Crystal Structure at 3.5 Å Resolution of HIV-1 Reverse Transcriptase Complexed with an Inhibitor

L. A. Kohlstaedt, J. Wang, J. M. Friedman, P. A. Rice, T. A. Steitz

A 3.5 angstrom resolution electron density map of the HIV-1 reverse transcriptase heterodimer complexed with nevirapine, a drug with potential for treatment of AIDS, reveals an asymmetric dimer. The polymerase (pol) domain of the 66-kilodalton subunit has a large cleft analogous to that of the Klenow fragment of *Escherichia coli* DNA polymerase I. However, the 51-kilodalton subunit of identical sequence has no such cleft because the four subdomains of the pol domain occupy completely different relative positions. Two of the four pol subdomains appear to be structurally related to subdomains of the Klenow fragment, including one containing the catalytic site. The subdomain that appears likely to bind the template strand at the pol active site has a different structure in the two polymerases. Duplex A-form RNA-DNA hybrid can be model-built into the cleft that runs between the ribonuclease H and pol active sites. Nevirapine is almost completely buried in a pocket near but not overlapping with the pol active site. Residues whose mutation results in drug resistance have been approximately located.

HIV reverse transcriptase (RT) is the target of 3′-azido-deoxythymidine (AZT) and dideoxyinosine (ddI), which are anti-AIDS drugs that function by terminating the DNA during its synthesis (1). Also, a class of non-nucleotide inhibitors of RT shows potential as less toxic drugs (2, 3). However, the effectiveness of all current compounds in the treatment of AIDS is limited by the toxicity of nucleotide analogues and by resistance mutations in the RT that render it insensitive to both classes of these inhibitors (4, 5). Because the structure of HIV RT is anticipated to facilitate the design of new inhibitors that might prove to be effective drugs to control AIDS, many investigators have attempted with varying degrees of success to grow crystals of RT suitable for high resolution structural analysis (6, 7). We now report a 3.5 Å resolution crystal structure of RT complexed with a non-nucleotide inhibitor derived from cocrystals that diffract to 3.1 Å resolution (8).

RT is a DNA polymerase that can employ either RNA or DNA as a template, yielding either RNA-DNA hybrid or duplex DNA products (7, 9). Although a few very weak amino acid sequence similarities between RT and other polymerases have been noted (10–17), the extent to which RT resembles the pol domain of the *Escherichia coli* Klenow fragment, the only polymerase whose structure is established, has remained unclear. Also unknown is why this class of DNA polymerases has a high error rate compared to other DNA polymerases (18–20). The lack of an editing 3′,5′-exonuclease activity associated with RT accounts for part of the decreased fidelity, but not all of it.

The HIV RT is processed initially from the pol gene product as a 66-kD polypeptide that has both a pol and an RNase H (ribonuclease H) domain. Subsequent proteolytic cleavage of a homodimer of the 66-kD subunits removes the RNase H domain from one subunit leaving a heterodimer containing one 66-kD subunit (p66) and one 51-kD subunit (p51) (21). The p66-p51 heterodimer appears to have only one pol active site, one RNase H active site, one tRNA binding site (22–24), and one Nevirapine binding site (25). The HIV RT catalyzed DNA synthesis is initiated in vivo by human $tRNA_3^{Lys}$ whose 3′ end partially unfolds and forms 18 base pairs of duplex with the viral RNA primer binding site (26). While reverse transcription from the viral genome is reported to require $tRNA_3^{Lys}$ and the gag protein (23), either *E. coli* $tRNA_2^{Gln}$ or human $tRNA_3^{Lys}$ can initiate reverse transcription by HIV RT alone provided the 3′ end of the template terminates with an 18-nucleotide (nt) primer binding site whose sequence is complementary to the 3′ end of the appropriate tRNA (27). In the heterodimer, the pol domain of p66 catalyzes the pol reaction while the same sequence in p51 does not (22, 24), making the role of p51 in the reaction unclear. We postulate that p51 forms part of the binding site for the tRNA primer as well as a portion of the template-primer binding site.

We describe here the three-dimensional fold of the p66-p51 heterodimer of HIV RT derived from electron density maps calculated at 3.5 Å resolution. The dimer is asymmetric. The two subunits are not related by a simple rotation axis; furthermore, the two pol domains have very different conformations in spite of having the same sequence. A cleft sufficient in size to accommodate A-form RNA-DNA hybrid runs between the RNase H and the pol active sites. Nevirapine (2), an inhibitor of HIV-1 RT, lies in a deep pocket adjacent to the anticipated pol active site but not overlapping it. Portions of the pol domain of RT, including the catalytic site, have similar structures to those in Klenow fragment (KF), but the rest is completely different.

Structure determination. HIV RT was purified (27) from an overexpression clone prepared by D'Aquila and Summers (28). Crystals were grown by vapor diffusion of a solution (6 to 8 mg/ml) of RT containing 25 mM bis tris propane, pH 7.0, 50 mM $(NH_4)_2SO_4$, 0.1 percent (w/v) β-octylglucoside, 5 percent (v/v) glycerol, 7 percent (w/v) polyethylene glycol 8000, 0.01 percent sodium azide against a solution that was double in the concentration of all compounds except RT. Nevirapine was present in twofold molar excess over RT. Crystals belong to the monoclinic space group C2 with cell dimensions a = 223.5 Å, b = 70.2 Å, c = 106.5 Å, and β = 105.3 degrees. One asymmetric unit contains a single p51-p66 heterodimer and a solvent content of 64 percent by volume. X-ray diffraction intensities were measured and processed with the use of Xuong-Hamlin multiwire area detector systems (29). Data were scaled and merged with the program SCALEPACK (30). Heavy atom derivatives were prepared by soaking crystals in solutions that contained the heavy atoms (Table 1). Additional high resolution data from both native and derivative crystals were measured on an *R*-axis imaging plate system (Molecular-Structure Corporation) and processed with the program DENZO (30). Although several heavy atom derivatives were obtained, many had binding sites in common but with varying occupancies (Table 1) (31–33).

An electron density map calculated with multiple isomorphous replacement (MIR) phases revealed the molecular boundary, domain structure, and much of the secondary structure. This map was improved with a solvent-flattening procedure (34). The coordinates of the RNase H domain (35, 36) were easily fitted to this map as a rigid

The authors are with the Departments of Molecular Biophysics and Biochemistry and Chemistry and the Howard Hughes Medical Institute, Yale University, New Haven, CT 06511.

body (Fig. 1A). The phases are very well determined at lower resolution (out to 4.5 Å) but less well in the 4.5 to 3.5 Å resolution range, resulting in incomplete side chain information.

Three different approaches to improving the resolution and accuracy of this experimental map were taken at various stages with varying success—map-editing, subdomain averaging, and application of the tangent formula. The electron density of the MIR phased map was edited in several ways in addition to solvent flattening; the edited map was subjected to Fourier transformation and a new map calculated with $2F_o$-F_c and weighted calculated phases as coefficients. In the most generally useful edited map the electron density corresponding to the RNase H domain was replaced by electron density calculated from the coordinates of RNase H. Later, electron density calculated from a polyalanine model of p51 was used to replace electron density of the MIR map, leaving the MIR density for side chains and the p66 pol domain. The new map resulting from this editing procedure showed improved electron density for the p66 domain and the side chains of p51. Because the p51 subunit and the pol domain of p66 do not have the same structure, it was not possible to improve the map by simply averaging the density of the two subunits. Rather, it was necessary to average four subdomains of the pol domain separately (*37*).

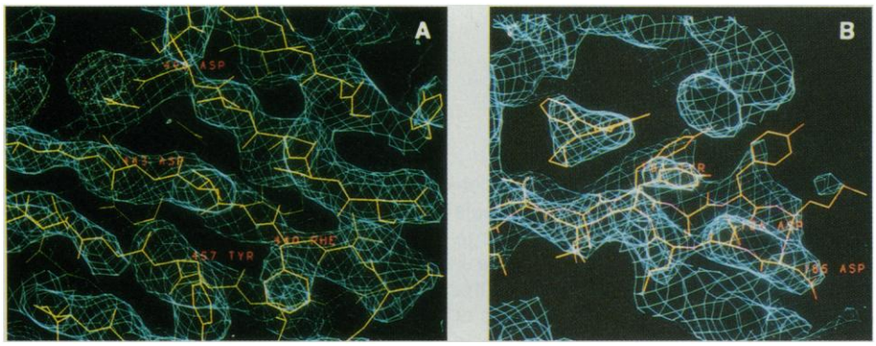

Fig. 1. Views of some of the electron density maps used to determine the structure of HIV RT. (**A**) A portion of the solvent flattened, MIR-phased electron density map at 3.5 Å resolution with a model of RNase H (*35, 36*) superimposed. The structure of RNase H (*36*) fits the electron density well everywhere, including the two divalent metal ions shown as two spheres. (**B**) An electron density map at 3.2 Å resolution that has been improved by cycles of density modification including insertion of the RNase H calculated density and solvent flattening followed by subdomain averaging alternated with phase calculation and map calculation with $2F_o$-F_c as coefficients. Density corresponding to the bound inhibitor Nevirapine and neighboring protein including Tyr^{181}, Tyr^{188}, and Asp^{185} is shown.

Table 1. Heavy atom statistics. Each column contains the following information: Derivative, the names of derivatives that were separately refined; Resolution, limits (in angstroms) of each set used; I/σ, ratio of intensity to sigma for the highest resolution bins; R_m, fractional difference in structure-factor amplitude between a given derivative data set and its corresponding native data set; reflections, the unique reflections to the resolution limits and their completeness in percentage; F_h/E, phasing power of each derivative (heavy atom structure factor, F_h/E; root-mean-square lack of closure error, E) listed in resolution shells; "FOM(%)", figure of merit. Derivatives AUCL, BAKE, DMA, HGTL, PIP, and PTCL are gold potassium chloride, Baker's dimercurial, dimercuriacetate, mercury thallium nitrate, di-μ-iodobis (ethylenediammine)-di-platinum (II), potassium platinum (II) tetrachloride, respectively. Data from different crystals of the same heavy atom derivatives were kept separated if they failed to merge, denoted as No. 1, No. 2, No. 3, and No. 4. High resolution data for DMA and HGTL were collected on an *R*-axis imaging plate system from two crystals each, as labeled "*R*-axis." Heavy atom derivatives were prepared by soaking crystals in heavy atom–containing stabilization solution for a period of 2 days to 1 week (over 4 weeks for HGTL No. 1 and HGTL No. 2). Concentrations of heavy atom compound were all less than or equal to 1 mM.

Derivative	Resolution (Å)	I/σ	R_m (%)	R_d (%)	Reflections		F_h/E at each resolution bin (Å)								
					N	%	Avg.	7.9	6.5	5.6	4.9	4.3	3.9	3.5	3.2
Native	3.2	1.8	10.4		24401	99									
HGTL															
No. 1	4.3	3.8	6.2	33	6540	60	2.07	2.4	2.5	2.0	1.9	1.8			
No. 2	4.3	2.3	9.6	37	7929	72	1.21	1.1	1.4	1.3	1.1				
No. 3	3.5	1.2	8.9	36	16172	80	1.64	2.2	2.4	2.4	2.0	1.4	1.2	0.9	0.9
No. 4	3.5	1.5	11.2	35	10557	52	1.36	1.8	1.6	1.7	1.5	1.2	1.0	1.0	1.0
R-axis	3.4	1.3	12.4	37	18493	99	1.81	2.7	3.3	2.9	2.3	1.7	1.4	1.0	0.8
DMA															
No. 1	4.5	3.7	16.5	32	6145	62	1.76	1.9	1.9	1.7	1.6	1.3			
No. 2	4.3	1.8	10.0	27	6015	44	1.34	1.2	1.5	1.5	1.4	1.1			
No. 3	3.5	1.5	14.4	39	8756	44	1.41	1.9	1.8	1.9	1.6	1.2	1.0	1.0	0.9
No. 4	4.0	3.1	11.4	35	8579	63	1.47	1.8	1.9	2.0	1.8	1.3	1.0	1.0	0.9
R-axis	3.5	1.2	11.1	37	17323	99	1.57	2.2	2.5	2.4	1.7	1.3	1.2	1.1	0.9
AUCL	5.5	2.1	16.6	30	4069	76	0.88	1.9	1.3	1.2	0.6	0.5			
PTCL															
No. 1	4.0	2.0	16.9	22	6855	50	0.63	1.0	0.9	0.9	0.5	0.4			
No. 2	4.0	2.1	9.5	16	7532	55	0.85	1.6	1.0	0.8	0.4				
BAKE	5.5	1.8	6.6	12	2860	54	0.96	1.0	1.0	0.9	0.6				
PIP	4.4	1.5	6.8	25	7953	61	0.81	1.3	0.9	0.6	0.3				
FOM(%)							56	88	86	81	76	65	53	37	24

The continuity of the polypeptide backbone electron density in the solvent-flattened, RNase-edited map allowed the polypeptide backbones to be fitted to p51 and p66 independently by separate authors. The two separate interpretations resulted in the same connectivity of the polypeptide backbone, which was further supported by the subdomain-averaged electron density map. In some regions side chains have been added with a high degree of confidence because of good electron density and additional constraints. The sequence can be fitted to the electron density at the NH_2- and COOH-termini of the pol domain, to Cys^{38} and Cys^{280} which react with mercury ions, and to the region of the conserved -YMDD- (Tyr-Met-Asp-Asp) sequence which has been cross-linked to Nevirapine (*38*). In two regions, the exact position of the backbone is at present uncertain: helix G of the "thumb" of p66 and the peptide extending from the β strand 11 of the "palm" of p51 through helix G of the "thumb" (Fig. 2, A and B).

Description of structure. The p66 subunit is folded into five separate subdomains, the RNase H domain and four subdomains of the pol domain (Figs. 2 and 3). Because these subdomains are largely arranged side by side, the subunit is highly elongated and curved, having dimensions of approximately 110 by 30 by 45 Å. A prominent feature of the pol domain is a large cleft, reminiscent of the cleft in the pol domain of the KF of *E. coli* DNA pol I (*39*). Its anatomical resemblance to a right hand has led to naming the subdomains as fingers, palm, and thumb. The fourth pol subdomain lies between the rest of the pol and the RNase H domain, leading it to be called the "connection" subdomain (Fig. 2A).

The palm and connection subdomains consist of five-stranded β sheets with two α helices on one side. The "thumb" is most

likely a four-helix bundle. A large family of seemingly unrelated proteins contains four-helix bundles including the *rop* protein dimer which binds RNA (*40*). The "fingers" domain contains a mixed β sheet and three α helices. The five β strands of the "palm" domain hydrogen bond to the four β strands at the base of the "thumb" which in turn hydrogen bond to the "connection" β sheet. This 14-stranded intersubdomain β sheet is partly unraveled in the p51 subunit.

The conformation of the pol domain of p66 is astonishingly different from that of p51 considering that the two proteins have the same amino acid sequence. The differences can be most easily appreciated by orienting the "palm" subdomains of the two subunits identically and noting the differences in the relative positions of the other subdomains (Fig. 2, A and B, and Table 2). Unlike p66, p51 has no cleft, and residues thought to be involved in catalysis (Asp^{185}, Asp^{186}, Asp^{110}) are essentially buried. The fingers move toward the palm while the thumb domain is further from the fingers and palm domains. The connection domain lies within and fills the expanded cleft between them. The tertiary structures of the four corresponding subdomains are largely the same in the two subunits, although some significant differences exist; these include the positions of β strand 20 and α helix L of the connection subdomain, β strands 12 to 14, which are not seen in p51 and perhaps some aspects of the thumb.

The most dramatic differences in relative position and interdomain packing contacts are seen for the "connection" subdomain (Fig. 3). In p66 it contacts RNase H, the bottom of the thumb and the connection of the p51 subunit, whereas in p51 it contacts all three of the other p51 subdomains as well as the

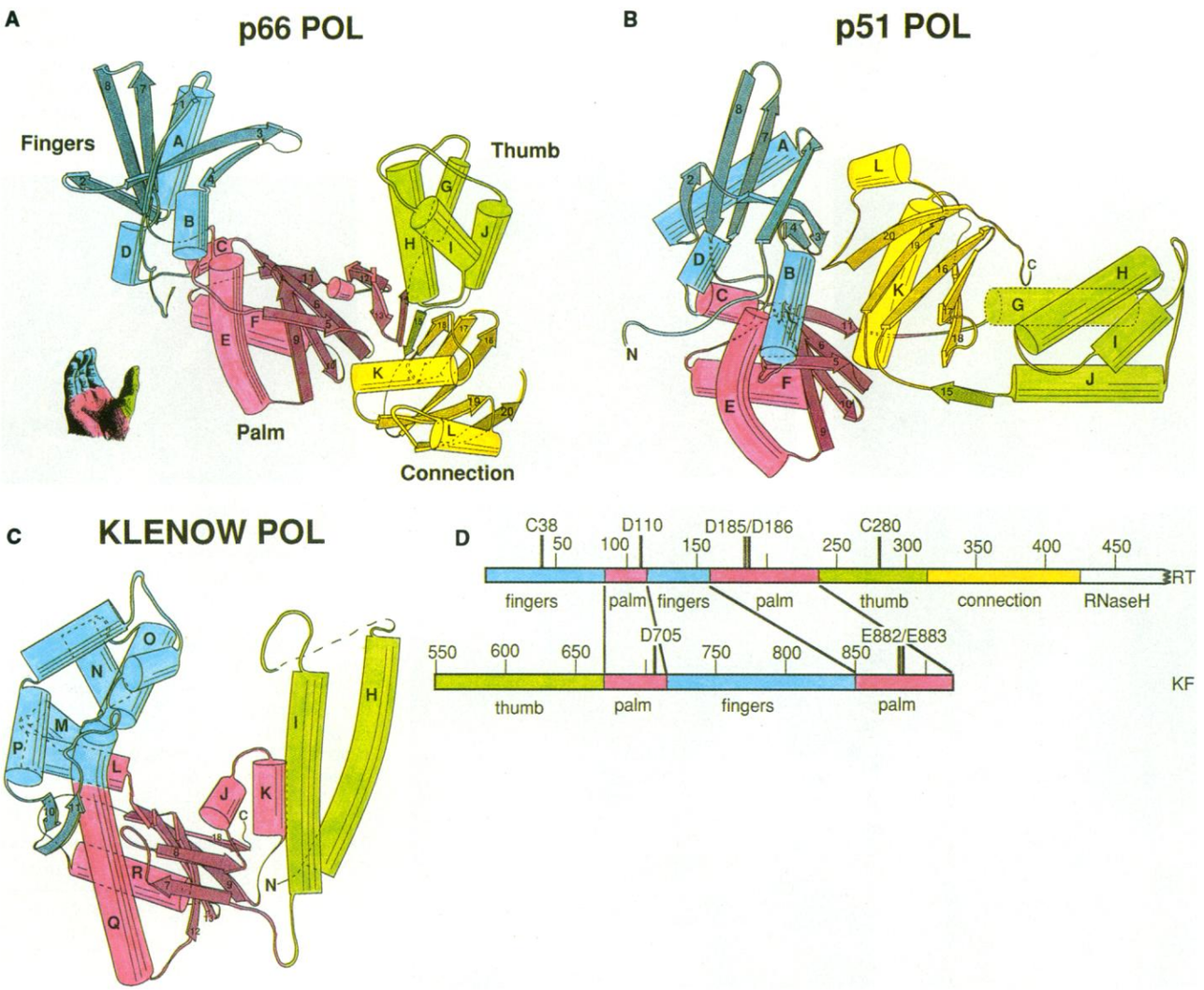

Fig. 2. The polymerase domains of HIV-1 RT and Klenow fragment in which helical regions are represented as tubes lettered and β strands as arrows numbered from the amino terminus based on RIBBON drawings (*57*). (**A**) The polymerase domain of p66. The finger subdomain is blue, the palm is pink, the thumb is green, and the connection is yellow. The anatomical analogy to a right hand [modified from (*58*)] is indicated. (**B**) The p51 subunit with its pink palm subdomain oriented identically to the corresponding subdomain in p66. The different relative orientations of the four subdomains in the two subunits is clear. Electron density corresponding to helix G is weak and ambiguous so that its presence is indicated by a dashed outline. (**C**) Pol domain of KF with its palm subdomain oriented identically to that of p66. The lettering and numbering is as given in (*39*). The blue finger subdomain is almost all α helix in contrast to the finger subdomain of RT. (**D**) The approximate locations of the subdomains in the linear sequence of reverse transcriptase (RT) and Klenow fragment (KF). The positions of the two Cys residues and the catalytic carboxyl groups in RT as well as the corresponding carboxyls in KF are shown. In both cases the palm is interrupted by an insertion of the finger subdomain, but the thumb occurs on opposite sides of the palm sequence.

"connection" of p66. This difference in position represents a rotation of one connection subdomain by about 155° relative to the other along with a 17 Å translation (Table 2). An equally amazing feat of versatility is the fact that in the two subunits different surfaces of this subdomain face the single long groove that is presumed (see below) to be the site of primer and template binding. In p66 the α helices of the connection subdomain face this groove while in p51 the β sheet faces it.

The pol domain of p66 is not related to p51 by a twofold rotation axis; rather, the two domains interact in a more head-to-tail arrangement that results from a 16 Å translation and an approximately 74° rotation of one palm relative to the other (Fig. 3). While nonsymmetric dimers, though rare, are not unprecedented (*41*), the degree of asymmetry seen here is. The tip of the fingers and the connection subdomains of p51 interact with the palm and connection subdomains of p66 while the very extended thumb of p51 contacts the RNase H domain of p66.

It is of interest to consider why the RNase H domain is cleaved from only one of the two subunits and why the protein does not simply polymerize indefinitely through the head-to-tail interaction. The interactions between the subunits of this dimer are such that the formation of a p66 homodimer by making use of the interactions seen here requires that one p66 subunit would assume the conformation of p51 with its connection domain in the cleft. From the location of the COOH-terminus of p51, the RNase H domain prior to cleavage would come out the "back" as viewed in Figs. 2B and 3, A and B. To achieve this, the polypeptide containing the protease cleavage site needs to be unraveled, presumably rendering it sensitive to cleavage as proposed earlier (*36*, *42*). The other subunit of a p66 pair would be stabilized as we see p66 here with its site of protease cleavage buried within a β sheet. Once p66 has interacted with p51, it can no longer assume the conformation necessary to interact with another p66. The heterodimer does not polymerize further because p66 does not and cannot have the necessary second dimerization surface.

Why has the pol domain of HIV RT evolved to adopt two such very different conformations? It would appear from the structure that each pol domain performs a different function (see below). Most nonviral enzymes would use two subunits of differing sequence or a large subunit with two domains of differing sequences to achieve

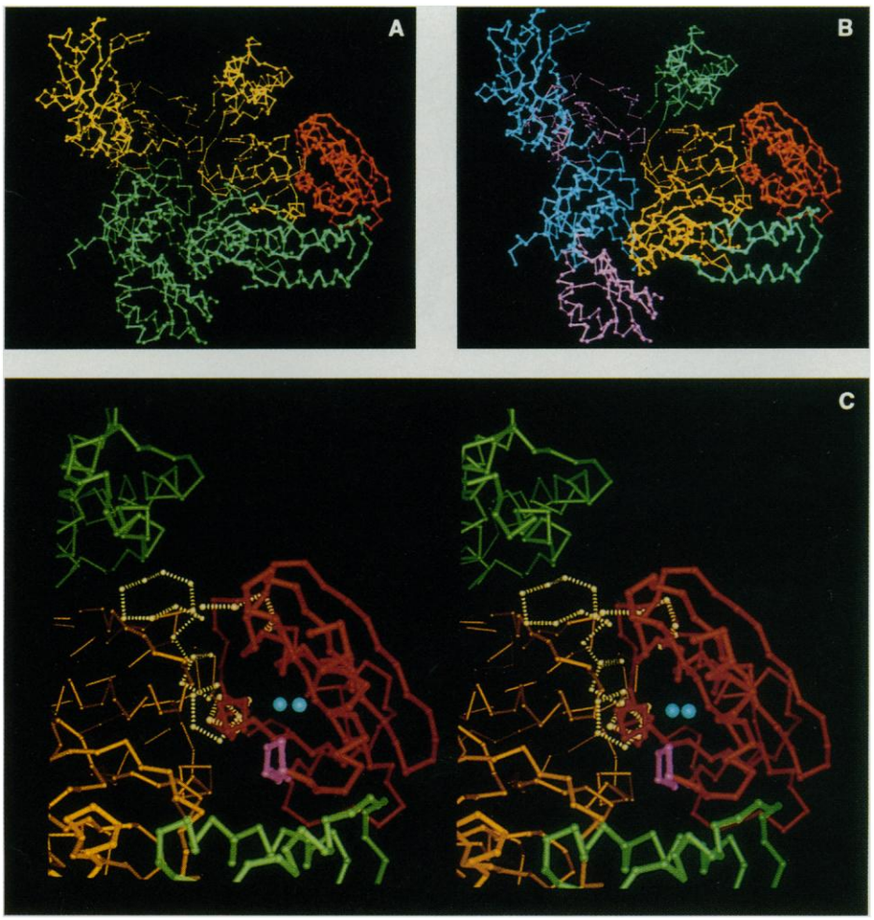

Fig. 3. α-Carbon backbone (*59*) of HIV RT dimer. (**A**) p51 is green and the pol domain of p66 is yellow while the RNase H domain is orange. (**B**) The pol subdomains are colored as in Fig. 2 ("finger" is blue; palm is purple; thumb is green; connection is yellow, ribonuclease H is orange). A major intersubunit interaction occurs between the two "connection" subdomains, as well as between the tip of the p51 fingers and the base of the p66 palm and between the tip of the p51 thumb and the RNase H domain. (**C**) A stereo α-carbon drawing of RNase and the surrounding region, oriented and colored as in (B). The two divalent metal ions are shown in blue and the peptide that becomes ordered in the whole protein is pink. The extra residues that exist in *E. coli* RNase H are shown in dashed white.

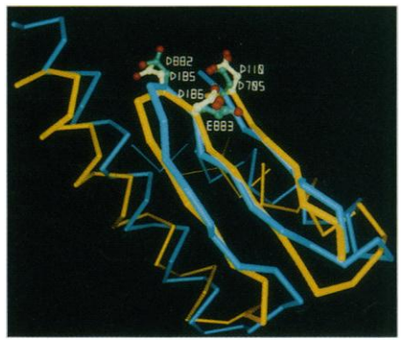

Fig. 4. Superposition of the α-carbon backbone of three β strands, 12 to 14, and two α helices, Q and F, of the KF (blue) on the corresponding structure in the palm subdomain of p66 (yellow). The palm subdomains were aligned by refining the model of the KF palm into the RT palm electron density in space group P1. The side chains of the catalytically important Asp^{882}, Glu^{883}, and Asp^{710} of KF (green) superimpose on Asp^{185}, Asp^{186}, Asp^{110} of RT (white). In both enzymes the three residues form a tripod of carboxyl groups capable of binding divalent metal ions (*50*).

Table 2. Positional relation between corresponding subdomains.

Subdomain	In the heterodimer		With palm subdomains superimposed	
	Rotation (deg)	Translation* (Å)	Rotation (deg)	Translation* (Å)
Fingers	85.0	8.5	51.4	0.1
Palm	74.2	16.3	0.0	0.0
Thumb	133.2	36.2	93.4	7.9
Connection	155.1	16.9	132.7	3.5

*The translation is in the direction of the rotation axis. The directions of the four rotation axes relating the subdomains are all different.

this goal of functional diversity. Virus genome length is limited to that which can be packaged within the viral particle. The evolutionary selective pressure to increase the information density of viral genomes has resulted in two strategies: overlapping out of frame genes and polypeptide sequences that can adopt more than one structure/function. Examples of the latter include the coat protein of spherical viruses, such as tomato bushy stunt. A single protein is able to form an icosohedral T=3 lattice by adopting two conformations that can pack differently (*43*).

The crystal structure of an isolated HIV RNase H domain has been determined at 2.8 Å resolution (*36*), however, the isolated domain is devoid of RNase H activity. The coordinates of HIV-1 RNase H including the two divalent metal ions (*35, 36*) fit into our electron density map well (Fig. 1A), except for a short polypeptide loop that contains the invariant His^{539} and a few other conserved residues. This region of RNase H is disordered in the crystal of the domain alone but it is ordered in the RT heterodimer and interacts with a portion of the thumb of p51. The His^{539}-containing loop lies close to the two divalent metal ions at the enzyme's catalytic center and could be involved in the RNase H activity (Fig. 3C). If so, its disorder in the RNase H fragment could account, at least in part, for the lack of enzymatic activity of the isolated domain. Additionally, it is clear that most of primer-template binding site is provided by the pol domain of p66 as well as by the thumb of p51 (see below), so that the RNA-DNA hybrid substrate may not be properly bound and oriented on the isolated RNase H domain. This interpretation is consistent with the location (Fig. 3C) of additional residues that exist in *E. coli* RNase H (*44*). These possibilities are consistent with the observation that addition of p51 to an RNase H fragment activates RNase H activity (*45*).

Comparison with other polymerases. The only other DNA polymerase whose three-dimensional structure is known is that of the KF of *E. coli* DNA pol I (*39*). The KF has two domains, a 200–amino acid domain that catalyzes the editing 3′,5′-exonuclease activity and a 400-residue domain having a large cleft that contains the active site for the DNA pol activity (*39, 46*). While early sequence comparisons hinted that DNA polymerases and even RNA polymerases may be related to each other (*47*), only with the large number of polymerase sequences that have become available in recent years has it been possible to find convincing, although extremely weak, sequence similarities among all classes of polymerases (*12–17*). The sequence similarity between HIV RT and KF is almost nonexistent; however, two short stretches of 15 to 20 amino acids have been aligned with 4 of 14 identities in motif C and 3 of 24 identities in motif A (*12*), not really sufficient to prove a relation in the absence of additional data.

Comparison of the backbone structure of the p66 pol domain with that of KF shows an obvious similarity between the palm subdomains (Fig. 2). The five β strands 5, 9, 10, 6, and 11 of the "palm" subdomain of RT superimpose on β strands 8, 12, 13, 9, and 14 of KF, and its α helices E and F superimpose on helices Q and R of KF. An antiparallel β-hairpin structure (motif C) that contains the catalytically important Asp^{882} and Glu^{883} of KF superimposes on a nearly identical hairpin in RT that contains the sequence Tyr-Met-Asp-Asp that is highly conserved with Asp^{185} and Asp^{186} occupying the same positions as the two acid residues in KF. In addition, Asp^{705} (motif A) in KF superimposes on Asp^{110} in RT (Fig. 4). Since mutation of Asp^{185}, Asp^{186}, or Asp^{110} in RT (*11, 48*) or mutation of Asp^{882}, Glu^{883} or Asp^{705} in KF (*49*) severely reduces pol activity, it is plausible to expect that these three acid residues play the same essential role in the catalysis of

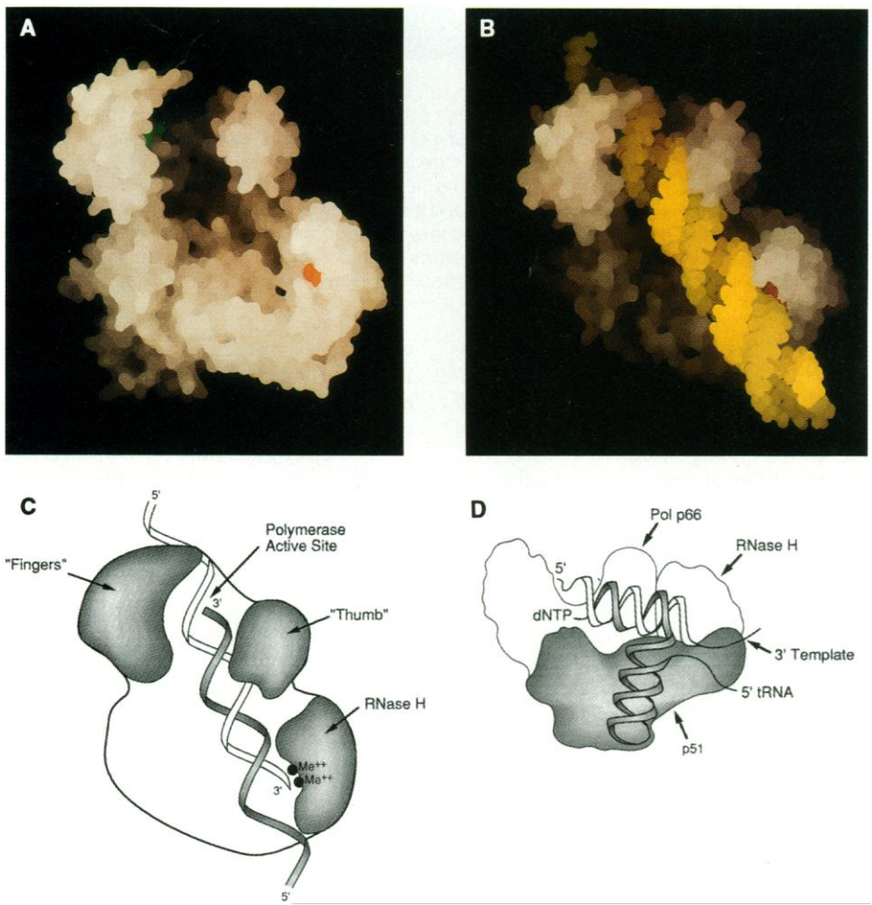

Fig. 5. (**A**) Solvent accessible surface model (*60*) of HIV RT heterodimer showing the deep cleft that runs between the RNase H active site in the lower right (marked by two divalent metal ions in red) and the pol active site at the upper left (marked by green Asp^{185}). (**B**) Solvent accessible surface model of HIV RT heterodimer with model built A-form RNA · DNA hybrid placed in the cleft such that the template strand in yellow contacts the two divalent metal ions (red) observed in the RNase H active site and the 3′ terminus of primer strand in lime is near to the Asp^{185} (green) in the pol active site. (**C**) A schematic drawing of RT with model built RNA · DNA hybrid as shown in (B). The possibility that the RNase H domain cleaves the RNA template strand progressively and in concert with DNA synthesis on the primer strand is shown. About 20 bp of duplex RNA · DNA hybrid having a rise per residue of 3.0 Å fit between the two active sites, somewhat longer than deduced from biochemical studies (*53*). If B-form DNA is bound, the distance between the pol and RNase H sites is 18 bp. (**D**) A schematic drawing showing a hypothetical complex between a tRNA primed template and RT. The possibility that the anticodon and dihydrouridine stems and loops bind to p51 is indicated schematically. Consistent with this model, tRNA, including its anticodon stem and loop, can be cross-linked to p51 and p66, although it is unknown which subunit interacts with the anticodon loop (*23*).

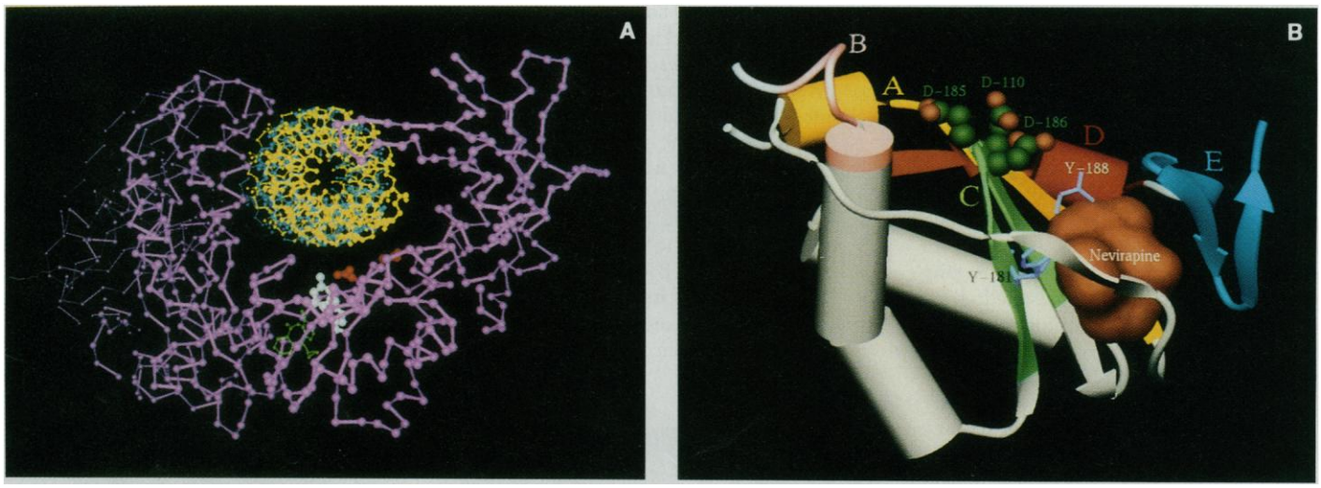

Fig. 6. The position of the inhibitor, Nevirapine (*2*), bound to the p66 pol active site near to the catalytic site and the expected DNA primer terminus. (**A**) α carbon backbone (purple) of p66 pol with model built DNA (blue and yellow) showing Asp^{185} and Asp^{186} (red) nearby. Nevirapine (green) binds near Tyr^{181} and Tyr^{188} (white). The view is from the "back" of Fig. 2A so that the "fingers" are on the right and the "thumb" on the left. (**B**) A schematic drawing (*61*) of the palm subdomain of p66 showing the five regions (A to E) that exhibit sequence similarity among the RNA-dependent DNA polymerases (*12*) with Nevirapine (yellow) superimposed. RNA-dependent RNA polymerases (*12*) and telomerase (*17*) have similarities in only regions A to D. The DNA-dependent polymerases show sequence simliarities to regions C (containing Asp^{185}, Asp^{186} in HIV RT) and A. Region B in Delarue *et al.* (*12*) is not the same as B in Poch *et al.* (*14*). Nevirapine is snuggled into a pocket between the conserved regions A and C and the region E conserved among reverse transcriptases. The side chains of Tyr^{188} and Tyr^{181} are packed against the inhibitor. The view is the same as Fig. 2A.

the pol reaction. Furthermore, since Asp^{882}, Glu^{883}, and Asp^{705} have been observed to bind divalent metal ions to KF (*50*), a similar role might be postulated for the corresponding acidic residues in RT. The possibility of a two-metal ion-catalyzed phosphoryl transfer mechanism similar to that of the 3′,5′-exonuclease mechanism of KF has been hypothesized (*50, 51*).

The second subdomain of HIV RT that appears to bear some resemblance to KF is the "thumb". While the thumb of RT may be related to a four-helix bundle, in KF two α helices dominate. Recent interpretation of the less well ordered polypeptide between these helices indicates that two shorter helices exist (*50*). The crystal structure of a KF complex with duplex DNA (*50*) and model building duplex DNA onto RT (see below) suggest that the "thumbs" may share some similar roles in binding the duplex product resulting from DNA synthesis.

The "finger" subdomains of KF and HIV RT are completely unrelated, the former being predominantly α helical and the latter being mostly β sheet (Fig. 2). What is the role of the finger subdomains, and why are they so different in these two polymerases? The α helix O in the fingers of the KF has figured prominently in the cross-linking of deoxynucleotide (dNTP) analogues to this enzyme, which leads to the hypothesis that it forms part of the dNTP binding site (*52*). However, the binary complex between KF and the dNTP's may well not be the same as formed in the presence of the primer-template which, of course, forms the binding site for the base of dNTP. Furthermore, HIV RT does not contain an α helix corresponding to the O helix in KF (Fig. 2).

Model building of primer-template onto RT (see below) and onto KF (*50*) suggests a possible alternative role for the "fingers" subdomain; it may form the binding site for the template strand in the region of the primer terminus. In the case of both enzymes, placement of the primer terminus near to the highly conserved trio of carboxylate side chains in the palm subdomain that are at the center of the catalytic site necessarily puts the template strand in contact with the fingers. If, indeed, the fingers bind the template, a possible explanation emerges for the observation that only the DNA-dependent DNA polymerases and DNA-dependent RNA polymerases have a sequence that is similar to that of the O helix of KF (*12*): perhaps a "fingers" subdomain homologous to that of KF can use DNA templates, whereas the fingers subdomain of RT appears likely to bind a template capable of an A-form conformation, which either an RNA or a DNA template can assume.

The connection subdomain has no counterpart in the KF structure. Its location between the polymerase and RNase H active sites in p66 implies a role in holding the two domains together as well as binding the primer-template (see below). It also interacts nonsymmetrically with the connection subdomain of p51.

The structural comparison of HIV reverse transcriptase with KF supports the hypothesis based on amino acid sequence comparisons that all polymerases are to some extent related to each other. Perhaps not surprisingly, the conservation is greatest in the subdomain that forms the catalytic site. Sequence similarities among all categories of polymerases have been noted primarily in two motifs [called motif A and motif C (*12*) and in Fig. 6B] that contain the three carboxylates that are essential to the polymerase catalytic site.

More extensive amino acid sequence similarities (motifs B, D, and E in Fig. 6B) have been noted among all polymerases that employ an RNA template (*14*) and extended to included telomerase (*17*), which has a firmly bound RNA as a template for adding telomere sequences onto chromosomes. While bovine poly(A) (polyadenylate) polymerase, which adds poly(A) onto RNA, is reported to show these similar regions (*15*), yeast poly(A) polymerase does not contain two of the catalytic Asp residues implying another alignment (*16*). The five conserved regions are located in the palm subdomain and form a conserved surface at the bottom of the cleft, consistent with the hypothesis that this subdomain is responsible for catalysis of the polymerase reaction and found in all polymerases (Fig. 6B).

Model of primer-template binding. An obvious groove exists between the RNase H active site, as marked by two bound divalent metal ions, and the pol active site, marked by Asp^{185} and Asp^{186}, suggesting how the duplex product of DNA synthesis might bind (Fig. 5). Primed DNA synthesis

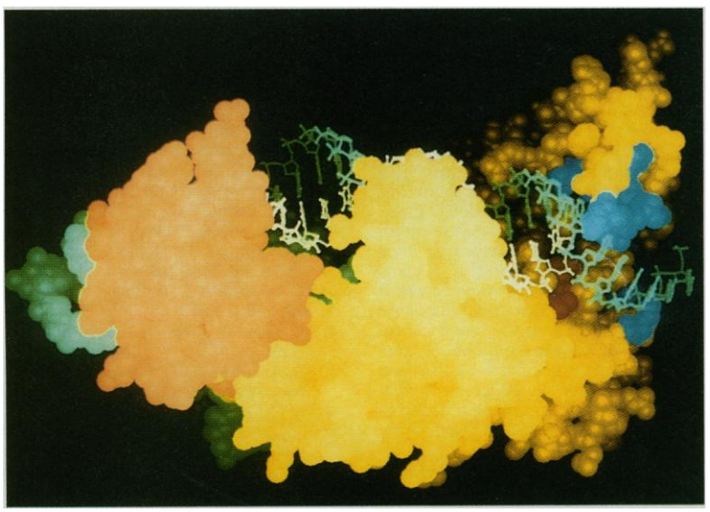

Fig. 7. A space-filling drawing of RT showing the positions of the AZT- and ddI-resistant mutations (blue) relative to the model built template (green)-primer (white). The view is approximately 90° from that of Figs. 2 and 3. The pol domain of p66 is yellow, p51 is green, and the RNase H domain is salmon. The vertical protrusion in the upper center is the thumb; the fingers are in back and to the right. The 3′ end of the white primer strand is adjacent to Asp^{185} Asp^{186} (red). Residues that confer AZT and ddI resistance are contained in a stretch of polypeptide from residues 66 through 75 as well as Thr^{215} and Lys^{219} and are shown in blue. It appears that they may be conferring resistance by affecting interactions between the protein and the template strand.

with the viral RNA template is accompanied by the cleavage of the RNA template by RNase H that has been estimated to be some 16 to 18 nucleotides down stream of synthesis (*53*). Thus, the RNA·DNA duplex product of synthesis, which is presumably the A form, must lie between these two active sites with the 3′ end of the template strand contacting the divalent metals of RNase H and the 3′ terminus of the primer strand near the DD sequence. Accordingly, a model of A-form DNA·RNA hybrid constructed with a rise per residue of 3.0 Å (*54*) was easily fit into the cleft. A backbone phosphate of the template strand can be positioned interacting with the two metal ions while the 3′ OH of the primer strand is near the three catalytically important aspartic acid residues (Figs. 5, B and D). In order to fit snugly into the groove along its full length, the DNA would have to be modestly bend toward the protein in the middle (which we have not done). In this model, there are about 20 base pairs of A-form DNA·RNA hybrid between the primer terminus and the RNase H active site. Consistent with this model is the cross-linking of dT_{15} to the RT homodimer (*55*); cross-links at L289-T290 and L295-T296 are in the thumb and in contact with the DNA·RNA at the minor groove.

The location of p51 relative to the model built A-form duplex implies a possible role for the smaller subunit in binding the $tRNA_3^{Lys}$-template complex (Fig. 5C). Particularly notable is the position of the p51 "connection" domain which would appear to be able to contact the tRNA–viral RNA duplex as well as the eighteen anticodon and D stems and loops of the tRNA that remain after the eighteen 3′ nucleotides unravel and base pair with the template. The RT heterodimer has been separately co-crystallized with tRNA and with duplex DNA which should allow these suggestions to be directly tested.

The orientation of DNA synthesis in the polymerase cleft (Fig. 5) is opposite to that hypothesized earlier for KF (*39*), but consistent with the recently determined crystal structure of a KF complexed with duplex DNA (*50*). It appears that single-stranded template (RNA or DNA) comes in from the top left of the cleft as viewed in Fig. 5 and the duplex product emerges from the bottom right as synthesis proceeds.

Drug binding sites. Nevirapine, developed by Boehringer Ingelheim Pharmaceuticals, is one of a class of non-nucleotide analogue inhibitors of HIV-1 RT that shows promise as a potential anti-AIDS drug (*2*). It shows noncompetitive inhibition and both the primer-template and dNTP substrates bind to the Nevirapine-RT complex. An azido modification of Nevirapine has been cross-linked (*25, 38*) to a peptide containing the sequence YMDD (183 to 186), which is highly conserved among DNA polymerases (*12*) and implicated in the catalytic site (*11, 48*).

Nevirapine binds in a deep pocket that lies between the β sheets of the "palm" and at the base of the "thumb" subdomains close to the expected primer terminus, but not overlapping the anticipated DNA binding sites (Fig. 6A). This pocket does not exist in p51 and accordingly Nevirapine does not bind to the smaller subunit (*25*). The inhibitor lies on top of a β hairpin motif that contains the Asp^{185}, Asp^{186} sequence at its β bend (Fig. 6B). The side chains of Tyr^{181} and Tyr^{188} are in contact with the hydrophobic inhibitor, consistent with the observation that mutation of Tyr^{181} to Ile reduces the affinity of enzyme for nevirapine (*38*).

The mechanism by which Nevirapine inhibits RT is unknown, but the structure of this complex suggests at least two possibilities. Nevirapine could be acting like sand in the gears of a machine by preventing movements of the "thumb" subdomain relative to the catalytic "palm" subdomain that may be essential to the catalytic cycle. Large changes in the position of the "thumb" of Klenow fragment have been observed in the presence and absence of a primed DNA substrate (*56*) and we anticipate similar substrate-induced changes in the "thumb" of RT. Additionally or alternatively, the inhibitor could be indirectly affecting the conformation of critical active, site aspartic acid residues that are nearby.

Residues whose mutation renders RT insensitive to AZT or ddI are at the polymerase active site and cluster in two regions: Met^{41}, Asp^{67}, Asp^{69}, Lys^{70}, Leu^{74} and Thr^{215}, Lys^{219} (*4*). The approximate locations of these two reigons, the former on the "fingers" subdomain and the latter on the "palm," suggest that these residues play a role in protein-template interaction (Fig. 7). Mutations in these residues are unlikely to be effecting their phenotype by altering direct interactions with the modified ribose of ddI or AZT. Rather, it appears likely that they alter the ability of this enzyme to discriminate among these modified nucleotides by altering interactions between the protein and the template.

The structure of RT should facilitate the design of new inhibitors of this enzyme and suggest strategies that might help to circumvent the ability of mutant RT molecules to escape the potency of inhibitors.

REFERENCES AND NOTES

1. H. Mitsuya, R. Yarchoan, S. Broder, *Science* **249**, 1533 (1990).
2. V. J. Merluzzi *et al.*, *Science* **250**, 1411 (1990).
3. R. Pauwels *et al.*, *Nature* **343**, 470 (1990).
4. B. A. Larder and S. D. Kemp, *Science* **246**, 1155 (1989); M. H. St. Clair *et al.*, *ibid.* **253**, 1557 (1991).
5. C-K. Shih *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 9878 (1991).
6. D. M. Lowe *et al.*, *Biochemistry* **27**, 8884 (1988); T. Unge *et al.*, *AIDS Res. Hum. Retroviruses* **6**, 1297 (1990); L. F. Lloyd *et al.*, *J. Mol. Biol.* **217**, 19 (1991); A. Jacobo-Molina *et al.*, *Proc. Natl. Acad.*

Sci. U.S.A. **88**, 10895 (1991).
7. A. Jacobo-Molina and E. Arnold, *Biochemistry* **30**, 6351 (1991).
8. At a resolution of 3.5 Å and with poorer phasing in the 4.0 to 3.5 Å resolution range, the course of the polypeptide in the two pol domains has been correctly traced, but the side-chain positions are less certain. Although a model with complete side chains has been built, thereby allowing Figs. 5 and 7 to be made, placement of residues may be off by 1 to 3 residues (or more) in many areas because it is often not possible to be certain how many residues exist in loops and turns and because the shape of side-chain electron density is rarely definitive. Thus, we have not put residue numbers on our models. The placement of Asp^{185} and Asp^{186} is more secure, since nevirapine has been cross-linked to a peptide containing these residues.
9. H. Varmus, *Sci. Am.* **257**, No. 3, 56 (1987); S. P. Goff, *J. Acquired Immune Def. Syndr.* **3**, 817 (1990).
10. M. S. Johnson, M. A. McClure, D-F. Feng, J. Gray, R. F. Doolittle, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 7648 (1986).
11. B. A. Larder, D. J. Purifoy, K. L. Powell, G. Darby, *Nature* **327**, 716 (1987).
12. M. Delarue, V. Pock, N. Tordo, D. Moras, P. Argos, *Protein Eng.* **3**, 461 (1990).
13. P. Argos, *Nucl. Acid Res.* **16**, 9909 (1988).
14. O. Poch, I. Sauvaget, M. Delarue, N. Tordo, *EMBO J.* **8**, 3867 (1989).
15. T. Raabe, F. J. Bollum, J. L. Manley, *Nature* **353**, 229 (1991).
16. J. Lingner, J. Kellermann, W. Keller, *Nature* **354**, 496 (1991).
17. V. Lundblad and E. H. Blackburn, *Cell* **60**, 529 (1990).
18. K. Bebenek and T. A. Kunkel, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 4946 (1990); Y. Takeuchi, T. Nagumo, H. Hoshino, *J. Virol.* **62**, 3900 (1988).
19. B. D. Preston, B. J. Poiesz, L. A. Loeb, *Science* **242**, 1168 (1988).
20. J. D. Roberts, K. Bebenek, T. A. Kunkel, *ibid.*, p. 1171.
21. F. DI Marzo Veronese *et al.*, *Science* **231**, 1289 (1986); E. M. Wondrak, J. Lower, R. Kurth, *J. Gen. Virol.* **67**, 2791 (1986); M. M. Lightfoote *et al.*, *J. Virol.* **60**, 771 (1986).
22. S. F. J. LeGrice, T. Naas, B. Wohlgensinger, O. Schatz, *EMBO J.* **10**, 3905 (1991).
23. C. Barat *et al.*, *ibid.* **8**, 3279 (1989).
24. Z. Hostomsky, Z. Hostomska, T.-B. Fu, J. Taylor, *J. Virol.* **66**, 3179 (1992).
25. J. C. Wu *et al.*, *Biochemistry* **30**, 2022 (1991).
26. R. Weiss *et al.*, Eds., *RNA Tumor Viruses* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, ed. 2, 1982).
27. L. A. Kohlstaedt and T. A. Steitz, *Proc. Natl. Acad. Sci. U.S.A.*, in press.
28. R. D'Aquila and W. C. Summers, *J. Acc. Imm. Def.* **2**, 579 (1989).
29. Ng.-h. Xuong, D. Sullivan, C. Nielsen, R. C. Hamlin, D. H. Anderson, *Acta Cryst.* **B41**, 267 (1985); Ng.-h. Xuong, C. Nielsen, R. C. Hamlin, D. H. Anderson, *J. Appl. Cryst.* **18**, 342 (1985).
30. Program written by Z. Otwinowski.
31. To reduce errors in the heavy atom parameter refinement among derivatives sharing the same major sites, the correlation matrix of Sygusch (32) was introduced in the phase calculation. This improved the phases and eliminated "ghost" peaks at the heavy atom sites. Since phase probability profiles derived from more than three derivatives are rather complicated, a 10° rms phase difference was observed between centroid phases calculated directly from the distributions and phases calculated from the Hendrickson-Lattman representations of such probability profiles (*33*). To improve the accuracy, the third and fourth orders of a Fourier expansion of the logarithm of these phase distributions were added to the usual first and second orders used in the Hendrickson-Lattman coefficients.
32. J. Sygusch, *Methods Enzymol.* **115**, 15 (1985).
33. W. A. Hendrickson and E. E. Lattman, *Acta Cryst.* **B26**, 136 (1970).
34. B. C. Wang, *Methods Enzymol.* **115**, 90 (1985).
35. Coordinates kindly provided by D. Matthews.
36. J. F. Davies, II, Z. Hostomska, Z. Hostomsky, S. R. Jordan, D. A. Mattews, *Science* **252**, 88 (1991).
37. To optimize the relative orientations of related subdomains of the heterodimer, an approximately rotated and translated map was first converted into pseudoatoms at each grid point. This set of pseudoatoms was refined as a rigid body in the space group *P1* against the back transform of a selected region of the unrotated MIR electron density map. Once refinement versus the initially selected electron density was complete, electron density selection was repeated based on the newly determined relative subdomain positions, and further cycles of refinement were carried out. All map rotations and translations that were used for orientational refinement and averaging were performed with modified versions of the SKEWPLANES procedure in the CCP4 programming package. Nine cycles of averaging and structure factor combination improved the electron density corresponding to the region of nevirapine which is bound only to the p66 monomer and was an unaveraged region of the map (Fig. 1B). While the continuity of the backbone was generally improved in this averaged map, the side-chain electron density was not always better because of the inexact similarities of the subdomain structures.
38. K. A. Cohen *et al.*, *J. Biol. Chem.* **266**, 14670 (1991).
39. D. L. Ollis, P. Brick, R. Hamlin, N. G. Xuong, T. A. Steitz, *Nature* **313**, 762 (1985).
40. L. Castagnoli *et al.*, *EMBO J.* **8**, 621 (1989).
41. T. A. Steitz, R. J. Fletterick, W. F. Anderson, C. M. Anderson, *J. Mol. Biol.* **104**, 197 (1976); D. M. McKay and T. A. Steitz, *Nature* **290**, 744 (1981).
42. Z. Hostomska, D. A. Matthews, J. F. Davies, II, B. R. Nodes, Z. Hostomsky, *J. Biol. Chem.* **266**, 14697 (1991).
43. S. C. Harrison, *Trends Biochem. Sci.* **3**, 3 (1978).
44. W. Yang, W. A. Hendrickson, R. J. Crouch, Y. Satow, *Science* **249**, 1398 (1990); K. Katayanagi *et al.*, *Nature* **347**, 306 (1990).
45. Z. Hostomsky, Z. Hostomska, G. O. Hudson, E. W. Moomaw, B. R. Nodes, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 1148 (1991).
46. C. Joyce and T. A. Steitz, *Trends Biochem. Sci.* **12**, 288 (1987).
47. D. Ollis, C. Kline, T. A. Steitz, *Nature* **313**, 818 (1985).
48. B. A. Larder, S. D. Kemp, D. J. M. Purifoy, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 4803 (1989).
49. A. H. Polesky, T. A. Steitz, N. D. F. Grindley, C. M. Joyce, *J. Biol. Chem.* **265**, 14579 (1990); A. H. Polesky *et al.*, *ibid.* in press.
50. L. S. Beese and T. A. Steitz, unpublished data.
51. ———, *EMBO J.* **10**, 25 (1991).
52. J. Rush and W. H. Konigsberg, *J. Biol. Chem.* **265**, 4821 (1990); V. N. Pandey and M. J. Modak, *ibid.* **263**, 6068 (1988); A. Basu and M. J. Modak, *Biochemistry* **26**, 1704 (1987); A. Basu, R. S. Tirumalai, M. J. Modak, *J. Biol. Chem.* **264**, 8746 (1989).
53. F. Oyama, R. K. Kikuchi, R. J. Crouch, T. Uchida, *J. Biol. Chem.* **264**, 18808 (1989); E. S. Furfine and J. E. Reardon, *ibid.* **266**, 406 (1991).
54. Model RNA · DNA hybrid was built with a modified version of the program DNA-FIT, written by J. Warwicker.
55. A. Basu, K. K. Ahluwalia, S. Basu, M. J. Modak, *Biochemistry* **31**, 616 (1992).
56. J. Friedman, L. Beese, T. A. Steitz, unpublished observations.
57. J. P. Priestle, *J. Appl. Cryst.* **21**, 572 (1988).
58. C. Branden and J. Tooze, *Introduction to Protein Structure*, (Garland, New York, 1991).
59. Figure drawn with Szazam by A. Perlo and MAXIM by M. Rould, Yale University.
60. M. L. Connolly, *J. Mol. Graphics* **3**, 19 (1985).
61. M. Carson, *J. Appl. Cryst.* **24**, 958 (1991).
62. We thank T. Warren (Boehringer Ingelheim Pharmaceuticals, Inc.) and D. Boisvert who provided important technical assistance in the purification and crystallization of HIV RT, W. Summers and R. D'Aquila for providing their RT expression clone, S. Schultz for optimization of RT overexpression, and Z. Otwinowski for advice and discussions. Supported in part by NIH grant GM 39546. The coordinates of the α carbon atoms have been sent to the Protein Data Bank at Brookhaven.

19 March 1992; accepted 26 May 1992