Framework Model for DNA Polymerases†

David J. Keller* and James A. Brozik

Molecular Machines Laboratory, Department of Chemistry, University of New Mexico, Albuquerque, New Mexico 87131

Received October 27, 2004; Revised Manuscript Received February 20, 2005

ABSTRACT: DNA polymerases are complex machines with both chemical and mechanical functions. Recent crystal structures, ensemble kinetics, and single-molecule investigations have helped to elucidate the main properties of several DNA polymerases, all of which share common structural elements and a common basic mechanism, despite wide variations in amino acid sequence. The framework model is intended to aid in the understanding of these common features (and differences). It defines a class of models that automatically incorporates most of what is known about DNA polymerases within a single theoretical structure so that it is easier to make comparisons between them and to generate detailed models for specific polymerases. The framework model has three main elements: (1) a set of four key variables that describe the important motions within the protein−DNA−nucleotide complex, (2) a complete set of conformational states for the protein−DNA−nucleotide system, and (3) an approximate potential energy surface that controls the motions and transition rates between states. As an example application, we use the general framework ideas to build a detailed model for the HIV reverse transcriptase that is consistent with existing data, and predicts force−velocity curves and stepping-statistics histograms that can be directly compared to experiment.

The current understanding of DNA polymerases comes mainly from two types of information: structural information from X-ray crystallography and dynamical information from ensemble enzyme kinetics. Of these two, the kinetic information is often more directly useful; kinetic properties include rates of nucleotide incorporation, processivity, replication fidelity, etc. But it is natural to regard the structural information as more fundamental: it tells us how polymerases are built from the atoms; the dynamical properties could in principle be calculated from structure. Predicting high-level kinetic and thermodynamic properties from structure is a major goal of biophysics.

The DNA polymerases are complex, mechanical, protein machines. Predicting their properties from first principles is not practical at present. This limits the value of the otherwise excellent structural information available for polymerases, and leaves much of the kinetic data, useful as it is in its own right, essentially unconnected to structure. To help bridge the gap between purely phenomenological kinetic models and atomistic calculations, an intermediate stochastic approach has recently been developed, most notably by Oster and co-workers (*1*−*3*). The stochastic approach is based on a free energy surface defined on a state space with only a few important degrees of freedom. It has the advantage of including crucial structural and mechanical information directly, while minimizing the (often unnecessary) complications of atom-by-atom detail. The main features of the free energy surface can sometimes be estimated from a combination of experimental structure and kinetics, simple theory, and detailed simulations.

Here these ideas are applied to the DNA polymerases. On the basis of structure and kinetics, four key degrees of freedom and 16 potentially stable conformational states common to all DNA polymerases are identified. The allowed transitions among these conformational states are then restricted, again on the basis of kinetics, to provide a general "framework" kinetic/stochastic mechanism that should be capable of describing all polymerases, including their mechanical properties. The framework mechanism is then used to construct the basic shape of the free energy surface for a DNA polymerase. Finally, a simple discrete state model for the HIV reverse transcriptase is constructed and explored.

Common Structure and Mechanism

Recently, a large number of crystal structures of polymerase−DNA complexes have become available (*4*−*14*), including a number of binary polymerase−DNA complexes and ternary polymerase−DNA−NTP[1] complexes, for T7 DNA polymerase, *Escherichia coli* polymerase I, Taq DNA polymerase, RB69 DNA polymerase, human and rat polymerase β, and the HIV reverse transcriptase (see Figure 1). Together, these structures span the full range of living things (archaea, eubacteria, eukaria, and viruses) and all the major functions of DNA polymerases. Despite this enormous breadth, and despite the wide disparity in sequence, all six structures (and all others known) show a common basic structure: all have domains identified as palm, thumb, and fingers, which bear a similar spatial relationship to each other; all bind DNA to align the crucial 3′ end of the primer

† This work was supported by National Institutes of Health Grant GM63808 to D.J.K. and J.A.B.

* To whom correspondence should be addressed: Molecular Machines Laboratory, Department of Chemistry, University of New Mexico, Albuquerque, NM 87131. Phone: (505) 277-3621. Fax: (505) 277-2609. E-mail: dkeller@unm.edu.

[1] Abbreviations: HIV RT, human immunodeficiency virus reverse transcriptase; Taq, *Thermus aquaticus*; NTP, nucleoside triphosphate; PP_i, inorganic pyrophosphate.

10.1021/bi0477079 CCC: $30.25
Published on Web 04/12/2005

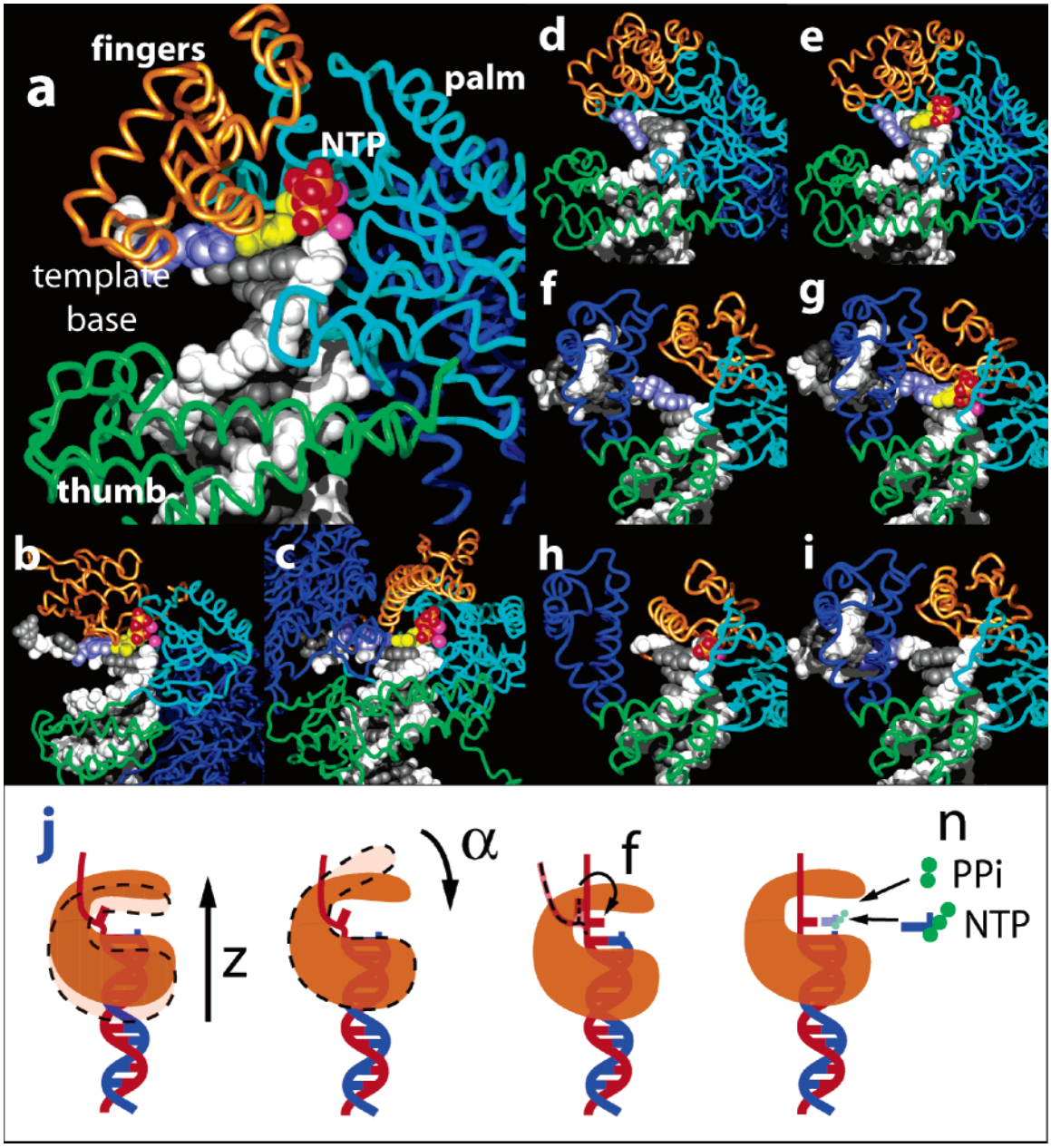

FIGURE 1: Comparison of crystal structures of DNA polymerases. Orange, cyan, and green indicate the fingers domain, palm domain, and thumb domain, respectively. The template base is light blue, and the nucleotide is yellow with a red and orange triphosphate moiety. Metal ions (usually magnesium) are magenta. Panels a−c show several DNA polymerases in the state just prior to catalysis, with the protein at the $n + 1$ position, the fingers domain closed, a nucleotide present, and the template base paired with the nucleotide (state 4, up, closed, stacked, NTP): (a) T7 DNA polymerase (PDB entry 1T7P), (b) HIV reverse transcriptase (PDB entry 1RTD), and (c) RB69 DNA polymerase (PDB entry 1IG9). Panels d−i show DNA polymerases in several other states of the catalytic cycle (see Figure 2): (d) Taq DNA polymerase (state 0; PDB entry 4KTQ) with protein at the $n + 1$ position, fingers open, NTP and PP_i absent, and template base unstacked; (e) same as panel d but with NTP present (state 2; PDB entry 2KTQ); (f) human polymerase β (state 1; PDB entry 1BPX) in the $n + 1$ position, fingers open, NTP and PP_i absent, and template base stacked; (g) same as panel f but with NTP present and fingers closed (state 4; PDB entry 1BPY); (h) rat polymerase β (state 5; PDB entry 1HUZ) in the n position with PP_i present and fingers closed (the template base is missing in this structure and also in panel i); and (i) human polymerase β (state 9; PDB entry 1BPZ) in the n position with fingers open and both NTP and PP_i absent. (j) Cartoons illustrating the main degrees of freedom in the framework model: (1) translocation along the DNA, z (z includes rotation about the helix axis that is not shown); (2) fingers domain closing angle, α; (3) template base stacking parameter, f; and (4) nucleotide occupation parameter, n (see the text for more complete definitions).

strand with the same spatial relationship to the palm, thumb, and fingers; all bind the incoming nucleotide so that it stacks on the end of the primer strand essentially as if it were already part of a double-stranded DNA; all organize the template base so that it base pairs with the incoming nucleotide and, like the nucleotide, adopts a conformation essentially identical to that of a double-stranded DNA; all have two metal ions bound to the palm domain and to the phosphate groups of the incoming nucleotide and in essentially identical positions with respect to the 3′-OH group of the primer; all seem to undergo a structural transition in which the fingers domain swings between an "open" and "closed" conformation; all have the ability to translocate from one site on the DNA to the next without unbinding (though they differ enormously in the average number of such processive steps between binding and dissociation).

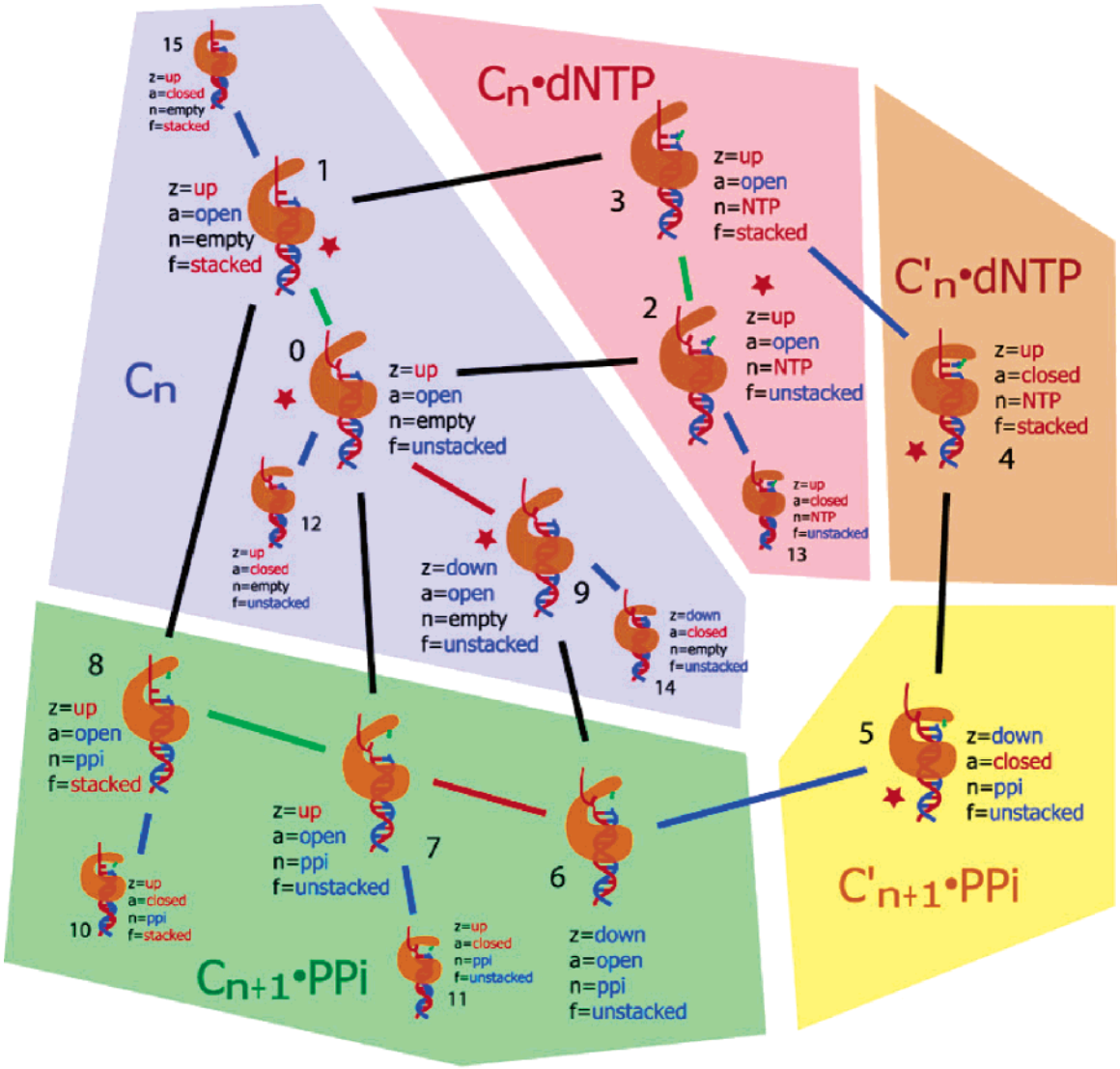

FIGURE 2: Discrete states of the framework model. Each state is indicated pictorially, and is labeled by framework parameters (z, α, n, and f) and by state index (0−15). Lines connecting pairs of states represent reversible transitions. They are color-coded by type: blue for fingers closing, red for translocation, green for template base stacking, and black for chemical change (NTP binding, PP_i release, or catalysis). Background color blocks correspond to the "biochemical" states of the minimal kinetic mechanism: blue (C_n) for the protein−DNA complex with *n* base pairs, red (C_n•dNTP) for C_n with a bound nucleotide, orange (C'_n•dNTP) for C_n•dNTP with a closed fingers domain, yellow (C'_{n+1}•PP_i) for C'_n with $n + 1$ base pairs and bound pyrophosphate (immediately after formation of a new phosphodiester bond), and green (C_{n+1}•PP_i) for C'_{n+1}•PP_i with an open fingers domain. Ten states (0−9) are on pathway, and six states (10−15) are off pathway traps or fingers fluctuations. Red stars denote states represented by crystal structures (see Figure 1).

Given these structural similarities, it is not surprising that all known DNA polymerases and reverse transcriptases seem also to share similar kinetic mechanisms (*15*−*21*). In the standard kinetic mechanism [based mainly on pre-steady state kinetics of nucleotide turnover (*15*)], the polymerase−DNA complex binds a nucleotide, then undergoes a (often slow) structural transition, then catalyzes the formation of a new phosphodiester bond, then undergoes a second structural transition, and finally releases pyrophosphate, returning to a state ready for the next nucleotide (see also Figure 2):

$$E{\cdot}DNA_n + NTP \rightarrow E{\cdot}DNA_n{\cdot}NTP \rightarrow E'{\cdot}DNA_n{\cdot}NTP \rightarrow E'{\cdot}DNA_{n+1}{\cdot}PP_i \rightarrow E{\cdot}DNA_{n+1}{\cdot}PP_i \rightarrow E{\cdot}DNA_{n+1} + PP_i$$

Some of these steps may be fast in certain polymerases (and so may not contribute to the overall polymerization rate or appear in some kinetic models), but all must be present in a complete turnover cycle. Additional steps are also needed to account for binding and release of DNA to and from polymerase, and for the proofreading nuclease activities of some polymerases, but the mechanism described above is sufficient for processive polymerization, which is the focus of this paper.

Key Degrees of Freedom

These structural and mechanistic similarities suggest (at least) four important degrees of freedom for a DNA polymerase, as illustrated in the various structures and cartoons in Figure 1.

(1) Translocation of the Protein along the DNA. In all crystal structures with bound DNA, the protein contacts the DNA along the minor groove, and follows the helix as it moves from one site to the next like a nut on a bolt (compare, for example, panels f and i in Figure 1). In the framework model, this helical motion is described by a helical position parameter *z*.

(2) Opening and Closing of the Protein, Described by the Conformational Parameter α. Crystal structures and ensemble kinetics show that DNA polymerases can adopt an open conformation which can admit new nucleotides and presumably allows movement of the protein along the DNA, or a closed conformation with a tight, well-ordered catalytic pocket [compare panels d−f and i (open) to panels a−c, g, and h (closed) in Figure 1]. The transition from open to closed is characterized mainly by a hinged movement of the fingers domain but also involves smaller movements by other domains, especially the thumb.

(3) Stacking and Unstacking of the n + 1 Template Base, Described by Stacking Parameter f. For polymerization to occur, the downstream base on the template strand must be stacked on the previous template base (which is now part of the double-stranded DNA), and correctly oriented for base pairing with the incoming nucleotide (compare panels d and e with panels a−c, f, and g in Figure 1).

(4) Binding and Unbinding of Nucleotide and Pyrophosphate, Described by Discrete Variable n with Three Values. n = empty for no nucleotide or pyrophosphate present. n = NTP for full nucleotide bound. n = PP_i for only pyrophosphate bound.

The precise definitions of z, α, and f in terms of molecular structure will vary from one application and polymerase to another. For many purposes, simple definitions will suffice, e.g., z as a rigid body helical rotation plus translation of the protein relative to the DNA, α as the hinge angle for the fingers bending upon opening and closing, f as the rms difference from a stacked template base, etc. Also, for some applications, other parameters may be necessary (e.g., binding of DNA to polymerase), but these four seem to be adequate to account for the main properties of processive polymerization.

Discrete States and Transitions

Though the first three parameters, z, α, and f, are continuous variables, it will be convenient in many cases to think of them as restricted to discrete values; let z = down correspond to the "initial" state of the protein before translocation [with no gap between the fingers and the blunt end of the double-stranded DNA (Figure 1i)] and z = up correspond to the "final" state of the protein after translocation to the next position (leaving a gap for a new nucleotide between fingers and DNA, e.g., Figure 1f). Likewise, let α = open correspond to the open state of the polymerase and α = closed correspond to the closed state. Finally, let f = stacked represent the oriented, stacked state (Figure 1a−c,f,g) and f = unstacked the disordered, unstacked state (Figure 1d,e).

The discrete values of the four conformational parameters nominally yield 2 × 2 × 2 × 3 = 24 discrete conformational states of the polymerase−DNA−nucleotide complex. Of these 24 nominal states, eight would involve overlaps or close contacts and so are not physically realistic. For example, whenever the protein is in the z = down, α = closed state, there is no gap between the fingers and the end of the double-stranded DNA, so it is not possible to insert a template base to be stacked or to insert a new nucleotide. Similarly, we omit all states with a z = down and stacked template base or NTP present even when the protein is in the open state.

This leaves 16 physically reasonable states, as shown in Figure 2. For convenience, the states are labeled both by the discrete values of z, α, f, and n and by simple integer *i*, e.g., (z, α, f, n) = (up, closed, stacked, NTP) = state 4 in Figure 2. Six of these states [denoted with red stars; state 0, Taq DNA polymerase (*6*), *Bacillus stearothermophilus* Pol I (*7*), and HIV RT (*10*); state 1, HIV RT (*9*, *11*) and polymerase β (*13*); state 2, Taq DNA polymerase (*6*); state 4, T7 DNA polymerase (*4*), Taq DNA polymerase (*5*, *6*), HIV RT (*8*), RB69 Pol α (*12*), and Pol β (*13*, *14*); state 5, Pol β (*14*), but the DNA has no template base, so the stacking state is undefined; state 9, Pol β (*13*), but the DNA has no template base] are represented directly by crystal structures (Figure 1); the other states can all be generated by simple changes from the known states. For example, state 10 z = up, α = closed, f = stacked, n = PP_i (bottom left of Figure 2) can be generated from any of the several crystal structures for the closed ternary complex (state 4, z = up, α = closed, f = stacked, n = NTP; top right of Figure 2) by exchanging the NTP for a pyrophosphate. The 16 states of the framework model thus seem to be quite well supported experimentally.

In principle, a transition pathway connects each pair of the 16 states, but not all pathways are physically reasonable. The lines connecting the states in Figure 2 represent all transition pathways that satisfy three requirements.

(1) Only one variable can change its value in a single transition. This eliminates all double changes and higher, such as a change from state {up, open, stacked, NTP} to state {up, closed, unstacked, NTP}, which involves changes in both α and f.

(2) No translocation, stacking or unstacking, or change in NTP binding site occupation can occur when the protein is in the closed state. The presumption here is that the closed state is tightly clamped.

(3) The chemical reaction, forming a new phosphodiester bond between the α-phosphate of the nucleotide and the 3′-OH of the primer strand, can occur only when a complete catalytic complex is formed; i.e., the protein must be in the closed conformation, and the template base must be stacked and base paired to a nucleotide.

The first rule says, in effect, that the fastest transitions are along the four parameters, z, α, f, and n, so that the fastest route between states that differ in two parameters is a two-step process rather than a single step. This effectively rules out concerted processes (such as simultaneous opening and translocation for example) and is perhaps the strongest assumption of the framework model. The second and third rules are strongly suggested by either ensemble kinetics, structure, or both. For example, though there are few data to suggest where protein translocation occurs in the catalytic cycle, all crystal structures with closed conformations show very tight organization of DNA, template base, and nucleotide (*4*−*6*, *8*, *12*−*14*). This argues for a tightly bound, immobile complex when the protein is closed. Similarly, the rule that the catalytic step occurs only in a fully organized catalytic complex is well supported by both structure and ensemble kinetics. Finally, it should be noted that these rules have been chosen to give a maximal set of transitions capable of describing all polymerases, but any specific polymerase may employ a smaller set of transitions.

Discrete Framework Model

The allowed states and transitions are shown in Figure 2, which represents the most general discrete framework mechanism. Each of the 16 states is labeled by the values of its four parameters (e.g., up, open, stacked, empty) and is connected to other states by transition pathways that satisfy the three rules listed above. Generally, the allowed transitions are sparse on the right side of the diagram, where the protein is closed, and numerous on the left side, where the protein is open. This reflects the fact that in the open state motion is less restricted, and the various steps needed to complete a catalytic cycle (NTP binding, translocation, fingers closing, etc.) may happen in a variety of different orders.

The "biochemical" states of the standard kinetic model are indicated by the color blocks in Figure 2. Thus, the single biochemical state E·DNA$_n$ is represented by six states (0, 1, 9, 12, 14, and 15) in the framework model, and the single transition E·DNA$_n$ → E·DNA$_n$·dNTP in the biochemical scheme is represented by two transitions in the framework model. The extra states and transitions in the framework model represent structural changes such as translocation and template base stacking that are not directly detected in ensemble kinetics experiments and so are not present in kinetic models.

Each allowed transition pathway is color-coded to indicate the physical nature of the change. Red indicates translocation from site to site along the DNA. Blue indicates opening or closing. Green indicates template base stacking or unstacking. Black indicates a purely chemical change such as nucleotide binding or pyrophosphate release. The mechanical transitions (red, blue, and green) are capable of converting chemical energy into mechanical work and will be affected by external forces, and so are crucial to understanding the machine-and-motor-like properties of polymerases.

Of the mechanical transitions, eight are open−close motions, two are translocations, and four are template base stackings. Six of the eight open−close motions are off-pathway "traps" or "dead-ends" connected to on-pathway open states. They may be in rapid equilibrium with their corresponding on-pathway states (so that both together act as a single state), may have unfavorable energy (and so contribute little to polymerase properties), or may represent true traps capable of sequestering part of the enzyme in an inactive form. Collectively, all six states represent noncatalytic open−close fluctuations during the overall turnover cycle.

The two remaining open−close transitions (connecting states 3 and 4 and states 5 and 6) are the conformational changes that immediately precede and immediately follow catalysis. The 3 → 4 closing transition is often the rate-limiting step in the turnover mechanism. It forms the essential, complete catalytic complex, state 4, in the model. It is commonly suggested that the formation of this state is the main mechanism by which the polymerase distinguishes the correct nucleotide from those that are not complementary to the template base, and hence is the main factor in determining polymerase fidelity. The idea is that unless the nucleotide and template base are correctly oriented in the binding pocket and correctly base paired to each other, the fingers domain cannot close, catalysis cannot take place, and the incorrect nucleotide eventually leaves the active site. Because this recognition is fundamental to the whole genetic apparatus, understanding this step in molecular detail and perhaps predicting the effects of mutations and external perturbations (inhibitors, forces, drugs, environment, etc.) are of particular interest.

After the catalytic step, the protein cannot translocate to the next position along the DNA until the complex opens, and (presumably because of steric clashes) the protein must translocate before the template base can stack or the nucleotide can bind in the active site. The two translocation steps (connecting states 6 and 7 and states 9 and 0) thus occur either immediately after opening or after opening and pyrophosphate release, and before stacking or nucleotide binding. The two translocation pathways differ only by the presence or absence of pyrophosphate, and so are probably similar in rate constants, sensitivity to load forces, effects of inhibitors, etc.

State Space and Free Energy Landscape

Three of the four parameters in the framework model (z, a, and f) involve physical motions and so can both generate force and respond to external force. The rate constants associated with transitions along these variables will, in general, depend on force. These three parameters are also (at least formally) continuous variables. Together, they define a three-dimensional continuous space of conformations for the polymerase−DNA−nucleotide system. The fourth parameter, n, is formally discrete (with three values), so the full state space for the framework model can be considered a set of three, three-dimensional volumes. For purposes of visualization, however, it is convenient to take parameter f also discrete, with f values of stacked and unstacked. Then the state space can be represented by six two-dimensional planes, as shown in Figure 3. Each plane represents the variables z and α, that is, all possible combinations of translocation and open−close conformations, for a particular state of base stacking and nucleotide occupation.

Associated with each point in the state space is a statistical free energy, which may be taken to be a Helmholtz free energy $A(z,a,f,n;T,F)$. Thus for each of the three volumes, there is a three-dimensional free energy surface $A_n(z,\alpha,f)$ [or, for each of the six planes, a two-dimensional free energy surface $A_{nf}(z,\alpha)$]. The polymerase machine moves on these surfaces, and the shapes of the surfaces determine both the stable states and the allowed transitions among them. Conversely, information about the stable states and transitions tells us about the shape of the free energy surface. Each biochemically relevant stable state corresponds to a minimum in the free energy surface, and each relevant transition between stable states corresponds to a low-energy path between minima. Saddle points along the low-energy paths correspond to transition states, and energy differences between saddle points and well bottoms correspond to activation energies.

Each of the z, α planes in Figure 3 is divided into four regions, one for each of the four combinations of z = up and down and α = open and closed. The six planes thus represent 24 states altogether, but eight of these (bright yellow in Figure 3) involve high-energy steric clashes and so play no role in the polymerase mechanism. To contribute to ensemble kinetics, a discrete state must have a significant

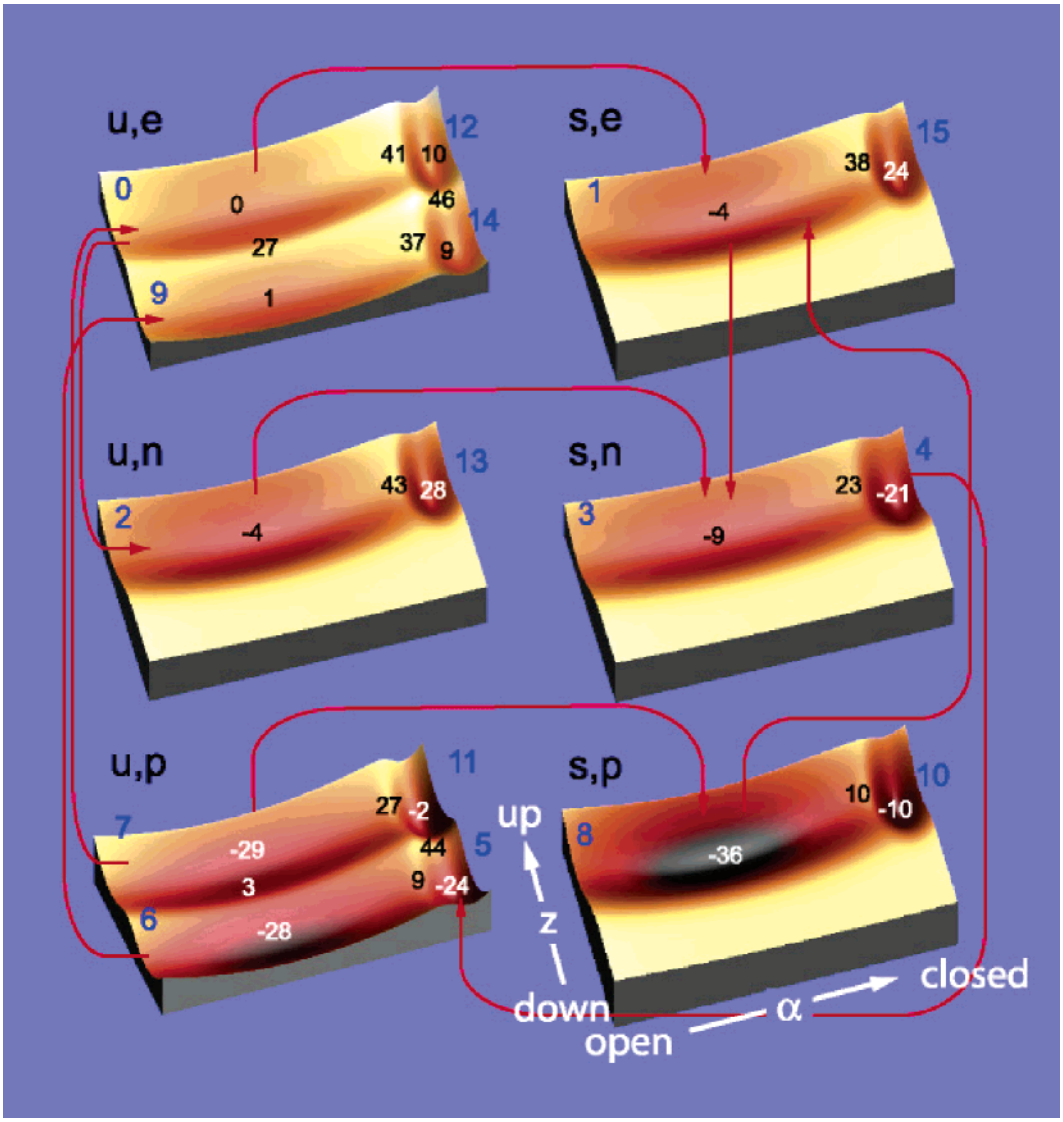

FIGURE 3: Example potential energy surface for the framework model. Each of the six two-dimensional surfaces shown represents one slice (along a plane where f is constant; see Table 1) through a three-dimensional free energy function $A_n(z,\alpha,f)$, as described in the text. Each surface is labeled by template base stacking state and nucleotide occupation state (e.g., u,e = unstacked, empty). For each surface, the vertical axis is translocation (z) and the horizontal axis is fingers movement (α). Each of the sixteen discrete states in Figure 2 is represented by a potential energy well (local energy minimum) labeled by its state index. High-energy disallowed states are bright yellow with no well. Transitions between states within a single surface are associated with low-energy passes between minima, and red arrows show transitions between surfaces (either purely chemical changes such as NTP binding or stacking−unstacking transitions). The numbers near minima and saddle points are free energy values (relative to state 0) for the example model for HIV reverse transcriptase, and give approximate activation barrier heights and free energy changes for each transition.

lifetime (on the millisecond time scale or longer), and so must be associated with a deep well in the continuum free energy surface. Conversely, any group of discrete states that is not separated from all others by significant barriers, and thus forms a single large well spanning more than one minimum, will effectively act as a single state.

Because the state space variables z, α, and f are closely connected to structure, many of the dimensions of the free energy surface can be estimated directly from crystal structures. For example, the distance between minima along the z direction (DNA translocation) should be about 3.4 Å, the length of one step along double-stranded DNA, and the two minima are probably similar in shape and size, with the barrier top roughly halfway between them. Along the α direction (fingers closing), the distance between minima should be similar to the distance moved by the tip of the fingers domain between open and closed states (~10 Å for HIV RT and Pol β and up to 15 Å for Taq DNA polymerase). Furthermore, the crystal structures are consistent with the idea that the forces that hold the fingers in the closed position are due to contacts between residues in the fingers and nearby structures such as the incoming nucleotide and the template base, and residues in the palm and thumb. If so, the potential well corresponding to the closed position (at large α in each of the surfaces in Figure 3) should be narrow (representing a "contact stick"), and the barrier separating open and closed states should be asymmetrically positioned nearer to the minimum in the closed state than the minimum in the open state. This choice is also consistent with measurements of how motor forces depend on external forces in T7 DNA polymerase (*22*) and HIV RT (*30*).

If the free energy surface is smooth near well bottoms and barrier tops, these basic dimensions largely determine the shape of the potential everywhere. The well bottoms will

Table 1: Calculated Rate Constants at Several Load Forces

transition	process	ΔG^a	$k_f(F=0)$	$k_r(F=0)$	$k_f(F=5)$	$k_r(F=5)$	$k_f(F=10)$	$k_r(F=10)$	$k_f(F=20)$	$k_r(F=20)$
0 → 1	stacking	−3.8	2760	601	1560	745	866	919	261	1400
2 → 3	stacking	−14.4	1020	3.1	358	4.7	123	7.0	12.9	15.7
3 → 4	closing	−16.2	102	0.15	38.5	0.22	14.7	0.33	1.7	0.74
5 → 6	opening	−11.8	1160	9.9	1360	5.4	1570	3.0	2140	0.83
6 → 7	translocation	−2.6	2670	921	3270	747	4000	606	5960	393
7 → 8	stacking	−2.5	2590	941	1420	1160	813	1420	244	2170
9 → 0	translocation	−3.1	2920	844	2370	1040	1940	1280	1290	1920
1 → 3	NTP binding	−4.0[b]	50 (NTP)	10						
0 → 2	NTP binding	+6.6[b]	70 (NTP)	1000						
4 → 5	catalysis	−3.4	20000	5000						
6 → 9	PP_i release	−	3000	0						
7 → 0	PP_i release	−	3000	0						
8 → 1	PP_i release	−	3000	0						

[a] ΔG in units of kJ/mol at 298.15 K; rate constants in units of s^{-1} except for the two second-order rate constants (k_f for 1 → 3 and 0 → 2 transitions) which are in units of $s^{-1}\ \mu M^{-1}$. Force in units of pN. No rate constants for PP_i binding are given, since it is assumed that the PP_i concentration is zero. Rate constants for the nonmechanical transitions (1 → 3 and below) were chosen to be consistent with previous measurements and modeling efforts (see refs *16*−*29*). [b] At 1 μM NTP.

be roughly harmonic, with curvatures (or force constants) determined mainly by the distances from minima to saddle points (barrier tops), and the saddle points themselves will also be roughly harmonic, with the maximum curvature limited by (at least) the requirement that all interactions occur over atomic distances of a few angstroms.

Transitions between Surfaces

The final element in the framework model is a set of transitions in the discrete variable n, corresponding to binding and unbinding of nucleotides, reaction of nucleotides to form new phosphodiester bonds, and binding and unbinding of pyrophosphate. Since the variable is discrete, these transitions are described by discrete transition rates, represented by red arrows in Figure 3. These rates are measured experimentally, or, at least in principle, estimated theoretically.

Constructing a model for a specific DNA polymerase then amounts to adjusting the positions of the minima and saddle points to suit the particular structure, estimating well depths and barrier heights, and adjusting the values of curvatures near the extrema.

Detailed Model for the HIV-1 Reverse Transcriptase

The sections above describe the general elements of the framework model. As an example of how the framework model is used for a specific polymerase, we have developed a detailed model for the HIV-1 reverse transcriptase. The model is quite preliminary (many values can only be crudely estimated) but consistent with all structural and kinetic data known to date. We use the model to predict (1) rate constants for all transitions in the framework model (Figure 2) as functions of temperature and applied force, (2) polymerization velocity versus applied load force for the polymerase motor, as a function of temperature and nucleotide concentration, and (3) the statistical distribution of polymerization times for stochastic polymerase motion, as a function of temperature, load force, and template length.

The calculations were carried out as described in the Appendix. Briefly, a specific mathematical form was chosen for the free energy surface, and the depths of free energy wells and the heights of activation energy barriers were adjusted to be consistent with experimentally measured rates and equilibrium constants (i.e., transition free energy changes). Then the free energy surface was used to calculate rate constants and state-to-state free energies of transition using eqs A.3 and A.4, and these were in turn used to calculate force−velocity curves and stepping time distributions (see below).

Table 1 lists the calculated values of the free energy changes and rate constants for the 13 transitions of the model, at several values of the load force. All but one ΔG is negative, but in this example model, most are near neutral. The pyrophosphate release transitions are assumed to occur in the absence of PP_i, so their free energies are formally undefined and functionally infinitely favorable. Aside from PP_i release, the driving transitions are (1) stacking in the presence of a bound nucleotide, (2) fingers closing, and (3) fingers opening. At high NTP concentrations, the NTP binding steps (1 → 3 and 0 → 2) also become strongly endergonic, especially for the 1 → 3 transition, where the NTP binds to an already stacked template base. The neutral steps may fluctuate forward and backward, but the driving steps ensure that the overall turnover cycle moves in the net forward direction. The driving steps, being effectively "irreversible", select the forward fluctuations of the neutral steps and, ratchet-like, force the motor ahead.

Force−Velocity Curves

Using the rate constants from eq A.4, the steady state turnover velocity for HIV RT was calculated as a function of applied load force, F, and as a function of temperature. Figure 4 shows several of these force−velocity curves. The fact that the rate-limiting step (fingers closing) is also a mechanical step in this model means that even a small force is sufficient to cause slowing of the polymerase, and all force−velocity curves are monotonic decreasing. If the model is changed so that force acts primarily on some fast step, the $F-V$ curves will have a region near zero force where the velocity is approximately independent of force (i.e., a plateau near $F = 0$).

At high load, where the rate-limiting step alone dominates the overall rate, the force−velocity curve will be approximately proportional to $e^{-F\Delta s_i/kT}$, where Δs_i is the barrier width (distance from the minimum on the reactant side to the barrier maximum; see eq A.6) or effective motion during one turnover cycle. If there is a dominant mechanical step

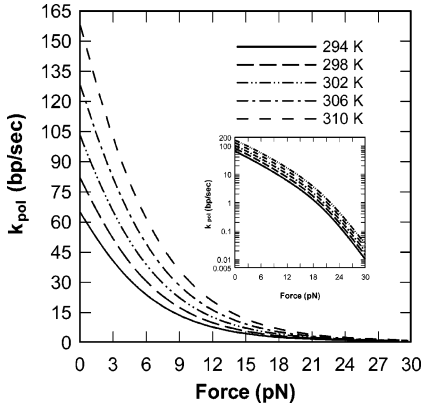

FIGURE 4: Velocity vs force plot at several temperatures for the example model. The overall rate constant, k_{pol}, is the saturating turnover velocity. The immediate drop in velocity at low force represents the fact that the slowest step (state 3 → state 4) is mechanical and so depends on force. The relatively rapid drop with an increase in load reflects a large effective overall motion on one full turnover cycle. In the model presented here, this includes contributions from all the mechanical steps, including fingers closing, template base stacking, and translocation. The inset is a log plot showing increasing slope and hence larger effective overall motion as force increases and additional mechanical steps become significant.

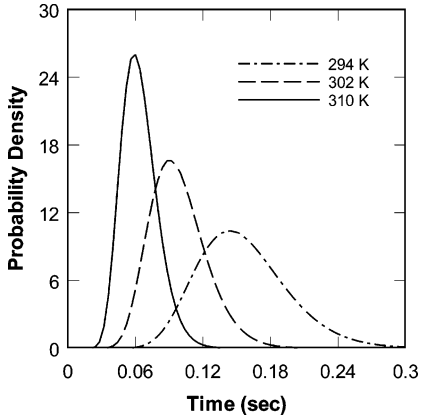

FIGURE 5: Replication time probability distributions for a 10-base template at three temperatures. The distributions are approximately Poissonian, with peaks near $t = 10/k_{pol}$. As the temperature increases, k_{pol} increases and mean replication time decreases.

therefore, a log plot should be linear and the slope provides an estimate of the effective motion or barrier width, but if several mechanical steps contribute to the load dependence, the log plot will be curved. The inset in Figure 4 shows that for the present model the log(velocity) versus F relationship is a curve, indicating contributions from several mechanical steps, and no single dominant step at all forces. This feature comes rather directly from the dimensions and shapes of the potential energy surfaces (Figure 3) which come in turn from the basic dimensions of DNA polymerase structure, and so may be a rather robust result. At low loads, the effective movement is ~0.8 nm, representing mainly the fingers closing motion. At high loads, the effective movement is more than twice as large, ~1.8 nm, representing contributions from all mechanical steps (fingers closing, translocation, and template base stacking). As the temperature increases, the velocity also rises at all load forces. If the rate is dominated by a single step, eq A.6 shows that by plotting ln(velocity) versus $1/T$ it is possible to estimate the effective barrier height at each value of the applied force.

Stepping Time Distributions

The stepping time distribution is the probability (density) for the time to complete one enzymatic turnover. The stepping time distribution can be compared directly to experimental histograms for turnover times from single-molecule fluorescence experiments (see below), and so is directly accessible experimentally. Formally, the stepping time distribution is the conditional probability that the system finished exactly one turnover cycle between times t and $t + dt$, given that it was in a designated initial state at time zero. We take any of states 0, 1, and 9 (the three termini of pyrophosphate release processes in the framework model) as the initial states, and to ensure that the system starts in one of these states, we require the total population of all three to be 1 and all other populations to be zero at time zero. It is assumed that the three possible initial states are in rapid equilibrium, so their relative populations are determined by their relative free energies (as determined by integrating over the corresponding regions of the potential energy surface). Finally, we modify the model by adding final states 0′, 1′, and 9′ that are distinct from states 0, 1, and 9 but identical in properties. This allows us to define a complete cycle as any series of transitions that leads from any of the initial states (0, 1, and 9) to any of the final states (0′, 1′, and 9′). To ensure that the system completes one and only one cycle, we make the pyrophosphate release transitions (state 6 to state 9, state 7 to state 0, or state 8 to state 1) irreversible (equivalent to setting the pyrophosphate concentration to zero, as in most in vitro experiments). Then the stepping time probability is the fraction of systems to arrive at state 0′, 1′, or 9′ during the time interval from t to $t + dt$, which is the same thing as the total rate of arrival of the system in these states times dt:

$$p_{step}(t)\,dt = \left[\frac{dp_{0'}(t)}{dt} + \frac{dp_{1'}(t)}{dt} + \frac{dp_{9'}(t)}{dt}\right] dt \qquad (1)$$

More generally, for any discrete model, the single-step time distribution is the derivative of the probability for the final state(s), for a process that started in the initial state(s) at time zero, with all transitions leading to the final state(s) being irreversible.

For the replication of a template of n bases, the population for the ith turnover, $\rho_i(t)$, is computed by multiple convolutions with a one-base stepping time distribution:

$$\rho_{i+1}(t) = \int_0^t \rho_i(t')\rho_{1\ step}(t - t')\,dt'$$

where $\rho_{1\ step}(t)$ is the one-base turnover stepping time probability density (eq 1 above). Figure 5 shows stepping time distributions at several temperatures, for a template 10 bases long. Despite the fact that each individual turnover involves many steps, the curves are essentially Poissonian, with a prominent peak at approximately $t = n/k_{pol}$. (A Poisson process results when each base incorporation is controlled by a single rate, k_{pol}. In this case, the distributions are approximately Poissonian because there is a reasonably dominant rate-limiting step in the single-base turnover cycle.) At non-zero loads, all mechanical steps are slowed, causing the distribution to broaden and the maximum to move to

longer times. Likewise, at higher temperatures, the rate-limiting forward rate constant is increased, and the maximum moves to shorter times. The presence of multiple substeps within a single-turnover cycle is less apparent in these multiple-turnover stepping time distributions than it would be, say, in a single-step time distribution, but their presence could be made apparent (and their transition rates measured) by varying the substrate concentration, by varying the types of forces applied (to affect the various different mechanical steps separately), or by more sophisticated fluorescence (or kinetic) experiments that probe the substeps directly.

Discussion and Conclusions

The intent of the framework model is to take a step from purely phenomenological models based on ensemble kinetics toward a directly predictive model based on a combination of kinetics, structure, and, ultimately, atomistic simulations. The defining elements of the model are the four degrees of freedom (z, α, f, and n) and the rules for transitions among the discrete states, all of which are strongly suggested by structure and kinetics. Together, these place significant restrictions on the range of possible mechanisms, and on the form of the potential energy surface for DNA polymerase machines.

The framework model has three kinds of *mechanical* steps (translocation, opening and closing, and stacking and unstacking) and three kinds of purely *chemical* steps (nucleotide binding, phosphodiester bond formation, and pyrophosphate release). The sizes of the energy drops on these two classes of transitions determine the basic mechanism by which polymerase motors generate forces. When the energy change across a mechanical transition is large and favorable, the motor moves (at least on the step in question) by means of an internally generated force, and the transition can be considered to be a "power stroke". On the other hand, if the free energy change is small or unfavorable, the motor may still generate forces, but will now do so by a "fluctuation ratchet" mechanism, in which other steps with favorable free energy prevent backward fluctuations and thus, ratchet-like, effectively select forward fluctuations. In motors such as the DNA polymerases, which have several mechanical steps, the overall mechanism may involve both power stroke and fluctuation ratchet processes. In the model presented here, the main fingers-closing step (state 3 → state 4) and the stacking step for a complex with NTP already present (state 2 → state 3) both have significant favorable free energy changes and so are power strokes, while the remaining mechanical transitions (including both translocation steps) have very small free energy changes and so are driven forward mainly by the ratcheting action of other steps (Table 1).

The most important feature of the DNA polymerases is the moveable fingers domain. The opening and closing of the fingers ensures the proper order for the crucial steps in the mechanism: NTP binding is necessary for fingers closing; fingers closing is necessary for the catalytic reaction; reaction is necessary for fingers opening; fingers opening is necessary for the cluster of steps involving PP_i release, template base stacking, and translocation. This sequence of dependencies essentially defines the turnover cycle.

The key aspects of the potential energy surface for a polymerase are thus those that generate the sequence: the favorable energy for the closed state (over the open state) in the s, n surface [when a nucleotide is present and the template base is stacked (Figure 3)] compared to the unfavorable energy of the closed state in the s, e or u, n surfaces (when either the nucleotide is absent or the template base is unstacked); the catalytic reaction only in the closed state [represented by the discrete transition (red arrow) from state 4 to state 5 in Figure 3]; the unfavorable energy for the closed state in the u, p surface immediately after the catalytic reaction; the high barriers between state 5 in the u, p surface and all other states except the open state 6 (and the lack of discrete transitions leading out from state 6), so that the fingers domain must open before any further transitions occur. The absolute sizes of energy differences and barriers can vary, but these qualitative features are necessary for any polymerase free energy surface that is consistent with existing experimental results.

A main goal of any modeling effort is to relate the macroscopic properties of the molecular machine to its molecular structure, say, by a series of calculations with decreasing molecular detail: interactions at the atomic scale → potential energy surface and diffusion constants → state-to-state rate constants and free energies → global kinetic mechanism. The framework model, by a combination of structural information and experimental kinetic data, attempts to fill in the final two steps, but does not make direct quantitative predictions from structure. Nonetheless, in light of the comments in the previous paragraph, the important parts of the potential energy surface can help us identify the crucial interactions and features in the protein structure itself. The fact that the closed state is favored when the nucleotide is present and the template base is stacked, but not otherwise, indicates that there must be strong favorable interactions between the fingers and the nucleotide·template base pair. The fact that the open state is favored after reaction, when the β- and γ-phosphates have been severed from the nucleotide, suggests that a large part of this favorable interaction is with the phosphate tail, and that there is a tension in the α-phosphate−β-phosphate linkage that is released when the linkage is broken. Similarly, the size of the barrier to fingers closing is determined largely by the stiffness of the hinge region of the fingers domain: the stiffer the hinge, the larger the curvature of the free energy surface in the open state of the s, n surface and the higher the barrier to closing. It may be possible to estimate some of these interactions by direct simulation, thus closing the last link between atomic detail and macroscopic properties.

Appendix

Calculation Methods for the HIV RT Model. (1) The free energy surface was represented as the logarithm of a sum of Gaussian functions centered at the minima in each three-dimensional surface:

$$A_n(z,\alpha,f;T) = -m_n kT \ln \times \left\{\sum_i e^{-[A_{0ni}+(1/2)\kappa_{ni}(z-z_{ni})^2+(1/2)\chi_{ni}(\alpha-\alpha_{ni})^2+(1/2)\lambda_{ni}(f-f_{ni})^2]/m_n kT}\right\} \quad \text{(A.1)}$$

where the sum is over all minima, i, in the potential energy surface for a given value of n, m_n is an integer scale factor,

T is the absolute temperature, k is Boltzmann's constant, A_{0ni} is the (approximate) well depth for the ith minimum with nucleotide occupation state n, κ_{ni}, χ_{ni}, and λ_{ni} are force constants for (approximately harmonic) well i in occupation state n, and z_{ni}, α_{ni}, and f_{ni} are the (approximate) locations in z, α, f space of the minimum for well i in occupation state n. This form yields smooth, reasonable surfaces that are close to harmonic near the minima and saddle points. The scale factor, m_n, makes it possible to adjust the depth of the wells and the curvature at the saddle points relatively independently. The corresponding probability density

$$\rho_n(z,\alpha,f;T) \propto e^{-A_n(z,\alpha,f;T)/kT} = \left\{\sum_i e^{-[A_{0ni}+(1/2)\kappa_{ni}(z-z_{ni})^2+(1/2)\chi_{ni}(\alpha-\alpha_{ni})^2+(1/2)\lambda_{ni}(f-f_{ni})^2]/m_nkT}\right\}^{m_n} \quad (A.2)$$

is relatively easily integrated to find transition free energies and rate constants (see below). The six surfaces shown in Figure 3 were found by evaluating $A_n(z,\alpha,f;T)$ in eq 1 above at f = stacked ($f = 1.0$, defined as the rms difference from the stacked position in 1RTD) and at f = unstacked ($f = 7.5$) for each of the three, three-dimensional surfaces (n = NTP, PP_i, or empty).

(2) The well depths and force constants were adjusted so that the equilibrium constants, K_{ij}, and free energy changes, $\Delta A_{ij} = -kT \ln K_{ij}$, for each transition, $i \rightleftharpoons j$, agree with experimental values, wherever available. The equilibrium constant is

$$K_{ij}(T) = e^{-\Delta A_{ij}(T)/kT} = \frac{\int_{\text{region } j} e^{-A_n(z,\alpha,f;T)/kT}\, dz\, d\alpha\, df}{\int_{\text{region } i} e^{-A_n(z,\alpha,f;T)/kT}\, dz\, d\alpha\, df} \quad (A.3)$$

If the potential wells are many kT's deep, only the regions near the bottoms of the wells contribute to the integrals. The free energy surfaces in these regions are approximately harmonic, and the equilibrium constant is then $K_{ij} \cong \sqrt{(\kappa_{ni}/\kappa_{nj})(\chi_{ni}/\chi_{nj})(\lambda_{ni}/\lambda_{nj})}\, e^{-(A_{0nj}-A_{0ni})/kT}$, where κ_{ni}, χ_{ni}, etc. are the same as in part 1. Measured or estimated equilibrium constants for HIV RT include those for nucleotide dissociation ($\sim$10 μM) (*16*, *21*, *23*−*29*), for fingers closing, for the catalytic step, and for pyrophosphate release ($\sim$7 mM) (*16*). In this model, the six "trap" states, 10−15, are considered high-energy states, so only 10 of the full 16 states contribute significantly to polymerase properties.

(3) The force constants and the scale factor m_n were adjusted to make the saddle points occur at approximately the right positions (based on structure; see above) with reasonable curvatures, and to make the barrier heights have values that gave agreement with measured rate constants and to be consistent with the overall single-step rate for HIV RT. The rate constants for transition $i \rightleftharpoons j$ were calculated using (*3*)

$$k_{ji}(F) = D\frac{N_j}{\Delta_{ij}},\ k_{ij}(F) = D\frac{N_i}{\Delta_{ij}} \quad (A.4a)$$

where D is a diffusion constant for motion along the reaction coordinate, $\Delta_{ij} = N_i\Sigma_j' + N_j\Sigma_i'$, and

$$N_i = \int_{\text{region } i} e^{-A_n(s)/kT}\, ds,\ N_j = \int_{\text{region } j} e^{-A_n(s)/kT}\, ds$$

$$\Sigma_i' = \int_{\text{region } j} e^{-A_n(s)/kT}\, \sigma_i(s)\, ds,\ \Sigma_j' = \int_{\text{region } i} e^{-A_n(s)/kT}\, \sigma_j(s)\, ds \quad (A.4b)$$

$$\sigma_i(s) = \int_s^{s_{ij}} e^{\beta A_n(s'')}\, ds'',\ \sigma_j(s) = \int_{s_{ij}}^{s} e^{\beta A_n(s'')}\, ds''$$

and s is one of the three variables z, α, and f, with s_{ij} the dividing point (the barrier maximum) between regions i and j. The quantity $A_n(s)$ is a one-dimensional potential of mean force obtained from the full free energy surface by integrating the probability density over two of the three variables (z, α, and f). For example, for a translocation along the variable z from state i to state j, the one-dimensional potential is approximately

$$e^{-A_n(z)/kT} = C\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left\{\sum_i e^{-[A_{0ni}+(1/2)\kappa_{ni}(z-z_{ni})^2+(1/2)\chi_{ni}(\alpha-\alpha_{ni})^2+(1/2)\lambda_{ni}(f-f_{ni})^2]/m_nkT}\right\}^{m_n} d\alpha\, df$$

$$= C\sum_{\upsilon=0}^{m_n}\left[\frac{m_n!}{\upsilon!(m_n-\upsilon)!}\sqrt{\frac{2\pi m_nkT}{\upsilon\chi_{ni}+(m_n-\upsilon)\chi_{nj}}}\sqrt{\frac{2\pi m_nkT}{\upsilon\lambda_{nj}+(m_n-\upsilon)\lambda_{nj}}} \times e^{-\{(1/2)(\alpha_{ni}-\alpha_{nj})^2/[1/\upsilon\chi_{ni}+1/(m_n-\upsilon)\chi_{nj}]+(1/2)(f_{ni}-f_{nj})^2/[1/\upsilon\lambda_{ni}+1/(m_n-\upsilon)\lambda_{nj}]\}/m_nkT} \times e^{-\{\upsilon[A_{0_{ni}}+(1/2)\kappa_{ni}(z-z_{ni})^2]+(m_n-\upsilon)[A_{0_{nj}}+(1/2)\kappa_{nj}(z-z_{nj})^2]\}/m_nkT}\right] \quad (A.5)$$

where C is a normalization factor that does not affect the rate constants. In arriving at the last expression in eq A.5, we assumed that path s passes through the two minima, i and j, and only the two Gaussians that lie on these minima contribute significantly to the probability density (eq A.2).

Unlike the transition free energies, which depend mainly on the energy differences between well minima, the rate constants depend mainly on barrier heights and widths. If the wells are many kT's deep, the formulas above can be approximated by simple Arrhenius-like expressions:

$$k_{ji}(F) \cong \frac{D}{2\pi kT}\sqrt{\kappa_i\kappa_b}\, e^{-(\Delta A_{ji}+F\Delta s_i)/kT}$$

$$k_{ij}(F) \cong \frac{D}{2\pi kT}\sqrt{\kappa_j\kappa_b}\, e^{-(\Delta A_{ij}+F\Delta s_j)/kT} \quad (A.6)$$

where κ_i and κ_j are the force constants [second derivatives of $A_n(s)$] at the minimum of wells i and j, respectively, and κ_b is the negative of the second derivative of $A_n(s)$ at the barrier maximum.

Table 2 lists values of z_{ni}, α_{ni}, f_{ni}, κ_{ni}, χ_{ni}, λ_{ni}, and A_{0ni} for each well i and for each occupation state n. In all cases, $m_n = 20$. Diffusion coefficients were roughly estimated using the Einstein formula $D = kT/\gamma$, where γ is an effective friction coefficient (assumed to be approximately independent of temperature) calculated using Stoke's law, $\gamma = 6\pi\eta R$, with R chosen to match the size of the feature associated with each degree of freedom (e.g., large R for motions along z, where the entire protein translocates, and small R for f, where only the template base moves). For motions along z, $\gamma = 5.0 \times 10^{-11}$ N m^{-1} s^{-1} (viscosity = 0.001 kg s m^{-1}, effective radius = 2.6 nm). For motions along α, $\gamma = 3.0 \times 10^{-11}$ N m^{-1} s^{-1} (effective radius = 1.6 nm). For motions along f, $\gamma = 1.0 \times 10^{-11}$ N m^{-1} s^{-1} (effective radius = 0.53 nm).

Table 2: Parameters for the Potential Energy Surface

Gaussian no.	position (nm) α	position (nm) z	position (nm) f	curvature (kJ mol^{-1} nm^{-2}) k_α	curvature (kJ mol^{-1} nm^{-2}) k_z	curvature (kJ mol^{-1} nm^{-2}) k_f	nominal depth (kJ/mol) A_0
Potential A (n = empty, $m_n = 20$							
0	0.50	0.17	0.10	2.5	2.5	2.5	80.0
1	0.50	0.51	0.10	29.7	249	509	11.0
2	1.125	0.51	0.10	1559	390	998	32.0
3	1.125	0.17	0.10	2.5	2.5	2.5	80.0
4	0.50	0.17	0.75	27.8	265	62.4	11.8
5	0.50	0.51	0.75	27.8	265	20.4	10.1
6	1.125	0.51	0.75	1559	390	62.4	32.0
7	1.125	0.17	0.75	998	390	62.4	32.0
Potential B (n = NTP, $m_n = 20$)							
0	0.50	0.17	0.10	2.5	2.5	2.5	80.0
1	0.50	0.51	0.10	27.3	249	509	−32.1
2	1.125	0.51	0.10	509	249	998	−48.1
3	1.125	0.17	0.10	2.5	2.5	2.5	80.0
4	0.50	0.17	0.75	2.5	2.5	2.5	80.0
5	0.50	0.51	0.75	27.7	249	21.2	−20.1
6	1.125	0.51	0.75	1559	249	62.4	32.0
7	1.125	0.17	0.75	2.5	2.5	2.5	80.0
Potential C (n = PP_i, $m_n = 20$)							
0	0.50	0.17	0.10	2.5	2.5	2.5	80.0
1	0.50	0.51	0.10	27.7	249	509	−12.0
2	1.125	0.51	0.10	998	249	998	32.0
3	1.125	0.17	0.10	2.5	2.5	2.5	80.0
4	0.50	0.17	0.75	22.9	265	62.4	−13.0
5	0.50	0.51	0.75	27.7	265	20.4	−14.5
6	1.125	0.51	0.75	998	1559	27.7	32.0
7	1.125	0.17	0.75	998	1559	62.4	2.80

For completeness, a load force, F, has been added to the potential in eq A.6: $A_n(s,F) = A_n(s,0) + Fs$. This load force is defined so that positive forces oppose the "forward" transition from well i to well j, and negative forces assist the forward transition (and vice versa for the backward transition). The quantities Δs_i and Δs_j in eq A.6 are the distances along s from the barrier top to the minimum in well i and in well j, respectively. Comparison of calculations done using eq A.4 with the approximate formulas in eq A.6 shows them to be good to a few percent for typical one-dimensional free energy functions with two wells separated by a barrier, as is the case here. Finally, the transition rates for the discrete transitions in n (between surfaces) were estimated from literature values (see Table 2 and refs *16*−*29*).

REFERENCES

1. Bustamante, C., Keller, D., and Oster, G. (2001) The Physics of Molecular Motors, *Acc. Chem. Res. 34*, 412−420.
2. Oster, G., and Wang, H. Y. (2003) Rotary Protein Motors, *Trends Cell Biol. 13*, 114−121.
3. Keller, D., and Bustamante, C. (2000) The Mechanochemistry of Molecular Motors, *Biophys. J. 78*, 541−556.
4. Doublie, S., Tabor, S., Long, A. M., Richardson, C. C., and Ellenberger, T. (1998) Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 Å resolution, *Nature 391*, 251−258.
5. Li, Y., Mitaxov, V., and Waksman, G. (1999) Structure-based Design of Taq DNA Polymerase with Improved Properties of Dideoxynucleotide Incorporation, *Proc. Natl. Acad. Sci. U.S.A. 96*, 9491−9496.
6. Kim, Y., Eom, S. H., Wang, J., Lee, D. S., Suh, S. W., and Steitz, T. A. (1995) Crystal structure of *Thermus aquaticus* DNA polymerase, *Nature 376*, 612−616.
7. Kiefer, J. R., Mao, C., Braman, J. C., and Beese, L. S. (1998) Visualizing DNA replication in a catalytically active *Bacillus* DNA polymerase crystal, *Nature 391*, 304−307.
8. Huang, H., Chopra, R., Verdine, G. L., and Harrison, S. C. (1998) Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: Implications for drug resistance, *Science 282*, 1669−1675.
9. Sarafianos, S. G., Das, K., Clark, A. D., Jr., Ding, J., Boyer, P. L., Hughes, S. H., and Arnold, E. (1999) Lamivudine (3TC) Resistance in HIV-1 reverse transcriptase involves steric hindrance with β-branched amino acids, *Proc. Natl. Acad. Sci. U.S.A. 96*, 10027−10032.
10. Sarafianos, S. G., Das, K., Tantillo, C., Clark, A. D., Jr., Ding, J., Whitcomb, J., Boyer, P. L., Hughes, S. H., and Arnold, E. (2001) Crystal structure of HIV-1 reverse transcriptase in complex with a polypurine tract, RNA:DNA, *EMBO J. 20*, 1449−1461.
11. Ding, J., Das, K., Hsiou, Y., Sarafianos, S. G., Clark, A. D., Jr., Jacobo-Molina, A., Tantillo, C., Hughes, S. H., and Arnold, E. (1998) Structure and functional implications of the polymerase active site region in a complex of HIV-1 RT with a double stranded DNA template primer and an antibody fAb fragment at 2.8 Å resolution, *J. Mol. Biol. 284*, 1095−1111.
12. Franklin, M. C., Wang, J., and Steitz, T. A. (2001) Structure of the replicating complex of a pol α family DNA polymerase, *Cell 105*, 657−667.
13. Sawaya, M. R., Prasad, R., Wilson, S. H., Kraut, J., and Pelletier, H. (1997) Crystal structures of human DNA polymerase β complexed with gapped and nicked DNA: Evidence for an induced fit mechanism, *Biochemistry 36*, 11205−11215.
14. Arndt, J. W., Gong, W., Zhong, X., Showalter, A. K., Liu, J., Dunlap, C. A., Lin, Z., Paxson, C., Tsai, M.-D., and Chan, M. K. (2001) Insight into the catalytic mechanism of DNA polymerase β: Structures of intermediate complexes, *Biochemistry 40*, 5368−5375.
15. Johnson, K. A. (1993) Conformational Coupling in DNA Polymerase Fidelity, *Annu. Rev. Biochem. 62*, 685−713.
16. Hseih, J.-C., Zinnen, S., and Modrich, P. (1993) Kinetic Mechanism of the DNA-dependent DNA Polymerase Activity of Human Immunodeficiency Virus Reverse Transcriptase, *J. Biol. Chem. 268*, 24607−24613.
17. Furge, L. L., and Guengerich, F. P. (1997) Analysis of Nucleotide Insertion and Extension at 8-Oxo-7,8-dihydroguanine by Replicative T7 Polymerase exo− and Human Immunodeficiency Virus-1 Reverse Transcriptase Using Steady-State and Pre-Steady-State Kinetics, *Biochemistry 36*, 6475−6487.
18. Woodside, A. M., and Guengerich, F. P. (2002) Effect of the O6 Substituent on Misincorporation Kinetics Catalyzed by DNA Polymerases at O6-Methylguanine and O6-Benzylguanine, *Biochemistry 41*, 1027−1038.
19. Woodside, A. M., and Guengerich, F. P. (2002) Misincorporation and Stalling at O6-Methylguanine and O6-Benzylguanine: Evidence for Inactive Polymerase Complexes, *Biochemistry 41*, 1039−1050.
20. Pop, M. P., and Biebricher, C. K. (1996) Kinetic Analysis of Pausing and Fidelity of Human Immunodeficiency Virus Type 1 Reverse Transcription, *Biochemistry 35*, 5054−5062.
21. Kati, W. M., Johnson, K. A., Jerva, L. F., and Anderson, K. S. (1992) Mechanism and Fidelity of HIV Reverse Transcriptase, *J. Biol. Chem. 267*, 25988−25997.
22. Wuite, G. J. L., Smith, S. B., Young, M., Keller, D., and Bustamante, C. (2000) Single-Molecule Studies of the Effect of Template Tension on T7 DNA Polymerase Activity, *Nature 404*, 103−106.
23. Krebs, R., Immendorfer, U., Thrall, S. H., Wohrl, B. M., and Goody, R. S. (1997) Single-Step Kinetics of HIV-1 Reverse Transcriptase Mutants Responsible for Virus Resistance to Nucleoside Inhibitors Zidovudine and 3-TC, *Biochemistry 36*, 10292−10300.
24. Kerr, S. G., and Anderson, K. S. (1997) Pre-Steady-State Kinetic Characterization of Wild-Type and 3′-Azido-3′-deoxythymidine (AZT) Resistant Human Immunodeficiency Virus Type 1 Reverse Transcriptase: Implication of RNA Directed DNA Polymerization in the Mechanism of AZT Resistance, *Biochemistry 36*, 14064−14070.
25. Jeffrey, J. L., Feng, J. Y., Qi, C. C. R., Anderson, K. S., and Furman, P. A. (2003) Dioxolane Guanosine 5′-Triphosphate, an

Alternative Substrate Inhibitor of Wild-type and Mutant HIV-1 Reverse Transcriptase, *J. Biol. Chem. 278*, 18971−18979.
26. Wohrl, B. M., Krebs, R., Thralli, S. H., Le Grice, S. F. J., Scheidig, A. J., and Goody, R. S. (1997) Kinetic Analysis of Four HIV-1 Reverse Transcriptase Enzymes Mutated in the Primer Grip Region of p66, *J. Biol. Chem. 272*, 17581−17587.
27. Zinnen, S., Hsieh, J.-C., and Modrich, P. (1994) Misincorporation and Mispaired Primer Extension by Human Immunodeficiency Virus Reverse Transcriptase, *J. Biol. Chem. 269*, 24195−24202.
28. Spence, R. A., Anderson, K. S., and Johnson, K. A. (1996) HIV-1 Reverse Transcriptase Resistance to Nonnucleoside Inhibitors, *Biochemistry 35*, 1054−1063.
29. Deval, J., Selmi, B., Boretto, J., Egloff, M. P., Guerreiro, C., Sarfati, S., and Canard, B. (2002) The Molecular Mechanism of Multidrug Resistance by the Q151M Human Immunodeficiency Virus Type 1 Reverse Transcriptase and Its Suppression Using Boranophosphate Nucleotide Analogues, *J. Biol. Chem. 277*, 42097−42104.
30. Lu, H., Macosko, J., Habel-Rodriquez, D., Keller, R. W., Brozik, J. A., and Keller, D. J. (2004) Closing of the Fingers Domain Generates Motor Forces in the HIV Reverse Transcriptase, *J. Biol. Chem. 279*, 54529−54532.

BI0477079