Michael A. Geeves, Cynthia Perreault-Micale, and Lynne M. Coluccio. Kinetic analyses of a truncated mammalian myosin I suggest a novel isomerization event preceding nucleotide binding. *J. Biol. Chem.*, 275:21624–21630, 2000.

[23] James D. Jontes and Ronald A. Milligan. Brush border myosin-I structure and ADP-dependent conformational changes revealed by cryelectron microscopy and image analysis. *J. Cell. Biol.*, 139:683–693, 1997.

[24] E. Michael Ostap and Thomas D. Pollard. Biochemical kinetic characterization of the *Acanthamoeba* myosin-I ATPase. *J. Cell. Biol.*, 132:1053–1060, 1996.

[25] Martin Bähler. Are class III and class IX myosins motorized signalling molecules? *Biochim. Biophys. Acta*, 1496:52–59, 2000.

[26] Sotaro Uemura, Hideo Higuchi, Adrian O. Olivares, Enrique M. De La Cruz, and Shinichi Ishiwata. Mechanochemical coupling of two substeps in a single myosin V motor. *Nature Struct. Mol. Biol.*, 11:877–883, 2004.

[27] Samara Reck-Peterson, D. William Provance, Jr., Mark S. Mooseker, and John A. Mercer. Class V myosins. *Biochim. Biophys. Acta*, 1496:36–51, 2000.

[28] Fusheng Tang, Emily J. Kauffman, Jennifer L. Novak, Johnathan J. Nau, Natalie L. Catlett, and Lois S. Weisman. Regulated degradation of a class V myosin receptor directs movement of the yeast vacuole. *Nature*, 422:87–92, 2003.

[29] David Altmann, H. Lee Sweeney, and James A. Spudich. The mechanism of myosin VI translocation and its load-induced anchoring. *Cell*, 116:737–749, 2004.

[30] Michele C. Kieke and Margaret A. Titus. The myosin superfamily: an overview. In Manfred Schliwa, editor, *Molecular motors*, pages 3–44. Wiley-VCH, Weinheim, 2003.

[31] Justin E. Molloy and Claudia Veigel. Myosin motors walk the walk. *Science*, 300:2045–2046, 2003.

[32] Amber L. Wells, Abel W. Lin, Li-Qiong Chen, Daniel Safer, Shane M. Cain, Tama Hasson, Bridget O. Carragher, Ronald A. Milligan, and H. Lee Sweeney. Myosin VI is an actin-based motor that moves backwards. *Nature*, 401:505–508, 1999.

[33] Manfred Schliwa. Myosin steps backwards. *Nature*, 401:431–432, 1999.

[34] Ahmet Yildiz, Joseph N. Forkey, Sean A. McKinney, Taekjip Ha, Yale E. Goldman, and Paul R. Selvin. Myosin V walks hand-over-hand: single fluorophore imaging with 1.5-nm localisation. *Science*, 300:2061–2065, 2003.

[35] Günther Woehlke and Manfred Schliwa. Directional motility of kinesin motor proteins. *Biochim. Biophys. Acta*, 1496:117–127, 2000.

[36] Joseph R. Marszalek and Lawrence S. B. Goldstein. Understanding the functions of kinesin-II. *Biochim. Biophys. Acta*, 1496:142–150, 2000.

[37] F. Jon Kull, Elena P. Sablin, Rebecca Lau, Robert J. Fletterick, and Ronald D. Vale. Crystal structure of the kinesin motor domain reveals a structural similarity to myosin. *Nature*, 380:550–559, 1996.

[38] Michael A. Geeves and Kenneth C. Holmes. Structural mechanism of muscle contraction. *Ann. Rev. Biochem.*, 68:687–728, 1999.

[39] Kenneth C. Holmes and Michael A. Geeves. The structural basis of muscle contraction. *Philos. Trans. R. Soc. B*, 355:419–431, 2000.

[40] Michael A. Geeves and Kenneth C. Holmes. The molecular mechanism of muscle contraction. *Adv. Protein Chem.*, 71:161–193, 2005.

[41] Roger S. Goody. The missing link in the muscle cross-bridge cycle. *Nature Struct. Biol.*, 10:773–775, 2003.

[42] Hans M. Warrick, Arturo DeLozanne, Leslie A. Leinwand, and James A. Spudich. Conserved protein domains in a myosin heavy chain gene from *Dictyostelium discoideum*. *Proc. Natl. Acad. Sci.*, 83:9433–9437, 1986.

[43] Arturo DeLozanne, Mindy Lewis, James A. Spudich, and Leslie A. Leinwand. Cloning and characterization of a nonmuscle myosin heavy chain cDNA. *Proc. Natl. Acad. Sci.*, 82:6807–6810, 1985.

[44] http://ca.expasy.org/cgi bin/niceprot.pl?P08799. Swissprot entry for *Dictyostelium discoideum* myosin II heavy chain (accession number p08799).

[45] D. Mornet, P. Pantel, E. Audemard, and R. Kassab. The limited tryptic cleavage if chymotryptic S-1: an approach to the characterization of the actin site in myosin heads. *Biochem. Biophys. Res. Comm.*, 89:925–932, 1979.

[46] Shinji Itakura, Hisashi Yamakawa, Yoko Yano Toyoshima, Akihiko Ishijima, Takaaki Kojima, Yoshie Harada, Tosio Yanagida, Takeyuki Wakabayashi, and Kazuo Sutoh. Force-generating domain of myosin motor. *Biochem. Biophys. Res. Comm.*, 196:1504–1510, 1993.

[47] Peteranne B. Joel, H. Lee Sweeney, and Kathleen M. Trybus. Addition of lysines to the 50/20 kda junction of myosin strengthens weak binding to actin without affecting the maximum ATPase activity. *Biochemistry*, 42:9160–9166, 2003.

[48] Naoya Sasaki, Reiko Ohkura, and Kazuo Sutoh. *Dictyostelium* myosin II as a model to study the actin-myosin interactions during force generation. *J. Muscle Res. Cell Mot.*, 23:697–702, 2002.

[49] Taro Q. P. Uyeda, Bruce Patterson, Leonardo Mendoza, and Yuichi Hiratsuka. Amino acids 519-524 of *Dictyostelium* myosin II form a surface loop that aids actin binding by facilitating a conformational change. *J. Muscle Res. Cell Mot.*, 23:685–695, 2002.

[50] M. Anson, Michael A. Geeves, S. E. Kurzawa, and Dietmar J. Manstein. Myosin motors with artificial lever arms. *EMBO J.*, 15:6069–6074, 1996.

[51] J. C. Seidel. The effects of nucleotides and Mg^{2+} on the electron spin resonance spectra of myosin spin lebeled at the s_2 thiol groups. *Arch. Biochem. Biophys.*, 152:839–848, 1972.

[52] J. C. Seidel, M. Chopek, and J. Gergely. Effects of nucleotides and pyrophosphate on apin labels bound to S_1 thiol groups of myosin. *Biochemistry*, 9:3265–3272, 1970.

[53] Clyde A. Smith and Ivan Rayment. Active site comparisons highlight structural similarities between myosin and other P-loop proteins. *Biophys. J.*, 70:1590–1602, 1996.

[54] Sungjo Park, Katalin Ajtai, and Thomas P. Burghardt. Inhibition of myosin ATPase by metal fluoride complexes. *Biochim. Biophys. Acta*, 1430:127–140, 1999.

[55] F. Jon Kull, Ilme Schlichting, Andreas Becker, Martin Kollmar, Dietmer J. Manstein, and Kenneth C. Holmes. An alternate conformation for MgADP-beryllium fluoride bound myosin II.

[56] M. Jamie T. V. Cope, James Whisstock, Ivan Rayment, and John Kendrick-Jones. Conservation within the myosin motor domain: implications for structure and function. *Structure*, 4:969–987, 1996.

[57] Andrew J. Fisher, Clyde A. Smith, James Thoden, Robert Smith, Kazuo Sutoh, Hazel M. Holden, and Ivan Rayment. X-ray structures of the myosin motor domain of *Dictyostelium discoideum* complexed with MgADP.BeF_x and MgADP.AlF_4^-. *Biochemistry*, 34:8960–8972, 1995.

[58] Yale E. Goldman. Kinetics of the actomyosin ATPase in muscle fibers. *Ann. Rev. Physiol.*, 49:637–654, 1987.

[59] David R. Trentham, J. F. Eccleston, and Clive R. Bagshaw. Kinetic analysis of ATPase mechanisms. *Q. Rev. Biophys.*, 9:217–281, 1976.

[60] Ronald D. Vale. Switches, latches, and amplifiers: common themes of G proteins and molecular motors. *J. Cell. Biol.*, 135:291–302, 1996.

[61] F. Jon Kull, Ronald D. Vale, and Robert J. Fletterick. The case for a common ancestor: Kinesin and myosin motor proteins and G proteins. *J. Muscle Res. Cell Mot.*, 19:877–886, 1998.

[62] Ivan Rayment, Wojciech R. Rypniewski, Karen Schmidt-Bäse, Robert Smith, Diana R. Tomchick, Matthew M. Benning, Donald A. Winkelmann, Gary Wesenberg, and Hazel M. Holden. Three-dimensional structure of myosin subfragment-1: a molecular motor. *Science*, 261:50–58, 1993.

[63] A. Houdusse, Andrew G. Szent-Györgyi, and Carolyn Cohen. Three conformational states of scallop myosin S1. *Proc. Natl. Acad. Sci.*, 97:11238–11234, 2000.

[64] Anne Houdusse, Vassilios N. Kalabokis, Daniel Himmel, Andrew Szent-Györgyi, and Carolyin Cohen. Atomic structure of scallop myosin subfragment S1 complexed with MgADP: a novel conformation of the myosin head. *Cell*, 97:459–470, 1999.

[65] Daniel M. Himmel, S. Gourinath, L. Reshetnikova, Y. Shen, Andreq G. Szent-Györgyi, and Carolyn Cohen. Crystallographic findings on the internally uncoupled and near-rigor states of myosin: further insights into the mechanics of the motor. *Proc. Natl. Acad. Sci.*, 99:12645–12650, 2002.

[66] Leslie E. W. LaConte, Josh E. Baker, and David D. Thomas. Transient kinetics and mechanics of myosin's force-generating rotation in muscle: Resolution of millisecond rotational transitions in the spin-labeled myosin light-chain domain. *Biochemistry*, 42:9797–9803, 2003.

[67] Niels Volkmann, Greta Ouyang, Kathleen Trybus, David J. DeRosier, Susan Lowey, and Dorit Hanein. Myosin isoforms show unique conformations in the actin-bound state. *Proc. Natl. Acad. Sci.*, 100:3227–3232, 2003.

[68] Niels Volkmann, Dorit Hanein, Greta Ouyang, Kathleen M. Trybus, David J. DeRosier, and Susan Lowey. Evidence for cleft closure in actomyosin upon ADP release. *Nature Struct. Biol.*, 7:1147–1155, 2000.

[69] Malcolm Irving, Gabriella Piazzesi, Leonardo Lucii, Yin-Biao Sun, Jeffrey J. Harford, Ian M. Dobbie, Michael A. Ferenczi, Massimo Reconditi, and Vincenzo Lombardi. Conformation of the myosin motor during force generation in skeletal muscle. *Nature Struct. Biol.*, 7:482–485, 2000.

[70] Andrew M. Gulick, Cary B. Bauer, James B. Thoden, Ed Pate, Ralph G. Yount, and Ivan Rayment. X-ray structures of the *Dictyostelium discoideum* myosin motor domain with six non-nucleotide analogs. *J. Biol. Chem.*, 275:398–408, 2000.

[71] Cary B. Bauer, Hazel M. Holden, James B. Thoden, Robert Smith, and Ivan Rayment. X-ray structures of the apo and MgATP-bound states of *Dictyostelium discoideum* myosin motor domain. *J. Biol. Chem.*, 275:38494–38499, 2000.

[72] Werner Kliche, Setsuko Fujita-Becker, Martin Kollmar, Dietmar J. Manstein, and F. Jon Kull. Structure of a genetically engineered molecular motor. *EMBO J.*, 20:40–46, 2001.

[73] Cary B. Bauer, Philip A. Kuhlman, Clive R. Bagshaw, and Ivan Rayment. X-ray crystal structure and solution fluorescence characterization of Mg.2'(3')-O-(N-methylanthraniloyl) nucleotides bound to the *Dictyostelium discoideum* myosin motor domain. *J. Mol. Biol.*, 274:394–407, 1997.

[74] Andrew M. Gulick, Cary B. Bauer, James B. Thoden, and Ivan Rayment. X-ray structures of the MgADP, MgATPγS, and MgAMPPNP complexes of the *Dictyostelium discoideum* myosin motor domain. *Biochemistry*, 36:11619–11618, 1997.

[75] Clyde A. Smith and Ivan Rayment. X-ray structure of the magnesium(II)-pyrophosphate complex of the truncated head of *Dictyostelium discoideum* myosin to 2.7 å resolution. *Biochemistry*, 34:8973–8981, 1995.

[76] Thomas Reubold, Susanne Eschenburg, Andreas Becker, F. Jon Kull, and Dietmar J. Manstein. A structural model for actin-induced nucleotide release in myosin. *Nature Struct. Biol.*, 10:826–830, 2003.

[77] Clyde A. Smith and Ivan 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.

[78] Frank Garland and Herbert C. Cheung. Fluorescence stopped-flow study of the mechanism of nucleotide binding to myosin subfragment 1. *Biochemistry*, 18:5281–5289, 1979.

[79] Kathleen M. Trybus and Edwin W. Taylor. Kinetics of ADP and AMP-PNP binding to SF-1. *Biophys. J.*, 25:21a, 1979.

[80] John W. Shriver and Brian D. Sykes. Phosphorus-31 nuclear magnetic resonance evidence for two conformations of myosin subfragment-1.nucleotide complexes. *Biochemistry*, 20:2004–2012, 1981.

[81] András Málnási-Csizmadia, Robert J. Woolley, and Clive R. Bagshaw. Resolution of conformational states of *Dictyostelium* myosin II motor domain using tryptophan (W501) mutants: implications for the open-closed transition identified by crystallography. *Biochemistry*, 39:16135–16146, 2000.

[82] Massimo Reconditi, Marco Linari, Leonardo Lucii, Alex Stewart, Yin-Biao Sun, Peter Boesecke, Theyencheri Narayanan, Robert F. Sischetti, Tom Irving, Gabriella Piazzesi, Malcolm irving, and Vincenzo Lombardi. The myosin motor in muscle generates a smaller and slower working stroke at higher load. *Nature*, 428:578–581, 2004.

[83] A. Shephard and J. Borejdo. Correlation between mechanical and enzymatic events in contracting skeletal muscle fiber. *Biochemistry*, 43:2804–2811, 2004.

[84] Andras Muhlrad, Y. Michael Peyser, Mahta Nili, Katalin Ajtai, Emil Reisler, and Thomas P. Burghardt. Chemical decoupling of ATPase activation anf force production from the contractile cycle in myosin by steric hindrance of lever-arm movement. *Biophys. J.*, 84:1047–1056, 2003.

[85] Mihály Kovács, Judit Tóth, András Málnási-Csizmadia, Clibe Bagshaw, and László Nyitray. Engineering lysine reactivity as a conformational sensor in *Dictyostelium* myosin II motor domain. *J. Muscle Res. Cell Mot.*, 25:95–102, 2004.

[86] Tomoki P. Terada, Masaki Sasai, and Tetsuya Yomo. Conformational change of the actomyosin complex drives the multiple stepping movement. *Proc. Natl. Acad. Sci.*, 99:9202–9206, 2002.

[87] Michael A. Geeves. Stretching the lever-arm theory. *Nature*, 415:129–130, 2002.

[88] Dipesh Risal, S. Gourinath, Daniel M. Himmel, Andrew G. Szent-Györgyi, and Carolyn Cohen. Myosin subfragment 1 structures reveal a partially bound nucleotide and a complex salt bridge that helps couple nucleotide and actin binding. *Proc. Natl. Acad. Sci.*, 101:8930–8935, 2004.

[89] Lisa K. Nitao, Rachel R. Ogorzalek Loo, Elizabeth O'Neill-Hennessey, Joseph A. Loo, Andrew G. Szent-Györgyi, and Emil Reisler. Conformation and dynamics of the SH1-SH2 helix in scallop myosin. *Biochemistry*, 42:7663–7674, 2003.

[90] Renu Batra, Michael A. Geeves, and Dietmar J. Manstein. Kinetic analysis of *Dictyostelium discoideum* myosin motor domains with glycine-to-alanine mutations in the reactive thiol region. *Biochemistry*, 38:6126–1634, 1999.

[91] Thomas P. Burghardt, Susanna P. Garamszegi, Sungjo Prak, and Katalin Ajtai. Tertiary structural changes in the cleft containing the ATP sensitive tryptophan and reactive thiol are consistent with pivoting of the myosin heavy chain at Gly699. *Biochemistry*, 37:8035–8047, 1998.

[92] Stefan Fischer, Björn Windshügel, Daniel Horak, Kenneth C. Holmes, and Jeremy C. Smith. Structural mechanism of the recovery stroke in the Myosin molecular motor. *Proc. Natl. Acad. Sci.*, 102:6873–6878, 2005.

[93] Paul B. Conibear, Clive Bagshaw, Piotr G. Fajer, Mihály Kovács, and András Málnási-Csizmadia. Myosin cleft movement and its coupling to actomyosin dissociation. *Nature Struct. Biol.*, 10:831–835, 2003.

[94] J. David Lawson, Edward Pate, Ivan Rayment, and Ralph G. Yount. Molecular dynamics analysis of structural factors influencing back door P_i release in myosin. *Biophys. J.*, 86:3794–3803, 2004.

[95] Kenneth C. Holmes, Isabel Angert, F. Jon Kull, Werner Jahn, and Rasmus R. Schröder. Electron cryo-microscopy shows how strong binding of myosin to actin releases nucleotide. *Nature*, 425:423–427, 2003.

[96] Pierre-Damien Coureux, Amber L. Wells, Julie Ménétrey, Christopher M. Yengo, Carl A. Morris, H. Lee Sweeney, and Anne Houdusse. A structural state of the myosin V motor without bound nucleotide. *Nature*, 425:419–423, 2003.

[97] Adam Kuspa, Richard Sucgang, and Gad Shaulsky. The promise of a protist: the *Dictyostelium* genome project. *Funct. Integr. Genomics*, 1:279–293, 2001.

[98] Alan R. Kimmel and Carole A. Parent. The signal to move: *D. discoideum* go orienteering. *Science*, 300:1525–1527, 2003.

[99] http://www.dictybase.org. dictyBase: an online resource for *Dictyostelium*.

[100] Shi Shu, Xiong Liu, and Edward D. Korn. *Dictyostelium* and *Acanthamoeba* myosin II assembly domains go to the cleavage furrow of *Dictyostelium* myosin II-null cells. *Proc. Natl. Acad. Sci.*, 100:6499–6504, 2003.

[101] Markus Fischer, Ilka Haase, Evelyn Simmeth, Günther Gerisch, and Annette Müller-Taubenberger. A brilliant monomeric red fluorescent protein to visualize cytoskeleton dynamics in *Dictyostelium*. *FEBS Lett.*, 577:227–232, 2004.

[102] Eric Shelden and David A. Knecht. Mutants lacking myosin II cannot resist forces generated during multicellular morphogenesis. *J. Cell Science*, 108:1105–1115, 1995.

[103] Eva C. Schwarz, Heidrun Geissler, and Thierry Soldati. A potentially exhaustive screening strategy reveals two novel divergent myosins in *Dictyostelium*. *Cell Biochem. Biophys.*, 30:413–435, 1999.

[104] Robert A. Alberty. Standard Gibbs free energy, enthalpy, and entropy changes as a function of pH and pMg for several reactions involving adenosine phosphates. *J. Biol. Chem.*, 244:3290–3302, 1968.

[105] Robert A. Alberty. Degrees of freedom in biochemical reaction systems at specified pH and pMg. *J. Phys. Chem.*, 96:9614–9621, 1992.

[106] Robert A. Alberty. Calculation of transformed thermodynamic properties of biochemical reactants at specified pH and pMg. *Biophys. Chem.*, 43:239–254, 1992.

[107] Robert A. Alberty. Equilibrium calculations on systems of biochemical reactions at specified pH and pMg. *Biophys. Chem.*, 42:117–131, 1992.

[108] Robert A. Alberty. Inverse Legendre transform in biochemical thermodynamics: illustrated with the last five reactions of glycolysis. *J. Phys. Chem. B*, 106:6594–6599, 2002.

[109] Robert A. Alberty. Thermodynamics of the hydrolysis of adenosine triphosphate as a function of temperature, pH, pMg, and ionic strength. *J. Phys. Chem. B*, 107:12324–12330, 2003.

[110] Robert A. Alberty. Effect of temperature on the standard transformed thermodynamic properties of biochemical reactions with emphasis on the Maxwell equations. *J. Phys. Chem. B*, 107:3631–3635, 2003.

[111] Robert A. Alberty. Change in the binding of hydrogen ions and magnesium ions in the hydrolysis of ATP. *Biophys. Chem.*, 70:109–119, 1998.

[112] Robert N. Goldberg and Yadu B. Tewari. Thermodynamics of the disproportionation of adenosine 5'-disphosphate to adenosine 5'-triphosphate and adenosine 5'-monophosphate. I. equilibrium model. *Biophys. Chem.*, 40:241–261, 1991.

[113] Yadu B. Tewari and Robert N. Goldberg. Thermodynamics of the disproportionation of adenosine 5'-disphosphate to adenosine 5'-triphosphate and adenosine 5'-monophosphate. II. experimental data. *Biophys. Chem.*, 40:263–276, 1991.

[114] Martin J. Kushmerick. Multiple equilibria of cations with metabolites in muscle bioenergetics. *Am. J. Physiol. (Cell Physiol.)*, 41:C1739–C1747, 1997.

[115] J. W. Larson, Yadu B. Tewari, and Robert N. Goldberg. Thermochemistry of the reactions between adenosine, adenosine 5'-monophosphate, inosine, and inosine 5'-monophosphate; the conversion of L-histidine to (urocanic acid+ammonia). *J. Chem. Thermodynamics*, 25:73–90, 1993.

[116] H. B. Clarke, D. C. Cusworth, and S. P. Datta. The dissociation constantsof the magnesium salts of phosphoric acid, gucose 1-phosphoric acid and glycerol 2-phosphoric acid. *Biochem. J.*, 56:xiv–xv, 1954.

[117] H. B. Clarke, D. C. Cusworth, and S. P. Datta. Thermodynamic quantities for the dissociation equilibria of biologically important compounds. *Biochem. J.*, 58:146–154, 1954.

[118] Juliana Boerio-Goates, Michael R. Francis, Robert N. Goldberg, Manuel A. V. Ribeiro da Silva, Maria D. M. C. Ribeiro da Siva, and Yadu B. Tewari. Thermochemistry of adenosine. *J. Chem. Thermodynamics*, 33:929–947, 2001.

[119] Kevin Range, Matthew J. Mc Grath, Xabier Lopez, and Darrin M. York. The structure and stability of biological metaphosphate, phosphate, and phosphorane compounds in the gas phase and in solution. *J. Am. Chem. Soc.*, 126:1654–1665, 2004.

[120] Ho-Tae Kim. Ab initio calculations for exploring hydration patterns of adenine. *J. Mol. Struct. Theochem*, 673:121–126, 2004.

[121] Louis H. Schliselfeld, C. Tyler Burt, and Richard J. Labotka. ^{31}p nuclear magnetic resonance of phosphonic acid analogues of adenosine nucleotides as functions of ph and magnesium ion concentration. *Biochemistry*, 21:317–320, 1982.

[122] Mildred Cohn and Thomas R. Hughes, Jr. Phosphorus magnetic resonance spectra of adenosine di- and triphosphate. I. effect of pH. *J. Biol. Chem.*, 235:3250–3253, 1960.

[123] Mildred Cohn and Thomas R. Hughes, Jr. Nuclear magnetic resonance spectra of adenosine di- and triphosphate. II. effect of complexing with divalent metal ions. *J. Biol. Chem.*, 237:176–181, 1962.

[124] Raj Gupta, Pratima Gupta, Wasley D. Yushok, and Zelda B. Rose. Measurement of the dissociation constant of MgATP at physiological nucleotide levels by a combination of ^{31}P NMR and optical absorbance spectroscopy. *Biochem. Biophys. Res. Comm.*, 117:210–216, 1983.

[125] Weiqi Zhang, Anita C. Truttmannand Daniel Lüthi, and John A. S. McGuigan. Apparent Mg^{2+}-adenosine 5-triphosphate dissociation constant measured with Mg^{2+} macroelectrodes under conditions pertinent to ^{31}P NMR ionized magnesium determination. *Anal. Biochem.*, 251:246–250, 1997.

[126] Daniel Lüthi, Dorothee Günzel, and John A. S. McGuigan. Mg-ATP binding: its modification by spermine, the relevance to cytosolic Mg^{2+} buffering, changes in the intracellular ionized Mg^{2+} concentration and the estimation of Mg^{2+} by ^{31}P-NMR. *Exp. Physiol.*, 84:231–252, 1999.

[127] Hans Åström, Ethel Limém, and Roger Strömberg. Acidity of secondary hydroxyls in ATP and adenosin analogues and the question of a 2',3'-hydrogen bond in ribonucleosides. *J. Am. Chem. Soc.*, 126:15710–14711, 2004.

[128] Nohad Gresh and Gen-Bin Shi. Conformation-dependent intermolecular interaction energies of the triphosphate anion with divalent metal cations. application to the ATP-binding site of a binuclear bacterial enzyme. a parallel quantum chemical and polarizable molecular mechanics investigation. *J. Comput. Chem.*, 25:160–168, 2004.

[129] Peiming Wang, Reef M. Izatt, John L. Oscarson, and Sue E. Gillespie. ^{1}H NMR study of protonation and Mg(II) coordination of AMP, ADP, and ATP at 25, 50, and 70°C. *J. Phys. Chem.*, 100:9556–9560, J. Phys. Chem.

[130] Robert A. Alberty and Robert N. Goldberg. Standard thermodynamic formation properties for the adenosine 5'-triphosphate series. *Biochemistry*, 31:10610–10615, 1992.

[131] Robert A. Alberty. *Thermodynamics of biochemical reactions*. John Wiley and Sons, Hoboken, New Jersey, 2003.

[132] David L. Miller and F. H. Westheimer. The hydrolysis of γ-phenylpropyl di- and triphosphates. *J. Am. Chem. Soc.*, 88:1507–1511, 1966.

[133] Albert Mildvan. Mechanisms of signaling and related enzymes. *Proteins Str. Func. Gen.*, 29:401–416, 1997.

[134] Linus Pauling. *The Nature of the Chemical Bond*. Cornell University Press, Ithaca, NY, 3 edition, 1960.

[135] Suzanne J. Admiraal and Daniel Herschlag. Mapping the transition state for ATP hydrolysis: implications for enzymatic catalysis. *Chem. Biol.*, 2:729–739, 1995.

[136] Jan Florián, Johan Åqvist, and Arieh Warshel. On the reactivity of phosphate monoester dianions in aqueous solution: Bronsted linear free-energy relationships do not have an unique mechanistic interpretation. *J. Am. Chem. Soc.*, 120:11524–11525, 1998.

[137] W. W. Cleland and Alvan C. Hengge. Mechanisms of phosphoryl and acyl transfer. *FASEB J.*, 9:1585–1594, 1995.

[138] Philip George, Robert J. Witonsky, Mendel Trachtman, Clara Wu, William Dortwart, Linda Richman, William Richman, Fahd Shurayh, and Barry Lentz. Squiggle-H_2O: an enquiry into the importance of solvation effects in phosphate ester and anhydride reactions. *Biochim. Biophys. Acta*, 223:1–15, 1970.

[139] Xabier Lopez, Darrin M. York, Annick DeJaegere, and Martin Karplus. Theoretical studies on the hydrolysis of phosphate diesters in the gas phase, solution and RNase A. *Int. J. Quantum Chem.*, 86:10–16, 2002.

[140] Paul C. Haake and F. H. Westheimer. Hydrolysis and exchange in esters of phosphoric acid. *J. Am. Chem. Soc.*, 83:1102–1109, 1961.

[141] A. Dejaegere and M. Karplus. Hydrolysis rate difference between cyclic an dacyclic phosphate esters: Solvation versus strain. *J. Am. Chem. Soc.*, 115:5316–5317, 1993.

[142] Carmay Lim and Martin Karplus. Nonexistence of dianionic pentacovalent intermediates in an ab initio study of the base-catalyzed hydrolysis of ethylene phosphate. *J. Am. Chem. Soc.*, pages 6872–5873, 1990.

[143] Annick Dejaegere, Xiaoling Liang, and Martin Karplus. Phosphate ester hydrolysis: calculation of gas-phase reaction paths and solvation effects. *J. Chem. Soc. Faraday Trans.*, 90:1763–1770, 1994.

[144] Xi Chen and Chang-Guo Zhan. Theoretical determination of activation free energies for alkaline hydrolysis of cyclic and acyclic phosphodiesters in aqueous solution. *J. Phys. Chem. A*, 108:6407–6413, 2004.

[145] Jose M. Mercero, Paul Barrett, Cheuk W. Lam, Joseph E. Fowler, Jesus M. Ugalde, and Lee G. Pedersen. Quantum mechanical calculations on phosphate hydrolysis reactions. *J. Comput. Chem.*, 21:43–51, 2000.

[146] Anthony J. Kirby, Marcelo F. Lima, Davi da Silva, and Faruk Nome. Nucleophilic attack by oxyanions on a phosphate monoester dianion: The positive effect of a cationic general acid. *J. Am. Chem. Soc.*, pages 1350–1351, 2004.

[147] Frank H. Westheimer. Pseudo-rotation in the hydrolysis of phosphate esters. *Acc. Chem. Res.*, 1:70–78, 1968.

[148] Piotr Grzyska, Przemyslaw G. Czyryca, Jamie Purcell, and Alvan C. Hengge. Transition state differences in the hydrolysis reactions of alkyl versus aryl phosphate monoester monoanions. *J. Am. Chem. Soc.*, 125:13106–13111, 2003.

[149] Chetan Lad, Nicholas H. Williams, and Richard Wolfenden. The rate of hydrolysis of phosphomonoester dianions and the exceptional catalytic proficiencies of protein and inositol phosphatases. *Proc. Natl. Acad. Sci.*, 100:5607–5610, 2003.

[150] C. A. Bunton, D. R. Llewellyn, K. G. Oldham, and C. A. Vernon. The reactions of organic phosphates. Part I. the hydrolysis of methyl dihydrogen phosphate. *J. Chem. Soc.*, page 3574, 1958.

[151] Richard Wolfenden, Caroline Ridgway, and Gregory Young. Spontaneous hydrolysis of ionized phosphate monoesters and diesters and the proficiencies of phosphatases and phosphodiesterases as catalysts. *J. Am. Chem. Soc.*, 120:833–834, 1998.

[152] Ching-Han Hu and Tore Brinck. Theoretical studies of the hydrolysis of the methyl phosphate anion. *J. Phys. Chem. A*, 103:5379–5386, 1999.

[153] Yan-Ni Wang, Igor A. Topol, Jack R. Collins, and Stanley Burt. Theoretical studies on the hydrolysis of mono-phosphate and tri-phosphate in gas phase and aqueous solution. *J. Am. Chem. Soc.*, 125:13265–13273, 2003.

[154] Matthew W. Palascak and George C. Shields. Accurate experimental values for the free energies of hydration of H^+, OH^-, and H_3O^+. *J. Phys. Chem. A*, 108:3692–3694, 2004.

[155] Jan Florián and Arieh Warshel. A fundamental assumption about OH^- attack in phosphate ester hydrolysis is not fully justified. *J. Am. Chem. Soc.*, 119:5473–5474, 1997.

[156] Junji Kumamoto, James R. Cox, and Frank H. Westheimer. Barium ethylene phosphate. *J. Am. Chem. Soc.*, 78:4858–4860, 1956.

[157] C. A. Bunton, M. M. Mhala, K. G. Oldham, and C. A. Vernon. The reactions of organic phosphates. part III. the hydrolysis of dimethyl phosphate. *J. Chem. Soc.*, pages 3293–3301, 1960.

[158] Jan Florián, Marek Strajbl, and Arieh Warshel. Conformational flexibility of phosphate, phosphonate, and phosphorothioate methyl esters in aqueous solution. *J. Am. Chem. Soc.*, 120:7959–7966, 1998.

[159] A. Dejaegere, C. Lim, and M. Karplus. Dianionic pentacoordinate species in the base-catalyzed hydrolysis of ethylene and dimethyl phosphate. *J. Am. Chem. Soc.*, 113:4353–4355, 1991.

[160] Petra Imhof, Stefan Fischer, Roland Krämer, and Jeremy C. Smith. Density functional theory analysis of dimethylphosphate hydrolysis: effect of solvation and nucleophile variation. *J. Mol. Struct. Theochem*, 713:1–5, 2004.

[161] J. Peter Guthrie. Hydration and dehydration of phosphoric acid derivatives: free energies of formation of the pentacoordinate intermediates for phosphate ester hydrolysis and of monomeric metaphosphate. *J. Am. Chem. Soc.*, 99:3991–4001, 1977.

[162] Buyong Ma, Cynthia Meredith, and Henry F. Schaefer III. Pyrophosphate structures and reactions: evaluation of electrostatic effects on the pyrophosphates with and without alkali cations. *J. Phys. Chem.*, 98:8216–8223, 1994.

[163] Buyong Ma, Cynthia Meredith, and Henry F. Schaefer III. The quest for a metaphosphate intermediate. the mechanisms for hydrolysis of pyrophosphates with and without catalysis. *J. Phys. Chem.*, 99:3815–3822, 1995.

[164] Humberto Saint-Martin and Luis E. Vicent. Ab initio study of the hydrolysis reactions of neutral and anionic Mg-pyrophosphate complexes in the gas phase. *J. Phys. Chem. A*, 103:6862–6872, 1999.

[165] Humberto Saint-Martin, Luis E. Ruiz-Vicent, Alejandro Ramírez-Solís, and Iván Ortega-Blake. Toward an understanding of the hydrolysis of Mg-PP_i. an *ab initio* study of the isomerization reactions of neutral and anionic Mg-pyrophosphate complexes. *J. Am. Chem. Soc.*, 118:12167–12173, 1996.

[166] James I. Watters, E. Dan Loughran, and Sheldon M. Lambert. The acidity of triphosphoric acid. *J. Am. Chem. Soc.*, 78:4855–4858, 1956.

[167] Xue-Bin Wang, Erich R. Vorpagel, Xin Yang, and Lai-Sheng Wang. Experimental and theoretical investigations of the stability, energetics, and structures of $H_2PO_4^-$, $H_2P_2O_7^{2-}$, and $H_3P_3O_{10}^{2-}$ in the gas phase. *J. Phys. Chem. A*, 105:10468–10474, 2001.

[168] J. Akola and R. O. Jones. ATP hydrolysis in water - a density functional study. *J. Phys. Chem. B*, 107:11774–11783, 2003.

[169] Jan Florián and Arieh Warshel. Phosphate ester hydrolysis in aqueous solution: associative versus dissociative mechanisms. *J. Phys. Chem. B*, 102:719–734, 1998.

[170] Johan Åqvist, Karin Kolmodin, Jan Florián, and Arieh Warshel. Mechanistic alternatives in phosphate monoester hydrolysis: what conclusions can be drawn from available experimental data? *Chem. Biol.*, 6:R71–R80, 1999.

[171] Piotr K. Grzyska, Youngjoo Kim, Michael D. Jackson, Alvan C. Hengge, and John M. Denu. Probing the transition-state structure of dual-specificity protein phosphatases using a physiological substrate mimic. *Biochemistry*, 43:8807–8814, 2004.

[172] Michael C. Hutter and Volkhard Helms. Phosphoryl transfer by a concerted mechanism in UMP/CMP-kinase. *Protein Science*, 9:2225–2231, 2000.

[173] Natalia Díaz and Martin J. Field. Insights into the phosphoryl-transfer mechanism of cAMP-dependent protein kinase from quantum chemical calculations and molecular dynamics simulations. *J. Am. Chem. Soc.*, 126:529–542, 2004.

[174] Marat Valiev, R. Kawai, Joseph A. Adams, and John H. Weare. The role of the putative catalytic base in the phosphoryl transfer reaction in a protein kinase: first-principles calculations. *J. Am. Chem. Soc.*, 125:9926–9927, 2003.

[175] Richard Wolfenden. Analog approaches to the structure of the transition state in enzyme reactions. *Acc. Chem. Res.*, 5:10–18, 1972.

[176] Mariano Barbacid. *ras* genes. *Ann. Rev. Biochem.*, 56:779–827, 1987.

[177] Anne B. Vojtek and Channing J. Der. Increasing complexity of the ras signaling pathway. *J. Biol. Chem.*, 273:19925–19928, 1998.

[178] D. R. Lowy and B. M. Willumsen. Function and regulation of ras. *Ann. Rev. Biochem.*, 62:851–891, 1993.

[179] Gretchen L. Temeles, Jackson B. Gibbs, Jill S. D'Alonzo, Irving S. Sigal, and Edward M. Scolnick. Yeast and mammalian ras proteins have conserved biochemical properties. *Nature*, 313:700–703, 1985.

[180] Timothy M. Glennon, Jordi Villà, and Arieh Warshel. How does GAP catalyze the GTPase reaction of Ras?: a computer simulation study. *Biochemistry*, 39:9641–9651, 2000.

[181] P Gideon, J John, M Frech, A Lautwein, R Clark, JE Scheffler, and Alfred Wittinghofer. Mutational and kinetic analyses of the GTPase-activating protein (GAP)-p21 interaction: the C-terminal domain of GAP is not sufficient for full activity. *Mol. Cell. Biol.*, 12:2050–2056, 1992.

[182] Emil F. Pai, Ute Krengel, Gregory A. Petsko, Roger S. Goody, Wolfgang Kabsch, and Alfred Wittinghofer. Refined crystal structure of the triphosphate conformation of H-*ras* p21 at 1.35 Å resolution: implications for the mechanism of GTP hydrolysis. *EMBO J.*, 9:2351–2359, 1990.

[183] Klaus Scheffzek, Mohammad Reza Ahmadian, Wolfgang Kabsch, Lisa Wiesmüller, Alfred Lautwein, Frank Schmitz, and Alfred Wittinghofer. The Ras-RasGAP complex: Structural basis for GTPase activation and its loss in oncogenic Ras mutants. *Science*, 277:333–338, 1997.

[184] Karen A. Maegley, Suzanne J. Admiraal, and Daniel Herschlag. Ras-catalyzed hydrolysis of GTP: A new perspective from model studies. *Proc. Natl. Acad. Sci.*, 93:8160–8166, 1996.

[185] Florian Hollfelder and Daniel Herschlag. The nature of the transition state for enzyme-catalyzed phosphoryl transfer. hydrolysis of o-aryl phosphorothioates by alkaline phosphatase. *Biochemistry*, 34:12255–12264, 1995.

[186] Ivana Nikolic-Hughes, Douglas C. Rees, and Daniel Herschlag. Do electrostatic interactions with positively charged active site groups tighten the transition state for enzymatic phosphoryl transfer? *J. Am. Chem. Soc.*, 126:11814–11819, 2004.

[187] V. Cepus, A. J. Scheidig, Roger S. Goody, and Klaus Gerwert. Time-resolved FTIR studies of the GTPase reaction of H-Ras P21 reveal a key role for the β-phosphate. *Biochemistry*, 37:10263–10271, 1998.

[188] Christoph Allin, Mohammad Reza Ahmadian, Alfred Wittinghofer, and Klaus Gerwert. Monitoring the GAP catalyzed H-Ras GTPase reaction at atomic resolution in real time. *Proc. Natl. Acad. Sci.*, 98:7754–7759, 2001.

[189] Xinlin Du, Gavin E. Black, Paolo Lecchi, Fred P. Abramson, and Stephen R. Sprang. Kinetic isotope effects in Ras-catalyzed GTP hydrolysis: evidence for a loose transition state. *Proc. Natl. Acad. Sci.*, 101:8858–8863, 2004.

[190] Igor A. Topol, Raul E. Cachau, Alexander V. Nemukhin, Bella L. Grigorenko, and Stanley K. Burt. Quantum chemical modeling of the GTP hydrolysis by the RAS-GAP protein complex. *Biochim. Biophys. Acta*, 1700:125–136, 2004.

[191] R. W. Lymn and Edwin W. Taylor. Mechanism of adenosin triphosphate hydrolysis by actomyosin. *Biochemistry*, 10:4617–4624, 1971.

[192] Fumi Morita. Interaction of heavy meromyosin with substrate. *J. Biol. Chem.*, 242:4501–4506, 1967.

[193] Neil C. Millar and Michael A. Geeves. Protein fluorescence changes associated with ATP and adenosine 5'-[γtriphosphate] binding to skeletal muscle myosin subfragment 1 and actomysoin subfragment 1. *Biochem. J.*, 249:735–743, 1988.

[194] Kenneth A. Johnson and Edwin W. Taylor. Intermediate states of subfragment 1 and actosubfragment 1 ATPase: reevaluation of the mechanism. *Biochemistry*, 17:3432–3442, 1978.

[195] Michael A. Geeves and Tracy E. Jeffries. The effect of nucleotide upon a specific isomerization of actomyosin subfragment 1. *Biochem. J.*, 256:41–46, 1988.

[196] Stuart Wakelin, Paul B. Conibear, Robert J. Woolley, David N. Floyd, Clive R. Bagshaw, Mihály Kovaács, and András Málnási-Csizmadia. Engineering *Dictyostelium discoideum* myosin II for the introduction of site-specific fluorescence probes. *J. Muscle Res. Cell Mot.*, 23:673–683, 2002.

[197] Richard J. Clark, Miklós Nyitrai, Martin R. Webb, and Michael A. Geeves. Probing nucleotide dissociation from myosin *in vitro* using microgram quantities of myosin. *J. Muscle Res. Cell Mot.*, 24:315–321, 2003.

[198] Edwin W. Taylor. Transient phase of adenosine triphosphate hydrolysis by myosin, heavy meromyosin, and subfragment 1. *Biochemistry*, 16:732–740, 1977.

[199] Neil C. Millar and Michael A. Geeves. The limiting rate of the ATP-mediated dissociation of actib from rabbit skeletal muscle myosin subfragment 1. *FEBS Lett.*, 160:141–148, 1983.

[200] Raymond F. Siemankowski and Howard D. White. Kinetics of the interaction between actin, ADP, and cardiac myosin-S1. *J. Biol. Chem.*, 259:5045–5053, 1984.

[201] S. B. Marston and Edwin W. Taylor. Comparison of the myosin and actomyosin ATPase mechanisms the four types of vertebrate muscles. *J. Mol. Biol.*, 139:573–600, 1980.

[202] S. E. Kurzawa-Goertz, Cynthia L. Perreault-Micale, Kathleen M. Trybus, Andrew Szent-Györgyi, and Michael A. Geeves. Loop I can modulate ADP affinity, ATPase activity, and motility of different scallop myosins. transient kinetic analysis of S1 isoforms. *Biochemistry*, 37:7517–7525, 1998.

[203] Renu Batra and Dietmar J. Manstein. Functional characterisation of *Dictyostelium* myosin ii with conserved tryptophanyl residue 501 mutated to tyrosine. *Biol. Chem.*, 380:1017–1023, 1999.

[204] Toshiaki Hiratsuda. Spatial proximity of ATP-sensitive tryptophanyl residue(s) and Cys-697 in myosin ATPase. *J. Biol. Chem.*, 267:14949–14954, 1992.

[205] Sungjo Park, Katalin Ajtai, and Thomas P. Burghardt. Mechanism for coupling free energy in ATPase to the myosin active site. *Biochemistry*, 36:3368–3372, 1997.

[206] András Málnási-Csizmadia, Mihály Kovács, Robert J. Woolley, Stanley W. Botchway, and Clive R. Bagshaw. The dynamics of the relay loop tryptophan residue in the *Dictyostelium* myosin motor domain and the origin of spectroscopic signals. *J. Biol. Chem.*, 276:19483–19490, 2001.

[207] Thomas P. Burghardt, Sungjo Park, Wen-Ji Dong, Jun Xing, Herbert C. Cheung, and Katalin Ajtai. Energy transduction optical sensor in skeletal myosin. *Biochemistry*, 42:5877–5884, 2003.

[208] Clive R. Bagshaw and David R. Trentham. The characterization of myosin-product complexes and of product-release steps during Magnesium ion-dependent adenosine triphosphatase reaction. *Biochem. J.*, 141:331–349, 1974.

[209] András Málnási-Csizmadia, David S. Pearson, Mihály Kovács, Robert J. Woolley, Michael Geeves, and Clive R. Bagshaw. Kinetic resolution of a conformation transition and the ATP hydrolysis step using relaxation methods with a *Dictyostelium* myosin II mutant containing a single tryptophan residue. *Biochemistry*, 40:12727–12737, 2001.

[210] Wei Zeng, Paul B. Conibear, Jane L. Dickens, Ruth A. Cowie, Stuart Wakelin, András Málnási-Csizmadia, and Clive R. Bagshaw. Dynamics of actomyosin interactions in relation to the cross-bridge cycle. *Philos. Trans. R. Soc. B*, 359:1843–1855, 2004.

[211] Margery Beinfeld, David A. Bryce, Daniel Kochavy, and Anthony Martinosi. The binding of divalent cations to myosin. *J. Biol. Chem.*, 250:6282–6287, 1975.

[212] Bernard A. Connolly and Fritz Eckstein. Structures of the mono- and divalent metal nucleotide complexes in the myosin ATPase. *J. Biol. Chem.*, 256:9450–9456, 1981.

[213] Katherine Polosukhina, Don Eden, Marc Chinn, and Stefan Highsmith. CaATP as a substrate to investigate the myosin lever arm hypothesis of force generation. *Biophys. J.*, 78:1474–1481, 2000.

[214] Clive R. Bagshaw. The kinetic mechanism of the manganous ion-dependent adenosine triphosphatase of myosin subfragment 1. *FEBS Lett.*, 58:197–201, 1975.

[215] Clive R. bagshaw and George H. Reed. Investigations of equilibrium complexes of myosin subfragment 1 with the manganous ion and adenosine diphosphate using magnetic resonance techniques. *J. Biol. Chem.*, 251:1975–1983, 1976.

[216] Howard D. White. Special instrumentation and techniques for kinetic studies of contractile systems. *Methods Enzymol.*, 85:698–708, 1982.

[217] Dietmar J. Manstein, Kathleen M. Ruppel, and James A. Spudich. Expression and characterization of a functional myosin head fragment in *Dictyostelium discoideum*. *Science*, 246:656–658, 1989.

[218] Philip A. Kuhlman and Clive R. Bagshaw. ATPase kinetics of the *Dictyostelium discoideum* myosin II motor domain. *J. Muscle Res. Cell Mot.*, 19:491–504, 1998.

[219] S. E. Kurzawa, Dietmar J. Manstein, and Michael A. Geeves. *Dictystelium discoideum* myosin II: characterization of functional myosin motor fragments. *Biochemistry*, 36:317–323, 1997.

[220] Sally K. A. Woodward, Michael Geeves, and Dietmar J. Manstein. Kinetic characterization of the catalytic domain of *Dictyostelium discoideum* myosin. *Biochemistry*, 34:16056–16064, 1995.

[221] Marcie Ritchie, Michael A. Geeves, Sally K. A. Woodward, and Dietmar J. Manstein. Kinetic characterization of a cytoplasmic myosin motor domain expressed in *Dictyostelium discoideum*. *Biochemistry*, 90:8619–8623, 1993.

[222] W. Jahn, Claus Urbanke, and John Wray. Fluorescence temperature jump studies on myosin s1. *J. Anat.*, 194:601, 1999.

[223] Claus Urbanke and John Wray. A fluorescence temperature-jump study of conformational transitions in myosin subfragment 1. *Biochem. J.*, 358:165–173, 2001.

[224] Howard D. White, Betty Belknap, and Martin R. Webb. Kinetics of nucleoside triphosphate cleavage and phosphate release steps by associated rabbit skeletal actomyosin, measured using a novel fluorescent probe for phosphate. *Biochemistry*, 36:11828–11836, 1997.

[225] A. L. Friedman, Michael A. Geeves, Dietmar J. Manstein, and James A. Spudich. Kinetic characterisation of myosin head fragment with long-lived myosin.ATP states. *Biochemistry*, 37:9679–9687, 1998.

[226] Takashi Shimada, Naoya Sasaki, Reiko Ohkura, and Kazuo Sutoh. Alanine scanning mutagenesis of the switch I region in the ATPase site of *Dictyostelium discodeum* myosin II. *Biochemistry*, 36:14037–14043, 1997.

[227] Naoya Sasaki, Takashi Shimada, and Kazuo Sutoh. Mutational analysis of the switch II loop of *Dictyostelium* myosin II. *J. Biol. Chem.*, 273:20334–20340, 1998.

[228] Hirofumi Onishi, Shin-Ichiro Kojijama, Kszuo Sutoh, Keigi Fujiwara, Hugo M. Martinez, and Manuel F. Morales. Functional transitions in myosin: formation of a critical salt-bridge and transmission of effect to the sensitive tryptophan. *Proc. Natl. Acad. Sci.*, 95:6653–6658, 1998.

[229] Taketoshi Kambara, Troy E. Rhodes, Reiko Ikebe, Misato Yamada, Howard D. White, and Mitsuo Ikebe. Functional significance of the conserved residues in the flexible hinge region of the myosin motor domain. *J. Biol. Chem.*, 274:16400–16406, 1999.

[230] Marcus Furch, Setsuko Fujita-Becker, Michael A. Geeves, Kenneth C. Holmes, and Dietmar J. Manstein. Role of the salt-bridge between switch-1 and switch-2 of *Dictyostelium* myosin. *J. Mol. Biol.*, 290:797–809, 1999.

[231] Xiang-Dong Li, Troy E. Rhodes, Reiko Ikebe, Taketoshi Kambara, Howard D. White, and Mitsuo Ikebe. Effects of mutations in the γ-phosphate binding site of myosin on its motor function. *J. Biol. Chem.*, 273:27404–27411, 1998.

[232] M. Ikebe and David J. Hartshorne. Proteolysis of smooth muscle myosin by Staphylococcus aureus protease: preparation of heavy meromyosin and subfragment 1 with intact 20,000-dalton light chains. *Biochemistry*, 24:2380–2387, 1985.

[233] Coleen Murphy, Ronald S. Rock, and James A. Spudich. A myosin II mutation uncouples ATPase activity from motility and shortens step size. *Nature Cell Biol.*, 3:311–315, 2001.

[234] Neil C. Millar, J. Victor Howarth, and Herbert Gutfreund. A transient kinetic study of enthalpy changes during the reaction of myosin subfragment 1 with ATP. *Biochem. J.*, 248:683–690, 1987.

[235] Hans-Jürgen Hinz, Peter Pollwein, Renate Schmidt, and Franz Zimmermann. The enthalpy changes upon hydrolysis of guanosine triphosphate anhydride and ester bonds. *Arch. Biochem. Biophys.*, 212:72–77, 1981.

[236] R. M. Simmons and Terrell L. Hill. Definition of free energy levels in biochemical reactions. *Nature*, 263:615–618, 1976.

[237] Hans Georg Mannherz, Helmut Schenck, and Roger S. Goody. Synthesis of ATP from ADP and inorganic phosphate at the myosin-subfragment 1 active site. *Eur. J. Biochem.*, 48:287–295, 1974.

[238] Roger S. Goody, Waltraud Hofmann, and Hans G. Mannherz. The binding constant of ATP to myosin S1 fragment. *Eur. J. Biochem.*, 78:317–324, 1977.

[239] Martin R. Webb and David R. Trentham. Chemical mechanism of myosin-catalyzed ATP hydrolysis. In L.D.Peachey, R.H.Adrian, and S.R. Geiger, editors, *Handbook of physiology: skeletal muscle*, pages 237–255. American Physiological Society, 1983.

[240] Mary E. Dempsey and Paul D. Boyer. Catalysis of an inorganic phosphate-H_2O^{18} exchange by actomyosin and myosin. *J. Biol. Chem.*, 236:PC6–PC7, 1961.

[241] Mary E. Dempsey, Paul D. Boyer, and E. S. Benson. Characteristics of an orthophosphate oxygen exchange catalysed by myosin, actomyosin, and muscle fibers. *J. Biol. Chem.*, 238:2708–2715, 1963.

[242] J. Robert Swanson and Ralph Yount. The properties of heavy meromyosin and myosin catalyzed medium and intermediate ^{18}O-phosphate exchange. *Biochem. Z.*, 345:395–409, 1966.

[243] L. Sartorelli, H. J. Fromm, R. W. Benson, and Paul D. Boyer. Direct and ^{18}O-exchange measurements relevant to possible activated or phosphorylated states of myosin. *Biochemistry*, 5:2877–2884, 1966.

[244] John H. Young, Jerome McLick, and Ephraim F. Korman. Pseudorotation mechanism of ATP hydrolysis in muscle contraction. *Nature*, 249:474–476, 1974.

[245] Carlos Silva López, Olalla Nieto Faza, Angel R. de Lera, and Darrin M. York. Pseudorotation barriers of biological oxyphosphoranes: a challenge for simulations of ribozyme catalysis. *Chem. Eur. J.*, 11:2081–2093, 2005.

[246] Clive R. Bagshaw, David R. Trentham, R. G. Wolcott, and Paul D. Boyer. Oxygen exchange in the γ-phosphoryl group of protein-bound ATP during Mg^{2+}-dependent adenosine triphosphatase activity of myosin. *Biochemistry*, 72:2592–2596, 1975.

[247] John A. Sleep, David D. Hackney, and Paul D. Boyer. The equivalence of phosphate oxygens for exchange and the hydrolysis characteristics revealed by the distribution of $[^{18}O]P_i$ species formed by myosin and actomyosin ATPase. *J. Biol. Chem.*, 255:4094–4099, 1980.

[248] Martin R. Webb and David R. Trentham. The mechanism of ATP hydrolysis catalysed by myosin and actomyosin, using rapid reaction techniques to study oxygen exchange. *J. Biol. Chem.*, 256:10910–10916, 1981.

[249] Marsha P. Dale and David D. Hackney. Analysis of positional isotope exchange in ATP by cleavage of the βP-OγP bond. demonstration of negligible positional isotope exchange by myosin. *Biochemistry*, 26:8365–8372, 1987.

[250] Martin R. Webb and David R. Trentham. The stereochemical course of phosphoric residue transfer during the myosin ATPase reaction. *Biochem. J.*, 255:8629–8632, 1980.

[251] Hua Deng, Jianghua Wang, Robert H. Callender, Jean C. Grammer, and Ralph G. Yount. Raman difference spectroscopic studies of the myosin S1·MgADP·Vanadate complex. *Biochemistry*, 37:10972–10979, 1998.

[252] Ivan Rayment. The structural basis of the myosin ATPase activity. *J. Biol. Chem.*, 271:15850–15853, 1996.

[253] Hiroshi Kagawa and Kazuhide Mori. Molecular orbital study of the interaction between MgATP and the myosin motor domain: the highest occupied molecular orbitals indicate the reaction site of ATP hydrolysis. *J. Phys. Chem. B*, 103:7346–7352, 1999.

[254] Tshioaki Hiratsuka and Tsuyoshi Katoh. Chemical identification of serine 181 at the ATP-binding site of myosin as a residue esterified selectively by the fluorescent agent 9-anthroylnitrile. *J. Biol. Chem.*, 278:31891–31894, 2003.

[255] Noriaki Okimoto, Kazunori Yamanaka, Junko Ueno, Masayuki Hata, Tyuji Hoshino, and Minoru Tsuda. Theoretical studies on the ATP hydrolysis mechanism of myosin. *Biophys. J.*, 81:2786–2794, 2001.

[256] Todd J. Minehardt, Nicola Marzari, Roger Cooke, Edward Pate, Peter A. Kollmann, and Roberto Car. A classical an ab initio study of the interaction of the myosin triphosphate binding domain with ATP. *Biophys. J.*, 82:660–675, 2002.

[257] Guohui Li and Qiang Cui. Mechanochemical coupling in myosin: a theoretical analysis with molecular dynamics and combined QM/MM reaction path calculations. *J. Phys. Chem. B*, 108:3342–3357, 2004.

[258] Hirofumi Onishi, Naoki Mochizuki, and Manuel F. Morales. On the myosin catalysis of ATP hydrolysis. *Biochemistry*, 43:3757–3763, 2004.

[259] David E. Coleman and Stephen R. Sprang. Reaction dynamics of G-protein catalyzed hydrolysis of GTP as viewed by X-ray crystallographic snapshots of $G_{i\alpha 1}$. *Methods Enzymol.*, 308:70–92, 1999.

[260] Tamar Schlick. *Molecular Modeling and simulation. An interdisciplinary guide.* Springer, New York, 2002.

[261] Attila Szabo and Neil S. Ostlund. *Modern Quantum Chemistry.* Dover, New York, 1996.

[262] Peter W. Atkins and R. S. Friedman. *Molecular quantum mechanics.* Oxford University Press, Oxford, 3rd edition, 1997.

[263] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Phys. Rev. B*, 136:864, 1964.

[264] Frank Jensen. *Introduction to computational chemistry.* John Wiley and Sons Ltd., New York, Weinheim, 1999.

[265] Axel D. Becke. Density-functional thermochemistry. III. the role of exact exchange. *J. Chem. Phys.*, 98:5648–5652, 1993.

[266] Axel D. Becke. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A*, 38:3098–3100, 1988.

[267] Chengteh Lee, Weitao Yang, and Robert G. Parr. Development of the colle-salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B*, 37:785–789, 1988.

[268] Arieh Warshel and Michael Levitt. Theoretical studies of enzymic reactions. *J. Mol. Biol.*, 103:227–249, 1976.

[269] Jiali Gao and Donald G. Truhlar. Quantum mechanical methods for enzyme kinetics. *Annu. Rev. Phys. Chem.*, 53:467–505, 2002.

[270] Martin J. Field, Paul A. Bash, and Martin Karplus. A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations. *J. Comput. Chem.*, 11:700–733, 1990.

[271] Sonja M. Schwarzl, Jeremy Smith, and Stefan Fischer. Criterion for usefulness applied to a link-atom method for combined quantum mechanical/molecular mechanical calculations. *J. Comput.-Aided Mol. Des.*

[272] Stephen J. Benkovic and Sharon Hammes-Schiffer. A perspective on enzyme catalysis. *Science*, 301:1196–1202, 2003.

[273] Mireia Garcia-Viloca, Jiali Gao, Martin Karplus, and Donald G. Truhlar. How enzymes work: Analysis by modern rate theory and computer simulations. *Science*, 303:186–195, 2004.

[274] Jordi Villà and Arieh Warshel. Energetics and dynamics of enzymatic reactions. *J. Phys. Chem. B*, 105:7887–7907, 2001.

[275] Arieh Warshel and Robert M. Weiss. An empirical valence bond approach for comparing reactions in solutions and in enzymes. *J. Am. Chem. Soc.*, 102:6218–6226, 1980.

[276] Fernando Bernardi, Massimo Olivucci, and Michael A. Robb. Simulation of MC-SCF results on covalent organic multi-bond reactions: molecular mechanics with valence bond (MM-VB). *J. Am. Chem. Soc.*, 114:1606–1616, 1992.

[277] Jörg Bentzien, Richard P. Muller, Jan Florián, and Arieh Warshel. Hybrid ab initio quantum mechanics/molecular mechanics calculations of free energy surfaces for enzymatic reactions: The nucleophilic attack in subtilisin. *J. Phys. Chem. B*, 102:2293–2301, 1998.

[278] U. Chandra Singh and Peter A. Kollman. A combined *ab initio* quantum mechanical and molecular mechanical method for carrying out simulations on complex molecular systems: Applications to the $CH_3 + Cl^-$ exchange reaction and gas phase protonation of polyethers. *J. Comput. Chem.*, 7:718–730, 1986.

[279] Kirsten P. Eurenius, David C. Chatfield, Bernard R. Brooks, and Milan Hodoscek. Enzyme mechanisms with hybrid quantum and molecular mechanical potentials. I. theoretical considerations. *Int. J. Quantum Chem.*, 60:1189–1200, 1996.

[280] Vincent Théry, Daniel Rinaldi, Jean-Louis Rivail, Bernard Maigret, and György G. Ferenczy. Quantum mechanical computations on very large molecular systems: The local self-consistent field method. *J. Comput. Chem.*, 15:269–282, 1994.

[281] Gérald Monard, Michel Loos, Vincent Théry, Kristofor Baka, and Jean-Louis Rivail. Hybrid classical quantum force field for modeling very large molecules. *Int. J. Quantum Chem.*, 58:153–159, 1996.

[282] Jiali Gao, Patricia Amara, Cristobal Alhambra, and Martin J. Field. A generalized hybrid orbital (GHO) method for the treatment of boundary atoms in combined QM/MM calculations. *J. Phys. Chem. A*, 102:4714–4721, 1998.

[283] Jingzhi Pu, Jiali Gao, and Donald G. Truhlar. Combining self-consistent-charge density-functional tight-binding (SCC-DFTB) with molecular mechanics by the generalized hybrid orbital (GHO) method. *J. Phys. Chem. A*, 108:5454–5463, 2004.

[284] Jingzhi Pu, Jiali Gao, and Donald G. Truhlar. Generalized hybrid orbital (GHO) method for combining ab initio hartree-fock wave functions with molecular mechanics. *J. Phys. Chem. A*, 108:632–650, 2004.

[285] Martin J. Harrison, Neil A. Burton, and Ian H. Hillier. Catalytic mechanism of the enzyme papain: Predictions with a hybrid quantum mechanical/molecular mechanical potential. *J. Am. Chem. Soc.*, 119:12285–12291, 1997.

[286] V. V. Vasilyev. Tetrahedral intermediate formation in the acylation step of acetylcholinesterases. a combined quantum chemical and molecular mechanical model. *J. Mol. Struct. Theochem*, 304:129–141, 1994.

[287] Debananda Das, Kirsten P. Eurenius, Eric M. Billings, Paul Sherwood, David C. Chatfield, Milan Hodosçek, and Bernard R. Brooks. Optimization of quantum mechanical molecular mechanical partitioning schemes: Gaussian delocalization of molecular mechanical charges and the double link atom method. *J. Chem. Phys.*, 117:10534–10546, 2002.

[288] Nathalie Reuter, Annick Dejaegere, Bernard Maigret, and Martin Karplus. Frontier bonds in QM/MM methods: A comparison of different approaches. *J. Phys. Chem. A*, 104:1720–1735, 2000.

[289] Bernard R. Brooks, Robert E. Bruccoleri, Barry D. Olafson, David J. States, S. Swaminathan, and Martin Karplus. CHARMM: A program for macromolecular energy, minimization and dynamics calculations. *J. Comput. Chem.*, 4:187–217, 1983.

[290] CFS Ltd. http://www.dl.ac.uk/CFS/, 2004.

[291] Brent A. Gregersen, Xabier Lopez, and Darrin M. York. Hybrid QM/MM study of thio effects in transphosphorylation reactions. *J. Am. Chem. Soc.*, 125:7178–7179, 2003.

[292] Markus J. Loferer, Hannes H. Loeffler, and Klaus R. Liedl. A QMMM interface between CHARMM and TURBOMOLE: Implementation and application to systems in bulk phase and biologically active systems. *J. Comput. Chem.*, 24:1240–1249, 2003.

[293] Michael W. Schmidt, Kim K. Baldridge, Jerry A. Boatz, Steven T. Elbert, Mark S. Gordon, Jan H. Jensen, Shiro Koseki, Nikita Matsunaga, Kiet A. Nguyen, Shujun Su, Theresa L. Windus, Michel Dupuis, and John A. Montgomery, Jr. General atomic and molecular electronic structure system. *J. Comput. Chem.*, 14:1347–1363, 1993.

[294] NIST. http://webbook.nist.gov/chemistry/, 2004.

[295] Tanya M. Ramond, Gustavo E. Davico, Rebecca L. Schwartz, and W. Carl Lineberger. Vibronic structure of alkoxy radicals via photoelectron spectroscopy. *J. Chem. Phys.*, 112:1158–1169, 2000.

[296] Michael J. Haas and Alex G. Harrison. The fragmentation of proton-bound cluster ions and the gas-phase acidities of alcohols. *Int. J. Mass Spectrom. Ion Proc.*, 124:115–124, 1993.

[297] John E. Bartmess, Judith A. Scott, and Robert T. McIver. Scale of acidities in the gas phase from methanol to phenol. *J. Am. Chem. Soc.*, 101:6046–6056, 1979.

[298] Vincent F. DeTuri and Kent M. Ervin. Competitive threshold collision-induced dissociation: Gas-phase acidities and bond dissociation energies for a series of alcohols. *J. Phys. Chem. A*, 103:6911–6920, 1999.

[299] E. P. Hunter and S. G. Lias. Evaluated gas phase basicities and proton affinities of molecules: An update. *J. Phys. Chem. Ref. Data*, 27:413–656, 1998.

[300] Robert W. Taft and Ronald D. Topsom. The nature and analysis of substituent effects. *Prog. Phys. Org. Chem.*, 16:1, 1987.

[301] John B. Cumming and Paul Kebarle. Summary of gas phase measurements involving acids ah. entropy changes in proton transfer reactions involving negative ions. bond

dissociation energies D(A-H) and electron affinities EA(A). *Can. J. Chem.*, 56:1, 1978.

[302] M. Fujio, Robert T. McIver, and Robert W. Taft. Effects of the acidities of phenols from specific substituent-solvent interactions. inherent substituent parameters from gas-phase acidities. *J. Am. Chem. Soc.*, 103:4017–4029, 1981.

[303] Mars V. Muftakhov, Yury V. Vasilev, and Victor A. Mazunov. Determination of electron affinity of carbonyl radicals by means of negative ion mass spectrometry. *Rapid Commun. Mass Spectrom.*, 13:1104–1108, 1999.

[304] Péter Hudáky, Tamaás Beke, and András Perczel. Peptide models XXXIV. side-chain conformational energy surfaces associated with all major backbone folds of neutral tautomers of n- and c-protected L-histidine. an ab initio study on ethylimidazole and *n*-formyl-L-histidinamide. *J. Mol. Struct. Theochem*, 583:117–135, 2002.

[305] Péter Hudáky, Ilona Hudáky, and András Perczel. Peptide models XXXV. protonated and deprotonated *n*-formyl-L-histidinamide: an ab initio study on side-chain potential energy surfaces of all major backbone conformers. *J. Mol. Struct. Theochem*, 583:199–213, 2002.

[306] Kaushik Raha and Kenneth M. Merz, Jr. A quantum mechanics-based scoring function: Study of zinc ion-mediated ligand binding. *J. Am. Chem. Soc.*, 126:1020–1021, 2004.

[307] Steven L. Dixon and Kenneth M. Merz, Jr. Semiempirical molecular orbital calculations with linear system size scaling. *J. Chem. Phys.*, 104:6643–6649, 1996.

[308] Marek Sierka, Annika Hogekamp, and Reinhart Ahlrichs. Fast evaluation of the Coulomb potential for electron densities using multipole accelerated resolution of identity approximation. *J. Chem. Phys.*, 118:9136–9148, 2003.

[309] Aaron R. Dinner, Xabier Lopez, and Martin Karplus. A charge-scaling method to treat solvent in QM/MM simulations. *Theor. Chem. Acc.*, 109:118–124, 2003.

[310] Sonja M. Schwarzl, Danzhi Huang, Jeremy C. Smith, and Stefan Fischer. Non-uniform charge scaling (NUCS): a practical approximation of solvent electrostatic screening in proteins. *J. Comput. Chem.*, 26:1359–1371, 2005.

[311] Péter Hudáky and András Perczel. Conformation dependence of pKa: Ab initio and DFT investigation of histidine. *J. Phys. Chem. A*, 108:6195–6205, 2004.

[312] Gert D. Billing and Kurt V. Mikkelsen. *Introduction to molecular dynamics and chemical kinetics*. John Wiley and Sons, New York, 1996.

[313] Gert D. Billing and Kurt V. Mikkelsen. *Advanced molecular dynamics and chemical kinetics*. John Wiley and Sons, New York, 1997.

[314] Makhtar S. Bazaraa, Hanif D. Sherali, and C. M. Shetty. *Nonlinear programming. Theory and algorithms*. John Wiley & Sons, 2nd edition, 1993.

[315] Stefan Fischer and Martin Karplus. Conjugate peak refinement: an algorithm for finding reaction paths and accurate transition states in systems with many degrees of freedom. *Chem. Phys. Lett.*, 194:252–261, 1992.

[316] Frank Noé, Fabian Ille, Jeremy C. Smith, and Stefan Fischer. Automated computation of low-energy pathwys for complex rearrangements in proteins: application to the conformational switch of Ras p21. *Proteins Str. Func. Gen.*

[317] Stefan Fischer, Stephen Michnick, and Martin Karplus. A mechanism for rotamase catalysis by the FK506 binding protein (FKBP). *Biochemistry*, 32:13830–13837, 1993.

[318] A. Nicoleta Bondar, Markus Elstner, Sandor Suhai, Jeremy C. Smith, and Stefan Fischer. Mechanism of primary proton tranfer in bacteriorhodopsin. *Structure*, 12:1281–1288, 2004.

[319] Andreea Gruia, Ana Nicoleta Bondar, Jeremy C. Smith, and Stefan Fischer. Mechanism of a molecular valve in the halorhodopsin chloride pump. *Structure*, 13:617–627, 2005.

[320] Raimund Dutzler, Tilman Schirmer, Martin Karplus, and Stefan Fischer. Translocation mechanism of long sugar chains across the maltoporin membrane channel. *Structure*, 10:1273–1284, 2002.

[321] Walter J. Moore and Dieter O. Hummel. *Physikalische Chemie*. Walter de Gruyter, Berlin, New York, 4th edition, 1986.

[322] Peter G. Bolhuis, David Chandler, Christoph Dellago, and Phillip L. Geissler. Transition path sampling: throwing ropes over rough mountain passes, in the dark. *Annu. Rev. Phys. Chem.*, 53:291–318, 2002.

[323] Peter Hänggi, Peter Talkner, and Michal Borkovec. Reaction-rate theory: fifty years after kramers. *Rev. Mod. Phys.*, 62:251–342, 1990.

[324] Hendrik A. Kramers. Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica*, 7:284–303, 1940.

[325] Sonja M. Schwarzl, Danzhi Huang, Jeremy C. Smith, and Stefan Fischer. How well does charge reparameterisation account for solvent screening in molecular mechanics calculations? the example of myosin. *In Silico Biology*, 3:187–196, 2003.

[326] Celeste Sagui and Thomas A. Darden. Molecular dynamics simulations of biomolecules: long-range electrostatic effects. *Annu. Rev. Biophys. Biomol. Struct.*, 28:155–179, 1999.

[327] Barry Honig and Anthony Nicholls. Classical electrostatics in biology and chemistry. *Science*, 268:1144–1149, 1995.

[328] Malcolm E. Davis, Jeffry D. Madura, Brock A. Luty, and J. Andrew McCammon. Electrostatics and diffusion of molecules in solution: simulations with the university of houston brownian dynamics program. *Comp. Phys. Comm.*, 62:187–197, 1991.

[329] Jeffry D. Madura, James M. Briggs, Rebecca C. Wade, Malcolm E. Davis, Brock A. Luty, Andrew Ilin, Jan Antosiewicz, Michael K. Gilson, Babak Bagheri, L. Ridgway Scott, and J. Andrew McCammon. Electrostatics and diffusion of molecules in solution: simulations with the university of houston brownian dynamics program. *Comp. Phys. Comm.*, 91:57–95, 1995.

[330] Anthony Nicholls and Barry Honig. A rapid finite difference algorithm, utilizing successive over-relaxation to solve the poisson-boltzmann equation. *J. Comput. Chem.*, 12:435–445, 1991.

[331] Wonpil Im, Dmitrii Beglov, and Benoît Roux. Continuum solvation model: computation of electrostatic forces from numerical solutions to the poisson-boltzmann equation. *Comp. Phys. Comm.*, 111:59–75, 1998.

[332] Donald Bashford and David A. Case. Generalised Born models of macromolecular solvation effects. *Annu. Rev. Phys. Chem.*, 51:129–152, 2000.

[333] W. Clark Still, Anna Tempczyk, Ronald C. Hawley, , and Thomas Hendrickson. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.*, 112:6127–6129, 1990.

[334] Michael Schaefer and Martin Karplus. A comprehensive analytical treatment of continuum electrostatics. *J. Phys. Chem.*, 100:1578–1599, 1996.

[335] Michael Schaefer, Christian Bartels, and Martin Karplus. Solution conformations and thermodynamics of structured peptides: molecular dynamics simulation with an omplicit solvation model. *J. Mol. Biol.*, 284:835–848, 1998.

[336] Di Qiu, Peter S. Shenkin, Frank P. Hollinger, and W. Clark Still. The BG/SA continuum model for solvation. a fast analytical method for the calculation of approximate Born radii. *J. Phys. Chem. A*, 101:3005–3014, 1997.

[337] Brian N. Dominy and Charles L. Brooks, III. Development of a Generalised Born model parameterization for proteins and nucleic acids. *J. Phys. Chem. B*, 103:3765–3773, 1999.

[338] Jiang Zhu, Emil Alexov, and Barry Honig. Comparative study of generalized born models: Born radii and peptide folding. *J. Phys. Chem. B*, 109:3008–3022, 2005.

[339] Philipp Werner and Amadeo Caflisch. A sphere-based model for the electrostatics of globular proteins. *J. Am. Chem. Soc.*, 125:4600–4608, 2003.

[340] Michael Feig, Alexey Onufriev, Michael S. Lee, Wonpil Im, David A. Case, and Charles L. Brooks III. Performance comparison of eneralized Born and Poisson methods in the calculation of electrostatic solvation energies for protein structures. *J. Comput. Chem.*, 25:265–284, 2004.

[341] Nicolas Calimet, Michael Schaefer, and Thomas Simonson. Protein molecular dynamics with the Generalized Born/ACE solvent model. *Proteins Str. Func. Gen.*, 45:144–158, 2001.

[342] Marcin Król. Comparison of various implicit solvent models in molecular dynamics simulations of immunoglobulin G light chain dimer. *J. Comput. Chem.*, 24:531–546, 2004.

[343] Sergio A. Hassan, Ernest L. Mehler, Daqun Zhang, and Harel Weinstein. Molcular dynamics simulations of peptides and proteins with a continuum electrostatic model based on screened Coulomb potentials. *Proteins Str. Func. Gen.*, 51:109–125, 2003.

[344] Peter J. Steinbach and Bernard R. Brooks. New spherical-cutoff methods for long-range forces in macromolecular simulation. *J. Comput. Chem.*, 15:667–683, 1994.

[345] Scott H. Northrup, Michael R. Pear, John D. Morgan, J. Andrew McCammon, and Martin Karplus. Molecular dynamics of ferrocytochrome c: Magnitude and anisotropy of atomic displacements. *J. Mol. Biol.*, 153:1087–1109, 1981.

[346] Thomas Simonson, Georgios Archontis, and Martin Karplus. Continuum treatment of long-range interactions in free energy calculations. application to protein-ligand binding. *J. Phys. Chem. B*, 101:8349–8362, 1997.

[347] Alexander D. MacKerell, Jr., Donald Bashford, M. Bellott, Roland L. Dunbrack, Jr., J. D. Evanseck, Martin J. Field, Stefan Fischer, Jiali Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, Krzysztof Kuczera, F. T. K. Lau, C. Mattos, Stephen Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, III, Benoît Roux, M. Schlenkrich, Jeremy C. Smith, Roland Stote, John Straub, M. Watanabe, J. Wirkiewicz-Kuczera, D. Yin, and Martin Karplus. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B*, 102:3586 –3616, 1998.

[348] Richard J. Loncharich and Bernard R. Brooks. The effects of truncating long-range forces on protein dynamics. *Proteins Str. Func. Gen.*, 6:32–45, 1989.

[349] Jeanmarie Guenot and Peter A. Kollman. Conformational and energetic effects of truncating nonbonded interactions in an aqueous protein dynamics simulation. *J. Comput. Chem.*, 14:295–311, 1993.

[350] K. Tasaki, S. McDonald, and J. W. Bradley. Observations concerning the treatment of long-range interactions in molecular dynamics simulations. *J. Comput. Chem.*, 14:278–284, 1993.

[351] S. Parkin, B. Rupp, and H. Hope. Structure of bovine pancreatic trypsin inhibitor at 125 k: Definition of carboxyl-terminal residues gly57 and ala58. *Acta Crystallogr. D Biol. Crystallogr.*, 52:18–29, 1996.

[352] M. C. Vaney, S. Maignan, M. Ries-Kautt, and A. Ducruix. High-resolution structure (1.33 angstrom) of a hew lysozyme tetragonal crystal grown in the apcf apparatus. data and structural comparison with a crystal grown under microgravity from spacehab-01 mission. *Acta Crystallogr. D Biol. Crystallogr.*, 52:505–517, 1996.

[353] T. R. Hynes and R. O. Fox. The crystal structure of staphylococcal nuclease refined at 1.7 Å resolution. *Proteins Str. Func. Gen.*, 10:92–105, 1991.

[354] Mafalda Nina, Dmitri Beglov, and Benot Roux. Atomic radii for continuum electrostatics calculations based on molecular dynamics free energy simulations. *J. Phys. Chem. B*, 101:5239–5248, 1997.

[355] K. A. Sharp, Anthony Nicholls, R. F. Fine, and Barry Honig. Reconciling the magnitude of the microscopic and macroscopic hydrophobic effect. *Science*, 252:106–109, 1991.

[356] I. Tunon, E. Silla E, and J. L. Pascual-Ahuir. Molecular surface area and hydrophobic effect. *Protein Eng.*, 5:715–716, 1992.

[357] Andreea D. Gruia, Stefan Fischer, and Jeremy C. Smith. Molecular dynamics simulation reveals a surface salt bridge forming a kinetic trap in unfolding of truncated Staphylococcal nuclease. *Proteins Str. Func. Gen.*, 50:507–515, 2003.

[358] Michael Schaefer, Christian Bartels, Fabrice Leclerc, and Martin Karplus. Effective atom volumes for implicit solvent models: comparison between Voronoi volumes and minimum fluctuation volumes. *J. Comput. Chem.*, 22:1857–1879, 2001.

[359] E. Neria, Stefan Fischer, and Martin Karplus. Simulation of activation free energies in molecular systems. *J. Chem. Phys.*, 105:1902–1921, 1996.

[360] Sampath Koppole, Jeremy C. Smith, and Stefan Fischer.

[361] Scott J. Weiner, Peter A. Kollman, David A. Case, U. Chandra Singh, Caterina Ghio, Guliano Alagona, Salvatore Profeta, and Paul Weiner. A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.*, 106:765–784, 1984.

[362] Wendy D. Cornell, Piotr Cieplak, Christopher I. Bayly, Ian R. Gould, Kenneth M. Merz, David M. Ferguson, David C. Spellmeyer, Thomas Fox, James W. Caldwell, and Peter A. Kollman. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.*, 117:5179–5197, 1995.

[363] Gottfried Otting, Edvards Liepinsh, and Kurt Wüthrich. Disulfide bond isosmerization in BPTI and BPTI(G36S): an NMR study of correlated mobility in proteins. *Biochemistry*, 32:3571–3582, 1993.

[364] G. Wagner, A. DeMarco, and Kurt Wüthrich. Dynamics of the aromatic amino acid residues in the globular conformation of the basic pancreatic trypsin inhibitor (BPTI). I. ^{1}H NMR studies. *Biophys. Struct. Mechanism*, 2:139–158, 1976.

[365] Amedeo Caflisch, Stefan Fischer, and Martin Karplus. Docking by monte carlo minimization with a solvation correction: Application to an FKBP - substrate complex. *J. Comput. Chem.*, 18:723–743, 1997.

[366] Thomas Simonson, Gerogios Archontis, and Martin Karplus. Free energy simulations com eof age: protein-ligand recognition. *Acc. Chem. Res.*, 35:430–437, 2002.

[367] W. Yang, Y. Q. Gao, Qiang Cui, J. Ma, and Martin Karplus. The missing link between thermodynamics and structure in F_1-ATPase. *Proc. Natl. Acad. Sci.*, 100:874–879, 2003.

[368] Alexey Onufriev, David A. Case, and Donald Bashford. Effective Born radii in the generalized Born approximation: The importance of being perfect. *J. Comput. Chem.*, 23:1297–1304), 2002.

[369] Scott H. Northrup, Michael R. Pear, Chyuan-Yih Lee, J. Andrew McCammon, and Martin Karplus. Dynamical theory of activated processes in globular proteins. *Proc. Natl. Acad. Sci.*, 79:4035–4039, 1982.

[370] Martin Karplus. Aspects of protein reaction dynamics: deviations from simple behavior. *J. Phys. Chem. B*, 104:11–27, 2000.

[371] Jana Sopkova-de Oliveira Santos, Stefan Fischer, Christophe Guilbert, Anita Lewit-Bentley, and Jeremy C. Smith. Pathway for large-scale conformational change in annexin V. *Biochemistry*, 39:14065–14074, 2000.

[372] Yingkai Zhang, Jeremy Kua, and J. Andrew McCammon. Influence of structural fluctuation on enzyme reaction energy barriers in combined quantum mechanical/molecular mechanical studies. *J. Phys. Chem. B*, 107:4459–4463, 2003.

[373] Sergio Martí, Juan Andrés, Vicent Moliner, Estanislao Silla, Iñaki Tuñón, and Juan Bertrán. Preorganization and reorganization as related factors in enzyme catalysis: the chorismate mutase case. *Chem. Eur. J.*, 9:984–991, 2003.

[374] Gregory King, Frederick S. Lee, and Arieh Warshel. Microscopic simulations of macroscopic dielectric constants of solvated proteins. *J. Chem. Phys.*, 95:4366–4377, 1991.

[375] Thomas Simonson and Charles L. Brooks III. Charge screening and the dielectric constant of proteins: insights from molecular dynamics. *J. Am. Chem. Soc.*, 118:8452–8458, 1996.

[376] Thomas Simonson and David Perahia. Internal and interfacial dielectric properties of cytochrome *c* from molecular dynamics in aqueous solution. *Proc. Natl. Acad. Sci.*, 92:1082–1086, 1995.

[377] Paul E. Smith, Roger M. Brunne, Alan E. Mark, and Wilfried F. van Gunsteren. Dielectric properties of trypsin inhibitor and lysozyme calculated from molecular dynamics simulations. *J. Phys. Chem.*, 97:2009–2014, 1993.

[378] Michael W. Schmidt, Kim K. Baldridge, Jerry A. Boatz, Steven T. Elbert, Mark S. Gordon, Jan H. Jensen, Shiro Koseki, Nikita Matsunaga, Kiet A. Nguyen, Shujun Su, Theresa L. Windus, Michel Dupuis, and John A. Montgomery Jr. General atomic and molecular electronic structure system. *J. Comput. Chem.*, 14:1347–1363, 1993.

[379] J. Stephen Binkley, John A. Pople, and Warren J. Hehre. Self-consistent molecular orbital methods. 21. small split-valence basis sets for first-row elements. *J. Am. Chem. Soc.*, 102:939–947, 1980.

[380] Mark S. Gordon, J. Stephen Binkley, John A. Pople, William J. Pietro, and Warren J. Hehre. Self-consistent molecular-orbital methods. 22. small split-valence basis sets for second-row elements. *J. Am. Chem. Soc.*, 104:2797–2803, 1982.

[381] William J. Pietro, Michelle M. Francl, Warren J. Hehre, Douglas J. DeFrees, John A. Pople, and J. Stephen Binkley. Self-consistent molecular orbital methods. 24. supplemented small split-valence basis sets for second-row elements. *J. Am. Chem. Soc.*, 104:5039–5048, 1982.

[382] Warren J. Hehre, R. Ditchfield, and John A. Pople. Self-consistent molecular orbital methods. XII. Further extensions of Gaussian-type basis sets for use in molecular orbital studies of organic molecules. *J. Chem. Phys.*, 56:2257–2261, 1972.

[383] Michelle M. Francl, William J. Pietro, Warren J. Hehre, J. Stephen Binkley, Mark S. Gordon, Douglas J. DeFrees, and John A. Pople. Self-consistent molecular orbital methods. XXIII. A polarization-type basis set for second-row elements. *J. Chem. Phys.*, 77:3654–3665, 1982.

[384] Timothy Clark, Jayaraman Chandrasekhar, Günther W. Spitznagel, and Paul von Ragué Schleyer. Efficient diffuse function-augmented basis sets for anion calculations. III. the 3-21+G basis set for first-row elements, Li-F. *J. Comput. Chem.*, 4:294–301, 1983.

[385] Yan Zhao, Núria González-García, and Donald G. Truhlar. Benchmark database of barrier heights for heavy atom transfer, nucleophilic substitution, association, and unimolecular reactions and its use to test theoretical methods. *J. Phys. Chem. A*, 109:2012–2018, 2005.

[386] François P. Rotzinger. Performance of molecular orbital methods and density functional theory in the computation of geometries of metal aqua ions. *J. Phys. Chem. B*, 109:1510–1527, 2005.

[387] Larry A. Curtiss, Krishnan Raghavachari, Paul C. Redfern, and John A. Pople. Assessment of Gaussian-2 and density functional theories for the computation of enthalpies of formation. *J. Chem. Phys.*, 106:1063–1079, 1997.

[388] Jack Simons. Electron propagator studies of molecular anions. *Int. J. Quantum Chem.*, 16:575–581, 1982.

[389] Piotr Skurski, Maciej Gutowski, and Jack Simons. How to choose a one-electron basis set to reliably describe a dipole-bound anion. *Int. J. Quantum Chem.*, 80:1024–1038, 2000.

[390] K. Raghavachari and J. B. Anderson. Electron correlation effects in molecules. *J. Phys. Chem.*, 100:12960, 1996.

[391] Frauke Gräter, Sonja M. Schwarzl, Annick Dejaegere, Stefan Fischer, and Jeremy C. Smith. Protein/ligand binding free energies calculated with quantum mechanica/molecular mechanics. *J. Phys. Chem. B*, 109:10474–10483, 2005.

[392] Elena Soriano, Sebastián Cerdán, and Paloma Ballesteros. Computational determination of pK_a values. a comparison of different theoretical approaches and a novel procedure. *J. Mol. Struct. Theochem*, 684:121–128, 2004.

[393] J. E. Davies, N. L. Doltsinis, A. J. Kirby, C. D. Roussev, and M. Sprik. Estimating pK_a values for pentaoxyphosphoranes. *J. Am. Chem. Soc.*, 124:6594–6599, 2002.

[394] Paul A. Bash, Martin J. Field, R. C. Davenport, Gregory A. Petsko, D. Ringe, and Martin Karplus. Computer simulation and analysis of the reaction pathway of triosephosphate isomerase. *Biochemistry*, 30:5826–5832, 1991.

[395] R. S. Mulliken. Electronic population analysis on LCAO-MO molecular wave functions. I. *J. Chem. Phys.*, 23:1833–1840, 1955.

[396] Mark E. Tuckerman, Dominik Marx, and Michele Parrinello. The nature and transport mechanism of hydrated hydroxide ions in aqueous solution. *Nature*, 417:925–929, 2002.

[397] W. J. Stevens, M. Krauss, Harold Basch, and P.G. Jasien. Relativistic compact effective potentials and efficient, shared-exponent basis sets the third-, fourth- and fifth-row atoms. *Can. J. Chem.*, 70:612, 1992.

[398] C. Møller and M. S. Plesset. Note on an approximation treatment for many-electron systems. *Phys. Rev*, 46:618, 1934.

[399] Keld L. Bak, Poul Jrgensen, Jeppe Olsen, Trygve Helgaker, and Wim Klopper. Accuracy of atomization energies and reaction enthalpies in standard and extrapolated electronic wave function/basis set calculations. *J. Chem. Phys.*, 112:9229–9242, 2000.

[400] Michael J. S. Dewar, Eve G. Zoebisch, Eamonn F. Healy, and James J. P. Stewart. Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.*, 107:3902–3909, 1985.

[401] Alexander A. Voityuk and Notker Rösch. AM1/d parameters for molybdenum. *J. Phys. Chem. A*, 104:4089–4094, 2000.

[402] Michael J. S. Dewar and Walter Thiel. Ground states of molecules. 38. the MNDO method. approximations and parameters. *J. Am. Chem. Soc.*, 99:4899–4907, 1977.

[403] Michael J. S. Dewar and Walter Thiel. Ground states of molecules. 39. MNDO results for molecules containing hydrogen, carbon, nitrogen, and oxygen. *J. Am. Chem. Soc.*, 99:4907–4917, 1977.

[404] Walter Thiel and Alexander A. Voityuk. Extension of the MNDO formalism to *d* orbitals: integral approximations and preliminary numerical results. *Theor. Chim. Acta*, 81:391–404, 1992.

[405] Walter Thiel and Alexander A. Voityuk. Erratum: Extension of the MNDO formalism to *d* orbitals: integral approximations and preliminary numerical results. *Theor. Chim. Acta*, 93:315, 1996.

[406] Walter Thiel and Alexander A. Voityuk. Extension of MNDO to d orbitals: Parameters and results for the second-row elements and for the zinc group. *J. Phys. Chem.*, 100:616–626, 1996.

[407] Marcus Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, Thomas Frauenheim, Sandor Suhai, and G. Seifert. Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties. *Phys. Rev. B*, 58:7260–7268, 1998.

[408] Marcus Elstner, Thomas Frauenheim, Efthimios. Kaxiras, G. Seifert, and Sandor Suhai. A self-consistent charge density-functional based tight-binding scheme for large biomolecules. *Phys. Stat. Sol.*, 217:357–376, 2000.

[409] Qiang Cui, Marcus Elstner, Efthimios Kaxiras, Thomas Frauenheim, and Martin Karplus. A QM/MM implementation of the self-consistent charge density functional tight binding (SCC-DFTB) method. *J. Phys. Chem. B*, 105:569–585, 2001.

[410] Xabier Lopez and Darrin M. York. Parameterization of semiempirical methods to treat nucleophilic attacks to biological phosphates: AM1/d parameters for phosphorus. *Theor. Chem. Acc.*, 109:149–159, 2003.

[411] Paul Winget, Anselm H. C. Horn, Cenk Selcuki1, Bodo Martin, and Timothy Clark. AM1* parameters for phosphorus, sulfur and chlorine. *J. Mol. Mod.*, 9:408–414, 2003.

[412] Péter Várnai, W. Graham Richards, and Paul D. Lyne. Modelling the catalytic reaction in human aldose reductase. *Proteins Str. Func. Gen.*, 37:218–227, 1999.

[413] Péter Várnai and Arieh Warshel. Computer simulation study of the catalytic mechanism of human aldose reductase. *J. Am. Chem. Soc.*, 122:3849–3860, 2000.

[414] Charles E. Grimshaw, Kurt M. Bohren, Chung-Jeng Lai, and Kenneth H. Gabbay. Human aldose reductase: Rate constants for a mechanism including interconversion of ternary complexes by recombinant wild-type enzyme. *Biochemistry*, 34:14356–14365, 1995.

[415] Markus Dittrich, Shigehiko Hayashi, and Klaus Schulten. On the mechanism of ATP hydrolysis in F_1-ATPase. *Biophys. J.*, 85:2253–2266, 2003.

[416] Markus Dittrich, Shigehiko Hayashi, and Klaus Schulten. ATP hydrolysis in the β_{TP} and β_{DP} catalytic sites of F_1ATPase. *Biophys. J.*, 87:2954–2967, 2004.

[417] Douglas R. Davies and Wim G. J. Hol. The power of vanadate in crystallographic investigations of phosphoryl transfer enzymes. *FEBS Lett.*, 577:315–321, 2004.

[418] Sushmita D. Lahiri, Guofeng Zhang, Debra Dunaway-Mariano, and Karen N. Allen. The pentacovalent phosphorus intermediate of a phosphoryl transfer reaction. *Science*, 299:2067–2071, 2003.

[419] Robert R. Holmes. Phosphoryl transfer enzymes and hypervalent phosphorus chemistry. *Acc. Chem. Res.*, 37:746–753, 2004.

[420] Jenny P. Glusker, Amy K. Katz, and Charles W. Bock. Metal ions in biological systems. *The Rigaku Journal*, 16:8–16, 1999.

[421] Charles W. Bock, Amy Kaufmann, and Jenny Glusker. Coordination of water to magnesium cations. *Inorg. Chem.*, 33:419–427, 1994.

[422] Charles W. Bock, Amy Kaufmann Katz, George D. Markham, and Jenny Glusker. Manganese as a replacement for magnesium and zinc: Functional comparison of divalent cations. *J. Am. Chem. Soc.*, 121:7360–7372, 1999.

[423] Hirofumi Onishi, Takashi Ohki, Naoki Mochizuki, and Manuel F. Morales. Early stages of energy transduction by myosin: Roles of Arg in Switch I, of Glu in Switch II, and of the salt-bridge between them. *Proc. Natl. Acad. Sci.*, 99:15339–15344, 2002.

[424] Oliver Daumke, Michael Weyand, Partha P. Chakrabarti, Ingrid Vetter, and Alfred Wittinghofer. The GTPas-activating protein Rap1GAP uses a catalytic asparagine. *Nature*, 429:197–201, 2004.

[425] Kenneth M. Doll, Bruce R. Bender, and Richard G. Finke. The first experimental test of the hypothesis that enzymes have evolved to enhance hydrogen tunneling. *J. Am. Chem. Soc.*, 125:10877–10884, 2003.

[426] Beáta Flachner, Zoltán Kovári, Andrea Varga, Zoltán Gugolya, Ferenc Vonderviszt, Gábor Náray-Szábo, and Mária Vas. Role of phosphate chain mobility of MgATP in completing the 3-phosphoglycerate kinase catalytic site: binding, kinetic, and crystallographic studies with ATP and MgATP. *Biochemistry*, 43:3436–3449, 2004.

[427] George S. Hammond. A correlation of reaction rates. *J. Am. Chem. Soc.*, 77:334–338, 1955.

[428] Pierre-Damien Coureux, H. Lee Sweeney, and Anne Houdusse. Three myosin V structures delineate essential features of chemo-mechanical transduction. *EMBO J.*, 23:4527–4537, 2004.

[429] Minko Dudev, Jonathan Wang, Todor Dudev, and Carmay Lim. Factors governing the metal coordination number in metal complexes from cambridge structural database analyses. *J. Phys. Chem. B*, 110:1889–1895, 2006.

[430] Ralph G. Yount and D. E. Koshland, Jr. Properties of the O^{18} exchange reaction catalyzed by heavy meromyosin. *J. Biol. Chem.*, 238:1708–1713, 1963.

[431] Martin A. Webb, David E. Ash, Thomas S. Leyh, David R. Trentham, and George H. Reed. Electron paramagnetic resonance studies of Mn(II) complexes with myosin subfragment 1 and oxygen 17-labeled ligands. *J. Biol. Chem.*, 257:3068–3072, 1982.

[432] Gerber Molecular Design. Moloc: a molecular design software suite.

[433] W. Humphrey, A. Dalke, and Klaus Schulten. VMD - Visual Molecular Dynamics. *J. Mol. Graph.*, 14:33–38, 1996.

Danke!

Mein herzlicher Dank gilt allen, die mich in den letzten Jahren fachlich und persönlich unterstützt haben, besonders Stefan Fischer, der die Dissertation betreut hat, Jeremy Smith, Peter Comba, Sidonia Mesentean, Petra Imhof, David Nutt, Thomas Splettstößer, Frank Noé, Sampath Koppole, Bogdan Costescu, Thomas Efferth, Gabriel Wittum, Claudia Felser, Petra Mischnick, Birgitta Krumm, Frauke Zbikowski, Jonny Stadler, David Haase, Angelika Glöckner, Chetan Erbe, Ingrid Haarmann, Eva Stortz, Frank Ströber, meinen Eltern, besonders für die Finanzierung dieses Buches, Frauke Gräter, Alexandra Holten, Chunying Niu, Barbara Schreiber und allen anderen, denen ich begegnen durfte. Danke!