Structure-Based Model of the Stepping Motor of PcrA Helicase

Jin Yu,† Taekjip Ha,†‡ and Klaus Schulten†‡

*Beckman Institute, †Department of Physics, and ‡Howard Hughes Medical Institute, University of Illinois at Urbana-Champaign, Urbana, Illinois

ABSTRACT DNA helicases are ubiquitous molecular motors involved in cellular DNA metabolism. They move along single-stranded DNA (ssDNA) and separate duplex DNA into its component strands, utilizing the free energy from ATP hydrolysis. The PcrA helicase from Bacillus stearothermophilus translocates as a monomer progressively from the 3’ end to the 5’ end of ssDNA and is one of the smallest motor proteins structurally known in full atomic detail. Using high-resolution crystal structures of the PcrA-DNA complex, we performed nanosecond molecular dynamics simulations and derived potential energy profiles governing individual domain movement of the PcrA helicase along ssDNA. Based on these profiles, the millisecond translocation of the helicase along ssDNA was described through Langevin dynamics. The calculations support a domain stepping mechanism of PcrA helicase, in which, during one ATP hydrolysis cycle, the pulling together and pushing apart of domains 2A and 1A are synchronized with alternating mobilities of the individual domains in such a fashion that PcrA moves unidirectionally along ssDNA. By combining short timescale (nanoseconds) molecular dynamics and long timescale (milliseconds) stochastic-dynamics descriptions, our study suggests a structure-based mechanism of the ATP-powered unidirectional movement of PcrA helicase.

INTRODUCTION DNA helicases are ubiquitous motor proteins which separate duplex DNA into their component strands using energy released from ATP hydrolysis (1–5). The helicases are involved in almost all aspects of DNA metabolism, including transcription, replication, and recombination. Defects in helicase functioning in humans can lead to genomic instability and predisposition to cancer (6). To achieve their functions, helicases move in a unidirectional manner along single-stranded DNA (ssDNA) (7); when a helicase continues its translocation along ssDNA encountering a junction formed by duplex DNA, the duplex DNA becomes unwind.

In their functional forms, helicases assemble as hexamers, tetramers, dimers, or monomers (1,3,4). PcrA helicase from Bacillus stearothermophilus (B. stearothermophilus) has been proposed to work as a monomer (8). Belonging to the superfamily 1 (SF1) helicases (5,9), monomeric PcrA (~80 kDa) is composed of four domains (1A, 2A, 1B, and 2B), resembling other SF1 helicases in their monomeric forms, e.g., Rep and UvrD, although Rep and UvrD are oligomers in their functional forms (10–12). The PcrA, Rep, and UvrD monomers exhibit ~40% sequence identity, and all have been demonstrated in experiments as being capable of translocating progressively 3’ to 5’ on ssDNA (13–16).

PcrA helicase from B. stearothermophilus has been crystallized and resolved at high resolution (8), in both a substrate (with ATP bound) and a product (without ATP/ADP bound) state. The structures, as shown in Fig. 1, were crystallized in the presence of a DNA junction, i.e., duplex DNA flanked by a piece of 3’ ssDNA, with the duplex bound to the side of domain 2B and an elongated piece of ssDNA crossing above the two RecA-like domains 2A and 1A. Domains 1A and 2A, conserved among a class of helicase-like proteins (2,9), are proposed to play the most essential role in the translocation. ATP binds into a cleft between domains 2A and 1A, the binding site being lined by amino acids that are highly conserved among SF1 helicases (5,9). Two of the motifs in the ATP binding site (Walker A and Walker B) are highly conserved among all ATPases. Indeed, superimposing the ATP binding pocket of PcrA helicase with that of F1-ATPase shows that the ATP binding sites have high structural identity, suggesting that a closely related ATP hydrolysis mechanism may be at work (17).

The study reported in this article seeks to identify through a computational modeling approach the molecular mechanism underlying the fundamental function of helicase, the ATP hydrolysis powered unidirectional translocation along an ssDNA track. Such an approach was employed in previous studies, for example, in Chennubhotla et al. (18), Aksimentiev et al. (19), Ma et al. (20), and Wang and Oster (21) for molecular machines. PcrA helicase is selected here, since it is one of the smallest linear motors with full atomic scale structures available as well as experimental information on velocity and step size. The helicase system also involves protein-DNA interaction and recognition in a very confined space. Understanding the basic mechanism of PcrA helicase may facilitate understanding of more complex molecular motors.

Prior theoretical work investigated helicase function in a generic framework employing certain mathematical models that can describe helicase unwinding duplex DNA (22–24). This work proved very useful for this study, yet it postulated ad hoc the fundamental steps in helicase motor
function, namely, different forward and backward translocation rates. The authors also addressed only generic helicases, not any particular one. In contrast, the present work seeks to establish the helicase motor mechanism from the structural and physical properties of a particular helicase, PcrA, the relevant physical properties being established through molecular dynamics (MD) simulations.

We based our study on the crystallographic structures reported in Velankar et al. (8) and mentioned above. We assume that the two structures, one with an ATP (analog) bound and one without ATP/ADP, present key states lying along the pathway of the PcrA translocation process. This is by no means certain since artifacts, in particular due to crystal packing, might shift the protein away from its mechanistically relevant conformations. The goal of our study was to identify through molecular dynamics simulations the mechanism of PcrA translocation. Two main translocation models had been discussed in the literature, an active rolling model, suggested actually for a dimeric helicase (25), and an inchworm model (26). As long as PcrA translocates as a monomer, the inchworm model is presently the only candidate. Our study assumes that PcrA works as a monomer and, hence, it focuses only on the inchworm model. This model was also proposed by the crystallographers who solved the structure of PcrA (8). According to the model, PcrA moves by alternating affinities between its translocation domains (2A and 1A) and ssDNA. So far, however, the inchworm model, while eminently insightful, was based on intuition, i.e., it is mainly qualitative and does not result from quantitative physical properties of PcrA derived from its crystallographic structures. Our study seeks to provide the missing physical basis for the inchworm model.

Experimental data showed that the translocation speed of PcrA along ssDNA is \( \sim 50 \) nucleotides (nt) per second, presumably consuming one ATP for 1-nt distance (13). Based on this information, we propose a schematic model (see bottom left panels of Fig. 1, a and b), in which the substrate state exhibits lower energy barriers for domain movement of 2A along ssDNA (red curve) and higher ones for 1A (green curve), while the product state exhibits lower energy barriers for domain movement of 1A along ssDNA and higher ones for 2A; coupling these changes in mobilities to attraction (upon ATP binding) and repulsion (upon ADP + Pi dissociation) between the two domains leads to directed translocation. We use in this figure and other figures the one-letter code for amino acids.
translocation. This model followed from the molecular dynamics (MD) simulations reported here, from which we derived the potential for individual domain (2A and 1A) motions along ssDNA as shown in Fig. 1. For this model we develop then a stochastic dynamics description of PcrA translocation. Our study suggests that the unidirectional translocation (3' to 5') of PcrA is a direct consequence of alternating high and low energy barriers experienced by 2A and 1A during each ATP hydrolysis cycle. Our work provides microscopic justification for the alternating affinities between domains and ssDNA proposed by the inchworm model (8).

In the following, we first introduce in Methods the stochastic model for the millisecond domain movements of PcrA. Then we describe a scheme to determine from MD simulations the potentials that govern the domain movements. The potentials obtained from MD simulations are provided in Results, followed by the demonstration that the potentials coupled to the ATP hydrolysis cycle lead to unidirectional movement. We finally identify the key amino-acid residues coordinating the unidirectional translocation in PcrA helicase, and provide further evidence for dynamic asymmetries of PcrA domains with bound ssDNA.

**METHODS**

Naturally, one would like to simulate the processive motion of PcrA along ssDNA entirely by MD simulations. Such simulations should employ nothing but the crystallographic structure of PcrA, heuristic information on atomic level interactions of biopolymers, and the laws of classical mechanics, i.e., the approach taken should be unbiased in regard to the translocation mechanism. Unfortunately, such an approach is computationally too demanding since MD simulations can cover at best microsecond timescales, i.e., they are at least a factor-1000 too slow for the present problem. To investigate the mechanism of PcrA we take a different route, approaching the description of PcrA from the short timescale by means of MD and from the long timescale by means of stochastic dynamics, both approaches being specified below. The stochastic dynamics method is based partially on results from short time MD simulations, partially on observed properties like ATP hydrolysis rate or PcrA translocation speed. This permits one to bridge the time gap between the MD description and the actual function of PcrA. In the following, we outline first the long-time stochastic dynamics approach and subsequently the short-time MD approach.

**Stochastic dynamics description of PcrA translocation**

Here we seek to describe the translocation of PcrA along ssDNA by means of stochastic dynamics theory. For this purpose, we will assume two limiting scenarios, expecting that the most realistic model falls between the two limits. We envision that the sliding of PcrA along ssDNA comes about through an inchworm motion involving separate, but coupled translocations of its 2A and 1A domains as suggested in Velankar et al. (8). Three factors govern the linked motion of domains 2A and 1A:

1. There exist geometrical constraints that forbid the domains to pass each other as well as to separate too far.
2. Binding of ATP favors a narrower separation between domains 2A and 1A while unbinding of ADP favors a wider separation between the domains (as revealed from the crystallographic structures).
3. Depending on the state of PcrA (substrate/product ρ) the domains experience different effective potentials characterizing the energetics of individual domains translocating along ssDNA, e.g., in the ATP bound state (ρ) 2A can glide easily (low energy barriers) and 1A can hardly glide (high energy barriers).

Below we introduce two scenarios that demonstrate how 1–3 can endow PcrA with unidirectional motion.

The motion of domains 2A and 1A translocating along ssDNA (coordinates x1 and x2, respectively) is a stochastic process, that, on the relevant timescale, can be described by means of a Langevin equation in the strong friction limit (27,28),

\[
\gamma \dot{x}_i = -\frac{\partial W(x_1, x_2)}{\partial x_i} + \xi.
\]

Equation 1 holds in a particular state of PcrA, e.g., the s or p state. Here γ is the friction coefficient; f is the fluctuating force, represented through a so-called Gaussian white noise (29,30); W(x1, x2) is the potential governing the movement of domains 1A and 2A along ssDNA; x1, i = 1, 2, are defined through

\[
x_1 = x_1^{CM} - L_d/2,
\]

\[
x_2 = x_2^{CM} + L_d/2,
\]

\[
L_d = (\langle x_1^{CM} - x_2^{CM} \rangle)^p - \xi,
\]

where x1CM and x2CM are the coordinates of the centers of mass of domains 1A and 2A along ssDNA, \(\langle \cdot \rangle_p\) denotes the average in the p state, and \(\xi\) is defined as 1-nt distance (~6.5 Å). We note that \(-\dot{x}\) defines the forward, i.e., toward the DNA junction, direction. The Langevin equation, Eq. 1, ignores possible memory effects; at this exploratory stage of the investigation the neglect of memory effects seems to be justified. The approach, of course, would be wrong for extreme memory effects, e.g., if translocation in step 1, 3, . . . follows a different mechanism from translocation in steps 2, 4, . . . The dynamics adopted ignores also hydrodynamic effects of the domain movements and translocation.

We define potentials, \(U_s(x_1, x_2)\), governing individual (i = 1, 2 for 1A and 2A, respectively) domain motions in the s and the p states (s = s, p); the interaction potential between the two domains is written \(V_s(x_1, x_2)\) (s = s, p). The values s and \(s'\) are independent indices, i.e., below we will combine \(U_s\) and \(V_s\) with different indices (states) s, s'. The value \(W(x_1, x_2)\) in Eq. 1, labeled by indices ss', is then decomposed into three contributions:

\[
W_{ss'}(x_1, x_2) = U_{1s}(x_1) + U_{2s}(x_2) + V_{ss'}(x_1, x_2).
\]

The individual potentials \(U_s(x_1, x_2)\) are derived from MD simulations (described below) and are shown in insets of Fig. 2.

In the first scenario, referred to as the weak coupling case, domains 2A and 1A move without interaction, i.e., we set \(V_{ss'}(x_1, x_2) = 0\). Therefore, there are only two independent Langevin equations (Eq. 1) for the s and p states, respectively. Geometrical constrains (1) (see above) still apply, e.g., it holds that \(0 < x_1 - x_2 \leq \xi\); the condition is satisfied through reflection boundaries. Starting from the p state, once 1A gets close enough to 2A, e.g., \(x_1 - x_2 < \xi/3\), PcrA can change rapidly to the s state (ATP bound); starting from the s state, once 2A moves far enough from 1A, e.g., \(x_1 - x_2 > 2\xi/3\), PcrA can change back rapidly to the p state (no ATP/ADP bound).

In the second scenario, referred to as the strong coupling case, the domains are pulled together when ATP binds and then are pushed apart when ADP and Pi unbind by means of the interaction potential \(V_{ss'}(x_1, x_2)\). This potential is modeled through

\[
V_{ss'}(x_1, x_2) = \frac{1}{2}K(x_1 - x_2 - \xi_{ss'}^2),
\]

where the force constant K adopts a value of \(1 \text{ k_BT/Å}^2\), empirically determined from MD simulations; the equilibrium length \(\xi_{ss'}\) for \(s' = s, p\) is chosen as \(\xi_p = \xi\) and \(\xi < \xi_p\), e.g., \(\xi_{ss'} = \xi_{ss').\) The form of the potential in Eq. 4

---

*Biophysical Journal 91(6) 2097–2114*
is the simplest choice and has been adopted due to lack of more detailed information. To improve the description, one might want to consider subunit distances for both crystal structures, but such sampling is presently unfeasible given the expected microsecond-to-millisecond timescale of the domain motion.

As mentioned above, \( \sigma' \) (in \( V_{\sigma'} \)) and \( \sigma \) (in \( U_{\sigma} \)) are independent indices; hence combinations of them (in \( W_{\sigma' \sigma} \)) can represent four different states. In the equilibrium substrate (or product) state \( \sigma = \sigma' = s \) (or \( p \)), which will be labeled \( ss \) (or \( pp \)). For \( \sigma = s \) and \( \sigma' = p \), the system is in transition from the equilibrium substrate to the product state. We will refer to the intermediate state as \( sp \); similarly we call the intermediate state from the substrate state to the product state \( ps \) (\( \sigma = p \) and \( \sigma' = s \)).

Correspondingly, there are four independent Langevin equations (Eq. 1) for these four states.

Transition from the \( pp \) state to the \( ps \) state is triggered by arrival of ATP and transition from the \( ss \) state to the \( sp \) state is triggered by the formation of ADP+Pi (through ATP hydrolysis). The two transitions happen at certain rates as specified below. For transitions from the \( ps \) or \( sp \) state to the equilibrium \( ss \) or \( pp \) state to happen, geometrical criteria were applied, such that \( ps \) or \( sp \) to \( ps \) as \( x_1-x_2 \) shrinks below \( l \), while \( sp \) or \( ps \) to \( pp \) as \( x_1-x_2 \) elongates beyond \( l \).

The Langevin equation (Eq. 1) can be solved numerically (28) when one assumes discrete time-steps \( \Delta t \), adopting the scheme

\[
x_i(t + \Delta t) = x_i(t) - \frac{1}{\gamma} \frac{\partial W_{\sigma \sigma'}}{\partial x_i} \Delta t + \sqrt{2D \Delta t} \mathcal{Z}.
\]

Here \( Z \) is a normal (Gaussian) random variable (29) (with mean 0 and variance 1); \( D \) is the diffusion coefficient, according to the fluctuation dissipation theorem (30) related to the friction coefficient \( \gamma \) through \( D = k_B T/\gamma \); and \( \Delta t \) is expected to assume a value \( \sim 10^4 \Delta t/\mu s \) (corresponding to a typical diffusion coefficient of a 3 nm-radius protein in solution (27)). Coupling between equations is achieved through a random process which we describe now.

In case of weak coupling, the transitions between states \( s \) and \( p \) (right after the domain motion) are assumed to be fast compared to the domain motion described by Eq. 1. Once \( x_1-x_2 \) satisfies the criteria for state transitions mentioned above, transitions are induced through a Poisson process using rate constants specified below. In the strong coupling case, the transitions between states \( ss \) and \( sp \) or between states \( pp \) and \( ps \) are assumed to be slow compared to the stochastic motion in Eq. 1.

The stochastic method adopted is related to the so-called kinetic and dynamic Monte Carlo scheme widely adopted in physics and chemistry (31–34). The simple choice of reaction coordinate \( x \) (path along ssDNA) could be improved by determining a reaction coordinate accounting for the forward reptation of ssDNA in PcrA by means of a so-called reaction path method (for a review see (35)), e.g., the ones suggested in Elber (36) and Straub (37).

### Nucleotide binding site energies and the site energy function

A key task in our study is the derivation of the potential \( U_{\sigma}(x_i) \) in Eq. 1, governing the movement of domains 1A and 2A. Since the translocation of helicase along ssDNA arises through ssDNA units (nucleotides) binding and unbinding sequentially to the domain surfaces, one strategy for determining the potential is to calculate the binding free energies \( E_B \) for individual nucleotide binding sites and then use these energies to obtain \( U_{\sigma}(x_i) \). The calculation of the binding free energy from MD simulations can be achieved by a method (38,39) which is based on 1), on the linear response approximation for electrostatic forces; and 2), on linear scaling between solvation energies and average van der Waals (vdW) energies. The method evaluates the absolute free energy solely from the difference between the average interaction energy between nucleotide and protein+water+ion (bound state) and the average interaction energy between nucleotide and water+ion (free state). The value \( E_B \) for a given binding site located at \( x_i \) is given as a weighted sum of the electrostatic and the vdW interaction energy.

\[
E_B(x_i) = \alpha \Delta E^{ele}(x_i) + \beta \Delta E^{vdW}(x_i)
\]

with weight coefficients \( \alpha \) and \( \beta \) discussed further below. Here \( \Delta E \) denotes the difference between average energies (stemming from simulations) of the bound and the free states, i.e.,

\[
\Delta E^{ele/vdW}(x_i) = E^{ele/vdW}(bound, x_i) - E^{ele/vdW}(free, x_i).
\]

In the present case, we will need only relative energies, not absolute ones, e.g., \( E_B(x_i) - E_B(x_j) \). Since, in the free state (nucleotides with water+ion, but without protein), the nucleotide-solvent interaction energies \( E^{vdW}(free, x_i) \) should not vary along ssDNA (poly-thymine), i.e., be independent of \( x \), it is not necessary to actually calculate \( E^{vdW}(free, x_i) \) and we set these energies to zero. We introduce a further approximation in evaluating \( E^{ele/vdW}(bound, x_i) \). For this purpose we split the energy terms as follows:
\[ E_{\text{ele}/vdW}^{\text{bound}}(x_i) = E_{\text{nucleotide-protein}}^{\text{bound}}(x_i) + E_{\text{nucleotide-water+ion}}^{\text{bound}}(x_i). \]  

For nucleotides buried inside PcrA, we assume that the contributions of nucleotides 19 and 20 (and those beyond 20) remain in the same binding description. One may suggest to replace cutoff distance of 60 Å by a continuous site energy function of corresponding nucleotides (numbered in the same way as in Fig. 1). The energies \( E_{\text{ele}} \) and \( E_{\text{vdW}} \) were calculated (with a cutoff distance of 60 Å) from MD trajectories, sampling every 100 ps and averaging over 2 ns, beginning after the first 1 ns of equilibration.

The weighted sums of \( E_{\text{ele}} \) and \( E_{\text{vdW}} \) for individual ssDNA nucleotide binding sites, calculated according to Eq. 6, are presented in Fig. 2 by site energy differences connected with moving nucleotides 15, 16, 17 and 18 (and those beyond 15, which along with 15 are assumed to be of equal energy) backward, as indicated by colored arrows in Fig. 2. This energy is

\[ U_2(\Delta x) = \sum_{i=15}^{18} [E_0(x_i + \Delta x) - E_0(x_i)] \quad \Delta x \in [0, 1/2]. \]  

In our description we assume that \( U_2(\Delta x) \) adopts a symmetrical form around \( \Delta x = 1/2 \), the position of an energy barrier separating the states before \( (\Delta x = 0) \) and after \( (\Delta x = 1) \) an ATP hydrolysis half-cycle. This assumption also applies to other \( U_p(\Delta x) \) derived below.

Now we can determine the boundary energy at site \( s_{ix} \). The emergence of the two new sites at \( \Delta x = 1 \) (as the system moves to the p state), \( A \) and \( B \), arises effectively from the disappearance of the middle site at \( s_{ix} \) and the boundary site at \( s_{ix} \); therefore, it holds that \( U_{iA}(1) = E_0^{A} + E_0^{B} - E_0(x_{i}) \), with \( E_0^{A} \) and \( E_0^{B} \) being the (unknown) binding energies at sites \( A \) and \( B \), respectively. Since \( A \) and \( B \) are sites close to \( x_{i+1} \) and \( x_{i-1} \), \( E_0^{A} + E_0^{B} \) might be approximated by \( E_0(x_{i+1}) + E_0(x_{i-1}); \) the periodicity condition \( U_{iA}(0) = U_{iA}(1) = 0 \) thus yields \( E_0(x_{i+1}) = E_0(x_{i-1}) \).

One can apply a similar reasoning to the backward (to the right in Fig. 1 a) motion of domain 1A and derive \( E_0(x_{i+1}) = E_0(x_{i-1}) \). In this case, nucleotides 16 and 17 (and those to the left beyond 16) will occupy their original binding sites while nucleotide 21 moves toward 20, 20 toward 19, 19 toward the new site \( B \), and 18 toward the new site \( A \). Thus, one can conclude

\[ U_{iB}(\Delta x) = \sum_{i=16}^{21} [E_0(x_i - \Delta x) - E_0(x_i)] \quad \Delta x \in [0, 1/2]. \]  

Similarly, we consider the energies for the p state, i.e., \( U_{iP} \). When domain 2A moves backward (to the right in Fig. 1 b), nucleotides 19, 20... remain at the same binding sites while nucleotide 15 moves toward 14, 16 toward 15, 17 toward 16, and 18 toward the new site \( C \) (which was occupied by 18 in the s state, see Fig. 1 a). Following the reasoning for the s state above, the energy \( U_{iP}(\Delta x) \) is

\[ U_{iP}(\Delta x) = \sum_{i=15}^{18} [E_0(x_i - \Delta x) - E_0(x_i)] \quad \Delta x \in [0, 1/2]. \]  

The periodicity condition \( U_{iP}(0) = U_{iP}(1) = 0 \) stipulates \( E_0(x_{i+1}) = E_0(x_{i-1}) + E_0(x_i) - E_0^{C} \). With \( E_0^{C} \) approximated by \( E_0(x_{i+1}), \) this gives \( E_0(x_{i}) = E_0(x_{i+1}) \). For a similar reason this condition applies also to the case that domain 1A moves forward with its energetics described by \( U_{iP} \). In that case, nucleotides 15, 16 occupy their original binding sites while nucleotide 17 moves toward the new site \( C \), 18 toward 19, 19 toward 20, and 20 toward 21. The energy is

\[ U_{iP}(\Delta x) = \sum_{i=17}^{20} [E_0(x_i + \Delta x) - E_0(x_i)] \quad \Delta x \in [0, 1/2]. \]  

Combining the definitions of \( U_{iP}(\Delta x) \) above, i.e., Eqs. 9–12, it can be recognized that each barrier \( U_{iP} \), defined as \( U_{iP}(\Delta x) \) at \( \Delta x = 1/2 \), can be written

\[ A_{iP} = \sum_{iP} + 4\delta, \]  

where \( \sum_{iP} \) is a combination of site energy terms, stated explicitly in Eq. 14. The site energies terms are determined from MD simulations as explained above. We add here a term \( \delta \) that controls the speed of PcrA translocation, i.e., ~1 nt per 20 ms, consistent with observation (13). The relationship of \( \delta \), which is chosen identical for both domains \( (i = 1, 2) \) and for both states \( (s, p) \), to the energy surface roughness (energy undulations) of \( E_0(x) \), is stated in Appendix A.

We emphasize that the construction and later use of the potentials \( U_{iP}(x) \) hinges on particular motions of ssDNA along the domains as specified above. If the actual translocation would involve different behavior of ssDNA, the
description adopted here would be called into question. The results of our study should be interpreted as proving the feasibility of the inchworm model once it is assumed.

**Molecular dynamics simulations**

Starting from the crystal structures of the PcrA helicase in the s (PDB code 3PJR) and p (PDB code 2PJR) state (8), we added missing residues in the protein as well as elongated both duplex DNA and ssDNA (poly-T). The ssDNA was elongated by first opening, i.e., swiveling by 30°, the 2B domain (45), adding the corresponding piece of ssDNA (five nucleotides) from the structurally homologous Rep complex (PDB code: 1UAA) (45) to the end of the original ssDNA, then swiveling the 2B domain back to its initial position by means of SMD (42), and letting the elongated ssDNA segment relax to fit to the PcrA domains.

The structures of the PcrA-DNA complex were solvated in a box of explicit water. Sodium, magnesium, and chloride ions were added to neutralize the negative charges of the PcrA-DNA complex and to control the ionic strength (0.1 M). The program Delphi (46) was used to locate positions of minimal electrostatic energies around the complex, where the water molecules were then replaced by ions. The whole simulated system of protein, DNA, water, and ions contained ~110,000 atoms.

All simulations used the program NAMD2 (47), the CHARMM27 force field (48), with an integration time step of 1 fs, and periodic boundary conditions. VdW energies were calculated using a smooth (10–12 Å) cutoff. The particle-mesh Ewald method (49) was employed for full electrostatics, with the density of grid points at least 1/Å in all cases. The particle-mesh Ewald electrostatic forces were computed every four time steps. The simulations were performed in the NpT ensemble, using the Nosé-Hoover Langevin piston method (50,51) for pressure control (1 atm), with an oscillation period of 100 fs and a damping time of 50 fs; Langevin forces (52) were applied to all backbone atoms of protein and DNA during each equilibration are shown in Fig. 3, demonstrating that the system was more or less stabilized after 1 ns.

The equilibrated configuration of double-stranded DNA is more regular and less distorted by the helicase in the s state than that in the p state (see Fig. 1), as evidenced by fewer contacts between double-stranded DNA and domain 2B in the s state. The binding between the translocation domains and ssDNA is illustrated in enlarged views in Fig. 1. One can recognize that nucleotide 18 points its base upward in the s state (Fig. 1 a) and downward into a binding pocket in the p state (Fig. 1 b). The pocket, formed by the side chains of Phe-64 and Tyr-257, is closed in the s state and open in the p state.

By monitoring an angle formed by two protein helices, residues 360–374 from 2A and residues 66–81 from 1A, the density of grid points at least 1/Å, the particle-mesh Ewald method (49) was employed for full electrostatics, with motions, i.e., heavy atoms for temperature control (310 K) with a coupling coefficient of 5/ps. 100 fs and a damping time of 50 fs; Langevin forces (52) were applied to all atoms for temperature control (310 K) with a coupling coefficient of 5/ps.

**RESULTS AND DISCUSSION**

**Equilibration by MD simulation**

The simulated system of the solvated PcrA-DNA complex is shown in Fig. 3 a. Each MD equilibration (substrate, s, and product, p), started after 5000 steps of energy minimization and lasted for ~3 ns. The root mean-square deviation curves for backbone atoms of protein and DNA during each equilibration are shown in Fig. 3, demonstrating that the system was more or less stabilized after 1 ns.

The equilibrated configuration of double-stranded DNA is more regular and less distorted by the helicase in the s state than that in the p state (see Fig. 1), as evidenced by fewer contacts between double-stranded DNA and domain 2B in the s state. The binding between the translocation domains and ssDNA is illustrated in enlarged views in Fig. 1. One can recognize that nucleotide 18 points its base upward in the s state (Fig. 1 a) and downward into a binding pocket in the p state (Fig. 1 b). The pocket, formed by the side chains of Phe-64 and Tyr-257, is closed in the s state and open in the p state.

By monitoring an angle formed by two protein helices, residues 360–374 from 2A and residues 66–81 from 1A, the density of grid points at least 1/Å, the particle-mesh Ewald method (49) was employed for full electrostatics, with motions, i.e., heavy atoms for temperature control (310 K) with a coupling coefficient of 5/ps. 100 fs and a damping time of 50 fs; Langevin forces (52) were applied to all atoms for temperature control (310 K) with a coupling coefficient of 5/ps.

**FIGURE 3 Equilibrium MD simulations of the PcrA-DNA complex.** The simulation system is shown on the left, with the PcrA-DNA complex (protein in dark blue, DNA in magenta) solvated in an explicit water box (light blue), together with ions (sodium in yellow, magnesium in green, and chloride in light blue); ATP, colored in red, is present in the system. Shown on the right are the root mean-square deviation value of protein and DNA backbone atoms with (top) and without (bottom) ATP bound.
during equilibration, it was found that the ATP binding cleft in-between domain 2A and 1A is $-10^\circ$ wider in the $p$ state than in the $s$ state. The opening and closing of the ATP binding cleft in different states (binding of ATP favors a narrower separation between domains 2A and 1A while unbinding of ADP and Pi favors a wider separation between the domains) turn out to be the most important structural feature directly supporting the PcrA translocation, as discussed below.

**Potentials $U_{i\sigma}$ from MD simulation**

Utilizing the equilibrium MD trajectories, we obtained the binding energies $E_b(x_i)$ of individual ssDNA nucleotides. The results, based on Eq. 6, are shown in Fig. 2 for the $s$ and the $p$ state, specifying both the electrostatic and the vdW contributions. It can be recognized that electrostatic energies, only due to the protein, differ significantly for individual nucleotides between $s$ and $p$ states, whereas the vdW energies do not. The standard deviation of the average value of the electrostatic (vdW) energy for each nucleotide is $\sim 2-4$ kcal/mol (0.5 kcal/mol). For each state, the continuous site energy function $E_b(x)$ (also shown in Fig. 2) connecting the discrete binding free energy values was constructed.

$E_b(x)$ permits one to derive the potential $U_{i\sigma}(x_i)$ governing the motion of individual domains along ssDNA (see Eqs. 9–12). The potential $U_{i\sigma}(x_i)$ is shown in the insets of Fig. 2. One can recognize that in the $s$ state the barrier height ($A_{2s}$) in the potential $U_{2s}$ for 2A motion is lower than the barrier ($A_{1s}$) in $U_{1s}$; the opposite is true in the $p$ state. The barrier heights $A_{i\sigma}$ can be expressed explicitly in terms of the site energy differences of nucleotides

$$
A_{2s} = E_b(x_{18}) - E_b(x_{17}) + E_b(x_{16}) - E_b(x_{17}) + 4\delta
$$

$$
A_{1s} = E_b(x_{18}) - E_b(x_{19}) + 4\delta
$$

$$
A_{2p} = E_b(x_{17}) - E_b(x_{18}) + E_b(x_{18}) - E_b(x_{16}) + 4\delta
$$

$$
A_{1p} = E_b(x_{18}) - E_b(x_{20}) + 4\delta,
$$

(14)

where $\delta$ is a tunable parameter (see Appendix A). Although the barrier heights $A_{i\sigma}$ individually depend on $\delta$, the difference between barrier heights of competitive (1A vs. 2A) domain motions in each state is independent of $\delta$, i.e., the value $A_{1s} - A_{2s} \sim 9$ kcal/mol in the $s$ state and $A_{2p} - A_{1p} \sim 12$ kcal/mol in the $p$ state are independent of $\delta$. The difference between barrier heights actually governs the timescale separation between the competing domain motions. From Eq. 14, one can recognize that $A_{2s} - A_{1s}$, after canceling identical terms in $A_{2s}$ and $A_{1s}$, is determined by energy imbalances (combined energy differences between neighboring sites) among some specific nucleotide binding sites, which are indicated in Fig. 2.

This difference between barriers $A_{2s}$ and $A_{1s}$ for domain motions is a key result of our study and was verified by us. In our analysis we accounted only for protein-DNA interactions without explicit solvent contributions. To partially verify the result we carried out an independent set of SMD simulations (sample movies of the SMD simulations are provided in Supplementary Material), pulling ssDNA one half-step forward and backward, along the protein-ssDNA interface across the helicase domains. The forces needed to pull the relevant nucleotides were monitored and are reproduced in Fig. 4. The results show that in the $s$ state, the average force needed to move nucleotides 15–18 rightward (corresponding to 2A motion toward the left) is smaller than that needed to move nucleotides 18–21 leftward (corresponding to 1A motion toward the right); vice versa, in the $p$ state, the average force needed to move nucleotides 15–18 leftward (corresponding to 2A motion toward the right) is larger than the force to move nucleotides 17–20 rightward (corresponding to 1A motion toward the left). This is consistent with the $U_{i\sigma}(x_i)$ potentials shown in Fig. 2. We note that the SMD simulations took into account interactions of ssDNA with both protein and solvent (water/ions) as well as the short time part of the entropic effect.

![FIGURE 4](image-url) Comparisons of SMD forces arising in ssDNA pulling simulations in PcrA with ATP bound (a) and without ATP bound (b). The green/red curve represents the average force needed to move relevant nucleotides (indicated by green/red arrows), corresponding to the movement of domain 1A/2A in the opposite direction. The thin curves were measured directly from simulations, while the thick curves were smoothed over every 10 data points. The results show that in panel 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 panel b.
Langevin dynamics of PcrA translocation in two scenarios

The goal of our study is to explain the physical mechanism underlying the unidirectional translocation of PcrA along ssDNA. Individual translocation steps per ATP hydrolysis require ~20 ms (13). Our MD simulations cover, however, only nanoseconds. As explained above, the timescale gap can be overcome through a stochastic, i.e., Langevin, dynamics description that builds on the potentials $U_i$ (see Eq. 3 and 9–12) estimated from the MD simulations. As explained in Methods, we simulated the millisecond domain movements of PcrA through Eq. 1 in corresponding states employing the potential function defined in Eq. 3. In case of the weak coupling scenario we choose for $V_s$ a vanishing potential; in case of the strong coupling scenario, we describe $V_s$ through Eq. 4. Our description includes random transitions between states. The results of this description are provided in Fig. 5.

Weak coupling scenario

This case is presented in Fig. 5 a. The left panel of Fig. 5 a shows the stochastic trajectories of domains 2A and 1A along ssDNA. Over the period of 10 hydrolysis cycles, i.e., 200 ms, the positions of 2A and 1A change by ~70 Å, reflecting 11 translocation steps with variable durations (due to the stochastic nature) in between. One can recognize that the domains move in a nearly synchronous fashion along the $-x$-direction (see movie provided in Supplementary Material). Underlying this motion is the scenario shown in the right panel of Fig. 5 a. The system traverses sequentially configurations $1' \rightarrow 2' \rightarrow 3' \rightarrow 4' \rightarrow 1''$ as identified in the figure. The configurations are also identified through state labels $p$ and $s$, introduced earlier. In configuration $1'$, PcrA is in the $p$ state, domains 2A and 1A are separated by 1-nt distance and move according to potentials $U_{2p}$ and $U_{1p}$, respectively; 1A experiences a low barrier and can move readily, while 2A experiences a high barrier and is essentially stuck. When 1A has moved forward (to the left) close enough to domain 2A (e.g., by $l_0/3$, see Methods) it reaches configuration $2'$. In this configuration the system has a high probability to transit to the $s$ state (in reality this corresponds to the approach of domains 2A and 1A, leading to an optimal binding geometry for ATP), reaching configuration $3'$. In the $s$ state, however, the potentials $U_{2s}$ and $U_{1s}$ differ qualitatively, in that now 2A is easy to move and 1A becomes stuck. When 2A moves forward (to the left) far enough from domain 1A (e.g., by $2l_0/3$), PcrA reaches configuration $4'$. In this configuration the system has a high probability of transiting to the $p$ state (this corresponds to ATP hydrolysis/
dissociation of ADP and Pi), reaching the configuration 1$. Since PcrA has now moved 1-nt distance to the left compared to the original configuration 1, we denote the new configuration by 1$.

In the weak coupling scenario, the barrier-crossing domain movements (governed solely by potential $U_{ip}$) are assumed rate-limiting. The barrier heights in configurations 1 and 2 assume the values $A_{2p} = 24.7$ kcal/mol, $A_{1p} = 12.8$ kcal/mol and in configurations 3 and 4 assume the values $A_{2s} = 10.7$ kcal/mol, $A_{1s} = 19.4$ kcal/mol (see Eq. 14). The values were chosen through their $\delta$-dependence in Eq. 14 ($\delta = -0.62$ kcal/mol) such that the simulated speed of translocation agrees with the observed speed (13).

**Strong coupling scenario**

The random motion of PcrA along ssDNA in the strong coupling scenario is presented in Fig. 5 b. The left panel of Fig. 5 b presents the corresponding stochastic motions of domains 2A and 1A. The motions are qualitatively similar to those in the weak coupling case. Over the period of 10 hydrolysis cycles the positions of domains 2A and 1A change by $\sim 65 \AA$, reflecting 10 translocation steps. One can recognize again that the domains move in a nearly synchronous fashion toward the $-x$-direction (see movie provided in Supplementary Material). The scenario underlying this motion is shown in the right panel of Fig. 5 b. The system traverses again sequentially configurations 1 to 2, 2 to 3, 3 to 4 and 4 to 1 as identified in the figure. The configurations are also characterized through state labels $pp$, $ps$, $ss$, and $sp$ introduced earlier. In configuration 1, PcrA is in the $pp$ state, domains 2A and 1A are separated by 1-nt distance ($l_p$ in $V_p$, see Eq. 4), and are both essentially “stuck” in potential $U_{ip}$ exhibiting high barriers. Arrival of ATP for binding leads to configuration 2 and the intermediate $ps$ state. In this state, domains 2A and 1A experience the interaction potential $V_s$ (corresponding to a harmonic spring, with $l_s \leq l_p$) while individual potentials $U_{ip}$ maintain the $pp$ state form. As a result, 1A experiences a lower barrier (combination of $U_{ip}$ and $V_s$) and can move readily; 2A also experiences a reduced barrier, but a higher one than 1A and still remains stuck. As $V_s$ quickly draws the domains 2A and 1A close to each other (through the motion of 1A, but not 2A), PcrA relaxes to the 3$^*$ configuration that corresponds to the $ss$ state. In this configuration, both domains are stuck by $U_{is}$ exhibiting high barriers. In the 3$^*$ configuration, ATP hydrolysis can take place. Once this happens and the hydrolysis products ADP and Pi start dissociating, configuration 4$, corresponding to the $sp$ state, is reached. In configuration 4 the interaction potential changes from $V_s$ back to $V_p$ while individual potentials $U_{ip}$ retain the prior form. As a result, domain 2A experiences a lower barrier (combination of $U_{2s}$ and $V_p$) and can move readily; 1A also experiences a reduced barrier, but a higher one than experienced by 2A, and remains stuck. The relaxation of PcrA under $V_p$ moves 2A to the left and brings the system back to configuration 1$. However, since PcrA has now moved 1-nt distance to the left compared to the original configuration 1$, we denote the new configuration by 1$.

In the strong coupling scenario, the barrier-crossing domain movements (under potential $U_{ip}$ and $V_{ip}$) are not rate-limiting. In the present case, the transitions 1$ \rightarrow 2^*$ and 3$ \rightarrow 4^*$ corresponding to the arrival of ATP and occurrence of ATP hydrolysis are assumed to be much slower than the domain motions and, therefore, are rate-limiting for the overall translocation process. This provides us with more leeway in the choice of the barriers $A_{ip}$. In fact, $\delta$-values in the range between $-1.2$ and $0.4$ kcal/mol give reasonable $A_{ip}$ values such that translocation occurs with a speed consistent with observation (13). We used $\delta = 0$ kcal/mol, which gives barrier heights $A_{2s} = 13.2$ kcal/mol, $A_{1p} = 21.9$ kcal/mol, $A_{2p} = 27.2$ kcal/mol, and $A_{1p} = 15.3$ kcal/mol (see Eq. 14).

**The intrinsic mechanism of unidirectional translocation**

Fig. 5 demonstrates that both suggested PcrA translocation mechanisms, involving weak coupling and strong coupling, lead to unidirectional motion along ssDNA. This behavior is described through the solution of coupled Langevin equations. Here we seek an alternative description in terms of rate equations that provide a more systematic explanation for the unidirectional PcrA translocation as well as a convenient mathematical formulation to represent the associated translocation velocity. Our approach is closely related to the generic description developed in Betterton and Jülicher (23).

We account for the motion of PcrA in terms of the average position $x$ of its two domains 1A and 2A, defining $x = (x_1 + x_2)/2$. The description needs to also attribute to the moving PcrA the various states and configurations underlying the translocation process. These states and configurations are defined for both scenarios in Fig. 5 (right panels). The linking between $x$ and states/configurations is presented in the schematic diagram in Fig. 6, again separately for the two scenarios. We begin with Fig. 6 a, which shows the weak coupling scenario (see Fig. 5 a). The figure places the states/configurations along the position axis. The states correspond to discrete values of position $x$, namely $x = j \cdot \frac{1}{2}, \frac{3}{2}, \frac{5}{2}, \frac{7}{2}, \ldots$ For the sake of better presentation, the states/configurations are arranged in two tiers, I and II. The correspondence with the configurations 1, 2, 3, 4, 1$^{-}$ in Fig. 5 a is indicated in Fig. 6 a: 1$^{-}$ corresponds to point $(j, 1)$, 2$'$ to point $(j - \frac{1}{2}, 1)$, 3$'$ to point $(j - \frac{1}{2}, 2)$, 4$'$ to point $(j - 1, 2)$, and 1$^{-}$ corresponds to point $(j, 1)$. The configurations capture the cyclic translocation dynamics of PcrA and repeat themselves along the position axis. For example, configuration 1$, located at $(j - 1, 1)$, is shifted by one unit along the position axis from configuration 1$. The translocating PcrA cycles through the states/configurations by undergoing transitions between them. The transitions are indicated in Fig. 5 a and include,
as well as the associated rate constants are labeled (see text for detail). Fast
indicated in Fig. 6
i.e., transitions that do not translocate PcrA, are fast relative
drawn in Fig. 6
and used 1/
and 1A cannot be
for example, 1′ ↔ 2′, 2′ ↔ 3′, 3′ ↔ 4′, and 4′ → 1′. The transitions will be described mathematically by linear rate
equations and associated rate constants. The rate constants
r_1, r_2 (s_1, s_2) correspond to the forward (backward) motion of
domains 1A and 2A, respectively (we note that the forward
direction in Fig. 6 is to the left); transitions starting from tier
I and II to the other are described by rate constants \omega_l, \omega_r; reverse transitions are denoted by primes, e.g., r_i′ and \omega_l′. In
the weak coupling scenario the transitions denoted by \omega_l, \omega_r, i.e., transitions that do not translocate PcrA, are fast relative
to the transitions that do translocate PcrA; this is also
indicated in Fig. 6 a. The latter transitions are governed by
the potentials \bar{U}_{ir}. The corresponding energy barriers are
drawn in Fig. 6 a to indicate which transitions are feasible
(only those connected with low barriers).

As stated in Table 1, the average time for the transition
1′ → 2′, which corresponds to domain 1A moving forward
and is governed by potential \bar{U}_{1pr}, i.e., 1/r_1, is estimated in
Appendix B to be ~19 ms; the average time for the transition
1′ → 2, which corresponds to domain 2A moving backward,
and 1/s_2, is estimated likewise to be ~10^{6} s. We note that 1A,
in principle, can move both forward and backward facing the
same low barrier height; however, due to the geometric
constraint (x_1 - x_2 \leq l_0, see Methods), 2A and 1A cannot be
separated too far and since the initial separation between 2A
and 1A in configuration 1′ is large (x_1 - x_2 = l_0), only the
forward motion of 1A is allowed. Values for transition rates
(timess), 1/r_2 and 1/s_1, are also stated in Table 1. One can
recognize that 1/r_1 + 1/r_2 assume a value of ~20 ms as we fit
it to the observed speed (13). The values r_i′ and \omega_l′ (i = 1, 2)
are not listed and they assume the same values (in the weak
coupling scenario) as \omega_l and \omega_r, respectively. The transition
rates between tiers I and II are unknown in this case; how-
ever, since the transitions denoted by \omega_l and \omega_r are relatively
fast, i.e., 1/\omega_l \ll 20 ms (i = I, II), and should be much faster
than the reverse transitions (primed), we concluded \omega_l′ \gg \omega_r′
and used 1/\omega_l = 100 ns and \omega_l′ = 0, accordingly.

We now consider the strong coupling scenario depicted in
Fig. 6 b. The presentation is analogous to that for the weak
coupling scenario; the numbering of configurations corre-
spending to Fig. 5 b (right panel) is: 1′ corresponds to point
(j, II), 2′ to point (j, I), 3′ to point (j - \frac{1}{2}, I), and 4′ to point
(j - \frac{1}{2}, II). However, in this case the domains experience an
interaction described by \bar{V}_{pr}(x_1, x_2), which leads to an
important variation. After transition 1′ → 2′, corresponding
to the arrival of ATP, domains 2A and 1A are still separated
(by l_p = l_0) while the potential \bar{V}_p(x_1, x_2) with a short
energy length (l_p = l_0/3) is switched on, placing the
system in an energized state. This is depicted in Fig. 6 b by
placing configuration 2′ upwards on the energy profile by
\Delta \bar{E} \sim k(l_p - l_0)^2/2. Likewise, configuration 4′, corresponding
to PcrA with ATP just hydrolyzed, is depicted in an upward
position on the energy axis again by \Delta \bar{E}, where domains 2A
and 1A are close while the potential \bar{V}_p(x_1, x_2) with a large
energy length (l_p = l_0) is switched on. Therefore, the transitions
2′ → 3′ and 4′ → 1′, which correspond, respectively,
to domain 1A and 2A moving forward, are energetically
favorable and happen fast.

The associated rate constants (times) for the transitions
involving domain movements in Fig. 6 b are also provided in
Table 1, with the values estimated in Appendix B. The
primed rate constants, not shown in the table, are estimated
as in Betterton and Jülicher (23): r_i′ = r_i \exp(-\Delta \bar{E}/k_B T) and
s_i′ = s_i \exp(-\Delta \bar{E}/k_B T), thus r_i′ and s_i′ (i = 1, 2) assume much
smaller values than \omega_l and \omega_r. For transitions between tiers I
and II (not involving domain movements), \omega_l′ = \omega_l is used
assuming an isoenergetic transition 3′ ↔ 4′ (55). We use
\omega_r′ = \omega_r for simplicity; \omega_l′ and \omega_r′ are related by
\omega_r′ = \omega_l′ \exp(\mu - \Delta \bar{E}/k_B T) due to the free energy changes, where \mu is
the chemical energy generated in one ATP hydrolysis cycle.

<table>
<thead>
<tr>
<th>Time</th>
<th>Physical correspondence</th>
<th>Weak coupling</th>
<th>Strong coupling</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/r_1</td>
<td>1A moving forward</td>
<td>19 ms</td>
<td>0.3 ms</td>
</tr>
<tr>
<td>1/s_2</td>
<td>2A moving backward</td>
<td>3 \times 10^{6} s</td>
<td>4 \times 10^{6} s</td>
</tr>
<tr>
<td>1/r_2</td>
<td>2A moving forward</td>
<td>0.8 ms</td>
<td>0.03 \mu</td>
</tr>
<tr>
<td>1/s_1</td>
<td>1A moving backward</td>
<td>6 \times 10^{2} s</td>
<td>14 ms</td>
</tr>
<tr>
<td>1/\omega_l</td>
<td>No domain movement</td>
<td>100 ns (assumed)</td>
<td>15.4 ms</td>
</tr>
<tr>
<td>1/\omega_r</td>
<td>No domain movement</td>
<td>100 ns (assumed)</td>
<td>4.6 ms</td>
</tr>
</tbody>
</table>

The corresponding domain movements (or lack of movement) of
the associated transitions are stated.
(with $\mu = 20 \, k_B T$ used). Since now the transitions between tiers I and II are rate-limiting, $1/\omega_I + 1/\omega_{II}$ has to assume a value of 20 ms according to the observed translocation speed (13).

The transitions shown in Fig. 6 capture the coupling of ATP binding and hydrolysis/dissociation to the translocation of PcrA. The transitions can be cast into a rate equation that describes the probabilities of finding PcrA in the configurations...1’, 2’, 3’, 4’,... shown in Fig. 6. For example, in the strong coupling scenario, the rate equation of the associated probabilities $P_1$, $P_2$, etc., reads ($\dot{P} = dP/dt$),

$$
\dot{P}_1 = J_{14} + J_{14} + J_{14}, \\
\dot{P}_2 = J_{23} + J_{23} + J_{24}, \\
\dot{P}_3 = J_{32} + J_{32} + J_{34}, \\
\dot{P}_4 = J_{41} + J_{41} + J_{44},
$$

as can be verified from inspecting Fig. 6b. The quantities $J_{\mu\nu}$ with $\mu, \nu = 1', 2', \ldots$ are the rate coefficients that can be identified from Fig. 6. For example, in the weak coupling scenario $k(2'\rightarrow 1') = r_1$ and $k(2'\rightarrow 3') = \omega_1$ ($= 0$).

The rate equation can be cast into the form

$$
\dot{\mathbf{P}} = \mathbf{K}\mathbf{P},
$$

where $\mathbf{P}$ is the infinite-dimensional probability vector (written here in transposed form),

$$
\mathbf{P}^T = (\ldots P_1P_2P_3P_4P_1', P_2', P_3', P_4', \ldots)
$$

(we assume for the sake of simplicity that the ssDNA is infinitely long). $\mathbf{K}$ is a matrix of rate coefficients that can be established from Eqs. 15 and 16. As often done in solving kinetic equations, we actually describe the steady state of translocating PcrA, which is characterized through $\dot{\mathbf{P}} = 0$.

For such a steady state, the probability vector (18) should assume a spatially periodic form, namely

$$
\mathbf{P}^T = (\ldots P_1P_2P_3P_4P_1P_2P_3P_4P_1P_2P_3P_4P_1P_2P_3P_4P_4\ldots)
$$

The probabilities $\dot{\mathbf{P}}^T = \mathbf{P}_1P_2P_3P_4$ obey then the four-dimensional linear equation

$$
\mathbf{M}\dot{\mathbf{P}} = 0
$$

as well as the condition $P_1 + P_2 + P_3 + P_4 = 1$. The matrix $\mathbf{M}$ is

$$
\mathbf{M} = \begin{pmatrix}
-\omega_I & -\omega_{II} & \omega_{II} & 0 \\
\omega_{II} & -\omega_I & -\omega_{II} - \omega_{II} & 0 \\
\omega_{II} & -\omega_I & -\omega_{II} - \omega_{II} & 0 \\
\omega_{II} & -\omega_I & -\omega_{II} - \omega_{II} & 0 \\
\omega_{II} & -\omega_I & -\omega_{II} - \omega_{II} & 0
\end{pmatrix}
$$

which follows from Eqs. 15 to 17. A solution for Eq. 20 exists since the condition $\det \mathbf{M} = 0$ is satisfied (the sum of the rows of $\mathbf{M}$ vanishes). The unique (null space of $\mathbf{M}$ has dimension 1) solution can be obtained through a numerical algorithm as provided, for example, in the Mathematica (56) package.

One can now determine the translocation velocity of PcrA along ssDNA. For this purpose we note first that the stationary condition gives the identities $J_{3\rightarrow 2} = J_{2\rightarrow 1} = J_{3\rightarrow 2} = J_{2\rightarrow 1} = J_{1\rightarrow 4} = J_{4\rightarrow 1} = \mathbf{P}^T J$. This is because, in the steady state, the probabilities $\mathbf{P}^T$ in Eq. 19 are translation invariant along ssDNA; the flux $J_{\mu\nu}$, which is only a combination of probabilities according to Eq. 16, should also be translation invariant. The corresponding translocation velocity is $v = \omega_{II}(J_1 + J_4)$ (28). One can obtain, e.g., using a symbolic programming language like Mathematica (56), an explicit expression for $v$ in terms of the rate constants shown in Fig. 6, $a$ and $b$. One finds

$$
v = -(r_1r_2 - s_1s_2)C
$$

where $C$ is positive; its analytic form is given in Appendix C.

The previous calculations provide both the probabilities to find PcrA in the configurations 1–4 as well as the translocation velocity $v$. For the weak coupling scenario, we determined the numerical values ($P_1, P_2, P_3, P_4 \sim (0.85, 0.0, 0.15, 0.0)$, implying that 85% of the time PcrA is in configuration 1 and 15% of the time it is in configuration 3 (see Fig. 5a); configurations 2 and 4 are short-lived transition states. This behavior, i.e., that the system is predominantly in configuration 1 (equilibrium product state $p$), can be recognized in Fig. 5a (left panel) as well as in a movie (showing the actual Langevin dynamics description) supplied in Supplementary Material. The calculated velocity is $v \sim 6.5 \text{ Å}/20 \text{ ms}$, a value that results from adjusting the $\delta$-parameter appropriately.

In the strong coupling case the probabilities for the configurations 1–4 defined in Fig. 5b are ($P_1, P_2, P_3, P_4 \sim (0.23, 0.01, 0.76, 0.0)$, implying that 76% of the time PcrA is in configuration 3 and 23% of the time it is in configuration 1. This behavior, i.e., that the system is predominantly in configuration 3 (equilibrium substrate state $s$) can be recognized also in Fig. 5b (left panel) as well as in a movie (showing the actual Langevin dynamics description) supplied in Supplementary Material. The calculated velocity is again $v \sim 6.5 \text{ Å}/20 \text{ ms}$, since it resulted again from the adjustment of the $\delta$-parameter.

Although the numerical value of $v$ stems from the choice of $\delta$, the analytical expression for $v$, given in Eq. 22, reveals that
the direction of translocation depends on the rates $r_1$, $r_2$, $s_1$, $s_2$ through the factor $-(r_1 r_2 - s_1 s_2)$ (negative velocities correspond to the forward direction). The rate constants $r_1$, $r_2$, $s_1$, and $s_2$, denoted in Fig. 6, correspond to the motions of 1A moving forward, 2A moving forward, 1A moving backward, and 2A moving backward, respectively (see Table 1). The values of $r_1$, $r_2$, $s_1$, and $s_2$ are dictated, respectively, by the barrier heights $A_{1p}$, $A_{2p}$, $A_{1s}$, and $A_{2s}$ introduced before. One can recognize that condition $A_{2p} > A_{1p}$ and $A_{2s} < A_{1s}$ results in the forward motion of PcrA ($v < 0$, i.e., from $3'$ to $5'$). Obviously, by switching the barrier heights to $A_{2p} < A_{1p}$ and $A_{2s} > A_{1s}$, one can reverse the translocation direction (to $v > 0$, i.e., from $5'$ to $3'$). In the case $A_{2p} > A_{1p}$ and $A_{2s} > A_{1s}$, or $A_{2p} < A_{1p}$ and $A_{2s} < A_{1s}$, translocation may be unidirectional, but is slow or may get stalled. We note that, in addition to the barriers, geometric constraints between domains as well as transitions between tiers I and II (in Fig. 6) also play important roles in coordinating the unidirectional translocation.

It is not clear which scenario, weak or strong coupling, PcrA realizes. From studies of F1-ATPase and the close structural homology between F1-ATPase and PcrA (17), it is proposed that ATP binding is linked to a power stroke (20,57,58); likewise, in the strong coupling scenario, binding of ATP induces a power stroke, corresponding to an energetically enforced transition ($2' \rightarrow 3'$ in Fig. 6 b). Single-molecule experiments (S. Myong, I. Rasnik, T. Lohman, and T. Ha, unpublished) suggest that ADP dissociation is the rate-limiting step of Rep helicase translocation. However, ADP dissociation is a fast step in our strong coupling scenario due to release of energy $\Delta E (4' \rightarrow 1'$ in Fig. 6 b). Accordingly, it seems likely that PcrA translocation employs a mixed (strong and weak coupling) scenario: ATP binding linked to 1A motion follows a strong coupling scenario, while ATP hydrolysis or ADP+Pi dissociation linked to 2A motion follows a weak coupling scenario. A mixed scenario for the translocation of PcrA along ssDNA will be discussed in Conclusion.

**Key amino-acid residues affecting the unidirectional translocation**

As demonstrated above, it is the relative height of barriers for domain motions, namely, $A_{2p} > A_{1p}$ and $A_{2s} < A_{1s}$, that dictates the $3'$ to $5'$ translocation of PcrA. Now we want to show how this condition is achieved through the PcrA-DNA complex structure and identify key amino-acid residues that contribute prominently to the barriers. According to our procedure for deriving the domain potential $U_{ap}$ (Eqs. 9–12), we can track the components of the corresponding barrier $A_{ap}$ down to the interaction energy contribution from individual amino acids to the nucleotide binding sites. Eq. 14 shows that the barrier difference $A_{2ap} - A_{1ap}$, which dictates the timescale separation of the competing domain motions (and hence the unidirectional translocation), is actually brought about by site energy imbalances of several involved nucleotides. For each state ($\sigma = s$ and $p$), the right panel of Fig. 7 shows the interaction contributions, in both electrostatic and vdW energies, from individual amino acids to $A_{2ap} - A_{1ap}$. The residues that contribute most are labeled and the atomic details of the involved amino acids and nucleotides are shown in the left panel of Fig. 7.

In the $s$ state, $A_{1s} - A_{2s} = E_p(x_{18}) - E_p(x_{19}) - 2E_p(x_{17})$ holds, with $x_{16}$, $x_{17}$, and $x_{19}$ denoting the binding sites of nucleotides 16, 17, and 19 (see Fig. 2 a and Fig. 7 a). From the right panel of Fig. 7 a, one can recognize that Arg-260 contributes most prominently through both electrostatic (32 kcal/mol) and vdW (5 kcal/mol) energies, to $A_{1s} - A_{2s}$. In the $p$ state, $A_{2p} - A_{1p} = E_p(x_{15}) - E_p(x_{16}) + E_p(x_{17}) - 2E_p(x_{18}) + E_p(x_{20})$ holds, with $x_{15}$, $x_{16}$, $x_{17}$, $x_{18}$, and $x_{20}$ denoting the binding sites of nucleotides 15, 16, 17, 18, and 20 (see Fig. 2 b and Fig. 7 b). If Arg-260 contributed similarly in the $p$ state as it does in the $s$ state, it would have contributed a large negative component to $A_{2p} - A_{1p}$. However, Arg-260 contributes little in the $p$ state. On the other hand, Lys-385, which contributes little in the $s$ state, exhibits a prominent contribution in electrostatic energy (53 kcal/mol) to $A_{2p} - A_{1p}$ in the $p$ state.

Close examination shows that the conformations of Arg-260 in the $s$ and $p$ states are quite different. Without the conformational change of Arg-260, the barrier for domain 1A, $A_{1ap}$, may be always higher (or lower) than that for domain 2A, $A_{2ap}$; accordingly, domain 1A (2A) cannot move in either state such that PcrA is stuck on the ssDNA. Thus, Arg-260 is key for the unidirectional translocation of PcrA. The observation is consistent with experimental data that mutation R260A abolishes the helicase function of PcrA (60).

Lys-385 located on a loop region linking domains 2A and 2B is highly conserved among PcrA helicases. In the $p$ state it interacts closely with nearby P1 on ssDNA. A test simulation (data not shown) pulling ssDNA through PcrA showed that every time a phosphate group passed by Lys-385, a tight interaction is established (data not shown). For example, the ATP binding site of nucleotides 16, 17, and 19 (see Fig. 7 a) pulls ssDNA through PcrA. Thus, Arg-260 may be always higher (or lower) than that for domain 1A, $A_{1ap}$; accordingly, domain 1A (2A) cannot move in either state such that PcrA is stuck on the ssDNA. Thus, Arg-260 is key for the unidirectional translocation of PcrA. The observation is consistent with experimental data that mutation R260A abolishes the helicase function of PcrA (60).

**Further evidence for alternating motional asymmetry in PcrA**

Key for the inchworm model of PcrA translocation are the alternating domain mobilities reflected in the potentials
Structure-Based Model of PcrA Helicase for substrate (relaxation maps (colored according to the amplitude of desorption and reassociation events)) obtained from an MD simulation and where $r_i$ is the position of atom $i$ ($\alpha$ in protein and $P$ in DNA) obtained from an MD simulation and where $\langle \ldots \rangle$ denotes the time average over the MD simulation. The correlation maps (colored according to the amplitude of $C(i, j)$) for substrate ($s$) and product ($p$) states are shown in Fig. 8a. One can recognize that in state $s$ the motion inside domain 1A is more strongly correlated than that inside domain 2A, while the opposite is true in state $p$. This character suggests that domain 1A is more rigid and less mobile than 2A in state $s$ and becomes less rigid and more mobile than 2A in state $p$.

We also investigated the PcrA-DNA complex based on an elastic network model employing the method suggested in Zheng and Brooks (61) and provided as a web service (Web Server, AD-ENM. http://enm.lobos.nih.gov/). Following Zheng and Brooks (61), we calculated the residue contributions to the fluctuation of the ATP binding pocket in both the $s$ and $p$ states. The binding pocket includes altogether 80 residues on the seven conserved motifs (5,9), located between domains 1A and 2A; the residue contribution describes how the elastic motions of all residues are coupled to these motions of PcrA with ATP bound ($a$) and without ATP bound ($b$). (Left) Shown is a detailed view of ssDNA units (nucleotides, in licorice presentation) and relevant amino acids in its vicinity (in both licorice and transparent vdW presentations). The important amino acids, which energetically contribute most, are labeled. The same color code is used as in Fig. 1. Note that Arg-260 and Lys-385 contribute most prominently to the barrier difference $|A_{3p}\sim A_{1s}|$ in panel $a$ and $A_{2p}\sim A_{1s}$ in panel $b$, respectively; Lys-385 is located in a loop region, shown in transparent tube presentation, colored yellow. The black dashed curve circles around the ATP binding pocket. (Right) Shown are the contributions from individual amino acids to the barrier difference $A_{1p}\sim A_{2s}$ in panel $a$ and $A_{2p}\sim A_{1p}$ in panel $b$, separating electrostatic (top) and vdW (bottom) contributions.
can explain PcrA translocation; yet, the mobilities are actually difficult to identify on the nanosecond timescale of MD simulations. But the combined evidence from nucleotide binding energies, steered molecular dynamics (SMD), cross correlation, and dynamic coupling analyses all point to a clear alternation of domain-ssDNA behavior that appears to be fundamental for PcrA translocation.

CONCLUSION

Combining nanosecond MD simulation and millisecond stochastic modeling, we studied monomeric PcrA helicase translocating along ssDNA. Based on structural information on the PcrA-DNA complex and equilibrium interaction energetics sampled from MD simulation, we derived the potential governing the movement of PcrA domains and simulated the stochastic dynamics of the translocation in two limiting scenarios. The advantage of utilizing MD simulation here is that we can estimate more realistically the potential and at the same time, trace the energy contribution to individual amino acids/nucleotides.

According to the derived potentials, the two translocating domains, 1A and 2A, in turn experience lower and higher energy barriers during the ATP hydrolysis cycle, so that the domains alternatively move along ssDNA, each in the same direction. Two limiting scenarios are proposed for PcrA translocation. In the weak coupling scenario, there is no interaction potential between domains; the domain experiencing the lower energy barriers in one state will move forward purely through random thermal motion; once the domain movement happens, some physical event, e.g., ATP binding, hydrolysis, or ADP-Pi dissociation, is stabilized or triggered. This in turn causes the conformation of the protein to be converted to that of the other state, in which only the other domain can move forward; in this scenario, the domain movements are rate-limiting.

In the strong coupling scenario, the two domains are facing relatively high barriers so that without interaction potential, neither can move; the interaction potential is driven by physical events such as ATP binding or ADP + Pi dissociation, which reduces the barriers so that the domain experiencing the lower reduced barrier quickly moves forward; the rate-limiting step in this scenario is any transition not involving domain movements.

Although it is not clear which scenario PcrA employs in reality, evidence suggests that PcrA works in a mixed scenario, in which ATP binding serves as a powerstroke facilitating the forward movement of domain 1A (and assists PcrA conformational change so that 1A cannot move backward), while ATP hydrolysis or ADP + Pi dissociation triggers the thermally agitated forward movement of domain 2A (and also assists PcrA conformational change to prevent...
2A from moving backward). This scenario is shown schematically in Fig. 9.

Fig. 9 depicts one PcrA translocation cycle involving five steps. In the first configuration, domains 2A and 1A are separated without ATP or ADP bound. The bases of the ssDNA track are shown in the insert to the left, revealing a base inserted between side groups Tyr-257 and Phe-64. Domain 1A can move more easily along ssDNA than can domain 2A. At this point ATP arrives at the binding site and begins to insert itself. The binding of ATP exerts an attractive force between the domains that consequently approach each other, but through motion of domain 1A, rather than 2A, leading to configuration 3. The close approach of domains 1A and 2A pushes Tyr-257 and Phe-64 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 Arg-610 and Arg-287 into close contact with the γ-phosphate of ATP; the movement of Tyr-257 is transmitted to Gln-254. QM/MM calculations have shown that the three mentioned side groups, Arg-610, Arg-287, and Gln-254, control the hydrolysis of ATP in PcrA (17). In configuration 3 the juxtaposition of the three side groups presumably is optimal for hydrolysis of ATP to ADP + Pi, PcrA quickly reaching configuration 4. At this point the mobility of domains 2A and 1A has reversed, 2A being the one to move more easily. Thermal fluctuations lead 2A to separate then from 1A reaching configuration 5, permitting release of ADP and Pi, reaching configuration 1′. PcrA has advanced by one step and is ready, after binding another molecule of ATP, to continue its unidirectional movement.

In any scenario, the unidirectional translocation of PcrA is controlled by the barrier difference between two competing domain motions and alternating high and low barriers of the domain during a translocation cycle. Through this mechanism we identify key amino acids that play important roles sustaining PcrA translocation. The study may facilitate understanding not only of PcrA and other helicases, but also of other ATPases and molecular motors.

**APPENDIX A: CONSTRUCTION OF CONTINUOUS SITE ENERGY FUNCTION**

Having determined the relative binding free energies at individual nucleotide binding sites from MD simulations, we constructed a continuous site-energy function $E_d(s)$ for each ($s$ and $p$) state as shown schematically in Fig. 2, $a$ and $b$, for the purpose of deriving the potentials $U_{dp}(s)$. The value $E_d(s)$ represents the (relative) binding energy experienced by a nucleotide at position $x$. The actual curve $E_d(s)$ is constructed by connecting the (five/six in ss) binding free-energy values measured above through third-order polynomial interpolation, subject to the constraints ($E' = dE/ds$; $E'' = d^2E/ds^2$):

$$E_d(s_i) = 0.5E_{p}^{ss} + E_{a}^{ss} - E_d(x_{i+1/2})$$

$$= \max\{E_d(x_i), E_d(x_{i+1})\} + \delta,$$  \hspace{1cm} (24)

$$E'_d(s_i) = E'_d(x_{i+1/2}) = 0, E''_d(s_i) = 1, E''_d(x_{i+1/2}) = -1,$$ \hspace{1cm} (25)

$$E_d(x_{i+1/2}) = E_b(x_{2i}) \left(\text{in ss/p}\right).$$  \hspace{1cm} (26)

The expressions $s_i, i = 15/14$ (in ss), $\ldots, 21$, as defined in Fig. 2, are discrete positions of individual ssDNA units (for the nucleotides, numbered for convenience, see Fig. 1). The expressions $x_{i+1/14}$ (in ss) and $x_{14}$ correspond to boundary sites exposed to solvent. The site energy values, $E_d(x_i)$, of individual binding sites at locations $x_i$ are given in Eq. 24; the energy values, $E_d(x_{i+1/2})$, at locations $x_{i+1/2}$ between binding sites $x_i$ and $x_{i+1}$ are also given in Eq. 24; the expression contains the term $\max\{E_d(x_i), E_d(x_{i+1})\}$ in order to mold an energy barrier into the function $U_{dp}(s)$ for a proper choice of $\delta$ as discussed in the main text. Here, $\delta$ is a single (the same for all $s_i$) tunable parameter that controls the size of undulation of the site.

**FIGURE 9** Five-step PcrA translocation cycle. The figure shows schematically a translocation cycle for a mixed scenario (weak coupling as well as strong coupling; see text) involving configurations 1–5 as well as configuration 1′, which is equivalent to configuration 1, except moved forward by one nt. The mixed scenario involves both a loaded spring (nonzero potential $V_a(x_i, x_{i+1})$, transition 2→3) and a step with a random thermal motion (vanishing potential $V_p(x_i, x_{i+1})$, transition 2→3). In configurations 1, 2, and 1′, domain 1A (green) moves more readily than domain 2A (red), while it is the opposite for configurations 3–5. The inset 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, Arg-287, and Arg-610 move close to the γ-phosphate of ATP; Gln-254 is linked closely to Tyr-257, which forms a key binding pocket for an ssDNA base, but squeezes out the base when the domains approach each other in binding ATP; Gln-254 has been identified as a key participant in ATP hydrolysis along with the mentioned arginines (17). The suggested mechanism therefore involves three steps: 1), binding of ATP that pulls domain 1A toward domain 2A; 2), insertion of Arg-287 and Arg-610 into an optimal (for hydrolysis) position in the ATP binding pocket along with Gln-254 linked to a key ssDNA interaction site; and 3), rapid hydrolysis of ATP that initiates separation of domains 2A and 1A through movement of 2A alone.
energy curve $E_d(x)$. The first and second derivatives of $E_d(x)$ at each site are given in Eq. 25. Periodic boundary conditions are enforced through Eq. 26, requiring both left and right boundary sites to have the same site energy; the value of this energy is discussed in the main text.

**APPENDIX B: RATE CONSTANTS FOR DOMAIN MOTIONS**

In Table 1 of the main text, we list the values of the rate constants for forward and backward barrier-crossing movements of domains 1A and 2A. Here we explain how these rate constants are evaluated. Each rate constant is evaluated as $1/\tau$, where $\tau$ is calculated as the mean first-passage time (9,27,63):

$$\tau = \frac{1}{D} \int_{x_1}^{x_d} dx \exp\left[\frac{W(x)}{k_BT}\right] \int_{x_1}^{x} dy \exp\left[-\frac{W(y)}{k_BT}\right].$$

(27)

Here, $D$ is the diffusion coefficient $D = k_BT/\gamma$ (see Eq. 5 in Methods) assumed to be independent of $x$; $x_1$ is defined as a reflection boundary position, $x$ is the starting position of the domain motion, and $x_d$ is the destination position (values of $x_1$, $x$, and $x_d$ are listed in Table 2). We utilize the potential $W_{ss}(x_1, x_2)$ (see Eq. 3 in Methods) in a specified state for only one domain to evaluate the corresponding rate constant, since the other domain remains stuck due to a high energy barrier. The explicit forms of $W(x)=W_{ss}(x, t_{	ext{time}})$ in calculating $\tau$ by Eq. 27 are presented below in both the weak coupling and the strong coupling scenarios. For convenience of calculation, we use a simple sinusoidal function with amplitude $A_{ss}$ to represent the $U_{ss}(x_1)$ term included in $W_{ss}(x_1, x_2)$, since it is mainly the amplitude of the potential barrier that determines the value of $\tau$ in Eq. 27.

**Weak coupling scenario**

To evaluate $r_1$ and $s_2$ (see tier I in Fig. 6a), which correspond to domain 1A moving forward and domain 2A moving backward, respectively, the following potentials are assumed:

$$W(x_1) = W_p(x_1, x_2 = 0) = A_{sp}\sin[\pi x_1/l_0]$$

$$W(x_2) = W_p(x_1 = l_0, x_2) = A_{sp}\sin[\pi x_2/l_0].$$

(28)

To evaluate $r_2$ and $s_1$ (see tier II in Fig. 6a), which correspond to 2A moving forward and 1A moving backward, respectively, the following potentials are assumed:

$$W(x_2) = W_p(x_1 = 0, x_2) = -A_{sp}\sin[\pi x_2/l_0]$$

$$W(x_1) = W_p(x_1, x_2 = 0) = A_{sp}\sin[\pi x_1/l_0].$$

(29)

**STRONG COUPLING SCENARIO**

To evaluate $r_1$ and $s_2$ (see tier I in Fig. 6b), which correspond to domain 1A moving forward and domain 2A moving backward, respectively, the following potentials are assumed (with parameter $l_0 = l_0/3$):

$$W(x_1) = W_p(x_1, x_2 = 0) = A_{sp}\sin[\pi x_1/l_0] + \frac{k}{2}(x_1 - l_0)^2$$

$$W(x_2) = W_p(x_1 = l_0, x_2) = A_{sp}\sin[\pi x_2/l_0] + \frac{k}{2}(l_0 - x_2 - l_0)^2.$$  

(30)

To evaluate $r_2$ and $s_1$ (see tier II in Fig. 6b), which correspond to 2A moving forward and 1A moving backward, respectively, the following potentials are assumed (with parameter $l_p = l_0$):

$$W(x_2) = W_p(x_1 = 0, x_2) = -A_{2p}\sin[\pi x_2/l_0] + \frac{k}{2}(x_2 + l_p)^2$$

$$W(x_1) = W_p(x_1, x_2 = 0) = A_{1p}\sin[\pi x_1/l_0] + \frac{k}{2}(l_0 - x_1)^2.$$  

(31)

**APPENDIX C: EXPlicit expression of VELOCITY**

In Eq. 22 of the main text, we give the expression of the velocity $v$ for PcrA translocation along ssDNA, which is derived from a calculation of stationary probability fluxes. The value $v$ is calculated to be $-(r_2/s_1 - s_2/C)$, as given in Eq. 22; the explicit expression of $C$, in terms of rate constants $r_1, r_2, s_1, s_2$ and $\omega_1, \omega_2, \omega_3$, is given below, for both the weak coupling and the strong coupling scenarios.

In case of the weak coupling scenario, $\omega_1 = \omega_2 = \omega$ holds

$$C = \omega(r_1 + r_2 + s_1 + s_2) + 4(r_1r_2 + s_1s_2 + r_1s_1 + r_2s_2)$$

(32)

In case of the strong coupling scenario and $\omega_1 = \omega_2 = \omega_3 = \omega$, with chemical potential $\mu$ and energy difference $\Delta E$ (see Fig. 6b), holds

$C = \omega e^{-\Delta E}(e^\mu - 1)/C^\prime$,  

(33)

where

$$C^\prime = (1 + e^{-\Delta E})(r_1 + s_1)(r_2 + s_2)(1 + 2e^{-\Delta E} + e^{2\Delta E})$$

$$+ \omega(r_2 + s_1 + e^{-\Delta E}(r_2 + r_2 + s_2 + 2s_1))$$

$$+ e^{2\Delta E}(r_2 + s_1 + (2 + e^{-\Delta E})(r_1 + s_2)).$$

(34)

**SUPPLEMENTARY MATERIAL**

An online supplement to this article can be found by visiting BJ Online at http://www.biophysj.org.

The authors thank M. Dittrich for fruitful discussions and for sharing data from a QM/MM study of PcrA (17). Molecular images in this article were generated with the molecular graphics program VMD (64).

The present work was supported by National Institutes of Health grants No. P41-RR05969 and No. GM065367. The authors also acknowledge computer time provided through grant No. LRAC MCA93S028.
REFERENCES


59. Reference deleted in proof.


62. Reference deleted in proof.
