Vertical-scanning Mutagenesis of a Critical Tryptophan in the Minor Groove Binding Track of HIV-1 Reverse Transcriptase

MOLECULAR NATURE OF POLYMERASE-NUCLEIC ACID INTERACTIONS

(Received for publication, June 10, 1998, and in revised form, August 6, 1998)

William A. Beard‡§, Katarzyna Bebenek§§, Thomas A. Darden‡, Leping Li‡, Rajendra Prasad‡, Thomas A. Kunkel¶, and Samuel H. Wilson¶¶

From the ‡Laboratory of Structural Biology and the §Laboratory of Molecular Genetics, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina 27709

While sequence-specific DNA-binding proteins interact predominantly in the DNA major groove, DNA polymerases bind DNA through interactions in the minor groove that are sequence nonspecific. Through functional analyses of alanine-substituted mutant enzymes that were guided by molecular dynamics modeling of the human immunodeficiency virus type 1-reverse transcriptase and DNA complex, we previously identified a structural element in reverse transcriptase, the minor groove binding track (MGBT). The MGBT is comprised of five residues (Ile94, Gln258, Gly262, Trp266, and Gln269) which interact 2–6 base pairs upstream from the polymerase active site in the DNA minor groove and are important in DNA binding, processivity, and frameshift fidelity. These residues do not contribute equally; functional analysis of alanine mutants suggests that Trp266 contributes the most to binding. To define the molecular interactions between Trp266 and the DNA minor groove, we have analyzed the properties of eight mutants, each with an alternate side chain at this position. A refined molecular dynamics model was used to calculate relative binding free energies based on apolar surface area buried upon complex formation. In general, there was a strong correlation between the relative calculated binding free energies for the alternate residue 266 side chains and the magnitude of the change in the properties which reflect template-primer interactions (template-primer dissociation rate constant, \( K_{T_{\text{RTP}}} \) processivity, and frameshift fidelity). This correlation suggests that hydrophobic interactions make a major contribution to the stability of the polymerase-DNA complex. Additionally, tyrosine and arginine substitutions resulted in mutant enzymes with DNA binding properties better than predicted by buried surface area alone, suggesting that hydrogen bonding could also play a role in DNA binding at this position.

Due to its central role in viral replication, HIV-1 reverse transcriptase (RT) is an ideal target for antiviral drug design. It has been extensively scrutinized both structurally and biochemically. RT is a heterodimer of 66- and 51-kDa polypeptides with three catalytic activities: RNA- and DNA-dependent DNA synthesis and RNase H endonuclease. The polymerase and RNase H active sites are located at opposite ends of the nucleic acid-binding cleft on p66 (1). The p51 subunit is derived by proteolytic carboxyl terminus truncation of p66 removing the RNase H subdomain. Although both p66 and p51 are composed of four common subdomains (fingers, palm, thumb, and connection), the spatial arrangement of the subdomains in the two subunits is very different. The connection subdomain, which joins the polymerase (fingers, palm, and thumb) and RNase H domains of p66, occludes the polymerase active site of p51.

RT lacks a 3' → 5' proofreading exonuclease activity (2), and it is not highly accurate (2, 3). The low RT fidelity is believed to play an important role in generating modified forms of RT with decreased inhibitor sensitivity. In addition to misincorporation errors, RT frequently produces frameshifts and base substitutions initiated by primer and template strand misalignment (4).

The fidelity and processivity of RT are strongly dependent on nucleic acid sequence. A correlation has been noted between sites of elevated termination and lower frameshift fidelity, and these parameters have been demonstrated to be influenced by the sequence of the DNA duplex template-primer (T-P) as much as 6 nucleotides from the 3'-primer terminus (5, 6). The x-ray crystal structure of RT complexed with DNA shows that DNA is A-like near the polymerase active site, and is B-form near the RNase H active site. A bend of 40–45° occurs at the junction between A-like and B-form DNA (7). The p66 thumb contacts the T-P over several base pairs upstream of the polymerase active-site. Two antiparallel \( \alpha \)-helices (H and I) of the thumb subdomain interact in the region of the DNA bend. Whereas \( \alpha \)-helix I appears to interact with the sugar-phosphate backbone of the template strand, \( \alpha \)-helix H is partially embedded in the widened DNA minor groove and interacts primarily with the primer strand. It has been suggested that these helices serve as a track along which the DNA moves during translocation (7). We have examined the role of residues in \( \alpha \)-helices H and I in nucleic acid binding by alanine-scanning mutagenesis and characterized the fidelity and kinetic properties of the mutant proteins (8–11). This analysis revealed that while alanine substitution of individual residues of \( \alpha \)-helix I did not influence T-P binding or fidelity, substitution to alanine at several residues of \( \alpha \)-helix H resulted in decreased nucleic acid binding affinity, processivity, AZTTP sensitivity (\( K_{T_{\text{AZTTP}}} \)), and frameshift fidelity (8, 10, 11).

Since the resolution of the crystallographic RT-DNA complex was insufficient to define detailed protein-DNA interactions (7), and since higher resolution structures of RT (apoenzyme and non-nucleoside inhibitor complexes) do not include DNA (1,
HIV-1 RT DNA Minor Groove Interactions

FIG. 1. Model of MGBT residues in the DNA minor groove. The van der Waals surface of the MGBT residues (orange: Ileadia, Gly258, Gly259, red: Trp266) are illustrated interacting with the solvent accessible surface of the DNA minor groove 2–6 base pairs from the active site carboxylates (blue). Trp266 primarily interacts with the primer nucleotide of the third base pair (white) upstream of the 3’-hydroxyl. The 5’ terminus of the template strand is indicated. This figure was made with GRASP (45).

12–16), molecular dynamics modeling was undertaken (11). Functional analysis of mutant RTs coupled with this molecular dynamics modeling identified a structural element, the minor groove binding track (MGBT), that is important for DNA binding and frameshift fidelity. The MGBT consists of five residues, four in the thumb (Gln258, Gly259, and Trp266) and one (Ileadia) in β-sheet 5b of the palm (Fig. 1). These five residues protrude into the minor groove over a distance from the second through the sixth base pair from the 3’-hydroxyl primer terminus and provide important RT interactions to the T-P during processive synthesis. Amino acid sequence alignments indicate that all five residues are highly conserved among lentiviral reverse transcriptases, suggesting that the MGBT is present in these enzymes.

RT mutants with a substitution to alanine at any of the MGBT residues have altered T-P interactions. However, the degree of the alteration varies depending on the amino acid substitution. While substitutions at Gln258, Gln259, and Ileadia have a modest effect on T-P binding and processivity, the strongest effect, a more than 400-fold increase in the T-P dissociation rate constant, is caused by replacement of the large tryptophan side chain at position 266, suggesting that this residue of the MGBT provides much of the binding of the RT to the T-P (8, 10). To understand the molecular interactions which occur between residue 266 and the DNA minor groove, alternate substitutions were made at this position. The mutant proteins were purified and their kinetic properties, nucleic acid binding, processivity, AZTTP sensitivity, and frameshift fidelity characterized. The results of the kinetic and fidelity characterization of these mutant RTs are discussed in light of a refined molecular dynamics model of the RT-DNA complex.

EXPERIMENTAL PROCEDURES

Materials—T4 polynucleotide kinase, poly(rA), dTTP, and dNTPs were from Pharmacia Biotech Inc. [α-32P]dTTP (3000 Ci/mmol) was from TEN Life Science Products Inc. Sodium heparin (170 USP units/mg) was from U. S. Biochemical Corp. Bacterial strains, phage, and other materials have been described (9). Molecular Modeling—A high resolution structure of the unliganded HIV-1 RT (Protein Data Bank entry 1trj, resolution = 2.35 Å (13), in combination with the Ca positions in a lower resolution structure of RT complexed with DNA (Protein Data Bank entry 1hmi, resolution = 3.0 Å (7)) was used to produce an all-atom model of the protein in the complex. The Ca positions of the 1rtj structure were slowly forced onto those of the 1hmi structure. This was performed in stages, over more than 100 ps, while immersing the protein in water plus neutralizing counterions. Residues whose secondary structure assignments agreed in the 1rtj and 1hmi structures, using the DSSP program (18), were constrained. Secondary structures assigned to 1rtj were preserved during this forcing process by constraining the appropriate hydrogen bonds. This simulation, as well as those described below, was performed with the AMBER simulation package (19) using the particle mesh Ewald option to accommodate long range interactions (20). The phosphate positions from the 1hmi structure were used to produce an all atom model of the DNA duplex in the complex. The DNA sequence was identical to that in the 1hmi structure. Starting from the structure of an ideal B-form DNA 19-mer duplex, the phosphate positions were slowly forced onto those of the 1hmi structure. Again, this was done in stages over several hundred picoseconds while immersing the DNA in water plus neutralizing counterions. During this process, the Watson-Crick base pair hydrogen bonds were constrained. After annealing the duplex to the 1hmi phosphate positions, the terminal primer strand residue was removed to produce a 1-base overhang as in the 1hmi structure. The DNA was then docked into the protein model. There were no steric conflicts involving protein backbone atoms, however, the side chains of residues in the binding interface were not yet optimally oriented. A conformational search of the relevant side chains was performed using a genetic algorithm with a vacuum molecular mechanics energy fitness function (21). The model complex was centered in a large box of water, and 29 sodium atoms were added to neutralize the system. The total system size was approximately 180,000 atoms. The solvated model system was carefully equilibrated over 150 ps. This was followed by 850 ps of unconstrained molecular dynamics at constant temperature and pressure. A representative structure was generated by averaging the coordinates of the last 600 ps of the trajectory, followed by energy minimization. This average structure was then used to estimate differences in binding free energy between the DNA and wild-type and mutant polymerases. The binding free energy of the DNA model bound to the wild-type or mutant proteins was estimated by calculating the amount of apolar solvent accessible surface area (SASA) buried upon complex formation, and multiplying by −25 cal/mol Å² (22). The conformation of the protein and DNA was assumed to be identical in the free and complexed state. Although this is not strictly correct (22), it was assumed that these errors would cancel one another when calculating differences in binding free energy between wild-type and mutant enzymes. Two methods of the wild-type complex was taken from the time averaged simulation structure, as described above. Conformations of the mutant RT-DNA complexes were obtained by optimal placement of the modified 266 side chains within the wild-type model complex, followed by limited energy minimization. The SASA buried upon complex formation is given by the SASA of the polymerase plus the SASA of the DNA minus the SASA of the complex. To calculate the SASA of an atom on an extended van der Waals surface of the molecule is extended by 1.4 Å, the assumed radius of a water molecule. The van der Waals radii for heavy atoms were taken from Chothia (23). The SASA of an atom in a molecule is then given by the amount of surface area on its extended van der Waals sphere which is not occluded by the extended spheres of other atoms. The apolar SASA for a molecule is obtained by summing the atomic SASA over the apolar atoms. Carbon, sulfur, and phosphorous atoms are considered to be apolar, while nitrogen and oxygen are considered polar. A discrete approximation was used (24), wherein each atom’s surface is represented by approximately 1000 points evenly distributed on its extended van der Waals sphere. Each such point on the surface is tested to see if it is occluded, i.e. within the extended van der Waals sphere of any other atoms in the molecule. The fraction of non-occluded points is multiplied by the surface area of the extended van der Waals sphere to get the SASA for that atom. The SASA of the RT, DNA, and RT-DNA complexes approximated this way are numerically accurate to within a fraction of a Å² (22).

Mutagenesis of the RT Gene—To introduce changes at residue 266 in the RT coding sequence, oligonucleotide-directed mutagenesis was performed, as described by Kunkel et al. (25). The complete coding sequence of the RT gene was sequenced to confirm the presence of the desired substitution and the absence of any other change.

Protein Purification—Mutants were purified as described previously (9) and the purified mutant enzymes were assayed for exonuclease activity. Additional purification to separate p66 homodimer, p66/p51 heterodimer, and/or exonuclease contaminants was accomplished with a ssDNA-cellulose column as described (8). The amount of contaminat-
ing 3' → 5' exonuclease activity on a mismatched primer for each enzyme preparation was determined as described (2). Enzyme preparations had at least 10-fold lower exonuclease activity than Klenow fragment.

Reverse Transcriptase Polymerization Assays—Enzyme activities were determined using a standard reaction mixture (50 μl) containing 50 mM Tris-HCl, pH 7.4 (22°C), 10 mM MgCl₂, and 100 μM KCl. Other reaction conditions are described in the figure legends. Reactions were initiated by addition of enzyme (expressed as concentration of dimer), incubated at 22°C, and stopped by the addition of 20 μl of 0.5 M EDTA, pH 8. Quenched reaction mixtures were spotted onto Whatman DE-81 filter discs and dried. Unincorporated [α-32P]dTTP was removed, and filters were counted as described (26). The K₅₀ for AZTTP, kcat, and substrate Michaelis constants were determined as described previously (9). Enzyme concentrations were determined from protein determinations (27) which had been calibrated by amino acid analysis.

The dissociation rate constant (kₕ₅₀) for poly(rA)-oligo(dT)₉₀ was determined as described (26). Because the dissociation rate constant for many of the Trp²⁵⁶ mutant enzymes was too rapid to measure accurately under these conditions, the pH during the challenge phase of the assay was reduced to pH 6.2. The reduction in pH resulted in a significant reduction in the dissociation rate constant (see “Results”). Enzyme was preincubated with T-P for 5 min before challenging free polymerase with heparin (zero time). The concentration of components in the challenge mixture was 300 nM RT, 300 nM T-P (expressed as primer 3' termini), 50 mM MES, pH 6.2, and 2 mg/ml heparin. To determine the concentration of RT remaining bound to T-P, 20-μl aliquots were removed and mixed with 20 μl of a mixture containing 100 mM Tris-HCl, pH 7.4, 20 mM MgCl₂, and 60 μM [α-32P]dTTP at time intervals after addition challenge. After an additional 10-min incubation, the reaction was stopped with EDTA, and the incorporation of radiolabeled dTMP was determined as described above. Although the dissociation rate constant and kinetic constants were determined on a RNA template and the molecular modeling performed with a DNA template, we have previously demonstrated a quantitative correlation between the magnitude of the effect on the T-P dissociation rate constant for mutant RTs with either a homopolymeric RNA template or a heteropolymeric DNA template (9).

Termination Probability Analysis—Termination probability represents the probability of the polymerase dissociating from the DNA at a defined site. The measurements were performed as described previously (5) using M13mp2 DNA as template primed with a 3'-P-5'-end-labeled 15-mer. The reactions (30 μl) were as described below for fidelity measurements except that they contained a severalfold molar excess of DNA over enzyme and were incubated at 37°C for 5–30 min. Aliquots (20 μl) were analyzed by agarose electrophoresis to ensure the product bands were quantified by phosphorimagery and terminal in parallel with products of sequencing reactions with the same template. The measurements were performed as described previously (5) using M13mp2 DNA as template primed with a 3'-P-5'-end-labeled 15-mer. The reactions (30 μl) were as described below for fidelity measurements except that they contained a severalfold molar excess of DNA over enzyme and were incubated at 37°C for 5–30 min. Aliquots (20 μl) were analyzed by agarose electrophoresis to ensure complete gap filling. All reactions generated products that migrated coincident with nicked, double-stranded DNA.

DNA Polymerase Reactions for Fidelity Studies—Reactions (25 μl) containing 20 mM Hepes, pH 7.8, 10 mM MgCl₂, 2 mM dithiothreitol, four dNTPs each at 1 mM, 32 fmol of gapped DNA, and 0.5–5 pmol of RT were incubated for 1 h at 37°C and terminated by addition EDTA to 15 mM. Aliquots (20 μl) were analyzed by agarose electrophoresis to ensure complete gap filling. All reactions generated products that migrated coincident with nicked, double-stranded DNA.

Fidelity Assay—The frameshift reversion assay uses gapped subtemplate containing 20 mM Hepes, pH 7.8, 10 mM MgCl₂, 2 mM dithiothreitol, four dNTPs each at 1 mM, 32 fmol of gapped DNA, and 0.5–5 pmol of RT were incubated for 1 h at 37°C and terminated by addition EDTA to 15 mM. Aliquots (20 μl) were analyzed by agarose electrophoresis to ensure complete gap filling. All reactions generated products that migrated coincident with nicked, double-stranded DNA.

RESULTS

To define the nature of the interactions occurring between Trp²⁵⁶ of the MGBT and the nucleic acid minor groove, eight alternate side chains were introduced at this position. These side chains differ in their size, hydrophobicity, and hydrogen bonding potential: aliphatic (alanine, valine, isoleucine, leucine), basic (arginine), acidic (glutamate), and aromatic (phenylalanine, tyrosine). The mutant enzymes were expressed in Escherichia coli and purified. Their DNA binding properties were surveyed and compared with calculated relative binding free energies determined from a RT-DNA model constructed by molecular modeling.

Molecular Modeling of the RT-DNA Complex—Using information from the HIV-1 RT structure of apoenzyme (1rtj, resolution = 2.35 Å) (13) and from the lower resolution structure with DNA (1hmi, resolution = 3.0 Å) (7), we have built an all atom fully solvated model of the RT complexed with a DNA duplex (T-P: 19–18-mer). An earlier version of this model helped identify a structural motif which interacts in the DNA minor groove several base pairs upstream of the polymerase active site which we called the MGBT (11). The duplex DNA in that model did not have a template overhang; rather the product of one step of polymerization was modeled as a base pair with the 5'-terminal template base. Before proceeding with detailed modeling of the residue 266 mutants, we decided to refine our previous model: removing the 3'-nucleotide at the primer terminus, applying very careful initial modeling procedures, and using a much larger solvent layer as described under “Experimental Procedures.” The calculation of the free energy of binding for a ligand such as DNA to a protein is a difficult task. In general, the gains in free energy from favorable polar interactions involved in binding are approximately canceled by the loss of solvent contact in the polar groups involved, whereas apolar binding interactions are always favorable (28). Thus, it is assumed that polar contributions to binding free energies are negligible. We estimated binding free energies by calculating the amount of apolar SASA buried upon complex formation, and multiplying by −25 cal/mol Å² (22). The calculated apolar SASA buried upon complex formation for each mutant is tabulated in Table I. From this, binding free energies were calculated and expressed relative to that of the alanine mutant. The aromatic side chains (phenylalanine and tyrosine) are suggested to be most similar to the wild-type tryptophan residue in the amount of apolar surface area buried upon DNA binding. The larger aliphatic side chains (leucine and isoleucine) also buried more surface area than valine and alanine. Finally, glutamate and arginine

| Side chain  | Apolar  | Polar  | ΔG_SASA  | Relative ΔG_SASA  |
|------------|---------|--------|----------|------------------|
| Alanine    | 25.3    | 0.0    | −40.57   | 0.0              |
| Arginine   | 42.1    | 70.6   | −41.27   | −0.7             |
| Glutamate  | 35.4    | 47.3   | −41.30   | −0.7             |
| Valine     | 49.9    | 0.0    | −41.50   | −0.9             |
| Isoleucine | 63.2    | 0.0    | −41.90   | −1.3             |
| Leucine    | 63.5    | 0.0    | −41.85   | −1.3             |
| Tyrosine   | 67.2    | 28.6   | −42.27   | −1.7             |
| Phenylalanine | 72.4   | 0.0    | −42.47   | −1.9             |
| Tryptophan | 74.4    | 8.9    | −42.73   | −2.2             |

* The individual side chains at position 266 were modeled as described under “Experimental Procedures.”

* Buried surface area of the solvent accessible residue of surface 266. Carbon, sulfur, and phosphorous were considered apolar, while nitrogen and oxygen were considered polar.

* An estimate of the free energy of binding was calculated from the amount of total apolar SASA buried upon complex formation where each Å² of buried apolar surface area contributes −25 cal/mol at 25°C (22).

* Calculated relative to alanine: [relative ΔG_SASA] = [ΔG_SASA]/ΔG_SASA, where ΔG_SASA refers to the free energy contribution of the apolar SASA of side chain X.
bury a significant amount of apolar surface area, although these side chains would be expected to be charged. The total (polar and apolar) buried surface areas for several of the 266 side chains are illustrated in Fig. 2. The modeling suggests that, except for arginine, the area of interaction with the other mutant 266 side chains overlaps with that observed for tryptophan and occurs exclusively with the sugar and bases of the primer strand three to four nucleotides upstream of the polymerase active site.

**Steady-state Kinetic Analysis**—To determine the influence of altering the chemical nature of the side chain of residue 266, we analyzed the steady-state kinetic, DNA binding, and frameshift fidelity properties of the altered enzymes. Trp266 mutation had very little effect on dNTP binding or catalysis as revealed by steady-state analysis. Vertical-scanning mutagenesis of Trp266 resulted in a less than 2-fold effect on catalytic efficiency ($k_{cat}/K_m$, dTTP) for the eight mutants relative to the wild-type enzyme (Table II). Under the assay conditions, $K_m$, dTTP and $k_{cat}$ are approximately equivalent to $K_d$, dTTP and $k_{3}$, respectively (Scheme 1) (8).

$$\text{RT} + \text{T-P} \rightleftharpoons k_1 \text{RT}^{\text{TP}} + \text{dTTP} \rightleftharpoons k_2 \text{RT}^{\text{TP},\text{p}} \rightleftharpoons k_3 \text{dATTP} \rightleftharpoons k_{\text{dATTP}} \text{RT} + \text{T-P},$$

**Scheme 1**

In contrast, mutagenesis of Trp266 resulted in a more pronounced effect on $K_m$, T-P consistent with its role in nucleic acid binding in the minor groove. The $K_m$, T-P of the W266E mutant increased 10-fold, while the alanine and valine mutations resulted in a greater than 3-fold increase (Table II). The other Trp266 mutants had a less than 2-fold effect on $K_m$, T-P.

The ability of a chain terminating nucleoside, such as AZTTP, to inhibit RT-dependent polymerization is a convenient and sensitive measure to determine whether substrate binding or catalysis is influenced by side chain substitution (8). For a processive polymerase ($k_3 > k_{d(T-P)}$), $K_i$, AZTTP = ($K_{d(\text{AZTTP})}k_{d(T-P)}/k_3$) is an
apparent equilibrium constant that only depends partially on $K_{\text{dAZTTP}}$. The $K_{\text{dAZTTP}}$ is also dependent on how well it is incorporated ($k_3$) and how tightly the polymerase binds nucleic acid ($k_{\text{off,T-P}}$). Alanine substitution for Trp$^{266}$ has previously been demonstrated to result in AZTTP resistance in vitro, and this resistance has been correlated with an elevated T-P dissociation rate constant (8). Substitution of Trp$^{266}$ with an acidic side chain (i.e. glutamate) resulted in a mutant enzyme that was less sensitive to AZTTP than when alanine was substituted at this position (Fig. 3, ordinate). In general, larger side chains had less of an effect on $K_{\text{dAZTTP}}$ (tyrosine < phenylalanine < arginine) than the smaller ones (leucine ~ alanine > valine) relative to wild-type enzyme (i.e. tryptophan). Interestingly, isoleucine substitution for Trp$^{266}$ resulted in a significantly lower $K_{\text{dAZTTP}}$ than leucine substitution, although both side chains are apolar, have the same van der Waals volume (124 Å$^3$), and bury similar SASA upon DNA binding (Table I). Since isoleucine branches at Cβ and leucine at Cγ, this suggests that the specific region of the SASA buried can influence T-P binding.

**T-P Dissociation Rate Constant ($k_{\text{off,T-P}}$)**—We previously demonstrated that alanine substitution for Gly$^{262}$ or Trp$^{266}$ resulted in weak nucleic acid binding because of a significant increase in the T-P dissociation rate constant (8). The rapid dissociation rate constant of the W266A mutant at pH 7.4 (1/2 $t_{1/2}$ < 3 s) precluded accurate comparison of many of the Trp$^{266}$ vertical scan mutants. Therefore, the conditions of the competition assay to determine $k_{\text{off,T-P}}$ were modified to decrease the rate of nucleic acid dissociation. Previous work has suggested that, in some instances, the processivity of a DNA polymerase may be enhanced at low pH (29, 30). These results suggest that the dissociation rate constant may be pH-sensitive and decreased at the lower pH. Indeed, lowering the pH of the challenge phase of the competition assay (see “Experimental Procedures”) resulted in a 10-fold reduction in the dissociation rate constant for wild-type enzyme (data not shown). This facilitated accurate determination of the dissociation rate constants for the Trp$^{266}$ mutants (Fig. 3, abscissa). In general, larger side chain substitutions at residue 266 resulted in a slower dissociation rate constant. However, the carboxylic acid substitution, W266E, displayed the most rapid T-P dissociation rate constant. Fig. 3 also indicates that both $k_{\text{off,T-P}}$ and $K_{\text{dAZTTP}}$ are influenced in a similar manner even though they were determined at pH 6.2 and 7.4, respectively. Thus, the dissociation rate constants determined at low pH are kinetically relevant to what occurs at pH 7.4.

**Processivity Measurements**—Our earlier work demonstrated that the W266A RT mutant had a significantly reduced processivity relative to the wild-type enzyme that was attributed to its elevated T-P dissociation rate constant (11). We therefore examined the processivity of the Trp$^{266}$ vertical scan mutant RTs. Primer extension reactions using the lacZ α-complementation sequence as template DNA were performed with six mutant RTs and the wild-type enzyme, under conditions that prevent reinitiation of synthesis on previously used template-primer sequences. In a single cycle of processive synthesis, the wild-type RT generates products that vary in length with the probability of termination of processive synthesis dependent on the DNA sequence (Fig. 4A). All mutant RTs have reduced processivity relative to the wild-type enzyme (Fig. 4B). However, the degree of the reduction varied depending on the mutant, suggesting that the RT-T-P interactions are modified depending on the nature of the side chain at residue 266.

The largest reduction in processivity is observed when a glutamate is present at position 266, with average product lengths even shorter than those generated by the W266A mutant. The lack of sufficient product formation beyond position 60 and 80 for W266A and W266E, respectively, precluded the calculation of termination probabilities at these positions. Substitution of tryptophan by a hydrophobic amino acid (i.e. leucine, isoleucine, or valine) results in RTs which have similar processivity. The processivity of these mutants is significantly lower than that of the wild-type enzyme but higher than the processivity of the alanine or glutamate mutants. The presence of a basic side chain (i.e. arginine) at position 266 results in a modest reduction of processivity, while RT with tyrosine has processivity similar to that of the wild-type enzyme.

As noted previously, wild-type RT displays a sequence dependent site-to-site variation in the amount of termination (4, 5) (Fig. 4A). This variation is also observed with the mutant RTs. Since termination probability is dependent on the dissociation rate constant at each position along the template, the termination probability pattern on a heteropolymeric template allows numerous sequence contexts to be easily examined. Interestingly, two of the strongest termination sites common to nearly all mutant RTs occur after insertion of A opposite the second T in the repeated sequence 3’-TGC-TGC-5’ (position 67 and 80) (Fig. 4B). These sites also exhibit an elevated termination probability with wild-type enzyme (Fig. 4E).

**Frameshift Fidelity of Mutant RT with Substitutions at Trp$^{266}$**—We have previously demonstrated a strong correlation between sites of strong termination in homopolymeric runs and sites of frameshift errors (4–6). The substitution to alanine at Trp$^{266}$ results in a RT with reduced fidelity for T-P slippage errors (10, 11). Thus, we have examined the frameshift fidelity of the seven RTs with other than alanine substitutions at this position. A M13mp2-based frameshift reversion assay was employed, designed to score single base deletions in short homopolymeric sequences. The target for measuring deletion events contains three short homopolymeric sequences, two 3-base T runs separated by a 4-base C run (see “Experimental Procedures”).

The results indicate that, except for the tyrosine mutant, all RTs with a substitution at position 266 have significantly reduced frameshift fidelity (Fig. 5). Dark blue revertants may result from a single-base deletion in one of the three adjacent homopolymeric sequences. Thus, to determine the site of the mutation, the DNA from 20 revertants generated in reactions with each of the mutant polymerases were analyzed by se-
This analysis revealed that the vast majority of revertants produced in reactions with each enzyme had a deletion in the T run located 3' to the four Cs (data not shown).

**DISCUSSION**

The MGBT is an important structural motif that contributes critical side chains for interaction in the DNA minor groove 2–6 base pairs from the active site (11). Loss of these interactions is accompanied by a decrease in DNA binding affinity and frameshift fidelity. To analyze the nature of the interactions between the MGBT and the nucleic acid, we have constructed and overproduced eight mutant forms of RT where 1 residue of the MGBT, Trp266, was altered to change the chemical nature of the side chain. In addition to the functional analysis of these enzymes, we have performed molecular dynamics modeling to gain insight into possible nucleic acid interactions occurring with each mutant. From our modeling efforts, binding free energies were calculated for each residue 266 side chain (Table I). The magnitude of the differences in the DNA binding free energy between the wild-type and mutant polymerases predicted that there would be changes in RT/T-P interactions with the mutants.

As observed previously with an alanine mutant of Trp266 (8), alternate substitutions for this residue did not have a large effect on the catalytic efficiency (Table II). However, the properties of the enzyme that reflect its interactions with the T-P, namely the T-P dissociation rate constant ($k_{\text{off}}$) (Fig. 3), $K_{\text{i,AZTTP}}$ (Fig. 3), processivity (Fig. 4), and frameshift fidelity (Fig. 5) were altered in a similar manner with the mutants relative to the wild-type RT. The magnitude of the alteration depended on the nature of the side chain at position 266.
HIV-1 RT DNA Minor Groove Interactions

Whereas the large aromatic side chains of tyrosine and phenylalanine yielded RTs with properties most similar to those of the wild-type enzyme, glutamate at position 266 drastically diminished DNA binding, processivity, and frameshift fidelity. Substitution with the aliphatic side chains of leucine, isoleucine, or valine resulted in an intermediate level of DNA binding, processivity, and frameshift fidelity. While the glutamate side chain, which buries slightly less apolar surface area than phenylalanine, has an apparent DNA binding affinity closer to that of the alanine substitution, the arginine mutant displayed DNA binding that was more similar to wild-type than predicted by the burial of surface area alone. This difference is not unexpected. The low binding affinity of the glutamate mutant could be the result of repulsion between the negatively charged side chain and the sugar-phosphate backbone of the nucleic acid. Since this destabilizing polar interaction is not accounted for in the binding free energy calculations, the dissociation rate constant for the glutamate mutant lies above the line in Fig. 6. In contrast, arginine, like tyrosine, has the ability to form a hydrogen bond with the DNA. In the model structure, one of the nitrogens of the guanidinium group is within hydrogen bonding distance to N-3 of the purine at the fifth template position. The ability to form a hydrogen bond may explain why the tyrosine side chain, which buries slightly less apolar surface area than phenylalanine, has an apparent DNA binding affinity closer to wild type than any other side chain substitution examined (Fig. 6). The hydroxyl group of Tyr266 is near N-2 of a primer strand guanosine at the fourth position. As with arginine and tyrosine, the dissociation rate constant for the wild-type tryptophan side chain, which can also donate a hydrogen bond, is slower (i.e. higher binding affinity) than would be predicted from the correlation observed in Fig. 6. Thus, although the binding free energy for RT depends on hydrophobic interactions, the ability to form a hydrogen bond may also contribute to complex stability.

To fulfill their function in DNA replication and repair, DNA polymerases should bind nucleic acid with little or no sequence specificity. X-ray crystal structures of several different DNA polymerases in complex with DNA indicate that like RT, the majority of interactions occur in the DNA minor groove and with the sugar-phosphate backbone (31–33). There are very few enzyme-DNA contacts observed in the major groove. Sequence-specific DNA-binding proteins generally take advantage of the specific hydrogen bonding pattern offered by the major groove. In contrast, the minor groove offers very little hydrogen bonding discrimination with regard to its nucleotide sequence. The nature and the extent of the minor groove interactions vary depending on the DNA polymerase. The DNA repair polymerase, DNA polymerase β, exhibits van der Waals and hydrogen bonding interactions with the sugar-phosphate backbone and the minor groove (31). In contrast with RT, these interactions are confined to the first 2–3 base pairs of the duplex. A high-resolution structure of Bacillus stearothermophilus DNA polymerase I in complex with DNA (33), as well as the ternary complex structure of T7 DNA polymerase (32), indicate that minor groove hydrogen bonding and sugar-phosphate backbone interactions extend several base pairs further upstream from the active site. In addition, extensive contacts with the sugar-phosphate backbone upstream from the active site have also been observed in the binary complexes of Taq DNA polymerase (34) and the Klenow fragment of E. coli DNA polymerase I (35).

Our work on the MGBT in RT indicates that enzyme-DNA interactions in the minor groove are functionally significant. Additional evidence for the functional importance of the MGBT is the fact that these residues are highly conserved among retroviral polymerases (11). Perturbing these interactions through protein modifications decrease RT binding affinity with a concomitant decrease in processive polymerization (Figs. 3 and 4B) and a lower frameshift fidelity (Fig. 5) (11, 17). Alteration of these minor groove interactions by altering the nucleic acid structure with site-specific DNA lesions has also
confirmed the functional importance of the MGBT residues in α-helix H (36, 37).

In contrast, deletion of putative electrostatic and van der Waals interactions between the sugar-phosphate backbone of the template strand and residues of α-helix I did not alter the DNA binding properties of the mutant RTs (9). However, alanine mutants of α-helix I exhibited a greater proportion of enzyme that was inactive when assayed for polymerase activity. The structures of the polymerase-DNA complexes discussed above suggest that hydrogen bonding may play a significant role in binding or fidelity. An arginine residue of DNA polymerase β (i.e. Arg283), which forms the van der Waals surface for the nascent base pair, is also within hydrogen bonding distance to the template sugar of the first base pair of the duplex. Alteration of these contacts by site-directed mutagenesis decreases DNA binding, dNTP binding, catalytic efficiency, and fidelity (38). In contrast, deletion of the hydrogen bonding in the DNA minor groove between the primer base and Tyr271 of DNA polymerase β by replacing this residue with phenylalanine had little or no influence on apparent DNA binding affinity, dNTP binding, catalytic activity, or fidelity. Therefore, the putative role(s) of each “observed interaction” in a model or crystal structure needs to be probed by experimental study.

The results presented here demonstrate that binding of RT to T-P may be modified by incremental changes in the nature of the interactions between a single amino acid side chain and the DNA. Since polymerases are sensitive to the DNA sequence (39–42), it is also important to consider how the interactions described here may be altered by the sequence-dependent structure of the nucleic acid (e.g. minor groove dimensions, secondary structure, and/or DNA flexibility). These changes in nucleic acid structure could alter critical protein-nucleic acid interactions resulting in mutational hot spots (4, 6) and/or strong sites for termination of processive synthesis (5). Structural studies and biochemical analysis indicates that repetitive sequences in DNA have a propensity for assuming nonconventional structures (43). A local structural aberration could be the cause of the strong termination of processive synthesis within the short trinucleotide repetitive sequence reported here. Since termination and mutational hot spots are often unique for each DNA polymerase (44), the protein-nucleic acid interactions are uniquely defined for each enzyme.

Acknowledgments—We thank Drs. William C. Copeland, Traci M. Hall, and Lars C. Pederson for critical reading of the manuscript.

REFERENCES

1. Kohlstaedt, L. A., Wang, J., Friedman, J. M., Rice, P. A., and Steitz, T. A. (1995) Science 266, 1783–1790
2. Roberts, J. D., Bebenek, K., and Kunkel, T. A. (1998) Science 242, 1171–1173
3. Preston, B. D., Poiesz, B. J., and Loeb, L. A. (1988) Science 242, 1168–1171
4. Bebenek, K., Abbotts, J., Roberts, J. D., Wilson, S. H., and Kunkel, T. A. (1989) J. Biol. Chem. 264, 16948–16956
5. Abbotts, J., Bebenek, K., Kunkel, T. A., and Wilson, S. H. (1993) J. Biol. Chem. 268, 10312–10323
6. Bebenek, K., Abbotts, J., Wilson, S. H., and Kunkel, T. A. (1993) J. Biol. Chem.