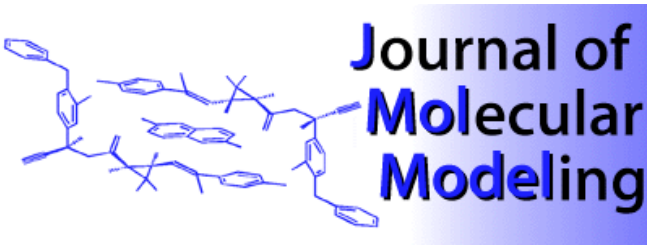
*J. Mol. Model.* **2000**, *6*, 575 – 586




**FULL PAPER**



# A Molecular Dynamics Model of HIV-1 Reverse Transcriptase Complexed with DNA: Comparison with Experimental Structures

**Leping Li[1], Lee G. Pedersen[1,2], William A. Beard[1], Katarzyna Bebenek[1,3], Samuel H. Wilson[1], Thomas A. Kunkel[1,3], and Thomas A. Darden[1]**

[1]Laboratory of Structural Biology, National Institute of Environmental Health Sciences, P.O. Box 12233, Research Triangle Park, NC 27709. Tel: 919-541-4933; Fax: 919-541-7880. E-mail: darden@g-rex.niehs.nih.gov.

[2]Department of Chemistry, University of North Carolina, Chapel Hill, NC 27709-7630

[3]Laboratory of Molecular Genetics, Nat. Institute of Environmental Health Sciences, Research Triangle Park, NC 27709

*Received: 17 February 2000/ Accepted: 14 August 2000/ Published: 17 November 2000*

**Abstract** We have built a molecular dynamics model for human immunodeficiency virus (HIV-1) reverse transcriptase (RT) complexed with a 19/18-mer template/primer by combining the structural information of a low resolution crystal structure of a HIV-1 RT/DNA complex (1hmi) with that of a high resolution crystal structure of unliganded HIV-1 RT (1rtj). The process involved slow forcing of the α-carbons of 1rtj onto those of 1hmi using constrained MD simulations, while immersing the protein in aqueous solution. A similar technique was used to build the bent all-atom DNA duplex, which was then docked into the modeled protein. The resulting model complex was refined using molecular dynamics simulation with the Particle-mesh Ewald method employed to accommodate long-range electrostatic interactions. New parameters of the Amber force field that affect DNA twist are tested and largely validated. The model has been used successfully to explain the results of vertical scanning mutagenesis of residue 266 (Trp266). Recently, the low resolution crystal structure of the HIV-1 RT/DNA complex has been refined to a 2.8 Å resolution (2hmi) and a crystal structure of a HIV-1/RT/dTTP ternary complex has been determined at 3.2 Å resolution (1rtd). A detailed structural comparison of the *prior* model structure and the two experimental structures becomes possible. Overall, the three structures share many similarities. The root mean square deviations of the α-carbons for the individual subdomains among the three structures are within the same ranges. The secondary structure assignments in the three structures are nearly identical. Key protein-DNA contacts such as those in the region of the primer grip are also similar in the three structures.

**Keywords** HIV-1 reverse transcriptase, Minor groove binding track, Particle-mesh Ewald

**Abbreviations** 1hmi: low resolution crystal structure of a HIV-1 RT/DNA/Fab complex; 2hmi: refined crystal structure of 1hmi; 1rtj: crystal structure of a high resolution unliganded HIV-1 RT; 1rtd: crystal structure of a HIV-1 RT/DNA/dTTP ternary complex

**Running title** Model of HIV-1 reverse transcriptase complexed with DNA.

*Correspondence to:* T. A. Darden

## Introduction

The reverse transcriptase (RT) of human immunodeficiency virus (HIV-1) is essential for viral replication. It is a heterodimer of 66- and 51-kDa polypeptides with RNA- and DNA-dependent DNA synthesis and RNase H endonuclease activities. The p51 subdomain is a carboxy-terminal truncation of p66 generated by HIV-1 protease. The p66 and p51 subunits share four common, but spatially different, subdomains designated as fingers, palm, thumb, and connection. The p66 subunit also contain a subdomain with RNase H activity at the carboxy-terminus. The fingers, palm, and thumb of p66 form a nucleic acid binding cleft, called the polymerase domain.

HIV-1 RT not only lacks a 3'→5' proofreading exonuclease activity, [1] but also is inaccurate. [1-3] The low replication fidelity of HIV-1 RT may account for its drug-resistance. It has been shown that the replication errors of HIV-1 RT such as single base deletions, additions, and substitutions in homopolymeric sequences, are largely mediated through a template-primer slippage mechanism. [4, 5] Frameshift error rates are DNA sequence-dependent and correlate with the probability of termination of processive synthesis, which depends on the sequence of the template-primer from one to six nucleotides away from the 3'-terminus. [3, 6] Functional studies of the wild type RT as well as its mutants have revealed that the molecular mechanisms for processivity and low replication fidelity are complex. [5,7-9] Clearly, these properties depend not only on the sequence and conformation of the template-primer, but also on the interactions between the enzyme and the template-primer. [5,7-9] The sequence effects may include conformational flexibility, minor groove width, hydrogen bonding specificity, and others. [10]

Analysis of the original low resolution X-ray crystal structure (1hmi) of HIV-1 RT complexed with a 19-mer/18-mer template-primer suggested that helix H in the thumb of p66 contacts the primer strand. [11] Alanine-scanning mutagenesis studies on helices H and I in p66 have confirmed the role of helix H in RT/template-primer binding, processivity, and frameshift fidelity. [5, 8, 9] In order to gain insight into the interactions between RT and template-primer, an all-atom structure of the protein/template-primer complex was required. Unfortunately, the only relevant experimental structure of the RT complexed with a template-primer (pdb code: 1hmi, resolution 3.0 Å) was incomplete with only α-carbons defined for the protein and phosphate atoms for the template-primer. [11] Although high-resolution structures of unliganded HIV-1 RT in different crystal forms [12-14] and those of HIV-1 RT complexed with non-nucleotide inhibitors [15-17] were available, those structures do not have information for the template-primer. Accordingly, a molecular modeling technique was sought.

In our initial attempt to find an appropriate all atom model, the method of Levitt [18] was used to build the full protein backbones of the p66 and p51 monomers, using the α-carbon positions of 1hmi [11] and a rotamer library [19] was used to build the sidechain positions. Since the template-primer is bent by approximately 45 degrees around the sixth base pair and the conformation of the DNA near the active site is A-like while that at the other end of the bend is B-like, the DNA was built by overlaying short fragments of duplex DNA of length three onto the phosphate positions from the 1hmi structure. For each fragment, ideal A-form and B-form DNAs were generated and the form that best fitted the corresponding phosphate positions from the 1hmi structure was chosen. Using these techniques, the first all-atom model of the RT/DNA complex was generated. [5] This model complex was used to analyze the results of alanine-scanning mutagenesis on helix H of the thumb domain of p66. [5] To improve the sidechain packing of the initial model, we modified the above procedure by using only fragments of the unliganded RT (pdb code: 1rtj, resolution 2.35 Å) [12] to build the protein. The resulting model complex was then solvated, and 200 picoseconds (ps) of unrestrained molecular dynamics was performed. This solvated model facilitated the identification of an important structural element for DNA binding and frameshift fidelity, the minor groove binding track (MGBT). The MGBT constitutes five residues (Ile94 in β-sheet β5, Gln258, Gly262, Trp266, and Gln269 in helix H). [7] Those residues are spatially close and protrude into the minor groove of the template/primer contacting the template-primer residues from the second to the sixth base pair upstream from the active site.

While many details of the second model appeared satisfactory, and the secondary structure assignments following molecular dynamics were clearly superior to the initial model, we felt that the model building approach could be yet improved, especially the technique used to build the side chains. A third model was developed by combining the structural information of the low resolution RT/DNA complex [11] with that of the high resolution unliganded RT. [12] Although the third model shares many similarities with the early models, improvements over the early models are apparent. The third model has been used successfully to explain the results of the vertical scanning mutagenesis of residue 266 (Trp266). [20]

After our third model was completed, the low-resolution structure of the RT/DNA/Fab complex was refined to a relatively high resolution (pdb code: 2hmi, resolution = 2.8 Å). [21] This structure was then used as the starting point for a series of in vacuum molecular dynamics simulations. [22] Almost simultaneously with the 2hmi structure publication, Huang et al. [23] reported the structure of a RT/DNA/dTTP ternary complex of 3.2 Å resolution (pdb code: 1rtd). To arrive at a stable ternary complex, one of the MGBT residues Gln258 in helix H of p66 was replaced with a cysteine residue that was later cross-linked to a thiol group of an engineered guanine residue in the minor group of DNA. This modification was motivated by analysis of our earlier model. [7] Thus, a detailed structural comparison among the three structures becomes possible. Herein, we report our latest modeling effort and discuss the structural similarities and

differences among the three structures in terms of their secondary structures, backbone chain traces, some of the important contacts among protein residues, and the protein/DNA interactions.

## Methods

### *Preparation of p66/p51*

The low-resolution structure of a HIV-1 RT/DNA/Fab complex (pdb code: 1hmi, resolution = 3.0 Å) [11] and the high resolution structure of an unliganded HIV-1 RT (pdb code: 1rtj, resolution = 2.35 Å) [12] were used to build the structure of the p66 and p51 monomers. Initially, hydrogen atoms were added to 1rtj and the missing loop (residues 218-230) of p51 in 1rtj was modeled using a combination of database fragments for backbone positions and a rotamer library [19] for sidechain positions. The molecule was then solvated in a large box of Monte Carlo TIP3P water with each box side at least 10.0 Å away from the nearest atoms of the complex. Six $Cl^-$ ions were included to neutralize the charge of the system**.** The ions were placed randomly (by replacing water molecules) in the system 10 Å away from each other and from the nearest protein/DNA atoms. The hydrogen atoms, newly introduced loop, counterions, and water molecules were then minimized. The water and counterions were equilibrated at 300 K for 50 picoseconds (ps). The simulation was carried out with periodic boundary conditions at constant volume. The α-carbons of 1rtj were then slowly forced onto those of 1hmi by a constrained molecular dynamics simulation using the Amber 4.1 package, [24] after the α-carbons of the 1hmi structure were overlaid onto those of 1rtj. This was accomplished by a series of 100 simulations of 2 ps each. During each simulation the α-carbons of the model were constrained to the reference α-carbons. The reference α-carbons were updated according to the following scheme:

$$\alpha\text{-carbons}_{ref} = \lambda \cdot \alpha\text{-carbons}_{1hmi} + (1.0 - \lambda) \cdot \alpha\text{-carbons}_{1rtj}$$

where λ varied from 0 to 1 in steps of 0.01. During the equilibration, the secondary structures in 1rtj were preserved by constraining the corresponding hydrogen bond angles and distances. Since the backbone chain traces for residues 284-295, 314-335, and 416-420 in p66 and residues 13-16, 44-46, 94-103, 276-295, 313-335, and 419-427 in p51 in 1rtj and 1hmi are significantly different, [17] the α-carbons of those residues of 1rtj were not forced onto those of 1hmi during the slow forcing process. After the slow forcing process was completed, the system was further equilibrated for 20 ps with the α-carbons constrained to those of 1hmi. An average structure was generated from the coordinates of the 20 ps simulation by a rms overlay of backbone atoms onto the starting structure (0 ps). The resulting average structure **was** used as the starting structure for the p66 and p51 monomers.

### *Preparation of DNA*

The same approach used to build the protein model was employed to construct the all atom bent template-primer. The phosphate positions of an all-atom canonical ideal B-DNA structure corresponding to the sequence of 1hmi were slowly forced as described above, onto those of 1hmi, while immersing the B-DNA in water. Sodium ions were included to neutralize the charge of the system. The ions and water molecules were equilibrated for 100 ps. During the slow forcing procedure, hydrogen bonds (both angles and distances) involved in the Watson-Crick base pairs were constrained. After the slow forcing was completed, the system was further equilibrated with the phosphate positions constrained to those of 1hmi for 200 ps. An average structure was generated from the last 100 ps simulation and the resulting average structure was used as the starting model for the DNA.

### *Docking DNA into p66/p51*

The protein and template-primer were merged according to the positions of the α-carbons and phosphate atoms in 1hmi. As expected, there was no steric clash between protein atoms or between DNA atoms. However, a few protein sidechains at the protein/DNA interface were too close to the DNA. The orientations of the offending sidechains were subsequently optimized using a genetic algorithm, [25] with energies evaluated *in vacuum* using a distance-dependent dielectric constant. In brief, the sidechain conformations of the residues that were in close contact with DNA and those of the surrounding residues were represented by binary chromosomes. We used six bits to represent each torsion angle, which is equivalent to a torsion angle resolution of approximately 5 degrees. Initially, a population of ”chromosomes” was constructed, in which the binary bits were randomly generated. The binary content of each chromosome was then decoded to real torsion angles followed by conformation generation. The non-bonded interaction energy was calculated for each conformation (chromosome) *in vacuum* using the Amber 94 force field [26] with a distance-dependent dielectric function and taken as the fitness score for the chromosome. Selection of chromosomes for the next generation consisted of two steps. The single best chromosome was passed on to the next generation deterministically while the remaining population was filled using a *roulette-wheel* selection procedure. The mutation rate for the bits was set to be 0.02. A sufficient number of runs were performed to result in a pseudoconvergence. The conformation with the lowest non-bonded energy that converged was taken as the starting conformation for the subsequent molecular dynamics refinement.

### *Molecular dynamics simulations*

The molecular dynamics simulation procedure was similar to that reported earlier. [25] Briefly, the simulations were performed using the Amber 4.1 package [24] with the smooth

Particle-mesh Ewald (PME) method [27, 28] employed to accommodate long-range electrostatic forces. The simulations utilized the Amber all-atom force field (parm94) [26] with a step size of 2 femtoseconds (fs). By using an extended list technique, the non-bonded interactions were effectively updated every step with a small overhead in computational cost. The non-bonded cutoff for van der Waals interactions was set to 8 Å. All covalent bonds involving hydrogen atoms were constrained using SHAKE. The starting complex was solvated in a large box of Monte Carlo TIP3P water. Sodium ions were included to neutralize the charge of the system. The ions were placed using a procedure similar to that for the $Cl^-$ ions (see above). The total number of atoms in the simulation system was approximately 180,000. The ions and water molecules (approximately 54,000 total) were then equilibrated for 100 ps at constant volume followed by energy minimization while freezing the model complex. For convenience, we refer to the minimized complex before molecular dynamics simulation as the starting model. Next, positional constraints of 10.0, 1.0, and 0.1 kcal/mol·Å$^2$ were placed on a group of selected α-carbons and all phosphate atoms during sequential simulation periods of 0-10, 11-50, and 51-100 ps, respectively. The α-carbons of residues 284-295, 314-335, 416-420 of p66 and residues 13-16, 44-46, 94-103, 276-295, 313-335 and 419-428 of p51 whose chain traces do not agree between 1rtj and 1hmi [17] were not constrained. During the latter simulations, the secondary structures of the protein and the Watson-Crick base pairs were constrained via the corresponding hydrogen bonds (both distances and angles). The positional constraint was then removed and the simulation was continued for another 150 ps with only the hydrogen bond constraint in place. Next, a *totally unrestrained* molecular dynamics simulation was performed for 850 ps at constant pressure.

Since a significantly reduced twist (compared to ideal B-DNA) of the DNA near the RNase H subdomain occurred during the course of MD simulation and since the reduced twist is possibly an artifact of the Amber 94 force field parameters for nucleic acids, the simulation was switched to a recently revised Amber 98 force field (parm98) [29] at 1,000 ps. The only differences between the two force fields are the torsion parameters for the sugar phosphate backbone. In parm98, an anomeric torsion of CT-OS-CT-N* has been added and the χ torsion parameters have been updated. In addition, the torsion for O-C-C-O of the sugar has been modified. The simulation was continued for additional 1,000 ps using parm98. An average structure was generated from the coordinates of the simulation periods 1,332 to 1,750 ps. The resulting average structure was energy minimized for 100 steps *in vacuum* using a distance-dependent dielectric function and is referred to as the model structure.

## Results and discussion

### *The overall structure*

The secondary structures for p66 and p51 domains are nearly identical to those of 2hmi, 1rtj and 1rtd (Tables 1 and 2). The rms deviation of the α-carbons of 2hmi compared to those of 1rtd is 2.13 Å. It is known that binding of an incoming nucleotide (dTTP in 1rtd) to polymerases induces conformational changes of the protein, especially in the regions of the fingers and palm subdomains. [30-33] Thus, the rms deviation of the α-carbons between the two crystal structures may not be surprising. The rms deviations of the α-carbons of the model compared to those of 2hmi and 1rtd are 2.46 and 2.94 Å, respectively. However, when only individual domains are

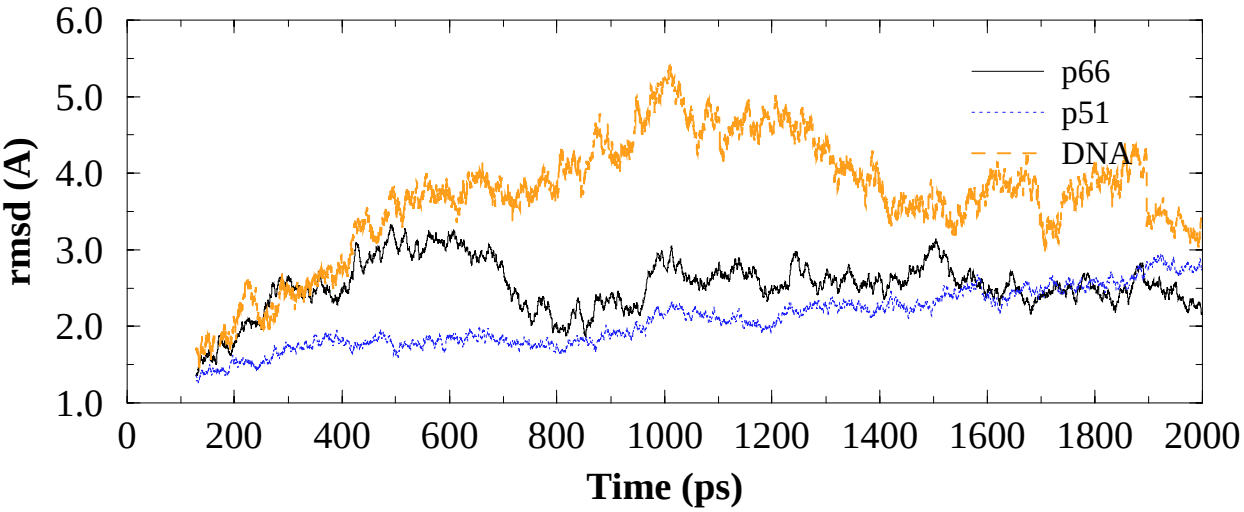



**Figure 1** *Root mean square deviations of the #a-carbons of the p66 and p51 monomers and the phosphate atoms of the 19-/18-mer template/primer compared to those of the starting structure as a function of simulation time*

**Table 1** *Secondary structural assignments for p66 domains [a]*

| 1hmi[b] | 2hmi | 1rtj | 1rtd | model |
|---|---|---|---|---|
| **Helix** | | | | |
| 28-44 | 28-44 | 28-43 | 28-43 | 28-43 |
| 78-83 | 78-81 | 78-82 | 80-83 | 78-83 |
| — | 97-99[c] | 97-99[c] | 97-99[c] | 97-99[c] |
| 114-117 | 114-116[c] | 112-117[c] | — | 111-116[c] |
| 122-127 | 122-125 | 125-126[c] | 124-128 | 122-128[c] |
| 155-174 | 155-173 | 156-174 | 156-173 | 155-174 |
| 195-212 | 195-211 | 195-210 | 195-211 | 195-210 |
| 255-268 | 254-267 | 254-267 | 254-269[d] | 254-269[d] |
| 278-286 | 277-283 | 277-283[e] | 277-283[e] | 277-282 |
| 298-311 | 296-311 | 297-311 | 297-310 | 297-309 |
| 364-382 | 364-382 | 364-383 | 364-383 | 364-382 |
| 395-404 | 395-404 | 395-402 | 395-404 | 395-402 |
| 474-488 | 474-487 | 474-488 | 474-487 | 474-488 |
| 500-508 | 500-507 | 501-508 | 500-507 | 500-508 |
| 516-527 | 516-527 | 516-526 | 516-527 | 516-527 |
| 544-555 | 545-552 | N/A[f] | 545-551 | 546-553 |
| **Sheet** | | | | |
| 7-12 | — | — | — | — |
| 18-24 | — | — | — | — |
| 49-51 | 47-49 | — | 47-48 | 47-49 |
| 56-63 | 60-62 | 60-65 | 60-64 | 60-64 |
| 73-77 | 73-75 | 70-75 | 71-75 | 71-75 |
| 86-90 | — | — | — | — |
| 94-96 | — | — | — | — |
| 105-112 | 105-110 | 105-110 | 105-110 | 105-109 |
| 128-134 | 130-132 | 130-132 | — | 130-132 |
| 141-147 | 142-146 | 142-144 | 145-146 | 142-146 |
| 178-183 | 179-183 | 179-183 | 179-183 | 178-183 |
| 186-191 | 186-191 | 186-191 | 186-191 | 186-191 |
| 214-217 | — | — | — | — |
| 219-222 | — | — | — | — |
| 227-229 | 227-229 | 227-228 | 227-229 | 227-229 |
| 232-235 | 232-234 | 232-234 | 232-234 | 232-234 |
| 238-242 | 239-240 | 239-241 | 238-241 | 238-240 |
| — | — | — | 252-253 | — |
| — | — | — | 292-293 | — |
| 316-321 | — | 315-316 | 316-317 | —- |
| 326-333 | 326-333 | 326-333 | 326-331 | 326-333 |
| 336-341 | 336-344 | 336-344 | 337-344 | 336-342 |
| 350-358 | 347-354 | 347-355 | 347-354 | 348-355 |
| — | — | 361-362 | — | 361-362 |
| 388-391 | 388-392 | 388-391 | 388-391 | 388-391 |
| 408-412 | — | — | — | — |
| — | 414-417 | 414-416 | 414-416 | 414-416 |
| 421-424 | — | — | — | — |
| 427-430 | — | — | — | — |
| 438-447 | 441-446 | 438-446 | 438-446 | 438-446 |
| 452-459 | 452-458 | 453-459 | 453-459 | 453-459 |
| 462-470 | 464-470 | 464-469 | 464-469 | 464-469 |
| 492-497 | 492-497 | 492-497 | 492-497 | 492-497 |
| — | — | 512-513 | — | 512-513 |
| 530-536 | 530-535 | 530-535 | 530-535 | 530-535 |

*[a] range of residues in secondary structure assigned using DSSP [37]*
*[b] adopted from [11]*
*[c] $3_{10}$ helix*
*[d] residues 267-269 in $3_{10}$ helix*
*[e] residues 281-283 in $3_{10}$ helix*
*[f] no structure*

**Table 2** *Secondary structural assignments for p51 domains [a]*

| 1hmi[b] | 2hmi | 1rtj | 1rtd | model |
|---|---|---|---|---|
| **Helix** | | | | |
| 28-44 | 28-44 | 28-43 | 28-43 | 28-43 |
| 78-84 | 78-83 | 78-83 | 78-83 | 78-83 |
| — | 88-91 | — | 85-89 | — |
| — | 98-100[c] | — | — | 100-102[c] |
| 112-115 | 112-116[c] | 112-117[c] | 111-117[c] | 112-117[c] |
| 122-127 | 125-127[c] | 125-128[c] | 122-128 | 125-128[c] |
| — | 135-137[c] | 135-137[c] | 135-137[c] | — |
| 155-174 | 155-174 | 155-172 | 156-173 | 155-174 |
| 198-212 | 195-211 | 195-212 | 195-211 | 195-212 |
| — | 228-231[c] | — | — | — |
| — | — | — | 236-238[c] | 236-238[c] |
| 254-270 | 254-267 | 254-267 | 254-267 | 254-267 |
| 277-283 | 277-282 | 277-283 | 277-281 | 277-281 |
| 298-310 | 298-309 | 297-309 | 297-309 | 297-310 |
| — | — | 360-362[c] | — | — |
| 364-381 | 364-381 | 364-381 | 364-383 | 364-382 |
| 395-404 | 395-398 | 395-401 | 395-401 | 395-401 |
| — | 403-405[c] | 403-405[c] | 402-404[c] | 402-404[c] |
| — | — | 421-427 | N/A[f] | 421-427 |
| **Sheet** | | | | |
| 7-12 | — | — | — | — |
| 19-22 | — | — | — | — |
| 49-51 | 47-49 | — | 47-49 | 47-50 |
| 56-63 | 60-63 | 60-63 | 60-64 | 60-64 |
| 72-76 | 72-75 | 72-75 | 71-75 | 71-75 |
| 87-90 | — | — | — | — |
| 94-96 | — | — | — | — |
| 105-111 | 105-110 | 106-110 | 105-110 | 105-110 |
| 128-134 | 130-132 | 130-132 | 130-132 | 130-132 |
| 141-147 | 142-146 | 142-144 | 142-146 | 142-146 |
| 179-183 | 179-183 | 178-183 | 179-183 | 179-183 |
| 186-191 | 186-191 | 186-191 | 186-191 | 186-191 |
| 214-219 | — | — | — | — |
| 239-242 | — | — | — | — |
| — | — | 252-253 | — | — |
| — | — | 292-293 | — | — |
| 316-319 | — | — | — | — |
| 325-333 | 326-330 | 326-333 | 326-333 | 326-333 |
| 336-343 | 337-344 | 336-344 | 336-344 | 336-342 |
| 350-358 | 347-354 | 347-355 | 347-352 | 348-355 |
| 386-392 | 388-391 | 388-391 | 388-391 | 388-391 |
| 410-416 | 414-416 | 413-416 | 414-416 | 414-416 |

*See Table 1 for footnotes*

compared, the rms deviations of the α-carbons are much smaller and within the same range (Table 3). Likely, the relatively large rms deviation of the α-carbons of the model compared to those of 2hmi resulted from the relative motions of the individual subdomains during the course of molecular dynamics simulation. Structural comparisons of the model with the starting model (the 0 ps structure) and 2hmi clearly showed that the RNase H subdomain and to a lesser degree the connection subdomain of p66 has drifted from their initial positions during the simulation. Meanwhile, the portion of the DNA at the RNase H subdomain has evolved away from its initial B-conformation, as indicated by the decrease of the helical twist from 34 degrees in the starting model to 28 degrees in the structure at 1,000 ps snapshot. The reduced twist of the DNA during the molecular dynamics simulation is believed to be due to the specified values of certain force

**Table 3** *Comparison of the rms deviations of the α-carbons for individual domains in 2hmi, the model, and 1rtd. Fingers: residues 1-88 and 121-146; palm: 89-120 and 147-242; thumb: 243-311, connection region: 312-425, and RNaseH subdomain: 426-560. Distances are in Å*

*[a] α-Carbons of 89-96 not included in rms calculations*

| | fingers | palm | thumb | connection | RNaseH |
|---|---|---|---|---|---|
| **p66** | | | | | |
| 2hmi – 1rtd | 1.76 | 1.27 | 2.09 | 1.10 | 1.00 |
| model – 2hmi | 1.81 | 1.61 | 2.21 | 1.22 | 1.08 |
| model – 1rtd | 2.14 | 1.67 | 2.32 | 1.02 | 0.97 |
| **p51** | | | | | |
| 2hmi – 1rtd | 1.44 | 1.58 (1.05)[a] | 1.57 | 1.30 | n/a |
| model – 2hmi | 1.28 | 2.08 (1.55)[a] | 1.49 | 1.42 | n/a |
| model – 1rtd | 1.90 | 2.33 (1.30)[a] | 0.97 | 1.61 | n/a |

field parameters (parm94) for nucleic acids. It is reasonable that relative motions of the individual domains of the protein are coupled to the conformation of the DNA with reduced twist during the molecular dynamics simulation.

When switched to the recently revised Amber 98 force field, the rms deviation of the phosphate atoms of DNA compared to those of the starting model dropped significantly from 5.5 Å at 1,000 ps to around 3.3 Å at 2,000 ps (Figure 1) whereas the helical twist of DNA increased from 28 degrees at 1,000 ps to 30 degrees at 2,000 ps. Despite the improvement, the helical twist of the DNA in the model remains somewhat smaller than that in 2hmi (34 degrees) and 1rtd (34 degrees). It is possible that the simulation time is not long enough to fully equilibrate the DNA. It is also possible that the conformation of the DNA in both 1rtd and 2hmi may be partially stabilized by crystal packing interactions not present in the simulation. Nonetheless, the reduced twist of the DNA apparently induced a drift of the RNase H subdomain and the connection subdomain of p66, to a lesser extent. The movement of the RNase H subdomain may also somewhat affect the relative orientations of other subdomains, as suggested by the collective motions study on HIV-1 RT. [34] A ribbon diagram of the overall model structure is shown in Figure 2. Worm diagrams of the fingers, thumb, and palm subdomains for the model, 1rtd, and 2hmi overlaid onto the palm subdomain of the model are shown in Figure 3.

*The primer grip*

The β12-β13 hairpin in the p66 palm is known as the primer grip because it interacts with nucleotides at the 3' end of the primer. [11] It has been suggested that the primer grip residues are important for both polymerase and RNase H activities. [35,36] Powell et al. [36] also demonstrated that the primer grip is involved in interactions with the RNA primers. The primer grip is in close proximity to residues 95-97 (Pro95-Pro97), Tyr181, the YMDD motif (Tyr183, Met184, Asp185, and Asp186), and Trp266 and Gln269 in helix H in all three structures. For instance, Trp229 is buried in a hydrophobic core. It is sandwiched by Tyr188 and Met230 while surrounded by Pro95-Pro97, Leu100, Tyr181-Tyr183, Asp186-Tyr188 and Leu234. The centroid to centroid distances between the aromatic rings of Tyr181, Tyr183, Tyr188, Phe227, Trp229, Tyr232, Trp239 and Trp266 are tabulated in Table 4. Although the distances between Tyr181 and Trp229 and that between Trp229 and Tyr232 in the model are larger than those in both 1rtd and 2hmi, overall, the relative positions of these residues in all three structures are similar. However, this is not the case for 1rtj, the unliganded structure employed to initiate the model building. For instance, in 1rtj, Tyr181 is approximately 10 Å from Trp229 and Tyr232 is approximately 8 Å from Trp266, while Tyr188 is approximately 9 Å away from Trp229. Although the relative positions of these residues are shorter in the starting model (the 0 ps structure) than those in 1rtj (the apo structure) with respect to those in the two DNA-bound crystal structures (1rtd and 2hmi), close agreement with the two experimental structures (1rtd and 2hmi) is present only in the final model.

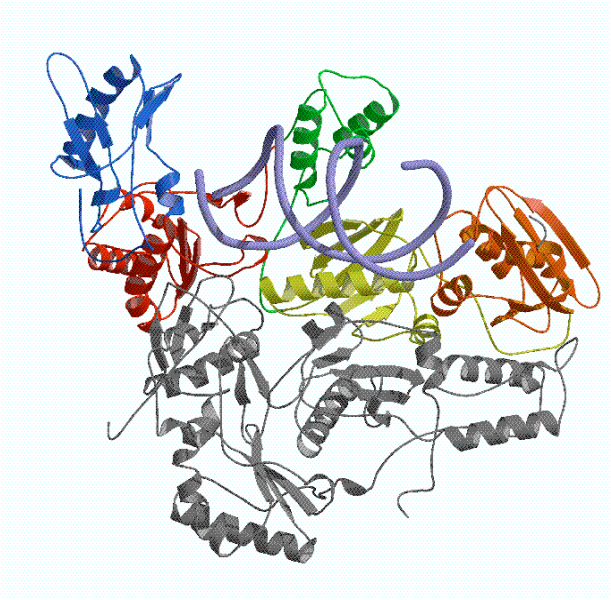


**Figure 2** *Ribbon diagram showing the overall model structure of the HIV-1 RT/DNA complex. The individual subdomains of the p66 subunit are colored as follows: fingers, blue; palm, red; thumb, green; connection, yellow; and RNase H, orange. The p51 subunit is in gray. The DNA duplex is shown as a worm diagram in light purple*

**Table 4** *Centroid to centroid distances between the aromatic side-chains of residues in the region of the primer grid [a]*

| Centroid to centroid /shortest C - C distances (Å) | | | | | |
|---|---|---|---|---|---|
| **Residue 1** | **Residue 2** | **2hmi** | **1rtd** | **model** | **1rtj** |
| Tyr181 | Tyr183 | 10.3/8.2 | 10.3/7.7 | 11.6/9.3 | 10.7/8.5 |
| | Tyr188 | 5.6/4.0 | 5.9/4.3 | 6.0/3.8 | 5.5/3.5 |
| | Trp229 | 6.3/3.9 | 6.0/4.0 | 7.7/5.4 | 12.6/10.3 |
| Tyr183 | Trp229 | 8.4/6.7 | 8.6/6.9 | 8.9/6.9 | 5.8/4.0 |
| Tyr188 | Phe227 | 7.6/6.1 | 7.5/6.0 | 8.4/6.3 | 8.0/6.3 |
| | Trp229 | 5.3/4.2 | 5.4/4.3 | 5.6/4.4 | 10.9/8.8 |
| Trp229 | Tyr232 | 6.8/5.2 | 6.4/4.4 | 10.8/8.2 | 7.6/5.1 |
| Tyr232 | Trp239 | 5.7/3.7 | 5.5/3.8 | 5.0/3.5 | 6.0/4.0 |
| | Trp266 | 6.1/4.1 | 5.8/4.1 | 6.8/4.7 | 9.1/7.5 |

*[a] Only atoms in the six-membered rings of trptophan residues were used in the calculation.*

*The His96-Tyr232-Trp266-Gln269 network*

The sidechain of Gln269 in p66 is in close proximity to His96, Tyr232, and Trp266 in the three structures (2hmi, 1rtd and the model) (Figure 4). The polar atoms of these sidechains form an extensive hydrogen bond network. In fact, in 1rtd, the OH of Tyr232 forms hydrogen bonds with the OE2 of Gln269 and the NE2 of His96. In 2hmi, the sidechain of His96 forms a hydrogen bond with that of Gln269 which in turn forms hydrogen bonds with the sidechains of Tyr232 and Trp266. In the model, the sidechain of His96 forms hydrogen bonds with Tyr232 and Gln269. These interactions may be dynamic in nature, as exchanges of interaction sites among those residues have been observed during the molecular dynamics simulation (data not shown). These interactions could be important in maintaining the structural integrity of the primer grip and its relative position to the primer strand. They may also play an important role in positioning of Trp266 in the minor groove. Interestingly, this hydrogen bond network is not present in 1rtj (apo structure) (Figure 4) but emerged as a result of the modeling. In fact, the sidechain of Tyr232 is approximately 7 and 9 Å away from Trp266 and Gln269, respectively. Likewise, the sidechains of Trp266 and Gln269 are approximately 6 Å away.

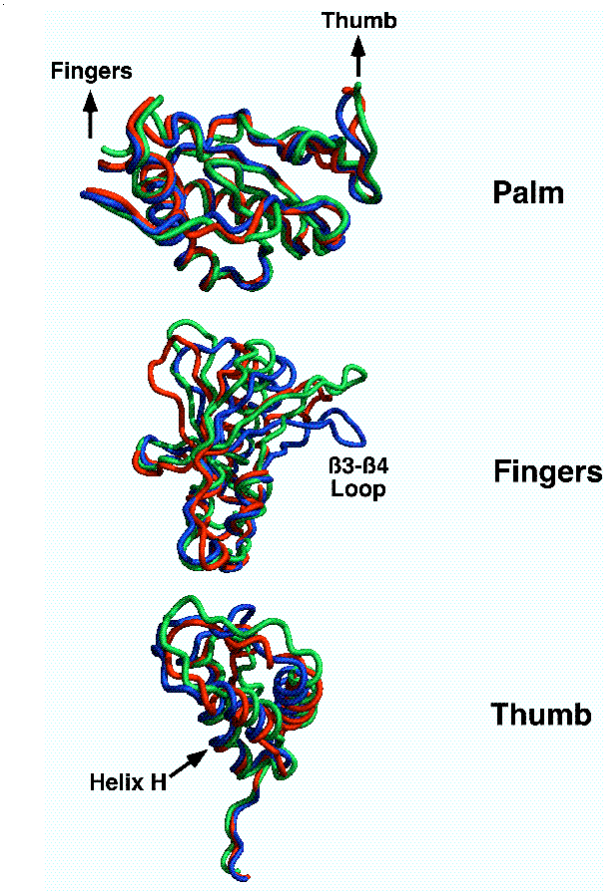



**Figure 3** *Comparison of the fingers, palm, and thumb subdomains for the model (green), 2hmi (red), and 1rtd (blue). In the worm diagrams, the palm subdomains of 2hmi and 1rtd were overlaid onto that of the model*

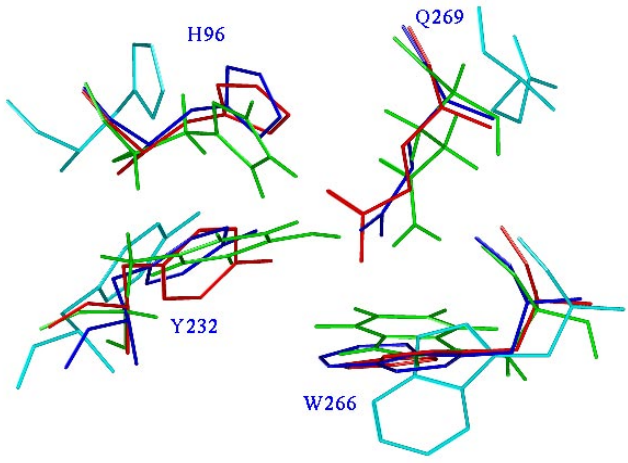



**Figure 4** *Comparison of His96, Tyr232, Trp266, and Gln269 in 1rtj (cyan), the model (green), 2hmi (red), and 1rtd (blue). The Cα positions were used for the rms overlay with 1rtd as the reference*

**Table 5** *Comparison of protein/DNA contacts among 1rtd, 2hmi and the model*

| | Distances[a] | | | Contacts[b] | | |
|---|---|---|---|---|---|---|
| | **1rtd** | **2hmi** | **model** | **1rtd** | **2hmi** | **model** |
| Phe61[e] | 3.59 | 7.92 | 8.97 | 2, 1[d] | — | — |
| Leu74[e] | 3.43 | 4.43 | 3.25 | 5[d] | — | 1 |
| Val75 | 3.44 | 4.29 | 5.25 | 2 | — | — |
| Asp76 | 2.66 | 3.73 | 6.14 | 7 | 1 | — |
| Arg78[e] | 3.08 | 5.14 | 3.58 | 3 | — | 1 |
| Asn81 | 3.41 | 4.04 | 3.03[c] | 1 | — | 6 |
| Glu89 | 3.16 | 3.30 | 3.64 | 8 | 6 | 1 |
| Gln91 | 3.12 | 4.15 | 6.99 | 2 | — | — |
| Leu92 | 3.24 | 4.75 | 3.80 | 1 | — | 1 |
| Ile94 | 3.33 | 3.90 | 4.09 | 2, 5[d] | — | — |
| Tyr115 | 3.80 | 4.15 | 6.63 | 1[d] | — | — |
| Gln151 | 3.96 | 3.68 | 3.72 | — | 2 | 2[d] |
| Gly152 | 3.11 | 3.11 | 3.43 | 6, 2[d] | 17 | 4 |
| Lys154 | 3.40 | 3.74 | 3.56 | 2 | 2 | 6 |
| Pro157 | 3.21 | 3.37 | 5.44 | 1 | 2 | — |
| Tyr183 | 2.67[c] | 3.47 | 3.12[c] | 4, 6[d] | 3, 2[d] | 3, 3[d] |
| Met184 | 3.58 | 3.65 | 3.40 | 2, 2[d] | 1 | 3, 3[d] |
| Asp185 | 3.65 | 3.24 | 2.60 | 1 | 3 | 5 |
| Asp186 | 3.80 | 3.68 | 3.36 | 1 | 1 | 2 |
| Met230 | 3.05 | 2.98[c] | 3.50 | 8 | 17 | 5 |
| Gly231 | 3.22 | 2.79 | 3.48 | 3 | 3 | 1 |
| Asn255 | 4.04 | 4.39 | 2.79[c] | — | — | 6 |
| Gln258[f] | 3.70 | 2.94 | 3.37 | 1 | 9 | 2, 1[d] |
| Lys259 | 3.13 | 2.99 | 2.63[c] | 4 | 4 | 10 |
| Gly262 | 3.28 | 3.69 | 3.49 | 7 | 2 | 3 |
| Lys263 | 3.49 | 3.24 | 2.65 | 2 | 9 | 6 |
| Asn265 | 3.54 | 3.23[c] | 4.81 | 5 | 4 | — |
| Trp266 | 3.25 | 3.29 | 3.63 | 13 | 12 | 4 |
| Val276 | 3.68 | 4.95 | 6.75 | 1 | — | — |
| Cys280[g] | 3.42 | 3.17 | 3.54 | 5 | 4 | 1 |
| Lys281[e] | 4.31 | 4.10 | 3.78 | — | — | 1 |
| Arg284[e] | 3.47 | 3.45 | 3.46 | 5 | 2 | 5 |
| Gly285[e] | 2.84[c] | 3.06 | 2.77[c] | 2 | 6 | 5 |
| Thr286[e] | 3.33 | 3.66 | 5.32 | 3 | 1 | — |
| Lys287[e] | 6.89 | 5.69 | 2.63[c] | — | — | 2 |
| Lys289[e] | 3.64 | 6.36 | 7.93 | 2 | — | — |
| Lys353 | 2.83[c] | 2.46[c] | 5.05 | 3 | 5 | — |
| Ala355 | 3.72 | 4.48 | 4.93 | 1 | — | — |
| Arg358 | 4.83 | 3.63 | 9.75 | — | 1 | — |
| Gly359 | 3.34 | 4.00 | 6.86 | 2 | — | — |
| Ala360 | 3.92 | 3.62 | 6.76 | — | 3 | — |
| His361 | 2.54[c] | 3.30 | 8.95 | 5 | 2 | — |
| Lys374 | 3.08[c] | 3.95 | 5.31 | 2 | — | — |
| Arg448 | 2.66[c] | 4.99 | 3.32 | 5, 6[d] | — | 7 |
| Lys451 | 3.07c] | 7.91 | 2.63[c] | 2 | — | 3 |
| Thr473 | 2.85[c] | 3.09[c] | 2.62[c] | 7 | 3 | 4 |
| Gln475 | 3.14[c] | 2.94[c] | 3.76 | 12, 1[d] | 4 | 1 |
| Lys476 | 3.52 | 3.50 | 3.33[c] | 3 | 1 | 2 |
| Gln500 | 3.11 | 5.74 | 9.54 | 8 | — | — |
| Tyr501 | 2.91[c] | 3.44 | 3.06 | 8 | 2 | 1 |
| Ile505 | 3.90 | 3.24 | 8.76 | — | 1 | — |
| His539 | 3.64 | 4.78 | 12.0 | 1 | — | — |
| Lys22 B | 4.07 | 5.17 | 2.59[c] | — | — | 2 |
| Lys395 B | 3.49 | 3.66 | 10.5 | 3 | 1 | — |
| Glu396 B | 5.94 | 3.30 | 10.7 | — | 2 | — |

*[a] The shortest distance between a protein residue and DNA, hydrogen atoms were not included in the calculations. Distances are in Å*
*[b] The number of contacts between a protein residue and DNA within 3.8 Å. Those that do not meet the cutoff criterion are labeled as '–'. The contact atoms on DNA refer to sugar and phosphate atoms unless stated*
*[c] Potential hydrogen bonds*
*[d] Contacts involve base pairs*
*[e] Disordered in 2hmi [21]*
*[f] Cysteine in 1rtd*
*[g] Serine in 2hmi*

*The protein/DNA contacts*

The residues in close contact with DNA (distance ≤ 3.8 Å) in all three structures (1rtd, 2hmi, and the model) are listed in Table 5. The template overhang nucleotides n+2 and n+3 in 1rtd were removed for consistency with 2hmi and the model. Most of the regions that are in close proximity to DNA in both 1rtd and 2hmi are predicted in the model, especially those in the palm and thumb of p66. However, some of the contacts predicted in starting model (the 0 ps structure) between protein residues and DNA in the RNase H and connection subdomains are not present in the model, possibly as a result of the reduced twist of the DNA in that region during the MD simulations.

Most of the contacts found between protein residues in the palm of p66 and DNA in 1rtd and 2hmi were present in the model, including those from the YMDD (residues 183-186) motif and the primer grip. The YMDD motifs in all three structures are in a very similar conformation, despite the energetically unfavorable backbone conformation of Met184. [17] The rms deviation of the α-carbons of the four residues between 1rtd and 2hmi, between 1rtd and the model, and between 2hmi and the model are 0.16, 0.11, and 0.12 Å, respectively. In all three structures, the sidechains of Tyr183 insert deeply into the minor groove and contact both the sugar and base of nucleotide n-1. The OH group of Tyr183 in both 1rtd and the model forms a hydrogen bond with the N2 of a guanine nucleotide n-1 (OH - N2 distance = 2.84 and 3.1 Å, respectively). In 2hmi, the hydrogen bonding interaction between the OH group of Tyr183 and DNA is weaker. Ding et al. [21] proposed that the hydroxyl group of Tyr183 interacts with both the template and primer bases by forming one or two hydrogen bonds depending on the nature of the base pair. Hydrogen bonding interactions between the OH group of Tyr183 and both the template and primer bases (n-1) have been found during the molecular dynamics simulations (data not shown). In addition to the YMDD motif and the primer grip, Glu89, Gln151, Gly152, and Lys154 contact the sugar-phosphate backbone of nucleotides n+1 to n-2 in all three structures. In 1rtd, the sidechain of Gln151 extends into the active site and forms a hydrogen bond with the O3' of the sugar of the incoming nucleotide, dTTP. Two residues (Ile94 and Pro157) in this subdomain are within 4.0 Å of the DNA only in 1rtd and 2hmi. The sidechain of Ile94 in all three structures inserts into the minor groove. In 1rtd, the sidechain of Ile94 not only contacts the sugar-phosphate backbone of the template strand, but also the base. Early functional analysis has identified Ile94 as one of the five MGBT residues that are important for RT binding, fidelity and processivity. [7] The Pro157 ring contacts the sugar-phosphate backbone of nucleotide n-1 in both 1rtd and 2hmi. For the model, this contact distance is more than 5 Å from the DNA. However, the contacts between Pro157 and nucleotide n-1 are present in the starting model. The distances between the Cγ of Pro157 and the 4'-oxygen atom of the sugar of nucleotide n-1 in 1rtd and 2hmi are 3.21 and 3.37 Å, respectively, while 3.64 Å in the starting model. Apparently, the Pro157 contacts are lost during the course of MD simulation. Interestingly, Gln91 and Tyr115 are within 4.0 Å of DNA only in 1rtd. The backbone oxygen atom of Gln91 is in close contact with the sugar-phosphate backbone of the template strand. In both 2hmi and the model, Gln91 is more than 4.0 Å away from the DNA. Although the sidechain of Gln91 is relatively near the DNA in 2hmi, (shortest distance = 4.15 Å), the backbone oxygen is more than 7 Å away from the DNA. In 1rtd, the backbone nitrogen of Tyr115 forms a hydrogen bond with the 3'-hydroxy of the sugar of the incoming nucleotide while the sidechain of Tyr115 stacks against the sugar ring of the incoming nucleotide. The hydroxyl group of Tyr115 contacts the base of the nucleotide n in the primer strand and the ring of Pro157. Although the relationship between Tyr115 and Pro157 is present in both 2hmi and the model, no near contacts (distance < 4.0 Å) between Tyr115 and DNA are found in the two structures. It is possible that Tyr115 in 1rtd moves nearer to the primer strand in response to the binding of the incoming nucleotide, dTTP.

Contacts between helices H (residues 255-268) and I (residues 278-283) and DNA in 1rtd and 2hmi are present in the model. In all three structures, the sidechains of residues Gln258 (Cys258 in 1rtd), Gly262 and Trp266 insert into the minor groove. In 1rtd, Gln258 was replaced by a cysteine residue that was then cross-linked to a thiol group of an engineered guanine residue in the minor groove. [23] In both 2hmi and the model, the sidechains of Gln258 contact the sugar-phosphate backbone of nucleotide n-5 in the primer strand. Although the contacts are van der Waals in nature, a fluctuating hydrogen bond interaction between the sidechain of Gln258 and N3 of nucleotide n-5 has been detected during the molecular dynamics simulation. Residue 262 (Gly262)

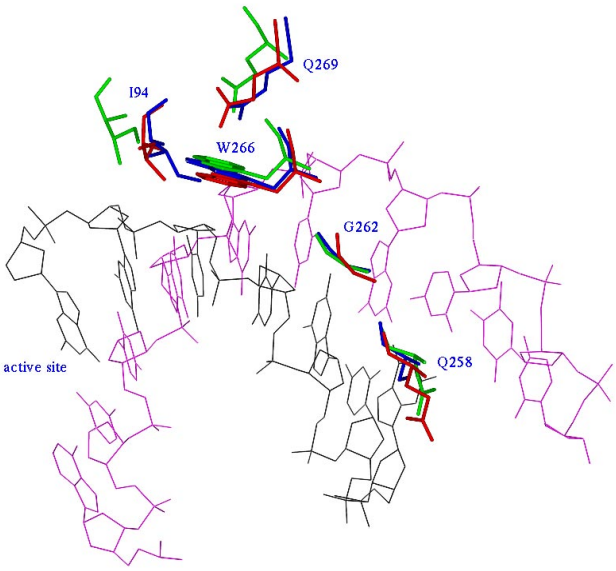



**Figure 5** *The MGBT residues in the model (green), 2hmi (red), and 1rtd (blue). Residue 258 in 1rtd is a cysteine. The phosphate positions of nucleotides n to n-5 were used for rms overlay with 1rtd as the reference. For clarity, only nucleotides n to n-5 of 1rtd are shown*

packs against the backbone of the primer strand in all structures. The sidechains of Trp266 insert into the minor groove in all three structures with extensive contacts with the sugar-phosphate of the primer strand (Figure 5). Although the number of contacts within 3.8 Å between Trp266 and DNA is less in the model than those in 1rtd and 2hmi, the positions and conformations of Trp266 relative to DNA in all structures are very similar (Figure 5) when 2hmi and the model are overlaid onto 1rtd (based on the phosphate positions of the corresponding nucleotides n to n-4). In addition to Ile94, Gln258, Gly262, Trp266, and Gln269 complete the five elements of the MGBT. Of the five MGBT residues, Trp266 is the most critical in terms of RT/DNA interactions, RT fidelity and processivity. [5, 7, 8, 20] Replacement of Trp266 by an alanine residue severely impairs DNA binding. [7, 20]

*The minor groove binding track (MGBT)*

It is not unanticipated that the precise interactions between the minor groove binding track residues (Ile94, Gly262, Gln258, Trp266, and Gln269) of HIV-1 RT and the DNA duplex were not predicted by our earlier modeling attempts. [7] Nonetheless, the nature of the interactions and relative positions of these residues in the minor groove were predicted correctly. The MGBT concept was based on experimental observations in context with the model. [7] It is strongly supported by the fact that alanine mutants of the five MGBT residues all have reduced nucleic acid binding affinity and synthesize DNA less processively and with altered frameshift fidelity. [5, 7, 8] These effects are particularly profound for alanine substitutions at positions 258, 262, and 266. These have template-primer dissociation rate constant elevated by 30-, 230-, and 430-fold, respectively. Additional evidence for the functional importance of the MGBT is the fact that these residues are highly conserved among retroviral polymerases (see Figure 1d in [7]). Taken together, the experimental results indicate that protein interactions with the DNA minor groove are critical for the ability of HIV-1 RT to bind template-primer with high affinity and to conduct processive and accurate polymerization.

## Conclusion

We have developed a protocol to combine the structural information of a low-resolution crystal structure of HIV-1 RT/DNA complex [11] with that of a high resolution crystal structure of an unliganded HIV-1 RT. [12] The process involved slowly forcing the α-carbons of 1rtj onto those of 1hmi using constrained molecular dynamics simulations, while immersing the protein in neutral aqueous solution. In principle, this technique should preserve the sidechain-packing of the protein and the resulting model should not have steric conflicts. A similar technique was also used to build the bent DNA duplex. The resulting complex was refined using molecular dynamics simulations with the PME method [27, 28] employed to accommodate long-range electrostatic interactions.

Overall, the model structure is similar to the two experimental structures (1rtd and 2hmi). We emphasize that these experimental structures are recent and were not available to construct our starting model. The rms deviations of the α-carbons for the individual domains among the three structures are within a similar range. The secondary structures present in the two crystal structures are also predicted in the model. Many of the key interactions among protein residues in 1rtd and 2hmi (not in 1rtj) are also present in the model, such as those around the primer grip. Residues in the palm and thumb of p66 in contact with DNA in the two experimental structures were also predicted in the model with similar detailed interactions. However, some of the contacts between protein residues and DNA in the connection and RNase H subdomains of p66 predicted in the model before molecular dynamics simulations were lost during the simulations due to the low twist of the DNA in the simulation. Implementation of improved force field parameters at the midpoint of the simulation led to some refinement of the model DNA structure towards that of the recent crystal structures.

**Acknowledgements** This work was funded in part by a grant to T.A.D from the NIH intramural AIDS Targeted Antiviral Program.

## References

1. Roberts, J. D.; Bebenek, K.; Kunkel, T. A. *Science* **1988**, *242*, 1171.
2. Preston, B. D.; Poiesz, B. J.; Loeb, L. A. *Science* **1988**, *242*, 1168.
3. Bebenek. K.; Kunkel, T. A. *Reverse Transcriptase*, Goff, S., and Skalka, A. M. Eds., Cold Spring Harbor laboratory, Cold Spring Harbor, NY. **1993**, pp 85-102.
4. Bebenek, K.; Kunkel, T. A. *Methods Enzymol.* **1995**, *262*, 217.
5. Bebenek, K.; Beard, W. A.; Casas-Finet, J. R.; Kim, H. R.; Darden, T. A.; Wilson, S. H.; Kunkel, T. A. *J. Biol. Chem.* **1995**, *270*, 19516.
6. Abbotts, J.; Bebenek, K.; Kunkel, T. A.; Wilson, S. H. *J. Biol. Chem.* **1993**, *268*, 10312.
7. Bebenek, K.; Beard, W. A.; Darden, T. A.; Li, L.; Prasad, R.; Luton, B. A.; Gorenstein, D. G.; Wilson, S. H.; Kunkel, T. A. *Nature Struct. Biol.* **1997**, *4*, 194.
8. Beard, W. A.; Stahl, S. J.; Kim, H. R.; Bebenek, K.; Kumar, A.; Strub, M. P.; Becerra, S. P; Kunkel, T. A.; Wilson, S. H. *J. Biol. Chem.* **1994**, *269*, 28091.
9. Beard, W. A.; Minnick, D. T.; Wade, C. L.; Prasad, R.; Won, R. L.; Knuar, A.; Kunkel, T. A.; Wilson, S. H. *J. Biol. Chem.* **1996**, *271*, 12213.
10. Kunkel, T. A.; Wilson, S. H. *Nature Struct. Biol.* **1998**, *5*, 95.
11. Jacobo-Molina, A.; Ding, J.; Nanni, R. G.; Clark, A. D. Jr; Lu, X.; Tantillo, C.; Williams, R. L.; Kamer, G.; Ferris,

A. L.; Clark, P.; Hizi, A.; Hughes, S. H.; Arnold, E. *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 6320.
12. Esnouf, R.; Ren, J.; Ross, C.; Jones, Y.; Stammers, D.; Stuart, D. *Nature Struct. Biol.* **1995**, *2*, 303.
13. Rodgers, D. W.; Gamblin, S. J.; Harris, B. A.; Ray, S.; Culp, J. S.; Hellmig, B.; Woolf, D. J.; Debouck, C.; Harrison, S. C. *Proc. Natl. Acad. Sci. USA* **1995**, 92, 1222.
14. Hsiou, Y.; Ding, J.; Das, K.; Clark, A. D. Jr; Hughes, S. H.; Arnold, E. *Structure* **1996**, *4*, 853.
15. Smerdon, S. J.; Jager, J.; Wang, J.; Kohlstaedt, L. A.; Chirino, A. J.; Friedman, J. M.; Rice, P. A.; Steitz, T. A. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 3911.
16. Ding, J.; Das, K.; Moereels, H.; Koymans, L.; Andries, K.; Janssen, P. A. J.; Hughes, S. H.; Arnold, E. *Nature Struct. Biol.* **1995**, *2*, 406.
17. Ren, J.; Esnouf, R.; Garman, E.; Somers, D.; Ross, C.; Kirby, I.; Keeling, J.; Darby, G.; Jones, Y.; Stuart, D.; Stammers, D. *Nature Struct. Biol.* **1995**, *2*, 293.
18. Levitt, M. *J. Mol. Biol.* **1992**, *226*, 507.
19. Ponder, J. W.; Richards, F. M. *J. Mol. Biol.* **1987**, *193*, 775.
20. Beard, W. A.; Bebenek, K.; Darden, T. A.; Li, L.; Prasad, R.; Kunkel, T. A.; Wilson, S. H. *J. Biol. Chem.* **1998**, *273*, 30435.
21. Ding, J.; Das, K.; Hsiou, Y.; Sarafianos, S. G.; Clark, A. D. Jr; Jacobo-Molina, A.; Tantillo, C.; Hughes, S. H.; Arnold, E. *J. Mol. Biol.* **1998**, *284*, 1095.
22. Madrid, M.; Jacobo-Molina, A.; Arnold, E. *Proteins* **1999**, *35*, 332.
23. Huang, H.; Chopra, R.; Verdine, G. L.; Harrison, S. C. *Science* **1998**, *282*, 1669.
24. Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E. III; Ferguson, D. M.; Seibel, L.; Singh, U. C.; Weiner, P. K.; Kollman, P. A. *Amber 4.1*, **1995**, University of California, San Francisco, CA.
25. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179.
26. Li, L.; Darden, T. A.; Freedman, S. J.; Furie, B. C.; Furie, B.; Baleja, J. D.; Smith, H.; Hiskey, R. G.; Pedersen, L. G. *Biochemistry* **1997**, *36*, 2132.
27. Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089.
28. Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577.
29. Cheatham, T. E.; Cieplak, P.; Kollman, P. A. *J. Biomol. Struct. Dyn.* **1999**, *16*, 845.
30. Pelletier, H.; Sawaya, M. R.; Wolfle, W.; Wilson, S. H.; Kraut, J. *Biochemistry* **1996**, *35*, 12742.
31. Doublié, S.; Tabor, S.; Long, A. M.; Richardson, C. C.; Ellenberger, T. *Nature* **1998**, *391*, 251.
32. Doublié, S.; Sawaya, M. R.; Ellenberger, T. *Structure* **1999**, *7*, R31.
33. Alley, S. C.; Shier, V. K.; Abel-Santos, E.; Sexton, D. J.; Soumillion, P.; Benkovic, S. J. *Biochemistry* **1999,** *38*, 7696.
34. Bahar, I.; Erman, B.; Jernigan, R. L.; Atilgan, A. R.; Covell, D. G. *J. Mol. Biol.* **1999**, *285*, 1023.
35. Ghosh, M.; Williams, J.; Powell, M. D.; Levin, J. G.; Le Grice, S. F. *Biochemistry* **1997**, *36*, 5758.
36. Powell, M. D.; Levin, J. G. *J. Virol.* **1996**, *70*, 5288.
37. Kabsch, W.; Sander, C. *Biopolymers* **1983**, *22*, 2577.