PcrA Helicase, a Molecular Motor Studied from the Electronic to the Functional Level

Markus Dittrich · Jin Yu · Klaus Schulten (✉)

Beckman Institute, University of Illinois at Urbana-Champaign, Urbana,
Illinois, IL 61801, USA
kschulte@ks.uiuc.edu

Abstract Molecular motors are adenosine tri-phosphate (ATP) hydrolysis-driven, cellular proteins responsible for a wide variety of different tasks, such as transport, energy metabolism, and DNA processing. Their operation cycle spans a wide range of length and time scales, from the localized and fast chemical reaction in the catalytic site(s) to the large scale and much slower conformational motions involved in the motors' physiological function. From a computational point of view, this means that currently there exists no single approach capable of capturing the whole spectrum of events during molecular motor function. In the present review, we show for PcrA helicase, a molecular motor involved in the unwinding of double-stranded DNA, how a combination of computational approaches can be used to examine PcrA's function in its entirety as well as in detail. Combined quantum mechanical/molecular mechanical simulations are used to study the catalytic ATP hydrolysis event and its coupling to protein conformational changes. Molecular dynamics simulations then provide a means of studying overall PcrA function on a nanosecond time scale. Finally, to reach physiologically relevant time scales, i.e., milliseconds, stochastic simulations are employed. We show that by combining the three stated approaches one can obtain insight into PcrA helicase function.

Keywords ATP hydrolysis · DNA helicase · Molecular dynamics simulations · Molecular motor · QM/MM · PcrA · Stochastic modeling

Abbreviations
ADP   adenosine di-phosphate
ATP   adenosine tri-phosphate
dsDNA   double-stranded DNA

DNA   deoxyribonucleic acid
MD   molecular dynamics
MM   molecular mechanics
nt   nucleotide
P_i phosphate
QM   quantum mechanics
QM/MM   quantum mechanical/molecular mechanical
RESP   restrained electrostatic potential
RNA   ribonucleic acid
SMD   steered molecular dynamics
ssDNA   single-stranded DNA
vdW   van der Waals

1
General Introduction

Molecular motors are proteins essential for cellular transport, metabolism, and signaling. They are powered by the energy stored in nucleotide tri-phosphates, mostly adenosine tri-phosphate (ATP). At the heart of molecular motor function is the conversion of the chemical energy stored in ATP into a mechanical force which drives the physiological function of the system. The chemo-mechanical coupling underlying the conversion of ATP's energy is still poorly understood. Indeed, even though several molecular motor proteins are known at a structural level, e.g., myosin, kinesin, F_1-ATPase, or helicase, and even though for each of them a vast amount of biochemical data is available, decades of study have not revealed how the chemo-mechanical coupling is being achieved in microscopic detail by any of these systems.

However, the availability of structural information allows computational studies to be conducted and a significant number have been reported over the past decade for such diverse systems as myosin [1–5], kinesin [3, 6], F_1-ATPase [7–12], and helicase [13–17]. For F_1-ATPase, e.g., molecular dynamics (MD) simulations provided insight into the inner workings of the rotation of the central stalk inside the hexameric head-unit [8]; combined quantum mechanical/molecular mechanical (QM/MM) simulations of the ATP hydrolysis reaction in the catalytic sites [10–12] on the other hand revealed a mechanism for efficient ATP catalysis and were able to identify several residues involved in the coupling of the chemical reaction to larger scale conformational changes of the protein. On the basis of these available structures and biochemical data, Oster and coworkers [7] have developed stochastic and kinetic models for F_1-ATPase function.

Molecular motor proteins present a number of formidable challenges to the computational researcher. First, their operation involves time scales covering several orders of magnitude, from the fast chemical reaction step which occurs over the course of a few femtoseconds all the way to the mechani-

cal motions responsible for the systems' physiological function taking place on a microsecond up to a millisecond time scale. Second, the events during motor function span a wide range of spatial scales, from the localized and electronic level changes accompanying the bond-breaking event in the catalytic binding site(s) during ATP catalysis ($\sim$ 1 Å) to the large scale structural changes involved in physiological function (10 Å to 10 nm).

From a methodological point of view this means that there is presently no single computational approach which is able to model the whole spatial and temporal scale of events underlying molecular motor function. At least three levels of computational methodology are necessary to capture the relevant degrees of freedom. A QM or QM/MM method is required to investigate the electronic events underlying the ATP hydrolysis reaction and its interaction with the protein environment. Typically, however, only localized structural changes and/or relatively short time scales can be accessed using QM/MM simulations and MD is needed to study the properties of the full protein in a solvent environment on a multi nanosecond time scale at the expense of neglecting electronic level processes such bond-breaking. In conjunction with methods such as steered MD (SMD) [48, 49] this allows one to study the larger scale motions underlying motor function, e.g., conformational changes coupled to binding of reactant ATP or unbinding of hydrolysis products ADP and P_i as well as protein movements relayed to the motors' binding partners such as actin in case of myosin. However, using even the most powerful computational resources currently available, the time scales accessible to all-atom MD simulations are still orders of magnitude away from the true physiological ones. Hence, in order to reach the latter regime one has to resort to an approach which focuses on a small number of essential degrees of freedom. One might, for example, model the motor function via the center of mass motion of the relevant protein domains on a low-dimensional potential energy surface governed by a Langevin-type stochastic equation.

The challenge from a modeling point of view, then, is to connect these approaches to obtain a coherent picture of the overall process. Ideally, the insight gained on each level is used to design and improve the methodology of the other levels, possibly in an iterative fashion. As an example, the essential degrees of freedom and the potentials governing the stochastic equations modeling the long time behavior of the system could be extracted from appropriately designed MD simulations. Conversely, insight gained from the stochastic modeling might then catalyze an improved setup of the MD simulations.

Here, we summarize progress on how such a comprehensive computational approach can be used to gain insight into the function of one particular molecular motor, the DNA helicase PcrA. This motor protein will be introduced in the next section. We will then lay out the computational approaches used and finally summarize and interpret recent computational findings.

2
PcrA Helicase, a Prototype Molecular Motor

DNA helicases are important participants in DNA metabolism, which involves processes such as replication, transcription, and repair. They catalyze the separation of double-stranded DNA (dsDNA) into its single-stranded DNA (ssDNA) components. Helicases are typically found to exist in a monomeric [18, 19], dimeric [20], or hexameric form [21] and unwind DNA either in a $3' \rightarrow 5'$ or $5' \rightarrow 3'$ direction.

PcrA is a monomeric $3' \rightarrow 5'$ helicase, and due to its relatively small size (80 kDa) constitutes an ideal candidate for computational study. Indeed, PcrA is one of the smallest molecular motors known to date. Several atomic resolution X-ray structures of PcrA have been reported [18, 19]. The present discussion will focus on two PcrA structures from *Bacillus stearothermophillus* that were determined by Velankar et al. [19] in a substrate (with bound ATP analogue AMP-PNP) and product (without ATP/ADP bound) state at 3.3 Å and 2.9 Å resolution, respectively. Figure 1 shows a molecular view of PcrA which consists of four major domains called 1A, 1B, 2A, and 2B. Also visible is a strand of ssDNA threading through the protein at the interface formed by the A and B domains and which is connected to a short piece of dsDNA bound to the 2B domain. ATP binds to the catalytic binding pocket located at the interface between the 1A and 2A domains.
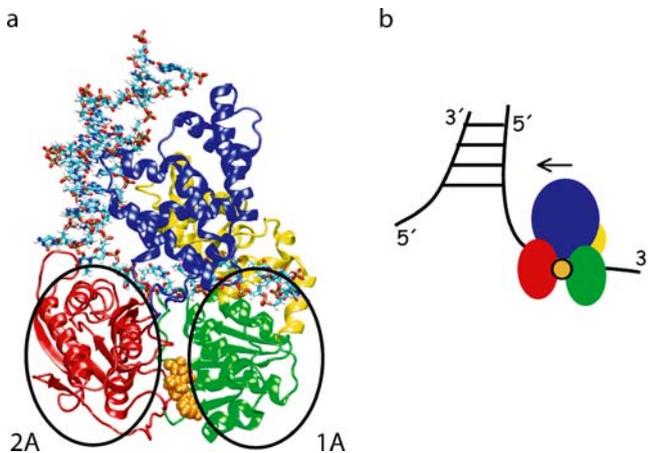
Fig. 1 Molecular view of PcrA helicase. **a** Depicted is the protein in cartoon representation. Also shown is a strand of ssDNA in licorice representation threading through the center of PcrA at the interface of the 1A (*green*) and 2A domains (*red*) and which is attached to a short piece of dsDNA adjacent to the 2B domain (*blue*); also shown is the 1B domain (*yellow*) and a single molecule of ATP (*orange*) bound to the catalytic site at the interface between the 1A and 2A domains (depicted in vdW representation). **b** Shown is a schematic view of PcrA translocating along ssDNA toward the dsDNA fork; domains 1A, 2A, 1B, and 2B are shown in *green*, *red*, *yellow*, and *blue*, respectively

In the following, we will focus on the mechanism of PcrA movement along ssDNA. On the basis of their crystal structure data and on the intuitive idea of alternating affinities, Velankar et al. [19] proposed an "inchworm" model for the translocation of PcrA along ssDNA that involves concerted movements of ssDNA bases across the interface with PcrA. This model does not, however, address how ATP hydrolysis is coupled to domain motion, what residues are most crucial during this process, how the proposed alternating affinities can be justified microscopically and finally, how uni-directionality is achieved. These are the questions that will be discussed in the following sections. We will start by looking at the ATP hydrolysis reaction in the catalytic site and then work our way up toward longer time and larger length scales.

3
Zooming in on ATP Hydrolysis in the Catalytic Site

The catalysis of ATP lies at the heart of PcrA function and takes place in a single catalytic site at the interface between the 1A and 2A domain (Fig. 1). Several important questions arise when considering the role of ATP catalysis during the overall translocation process of PcrA along ssDNA. First and from a purely enzymatic point of view, one would like to know how efficient ATP catalysis is being achieved by the protein. Second, one would like to investigate how the actual catalytic step is coupled to larger scale protein conformational changes eventually leading to PcrA translocation along ssDNA.

An investigation of both questions requires a proper treatment of the electronic degrees of freedom during the catalytic reaction in the biological context provided by the protein environment. QM/MM methods present a powerful and versatile way of addressing these issues, since they allow one to conduct high level electronic structure calculations of a particular region of interest in the presence of a classically treated protein environment. The following paragraph provides a brief outline of the principles behind QM/MM calculations, before we describe some of the insights into PcrA function that have emerged using this approach.

QM/MM Calculations

The underlying theme of QM/MM simulations [22–29] is a quantum mechanical treatment of a certain molecular core region that is undergoing electronic level changes in the presence of a (large) biomolecular environment given by, for example, protein, solvent, or DNA/RNA. The environment itself is assumed to be electronically inert and is, therefore, treated classically, typically with a molecular mechanics method. The main motivation for this separation into a quantum mechanically and a classically treated region is the fact that quantum mechanical calculations are computationally very expensive

and can only be applied to relatively small systems of at most a few hundred atoms. Hence, QM/MM simulations provide a compromise between the need to treat certain parts of the system at an electronic level and the requirement for the presence of the sizeable host biomolecule or biomolecular complex.

Depending on the particular QM/MM method chosen, it is sometimes necessary to limit the number of classically treated MM atoms. In the case of PcrA which contains more than 110 000 atoms in its solvated state (Sect. 4), the conducted QM/MM calculations described below were restricted to a reduced sub-system containing only 20 000 atoms that was centered around the nucleotide ATP bound to the catalytic site. This is depicted schematically in Fig. 2, which shows the full QM/MM system indicating the quantum mechanically (QM) and classically treated regions (MM), respectively.

The remainder of this section will provide a brief overview of QM/MM simulations in general and the implementation used in the studies summarized in the present manuscript in particular. In all of the following it is assumed that the Born–Oppenheimer approximation is valid, i.e., electronic and nuclear degrees of freedom can be separated. In this case, the QM/MM Hamiltonian describing the quantum mechanically modeled core region and the classically treated protein environment can formally be expressed as

$$H = H_{QM} + H_{MM} + H_{QM/MM} , \qquad (1)$$
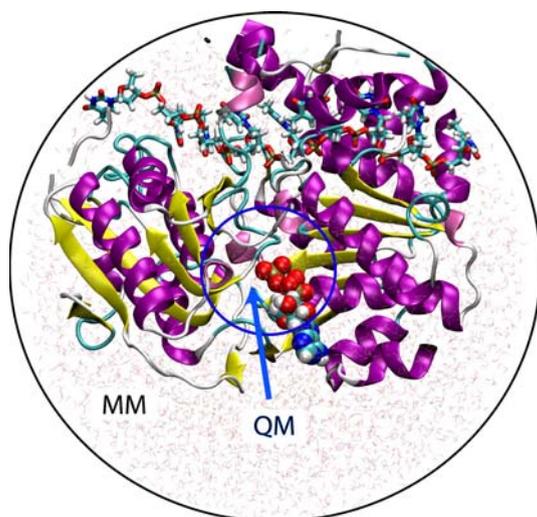
Fig. 2 QM/MM simulation system for PcrA. Shown is the complete model which is centered around the catalytic binding site located at the interface between the 1A and 2A domains and contains about 20 000 atoms. The protein is shown in cartoon representation, a piece of bound ssDNA is depicted in licorice representation, and bound ATP in vdW representation. The classically (MM) and quantum mechanically (QM) treated regions are indicated by *circles* and labeled accordingly

where H_{QM} can be identified with the Hamiltonian of the time-independent Schrödinger equation

$$H_{QM} \, |\psi(\{r\})\rangle = E \, |\psi(\{r\})\rangle \tag{2}$$

and the classical contribution H_{MM} is given by Eq. 14 below.

The main challenge for achieving a proper QM/MM description lies in the treatment of the interface between the quantum and classical regions, described by $H_{QM/MM}$. QM/MM studies of chemical reactions in solution allow for a straightforward separation of both regions, e.g., the solvent being treated using a classical description and the chemically active solute being modeled by a QM method. QM/MM simulations of biomolecular systems like, for example, enzymes, on the other hand, often lead to interfaces separating the quantum and classical regions that cut through covalent bonds, thereby, complicating the description significantly.

Broadly, one can distinguish three types of QM/MM coupling schemes. These are listed below in the order of increasing complexity:

- *Mechanical Embedding.* This represents the crudest level of description. The QM and MM regions are not connected by any covalent bonds and interact only via steric constraints, typically implemented in terms of a van der Waals (vdW) type interaction:

$$H_{QM/MM} = \sum_{a \in MM} \sum_{i \in QM} H_{vdW}^{ai} \, . \tag{3}$$

  The two sums extend over all nuclei in the classically (MM) and quantum mechanically (QM) treated region, respectively. The choice of vdW parameters is not straightforward. Often, the parameters provided by the classical forcefield are employed, but QM/MM studies with specifically parameterized vdW interactions have also been reported.

- *Electrostatic Embedding.* At this level of QM/MM coupling, in addition to steric constraints between the QM and MM regions, one also takes into account their electrostatic interaction, leading to a QM/MM Hamiltonian of the form:

$$H_{QM/MM} = \sum_{a \in MM} \sum_{i \in QM} \left(H_{vdW}^{ai} + H_{ele}^{ai} \right) \, . \tag{4}$$

  Here, H_{ele} describes the electrostatic interaction between the QM and MM regions. As will be discussed in more detail below, this term can be evaluated in several different ways.

- *Covalent Embedding.* This most sophisticated level of QM/MM scheme involves, in addition to steric and electrostatic interactions, covalent bonds between QM and MM atoms. Typically, this is the description necessary for simulations of biomolecular systems, and was used in our QM/MM approach.

The next section is devoted to a more in depth exposition of the particular covalent embedding scheme used in our calculations.

RESP Charge-Based Covalent Embedding

A number of different covalent embedding schemes have been proposed since the early QM/MM studies of Warshel [30]. Here, rather than trying to be comprehensive, we focus on the particular method employed in the QM/MM implementation on which the reported results are based and refer to the literature for additional information. See also the review on QM/MM methods by Senn and Thiel in this volume.

When describing quantum mechanical events inside biomolecules one frequently encounters covalent bonds across the QM/MM boundary connecting classically and quantum mechanically treated atoms. This is problematic for several reasons. First and foremost, "cutting" these bonds creates an open valence in the QM region. Several schemes have been devised to deal with this problem. The probably most straightforward method is the so-called link atom approach [23] that is also used in our QM/MM method. Here, the dangling bond is simply terminated by a "dummy" hydrogen atom. More sophisticated schemes have been proposed, for example, the local self-consistent field method by Théry et al. [31] in which a frozen hybrid orbital with a predetermined density is used to satisfy the valence at the boundary. However, even though the introduction of an additional (hydrogen) atom in the link atom approach can in principle cause artificial behavior at the boundary, it is currently not clear that alternative, more sophisticated methods provide a much improved description.

A second problematic issue arising from a QM/MM boundary that dissects covalent bonds is the difficulty of achieving a proper separation of the QM and MM interaction terms in order to avoid the double counting of certain interactions. This will be discussed in more detail below.

The QM/MM interface Hamiltonian of the covalent embedding scheme describing the quantum-classical interaction can be written as

$$H_{\text{QM/MM}} = \sum_{a \in \text{MM}} \sum_{i \in \text{QM}} \left(H_{\text{vdW}}^{ai} + H_{\text{ele}}^{ai} \right) + H_{\text{covalent}}^{\text{QM/MM}} . \tag{5}$$

The $H_{\text{covalent}}^{\text{QM/MM}}$ contribution maintains the proper conformation across the covalent quantum-classical boundary and is provided by the molecular mechanics force field. Quantum mechanically treated atoms in the boundary region experience all classical bonded force contributions that involve at least one classical atom.

In most conventional QM/MM approaches the electronic contribution to the electrostatic interaction H_{ele}^{ai} in Eq. 5 is given via the one electron

operator

$$H_{\text{ele,e}}^{ai} = \frac{q_a}{|\mathbf{R}_a - \mathbf{r}_i|} \,, \tag{6}$$

where i indexes the electron coordinates and q_a is the partial charge on the classical atom a at position $\mathbf{R}_a$. Unfortunately, as indicated above, this approach leads to problems at the covalent QM/MM boundary where electrostatic interactions between quantum and classical atoms are already implicitly accounted for via the classical bond, angle, and dihedral energies. Subtracting these interactions from the expression Eq. 6 is difficult since the electronic charge is de-localized over the whole QM segment and, therefore, can not be easily partitioned among atoms. In other words, there is no quantum mechanical charge operator that would allow the assignment of partial atomic charges. To address this shortcoming, our QM/MM approach introduces an effective charge operator $\hat{\xi}_p$ for each atomic site in the QM segment, that is evaluated via the restrained electrostatic potential (RESP) charge method [32]. This leads to the following expression for the electrostatic potential

$$H_{\text{ele}} = H_{\text{ele,e}} + H_{\text{ele,n}} \tag{7}$$

$$= \gamma \sum_{a \in \text{MM}} \sum_{p \in \text{QM}} \kappa_{ap} \frac{\hat{\xi}_p q_a}{|\mathbf{R}_a - \mathbf{R}_p|} + \sum_{a \in \text{MM}} \sum_{p \in \text{QM}} \kappa_{ap} \frac{Z_p q_a}{|\mathbf{R}_a - \mathbf{R}_p|} \,,$$

where $H_{\text{ele,e}}$ and $H_{\text{ele,n}}$ are the electronic and nuclear contribution, respectively. Here q_a and Z_p are the partial charges on the classical atoms and nuclear charges of the quantum mechanically modeled atoms, respectively. The κ_{ap} mediate the proper exclusion of the classical non-bonded interactions at the QM/MM boundary, and γ is a scaling factor to ensure the proper total charge of the QM/MM system.

Using the Hamiltonian Eq. 1 together with the QM/MM interaction energy Eq. 5, one obtains the following expression for the total energy E

$$E = \langle \Psi | H | \Psi \rangle \tag{8}$$

$$= E^{\text{QM}}(\mathbf{P}) + \gamma \sum_{a \in \text{MM}} \sum_{p \in \text{QM}} \kappa_{ap} \frac{\xi_p(\mathbf{P}) q_a}{|\mathbf{R}_a - \mathbf{R}_p|} + E^{\text{MM}} \,.$$

Here, $|\Psi\rangle$ is the electronic wave function and $\mathbf{P}$ the corresponding electronic density matrix. The term E^{QM} is the expression for the gas phase energy of the QM part, and E^{MM} includes all remaining non-electronic energy contributions. In the second term, $\xi_p(\mathbf{P})$ is the effective charge for atoms in the QM segment, which is evaluated via the RESP method by minimizing the follow-

ing expression

$$I = \sum_{\alpha} \omega_{\alpha} \left[\sum_{p \in QM} \frac{\xi_p(\boldsymbol{P})}{|\boldsymbol{R}_{\alpha} - \boldsymbol{R}_p|} - V_{\alpha}(\boldsymbol{P}) \right]^2 + 2\lambda_e \left[\sum_{p \in QM} \xi_p(\boldsymbol{P}) - N_e \right] \qquad (9)$$

$$+ \sum_{p \in QM} g_p \left(\xi_p(\boldsymbol{P}) + Z_p \right)^2 .$$

The first term on the right-hand side attempts to find effective charges $\xi_p(\boldsymbol{P})$ that best reproduce the exact electronic electrostatic potential $V_{\alpha}(\boldsymbol{P})$ subject to the constraints expressed in the second and third term. Here, α sums over the points of a grid located at positions $\boldsymbol{R}_{\alpha}$ surrounding the QM segment at which the electrostatic potential is evaluated. The ω_{α} are weight factors whose value depends on the particular choice for the computational grid. The second term enforces charge conservation and constrains the overall effective charge to be equal to the total number of electrons N_e via a Lagrange multiplier λ_e. Both $V_{\alpha}(\boldsymbol{P})$ and N_e have to take into account the presence of the dummy hydrogen atoms. Finally, the third term represents a harmonic penalty function that can be adjusted by the parameter g_p to prevent the appearance of unreasonably large effective charges [32].

The minimization of the expression given by Eq. 9 can be cast into a matrix equation for $\xi_p(\boldsymbol{P})$ which can then be used to formulate a new Fock operator

$$F^{QM/MM} = F + F^{solv} , \qquad (10)$$

where F is the gas-phase Fock operator and F^{solv} the so-called solvated Fock operator mediating the electrostatic interaction between the QM and MM segment of the QM/MM system [33]. The self-consistent solution of the Roothaan-Hall equations with the Fock operator given by Eq. 10 is the mathematical basis of the QM/MM approach used in the studies described in the remainder of this section.

Finally, we would like to point out that due to the computational cost involved, the energetics derived from ab initio QM/MM calculations typically are enthalphies rather than free energies. The latter are required, to, for example, derive true reaction free energies or equilibrium constants. Hence, for each particular system it is necessary to take into account possible entropic effects when arguing based on calculated enthalpies. There exist, however, semi-empirical QM/MM approaches that are capable of computing true free energy surfaces, at the possible expense of reduced accuracy and transferability [34, 35].

QM/MM Study of the Chemo-Mechanical Coupling in PcrA Helicase

The QM/MM approach just described was employed to investigate the ATP hydrolysis reaction in the catalytic site of PcrA helicase and its coupling to

protein conformational changes, i.e., the chemo-mechanical coupling [36].
The reactant state conformation of the quantum mechanically treated region
is shown in the upper left-hand side panel of Fig. 3 which also depicts import-
ant distances between atoms.

Figure 3 reveals that all binding pocket residues immediately adjacent to
the tri-phosphate part of ATP are either charged or polar, with several water
molecules surrounding the γ-phosphate group. A magnesium ion is located
between two oxygen atoms of the β and γ-phosphate groups and is hydrogen-
bonded to additional water molecules and binding pocket residues. Most of
the residues forming the catalytic binding sites are located in the 1A domain
of PcrA with the exception of two arginine residues, R287 and R610, which are
contributed by the neighboring 2A domain (c.f. bottom panel of Fig. 3).

Using QM/MM calculations, the ATP hydrolysis reaction pathway can be
calculated [36] and the conformations of the reactant, transition, and product
state are shown in the upper panel of Fig. 3 with the corresponding energies
depicted in Fig. 4.

The center column of Fig. 3 shows the transition state conformation during
ATP hydrolysis. It features a hydronium ion that evolved via proton transfer
from the nucleophilic water, W1, toward W2. During the conversion from the
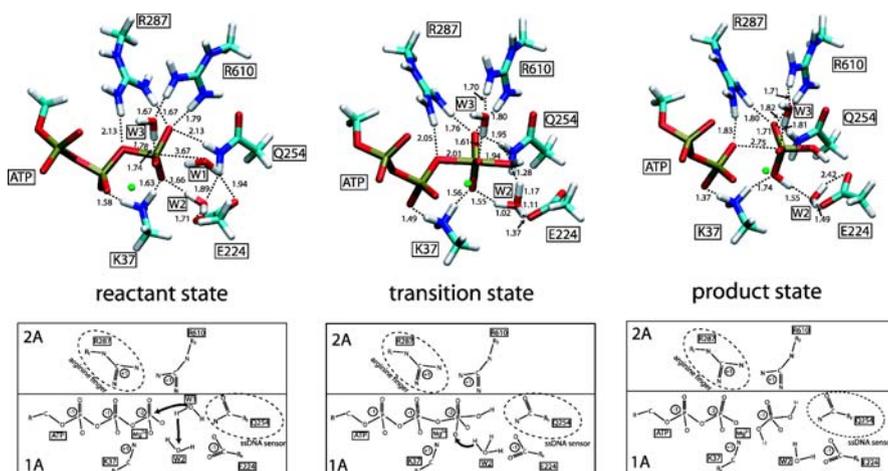
Fig. 3 Conformation of the quantum mechanically treated core region in the catalytic site
of PcrA. Shown in the *top panel* are molecular views of the structures of the reactant (*top
left*), transition (*top middle*), and final (*top right*) state along the ATP hydrolysis reac-
tion pathway. Important distances between atoms are indicated and given in units of Å.
The *bottom panel* depicts schematic views of the catalytic site in each of the three states
and shows which of the residues belong to domain 1A or 2A. The *bold arrows* sketch the
proton relay mechanism discussed in the text. The *color coding* is as follows: *white* - hy-
drogen; *light blue* – carbon; *dark blue* – nitrogen; *red* – oxygen; *gold* – phosphorus; *green*
– magnesium ion (adapted from [36])
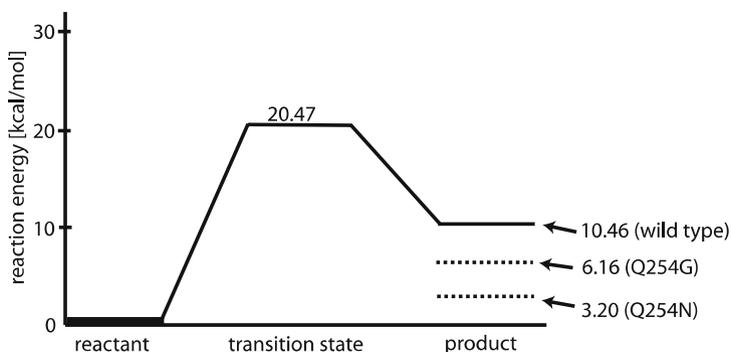
Fig. 4 Energetics of ATP hydrolysis in PcrA. Shown are the energies in kcal/mol of the reactant, transition, and product state during ATP hydrolysis in the catalytic sites of PcrA. For the product state, the energies for the wild type and two mutants, Q254N and Q254G, are provided (adapted from [36])

transition state toward the product conformation, the hydronium ion transfers a second proton to the γ-phosphate group to yield products ADP and P_i. This mechanism of proton transfer from the nucleophilic water toward the γ-phosphate group of ATP is sketched by the bold arrows in the bottom panel of Fig. 3 and was termed a proton relay mechanism [10], since it involves two water molecules, W1 and W2. Until recently, it was thought that direct proton transfer from the nucleophilic water, W1, toward one of the γ-phosphate oxygen atoms of ATP is the physiologically relevant pathway. However, QM/MM studies have shown that the proton relay mechanism is energetically more favorable by up to 20 kcal/mol [36]. Such a mechanism had also been proposed based on theoretical studies of other ATP-driven molecular motors, such as F_1-ATPase [10, 12] and myosin [1], and had been proposed also based on structural and biochemical evidence for ras [37] and F_1-ATPase [38]. All combined, this evidence suggests that such a proton relay mechanism is responsible for efficient hydrolysis in a wider class of ATP-driven molecular motors. Since the proton relay mechanism relies on the proper arrangement of water molecules in the binding pocket, which is in turn facilitated by the protein environment, it represents a genuinely enzymatic pathway.

Figure 4 gives the energies of important states along the ATP hydrolysis pathway in PcrA. Several features are noteworthy: First, the QM/MM calculations [36] reveal the presence of a modest transition state barrier of ~ 20 kcal/mol in good agreement with experimentally measured rates ($\sim 10^2$ s^{-1}). This moderate barrier is due to the proton relay mechanism as shown by comparison with calculations of direct proton transfer pathways. In addition, there are several binding pocket residues that contribute to a lowering of the transition state barrier. This is shown in Fig. 5 which depicts the change in electrostatic interaction between ATP and the protein environment upon going from the reactant to the transition state. Clearly, K37,
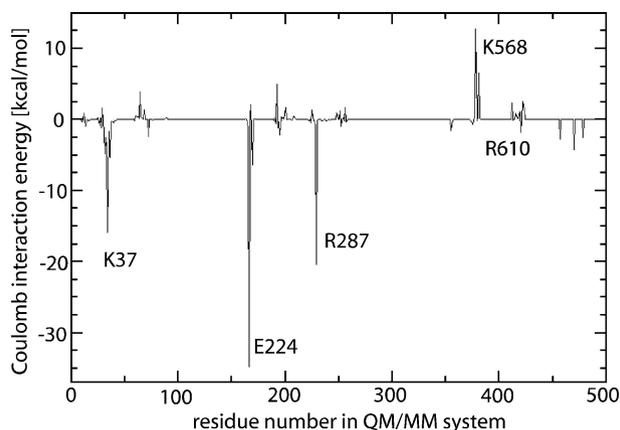
Fig. 5 Electrostatic interactions between catalytic core and protein environment. Shown are the changes in electrostatic interaction between the nucleotide and the protein environment upon going from the reactant to the transition state (adapted from [36])

E224, and R287 stabilize the transition state, whereas K568 leads to a slight de-stabilization. Somewhat surprisingly, even though R287 provides a large contribution, its immediate structural neighbor, R610, does not.

The second important feature to notice in Fig. 4 is the endothermicity of the reaction energy profile in the wild type system, implying that the catalytic site in the conformation studied has the ability to tightly bind ATP, but not yet to hydrolyze it. To investigate this in more detail and to identify means by which the protein could achieve a lowering of the product state energy, *in silico* mutation studies can be used [36]. Here, selected residues are mutated in the model system and the reaction energy profile is recalculated either fully or for selected points along the reaction pathway. In the present case, such an analysis shows, for example, that the mutations Q254N and Q254G exert a significant influence on the product state energy. This, in turn, allows one to hypothesize that motion of Q254 with respect to P_γ leads to the lowering of the product state energy to enable efficient ATP hydrolysis. Q254 links to a protein pocket known to be involved in ssDNA base flipping [19] during translocation via the inchworm mechanism. This suggests a direct connection between translocation along ssDNA and the actual catalytic step in the ATP binding pocket of PcrA via Q254.

Finally, based on the ATP hydrolysis reaction energy profile in PcrA shown in Fig. 4, the in silico mutation studies, and the findings from QM/MM simulations of ATP hydrolysis in F_1-ATPase [10, 12], the overall ATP catalysis reaction energy profile can be assumed to have an equilibrium constant of $K \sim 1$, corresponding to an iso-energetic reaction energy profile [36]. This has important implications, since it means that the actual chemical bond-breaking step in the catalytic sites of PcrA does not release any net free energy and is

hence not coupled to force generation. Similar to F_1-ATPase [7, 12, 39], it is likely that the binding of reactant ATP to the initially open catalytic cleft at the interface between the 1A and 2A domains leads to a closing motion which provides the actual power stroke propagating PcrA along ssDNA. This picture fits nicely with the above observation that insertion of R287 from domain 2A into the catalytic site located in the 1A domain is required for efficient ATP hydrolysis to take place. Such a coupling of large-scale protein conformation and the chemical reaction assures that ATP hydrolysis does not occur prematurely, which could destroy the directed motion of PcrA along ssDNA [36].

However, due to the static nature of the employed QM/MM method and the limited spatial extent of the QM/MM system, one is limited in one's ability to test the domain motion hypothesis and the coupling of Q254 to ssDNA base flipping on realistic time and length scales. This is where all-atom MD simulations and coarser-grained descriptions come into play. These methods will be the topic of the following sections.

4
Molecular Dynamics Simulations of PcrA Function

Continuing our study of molecular motor function in PcrA, we turn to MD simulations to investigate PcrA's larger scale conformational motions leading to translocation along ssDNA and to study the coupling of ATP hydrolysis to unidirectional translocation. Figure 6 shows a molecular view of an all-atom simulation system of PcrA, containing the protein, DNA, TIP3P water molecules, and a number of ions adding up to a total of more than $110\,000$ atoms. Before discussing some of the findings for PcrA function obtained via MD simulations we will give a brief exposition of MD simulation methods.

MD Simulations—A Brief Introduction

Classical simulations are based on the solution of the differential equation posed by Newton's second law for a system of N interacting particles

$$\ddot{x}_i = \frac{1}{m_i} F_i \quad (i = 1, ..., N) . \tag{11}$$

Here $\ddot{x}_i$ is the second time derivative of the position of particle i, m_i its mass, and F_i the total force acting on it. F_i depends on the positions of all other particles. The calculation of the forces F_i is the step requiring the most computational effort.

A number of algorithms have been developed to numerically integrate Eq. 11 and, thereby, obtain the particles' trajectories $x_i(t)$. Many integration schemes use the *Leap frog algorithm* for updating coordinates and velocities
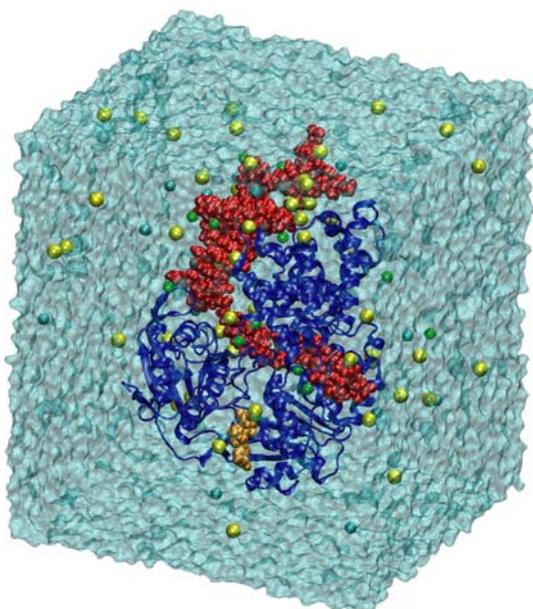
Fig. 6 PcrA all-atom simulation system. Shown is the protein in a *blue* cartoon representation with bound dsDNA and ssDNA shown in *red* vdW representation. The nucleotide ATP bound to the catalytic site at the interface between the 1A and 2A domains is shown as *gold*-colored vdW spheres. The protein-DNA complex is immersed in a large box of TIP3P water molecules containing Cl$^-$ (*cyan*), Na$^+$ (*yellow*), and Mg^{2+} (*green*) ions (adapted from [51])

of particle i:

$$r_i\left(t + \Delta t\right) = r_i(t) + \Delta t\, v_i\left(t + \Delta t/2\right) \tag{12}$$

$$v_i\left(t + \Delta t/2\right) = v_i\left(t - \Delta t/2\right) + \Delta t\frac{1}{m_i}F_i\,, \tag{13}$$

with Δt being the value of the discrete time-step. In biomolecular simulations, the value of Δt has to be a fraction of the fastest vibrational period in the system and, therefore, is in the 1–2 fs range.

All classical simulations are based on a classical approximation to the Hamilton operator in the Schrödinger Eq. 2. With this classical Hamiltonian one can then proceed to calculate energies, forces, and higher derivatives and use them to, for example, compute the minimum energy conformation or simulate the dynamical behavior via solving Eq. 11.

In the following we will focus on the classical Hamiltonians that are commonly used for simulations of biomolecular systems. They are referred to as *force fields* and are known under names such as CHARMM [40], GROMACS [41], and AMBER [42]. They all have in common a relatively simple mathematical structure to allow for a computationally efficient force evalu-

ation. Furthermore, these force fields are empirical, i.e., they contain fitting parameters, that are either determined by comparison to experiments, via quantum mechanical calculations, or both.

A typical force field is described by the following Hamiltonian

$$H_{\mathrm{MM}} = H_{\mathrm{bnd}} + H_{\mathrm{ang}} + H_{\mathrm{dih}} + H_{\mathrm{vdW}} + H_{\mathrm{ele}} . \tag{14}$$

Here, the first three terms represent the bonded contributions, namely the bond

$$H_{\mathrm{bnd}} = \sum_{a}^{\mathrm{bonds}} \frac{1}{2} k_a^{\mathrm{bnd}} \left(x_a - x_{0a} \right)^2 , \tag{15}$$

and angle terms

$$H_{\mathrm{ang}} = \sum_{a}^{\mathrm{angles}} \frac{1}{2} k_a^{\mathrm{ang}} \left(\theta_a - \theta_{0a} \right)^2 , \tag{16}$$

that are both simple harmonic functions of the bond lengths x_a and angles θ_a, respectively. The parameters x_{0a} and θ_{0a} denote the corresponding equilibrium values and the $k_a^{\mathrm{bnd/ang}}$ are the interaction strengths.

The contributions of the dihedral angles ϕ_a are periodic, typically of small magnitude, and are given by

$$H_{\mathrm{dih}} = \sum_{a}^{\mathrm{dihed}} k_a^{\mathrm{dih}} \left(1 + \cos \left(n\phi_a - \phi_0 \right) \right) . \tag{17}$$

Again, ϕ_0 is the equilibrium value of the dihedral angle, k_a^{dih} the interaction strength, and n determines the periodicity. The last two terms in Eq. 14 are the non-bonded contributions to the force field. The short range part is represented by a Lennard-Jones potential

$$H_{\mathrm{vdW}} = \sum_{a \neq b}^{N} 4\epsilon_{ab} \left[\left(\frac{\sigma_{ab}}{r_{ab}} \right)^{12} - \left(\frac{\sigma_{ab}}{r_{ab}} \right)^{6} \right] , \tag{18}$$

where r_{ab} is the distance between particles a and b, and the quantities ϵ_{ab} and σ_{ab} correspond to the depth of the Lennard-Jones potential and to the interparticle distance where the potential is equal to zero, respectively.

Finally, the last term represents the Coulomb interaction between pairs of atoms with partial charges q_a and q_b

$$H_{\mathrm{ele}} = \sum_{a \neq b}^{N} \frac{q_a q_b}{r_{ab}} . \tag{19}$$

Because of the $1/r$ dependence, the Coulomb interaction is long-range and, therefore, a straightforward calculation requires a computational effort of

order $O(N^2)$ in the number of particles N. Computationally, it is, therefore, the most expensive part in evaluating Eq. 14 and several numerical approximations, for example the Particle Mesh Ewald method, have been devised to reduce the computational cost of this term to $O(N \log N)$.

Using efficient parallelization schemes, such as that employed by the molecular dynamics program NAMD [43], it is presently possible to calculate the time evolution of a biomolecular system using Newton's equation of motion combined with an expression for the force field similar to the one given in Eq. 14 for a million atoms over a time of tens of ns [44]. MD calculations were traditionally performed in the micro-canonical (NVE) ensemble; currently, most MD simulations are conducted in the canonical (NVT) and grand-canonical (NPT) ensemble, i.e., under conditions of constant volume/pressure. The temperature of the system can be kept constant by coupling the system to an external heat bath at the desired temperature [45]; the pressure can be controlled by using, for example, the Nose-Hoover Langevin piston method [46, 47]. A rather comprehensive introduction into concepts and algorithms underlying classical MD simulations can be found in [43].

Molecular Dynamics Simulations of PcrA

Having said this, it becomes apparent that the main difficulty in applying MD methods to the study of molecular motors is the vast gap between the time-scales available to equilibrium MD simulations (~ 100 ns) and the ones relevant physiologically (μs to ms). Since the translocation of PcrA by a single nucleotide takes about 20 ms, the accompanying protein motions, among them, for example, the closing of the catalytic cleft between domains 1A and 2A or the motion of bound ssDNA bases, take place on a similar length time-scale. Hence, investigating PcrA function with equilibrium MD simulations is hampered by this gap in time-scale and different strategies are needed.

One possible approach to elucidate the coupling of ATP binding and hydrolysis to changes in PcrA conformation consists of the analysis of equilibrium MD trajectories of PcrA in the ATP and ADP+P$_i$ bound states. Figure 7 shows the per residue RMSD averaged over several ns for both states, clearly illustrating that at least on a ns time scale, there is no significant difference between the two, rendering this approach futile.

The time-scale gap can be partially overcome by applying an external biasing potential, enforcing a certain change in conformation, via, for example steered molecular dynamics simulations (SMD) [48, 49] or umbrella sampling [50]. The application of these approaches to molecular motors, however, suffers from several limitations. First, it requires a certain insight into what the physiologically relevant motions are, in order to bias the simulations properly. Second, biasing can be difficult for all but the simplest motions. Finally, enforcing ms time-scale protein motions on a ns computational time-scale can lead to artifacts or, in the worst case, destroy the protein.
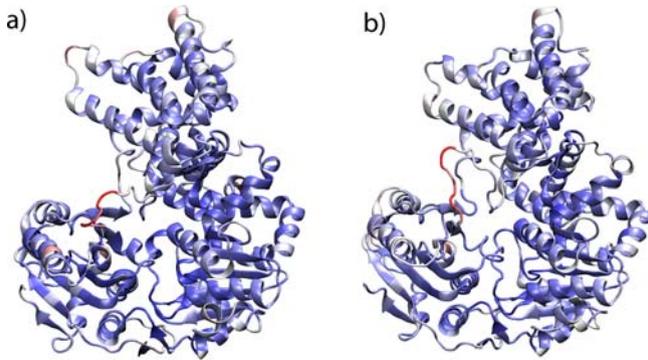
Fig. 7 Trajectory averaged per-residue RMSD in PcrA. Shown are the RMSD values based on ∼3.5 ns of MD simulations for the ATP (**a**) and ADP/P$_i$ (**b**) bound PcrA systems. The *color scale* ranges from *blue* to *red* and indicates increasing RMSD values. The most mobile part of the structure is a flexible loop connecting the 2A and 2B domains *colored in red*

Nevertheless, despite all these drawbacks, significant insight into the system can be gained if the approach is applied properly and can be used, for example, to calculate potentials of mean force or force-extension curves.

In the case of PcrA, such a biasing method can be employed to investigate the mechanism behind the directionality of ssDNA motion in the $3' \rightarrow 5'$ direction. As proposed in the inchworm translocation model [19] derived from structural data, the 1A and 2A domains each have different affinities for ssDNA, and both alternate their affinities between the substrate (with ATP bound, *s*) and product (without ATP/ADP bound, *p*) state. The model, however, does not provide any microscopic evidence to support such a mechanism. To test this idea further, SMD simulations have been used by attaching ten harmonic springs (force constants of 2 kcal/mol Å^2) to ten phosphorous atoms of the ssDNA bound to PcrA [51]. The ssDNA was then pulled one half nucleotide (nt) distance forward and backward across the protein-ssDNA interface. The direction of pulling in the *s* and *p* states was chosen according to the anticipated domain motion in each state corresponding to separation and closure of domains 1A and 2A, respectively. The required pulling forces were monitored and are reproduced in Fig. 8. These results show that in the *s* state, the average force needed to move nucleotides 15–18 to the right (corresponding to forward motion of 2A to the left) is smaller than that needed to move nucleotides 18–21 to the left (corresponding to backward motion of 1A to the right). Similarly, in the *p* state, the average force needed to move nucleotides 15–18 to the left (corresponding to backward motion of 2A to the right) is larger than the force to move nucleotides 17–20 to the right (corresponding to forward motion of 1A to the left). This directly supports the idea of alternating affinities between the protein domains and ssDNA in different PcrA states and can explain the directionality of PcrA translocation.
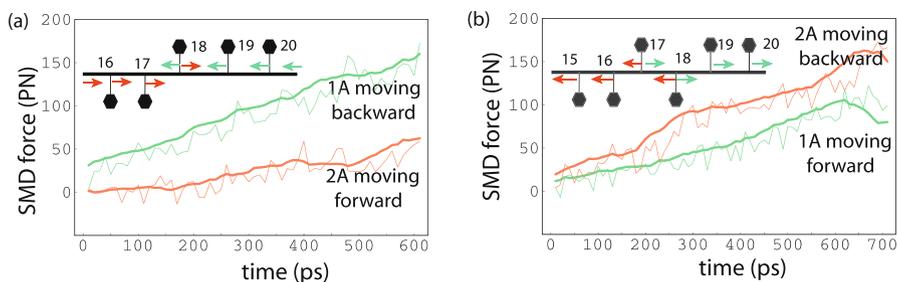
Fig. 8 Comparisons of SMD forces arising during ssDNA pulling simulations in PcrA. Shown are the results with ATP bound (*s* state) (**a**) and without ATP bound (*p* state) (**b**). The *green/red curve* represents the average force needed to move relevant nucleotides in a direction opposite to the corresponding movement of domain 1A/2A (indicated by *green/red arrows*). Numbers are used to label nucleotides on ssDNA. The *thin curves* were measured directly from simulations, while the *thick curves* were smoothed over every ten data points. The results show that in **a**, the average force needed to shift the relevant nucleotides, corresponding to the domain movement of 2A, is smaller than the average force needed to shift the relevant nucleotides corresponding to the domain movement of 1A; the opposite is true in **b** (adapted from [51])

In order to further investigate the microscopic mechanism underlying the alternating affinities between the 1A and 2A domains and the bound ssDNA, all-atom ns time-scale MD simulations were used to derive an effective potential, $U_{i\sigma}(x_i)$, governing individual domain motion of PcrA. Here, σ is a state index describing either *s* or *p*; $i = 1,2$ refers to domain 1A and 2A, respectively, with x_i being the position of each domain along the bound ssDNA. The effective potential $U_{i\sigma}(x_i)$ is then used to formulate a coarse-grained Langevin dynamics description of PcrA that will be described in the next section. This approach [51] allows one to describe the ATP hydrolysis coupled uni-directional ssDNA translocation of PcrA on the physiologically relevant ms time-scale.

Since the translocation of helicase along ssDNA arises through the sequential binding and unbinding of ssDNA units (nucleotides) to the PcrA domain surfaces, one possible strategy for determining $U_{i\sigma}$ is to calculate the binding free energies of individual ssDNA units with the protein, and then use an appropriate combination of these energies to obtain $U_{i\sigma}$. The calculation of the binding free energy from MD simulations is not straightforward, but can be accomplished by a semi-empirical method [52]. Here, the binding energy E_b is approximated as a weighted sum of electrostatic and vdW interactions between the protein and each individual nucleotide as extracted from equilibrium MD simulations. In the present case, only relative but not absolute energies are needed. Hence, the solvent (water and ion) contribution to the energy can be neglected by assuming a homogeneous solvent-nucleotide interaction along the bound ssDNA region (consisting of poly-thymine in the

simulation model) [51]. From the discrete set of nucleotide-protein inter-action energies derived from MD, a continuous site-energy function $E_b(x)$ for each state (s and p) was constructed by polynomial interpolation [51]. Figure 9 shows the continuous site energies $E_b(x)$ as well as the resulting potential energy functions $U_{i\sigma}$ in both the s and p states.

Now we give an example of how to derive $U_{i\sigma}$ from the binding energies $E_b(x)$ [51]. Starting from the s state, as domain 2A moves forward (to the left) by a distance Δx ($\Delta x \leq 0.5$, in unit of a single nt distance), nucleotides 15 to 18 move backward (to the right) by Δx, while the other nucleotides remain stationary. This scenario can be inferred from the available X-ray crystal structures. Hence, the potential U_{2s}, for example, can be expressed as

$$U_{2s}(\Delta x) = \sum_{j=15}^{18} \left[E_b(x_j + \Delta x) - E_b(x_j) \right] \quad \Delta x \in [0, 0.5] . \tag{20}$$

The other $U_{i\sigma}$ can be computed in a similar fashion. In the derivation, we assume a symmetric form of $U_{i\sigma}$ with a barrier at the center at $\Delta x = 0.5$, where $A_{i\sigma} = U_{i\sigma}(0.5)$ is defined as the barrier height. According to Eq. 20, $A_{i\sigma}$ in-
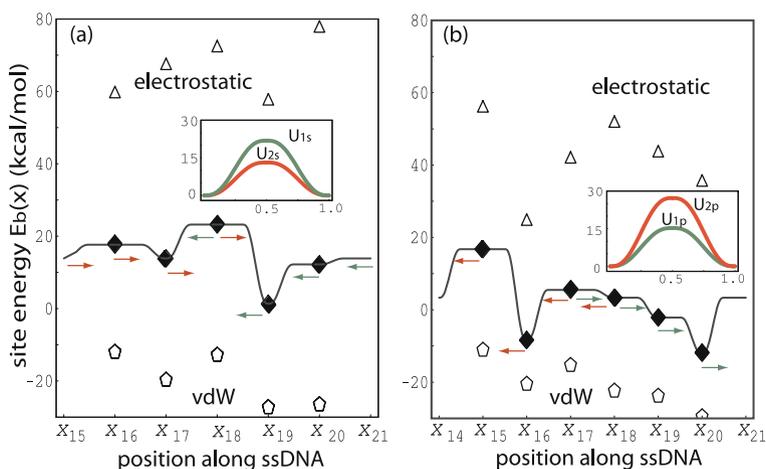
Fig. 9 Site energies of ssDNA nucleotides and individual domain potentials for PcrA. Shown are the result for the substrate (**a**) and the product state (**b**). *Filled dark diamonds* represent the relative binding free energies of nucleotide i at position x_i, i.e., the weighted sum of electrostatic and vdW energies between protein and individual nucleotides. The separate electrostatic and vdW contributions are indicated through *non-filled triangles* and *non-filled pentagons*, respectively. A smooth site energy function $E_b(x)$ is drawn through the *filled dark diamonds* using a third-order polynomial interpolation. The *inset* shows the potential $U_{i\sigma}(\Delta x)$ as derived from $E_b(x)$ that is experienced by domain 2A (*red solid curve*) and 1A (*green solid curve*) as the domains move along ssDNA; the length scale is in units of one nt distance (6.5 Å) (adapted from [51])

cludes a sum of site energy terms $E_b(x_j + 0.5) - E_b(x_j)$. Following [51] one can write this as

$$A_{i\sigma} = \Sigma_{i\sigma} + \Delta \,, \tag{21}$$

where $\Sigma_{i\sigma}$ are sums of the differences of binding energies $E_b(x_j)$. Δ is a single tunable parameter that accounts for the effect of additional friction due to ssDNA-protein interactions and is assumed to be independent of i and σ. The value of Δ is adjusted so that the translocation speed of PcrA is consistent with that measured from experiment, i.e., about one nt distance every 20 ms.

It is important to point out that the difference between amplitudes, $|A_{1\sigma} - A_{2\sigma}|$, is independent of Δ. The calculations show that $A_{1s} - A_{2s} \sim 9$ kcal/mol in the s state and $A_{1p} - A_{2p} \sim$ -12 kcal/mol in the p state [51]. This result, i.e., $A_{1s} > A_{2s}$ and $A_{1p} < A_{2p}$, derived solely on the basis of ns time-scale MD simulations, confirms the idea of alternating affinities between the domains and ssDNA in different states, first proposed by Velankar et al. [19]: in the ATP-bound (s) state domain 2A moves more easily than 1A while in the empty binding site case (state p) the opposite is true. As a result, when ATP binds and the domains 1A and 2A are being pulled together, only 1A moves. When the domains separate again, domain 2A moves. This scenario of alternating 1A and 2A mobilities along DNA results in a unidirectional translocation as one can readily realize by grasping with ones hands (representing domains 1A and 2A) a rope (representing ssDNA) and alternatingly making a tight and a loose fist with the left and right hand while pulling the rope with the tight fist (c.f. Fig. 6b in [19]). The derived potential energy functions $U_{i\sigma}$ (Fig. 9) will be utilized in a Langevin dynamics description presented below, demonstrating indeed uni-directional translocation along ssDNA.

5
Langevin Simulations of PcrA Function

From experimental studies it is known that PcrA translocates along ssDNA uni-directionally at a speed of about 50 nt/s, presumably consuming one ATP per single nucleotide step which, therefore, takes about 20 ms [53]. To describe PcrA function on this time-scale, a computational modeling approach different from QM/MM or MD is required for two major reasons. First, due to the explicit treatment of electronic and/or atomistic details, the time-scales accessible to QM/MM and MD simulations do not reach the long time-scales needed to model physiological PcrA function. Second, on larger length- and longer time-scales, the essential degrees of freedom governing the dynamics of the system are likely quite different and, therefore, largely inaccessible from simulations of smaller scale models. Therefore, in order to properly investigate the relevant mechanism on this functional level one needs to adopt a different approach.

One possible description of the PcrA ssDNA stepping motor replaces the two translocating domains 1A and 2A by two featureless beads [51], which move along ssDNA during each ATP hydrolysis cycle (Fig. 1). This simple view is based on the two basic conformations of the PcrA-DNA complex found in the available X-ray crystal structures [19].

A straightforward approach to characterize the bead movement along ssDNA is by means of a Langevin equation describing Brownian motion in a potential. Let us first suppose that the interaction between the bead and ssDNA is purely associative, i.e., the attachment of the bead to the DNA strand does not interfere with its movement along ssDNA. Under these conditions, the bead motion can be described by a 1-D Langevin equation,

$$m\ddot{x} = -\gamma\dot{x} + \tilde{f}(t) \,. \tag{22}$$

Here γ is the friction coefficient. The term $\tilde{f}(t)$ represents white noise and obeys the fluctuation–dissipation theorem [54]

$$\langle \tilde{f}(t) \rangle = 0 \,; \quad \langle \tilde{f}(t)\tilde{f}(t') \rangle = 2\gamma k_{\mathrm{B}}T\delta(t - t') \,. \tag{23}$$

Since a particle the size of a protein in solution has a very low mass to viscosity ratio (the Reynolds number) one can work in the strong friction limit in which the acceleration term $m\ddot{x}$ in Eq. 22 vanishes.

Next, we take into account the interactions between the protein and ssDNA which prevent the beads from sliding freely along ssDNA. The interactions can be modeled by additional potentials in Eq. 22 which govern the movement of individual beads. These additional potentials are just the $U_{i\sigma}$ in Eq. 20 that were derived from MD simulation and originate from the imbalance in binding energy of individual ssDNA nucleotides to PcrA [51]. Since the two beads, corresponding to domain 1A and 2A, move in a concerted fashion, it may be advantageous to introduce an extra interaction term acting between the two domains facilitating the movement of domains. This interaction potential, when non-vanishing, is modeled by a harmonic spring

$$V_{\sigma'} = \frac{1}{2}k(x_1 - x_2 - l_{\sigma'})^2 \,. \tag{24}$$

Here, σ' (s or p) is a state index varying independently from the σ in $U_{i\sigma}$. The force constant of the spring, k, adopts a value of $1k_{\mathrm{B}}T/\text{Å}^2$, a value empirically estimated from MD simulations [51]. $l_{\sigma'}$ is the equilibrium length of the spring, with $l_p > l_s$, since the two domains 1A and 2A are separated farther in the p state than in the s state. For the sake of simplicity, we use $l_p = l_0$ and $l_s = l_0/3$ with l_0 corresponding to the nucleotide separation in ssDNA (~ 6.5 Å).

Putting all of this together, the Langevin equation in the strong friction limit reads

$$\gamma\dot{x}_i = -\frac{\partial W_{\sigma\sigma'}(x_1, x_2)}{\partial x_i} + \tilde{f}(t) \,, \tag{25}$$

where $W_{\sigma\sigma'}(x_1, x_2) = U_{1\sigma}(x_1) + U_{2\sigma}(x_2) + V_{\sigma'}(x_1, x_2)$, with $x_i, i = 1,2$, being the position of the beads tracking domains 1A and 2A along ssDNA. Equation 25 can be solved numerically assuming discrete time steps of size Δt,

$$x_i(t + \Delta t) = x_i(t) - \frac{1}{\gamma} \frac{\partial W_{\sigma\sigma'}(x_1, x_2)}{\partial x_i} \Delta t + \sqrt{2D \, \Delta t} Z . \tag{26}$$

Here, Z is a standard normal random variable (with mean 0 and variance 1) and D is the diffusion coefficient, according to the fluctuation-dissipation theorem related to the friction coefficient through $D = k_B T/\gamma$. D is chosen to assume a value of $10^4 \, \text{Å}^2/\mu s$, typical for a 3 nm-radius protein in solution [55]. It is important to keep in mind, however, that D can vary significantly in different physiological environments.

Note that Eq. 25 holds in a particular state σ, denoted by s or p when $V_{\sigma'}$ is absent. Transitions between states are either triggered once some geometrical criterion is satisfied, or are simulated through a Poisson process with a specific rate constant. In the latter scenario one generates uniformly distributed random numbers Y (with $Y \in [0, 1]$) and transitions are accepted if $Y \leq \omega\Delta t$ ($\omega\Delta t \ll 1$), where ω is the rate constant for the transition and Δt is the discrete time step in Eq. 26. The rate constants are estimated [51] based on the rate limiting steps assumed in each scenario and the translocation speed observed experimentally [53].

In the following, we will assume two limiting scenarios [51], the weak and strong coupling scenario, expecting that the most realistic model will lie somewhere between the two. We envision that the translocation of PcrA along ssDNA comes about through an inchworm motion involving separate, but coupled translocations of its 1A and 2A domains (described by beads) that are governed by three factors: (1) There exist geometrical constraints that prohibit the domains to pass each other as well as to separate too far. (2) Binding of ATP favors a narrower separation between domains 1A and 2A while unbinding of ADP and P$_i$ favors a wider separation between the domains. (3) Depending on the state (s or p) of PcrA, the domains experience different effective potentials such that in the ATP bound state (s), 2A can move easily (low energy barrier A_{2s}) and 1A can hardly move (high energy barrier A_{1s}) and in the absence of ATP/ADP (p), 1A can move easily (low energy barrier A_{1p}) and 2A can hardly move (high energy barrier A_{2p}). Below we investigate PcrA translocation in the two scenarios, the weak and strong coupling scenario, and demonstrate how (1)–(3) can endow PcrA with unidirectional motion in the $3' \rightarrow 5'$ direction.

Weak Coupling Scenario

Here, domain 2A and 1A move with vanishing mutual interaction potential, i.e. $V_{\sigma'}(x_1, x_2) = 0$. However, the two domains are still coupled by geometrical constrains. We furthermore assume that the rate limiting steps are indi-

vidual domain movements, i.e., thermally activated barrier-crossing events governed by potentials $U_{i\sigma}$. The barrier-crossing time can be estimated by the mean first-passage time [55, 56], which in turn is mostly determined by $A_{i\sigma}$ given in Eq. 21. In order to obtain a translocation speed of about 6.5 Å/20 ms a value for the parameter Δ of –2.5 kcal/mol was used. The corresponding results of Langevin dynamics simulations are shown in Fig. 10a. The left panel depicts the stochastic trajectories describing the motion of 1A and 2A along ssDNA over a period of 10 hydrolysis cycles. The right-side panel shows
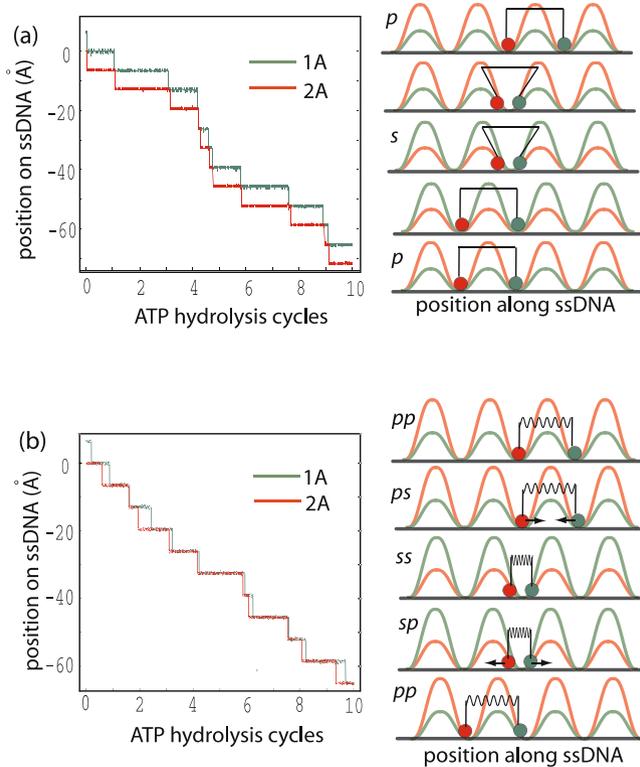
Fig. 10 Langevin simulations of ssDNA translocation in PcrA. Depicted are the trajectories for the weak (**a**) and the strong coupling scenario (**b**). *Left* Shown are trajectories of the two translocating domains, 1A (*green*) and 2A (*red*), moving along ssDNA in each scenario; the time is given in units of ATP hydrolysis cycles (one cycle lasts about 20 ms). *Right* Illustrated are the individual potentials $U_{i\sigma}$ experienced by domain 1A (*green*) and domain 2A (*red*) moving along ssDNA in different states (*p, s* in the weak coupling scenario or *pp, ps, ss,* and *sp* in the strong coupling scenario, defined in [51]) in each scenario. In the weak coupling scenario, two domains are shown as being connected by a rod, corresponding to the geometric constraint; in the strong coupling scenario (**b**), the domains are shown as being connected by an elastic spring with variable equilibrium lengths, corresponding to the non-vanishing interaction potential $V_{\sigma'}$ (adapted from [51])

schematically the sequence of configurations corresponding to the motion. In the p state, the two domains separate first; domain 1A experiences a low barrier (A_{1p}) and can move readily, while 2A experiences a high barrier (A_{2p}) and is essentially "stuck". When 1A has moved forward close to 2A, ATP binds into the cleft between 1A and 2A and the system transits to the s state. In the s state, the potentials U_{is} differ qualitatively from U_{ip}, in that now 2A is ready to move (low barrier A_{2s}) and 1A becomes "stuck" (high barrier A_{1s}). When 2A moves forward far enough, ADP and P$_i$ quickly dissociate and the system transits back to the p state. Overall, after one cycle, PcrA has shifted one nt distance compared to the starting state on the ssDNA.

Strong Coupling Scenario

In this scenario, domain 1A and 2A move under the influence of a non-vanishing interaction potential $V_{\sigma'}(x_1, x_2)$, modeled by a harmonic spring. Hence, the barrier-crossing events of each domain happen under the combined potential of $U_{i\sigma}$ and $V_{\sigma'}$. The states in the strong coupling scenario, accordingly, can be defined through two labels, σ and σ'. This leads to four possible states, namely, $(\sigma\sigma') = (ss), (sp), (ps)$, and (pp) as shown in Fig. 10b. Since $A_{1p} < A_{2p}$ as well as $A_{1s} > A_{2s}$ still holds, the barrier experienced by 1A/2A is lower than that experienced by 2A/1A during ATP binding/ADP + P$_i$ dissociation. Therefore, the unidirectional translocation of PcrA is maintained, as demonstrated in Fig. 10b. The domain movement itself, however, can happen fast in this scenario and we assume that the rate limiting steps are waiting for ATP arrival and ATP hydrolysis; the rate constants of these events can also be fitted to yield a translocation speed of about 6.5 Å/20 ms.

From our analysis, it is not clear which of the two limiting scenarios is the physiologically relevant one for PcrA. Experimental evidence suggests that PcrA may actually work in a mixed scenario, in which ATP binding serves as a power-stroke [57, 58], facilitating the forward movement of domain 1A and assisting PcrA conformational change so that 1A can not move backward, while ATP hydrolysis or ADP + P$_i$ dissociation triggers the thermally activated forward movement of domain 2A (Myong et al., 2006, personal communication), assisting PcrA conformational change to prevent 2A from moving backward.

No matter what the exact scenario is, one can recognize that the unidirectional translocation of PcrA stems from the fact that in the state without ATP bound, domain 1A experiences a lower energy barrier than 2A; while in the ATP-bound state, domain 2A experiences a lower energy barrier than 1A, that is $A_{1p} < A_{2p}$ and $A_{1s} > A_{2s}$. One can imagine that the opposite relationship $A_{1p} > A_{2p}$ and $A_{1s} < A_{2s}$ will lead to helicase translocation in the reversed direction.

Since the above Langevin dynamics simulations of PcrA translocation were conducted with the potentials $U_{i\sigma}$ derived from atomic-scale simulations, and

recognizing that it is the difference in barrier height $|A_{1\sigma} - A_{2\sigma}|$ that directs the unidirectional movement, it is interesting to evaluate the contributions of individual residues to this difference term. This reveals that two key amino acids, R260 and K385, contribute prominently to $A_{1s} - A_{2s}$ and $A_{2p} - A_{1p}$, respectively [51], and are therefore proposed to be most important in affecting the unidirectional translocation of PcrA.

6
Summary

Combining QM/MM calculation, nanosecond MD simulation, and millisecond stochastic modeling, the ATP hydrolysis coupled ssDNA translocation of a monomeric PcrA helicase has been investigated.

Figure 11 summarizes the combined results and depicts a single PcrA translocation cycle involving five steps. In configuration 1, domains 1A and 2A are separated without ATP or ADP bound. The arrangement of the bases of the ssDNA are shown in the insert to the left, revealing a single base intercalated between side groups of Y257 and F64. Initially, domain 1A can move more easily along ssDNA than domain 2A. At this point ATP arrives at the binding site and initiates binding. The binding of ATP causes an attractive force between the domains which consequently start approaching each other through motion of domain 1A, leading to configuration 3. The close approach of domains 1A and 2A pushes Y257 and F64 together and, thereby, squeezes the ssDNA base out, it moving from its former position to a new position as indicated by the arrows in the inserted figures. The close approach of 2A and 1A brings R610 and R287 into contact with the γ-phosphate of ATP; the movement of Y257 is transmitted to Q254. QM/MM calculations have shown that the three mentioned side groups, R610, R287, and Q254 control the hydrolysis of ATP in PcrA [36]. In configuration 3 the juxtaposition of the three side groups is optimal for hydrolysis of ATP to ADP + P_i, leading to configuration 4. At this point the mobility of domains 2A and 1A has reversed, 2A now being able to move more easily. Thermal fluctuations lead to the separation of 2A from 1A giving rise to configuration 5, and permitting release of ADP and P_i, reaching configuration 1′. PcrA has advanced by one step and is ready for the next cycle.

In summary, we have presented a comprehensive study of PcrA helicase function covering length scales ranging from the electronic domain using QM/MM simulations to the molecular level via MD and long time stochastic simulations. The time scales covered span several orders of magnitude from fast femtosecond time scale bond-breaking events to physiological protein motions on the millisecond time scale. The combination of methods employed in our computational modeling revealed a detailed picture of PcrA function, for example, a proton relay mechanism responsible for efficient ATP
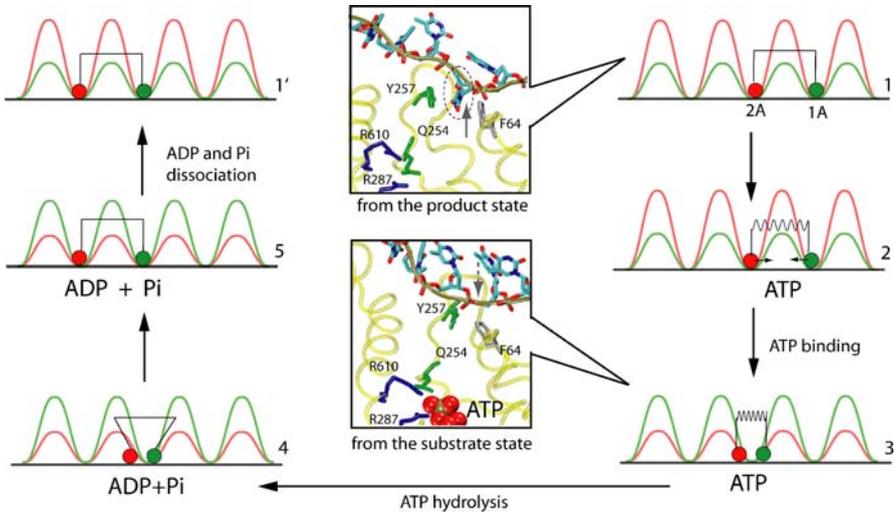
Fig. 11 Five-step PcrA translocation cycle. The figure shows schematically a translocation cycle for a mixed (see text) scenario involving five configurations 1, 2, 3, 4, 5 as well as configuration 1′ that is equivalent to configuration 1, except that the system has advanced by one base. The mixed scenario involves both a loaded spring (non-zero potential $V_\sigma(x_1, x_2)$, configuration 2 → 3) and a step with a random thermal motion (vanishing potential $V_\sigma(x_1, x_2)$, configuration 4 → 5). In configurations 1, 2, 1′ domain 1A (*green*) moves more readily than domain 2A (*red*), while it is the opposite for configurations 3, 4, 5. The *insert figures* show how the domain and ssDNA base motions are coupled to the chemistry at the ATP binding site: upon the approach of domains 2A and 1A, R287 and R610 move close to the γ-phosphate of ATP; Q254 is linked closely to Y257 that forms a key binding pocket for an ssDNA base, which is squeezed out when the domains approach each other in binding ATP; Q254 was identified as a key participant in ATP hydrolysis along with the mentioned arginines [36]. The suggested mechanism involves the binding of ATP that pulls domain 1A towards domain 2A; insertion of R287 and R610 into an optimal (for hydrolysis) position in the ATP binding pocket along with Q254 linked to a key ssDNA interaction site; rapid hydrolysis of ATP that initiates separation of domains 2A and 1A through movement of 2A alone (adapted from [51])

hydrolysis in the catalytic site and the means by which uni-directional motion along ssDNA is achieved, namely by a ratchet mechanism. Employing a similar strategy to investigate other molecular motors can be expected to reveal more about the mechanism by which these remarkable nano-machines perform their physiological function.

References

1. Li G, Cui Q (2004) J Phys Chem B 108:3342–3357
2. Minehard TJ, Marzari N, Cooke R, Pate E, Kollman PA, Car R (2002) Biophys J 82:660–675
3. Zheng W, Brooks BR (2005) Biophys J 89:167–178
4. Fischer S, Windshügel B, Horak D, Holmes KC, Smith JC (2005) Proc Natl Acad Sci USA 102:6873–6878
5. Kawakubo T, Okada O, Minami T (2005) Biophys Chem, 115:77–85
6. Wriggers W, Schulten K (1999) Proteins: Struct Func Gen 35:262–273
7. Oster G, Wang H (2000) Biochim Biophys Acta 1458:482–510
8. Böckmann RA, Grubmüller H (2002) Nature Struct Biol 9:198–202
9. Ma J, Flynn TC, Cui Q, Leslie AGW, Walker JE, Karplus M (2002) Structure 10:921–931
10. Dittrich M, Hayashi S, Schulten K (2003) Biophys J 85:2253–2266
11. Strajbl M, Shurki A, Warshel A (2003) Proc Natl Acad Sci USA 100:14834–14839
12. Dittrich M, Hayashi S, Schulten K (2004) Biophys J 87:2954–2967
13. Cox K, Watson T, Soultanas P, Hirst J (2003) Proteins: Struct Func Gen 52:254–262
14. Betterton M, Julicher F (2003) Phys Rev Lett 91:258103
15. Betterton M, Julicher F (2005) Phys Rev E 71:011904
16. Bhattacharjee S, Seno F (2003) J Phys A: Math Gen 36:L181-L187
17. Bhattacharjee S (2004) Europhys Lett 65:574–580
18. Subramanya H, Bird L, Brannigan J, Wigley D (1996) Nature 384:379–383
19. Velankar S, Soultanas P, Dillingham M, Subramanya H, Wigley D (1999) Cell 97:75–84
20. Wong I, Lohman TM (1992) Science 256:250–256
21. Stasiak A, Tsaneva IR, West SC, Benson CJ, Yu X, Egelman EH (1994) Proc Natl Acad Sci USA 91:7618–7622
22. Warshel A (1976) Nature 260:679–683
23. Singh UC, Kollman PA (1986) J Comp Chem 7:718–730
24. Field MJ, Bash PA, Karplus M (1990) J Comp Chem 11:700–733
25. Stanley RJ, Boxer SG (1995) J Phys Chem 99:859–863
26. Maseras F, Morokuma K (1995) J Comp Chem 16:1170–1179
27. Bakowies D, Thiel W (1996) J Phys Chem 100:10580–10594
28. Lyne PD, Hodoscek M, Karplus M (1999) J Phys Chem A 103:3462–3471
29. Lyne D, Walsh O (2001) Computer Simulation of Biochemical Reactions with QM/MM Methods. In: Becker O, Mac Kerell A Jr, Roux B, Watanabe M (eds) Computational Biochemistry and Biophysics. Marcel Dekker Inc., New York
30. Warshel A, Levitt M (1976) J Mol Biol 103:227–249
31. Théry V, Rinaldi D, Rivail J-L, Maigret B, Ferenczy GG (1994) J Comp Chem 15:269–282
32. Bayly C, Cieplak P, Cornell W, Kollman P (1993) J Phys Chem 100:10269–10280
33. Hayashi S, Ohmine I (2000) J Phys Chem B 104:10678–10691
34. Warshel A, Weiss RM (1980) J Am Chem Soc 102:6218
35. Warshel A (2003) Ann Rev Biophys Biomol Struct 32:425–443
36. Dittrich M, Schulten K (2006) Structure 14:1345–1353
37. Scheidig AJ, Burmester C, Goody RS (1999) Structure 7:1311–1324
38. Ahmad Z, Senior AE (2004) J Biol Chem 44:46057–46064
39. Dittrich M, Schulten K (2005) J Bioener Biomemb 37:441–444
40. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M (1983) J Comp Chem 4:187–217

41. Lindahl E, Hess B, van der Spoel D (2001) J Mol Mod 7:306–317
42. Pearlman DA, Case DA, Caldwell JW, Ross WS, Cheatham T, Debolt S, Ferguson D, Seibel G, Kollman P (1995) Comput Phys Commun 91:1–41
43. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K (2005) J Comp Chem 26:1781–1802
44. Freddolino PL, Arkhipov AS, Larson SB, Mc Pherson A, Schulten K (2006) Structure 14:437–449
45. Berendsen HJC, Postma JPM, van Gunsteren WF, Di Nola A, Haak JR (1984) J Chem Phys 81:3684–3690
46. Martyna GJ, Tobias DJ, Klein ML (1994) J Chem Phys 101:4177–4189
47. Feller SE, Zhang YH, Pastor RW, Brooks BR (1995) J Chem Phys 103:4613–4621
48. Izrailev S, Stepaniants S, Isralewitz B, Kosztin D, Lu H, Molnar F, Wriggers W, Schulten K (1998) Steered Molecular Dynamics. In: Deuflhard P, Hermans J, Leimkuhler B, Mark AE, Reich S, Skeel RD (eds) Computational Molecular Dynamics: Challenges, Methods, Ideas. Vol 4. Springer, Berlin Heidelberg New York
49. Isralewitz B, Gao M, Schulten K (2001) Curr Op Struct Biol 11:224–230
50. Torrie, Valleau (1977) J Comp Phys 23:187–199
51. Yu J, Ha T, Schulten K (2006) Biophys J 91:2097–2114
52. Åqvist J (1996) J Comp Chem 17:1587–1597
53. Dillingham M, Wigley D, Webb M (2000) Biochemistry 39:205–212
54. Kubo R (1966) Rep Progr Phys 29:255–284
55. Howard J (2001) Mechanics of Motor Proteins and the Cytoskeleton. Sinauer Associates, Inc., Sunderland, MA
56. Szabo A, Schulten K, Schulten Z (1980) J Chem Phys 72:4350–4357
57. Nishizaka T, Oiwa K, Noji H, Kimura S, Muneyuki E, Yoshida M, Kinosita K Jr (2004) Nat Struct Mol Biol 11:142–148
58. Shimabukuro K, Yasuda R, Muneyuki E, Hara KY, Kinosita K Jr, Yoshida M (2003) Proc Natl Acad Sci USA 100:14731–14736
59. Humphrey W, Dalke A, Schulten K (1996) J Mol Graphics 14:33–38