The EMBO Journal Vol.17 No.15 pp.4535–4542, 1998

The structure of HIV-1 reverse transcriptase complexed with an RNA pseudoknot inhibitor

Joachim Jaeger[1,2], Tobias Restle[1,3] and Thomas A.Steitz[1,4,5,6]

[1]Department of Molecular Biophysics and Biochemistry, [4]Department of Chemistry and [5]Howard Hughes Medical Institute, Yale University, New Haven, CT 06520-8114, USA

[2]Present address: School of Biochemistry and Molecular Biology, University of Leeds, UK
[3]Present address: Max-Planck-Institut für Molekulare Physiologie, Dortmund, Germany

[6]Corresponding author

J.Jaeger and T.Restle contributed equally to this work

Small RNA pseudoknots, selected to bind human immunodeficiency virus type 1 (HIV-1) reverse transcriptase tightly, are potent inhibitors of reverse transcriptase. The co-crystal structure of reverse transcriptase complexed with a 33 nucleotide RNA pseudoknot has been determined by fitting the ligand into a high quality, 4-fold averaged 4.8 Å resolution electron density map. The RNA is kinked between stems S1 and S2, thereby optimizing its contacts with subunits of the heterodimer. Its binding site extends along the cleft that lies between the polymerase and RNase H active sites, partially overlaps with that observed for duplex DNA and presumably overlaps some portion of the tRNA site. Stem S2 and loop L1 stabilize the 'closed' conformation of the polymerase through extensive electrostatic interactions with several basic residues in helix I of the p66 thumb and in the p66 fingers domain. Presumably, this RNA ligand inhibits reverse transcriptase by binding to a site that partly overlaps the primer–template binding site.

Keywords: crystal structure/HIV-1/reverse transcriptase/RNA pseudoknot

Introduction

Human immunodeficiency virus type 1 reverse transcriptase (HIV-1 RT) is responsible for the replication of the single-stranded RNA genome into double-stranded DNA. The polymerase domain of HIV-1 RT has been one of the main targets for antiviral therapy, which has focused on the use of chain-terminating nucleoside analogs such as AZT or ddC/ddI (Mitsuya *et al.*, 1990) and non-nucleoside inhibitors such as Nevirapine (Merluzzi *et al.*, 1990; Kohlstaedt *et al.*, 1992). However, cellular toxicity together with the emergence of drug-resistant variants of the virus compromise the efficacy of such compounds (Larder and Kemp, 1989; Goody *et al.*, 1991; St. Clair *et al.*, 1991), providing a need for new classes of antiviral agents and alternative strategies (Restle *et al.*, 1990; Divita *et al.*, 1994). Tuerk and Gold (1990) have used SELEX to isolate small, high-affinity ligands for RT from libraries of randomized RNA. The possibility that such inhibitors might be useful in the treatment of AIDS has been considered by others, though the problem of delivering the inhibitor to the target cells is an unsolved challenge. The SELEX procedure yielded two classes of consensus sequences (type I and II) that fold into a pseudoknot structure (Tuerk *et al.*, 1992). Among the several types of RNA pseudoknots classified (Pleij *et al.*, 1985, 1989), the H-type pseudoknots fold into two short stems connected by two loops of variable length. The consensus type II RNA inhibitor of RT forms a typical H-type RNA pseudoknot. We describe here the co-crystal structure of RNA pseudoknot 1.1 bound to HIV-1 RT (Tuerk *et al.*, 1992).

The structures of HIV-1 RT have been determined and refined in several different crystal forms at resolutions between 3.5 and 2.2 Å: (i) complexed with Nevirapine; (ii) complexed with other non-nucleoside inhibitors; (iii) complexed with duplex DNA; or (iv) without any ligands (Kohlstaedt *et al.*, 1992; Jacobo-Molina *et al.*, 1993; Smerdon *et al.*, 1994; Ren *et al.*, 1995; Rodgers *et al.*, 1995). HIV-1 RT forms an extended, strikingly asymmetric heterodimer with subunits of 66 and 51 kDa. The p66 domain shows an overall architectural similarity to the polymerase domain of *Escherichia coli* DNA polymerase I and, like other polymerases, has been compared with a right hand. Its domains have been termed fingers (residues 1–85, 118–155), palm (86–117, 156–237), thumb (238–318), connection (319–426) and RNase H (427–554). Although p51 consists of the same domains except for RNase H, they are oriented differently. The large subunit, consisting of fingers, palm, thumb, connection and RNase H domains, adopts an open, 'active' conformation that readily accommodates the large double-stranded substrate (Kohlstaedt *et al.*, 1992; Jacobo-Molina *et al.*, 1993). The catalytically important residues are largely inaccessible in the p51 subunit, and mutagenesis studies have demonstrated that this smaller subunit plays a largely structural role (Le Grice *et al.*, 1991; Wang *et al.*, 1994). The subunit p51 has been proposed to form part of the binding site for the initiating $tRNA^{Lys3}$ (Kohlstaedt *et al.*, 1992; Richter-Cook, 1992). This notion is consistent with data on the cross-linking of tRNA to HIV-1 RT (Mishima and Steitz, 1995).

The polymerase and RNase H active sites are separated by ~60 Å. A comparison of three different crystal forms of HIV-1 RT revealed that the molecule has a specific flexibility that allows rotation of the polymerase active site relative to the rest of the molecule. This swivel motion may allow the polymerase to accommodate the rotational and translational movements of the growing nucleic acid duplex during the abortive initiation phase while tRNA is still bound (Jaeger *et al.*, 1994).

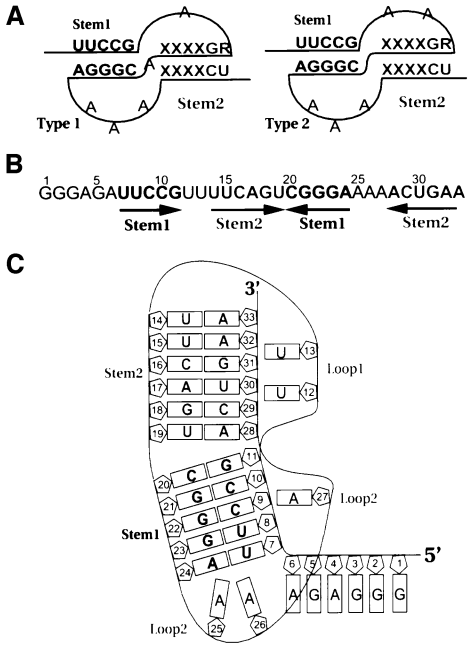


Fig. 1. Schematic diagrams of H-type RNA pseudoknots. (**A**) Type I and type II consensus sequences obtained by application of the SELEX procedure (Tuerk *et al.*, 1992). (**B**) The nucleotide sequence of the RNA pseudoknot co-crystallized with RT. The first five nucleotides at the 5′ end were introduced for efficient *in vitro* transcription by T7 RNA polymerase. (**C**) A schematic representation of the base pairs formed in the RT inhibitor pseudoknot.

Although the interactions between RNA pseudoknots and HIV-1 RT have been analyzed in some detail by biochemical studies and have been characterized further by chemical modification data (Burke *et al.*, 1996), the structure presented here is the first example of an RNA pseudoknot complexed with its target protein. This co-crystal structure may be of major interest since the molecular details governing the interactions between the pseudoknot and its target protein can provide a basis for rational modifications of the nucleic acid ligands, as well as additional randomization–selection strategies which could in turn lead to tighter binding inhibitors and possible pharmaceuticals.

Results

Crystallization and structure determination

HIV-1 RT was overexpressed in *E.coli* as previously described by D'Aquila and Summers (1989) and purified following a procedure developed by Kohlstaedt and Steitz (1992) including several modifications to eliminate host RNase activity that co-purifies with the RT (see Materials and methods). The 33 nucleotide RNA pseudoknot in Figure 1 was transcribed using T7 RNA polymerase and purified by gel electrophoresis. Its sequence differs from the previously published type II consensus sequence by inclusion of five nucleotides at the 5′ end to enhance the level of transcription. Gold and co-workers (Tuerk *et al.*, 1992) have shown that these pseudoknots bind to RT with nanomolar binding constants.

Crystals were grown at 4°C from 1.0 M citrate, 50 mM bis-Tris-propane (BTP) pH 7.0 and 100 mM ammonium sulfate using the hanging drop method. The co-crystals belong to the monoclinic space group C2 with four RNA–RT heterodimer complexes per asymmetric unit (420 kDa/asymmetric unit) with a solvent content of 71% (v/v). The structure of the complex was solved by molecular replacement using as a trial model a tetramer of RT heterodimers (Jaeger *et al.*, 1994; Rodgers *et al.*, 1995). The tetramer was generated by 'excising' two adjacent dimers of heterodimers from the crystal packing of a related orthorhombic form (F222) having unit cell dimensions of a = 162.2 Å, b = 172.2 Å, c = 634.2 Å and containing two RT heterodimers per asymmetric unit. The unit cell dimensions and the analyses of self-Patterson functions of both the monoclinic and the orthorhombic crystal forms indicated similarities in the packing arrangement of the heterodimers. The original phases for the pseudoknot complex derived by molecular replacement were improved dramatically by solvent flattening and 4-fold non-crystallographic symmetry averaging of the electron density of the four heterodimers. The resulting electron density map showed continuous density for most of the protein backbone atoms (Figure 2A) as well as additional density corresponding to the RNA, which lies between the p66 polymerase and RNase H domains. The current electron density maps, calculated using phases to 4.8 Å resolution derived from cycles of density averaging, show electron density $>2.5\ \sigma$ above the mean density for the RNA backbone and some additional features for several bases in the stems S1 and S2 (Figure 2B and C).

The solution NMR structure of a murine mammary tumor virus (MMTV) RNA pseudoknot (Shen and Tinoco, 1995), which causes ribosomal frameshifting, was used as an initial model to aid in electron density interpretation. While the nucleotide sequences and lengths of these two RNA pseudoknots are different, the stems of the NMR structure (the only known pseudoknot structure at the time of our analysis) superimpose reasonably well onto the averaged electron density maps.

Protein conformation

The four RT heterodimers forming the asymmetric unit of the RNA pseudoknot-complex crystal display a similar overall conformation, which most closely resembles that of the apo-enzyme (Rodgers *et al.*, 1995). The r.m.s. difference between the α-carbon positions of the p66 polymerase domain (residues 1–318) in the pseudoknot complex and those in the apo-RT heterodimer is 0.89 Å. In this closed form of p66 polymerase, the tips of the thumb and the fingers domains (residues 288–290 and 24–28, 58–65) are in contact (Figure 2). The size of the cavity or hole immediately above the active-site carboxylate residues (Asp110, Asp185 and Asp186) is 14×22 Å, too small to accommodate either stem of the pseudoknot or any other duplex nucleic acid substrate. The differences in the extent of domain closure among the four heterodimers is less than a 5° rotation of the thumb. The real space correlation coefficients between the electron density of the four heterodimers related by local symmetry is better than 0.88.

The relative orientation of the p66 polymerase domain with respect to the RNase H and p51 thumb domains corresponds to that observed in the orthorhombic crystal form of a Nevirapine complex (Jaeger *et al.*, 1994) and in the apo-enzyme (Rodgers *et al.*, 1995). Comparison of

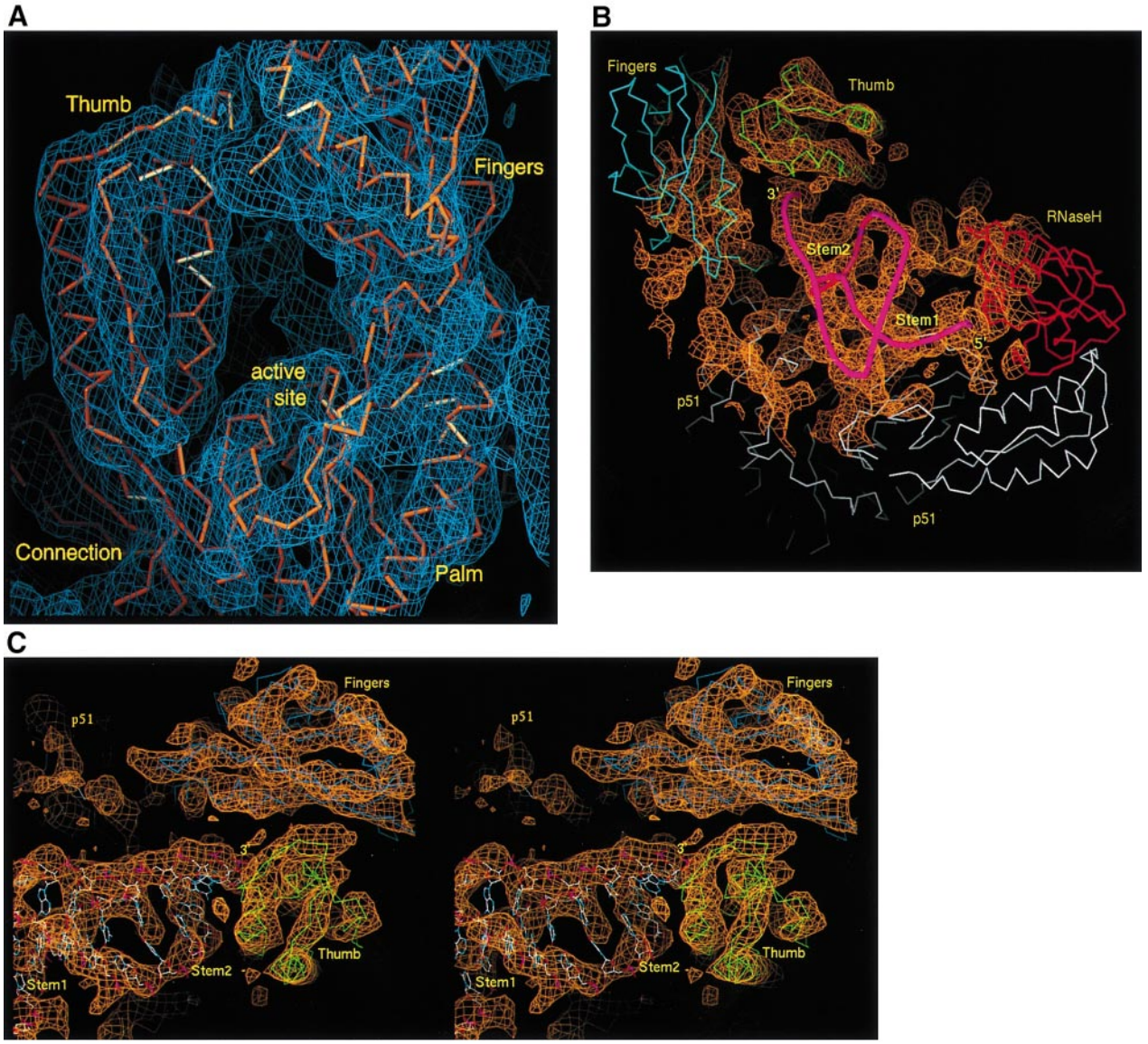
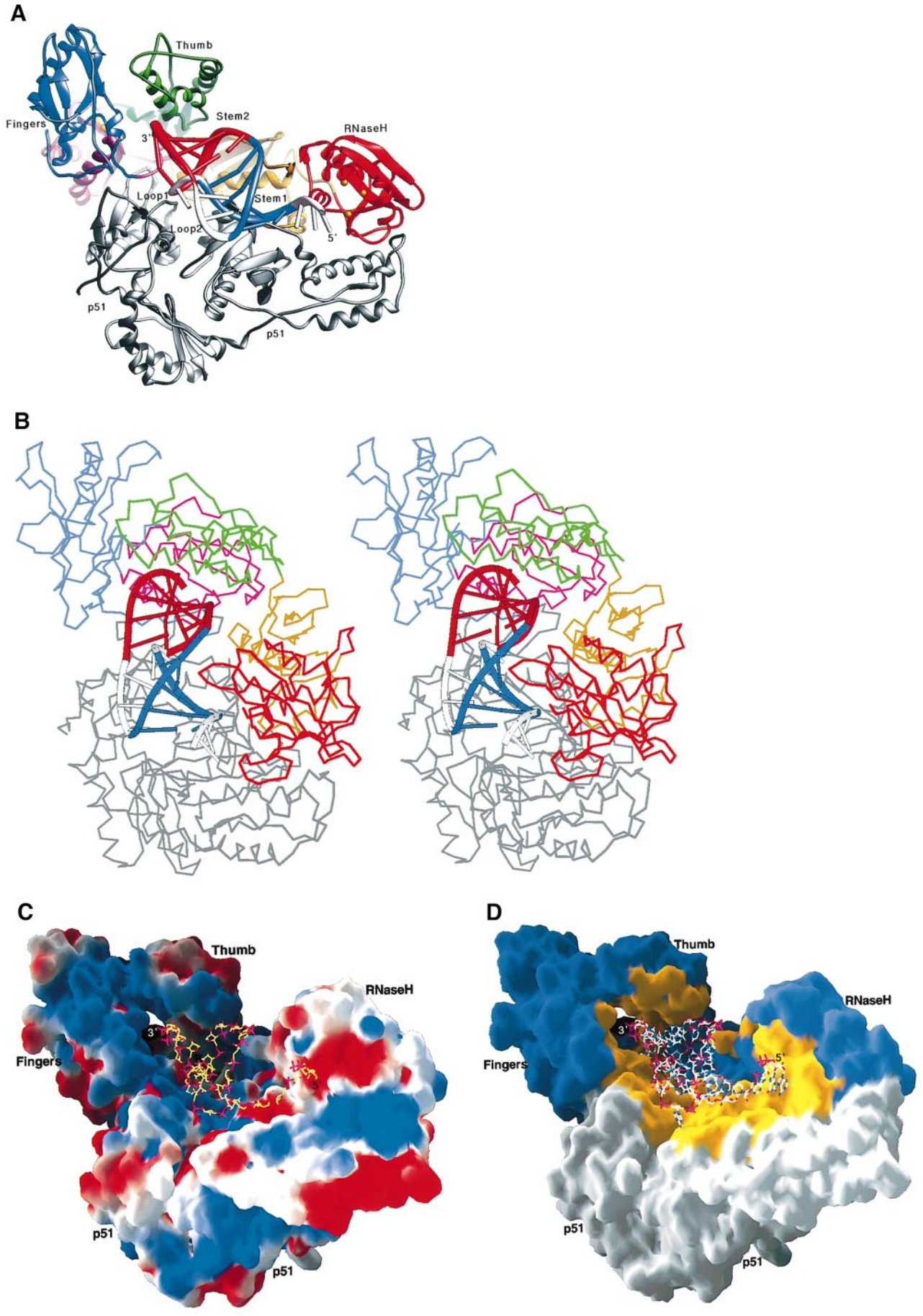


Fig. 2. Electron density maps of HIV-1 RT complexed with the RNA pseudoknot. (**A**) An α-carbon backbone of the p66 polymerase domain is superimposed on a 4-fold averaged $2F_o–F_c$ electron density map calculated at 4.8 Å resolution. The map is contoured at 1.2 σ above the mean. The extra electron density above the active site corresponds to stem S2 of the RNA pseudoknot (coordinates not shown). The size of the hole above the active site is too small to accommodate either stem of the pseudoknot or any other duplex nucleic acid substrate. (**B**) Four-fold averaged $2F_o–F_c$ electron density map contoured at 1.2 σ above the mean. The backbone coordinates of the p66 polymerase domain and the RNA pseudoknot are superimposed. The thumb (green) and fingers (blue) are in contact, thereby preventing access of primer–template substrate to the polymerase active site. A smooth purple backbone of the RNA is fitted to the extra density which appears when only the protein coordinates are included in the phase calculations. (**C**) The overall position of the RNA pseudoknot on the RT heterodimer with the corresponding 4-fold averaged $2F_o–F_c$ electron density map (contour: 1.2 σ above the mean). The backbone coordinates of HIV-1 RT and the RNA pseudoknot are superimposed. Stem S2 at the 3′ end of the pseudoknot forms strong electrostatic interactions with basic residues in helices H and I of the thumb domain (green) and the fingers domain (blue). The 5′ end is found in the vicinity of the RNase H active site. Loop L2 (foreground) interacts with the p51 connection domain.

the heterodimers in three different crystal forms previously demonstrated differences in protein conformation that correspond to a specific molecular swiveling motion around an axis that is nearly parallel to DNA when bound to the large active-site cleft (Jaeger *et al.*, 1994). The HIV-1 RT structures in the RNA pseudoknot co-crystal and that in the DNA complex display large differences in this molecular swivel. Superposition of the fingers and palm domains of the RT molecules in the RNA pseudoknot and the DNA complexes results in relative differences between corresponding α-carbon atoms of up to 6 Å in the RNase H domain.

RNA conformation

The electron density corresponding to the RNA backbone is continuous and contains high density features at the positions of the phosphate groups. Knowledge of structures of the two components in this complex from studies at higher resolution was essential for the interpretation of electron density features at 4.8 Å resolution. Thus, although electron density for the bases within the two pseudoknot stems is not always resolved, they have been modeled as Watson–Crick pairs as seen in the NMR solution structure. Density for the first three nucleotides at the 5′ end is not visible in the averaged density maps,


A
Thumb
Fingers
Stem2
3'
RNaseH
Loop1
Stem1
Loop2
5'
p51
p51
B
C
Thumb
RNaseH
3'
Fingers
p51
p51
D
Thumb
RNaseH
3'
5'
Fingers
p51
p51

but nucleotides 4–6 are observed ~10 Å away from the RNase H active site. Although there is strong electron density for all atoms of U12 in loop L1, density is only visible for the backbone of U13. The density corresponding to the backbone atoms of loop 2 is well defined. The bases in loops L1 and L2 have been placed, where possible, to fit the electron density or to form triple base pairs.

The electron density also indicates that the RNA pseudoknot is kinked between stems S1 and S2 by 59° from straight co-axial stacking, which is ~10° more than that observed in the NMR structure of the MMTV frameshifting pseudoknot. The kink generates a curved shape that optimizes the interactions between the RNA and the target protein. A similarly snug fit of the RNA onto the substrate-binding surface of RT would not occur if the stem S2 were co-axially stacked onto stem S1.

RNA interactions with protein

The pseudoknot RNA interacts extensively with both subunits of the HIV-1 RT heterodimer. Upon binding of the RNA, 2600 $Å^2$ of protein surface are buried (Figure 3B and C). Although the 33 nucleotide pseudoknot RNA is small compared with the 76 nucleotide $tRNA^{Lys3}$, it covers a large portion of the extended substrate-binding cleft of the heterodimer.

Stem S1 interactions. Stem S1 and the 5′ terminus interact with the connection domains of both subunits, the RNase H domain and the p51 fingers domain. Arg448, Gln500 and His538 of the p66 subunit are <5 Å from A4, G5 and A6 of the single-stranded 5′ extension. Residues Lys395 and Glu415–Asn418 of the p51 connection domain interact with U7, U8 and C10 of stem S1. Loop L2 interacts with residues located in a small cleft or canyon that lies between the p51 fingers and the p51 connection domain and has not been discussed previously; this cleft is packed with basic residues (lysines 13, 82, 154, 385, 388 and Arg78). Residues 78, 82 and 388 from this area contact the RNA. One is led to wonder whether the basic residues in this canyon may play a role in tRNA binding as well.

Stem S2 interactions. In contrast to stem S1, stem S2 and the 3′ terminus interact mostly with the p66 polymerase domain. Basic residues located both in helices H and I of the p66 thumb domain and in β-strands of the fingers domain form extensive contacts with the backbone of the RNA ligand. Helix I is oriented such that its helical axis runs across the minor groove of stem S2. Residues Lys275, Arg277, Lys281 and Arg284 are located <5 Å from the phosphate backbone of nucleotides U14–U19 and A28–A33 of stem S2 (Figure 1). U14 forms multiple interactions with Trp88, Glu89 (p66) and Lys22 (p51). Furthermore, lysine residues 13, 82 and 154 from the p66 fingers domain are in a position to interact with the backbone phosphate groups of nucleotides U30 and G31 at the 3′ end of the RNA. Loop L2 is oriented such that A25 is inserted in a shallow groove adjacent to Asp324, Lys388 and Glu413 on the p51 connection domain.

Discussion

In spite of the rather modest 4.8 Å resolution of this structure determination, it is possible to establish very reliably the conformation of the protein and RNA because first, the electron density map is of high quality due to averaging, and secondly the structures of the protein and a related RNA are already known in great detail. Similarly, the overall conformation of *E.coli lac* repressor bound to DNA was established at 4.8 Å resolution (Lewis *et al.*, 1996), again made possible by the prior determination of the high resolution structures of the *lac* repressor 'core' and the DNA-binding domain–operator DNA complex (Lamerichs *et al.*, 1989; Friedman *et al.*, 1995). In neither the RT nor repressor complexes with nucleic acid is it possible to establish with certainty the exact positions of the protein side chains and the precise nature of the hydrogen bonding interactions.

The SELEX procedure appears to have identified an RNA molecule whose uncomplexed solution structure is very similar to its structure when bound to RT and which binds to the unliganded conformation of RT. Thus, neither the RNA nor the protein undergoes a significant conformational change upon complex formation. This may be an

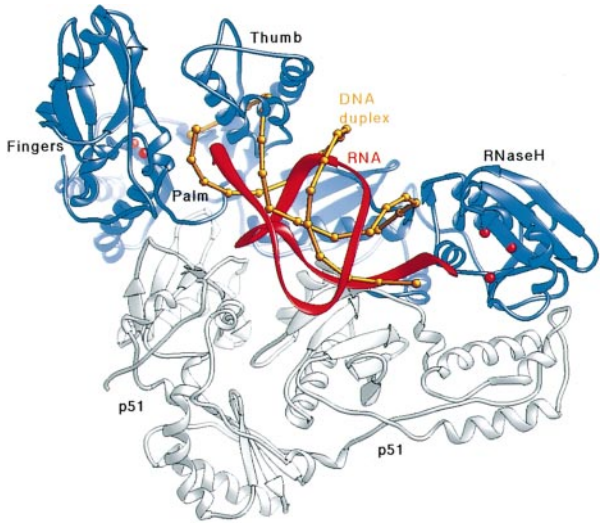


Fig. 4. Ribbon drawing (Carson, 1991) of the HIV-1 RT heterodimer complexed with duplex DNA and pseudoknot RNA. The RNA pseudoknot backbone is shown in red and the phosphorous atoms of the DNA duplex (Jacobo-Molina *et al.*, 1993) are depicted as yellow spheres. The color coding is as follows: p66 (blue), p51 (silver), RNA (red) and DNA (yellow). The DNA and RNA occupy the same overall site along the large substrate-binding cleft of the heterodimer. However, the interactions between the RNA and the p51 subunit are more extensive than in the DNA complex. The overlapping position of the RNA and DNA is consistent with the pseudoknot acting as a potent competitive inhibitor.

Fig. 3. The structure of the HIV-1 RT heterodimer complexed with a 33 nucleotide RNA pseudoknot. (**A**) A ribbon diagram (Carson, 1991) of RT color coded as follows: p66 fingers (blue), p66 palm (magenta), p66 thumb (green), p66 connection (yellow), RNase H (red) and p51 (silver). The schematic representation of the RNA is color coded as in Figure 1, with RNA stem S1 in blue and RNA stem S2 in red. (**B**) A stereo diagram of the RNA pseudoknot on the RT heterodimer. The protein is shown as an α-carbon backbone (color coding as in A). (**C**) Surface rendering of HIV-1 RT indicating the contact regions of the RNA. Surface residues <5 Å away from the pseudoknot inhibitor are colored in yellow; the two different subunits are shown in blue (p66) and silver (p51). (**D**) Surface rendering showing the electrostatic potential of the HIV-1 RT heterodimer. The solvent-accessible surfaces were calculated and rendered using the program GRASP (Nichols *et al*., 1991).

Table I. Data collection and refinement statistics

	Data set 1	Data set 2
X-ray source	CHESS beamline A1[a]	
Wavelength (Å)	0.92	
Resolution (Å)	30.0–4.75	30.0–5.40
Unit cell	166.7, 169.2, 331.4, 104.5	166.6, 168.2, 331.7, 104.0
Independent measurements	32 343	22 489
Completeness	80.9%	76.0%
(last shell)	(56.9%)	(55.5%)
R_{merge}[b]	6.4%	8.0%
(last shell)	(26.3%)	(22.7%)
$<F/\sigma F>$	15.7	12.9
(last shell)	(5.9)	(5.5)
R_{ave}[c]	22.8%	25.7%
R_{free}[d]	41.3%	42.1%

[a]All data were collected on 2×2 array CCD detector developed by Sol Gruner, Princeton.

[b] $$R_{merge} = \frac{\Sigma_{hkl}\,(\Sigma_{nsym}|I_{nsym.hkl} - K<I>|)}{\Sigma_{hkl}\,\Sigma_{nsym}\,<I>_{nsym.hkl}}$$

[c] $$R_{ave} = \frac{\Sigma_{hkl}|F_{obs} - F_{ave}\,|}{\Sigma_{hkl}\,F_{obs}}$$

[d] $$R_{free} = \frac{\Sigma_{hkl}|F^{free}_{obs} - F^{free}_{calc}\,|}{\Sigma_{hkl}\,F^{free}_{obs}}$$

F_{obs}, observed structure factor amplitudes; F_{calc}, calculated structure factor amplitudes from model coordinates; F_{ave}, structure factor amplitudes back-transformed from non-crystallographically averaged electron density.

important feature of selecting for tight binding ligands from large libraries, since any conformational change in either the protein or RNA would reduce the affinity, unless compensated by additional protein–RNA interactions.

The explanation of the sequence conservation observed by Tuerk *et al.* (1992) for nucleotides U7–G11 and C20–A24 within stem S1 does not arise from sequence-specific base interactions with HIV-1 RT. Except for A25 in loop L2, the possibility of other specific contacts between RT side chains and the nucleotide bases cannot be assessed unambiguously due to the limited resolution of the structure. The interactions between helix I and stem S2 appear to be more specific and to utilize more basic residues than those made to stem S1. Thus, it seems likely that the formation of base triplets between bases from loop L2 and base pairs in stem S1 rather than protein contacts are responsible for the sequence conservation in the selected inhibitor RNA pseudoknot.

Comparison of the RNA pseudoknot and DNA complexes with RT

The position of the pseudoknot RNA bound to RT largely overlaps the location of DNA when complexed to the enzyme (Figure 4). However, there are also some significant differences between these two nucleic acid complexes. The DNA complex can be made either by soaking duplex DNA into the crystal lattice formed by RT and a Fab antibody fragment or by co-crystallization. In this complex, the interactions seen between RT and the DNA duplex appear to be less extensive than those observed in the pseudoknot complex, although we cannot quantify this comparison since the full coordinates of the DNA complex are not available. The DNA appears to 'float' above the electrostatically positive cleft rather than sitting down snugly as the RNA does. Another difference between the RNA and DNA complexes is the far more extensive interaction between the RNA and the p51 subunit.

Comparison of the co-crystal structure of RT complexed to the pseudoknot with that of the RT complex with duplex DNA also reveals differences in a specific swivel motion between the p66 polymerase domain and the rest of the RT heterodimer. The structures of four RT heterodimers in three different crystal forms showed conformations that differed by the relative orientations of the polymerase active site and the rest of the molecule (Jaeger *et al.*, 1994). We had hypothesized previously that this molecular swivel motion allows the polymerase domain to move independently from the rest of the molecule in order to accommodate movements of the growing and translocating nucleic acid duplex in the initiation step when $tRNA^{Lys}$ is bound. The relative orientation of domains in the pseudoknot complex is very similar to that observed in the apo-RT structure and considerably different from the DNA–antibody–RT complex.

The H-type RNA pseudoknots fold into two short stems connected by two loops of variable length. The solution NMR structures of these types of pseudoknots have shown that the two stems are not stacked as a continuous duplex, as was suggested by modeling studies (Dumas *et al.*, 1987), but display an angle of ~120° between the two helix axes (or 60° less than co-axial) (Shen and Tinoco, 1995). The similar extent of kinking in both the MMTV and RT pseudoknots might be viewed as somewhat surprising, since the RT inhibitor studied here belongs to the consensus sequence that lacks the single-base insertion between the two stems. The MMTV pseudoknot, on the other hand, approximately corresponds to the type I consensus, which contains an adenine base inserted between the two stems, thereby generating the kink. However, since the type I and type II RNA pseudoknots presumably have the same overall shape in order to fit the RT-binding site, it seems likely that the pseudoknots can be kinked with or without the A inserted between the stems. Furthermore, the connecting L1 and L2 loops of the RT-inhibiting pseudoknot are shorter than those of the MMTV pseudoknot, which may in part be responsible for the kink in the RT pseudoknot. The longer loops of the MMTV pseudoknot are clearly incompatible with the averaged electron density maps, indicating that the electron density is free of bias that could have been introduced from the masks used in averaging. The structural similarity between the uncomplexed and complexed RNA pseudoknots would suggest that binding to RT does not change pseudoknot conformation.

It is not known to what extent the selected RNA pseudoknot inhibitors used in this study mimic the shape of the natural RNA substrates for RT. Some relationship might be anticipated. It had been hypothesized that the structure that is formed by stacking two short stems of base-paired regions on top of one another closely resembles features present in the tRNA fold (Pleij *et al.*, 1985; Dumas *et al.*, 1987); Westhof and co-workers (Felden *et al.*, 1996) have characterized the pseudoknotted three-

way junctions found in the 3′ terminus of tobacco mosaic virus that impose tRNA-like mimicry, as shown by their serving as substrates for histidine tRNA synthetase. However, it seems unlikely the RT pseudoknot is mimicking more than some aspects of the overall shape and charge distribution of the $tRNA^{Lys3}$ complex with the primer-binding site, since the underlying backbone structure must be completely different.

Materials and methods

Preparation of RNA

RNA synthesis was carried out by *in vitro* transcription with T7 RNA polymerase using synthetic DNA oligonucleotides as described in Milligan *et al.* (1987). T7 RNA polymerase was purified according to a protocol described by Jeruzalmi (1995) using the expression system of Studier (Davanloo *et al.*, 1984). The RNA product was purified on 20% denaturing polyacrylamide gels. Typically, 1–1.5 mg of homogeneous RNA could be prepared from a 10 ml reaction. Prior to use in the co-crystallization trials, the pseudoknot RNA was refolded at a concentration of 300–400 μM at 65°C for 5 min followed by slow cooling to room temperature in 20 mM Tris–HCl, 25 mM NaCl and 5 mM $MgCl_2$ or 20 mM cacodylate buffer pH 6.5, 25 mM NaCl and 5 mM $MgCl_2$.

Protein purification

HIV-1 RT was purified from *E.coli* containing an overexpression vector for the RT gene prepared by D'Aquila and Summers (1989). Protein extraction was performed according to a slightly modified protocol originally developed by Kohlstaedt and Steitz (1992). The RT-containing flow-through of a DEAE–Sephacel column was applied to a heparin column and developed with a linear gradient from 25 to 500 mM NaCl using an inert Waters LC 625 HPLC System. The eluted RT was precipitated with 60% saturated ammonium sulfate and redissolved in a small volume (~5 ml) of 50 mM BTP pH 7.0, 250 mM ammonium sulfate, 0.02% β-octalglucoside and applied to a gel filtration column (Superdex 75) equilibrated in the same buffer. The purified protein was concentrated and dialyzed in a Schleicher & Schuell UH 100 unit and stored at 4°C. The concentration was determined using an extinction coefficient of 3.54 mg/cm (Kohlstaedt and Steitz, 1992). A typical purification yielded ~30 mg of highly pure RT per 100 g (wet weight) of cells. In order to remove RNase contaminations, a fourth purification step using a phenyl Sepharose column (Pharmacia) was performed.

Analysis of RNase contaminations in the RT preparations

Since contaminating *E.coli* RNases were a major concern in co-crystallization trials, all RT and pseudoknot RNA preparations were incubated routinely at equimolar concentrations of 50 μM either at 37°C for 30 min or at 4°C up to several days, which more closely resembled the co-crystallization conditions, and analyzed by denaturing gel electrophoresis. Only those preparations showing negligible RNA degradation were used for co-crystallization experiments.

Co-crystallization of RT and RNA

Crystals were grown at 4°C by vapor diffusion in hanging drops. Protein and RNA were mixed in 50 mM BTP, pH 7.0, 100 mM ammonium sulfate, 0.02% (w/v) hexyl-β-D-glucopyranoside, 10% (w/v) glycerol, 0.02% (w/v) sodium azide, 7% 1,4-dioxane (Sigma) at a concentration of 100 and 120 μM, respectively, and incubated for at least 10 min on ice. The protein–RNA solution was mixed with an equal volume of well solution containing 1.0–1.1 M sodium citrate, pH 6.9, 100 mM magnesium acetate and 3% 1,4-dioxane. Small sword-shaped crystals could be observed typically after 4–7 days. These grow to 0.4×0.15×0.075 mm within ~2 weeks; in a few cases, crystals up to 0.8×0.25×0.15 mm have been obtained. The co-crystals belong to the monoclinic space group C2 with unit cell dimensions of typically a = 162 Å, b = 169 Å, c = 331 Å and β = 105°.

Data collection and molecular replacement

Since the RT–pseudoknot co-crystals are highly radiation sensitive, crystals were flash frozen in liquid propane after soaking them at 4°C for 3–10 min in 200 μl of a cryo-protectant solution containing 1.0 M sodium citrate pH 6.9, 100 mM magnesium acetate, 3% 1,4-dioxane, 20% sucrose, 10% xylitol. The best freezing results were obtained when crystals were transferred in a single step into large volumes (2 ml/crystal) of cryo-protectant. Microdialysis techniques or longer soaking times to exchange mother liquor with cryo-solution were not successful. X-ray data were collected at 100°K.

A 7 Å resolution data set collected using a laboratory X-ray source had a completeness of 78% and an overall R_{sym} of 18%, and was used to solve the structure by molecular replacement. The four tetramer RT heterodimers found in the F222 crystal form served as the trial model. Data were analyzed first with self-rotation and self-Patterson functions. The assumption of four molecules per asymmetric unit resulted in a V_M = 3.72 $Å^3$/Da and a solvent content of 68–70% (Matthews, 1968) and was supported by similarities to the packing arrangements found in the orthorhombic crystal form (F222). The tetrameric trial model was generated by 'selecting' four closely packed RT heterodimers from the F222 crystal lattice. Self-rotation functions were calculated using structure factors calculated from tetrameric trial models in which the local dyad axes of the tetramer were oriented parallel to a and c* axes. The patterns of the calculated self-rotation functions were compared with those of the self-rotation functions based on the observed data. The translation problem was solved manually using molecular replacement and packing analyses as implemented in the program O (Jones *et al.*, 1991). 'Local' translation function searches only along x and z in the vicinity of the 'plausible' solution followed by refinement of the Patterson correlation coefficient confirmed the predicted, computer-generated packing. The crystallographic and free R-factors stayed above 57% for the randomly placed molecules subjected to the same procedure, whereas both were substantially lower for the correctly placed molecule (34 and 42%, for R_{cryst} and R_{free}, respectively).

Several data sets extending to Bragg spacings between 5.4 and 4.8 Å from flash-frozen RT–RNA co-crystals were collected at the CHESS A1 beamline. The wavelength was 0.92 Å and data were recorded on a 2×2 CCD area detector (2048×2048 pixels) designed by S.Gruner, Princeton University. Two of these data sets have been used for further structural analyses and refinement (Table I).

Structure determination and refinement

Phases derived by molecular replacement for the structure factor amplitudes were improved by density averaging using the program suite RAVE (Kleywegt *et al.*, 1996, 1997). Three different schemes were used for the electron density averaging and phase extension procedure. First, extension of data phased between 15 and 6.5 Å to the 4.8 Å diffraction limit produced reasonable $2F_o–F_c$ density for the protein and additional density in the polymerase cleft. By contrast, extension of the 15–6.5 Å resolution phases to both high and very low resolution produced maps of inferior quality. In this case, the density corresponding to long helices in the connection domain became disconnected. Finally, applying the first approach with SIGMAA weighting resulted in improved electron density, but with considerably inflated figures of merit and correlation coefficients. The non-crystallographic averaging using data between 15.0 and 4.8 Å usually converged after 20–25 cycles. The averaging R-factors converged at ~26%. The electron density correlation coefficient between dimer 2 and dimer 1 was ~0.90, whereas it was 0.83 for dimer 3 compared with dimer 1, and 0.77 for dimer 4 compared with dimer 1. The non-crystallographic symmetry averaging of the four RT heterodimers was followed by domain averaging. Each heterodimer was subdivided into four domains: RNA, p66 thumb, p66 minus the thumb and p51. The 4-fold averaging of these four domains resulted in a further drop in the averaging R-factor to ~22%.

Model building

The protein coordinates were adjusted by rigid-body refinement of each domain and by a real space fitting of individual helices and contiguous β sheets to electron density using the program O. An ensemble of four NMR structures of the MMTV pseudoknot RNA was docked manually into the 4-fold averaged density. The loops in the NMR-determined structure were truncated and the sequence of the RNA was changed to that shown in Figure 1. Refitting of this revised pseudoknot model resulted in a small rotation with a slight improvement in the real space correlation coefficient for the new RNA position.

Coordinates of the NMR structures were not included at any stage of the non-crystallographic density averaging and/or any of the map calculations. Moreover, the RNA coordinates were randomized to remove any model bias in the calculations of averaging masks. Masks were recalculated using a 2.8 Å solvent probe radius for the RNA and a 1.4 Å radius for the protein portion. The combined masks calculated with the program MAMA (Kleywegt *et al.*, 1997) were used in the subsequent averaging and density modification procedures implemented in the program DM (Cowtan, CCP4 program suite, 1994). The coordinates

have been deposited in the Brookhaven Protein Data Bank with accession number 1 hvv.

Acknowledgements

We would like to thank I.Tinoco and L.X.Shen for providing unpublished coordinates of the MMTV RNA pseudoknot structure, and J.Pata for helpful discussions and critical reading of the manuscript. The unrefined coordinates will be sent to the Protein Data Bank at Brookhaven. This work has been supported by NIH grant #GM39546 to T.A.S. T.R. was supported by a stipend from the Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie (BMBF, Programm Infektionsforschung).

References

Burke,D.H., Scates,L., Andrews,K. and Gold,L.J. (1996) Bent pseudoknots and novel RNA inhibitors of type 1 human immunodeficiency virus (HIV-1) reverse transcriptase. *J. Mol. Biol.*, **264**, 650–666.

Carson,M. (1991) Ribbons 2.0. *J. Appl. Crystallogr.*, **24**, 958–961.

CCP4: Collaborative Computing Project Number 4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr.*, **D50**, 760–763.

D'Aquila,R.T. and Summers,W.C. (1989) HIV 1 reverse transcriptase/ribonuclease H: high level expression in *Escherichia coli* from a plasmid constructed using the polymerase chain reaction. *J. AIDS*, **2**, 579–587.

Davanloo,P., Rosenberg,A.H., Dunn,J.J. and Studier,F.W. (1984) Cloning and expression of the gene for bacteriophage T7 RNA polymerase. *Proc. Natl Acad. Sci. USA*, **81**, 2035–2039.

Divita,G., Restle,T., Goody,R.S., Chermann,J.C. and Baillon,J.G. (1994) Inhibition of human immunodeficiency virus type 1 reverse transcriptase dimerization using synthetic peptides derived from the connection domain. *J. Biol. Chem.*, **269**, 13080–13083.

Dumas,P., Moras,D., Florentz,C., Giege,R., Verlaan,P., Van Belkum,A. and Pleij,C.W. (1987) 3-D graphics modelling of the tRNA-like 3′-end of turnip yellow mosaic virus RNA: structural and functional implications. *J. Biomol. Struct. Dynam.*, **4**, 707–728.

Felden,B., Florentz,C., Giege,R. and Westhof,E.A. (1996) The central pseudoknotted three-way junction imposes tRNA-like mimicry and the orientation of three 5′ upstream pseudoknots in the 3′ terminus of tobacco mosaic virus RNA. *RNA*, **2**, 201–212.

Friedman,A.M., Fischmann,T.O. and Steitz,T.A. (1995) Crystal structure of lac repressor core tetramer and its implications for DNA looping. *Science*, **268**, 1721–1727.

Goody,R.S., Muller,B. and Restle,T. (1991) Factors contributing to the inhibition of HIV reverse transcriptase by chain-terminating nucleotides *in vitro* and *in vivo*. *FEBS Lett.*, **291**, 1–5.

Jacobo-Molina,A. *et al*. (1993) Crystal structure of human immunodeficiency virus type 1 reverse transcriptase complexed with double-stranded DNA at 3.0 Å resolution shows bent DNA. *Proc. Natl Acad. Sci. USA*, **90**, 6320–6324.

Jaeger,J., Smerdon,S.J., Wang,J., Boisvert,D.C. and Steitz,T.A. (1994) Comparison of three different crystal forms shows HIV-1 reverse transcriptase displays an internal swivel motion. *Structure*, **2**, 869–876.

Jeruzalmi,D.J. (1995) Crystallographic studies of the T7 RNA polymerase and the T7 RNA polymerase/T7 lysozyme complex. PhD Thesis, Yale University.

Jones,T.A., Zou,J.Y., Cowan,S.W. and Kjeldgaard,M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr.*, **A47**, 110–119.

Kleywegt,G.J. (1996) Making the most of your search model. *CCP4/ESF EACBM Newslett. Protein Crystallogr.*, **32**, 32–36.

Kleywegt,G.J. *et al*. (1997) The crystal structure of the catalytic core domain of endoglucanase I from *Trichoderma reesei* at 3.6 Å resolution and a comparison with related enzymes. *J. Mol. Biol.*, **272**, 383–397.

Kohlstaedt,L.A. and Steitz,T.A. (1992) Reverse transcriptase of human immunodeficiency virus can use either human tRNA (3Lys) or *Escherichia coli* tRNA (2Gln) as a primer in an *in vitro* primer-utilization assay. *Proc. Natl Acad. Sci. USA*, **89**, 9652–9656.

Kohlstaedt,L.A., Wang,J., Friedman,J.M., Rice,P.A. and Steitz,T.A. (1992) Crystal structure at 3.5 Å resolution of HIV-1 reverse transcriptase complexed with an inhibitor. *Science*, **256**, 1783–1790.

Lamerichs,R.M., Boelens,R., van der Marel,G.A., van Boom,J.H., Kaptein,R., Buck,F., Fera,B. and Ruterjans,H. (1989) H-NMR study of a complex between the lac repressor headpiece and a 22 base pair symmetric lac operator. *Biochemistry*, **28**, 2985–2991.

Larder,B.A. and Kemp,S.D. (1989) Multiple mutations in HIV-1 reverse transcriptase confer high-level resistance to zidovudine (AZT). *Science*, **246**, 1155–1157.

Le Grice,S.F., Naas,T., Wohlgensinger,B. and Schatz,O. (1991) Subunit-selective mutagenesis indicates minimal polymerase activity in heterodimer-associated p51 HIV-1 reverse transcriptase. *EMBO J.*, **10**, 3905–3911.

Lewis,M., Chang,G., Horton,N.C., Kercher,M.A., Pace,H.C., Schumacher,M.A., Brennan,R.G. and Lu,P. (1996) Crystal structure of the lactose operon repressor and its complexes with DNA and inducer. *Science*, **271**, 1247–1254.

Matthews,B.W. (1968) Solvent content of protein crystals. *J. Mol. Biol.*, **33**, 491–497.

Merluzzi,V.J. *et al*. (1990) Inhibition of HIV-1 replication by a nonnucleoside reverse transcriptase inhibitor. *Science*, **250**, 1411–1413.

Milligan,J.F., Groebe,D.R., Witherell,G.W. and Uhlenbeck,O.C. (1987) Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Res.*, **15**, 8783–8798.

Mishima,Y. and Steitz,J.A. (1995) Site-specific crosslinking of 4-thiouridine-modified human tRNA (3Lys) to reverse transcriptase from human immunodeficiency virus type I. *EMBO J.*, **14**, 2679–2687.

Mitsuya,H., Yarchoan,R. and Broder,S. (1990) Molecular targets for AIDS therapy. *Science*, **249**, 1533–1544.

Nicholls,A., Sharp,K.A. and Honig,B. (1991) Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins*, **11**, 281–296.

Pleij,C.W. and Bosch,L. (1989) RNA pseudoknots: structure, detection and prediction. *Methods Enzymol.*, **180**, 289–303.

Pleij,C.W., Rietveld,K. and Bosch,L. (1985) A new principle of RNA folding based on pseudoknotting. *Nucleic Acids Res.*, **13**, 1717–1731.

Ren,J. *et al.* (1995) High resolution structures of HIV-1 RT from four RT–inhibitor complexes. *Nature Struct. Biol.*, **2**, 293–302.

Restle,T., Muller,B. and Goody,R.S. (1990) Dimerization of human immunodeficiency virus type 1 reverse transcriptase. A target for chemotherapeutic intervention. *J. Biol. Chem.*, **265**, 8986–8988.

Richter-Cook,N.J., Howard,K.J., Cirino,N.M., Wohrl,B.M. and Le Grice, S.F. (1992) Interaction of tRNA (Lys-3) with multiple forms of human immunodeficiency virus reverse transcriptase. *J. Biol. Chem.*, **267**, 15952–15957.

Rodgers,D.W., Gamblin,S.J., Harris,B.A., Ray,S., Culp,J.S., Hellmig,B., Woolf,D.J., Debouck,C. and Harrison,S.C. (1995) The structure of unliganded reverse transcriptase from the human immunodeficiency virus type 1. *Proc. Natl Acad. Sci. USA*, **92**, 1222–1226.

Shen,L.X. and Tinoco,I.,Jr (1995) The structure of an RNA pseudoknot that causes efficient frameshifting in mouse mammary tumor virus. *J. Mol. Biol.*, **247**, 963–978.

Smerdon,S.J., Jager,J., Wang,J., Kohlstaedt,L.A., Chirino,A.J., Friedman,J.M., Rice,P.A. and Steitz,T.A. (1994) Structure of the binding site for nonnucleoside inhibitors of the reverse transcriptase of human immunodeficiency virus type 1. *Proc. Natl Acad. Sci. USA*, **91**, 3911–3915.

St. Clair,M.H., Martin,J.L., Tudor-Williams,G., Bach,M.C., Vavro,C.L., King,D.M., Kellam,P., Kemp,S.D. and Larder,B.A. (1991) Resistance to ddI and sensitivity to AZT induced by a mutation in HIV-1 reverse transcriptase. *Science*, **253**, 1557–1559.

Tuerk,C. and Gold,L. (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science*, **249**, 505–510.

Tuerk,C., MacDougal,S. and Gold,L. (1992) RNA pseudoknots that inhibit human immunodeficiency virus type 1 reverse transcriptase. *Proc. Natl Acad. Sci. USA*, **89**, 6988–6992.

Wang,J., Smerdon,S.J., Jager,J., Kohlstaedt,L.A., Rice,P.A., Friedman,J.M. and Steitz,T.A. (1994) Structural basis of asymmetry in the human immunodeficiency virus type 1 reverse transcriptase heterodimer. *Proc. Natl Acad. Sci. USA*, **91**, 7242–7246.

Received March 17, 1998; revised and accepted May 27, 1998