Enzyme Selectivity of HIV Reverse Transcriptase: Conformations, Ligands, and Free Energy Partition

ARTICLE *in* THE JOURNAL OF PHYSICAL CHEMISTRY B · JULY 2015

Impact Factor: 3.3 · DOI: 10.1021/acs.jpcb.5b05467 · Source: PubMed

3 AUTHORS, INCLUDING:

Serdal Kirmizialtin
New York University Abu Dhabi
26 PUBLICATIONS **406** CITATIONS

THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

Enzyme Selectivity of HIV Reverse Transcriptase: Conformations, Ligands, and Free Energy Partition

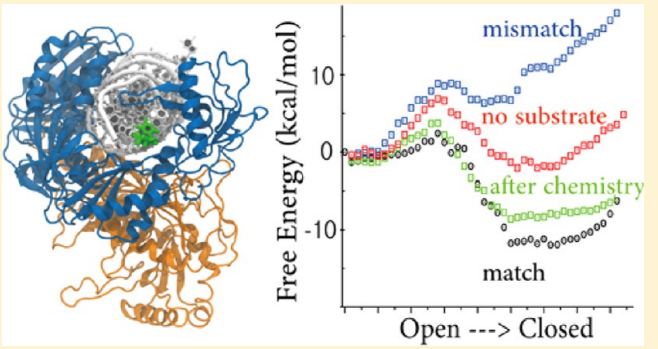
Serdal Kirmizialtin,[†] Kenneth A. Johnson,[‡] and Ron Elber*[,§]

[†]Chemistry Program, New York University at Abu Dhabi, PO Box 129188, Abu Dhabi, United Arab Emirates

[‡]Department of Molecular Biosciences and [§]Department of Chemistry and the Institute for Computational Engineering and Sciences, University of Texas at Austin, Austin, Texas 78712, United States

Supporting Information

ABSTRACT: Atomically detailed simulations of HIV RT are performed to investigate the contributions of the conformational transition to the overall rate and specificity of enzyme catalysis. A number of different scenarios are considered within Milestoning theory to provide a more complete picture of the process of opening and closing the enzyme. We consider the open to closed transition in the absence of and with the correct and incorrect substrates. We also consider the free energy profile and the kinetics of the conformational change after the chemistry step in which a new base was added to the DNA, but the DNA was not yet displaced. We partition the free energy along the reaction coordinate and analyze the importance of different protein domains. Strikingly, significant influence on the free energy profile is detected for amino acids far from the active site. The overall long-range impact is about 50 percent of the total. We also illustrate that the overall rate is not necessarily determined by the highest free energy barrier along the reaction path (with respect to the free enzyme and substrate) and that the specificity is not necessarily determined by the same reaction step that determines the rate.



I. INTRODUCTION

In our previous work we showed that a substrate-induced change in enzyme structure from an open to a closed state can be the major determinant of enzyme specificity even if the conformational change is not rate determining, and we examined the kinetics of this structural transition by computer simulations at atomic resolution.[1] Here we extend these studies to explore the conformational change in the absence of substrate and after the chemical reaction in order to provide a more complete description of the role of conformational dynamics in a complete enzyme cycle.

Because the relevance of the conformational change to enzyme specificity has been controversial, it is necessary for us to first address recent criticisms. If the conformational change step is rate-determining, it will be the major determinant of enzyme specificity,[2] and there seems to be no dispute of that conclusion. However, the disagreement arises when the conformational change is not rate-determining. In particular, Warshel[3] has made two claims that we disagree with (i) "the highest activation barrier (relative to the unbound state) absolutely determines the overall rate" (see legend to Figure 1 of ref 3), and (ii) "as long as the free energy barriers associated with any of the prechemistry steps are not rate limiting, they could not contribute to the catalysis and then to the fidelity" (see abstract of ref 3).

The error in these statements stems from equating *overall rate* of catalysis to specificity. Stated in terms familiar to enzymologists, enzyme specificity is a function of the apparent second-order rate constant for substrate binding (and subsequent product formation) and is defined by k_{cat}/K_m, the so-called specificity constant. In contrast, the *net rate* is defined by k_{cat}, the maximum rate of turnover at saturating substrate concentration. While enzyme *specificity* is determined by the highest activation barrier relative to the unbound state, the overall rate (or net rate) is a function of the highest absolute barrier relative to their local minima. Although in some cases these two aspects of enzyme catalysis (specificity and net rate) can be attributed to a single step in the reaction sequence, the specificity-determining step can differ from the rate-determining step when the specificity-determining step is largely irreversible and precedes the rate-determining step.

To illustrate this point, we provide counter examples to the claims above using two simple kinetic schemes below that serve as pedagogical examples. We first consider the elementary Michaelis–Menten model for enzyme kinetics for a single substrate and then consider competing reactions with two similar substrates.

We derive the results in considerable detail to remove any suspicion that our argument is merely verbal and to show that our

Received: June 8, 2015
Revised: July 27, 2015

DOI: 10.1021/acs.jpcb.5b05467
J. Phys. Chem. B XXXX, XXX, XXX−XXX

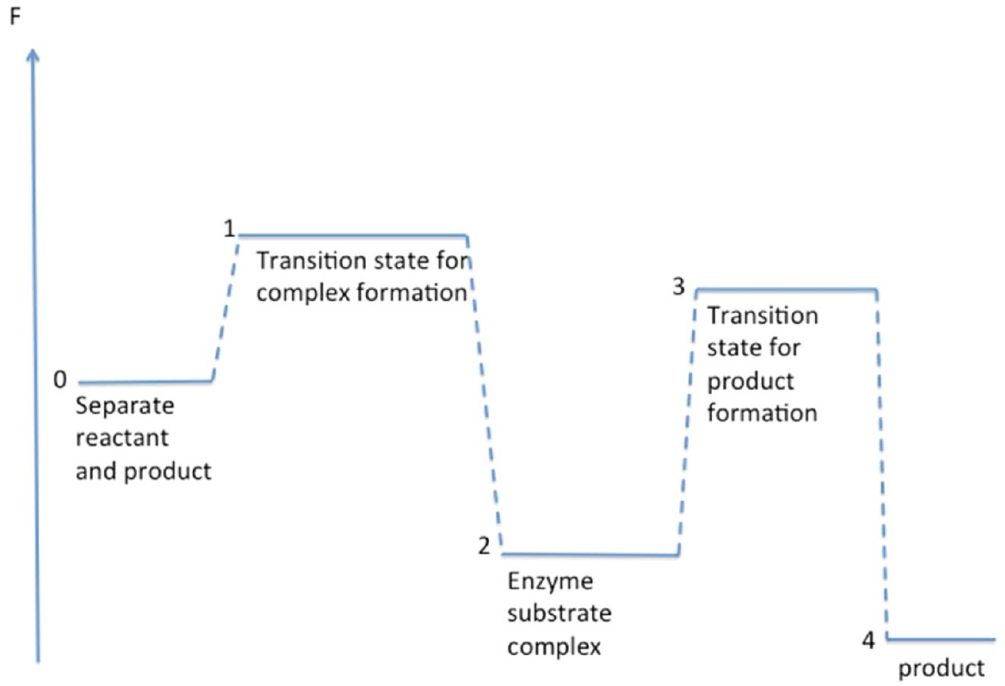


Figure 1. Schematic free energy diagram illustrating a case in which the rate determining state is not with the highest free energy barrier with respect to separate reactants and products. The dashed line is to guide the eye. See text for more details.

conclusions are based on simple mathematical analysis of a minimal model widely used in biochemistry. This analysis can easily be extended to a more complex kinetic scheme that includes an explicit step for the conformational transition preceding chemistry.

We first illustrate that the rate-determining step can be a step in which the barrier height is not the highest relative to the separated enzyme and substrate. Consider a simple scheme of a reaction of an enzyme E with substrate A.

$$\mathrm{E} + \mathrm{A} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \mathrm{EA} \overset{k_2}{\rightarrow} \mathrm{E} + P_\mathrm{A} \tag{1}$$

We use the steady-state approximation on the intermediate EA to write

$$\frac{\mathrm{d}[\mathrm{EA}]}{\mathrm{d}t} = 0 = k_1[\mathrm{E}][\mathrm{A}] - (k_{-1} + k_2)[\mathrm{EA}] \tag{2}$$

and obtain

$$[\mathrm{EA}] = \frac{k_1[\mathrm{E}][\mathrm{A}]}{k_{-1} + k_2} \tag{3}$$

An explicit expression for the amount of the free enzyme [E] is derived using the total concentration (or the number of enzyme molecules) $[\mathrm{E}_0]$, to give $[\mathrm{E}] = [\mathrm{E}_0] - [\mathrm{EA}]$ and therefore

$$[\mathrm{EA}] = \frac{k_1([\mathrm{E}_0] - [\mathrm{EA}])[\mathrm{A}]}{(k_{-1} + k_2)}$$

$$\left(1 + \frac{k_1[\mathrm{A}]}{k_{-1} + k_2}\right)[\mathrm{EA}] = \frac{k_1[\mathrm{E}_0][\mathrm{A}]}{(k_{-1} + k_2)}$$

$$[\mathrm{EA}] = \frac{k_1[\mathrm{E}_0][\mathrm{A}]}{k_{-1} + k_2 + k_1[\mathrm{A}]} \tag{4}$$

Finally, the expression for the rate is

$$\text{rate} = \frac{\mathrm{d}P_\mathrm{A}}{\mathrm{d}t} \cong \frac{k_2k_1[\mathrm{E}_0][\mathrm{A}]}{k_{-1} + k_2 + k_1[\mathrm{A}]} \tag{5}$$

This is the Michaelis−Menten equation as derived by Briggs and Haldane[4] and written with explicit rate constants rather than being reduced to k_{cat} and K_m parameters.

In order to demonstrate our argument, we consider the limit where $k_{-1} \ll k_2$

$$\text{rate} \underset{k_{-1} \ll k_2}{\cong} \frac{k_2k_1[\mathrm{E}_0][\mathrm{A}]}{k_2 + k_1[\mathrm{A}]} \tag{6}$$

Physically this limit means that the backward reaction after complex formation is unlikely and the process becomes essentially irreversible kinetically. This is an important limit found in the action of a number of enzymes, and we have illustrated such a process, computationally and experimentally, in ref 1.

We consider next two other limits which are a function of substrate concentration (i) $k_2 \gg k_1[\mathrm{A}]$ and (ii) $k_2 \ll k_1[\mathrm{A}]$. Note that $k_1[\mathrm{A}]$ is a pseudo first-order rate coefficient for a fixed concentration of substrate. We have for the two cases (i) rate $\cong k_1[\mathrm{E}_0][\mathrm{A}]$ and (ii) rate $\cong k_2[\mathrm{E}_0]$. It should be clear at this point that the term "overall rate" is useful only with reference to a state defined by the substrate concentration. Generally, the terms "net rate" or "overall rate" refer to the maximum rate achieved at high substrate concentration. Thus, the only reasonable interpretation of the term "overall rate" is to mean the maximum rate, which in this model is defined by rate $\cong k_2[\mathrm{E}_0]$ when $k_2 \ll k_1[\mathrm{A}]$. This corresponds to the definition of $k_{cat} = k_2$ for this simple model.

To make the connection to free energy we write the rate coefficients using Arrhenius expressions $k_x = \omega \exp(-\beta\Delta F_x)$ where for simplicity we use the same pre-exponential factor for all rate coefficients. Consider the free energy diagram in Figure 1.

According to the free energy diagram we have $k_1 = \omega \exp(-\beta(F_1 - F_0))$, $k_{-1} = \omega \exp(-\beta(F_2 - F_1))$, and $k_2 = \omega \exp(-\beta(F_3 - F_2))$. From the diagram it is obvious that we have

DOI: 10.1021/acs.jpcb.5b05467
J. Phys. Chem. B XXXX, XXX, XXX−XXX

$k_{-1} \ll k_2$ and $k_2 \ll k_1[A]$. The important conclusion from this analysis is that even though substrate binding (barrier 1) is faster than chemistry (barrier 3), the barrier to release the substrate (reverse reaction) is sufficiently high (barrier 2) such that the barrier for catalysis relative to free enzyme and substrate becomes lower than that for the binding step.

We consider next the fidelity of the enzyme. We show that under the same circumstances the selectivity is controlled by a step that is **not** rate-determining, namely, step 1.

Consider two substrates A and B that compete for the same enzyme E to produce $[P_A]$ and $[P_B]$, respectively. The reaction schemes are

$$\mathrm{E} + \mathrm{A} \underset{k_{-1}^a}{\overset{k_1^a}{\rightleftharpoons}} \mathrm{EA} \overset{k_2^a}{\rightarrow} \mathrm{E} + \mathrm{P_A}$$

$$\mathrm{E} + \mathrm{B} \underset{k_{-1}^b}{\overset{k_1^b}{\rightleftharpoons}} \mathrm{EB} \overset{k_2^b}{\rightarrow} \mathrm{E} + \mathrm{P_B} \tag{7}$$

We define the specificity as the ratio of the two rates

$$S = \frac{\mathrm{rate}(P_A)}{\mathrm{rate}(P_B)} = \frac{k_2^a[\mathrm{EA}]}{k_2^b[\mathrm{EB}]} \tag{8}$$

Assuming steady states for the two intermediates [EA] and [EB], we have

$$\frac{\mathrm{d}[\mathrm{EA}]}{\mathrm{d}t} = k_1^a[\mathrm{E}][\mathrm{A}] - (k_{-1}^a + k_2^a)[\mathrm{EA}] = 0$$

$$\frac{\mathrm{d}[\mathrm{EA}]}{\mathrm{d}t} = k_1^b[\mathrm{E}][\mathrm{B}] - (k_{-1}^b + k_2^b)[\mathrm{EB}] = 0$$

$$[\mathrm{EA}] = \frac{k_1^a[\mathrm{E}][\mathrm{A}]}{(k_{-1}^a + k_2^a)}[\mathrm{EB}] = \frac{k_1^b[\mathrm{E}][\mathrm{B}]}{(k_{-1}^b + k_2^b)} \tag{9}$$

As before, assume that $k_{-1} \ll k_2$

$$S \cong \frac{k_2^a k_1^a[\mathrm{E}][\mathrm{A}]}{k_2^a} \Big/ \frac{k_2^b k_1^b[\mathrm{E}][\mathrm{B}]}{k_2^b} = \frac{k_1^a[\mathrm{A}]}{k_1^b[\mathrm{B}]} \tag{10}$$

From this analysis we demonstrate that the specificity depends on the rate coefficients of step 1, which is not rate determining! Stated in terms that are commonly used in enzymology, the net rate is determined by $k_{cat} = k_2$, while specificity is determined by $k_{cat}/K_m = k_1$ (for this simple model). Our simple analysis here shows that when a binding step is largely irreversible, specificity and net rate are a function of different steps in the pathway. Moreover, this analysis provides clear exceptions to Warshel's general assertions. Specifically, we demonstrate that (i) the highest activation barrier (relative to the unbound state) does not necessarily determine the overall rate, (ii) even when the free energy barrier associated with a prechemistry step is not rate limiting, it can contribute to fidelity.

We have extended our analysis to a real-life example based upon our data defining the kinetics of DNA polymerization. DNA polymerases represent an ideal model for understanding enzyme specificity because their alternative substrates are well-known in the form of noncognate base pairs, and enzyme fidelity is biologically important so specificity in discriminating against similar substrates is quite high. In order to better understand the molecular details underlying enzyme conformational dynamics in governing specificity we previously examined the kinetics of closing the structure of the HIV reverse transcriptase (HIV-RT) with a bound nucleotide substrate[1,5] using atomically detailed simulations and the Milestoning theory.[6,7] We showed that while a correct substrate induces a fast conformational change to bind the substrate tightly and facilitates fast catalysis, an incorrect substrate fails to stabilize the closed state or organize the active site residues for catalysis. Thus, the conformational change is a major determinant of enzyme specificity. In this report, we extend these studies to examine the conformational dynamics in the absence of nucleotide and following nucleotide incorporation and pyrophosphate release.

We also consider the individual contributions of amino acids to the free energy change along the transition pathways. Such partitions can be a useful tool to identify the critical residues that can be targets of site directed mutagenesis experiments. We note, however, that partitioning of the free energy to individual components has long been a controversial issue computationally. In general, the partition depends on the path between two end states. The dependence of free energy partition on the path makes this type of analysis ill defined.[8] However, in the present case we have fixed the path. It is the minimum free energy coordinate for the conformational transition under considerations[9] that were determined using the String method.[10] The assumption that we made in the additional analysis discussed in this paper is of a small perturbation, such that the reaction coordinate does not change significantly from the initial guess in response to substrate and protein variations. Hence the analysis we present here is well-defined formally. It provides unambiguous information about the participation of different protein parts in the reaction once a reaction coordinate is available.

This manuscript is organized as follows. In Methods, we discuss the theory of minimum free energy calculation and Milestoning, the setup of the calculations and the novel analysis of free energy partition. In Results, we outline the kinetics and thermodynamics of enzyme conformational transition in different substrate conditions, and we discuss the potential impact of our free energy partition calculations for future experiments. Conclusions are in the last section.

II. METHODS

II.1. Theory. In this section we briefly review the computational methods used in this manuscript. We describe the locally updated planes (LUP)[11] and the finite temperature String method[12] as implemented in our molecular dynamics code MOIL,[13] to compute minimum free energy pathways. We also outline the computational aspects of the Milestoning method.[14−16] Finally, we consider the exact partition of the free energy along the reaction pathway.

II.1.1. Locally Updated Planes and String Methods. The string method is an iterative algorithm to refine a whole curve to determine a minimum free energy coordinate (MFEP). Consider a one-dimensional curvilinear coordinate and hypersurfaces perpendicular to it (Figure 2) that define a reaction coordinate. A frequent approximation to the hypersurfaces is hyperplanes that are orthogonal to the curvilinear path. In the present study we use hyperplanes similarly to previous investigations.[17−19] In our code, the String method is a direct extension of the LUP[11] that was published in 1990 to compute minimum energy pathways (MEPs). In LUP a discrete representation of the reaction coordinate is quenched, keeping the end points of the first and last structures (reactant and product) fixed. Let $x \in R^{3N}$ be the coordinate vector of the whole system, N is the number of atoms, and let x_R and x_P be the coordinate vectors of the reactant and product, respectively. Let $x(l, \tau)$ be a curve connecting the

DOI: 10.1021/acs.jpcb.5b05467
J. Phys. Chem. B XXXX, XXX, XXX−XXX

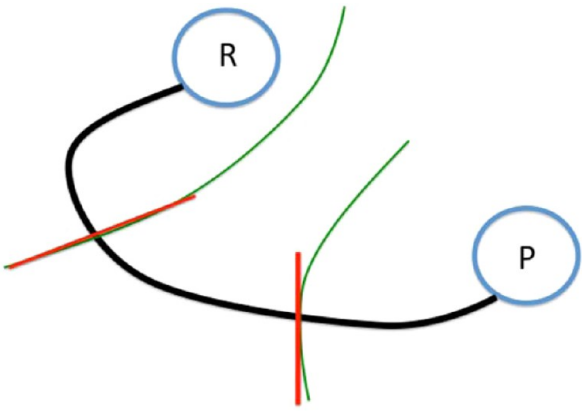


Figure 2. Schematic representation of the path connecting reactant (R) and product (P). Milestones (green curves) are hypersurfaces that are normal to the curvilinear reaction coordinate. In the practical application of the string method the hypersurfaces are replaced by hyperplanes (red straight lines) locally orthogonal to the line connecting the reactants and the products.

reactant and product and parametrized by l such that $x(l = 0, \tau)$ is the reactant and $x(l = L, \tau)$ is the product. Hence the two end points are fixed and are independent of the value of τ. The variable τ is a fictitious time that is used to monitor the path quenching process. The path is approaching the correct solution (the steepest descent path, SDP) monotonically as τ increases. We define the path slope, which is a unit vector along the direction of the path at l

$$e_l = \left(\frac{\mathrm{d}x(l,\tau)}{\mathrm{d}l}\right) \Big/ \left(\left|\frac{\mathrm{d}x(l,\tau)}{\mathrm{d}l}\right|\right) \tag{11}$$

The calculation of the minimum energy path is the minimization of the energy of each of the points along the curve, $x(l, \tau)$. The potential energy is $U(x)$ and the minimization is subject to the constraint that the point remains on the curve.

$$\frac{\mathrm{d}x(l,\tau)}{\mathrm{d}\tau} = -[\mathbf{I} - e_l \cdot e_l^t]\nabla U(x(l,\tau)) \tag{12}$$

The projection operator in eq 12, $(\mathbf{I} - e_l \cdot e_l^t)$, forces the displacement, dx, to remain in the hyperplane orthogonal to the current direction of the curvilinear coordinate. At the limit of $\tau \to \infty$, provided that the potential is bound from below, the curve is reaching a stationary solution in which the right-hand side is zero for all points along the path. The asymptotic path is one possible realization of the steepest descent path, also called the minimum energy path or the intrinsic reaction coordinate.

Let a discrete set of coordinates $x(l = i, \tau)$ $i = 0, \ldots, L$ approximate the continuous curve. In the discrete representation, $x(0, \tau)$ is the vector coordinate of the reactants, $x(L, \tau)$ is the coordinate vector of the products, and both of them are fixed. There are $L - 1$ intervals interpolating between the end states. We assume that the lengths of the intervals are sufficiently small to provide adequate representation of the continuous curve. This assumption can be enforced by additional constraints, which are either penalty functions[20] or holonomic constraints[9] to keep the distance between sequential points along the path the same;[21] for example

$$|x(i,\tau) - x(i-1,\tau)| = |x(i+1,\tau) - x(i,\tau)| \;\forall\; i \tag{13}$$

To find the minimum energy path we solve the quenched equation at the limit $\tau \to \infty$ for the discrete positions along the path

$$\frac{\mathrm{d}x(l_i,\tau)}{\mathrm{d}\tau} = -[\mathbf{I} - e_{l_i}e_{l_i}^t]\nabla U(x(l_i,\tau))\cdot \qquad \forall\; i \neq 2, L-1$$

$$e_{l_i} \approx [x(l_{i+1},\tau) - x(l_{i-1},\tau)]/|x(l_{i+1},\tau) - x(l_{i-1},\tau)| \tag{14}$$

We estimate the path slope using a finite difference between positions along the reaction coordinate as shown in the second line. Note that the path slope is changing dynamically as the individual discrete points along the path are modified and couple the coordinate vectors along the path.

To compute the minimum free energy path in the space of coarse variables, we consider the potential of mean force instead of the potential, and optimize the curve in a subspace of coarse variables. In general the coarse variables can be curvilinear coordinates, which complicates the analysis. We restrict our implementation in MOIL[13] to a subset of coarse variables, which are Cartesian coordinates. For example, we considered all the Cartesian coordinates of the C_α atoms of the amino acids in a protein. Let $y \in R^n$ be the vector of the coarse variables in n dimensions and let z be the vector of the coordinates that supplements y to obtain x. Let s_l be the path slope in the space of coarse variables (similarly to e_l in full space). To obtain the minimum free energy path in the space of coarse variables the String extension of LUP[12] uses the potential of mean force (PMF) instead of the full potential in eq 12. We have

$$\frac{\mathrm{d}y(l_i,\tau)}{\mathrm{d}\tau} = -[\mathbf{I} - s_{l_i}s_{l_i}^t]\langle\nabla_y U(x(l_i,\tau))\rangle_Z \qquad \forall\; i \neq 2, L-1$$

$$s_{l_i} \approx [y(l_{i+1},\tau) - y(l_{i-1},\tau)]/|y(l_{i+1},\tau) - y(l_{i-1},\tau)|$$

$$\langle\nabla_y U(x(l_i,\tau))\rangle_Z \equiv \frac{\int_Z \mathrm{d}z\cdot\nabla_y U(x(l_i,\tau))\cdot\exp[-\beta U(x(l_i,\tau))]}{\int_Z \mathrm{d}z\cdot\exp[-\beta U(x(l_i,\tau))]} \tag{15}$$

where the average in the lower equation is conducted on the subspace, Z. The joined Cartesian spaces Z and Y make the complete space X. We were able to write the average without a Jacobian factor for transformation between coordinate sets ($X \to Y, Z$) since we restrict the representation of the coarse variables to Cartesian only. In practice, we use constant temperature Molecular Dynamics simulations to perform the spatial average in eq 15. Hence, we replace the spatial average by a temporal average in accord with the ergodic hypothesis. The values of the coarse variables are kept fixed during the averaging process and are modified only when propagating the curve according to the top formula of eq 15 for one fictitious time step. Hence, the curve quenching is conducted only in the coarse space.

In summary, eq 15 is the String formula that is implemented in MOIL[13] and is a direct extension of the LUP algorithm. The result of these calculations is a sequential set of points in coarse space, $y_{i=1,\ldots,L}$. This set is a discrete approximation to the curve that defines the reaction coordinate.

The reaction coordinate is the set of hypersurfaces orthogonal to the curve computed earlier. In section II.1.1 we generated a sequence of points in coarse space that are parametrized by a single scalar variable l. The reaction coordinate is characterized

DOI: 10.1021/acs.jpcb.5b05467
J. Phys. Chem. B XXXX, XXX, XXX−XXX

by the values of the coarse variables at points, $y(l)$, and by the slope of the path, q_l. The change in the free energy along the path is given by

$$\begin{aligned}\Delta W(l_i, l_{i+1}) &\cong \langle \nabla U(x(l_i, \tau \to \infty)) \rangle_Z^t \cdot (y(l_{i+1}) - y(l_i)) \\ & i = 1, \ldots, L-1 \\ &= \sum_{j=1,\ldots,n} \langle \nabla U_j(x(l_i, \tau \to \infty)) \rangle_Z \cdot (y_j(l_{i+1}) - y_j(l_i)) \\ &= \sum_{j=1,\ldots,n} \Delta W_j(l_i, l_{i+1})\end{aligned} \tag{16}$$

The sum over j is over the coarse variables. Since we have chosen the coarse variables to be Cartesian, the individual contributions in the above sum can be separated and grouped to different subsets (e.g., atoms that belong to a particular amino acid or to a particular secondary structure element). The final expression shows that the free energy differences can be written as contributions from displacements along different coarse variables. Of course, this exact partition is conditioned on the availability of a reaction coordinate.

Milestoning is a theory and an algorithm that was introduced to compute thermodynamics and long time kinetics along reaction coordinates. While Milestoning at present is a general and rigorous theory[7] which is exact and is not limited to one reaction coordinate, here we exploit the simpler, approximate, and less expensive to compute versions of the algorithm.[6,14,22] Since Milestoning was discussed extensively elsewhere we only define below the important variables and introduce the final formulas.

Definitions in the context of the Milestoning theory are as follows:

(1) Milestones: Milestones are orthogonal hypersurfaces normal to the curve of the reaction coordinate (Figure 2). A trajectory is in state α if the last milestone that it crossed is α.

(2) Probability: The probability, $p_\alpha(x_\alpha, t)$, is the probability that at time t the last milestone crossed by a trajectory at position x_α is α. The stationary probability is defined as the limit $p_{\alpha,\mathrm{stat}}(x_\alpha) = \lim_{t\to\infty} p_\alpha(x_\alpha, t)$.

(3) Free energy: The free energy of state α is defined as $F_\alpha(x_\alpha) = -kT \log(p_{\alpha,\mathrm{stat}}(x_\alpha))$ where k is the Boltzmann constant and T is the temperature.

(4) Flux: The flux, $q_\alpha(x_\alpha, t)$, is the number of trajectories that pass through a milestone α at phase space point x_α in unit time. By construction it is always positive.

(5) Mean First Passage Time (MFPT): The MFPT, $\langle\tau\rangle$, is the time that it takes a trajectory, on the average, to start at the reactant and ends the product state for the first time. In systems that follow exponential kinetics, it is the inverse of the rate coefficient.

(6) Conditional or transition probability (kernel): The kernel $K_{\beta\alpha}(x_\beta, x_\alpha, t)$ is the probability that a trajectory will cross milestone α at point x_α and at time t given that it passes milestone β at point x_β at time 0. The kernel in current Milestoning implementation depends only on the time differences between crossing events.

The starting equation in Milestoning is the equation that accounts for conservation of the flux.

$$\begin{aligned}q_\alpha(x_\alpha, t) &= p_\alpha(x_\alpha, t \\ &= 0)\delta(t - 0^+) \\ &+ \sum_\beta \int_{\Gamma_\beta} \int_0^t \mathrm{d}x_\beta \cdot \mathrm{d}t' \cdot q_\beta(x_\beta, t') K_{\beta\alpha} \\ &(x_\beta, x_\alpha, t - t')\end{aligned} \tag{17}$$

The summation over β is over other milestones that can access directly milestone α (without crossing other milestones in between). The integration in the expression on the right is over the phase space of the milestone Γ_β, and of the time of entry to the other milestones, in an earlier time t'.

Consider a stationary version of eq 17, which we provide without proof. A complete discussion can be found elsewhere.[14,16] A stationary flux is obtained at sufficiently long times in which the system is approaching a steady state. The result of a long term analysis of eq 17 is a linear, integral, and exact equation for the stationary flux

$$\begin{aligned}q_{\alpha,\mathrm{stat}}(x_\alpha) &= \lim_{t\to\infty} q_\alpha(x_\alpha, t) \\ q_{\alpha,\mathrm{stat}}(x_\alpha) &= \sum_\beta \int_{\Gamma_\beta} \mathrm{d}x_\beta \cdot q_{\beta,\mathrm{stat}}(x_\beta) \cdot K_{\beta\alpha}(x_\beta, x_\alpha) \\ K_{\beta\alpha}(x_\beta, x_\alpha) &\equiv \int_0^\infty \mathrm{d}t \cdot K_{\beta\alpha}(x_\beta, x_\alpha; t)\end{aligned} \tag{18}$$

If the system is close to equilibrium in the NVT ensemble[7] one may assume that the initial flux at milestone $q_\beta(x)$ is distributed according to the canonical weight

$$q_{\beta,\mathrm{stat}}(x) = w_\beta \exp(-U(x)/kT) \qquad x \in \Gamma_\beta \tag{19}$$

where w_β is a weight to be determined. If the initial conditions for the trajectories are already sampled from the canonical ensemble conditioned to be at milestone β, the flux is

$$q_{\beta,\mathrm{stat}}(x) = \frac{w_\beta}{L} \sum_{i=1}^{L} \delta(x_i - x) \tag{20}$$

where i is the running index of the sampled trajectories and L is their total number.

Plugging eq 19 into eq 18 and integrating over x_α we have

$$\begin{aligned}w_\alpha &= \sum_\beta w_\beta K_{\beta\alpha} \\ (\mathbf{K})_{\beta\alpha} &= K_{\beta\alpha} = \int_{\Gamma_\beta} \int_{\Gamma_\alpha} \mathrm{d}x_\beta \cdot \mathrm{d}x_\alpha \cdot \exp\left(-\frac{U(x_\beta)}{kT}\right) K_{\beta\alpha}(x_\beta, x_\alpha)\end{aligned} \tag{21}$$

The first line of eq 21 is a linear matrix equation for the coefficients, w_α. The matrix element $K_{\beta\alpha}$ is estimated from short trajectories between the milestones. For convenience we also denote the full matrix with bold face, $\mathbf{K}$, of dimension $M \times M$ where M is the number of milestones. The physical interpretation of $K_{\beta\alpha}$ is the probability that a thermal trajectory initiated in Milestone β will hit milestone α before any other milestone.

For computational purposes we define two types of transition kernels. The first, $\mathbf{K}_\mathrm{C}$, is set with cyclic boundary condition. Trajectories that enter the product state are immediately returned to the reactant. This kernel conserves probability and is appropriate to study stationary flux. The second kernel, $\mathbf{K}_\mathrm{A}$, is absorbing or terminating; every trajectory that enters the final

DOI: 10.1021/acs.jpcb.5b05467
J. Phys. Chem. B XXXX, XXX, XXX−XXX

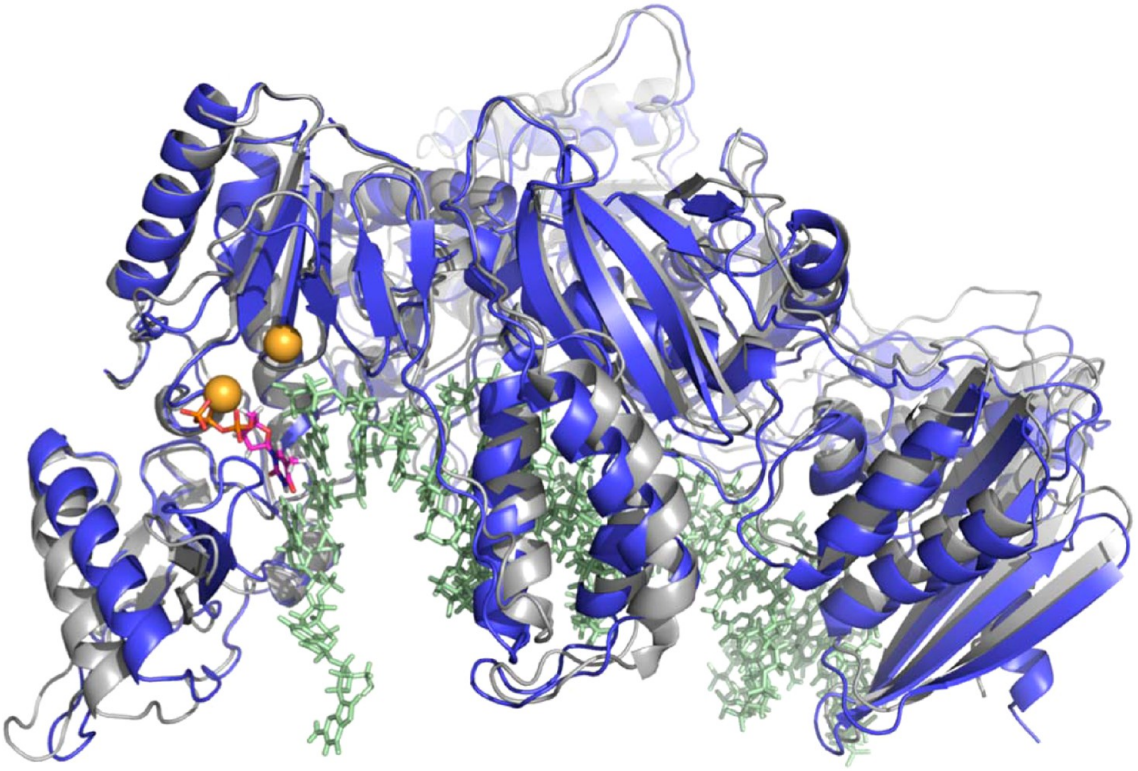


Figure 3. Molecular structure of the HIV reverse transcriptase that is used in molecular dynamic simulations. The protein−DNA complex is shown in open (gray) and closed (blue) forms. Shown as colored spheres are magnesium ions. The incoming nucleotide bound to the active site is shown with purple-orange sticks.

state disappears. The absorbing kernel does not conserve probability and is appropriate for the study of the mean first passage time as we outline below.

We also define the lifetime of the milestone, t_β as follows: Given that a trajectory started from a point x_β in Milestone β, we ask what is the average time for this trajectory to hit for the first time a milestone different from β.

We note that the kernel with periodic boundary conditions is a probability function and is normalized as

$$\sum_\alpha \int_{\Gamma_\alpha} \int_0^\infty K_{C,\beta\alpha}(x_\beta, x_\alpha, t)\mathrm{d}x_\alpha \mathrm{d}t = 1 \tag{22}$$

The lifetime is therefore (written for a point in a milestone and for an average over a milestone)

$$t_\beta(x_\beta) = \sum_\alpha \int_{\Gamma_\alpha} \int_0^\infty \mathrm{d}x_\alpha \mathrm{d}t \cdot t \cdot K_{C,\beta\alpha}(x_\beta, x_\alpha, t)$$

$$t_\beta = \int_{\Gamma_\beta} \int_{\Gamma_\alpha} \int_0^\infty \mathrm{d}x_\beta \mathrm{d}x_\alpha \mathrm{d}t \cdot t \cdot K_{C,\beta\alpha}(x_\beta, x_\alpha, t) \tag{23}$$

Given the stationary flux and the milestone lifetime, we have shown[7] that the stationary probability that the last milestone crossed and the free energy of α are

$$p_{\beta,\mathrm{stat}} = \int_{\Gamma_\beta} \mathrm{d}x_\beta \cdot t_\beta(x_\beta) q_{\beta,\mathrm{stat}}(x_\beta)$$

$$F_\beta = -kT \log\left[\int_{\Gamma_\beta} \mathrm{d}x_\beta \cdot t_\beta(x_\beta) q_{\beta,stat}(x_\beta)\right] \tag{24}$$

The final formula for this section is for the overall mean first passage time (definition 5), which is derived under similar assumptions as for eq 19

$$\langle \tau \rangle_{i\to f} = \sum_{\alpha\beta} p_\alpha(t=0)(\mathbf{I} - \mathbf{K}_\mathrm{A})^{-1}_{\alpha\beta} t_\beta \tag{25}$$

Note the use of an absorbing kernel in the derivation of the MFPT.

Equations 24 and 25 are the central expressions of this section.

The models of the open and closed forms of the enzyme were based on the structures 1RTD[23,24] and 1J5O,[24] respectively, from the Protein Data Bank.[25] The molecular model of the enzyme−DNA complex with incoming nucleotide is shown in Figure 3. The molecular modeling of the enzyme and DNA complex with an incoming nucleotide was discussed in detail in ref 1. This section discusses the modeling of the two other molecular systems: enzyme−DNA complex with no bound nucleotide (NN), and the complex after correct nucleotide undergoes chemical reaction denoted as After Chemistry (AC).

To model NN we use the existing crystal structures in the two end points mentioned above. We remove the matching nucleotide (here, TTP) from the crystal structure of the closed complex and perform a short minimization to relax the structure to a local energy minimum. The open state 1J5D does not have an incoming nucleotide bound so we used it as is. To model the AC structure, we added the matching nucleotide (thymine in our case) to the growing strand and removed the pyrophosphate group and the two catalytic magnesium ions at the active site. These models were used to compute the minimum free energy paths for open to closed transition of HIVRT. The paths were then used to compute the thermodynamics and kinetics of the conformational transitions in different substrate conditions.

DOI: 10.1021/acs.jpcb.5b05467
J. Phys. Chem. B XXXX, XXX, XXX−XXX

All calculations were performed with the MOIL suit of programs.[13] The OPLSAAL all-atom force field was used to model the protein and the nucleic interactions.[26,27] As detailed in ref 28, bonding terms of the nucleic acids were adopted from AMBER f99.[29] The simulated system consists of the enzyme (984 residue) in complex with 25bp DNA duplex with a sequence of 5′-GCCTCGCAGCCGTCCAACCAACTCA-3′ base pair with 3′-CGGAGCGTCGGCAGGTTGGTTGAG-TAGCAGCTAGGTTACGGCAGG-5′. The entire complex is embedded in 40 300 TIP3P water molecules. A periodic rectangular box of 108.5 × 108.5 × 118.5 A^3 was used. Magnesium and chloride ions were added to ensure neutrality and to mimic experimental conditions of concentration of 50 mM. The geometry of individual water molecules was fixed with matrix SHAKE.[30] Particle mesh Ewald summation[31] was used for long-range summation of electrostatics with a grid size of 64 Å in each direction. Velocity scaling was applied to keep the temperature constant at 311 K and configurations perpendicular to the reaction coordinate were sampled using Lagrangian constraints as implemented in ref 32. Initial phase space points were sampled at constant temperature, i.e., from the NVT ensemble. Short trajectories initiated from the sample points to estimate the transition kernel matrix followed the Newton's equations of motion.

We constructed the transition path that minimizes the free energy barrier using the string method. As an initial guess for NN and AC paths, we used our path for the correct match computed in ref 1. Each configuration of the enzyme along the discrete representation of the path in coarse space is modified to NN or AC condition and then solvated with water and ions. Solvated configurations were further equilibrated for about 200 ps by freezing the enzyme−DNA complex but allowing the solution atoms to move. We used the Cartesian coordinates of the α-carbons of the enzyme and all heavy atoms of the fingers domain (1−85, 115−150) as our coarse variable set. We evolved the configurations at constant temperature in the hyperplane orthogonal to the path[32] sampling conformations from canonical ensemble. We computed the average force perpendicular to the path in 5 ps time intervals. The average force was used to quench the path and adjust the distances between the points (eq 14). The path is updated about ∼250 times to achieve a converged path.

The calculations of the optimal curve (eq 15) are meaningful if the distances separating sequential points along the path are small such that the variations in the sequential forces are small as well. That is, the finite difference approximation of the curve is accurate. One way of avoiding a particularly large distance is to spread the points uniformly along the curve, which minimizes the worst distance. Uniform density of points along the path was proposed and enforced in the past by the addition of equidistance restraints,[21] Lagrange multipliers,[9] or path reparameterization.[33] We combined the use of restraints, which are more stable numerically, with a refinement step of the reparameterization.

To ensure the convergence of the path we computed the RMSD (root mean square deviation) of the path after each update. We stopped adjusting the path when the change in RMSD (root mean square difference) between two paths in a sequence was smaller than 0.001 Å.

Three-nanosecond simulations at the hyperplane normal to the reaction coordinate (the milestone) were used to sample configurations. Snapshots were saved every 0.5 ps and were examined to check if they are sampled from first hitting point distribution (FHPD). The FHPD points are identified by integration backward in time and verifying that the first milestone crossed during the backward integration is different from the originating milestone. If the trajectory first hits the initiating milestone, it is not sampled from FHPD and is removed from the statistics. The acceptance ratio varied from a milestone to a milestone and was between 0.1 and 0.9. Verified FHPD points were then integrated forward in time until they hit for the first time another milestone.[16] In the forward integration we allow for recrossing of the originating milestone. The typical time for the termination is about 10−30 ps. The identities of the initiating and terminating milestones were recorded as well as the lifetime of the trajectories. At each milestone a minimum of 200 trajectories are used to estimate the matrix element of the transition kernel.

We estimate the free energy change along the reaction coordinate by Milestoning and by integration of the mean force eqs 16 and 24. These two calculations are possible in a single Milestoning calculation since the sampling at the milestone for initial conditions for trajectories can be used to estimate the average force in the hypersurface orthogonal to the reaction coordinate. Estimating the free energy profile by two different approaches enhances the confidence in the conclusions of the computations.

III. RESULTS

In Figure 4, we summarize the main result of this article on the overall free energy landscapes of different binding modes of the ligand. The free energy profiles presented are computed using integration of the PMF. All the curves are matched at the left with the state of physical binding of the nucleotide to the open state. While differences in physical binding can be significant and will be addressed in future work, we focus in the present study on the conformational transition step. Interestingly, the right side of the free energy profile of binding a mismatch nucleotide (dATP, blue curve) is the highest of the curves. This makes the closed state of a bound mismatch unlikely, even with respect to the enzyme with no nucleotide bound. A conformational transition from an open to closed state without the presence of a ligand (red curve) is almost free-energy neutral. There is no strong preference of the enzyme for the open or closed state in the absence of ligand according to our findings. Correct ligand binding makes a significant change in free energy landscape shifting the equilibrium to the closed state (black curve).

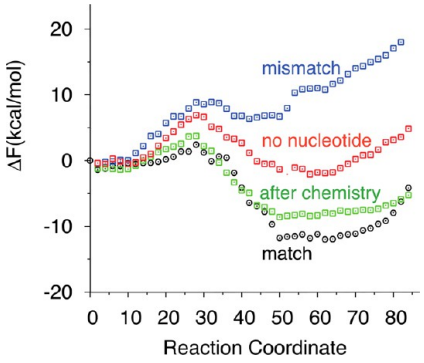


Figure 4. Potential of mean force along the open to closed transition of the HIV RT. Each color represents the PMF with a different substrate: bound correct nucleotide TTP (Black), mismatch dATP (Blue), correct nucleotide after addition to the growing strand and release of pyrophosphate group (Green); and no nucleotide (Red) .

DOI: 10.1021/acs.jpcb.5b05467
J. Phys. Chem. B XXXX, XXX, XXX−XXX

Addition of the nucleotide to the growing strand results in a moderate change of the free energy profile at the closed state (green curve). Release of the pyrophosphate group together with two magnesium ions at the conserved ion binding pockets also reduces the free energy barrier going from open to closed, causing a rapid collapse to the closed state when the enzyme is open, and at the same time, the absence of magnesium ions at the active site reduces the stability of the closed state relative to the enzyme with the correct nucleotide bound. However, the changes are still not enough to shift the equilibrium toward the opening after chemistry. Hence the translocation of the DNA after the chemical reaction may be necessary to destabilize the closed state.

In Figure 5 we compare the free energy profiles that are obtained by the two different computational approaches

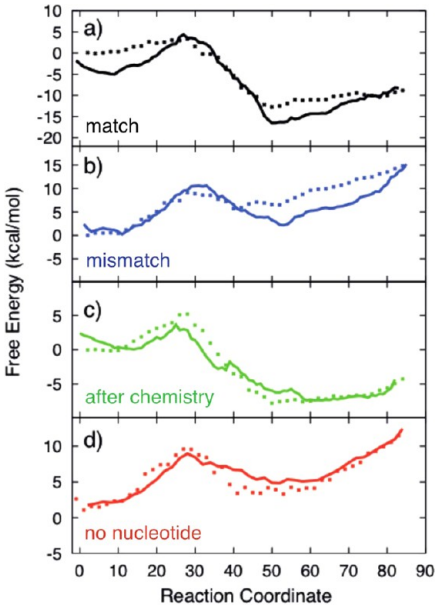


Figure 5. Comparison of the free energy calculations along the enzyme conformational transition pathway with different binding modes of incoming substrate using the two different methods: Points are computed from the average of the projection of force along the path (PMF). Solid lines are the free energy estimates from the stationary probability of milestones computed from unconstrained molecular dynamic simulations. The plots correspond to different substrates: a, correct nucleotide bound; b, mismatch bound; c, correct nucleotide after pyrophosphate release; d, no incoming nucleotide bound.

described in the Methods section. The widely used protocol is the integration of PMF along a reaction coordinate, eq 16, which we showed already in Figure 4. The second approach is based on the stationary flux in the Milestoning formulation, eq 24. The two entities are not exactly the same since in Milestoning the probability is defined by the last milestone that was passed by a trajectory (and not an average over an interval). However, for sufficiently small displacements between the milestones which is about 0.05 Å in our case we expect them to be similar, as we illustrate for a dipeptide in ref 14. Of course, the displacement should not be too small to violate the decorrelation assumption of the current Milestoning algorithm.

Besides the definition of free energy, the theories behind the calculations are markedly different. The PMF is based on equilibrium sampling of configurations, while Milestoning is based on kinetic analysis of trajectory fragments. The observation that the long term behavior of the kinetic matrix agrees with straightforward equilibrium calculations is a useful testimony for the convergence of the calculations.

In Table 1 we provide the rate constants for the conformational transitions computed by the Milestoning theory. In

Table 1. Rate Coefficients for the Transition between Open and Closed States (k_2 and k_{-2}) Computed by Milestoning[a]

substrate	k_2 (s^{-1})	k_{-2}(s^{-1})
correct	2500−20000 (2000)	40 (3)
mismatch	200−400 (>500)	4000 (>1200)
no-nucleotide	100−2000	40000
after chemistry	8 × 10$\hat{6}$	4400

[a]The experimental results are provided in parentheses when available.

parentheses we give the experimental rate coefficients, when available. Figure 6 shows the mean first passage time, computed with Milestoning, starting from the reactant and terminating at a specific position along the reaction coordinate.

We simulated the opening of the complex, after the chemical step is completed, by adjusting the structures of the path we defined previously.[1] We started from the closed state structure with a bound nucleotide and completed the bond between the 3′OH end of the primer strand and incoming nucleotide. We then removed the pyrophosphate and two metal ions, and used the string method to refine the path starting from the previously computed reaction coordinate for the correct nucleotide computed previously. With the reaction coordinate at hand we used Milestoning to compute kinetics and thermodynamics of the process going from an open to a closed state with the elongated DNA occupying the nucleotide binding site. The order of events in the pathway involving release of the bound metal ions and the pyrophosphate and the opening of the enzyme is not known. For this analysis we modeled the reaction as sequential with release of pyrophosphate preceding opening. First, the small molecules leave with no significant change in the protein structure, and second the conformational transition takes place without the presence of the ion or the leaving group. Our MD simulations suggest that opening is fast (4400 s^{-1}) relative to the rate of opening before chemistry (40 s^{-1}). In future work, we will consider the alternative reaction sequence with opening preceding pyrophosphate release.

To get further insight into the free energy profiles and to the molecular factors that determine the dramatic changes in the rate and equilibrium binding of nucleotide to the HIV RT, we exploit the exact partitioning of the mean force potential that was discussed earlier: If the molecular force is written as a sum, then the mean force potential along the reaction coordinate can also be written as a sum of the same type of terms (eq 16).

Figure 7 shows the results of the subdivision of the protein into individual domains thought to make important contributions to the free energy of the open to closed transition. In particular, for each condition we compare the contributions of the fingers domain (residues 1−88, 121−146) to the results obtained for the whole protein and the whole protein minus the fingers domain. In addition, we also consider the contribution of a small subset of key amino acids in the two loops of the fingers domain; the loops connecting $\beta3-\beta4$ and $\beta7-\beta8$ (residues 65−74 and 137−144). Figure 7a shows the partition of free energy change from open to the closed with correct nucleotide bound. Interestingly, the results indicate that the fingers domain contributes about half of the net free energy difference between open and closed states,

DOI: 10.1021/acs.jpcb.5b05467
J. Phys. Chem. B XXXX, XXX, XXX−XXX

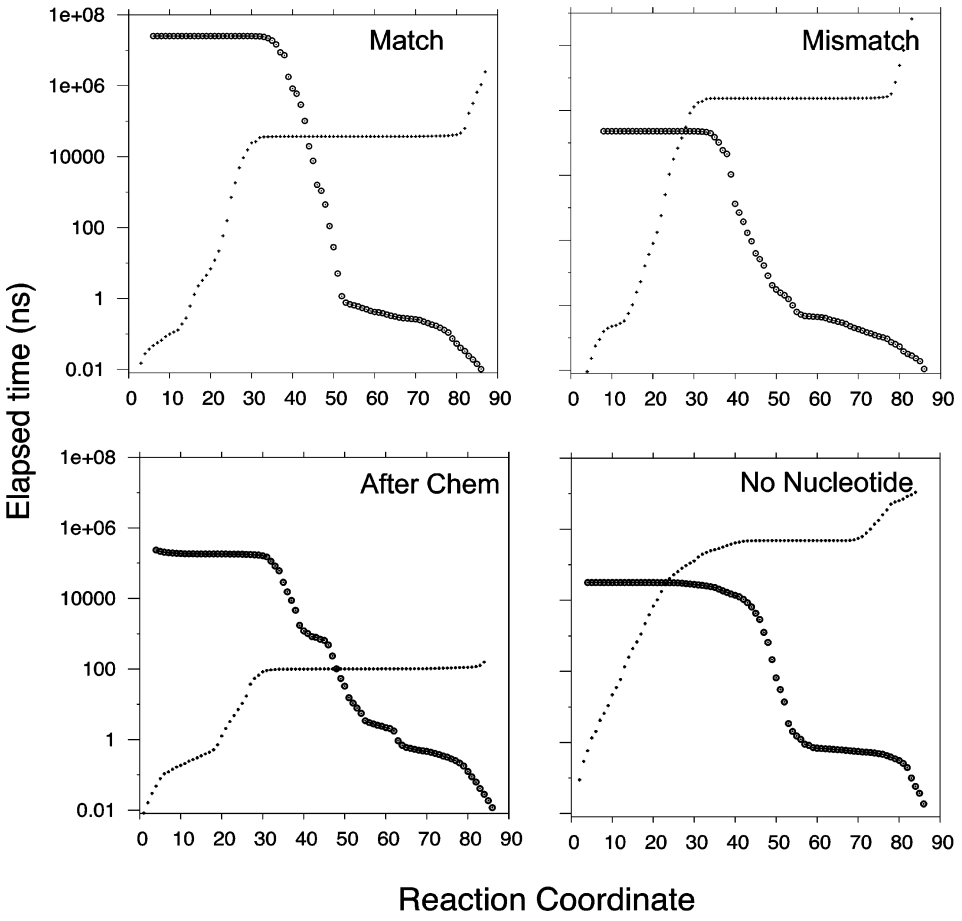


Figure 6. Mean first passage time for the conformational transition of HIV RT between two functional states with different substrate binding states. Open state is represented by 1 in the reaction coordinate while the closed structure is 85. Small dots show the elapsed time from open to closed while circles are the time for the reverse transition.

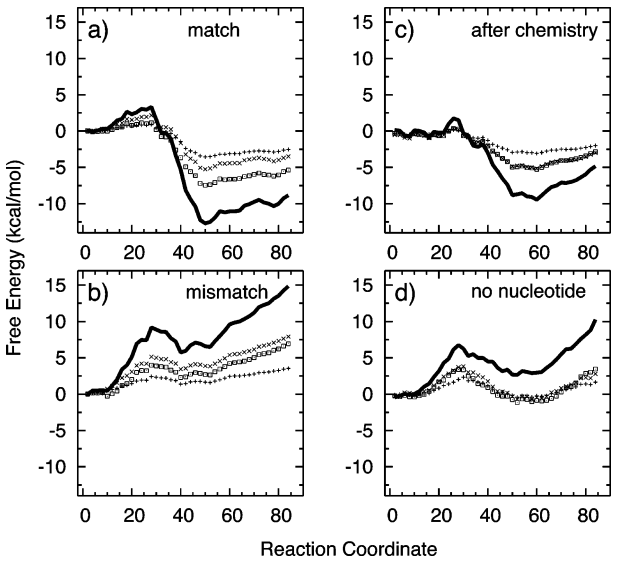


Figure 7. PMF for the systems studied (with the same order as in Figure 5). Solid lines are the PMF for the whole enzyme (residues 1−984). Points are the contributions from a subset of particles from the enzyme; (×) represents the contribution from fingers domain (residues 1−88, 121−146), (□) is the whole enzyme other than fingers domain, and (+) is the contribution from a subset of fingers domain (65−74, 137−144).

while changes in structure in the remainder of the protein contribute the rest. It is obvious that the nearest residues to the active site (finger domain) will have a significant contribution to the free energy profile. However, the significant contribution of the rest of the protein is surprising.

Figure 7b shows the results obtained with a mismatched bound nucleotide. Here, contributions of the fingers domain and the remainder of the protein to the unfavorable free energy difference between the open and closed states are approximately equal. In addition, the fingers domain and the remainder of the protein contribute comparable amounts to the free energy barrier at the transition state. The important conclusion of these studies is that the fingers domain and the remainder of the protein each contribute to the mismatch recognition, and this implies that a mismatched nucleotide−template interaction not only causes changes in protein structure at the local level due to changes in the fingers domain interaction energies, but also causes global changes in protein structure. These changes were already reported in our earlier work.[1] Figure 7c shows the results obtained for the open to closed transition after chemistry, that is, with the DNA primer terminus occupying the nucleotide-binding site. Again the contributions of the fingers domain and the remainder of the protein are approximately equal for both the net free energy difference and the activation barrier. Thus, both local and global protein and DNA structure elements tend to stabilize the closed state. Finally, in Figure 7d, we consider the change from the open to closed state in the absence of added

DOI: 10.1021/acs.jpcb.5b05467
J. Phys. Chem. B XXXX, XXX, XXX−XXX

Table 2. Residues that Contribute Most to the PMF[a]

correct			mismatch			no-nucleotide			after-chem		
Res ID	TS	Closed	Res ID	TS	Closed	Res ID	TS	Closed	Res ID	TS	Closed
N137	0.2	−1.1	*K66*	0.5	0.45	*K67*	0.75	0.4	*N137*	0	−0.7
K66	0.15	−0.5	*W71*	0.4	0.45	*N137*	0.4	−0.3	*K66*	0	−0.5
E138	0.05	−0.3	*N137*	0.25	0.4	*E138*	0.2	0.35	*S68*	0	−0.4
K32	0.15	−0.15	*R72*	0.2	0.4	*C38*	0.05	−0.25	*K65*	0	−0.3
K65	0.05	−0.25	*E138*	0.2	0.3	*K65*	0.35	0.2	*T139*	0	0.2

[a]We report the change in the free energy from open to transition state (TS) and from open to closed state. The locations of these residues are depicted in Figure 6. All values are in kcal/mol.

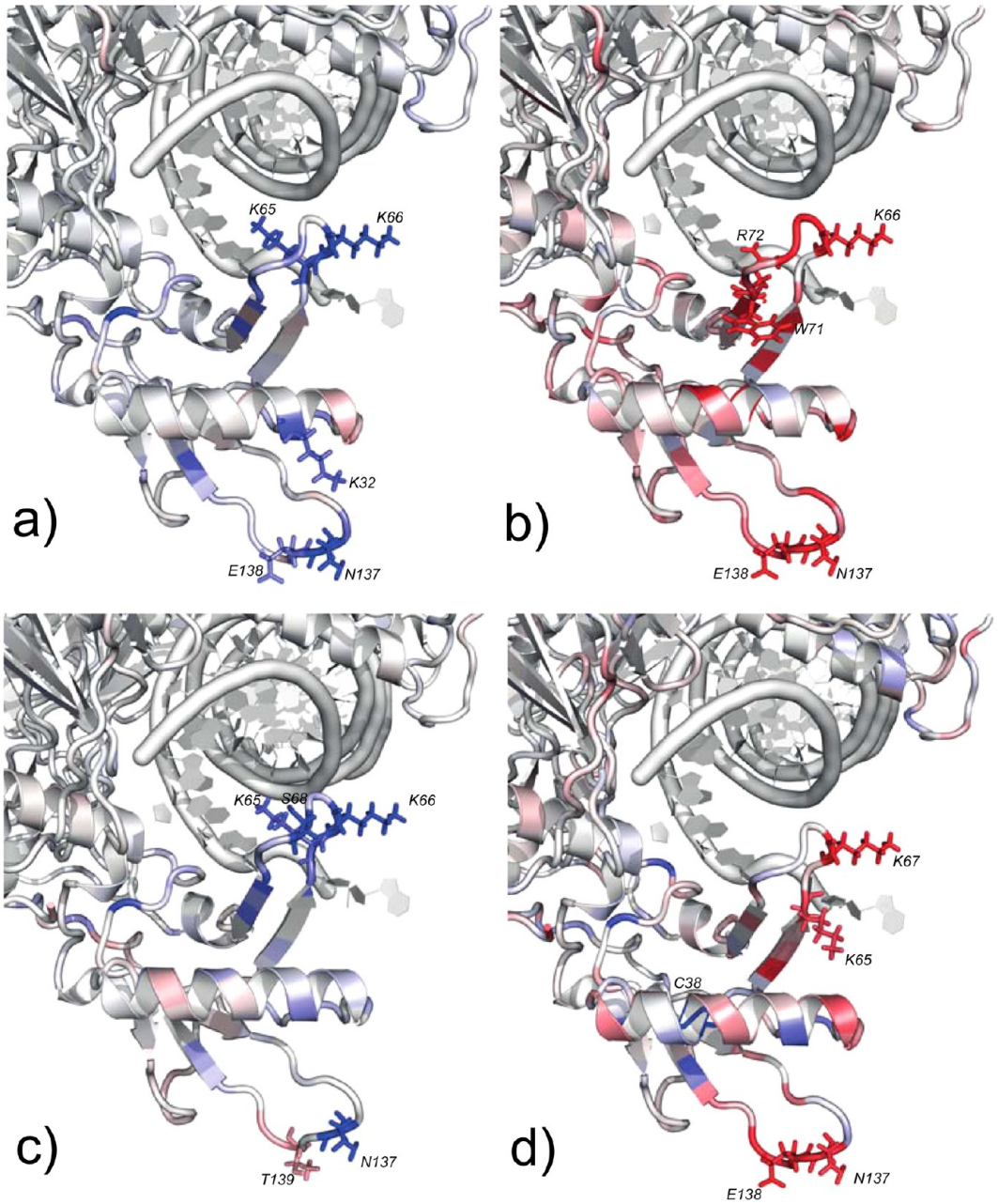


Figure 8. Contribution of each residue to the free energy difference going from open to closed states of the enzyme. Color changes in red−white−blue with red being free energy contribution of ≥ + 0.5 kcal/mol to the closing conformational change, white being 0.0 kcal/mol, and blue ≤ − 0.5 kcal/mol. Contributions of individual residues to the overall free energy changes with substrate states studied. The contributions are depicted on the open state structure of the enzyme for comparison. (a) Free energy difference when a correct match is bound; (b) for mismatch; (c) after the chemical reaction of the correct substrate; and (d) in the absence of incoming nucleotide. Residues that contribute the most to the free energy change summarized in Table 2 and their side chains are drawn here explicitly in stick representation.

DOI: 10.1021/acs.jpcb.5b05467
J. Phys. Chem. B XXXX, XXX, XXX−XXX

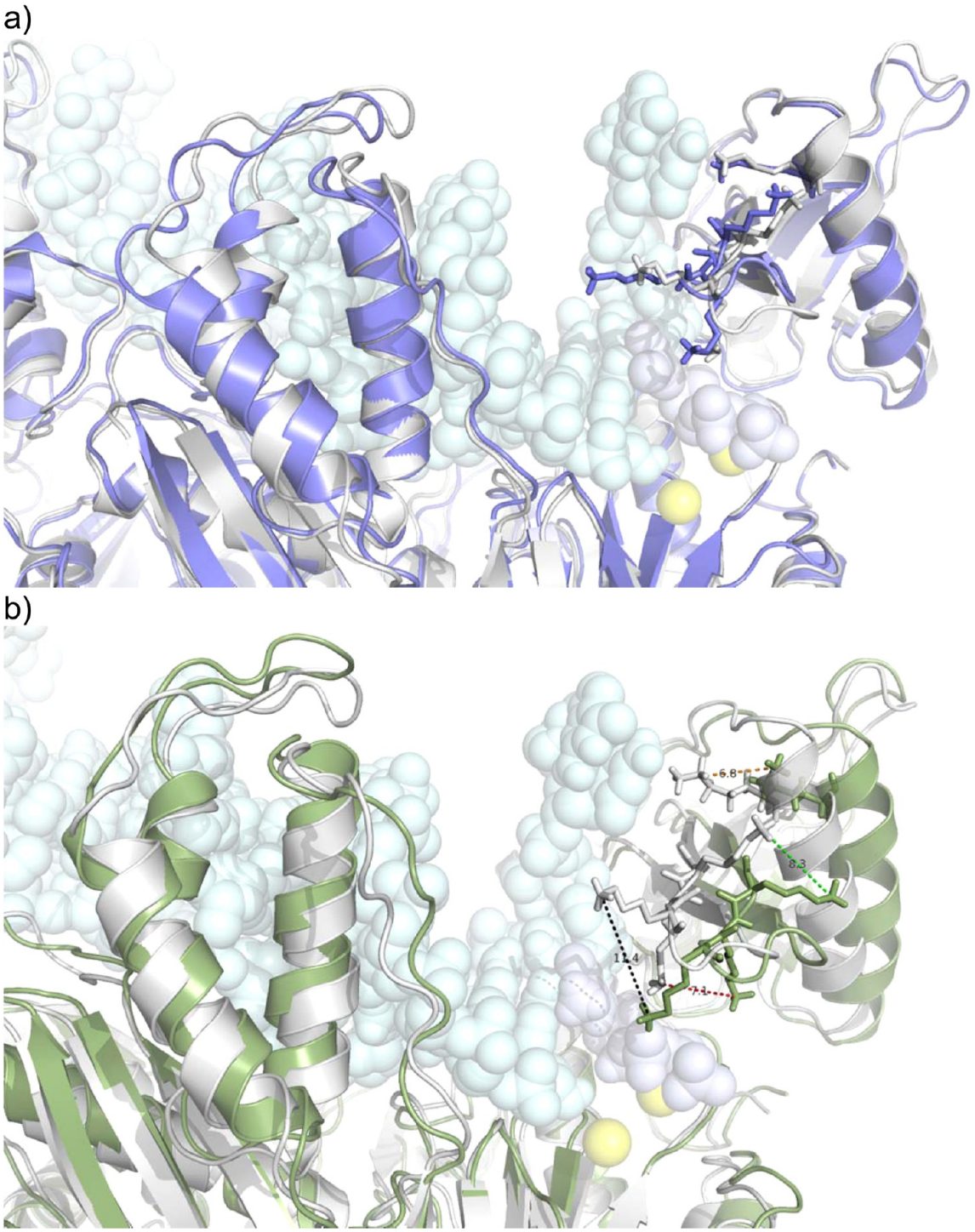


Figure 9. Structures selected from the transition state (TS) shown in Figure 5 for different substrate binding states. Gray is the structure of protein when the ligand is correct in comparison when the enzyme (a) after chemical reaction and (b) without a nucleotide bound. Lysine side chains in the fingers domain show the most variation and they are shown with sticks for comparison.

nucleotide, and again, the changes in free energy contributed by the fingers domain are approximately equal to the contributions of the remainder of the protein.

Clearly, these results indicate that motions throughout the protein are coupled to the reorganization of the residues in the immediate vicinity of the active site. Long distance interactions may act to gate the transition from the open to the closed state and could reflect the interaction of the DNA primer/template with the protein. For example, it is known that mismatches in the primer/template slow the rate of incorporation of a correct base pair.[34−37] This could be accomplished through these long-range interactions, or could be a result of misalignment of active site residues caused by the altered structure of the mismatched base pair in the DNA. Further experimental and computational analysis will be required to address the effect of mismatches in the DNA on nucleotide-induced conformational changes and incorporation. Our analysis provides novel predictions that will need to be tested experimentally in future work.

DOI: 10.1021/acs.jpcb.5b05467
J. Phys. Chem. B XXXX, XXX, XXX−XXX

Partitioning of the PMF calculations provide an estimate of the contributions of individual amino acids to conformational change in enzyme structure. In Table 2, we list the free energy contributions of the most important amino acids, both at the transition state (TS) and to the net free energy difference between open and closed states. In this list, some of the amino acids are in line with expectations regarding the importance of charged residues at the active site such as R72, K65, and K66. Other predictions are unexpected, but could be rationalized post hoc, and will require further experimental tests. In particular, during the open to closed transition, W71 slides along a hydrophobic face of the αA-helix. This motion was not noticed from examination of static structures, but is quite evident in a movie generated by the minimum free energy pathway. Charged residues N137 and E138 interact with the charged groups of αA-helix and come in close contact while enzyme closes. Figure 8 shows the location and contribution of each of these residues to the overall free energy change as a heat map. One striking result is that the side chains far from the active site play a significant role in nucleotide recognition and stabilization of the incoming cognate nucleotide, while the same residues destabilize non-cognate one leading to the opening of the complex.

In addition to the free energy contributions, we show snapshots of structures sampled from the transition state ensembles for the two different modes of binding considered (see Figure 9). Transition state (TS) structures are compared with the cases when the correct nucleotide is bound, which was described in detail in our earlier work.[1] TS structure of correct substrate is similar even after the chemical reaction (Figure 9a). In the TS ensemble, positively charged groups bring together the finger domain in close contact with the DNA and terminal nucleotide. Interestingly, a significant difference is observed in the TS structures of the transition in no nucleotide bound transition when compared with the correct substrate bound (see Figure 9b). The most significant difference is the displacement of the observed lysine side chains. Lysine chains did not align toward the DNA nor come close enough, an observation reported earlier for the mismatch.[1] Perhaps not surprisingly, this gives a dramatic change in the local charge distribution around the active site and caused the slow rate of closing reported here. Also αA-helix that is sandwiched by the two loops connecting β3−β4 and β7−β8 stayed farther from the active site relative to the correct nucleotide.

IV. DISCUSSION AND CONCLUSIONS

In summary, our investigation of the conformational transition (step 2) includes now the following scenarios

$$\begin{aligned} &\text{(a)}\ ED_nN_C \rightleftharpoons E'D_nN_C \\ &\text{(b)}\ ED_nN_I \rightleftharpoons E'D_nN_I \\ &\text{(c)}\ ED_n \rightleftharpoons E'D_n \\ &\text{(d)}\ E'D'_{n+1} \rightleftharpoons ED'_{n+1} \end{aligned} \tag{26}$$

where ED_n represents a complex of enzymes with DNA containing n nucleotides, E′ represents the closed enzyme state, D′ represents DNA where the primer terminus occupies the nucleotide binding site (the pretranslocation state), N_C represents a correct nucleotide, and N_I represents an incorrect nucleotide.

In scenario (a) we considered the conformational transition with a correct ligand bound to the protein, in (b) with an incorrect ligand bound, in (c) with no ligand bound, and in (d) after the chemical step, but without the displacement of the DNA to free space to a new substrate. Step (c) is of particular interest from the perspective of mechanisms. Calculation (c) helps differentiate between a sequential binding mechanism and a mechanism of a conformational selection in which the conformational change precedes substrate binding. The conformational selection picture assumes a rapid equilibrium between the different states of the proteins (open or close), which is accomplished without the presence of the substrate. If step (c) is indeed much faster than step (a) or (b), then conformational selection is a possible mechanism, but one key determinant is the rate of nucleotide binding to the open state, which we have not yet addressed. However, we have measured the maximum rate of the observed conformational change with different ligands yielding different rates,[38,39] which argues against the conformational selection mechanism that predicts that the maximum rate of the observed conformational change will be independent of the ligand.

Step (d) is concerned with the completion of a cycle of the molecular machine. For the enzyme to be ready to accept another incoming nucleotide after the chemical step, the complex must transition to the beginning of the cycle. This transition includes two steps: The DNA must translocate further from the finger domain to free up the nucleotide binding site for the incoming nucleotide, and a second step involves the opening of the finger domain of the protein so that a new ligand can bind. But what is the order of events? Does the DNA slide while the protein is still in the closed form, or does it slide after the protein changes its conformation to the open state? Scenario (d) prepares the ground for a fuller investigation of this question. Here we consider the feasibility of sliding after the closed to open transition. The reverse of the scheme in (c) is indeed similar to the second scenario: there the transition is after the DNA is translocated and ready for the incoming nucleotide binding and suggests that translocation should proceed chemistry to destabilize the closed state and afford rapid opening of the enzyme for a new nucleotides.

Our analysis provides a novel insight into the molecular details underlying substrate recognition by HIV reverse transcriptase. By comparing the molecular trajectories of the open to closed transitions in the presence and absence of nucleotide, with a mismatched nucleotide and the process after the chemical reaction, we have revealed new features of the complex reactions' energy landscape. First, it is clear that the binding of a correct nucleotide precedes the open to closed transition in keeping with an induced-fit model. A view of the induced fit model according to the present calculations is a shift or tilting of the free energy landscape of the protein that follows the binding of the substrate. The tilting pushes the protein to a new alternative structural state, more stable from free energy perspective than the stable state of the unbound form, and hence it induces the conformational transition. Without the presence of the ligand, weakly attached to the surface of the protein, the energy landscape would have been less likely to motivate the conformational transition. Of course, for the case of enzymatic reaction, the open state of the unbound form must be present with high probability. If it is not open, the ligand may not be able to enter the active site and bind. Indeed Figure 4 indicates that the transition is more likely to occur after the substrate binds physically at the surface of HIV-RT. The closed form is about 0.5−1 kcal/mol higher in free energy than the open form in the

DOI: 10.1021/acs.jpcb.5b05467
J. Phys. Chem. B XXXX, XXX, XXX−XXX

absence of a substrate. The same form is more stable (lower in free energy) by about 10 kcal/mol once the ligand is bound.

Our study also provides a method to estimate the free energy contribution of each individual amino acid to the observed free energy profile. The assumption is that the reaction coordinate following point mutations of the HIV RT protein is stable and the changes can be thought of as perturbations. Hence the mechanism of the conformational transition is not changing significantly. This analysis makes predictions that can be approached by site-directed mutagenesis, but experimental data must also be interpreted with caution. Conservative mutations to disrupt hydrogen bonds, ionic interactions or hydrophobic interactions generally lead to changes in apparent free energy (as a $\Delta\Delta G$ based upon changes in rate or equilibrium constants) in the range of 1−3 kcal/mol, and these results overestimate the net contribution of an individual interaction due to secondary effects of the amino acid substitution. For example, the sum of the effects of mutations of all residues surrounding the active site of tyrosyl tRNA synthetase leads to a net change in free energy, which is twice that measured directly.[40] In our analysis, each amino acid is predicted to contribute less than 1 kcal/mol of interaction energy so that the sum of all interactions is consistent with the net free energy difference. Further experimental and computational approaches are required to resolve the contributions of individual amino acids, and the current results present a novel approach that has the potential to provide new insights into protein structure/function relationships. For example, in some cases single or multiple amino acid substitutions in HIV reverse transcriptase lead to resistance to nucleoside analogues used to treat HIV infections. The computational methods hold the potential to provide new theories to understand the changes in protein structure leading to resistance based upon computation of the free energy profile comparing the native nucleotides with the nucleotide analogs for wild-type and mutant forms of the enzyme. This extensive computational effort will be facilitated by PMF methods.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpcb.5b05467.

Open to closed transition of HIV reverse transcriptase along the minimum free energy path is shown with a supplementary movie. Residues represented in stick are key residues in the motion for the correct incoming nucleotide (see main text for more detail) (ZIP)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: ron@ices.utexas.edu.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This research was supported by grants from the NIH GM59796 and Welch Foundation F-1783 to RE, and by grants from the NIH GM084741 and Welch Foundation F-1604 to K.A.J. and AD181 faculty research grant to S.K.

■ REFERENCES

(1) Kirmizialtin, S.; Nguyen, V.; Johnson, K. A.; Elber, R. How Conformational Dynamics of DNA Polymerase Select Correct Substrates: Experiments and Simulations. *Structure* **2012**, *20*, 618−627.

(2) Johnson, K. A. Conformational Coupling in DNA Polymerase Fidelity. *Annu. Rev. Biochem.* **1993**, *62*, 685−713.

(3) Prasad, B. R.; Kamerlin, S. C. L.; Florian, J.; Warshel, A. Prechemistry Barriers and Checkpoints do not Contribute to Fidelity and Catalysis as Long as They are not Rate Limiting. *Theor. Chem. Acc.* **2012**, *131*, 1288.

(4) Briggs, G.; Haldane, J. B. S. A Note on The Kinetics of Enzyme Action. *Biochem. J.* **1925**, *19*, 338−339.

(5) Elber, R.; Kirmizialtin, S. Molecular Machines. *Curr. Opin. Struct. Biol.* **2013**, *23*, 206−211.

(6) Kirmizialtin, S.; Elber, R. Revisiting and Computing Reaction Coordinates with Directional Milestoning. *J. Phys. Chem. A* **2011**, *115*, 6137−6148.

(7) Bello-Rivas, J. M.; Elber, R. Exact Milestoning. *J. Chem. Phys.* **2015**, *142*, 094102.

(8) Boresch, S.; Archontis, G.; Karplus, M. Free-Energy Simulations - the Meaning of the Individual Contributions from a Component Analysis. *Proteins: Struct., Funct., Genet.* **1994**, *20*, 25−33.

(9) E, W.; Ren, W. Q.; Vanden-Eijnden, E. String Method for the Study of Rare Events. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2002**, *66*, 052301.

(10) Maragliano, L.; Fischer, A.; Vanden-Eijnden, E.; Ciccotti, G. String Method in Collective Variables: Minimum Free Energy Paths and Isocommittor Surfaces. *J. Chem. Phys.* **2006**, *125*, 024106.

(11) Ulitsky, A.; Elber, R. A New Technique to Calculate Steepest Descent Paths in Flexible Polyatomic Systems. *J. Chem. Phys.* **1990**, *92*, 1510−1511.

(12) E, W.; Ren, W. Q.; Vanden-Eijnden, E. Finite Temperature String Method for the Study of Rare Events. *J. Phys. Chem. B* **2005**, *109*, 6688−6693.

(13) Elber, R.; Roitberg, A.; Simmerling, C.; Goldstein, R.; Li, H. Y.; Verkhivker, G.; Keasar, C.; Zhang, J.; Ulitsky, A. MOIL: A Program for Simulations of Macrmolecules. *Comput. Phys. Commun.* **1995**, *91*, 159−189.

(14) West, A. M. A.; Elber, R.; Shalloway, D. Extending Molecular Dynamics Time Scales with Milestoning: Example of Complex Kinetics in a Solvated Peptide. *J. Chem. Phys.* **2007**, *126*, 145104.

(15) Majek, P.; Elber, R. Milestoning Without a Reaction Coordinate. *J. Chem. Theory Comput.* **2010**, *6*, 1805−1817.

(16) Kirmizialtin, S.; Elber, R. Revisiting and Computing Reaction Coordinate with Directional Milestoning. *J. Phys. Chem. A* **2011**, *115*, 6137−6148.

(17) Elber, R.; West, A. Atomically Detailed Simulation of the Recovery Stroke in Myosin by Milestoning. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 5001−5005.

(18) Elber, R. A Milestoning Study of the Kinetics of an Allosteric Transition: Atomically Detailed Simulations of Deoxy Scapharca Hemoglobin. *Biophys. J.* **2007**, *92*, L85−L87.

(19) Vanden-Eijnden, E.; Venturoli, M. Revisiting the Finite Temperature String Method for the Calculation of Reaction Tubes and Free Energies. *J. Chem. Phys.* **2009**, *130*, 194103.

(20) Czerminski, R.; Elber, R. Self-Avoiding Walk between Two Fixed End Points as a Tool to Calculate Reaction Paths in Large Molecular Systems. *Int. J. Quantum Chem.* **1990**, *38*, 167−186.

(21) Elber, R.; Karplus, M. A Method for Determining Reaction Paths in Large Molecules - Application to Myoglobin. *Chem. Phys. Lett.* **1987**, *139*, 375−380.

(22) Shalloway, D.; Faradjian, A. K. Efficient Computation of the First Passage Time Distribution of the Generalized Master Equation by Steady-State Relaxation. *J. Chem. Phys.* **2006**, *124*, 054112.

(23) Huang, H. F.; Chopra, R.; Verdine, G. L.; Harrison, S. C. Structure of a Covalently Trapped Catalytic Complex of HIV-I Reverse Transcriptase: Implications for Drug Resistance. *Science* **1998**, *282*, 1669−1675.

DOI: 10.1021/acs.jpcb.5b05467
J. Phys. Chem. B XXXX, XXX, XXX−XXX

(24) Sarafianos, S. G.; Das, K.; Clark, A. D.; Ding, J. P.; Boyer, P. L.; Hughes, S. H.; Arnold, E. Lamivudine (3TC) Resistance in HIV-1 Reverse Transcriptase Involves Steric Hindrance with Beta-Branched Amino Acids. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 10027−10032.
(25) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucl. Acid Res.* **2000**, *28*, 235−242.
(26) Kaminski, G.; Friesner, R.; Tirado-Rives, J.; Jorgensen, W. L. Evaluation and Reparameterization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides. *J. Phys. Chem. B* **2001**, *105*, 6474−6487.
(27) Pranata, J.; Wierschke, S. G.; Jorgensen, W. L. OPLS Potential Functions for Nucleotide Bases- Relative Association Constants of Hydrogen-Bonded Base Pairs in Chloroform. *J. Am. Chem. Soc.* **1991**, *113*, 2810−2819.
(28) Kirmizialtin, S.; Elber, R. Computational Exploration of Mobile Ion Distributions Around RNA Duplex. *J. Phys. Chem. B* **2010**, *114*, 8207−8220.
(29) Wang, J. M.; Cieplak, P.; Kollman, P. A. How Well Does a Restrained Electrostatic Potential (RESP) Model Perform in Calculating Conformational Energies of Organic and Biological Molecules? *J. Comput. Chem.* **2000**, *21*, 1049−1074.
(30) Weinbach, Y.; Elber, R. Revisiting and Parallelizing SHAKE. *J. Comput. Phys.* **2005**, *209*, 193−206.
(31) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald an N.Log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089−10092.
(32) Elber, R. Calculation of the Potential of Mean Force Using Molecular Dynamics with Linear Constraints -An Application to a Conformational Transition in a Solvated Dipeptide. *J. Chem. Phys.* **1990**, *93*, 4312−4321.
(33) E, W. N.; Vanden-Eijnden, E. Transition-Path Theory and Path-Finding Algorithms for the Study of Rare Events. *Annu. Rev. Phys. Chem.* **2010**, *61*, 391−420.
(34) Donlin, M. J.; Patel, S. S.; Johnson, K. A. Kinetic Partitioning Between the Exonuclease and Polymerase Sites in DNA Error Correction. *Biochemistry* **1991**, *30*, 538−546.
(35) Wong, I.; Patel, S. S.; Johnson, K. A. An Induced-Fit Kinetic Mechanism for DNA Replication Fidelity: Direct Measurement by Single-Turnover Kinetics. *Biochemistry* **1991**, *30*, 526−537.
(36) Patel, S. S.; Wong, I.; Johnson, K. A. Pre-Steady-State Kinetic-Analysis of Processive DNA-Replication Including Complete Characterization of An Exonuclease-Deficient Mutant. *Biochemistry* **1991**, *30*, 511−525.
(37) Kati, W. M.; Johnson, K. A.; Jerva, L. F.; Anderson, K. S. Mechanism and Fidelity of HIV Reverse Transcriptase. *J. Biol. Chem.* **1992**, *267*, 25988−25997.
(38) Kellinger, M. W.; Johnson, K. A. Nucleotide-Dependent Conformational Change Governs Specificity and Analog Discrimination by HIV Reverse Transcriptase. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 7734−7739.
(39) Kellinger, M. W.; Johnson, K. A. Role of Induced Fit in Limiting Discrimination Against AZT by HIV Reverse Transcriptase. *Biochemistry* **2011**, *50*, 5008−5015.
(40) Johnson, K. A.; Benkovic, S. J. Analysis of Protein Function by Mutagenesis. In *The Enzymes*; Sigman, D. S., Boyer, P. D., Eds.; Academic Press Inc.: San Diego, CA, 1990; Vol. *XIX*, pp 159−211.

DOI: 10.1021/acs.jpcb.5b05467
J. Phys. Chem. B XXXX, XXX, XXX−XXX