Trapping of a catalytic HIV reverse transcriptase•template:primer complex through a disulfide bond

Huifang Huang[1]*, Stephen C Harrison[2] and Gregory L Verdine[1]

Background: HIV-1 reverse transcriptase (RT) is a major target for the treatment of acquired immunodeficiency syndrome (AIDS). Resistance mutations in RT compromise treatment, however. Efforts to understand the enzymatic mechanism of RT and the basis for mutational resistance to anti-RT drugs have been hampered by the failure to crystallize a catalytically informative RT–substrate complex.

Results: We present here experiments that allow us to understand the reason for the failure to crystallize such a complex. Based on this understanding, we have devised a new approach for using a combinatorial disulfide cross-linking strategy to trap a catalytic RT•template:primer•dNTP ternary complex, thereby enabling the growth of co-crystals suitable for high-resolution structural analysis. The crystals contain a fully assembled active site poised for catalysis. The cross-link itself appears to be conformationally mobile, and the surrounding region is undistorted, suggesting that the cross-link is a structurally passive device that merely acts to prevent dissociation of the catalytic complex.

Conclusions: The new strategy discussed here has resulted in the crystallization and structure determination of a catalytically relevant RT•template:primer•dNTP complex. The structure has allowed us to analyze possible causes of drug resistance at the molecular level. This information will assist efforts to develop new classes of nucleoside analog inhibitors, which might help circumvent current resistance profiles. The covalent trapping strategy described here may be useful with other protein–DNA complexes that have been refractory to structural analysis.

Addresses: [1]Department of Chemistry and Chemical Biology and [2]Howard Hughes Medical Institute and Department of Molecular and Cellular Biology, Harvard University, Cambridge MA 02138, USA.

*Present address: Department of Biochemistry, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA.

Correspondence: Gregory L Verdine
E-mail: verdine@chemistry.harvard.edu

Key words: chemical modification, crystal structure, disulfide cross-linking, HIV-1 reverse transcriptase, nucleic acid–protein interaction

Received: **24 January 2000**
Accepted: **24 February 2000**

Published: **25 April 2000**

Chemistry & Biology 2000, **7**:355–364

1074-5521/00/$ – see front matter

Introduction

The reverse transcriptase (RT) enzyme of the human immune deficiency virus (HIV) catalyzes the multistep conversion of the single-stranded viral RNA genome into a double-stranded DNA copy, which then integrates into the host genome [1,2]. Because of its central function in HIV proliferation, RT is a major target for therapeutic agents that combat HIV infection [3,4]. In the United States, seven RT inhibitors are currently licensed as drugs. On the basis of their mechanism of action, RT inhibitors are divided into two groups: non-nucleosides and nucleoside analogs. Non-nucleoside inhibitors are diverse in terms of chemical structure, but all bind to a common hydrophobic pocket in RT adjacent to the polymerase active site. Nucleoside analog inhibitors are prodrugs, which are converted into the active form, the 5′-triphosphate, by cellular enzymes. As a class, the nucleoside analog inhibitors lack the 3′-hydroxyl groups required for chain extension, and their incorporation by RT results in chain termination.

Development of resistance through mutations in RT seriously limits the long-term effectiveness of these drugs [5,6]. There is, therefore, widespread interest in understanding the mechanisms of resistance as a step toward developing more effective treatments for HIV infection. Structures have been available since 1993 for RT holoenzyme, RT complexed with a duplex DNA alone, and RT bound to a non-nucleoside inhibitor [7–14]. Until recently, however, attempts to understand the catalytic mechanism of RT and the influence of mutations on the enzyme have been hampered by the lack of a three-dimensional structure for the catalytic complex formed between RT, a template:primer, and a nucleotide triphosphate. The earlier structures yielded valuable insights into the overall architecture of the enzyme and provided a basis for understanding the mechanism underlying development of resistance to non-nucleoside inhibitors, but they left open key questions regarding the interaction of RT with an incoming nucleotide and with the template overhang. Moreover, the structures suggested, paradoxically, that many mutations that cause resistance to nucleoside analogs are located far outside the enzyme reactive site, leaving no obvious rationale for their effects on the enzyme. Resolution of this paradox is a pre-requisite for applying a rational approach to the development of new nucleoside analog drugs.

Despite concerted efforts in several laboratories, it has thus far not been possible to obtain high-quality co-crystals of RT bound noncovalently to a template:primer and a nucleoside triphosphate using conventional approaches. We describe here a biochemical analysis of RT–substrate binding that suggests that the previous inability to crystallize a noncovalent ternary complex resulted from insufficient specificity of RT for a polymerizable end in DNA. To channel the multiple equilibrating RT–substrate complexes into a homogenous catalytically relevant species, we have developed a strategy by which the enzyme is covalently cross-linked to the template:primer. An important aspect of this strategy is the choice of disulfide bond formation as the cross-linking chemistry, which allows the cross-link to be generated under equilibrium conditions, thereby avoiding kinetic trapping of strained or transient species. Using this approach, we have succeeded in crystallizing a catalytically relevant RT•template:primer•dNTP complex. The 3.2 Å structure of this complex has revealed a large domain shift that brings the residues responsible for resistance to nucleoside analog inhibitors close to the enzyme active site ([15]; PDB code 1RTD). The structure also revealed that the single-stranded portion of the template does not pass through a cleft near the active site as previously believed, but rather runs alongside the surface of the enzyme [15]. The detailed biochemical description of RT–DNA cross-linking provided here reveals that the cross-linking chemistry is remarkably sensitive to even subtle differences in the relative positions of the two reacting partners. The conclusion is reinforced by analysis of an extension of the cross-linking strategy to include disulfide trapping of RT on an RNA:DNA heteroduplex template:primer. Our detailed biochemical studies provide a blueprint for use of disulfide cross-linking in structural analysis of other challenging protein–nucleic acid complexes.

Results

Overexpression strategy

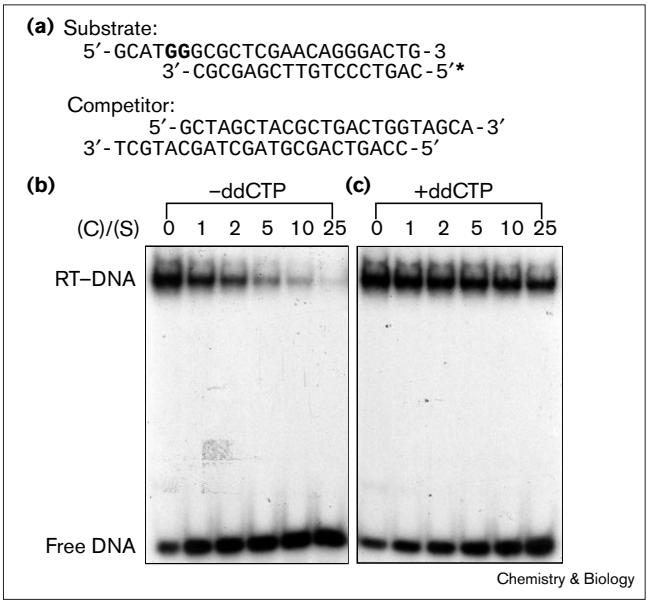
To facilitate the high-level overexpression and subsequent purification of wild-type and mutant RT enzymes, we constructed a new *Escherichia coli* overexpression system in which the two RT subunits, p66 and p51, are biosynthesized simultaneously. The coding sequences for both p66 and p51, each linked to an independent T7 promoter, were introduced into the plasmid pLM1 [16,17] oriented in opposite directions. A His_6 tag was attached at the carboxyl terminus of p51, to enable separation by Ni-NTA chromatography of homogenous RT heterodimer containing full-length p51 from the heterogeneous heterodimer that can arise from intracellular proteolysis of the p66•p66 homodimer. An additional advantage of encoding p66 and p51 independently, as opposed to relying on intracellular proteolysis of p66, is that it allows mutations to be made in a particular subunit. In our hands, the co-expression system routinely yields ~1 mg/l of highly purified RT.

Electrophoretic mobility shift assay (EMSA)

In an attempt to understand the reason for the failure to co-crystallize a noncovalent ternary RT•template:primer•dNTP complex, we decided to analyze the robustness of this complex using EMSA in a competition format. RT was briefly incubated with a radiolabeled polymerizable substrate (i.e. a duplex oligonucleotide that has a recessed 3′-end, Figure 1a). To this was added varying amounts of a nonpolymerizable competitor (i.e. a duplex oligonucleotide that has 3′-overhangs, Figure 1a). After allowing sufficient time for the establishment of equilibrium, the reaction mixture was run on a nondenaturing polyacrylamide gel to separate the RT-bound substrate from unbound substrate. This procedure was carried out in the absence of added nucleotide — a condition in which no polymerization takes place — and in the presence of ddCTP — a condition in which one cycle of polymerization takes place and an additional ddCTP molecule can bind noncovalently to the terminated RT•template:primer complex (Figure 1a).

In the absence of nucleotide, the nonpolymerizable DNA competed on a equal basis with the radiolabeled polymerizable substrate (Figure 1b); for example, addition of one mole equivalent of competitor reduced the intensity of the RT–substrate band by 50%, and addition of 25 equivalents reduced the intensity to 4% of the starting value. In

Figure 1



Analysis of affinity and specificity of RT–DNA substrate interaction using an electrophoretic mobility shift assay (EMSA). **(a)** Sequences of DNA substrate and competitor used in this study. **(b,c)** Competition assay of RT–DNA substrate interaction by increasing amounts of DNA competitor, as analyzed by nondenaturing polyacrylamide gel, in (b) the absence or (c) the presence of ddCTP.

the absence of nucleotide, therefore, RT shows no discrimination between a polymerizable DNA substrate and a nonpolymerizable competitor.

When ddCTP was added to the RT–substrate mixture, the ability of the nonpolymerizable DNA to compete was dramatically reduced (Figure 1b). For example, addition of a 25 equivalents of competitor over substrate resulted in only 44% reduction in the intensity of the RT•substrate complex. Taking into account all of the data in Figure 1b, we conclude that RT shows 6–10-fold preference to bind a polymerizable substrate over a nonpolymerizable one, in the presence of a terminating nucleotide co-substrate.

Encouraged by this result, we conducted extensive crystallization trials under nucleotide-addition conditions mirroring these used in EMSA, but the only crystals obtained did not contain DNA (data not shown). We therefore concluded that in the case of RT, 6–10-fold specificity is insufficient to produce a complex suitable for crystallization. We imagine that, with this limited level of specificity, there exist multiple equilibrating RT•substrate complexes under the conditions of crystallization, and this inhomogeneity impedes co-crystal formation.

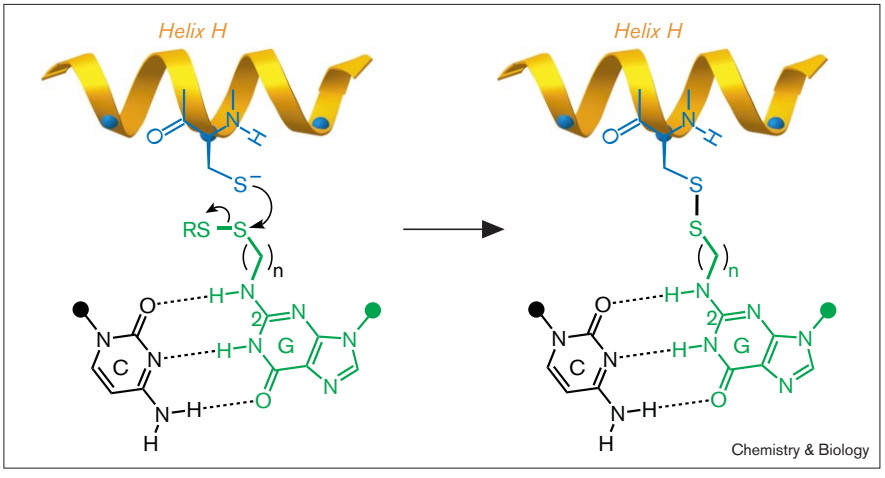
Strategy of disulfide cross-linking

Because the main problem appeared to be that RT and its co-substrates associate to form a heterogeneous mixture of complexes, we sought a means of channeling these into a single homogenous complex representing a catalytically relevant species. Covalent cross-linking seemed to represent an attractive possibility. Implementing such an approach is complicated, however, by the undesirable possibility of kinetically trapping a strained or otherwise catalytically irrelevant species. To avoid this problem, we elected to use an equilibrium (kinetically reversible) cross-linking scheme, in which the noncovalent interactions between RT and its co-substrates would dictate the formation of a covalent cross-link. In evaluating the various chemistries available for equilibrating cross-linking of RT to its template:primer, we chose disulfide bond formation, for the following reasons: the reaction through which disulfide bonds are formed, thiol-disulfide interchange, proceeds readily under physiological conditions [18,19]; disulfide bond formation is a mild and reversible chemical process, especially in the presence of a thiol reducing agent; and efficient chemistry exists for the site-specific attachment of tethered thiols into DNA [20–22]. Furthermore, thiol-bearing cysteine residues can be readily introduced into a protein through site-directed mutagenesis.

Selecting the positions of cross-link attachment in DNA and RT required an approximate notion of what base pairs and residues might lie near one another in the catalytic complex. Although there existed insufficient information to predict proximity with certainty, a hypothesis was presented by a published model [23] based on the crystal structure of RT•duplex DNA complex [8], together with the results of alanine-substitution metagenesis and energy-minimized calculations. This model proposed that amino acid residues positioned along one face of α-helix H in the RT p66 subunit (Gln258, Gly262 and Trp266) track along the minor groove of the template:primer. We reasoned that a cysteine residue introduced into this face of helix H should reach toward the 'floor' of the minor groove, where it could react with a thiol-bearing tether protruding from the minor groove floor toward helix H, thereby forming a disulfide cross-link and trapping the RT•DNA complex (Figure 2). To explore this possibility, we generated three variant RT proteins (Gln258→Cys-RT [Q258C-RT], Gly262→Cys-RT [G262C-RT], and Trp266→Cys-RT [W266C-RT], respectively) that have cysteine residues engineered individually into successive turns of helix H. At the same

Figure 2

Chemistry of disulfide bond formation. When appropriately aligned, the sidechain of an engineered cysteine residue (blue) in helix H (gold) of RT can react to a thiol group in the minor groove of DNA (activated as the mixed disulfide) to form a disulfide bond. The thiol group is tethered to N^2 of a dG (green) in the template:primer.



time, we used convertible nucleoside methodology [22] to introduce a thiol tether into the minor groove of DNA at the N^2-position of a guanine residue (N^2-G; refer to Figure 2) located several residues away from the recessed 3′-end of the template:primer.

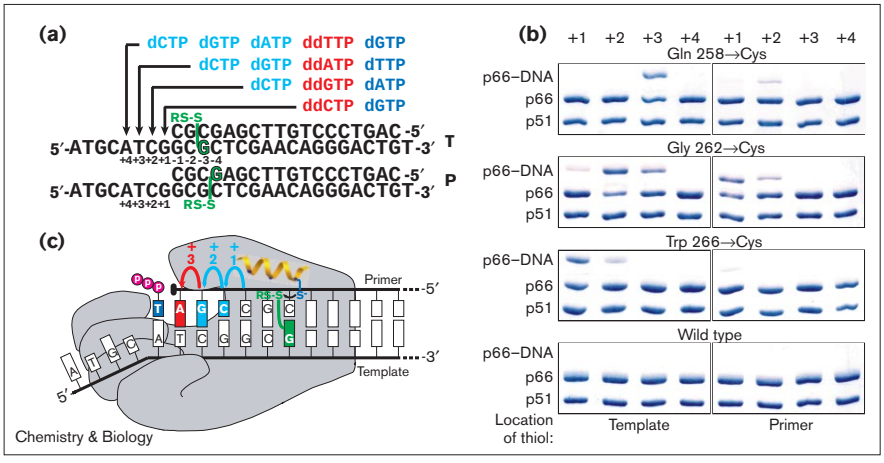
Taking into account the possibility that this model might deviate in certain details from the actual RT–DNA interaction, and expecting that the cross-linking reaction would be sensitive to factors of geometry and distance between the reacting partners, we devised a limited combinatorial approach whereby multiple potential cross-link configurations could be sampled (Figure 3a). We designed a template:primer that would allow defined cycles of polymerization followed by termination to be carried out simply by appropriate choice of terminating and nonterminating nucleotides. In this way, the enzyme could be 'ratcheted' forward on the template:primer one base pair at a time, while holding fixed position of the reactive thiol in DNA. Additional variations could be gained by transposing the location of the engineered cysteine along helix H, by tethering the reacting thiol in either the template strand or the primer strand, and by varying the tether length in DNA (Figure 2, n = 2 or 3).

Combinatorial cross-linking analysis

The various permutations of cysteine-mutant RT, extended/terminated template:primer, location of the tether, and tether length were analyzed for their ability to undergo cross-linking using denaturing polyacrylamide gel electrophoresis (Figures 3b,4a). The p51 and p66 subunits of RT are sufficiently different in size that they are clearly resolved by sodium dodecyl sulfate–polyacylamide gel electrophoresis (SDS–PAGE). Covalent linkage of either subunit to DNA would be expected to result in a loss of band intensity for that subunit, accompanied by the appearance of a new band that has retarded mobility, owing to the presence of an attached DNA strand. The results with the thiol group tethered in the template strand have been reported previously [15], and are reproduced here for completeness, as part of comprehensive analysis of the factors influencing cross-linking efficiency.

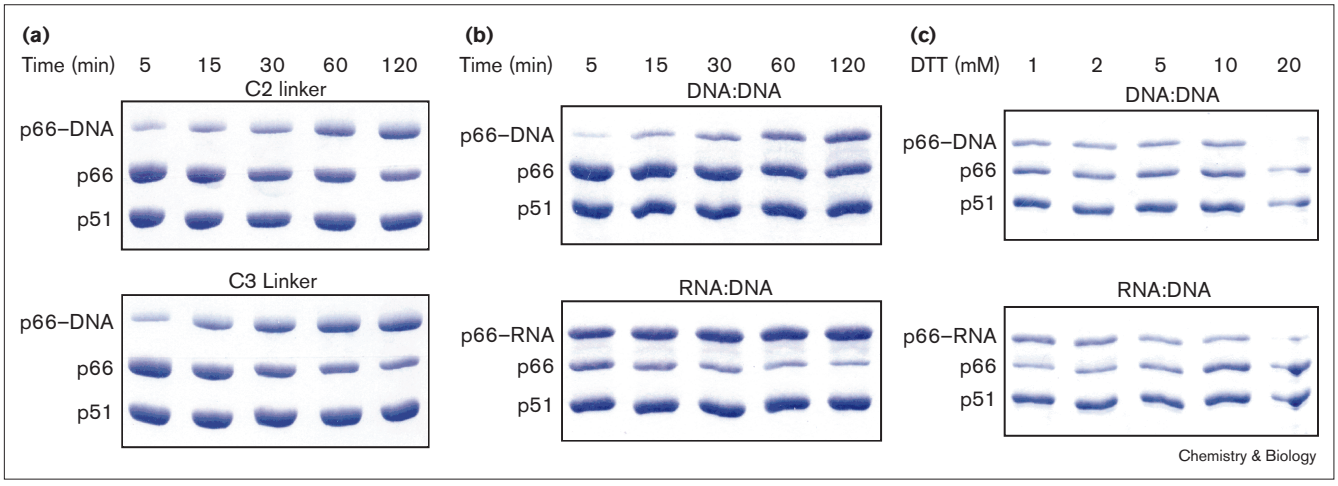
The analysis begins with a template:primer having a thiol group attached to the template strand (DNA T, Figure 3a). As indicated in the upper-left panel of Figure 3b, following one cycle of polymerization (+1 lane), no cross-linking was observed; nor was any observed after two cycles (+2). After three cycles of polymerization (+3), however, roughly half of the p66 subunit was found to be cross-linked to DNA. One further cycle of polymerization (+4) resulted in a complete loss of cross-linking. These results can be interpreted as follows: after three cycles of polymerization, the cysteine and DNA-thiol are optimally positioned to react; fewer or further cycles simply move the reacting partners away from the optimal configuration (Figure 3c).

Figure 3



Biochemical studies of RT–DNA cross-linking. **(a)** The sequence of the template:primer and the constituents of the dNTP/ddNTP mixtures used in this study. Polymerization was initiated and continued by successive extensions using dNTPs (cyan), and terminated through incorporation of a ddNTP (red). An additional dNTP (blue), present in the reaction mixture, occupies the active site of RT, pairing with the corresponding base of the template. A thiol group is either attached to the −3 dG in the template strand (DNA T), or attached to the −4 dG in the primer strand (DNA P). **(b)** SDS gel analysis of the reactions between the three cysteine-engineered RT mutants and the thiol-tethered template:primer under the four polymerization conditions illustrated in (a). In the nonreducing PAGE analysis used here with protein staining, disulfide cross-linking between the p66 subunit of RT and DNA results in the appearance of a new band having retarded mobility (p66–DNA), accompanied by reduced intensity of the p66 band. Slight difference of mobility of p66–DNA bands in the left and right panel of gels probably results from unequal size of the template strand relative to the primer strand attached. **(c)** Representation of the process leading to efficient cross-linking between Q258C-RT and DNA. The p66 subunit of RT is represented by the polymerase structural convention as a right hand (gray). After two cycles of polymeric reactions (cyan) and one cycle of termination (red), an additional dTTP (blue base, magenta for triphosphate) occupies the active-site pocket formed by the palm region and part of the finger region of p66. This configuration optimally positions the sidechain of Cys258 (blue) in helix H (gold) relative to the thiol group tethered to the dG (green) in the template strand, thereby resulting in efficient disulfide bond formation. The single stranded part of the template is depicted as passing by the surface of the finger region, rather than through the active-site groove as seen in the structure of RT•template:primer•dTTP ternary complex [15].

Figure 4



Kinetic studies of covalent complex formation of Q258C-RT with different template:primers under various conditions, as analyzed by denaturing SDS–PAGE under nonreducing conditions. **(a)** Covalent complex formation of Q258C-RT with a template:primer having either a C2- or a C3-tethered thiol group in the template strand of DNA. **(b)** Covalent complex formation of Q258C-RT with either a DNA:DNA or RNA:DNA template:primers. **(c)** Same as the 120 min time point in (b), except that the reaction was carried out at the presence of increasing amounts of reducing agent, dithiothreitol (DTT).

When the engineered cysteine residue in RT is transposed by one turn along helix H (Figure 3b, left G262C-RT panel), cross-linking is again positionally selective, but the highest efficiency of cross-linking now occurs after two cycles of polymerization. Remarkably, transposition of the engineered cysteine residue by one additional turn along helix H results in a further one-base-pair decrease in the optimal number of polymerization cycles (+1, left W266C-RT panel). Wild-type RT failed to cross-link the thiol-tethered template:primer under identical reaction conditions (Figure 3b, left wild-type panel), even though this form of the protein contains a cysteine residue on helix I (p66 C280) within ~5 Å of the DNA backbone.

When the tethered G was moved from position –3 of the template strand to position –4 of the primer strand (DNA P, Figure 3a), cross-linking was also positionally selective (Figure 3b, right-hand set of panels). Moving the tether by one base pair along the DNA, however, correspondingly reduced by one cycle the polymerization required to achieve optimal cross-linking (compare the paired sets of right-hand and left-hand panels in Figure 3b). Again, a monotonic decrease in the preferred number of polymerization cycles was observed as the position of the cysteine residue was transposed along successive turns of helix H, such that the W266C enzyme overshot its optimal cross-linking target after the addition of even one nucleotide. Although we did not analyze cross-linking in the absence of any added nucleotide, it is reasonable to expect that this will be the preferred cross-linking configuration for W266C-RT with the primer-tethered substrate. Shifting the tether from template (–3) to primer (–4) reduced the rate of cross-linking somewhat, as evidenced by reduced yields of cross-linked product, and therefore we chose the former for structural studies.

Taken together, these data demonstrate that the engineered thiol moieties on RT and on the template:primer participate in disulfide cross-linking, through a reaction that has exquisite positional selectivity.

Of the multiple cross-link configurations studied here, Q258C-RT, together with the template-tethered template:primer showed the greatest degree of positional selectivity and highest efficiency in disulfide cross-linking. This particular configuration was therefore chosen for further biochemical and structural studies.

Variation of the tether length in DNA

To investigate the dependence of the cross-linking reaction on the length of the tether in DNA, two template:primers that have identical sequences (DNA T, Figure 3a), but differ in whether they possess a C2 or C3 tether (refer to Figure 2, n = 2 or 3, respectively), were prepared and employed in parallel kinetic cross-linking analyses with Q258C-RT. As shown in Figure 4a, the template:primer with the C3 thiol tether (bottom panel) undergoes disulfide cross-linking to the cysteine engineered enzyme at a faster rate than that of the corresponding template:primer with the C2 linker (top panel), even though the latter enjoys an entropic advantage, owing to the presence of fewer rotative bonds in the tether. Based on these results, the template:primer with a C3 thiol-bearing tether, together with Q258C-RT, were chosen for high-resolution structural studies [15].

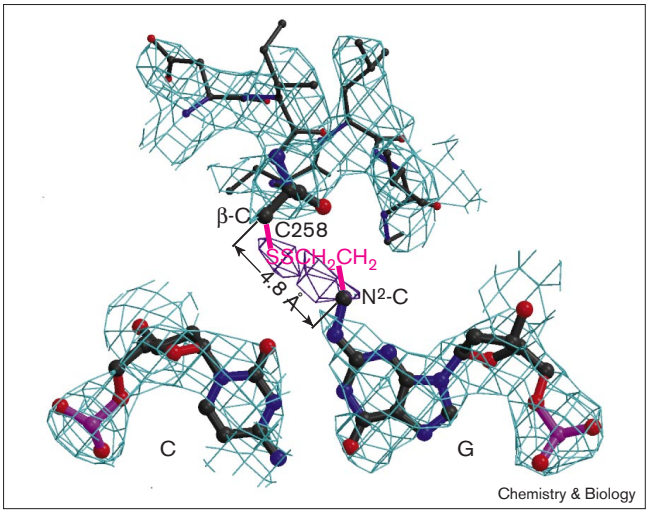
Comparison of disulfide cross-linking to DNA:DNA versus RNA:DNA template:primers

During the viral life cycle, RT uses both RNA and DNA templates to program the synthesis of a complementary DNA copy. It was, therefore, of interest to compare the structure and activity of RT on both forms of nucleic acid template. We examined the kinetics of disulfide cross-linking of Q258C-RT to an RNA versus DNA template. For this purpose, a thiol-bearing tether was linked to the N2-position of the corresponding guanine in the RNA template strand, using chemistry analogous to that employed in the DNA case [24,25]. As shown in Figure 4b, Q258C-RT cross-links to the RNA:DNA substrate (bottom panel) at a much faster rate than to the DNA:DNA substrate (top panel). For example, whereas the cross-linking reaction is more than 50% complete with RNA:DNA after 5 minutes, it takes roughly 2 hours to reach the same level of conversion with DNA:DNA. Following from these preliminary studies, we succeeded in growing diffraction quality crystals of Q258C–RT disulfide cross-linked to an RNA:DNA template:primer (H.H, S.C.H. and G.L.V., unpublished observations).

Stability of disulfide cross-link

A crucial aspect of our intermolecular cross-linking system is the use of disulfide bond formation, a process that is rapidly reversible. We were interested to determine how effective the RT–DNA cross-link is at competing with free thiols in solution under equilibrating conditions. To address this issue, we carried out cross-linking reactions in the presence of increasing concentrations of the potent reducing agent dithiothreitol (DTT), in addition to the usual fixed amount (2 mM) of β-mercaptoethanol. After allowing the reaction to proceed for 2 hours at room temperature, it was quenched by adding a rapid thiol-capping agent. The extent of cross-linking was determined by nonreducing SDS–PAGE analysis (Figure 4c). Remarkably, the disulfide cross-linking of RT to its duplex substrates takes place to an appreciable extent even in the presence of 10 mM DTT, conditions that are generally considered to maintain proteins in the fully reduced state. These findings are all the more striking when taking into account the fact that the 10 mM DTT represents a 1000-fold molar excess over template:primer (10 μM), and a 2500-fold molar excess over RT (4 μM). It is clear, therefore, that the high local concentration of reacting partners afforded by the noncovalent association of RT with the template:primer provides a powerful thermodynamic driving force for the disulfide cross-linking reaction.

Figure 5



$2F_o–F_c$ (cyan) and $F_o–F_c$ (purple) electron density maps at 3.2 Å resolution of the region surrounding the tethered G•C base pair in template:primer and Cys258 residue in helix H of RT. The maps were contoured at 1.2 σ ($2F_o–F_c$) and +3 σ ($F_o–F_c$), respectively. Oxygen atoms, red; nitrogen, blue; phosphorus, purple. The terminal atoms of the disulfide tether (Cys258 β-C and guanine N^2-C) and the distance separating them are shown explicitly; N^2-C was not included explicitly in the structure refinement, but was modeled here by reference to the positions of other atoms in the G base. The figure was prepared using Bobscript, a modification of Molscript [32].

Effect of the disulfide cross-linking on the structure of the RT•DNA duplex

The sensitivity of disulfide cross-link to geometric factors, together with the robustness of the bond thus formed, suggested that the cross-linking itself was comfortably accommodated in the RT•template:primer structure. To gain direct insight into this issue, we examined the cross-linked co-crystal structure carefully in the region around the covalent connection. Despite clear electron density for helix H and the DNA base pairs in the immediate vicinity of the cross-link, density for the disulfide linker itself is absent from $2F_o–F_c$ maps contoured at 1.2 σ (Figure 5, cyan), suggesting that the linker is significantly more mobile that its surroundings. Discontinuous electron density attributable to the disulfide linker could be observed in difference electron density maps ($F_o–F_c$) of the complex (Figure 5, purple). The distance between the terminal carbons of the tether in the crystal structure, ~4.8 Å, is less than the maximum calculated distance of ~7.5 Å for an unstrained tether, suggesting that the tether fits comfortably into the space between RT and DNA. Furthermore, the segment of helix H and the DNA base pairs in the immediate vicinity of the cross-link exhibit normal structural parameters. Taken together, these features strongly support the notion that the cross-link exerts little if any effect on the structure of the RT–DNA–dNTP complex, but merely acts to prevent dissociation of the components from one another.

The co-crystal structure of the RT–DNA–dNTP complex provides a clear rationale for the observed positional specificity of disulfide cross-linking. As shown in Figure 6, the α carbons of positions 258, 262 and 266 lie in nearly the

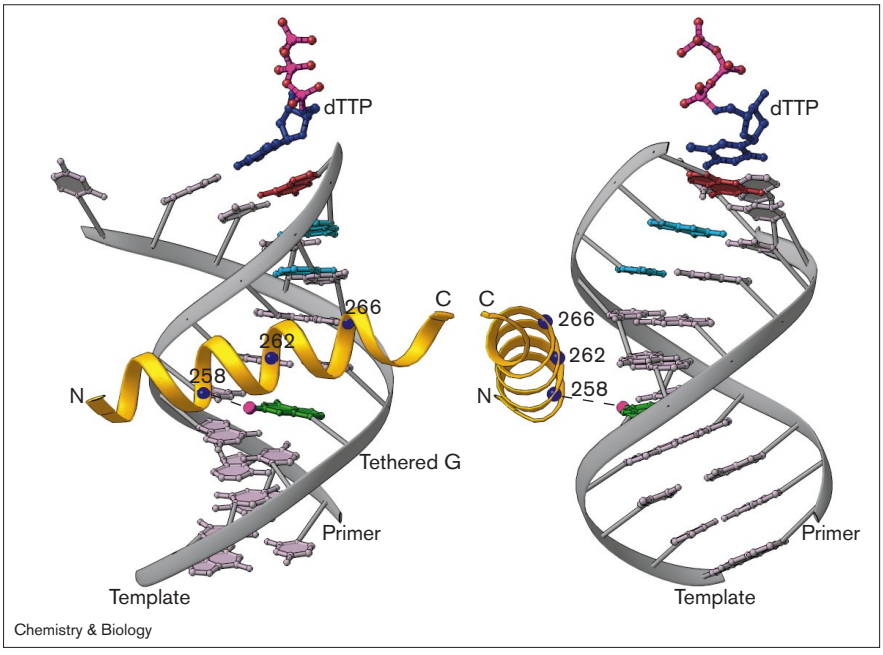
same plane as the 6th, 5th and 4th base pairs from the dTTP:dA pair bound in the enzyme active site. Even though an α helix rises much more steeply than base pairs in A-form DNA (~5.4 Å axial rise per helical turn versus ~3 Å per base pair), the tilt of helix H that permits it to track the minor groove also creates a near-perfect correspondence of successive α-helical turns with successive base pairs in the DNA. Based on the structure, it can be predicted that transposition of the cysteine residue by one turn on helix H should decrease by one base pair the optimal distance from the cross-link to the active site (refer to Figure 3b). Indeed, this is precisely what is observed: transposition of the cysteine residue from position 258→262→266 systematically reduces the cross-link–active-site distance from 6→5→4 base pairs (+3→+2→+1 nucleotides incorporated for optimal cross-linking; refer to Figure 3b). The efficiency and selectivity of cross-linking for each of these three positions are not identical, owing to subtle differences in geometric factors, such as DNA conformation or placement of the cysteine on the face of helix H, which exerts an influence on the cross-linking chemistry.

Discussion

The inability to obtain high quality co-crystals remains a major impediment to many studies of protein–nucleic acid interactions. An especially compelling example of this is HIV-RT, for which the lack of structural information on a catalytically relevant complex impeded understanding of drug resistance mutations and held back efforts to create novel inhibitors. The work reported here began with efforts to discover the reason for the failure of RT to co-crystallize with a template:primer and a nucleotide triphosphate, and then, on the basis of this information, turned to design of a system more predisposed toward crystallization. The present results point to the probable cause of the failure to grow co-crystals by conventional methods: RT exhibits only a 6–10-fold preference for a template:primer containing a polymerizable end over one that is incapable of being polymerized, even in the presence of a cognate nucleotide. Remarkably, no preference whatsoever for substrate versus nonsubstrate template:primer was observed in the absence of nucleotide (this work and [26]). The level of specificity shown by RT is, therefore, too low for the formation of a species sufficiently homogenous for crystallization. We imagine that noncovalent association of RT, template:primer and nucleotide yields complexes that differ in the positioning of the enzyme on the template:primer, and that these various complexes interconvert on the crystallization timescale. This analysis led us to the hypothesis that the RT crystallization problem might be solved by finding a way to funnel the equilibrating complexes into the single catalytically relevant one, which would be rendered incapable of dissociation. We have previously reported the use of disulfide cross-linking to obtain crystals of a ternary catalytic RT complex [15]; here we have explained in greater detail the chemistry of this cross-linking reaction, with an eye toward providing a robust basis for comparison in future studies with other proteins.

The combinatorial cross-linking system analyzed here has allowed us to explore independently the influence of several variables on the rate of the cross-linking reaction: the distance between the cross-link attachment points in the protein and template:primer; the length of the cross-link itself; and the nucleic acid composition (DNA:DNA or RNA:DNA) of the template:primer. As previously observed, the cross-linking reaction exhibits tremendous selectivity for the distance between the enzyme active site and the location of the cross-link. For example, Q258C-RT strongly prefers to cross-link a tethered G residue located on the template strand six base pairs away from the nucleotide bound in the active site (corresponds to +3 in Figure 3b, top left-hand panel). The co-crystal structure of the cross-linked RT catalytic complex clearly shows that in this configuration, the cross-linking partners are nearer to each other than any other register (Figure 6). The structure also provides a clear explanation for the positional dependence of the three cysteine RT mutants, in which the site of cross-linking is scanned along the helix exposed to the minor groove (Figure 6). Helix H is tilted in such a way that successive turns are almost perfectly aligned with successive base pairs in the template:primer. Residue 266 therefore lies one base pair closer to the active site than residue 262, which in turn is one base pair closer than residue 258. Correspondingly, W266C-RT requires the addition of one less nucleotide than G262C-RT for optimal cross-linking, and G262 requires one less than Q258C. Nevertheless, the three mutants differ clearly in the extent of their positional selectivity. Whereas Q258C-RT shows nearly perfect in cross-linking to G(–3), the G262C and W266C mutants are less selective (compare the number of bands along the p66•DNA position in Figure 3b). We attribute these differences in selectivity to two factors: because the period of an α helix is nonintegral (3.6 residues/turn), the sidechains of 258, 262 and 266 are slightly displaced with respect to one another along the face of helix H; and helix H lies over a segment of the template:primer that is in transition from A-form (carboxy-terminal end of helix H) to B-form (amino-terminal end of helix H). The local DNA conformation with respect to the cysteine sidechain is, therefore, subtly but significantly different at positions 258, 262 and 266. The minor cross-link bands represent RT•template:primer complexes in which the cysteine residue is displaced by one base pair from that containing the thiol tether, suggesting there is enough play in the cross-link to reach one base pair away. This conclusion is consistent with the results of the X-ray analysis (Figure 5), which shows that the cross-link itself is conformationally mobile and several angstroms longer than the shortest cross-linking distance.

Figure 6



Ribbon representation of **(a)** front and **(b)** side views of the location of helix H relative to template:primer in the RT–DNA crystal structure. Relevant bases are color-coded as in Figure 3, and the backbone of the template:primer is in gray. α Carbons of the mutated cysteine residues (blue) examined biochemically in helix H (gold) are highlighted to illustrate the relative positions of these amino acids to the surrounding base pairs. N^2 of thiol tethered dG is colored as a magenta sphere to illustrate its location relative to the α carbon of residue 258. Dash denotes the cross-linked positions in the crystallographically determined complex. The figure was prepared using RIBBONS [33].

Many proteins require external thiol to maintain their activity. Can disulfide cross-linking, which usually requires oxidizing conditions, be applied in such cases? We show here that the disulfide cross-linking of RT to template:primer is surprisingly efficient under conditions that are sufficiently reducing to keep most proteins active. This result presumably reflects the fact that the noncovalent binding of RT to DNA causes the reacting partners to be present in a high effective concentration, thereby providing thermodynamic driving force for the overall reaction. In a chemical sense, this is equivalent to disulfide formation during protein folding, where protein–protein interactions drive the formation of thermodynamically stable disulfide bonds. It is also analogous to the high reduction potential of protein thioredoxin [27], wherein two cysteines are closely juxtaposed in a β-turn structure. Another serious concern in any cross-linking study is the trapping of strained species. The ability to do disulfide chemistry under conditions in which disulfides are freely exchanging enables the preferential trapping of a stable complex. The structure of the cross-linked RT•DNA complex bears out this expectation. Specifically, the positional selectivity of the cross-linking reaction corresponds to the product with the most relaxed structure. Furthermore, the tether itself is unstrained and conformationally mobile. The lack of cross-linking-induced structural perturbation is also clearly evident in the X-ray structure. Both the helix H and the DNA near the cross-link attachment points are virtually superimposible on the corresponding elements of the noncatalytic, noncovalent RT•DNA complex.

These studies provide a blueprint for the use of disulfide cross-links to freeze protein–DNA complexes for structural studies. From a practical point of view, use of this method is facilitated by having some notion of amino acid residues that are likely to lie near the DNA. Indeed, for the reasons described above, formation of a stable cross-link itself constitutes strong evidence for the residue in question being located near DNA in the non-cross-linked state. The method can therefore be used to test hypotheses concerning the location of particular residues at the protein–DNA interface, in cases where the body of available structural information is insufficiently predictive. For reasons particular to the case of RT•template:primer complex, we chose to attach the enzyme to a G residue in the minor groove. In other cases, it may be advantageous to link the protein to some other portion of the DNA or RNA, such as a base in the major groove or the backbone. There exist many straightforward methods for incorporating thiols into nearly any position in DNA [20–22,25,28,29], and the required reagents are in many instances commercially available. We recommend the use of a screening approach akin to that employed here, to sample multiple cross-link configurations and to determine experimentally the one that is most stable and most likely to be physiologically relevant.

Significance

Many protein–nucleic acid interacting systems are difficult to study using high-resolution structural methods because they do not form stable homogenous complexes. We have shown that this is the case for HIV-1 reverse

transcriptase, which binds a polymerizable substrate only 6–10-fold more tightly than a nonpolymerizable substrate. Based on these observations, we formulated a strategy to freeze a catalytically relevant reverse transcriptase (RT)–DNA complex through equilibrium disulfide cross-linking. We have analyzed in detail a number of key parameters to understand their influence on the cross-linking reaction: relative positions of the tether in the template:primer and protein, length of the tether, DNA:DNA versus RNA:DNA template:primer and competition by external thiols. These studies provide a blueprint for the application of disulfide cross-linking techniques to other protein–nucleic acid systems.

Materials and methods

Materials

T4 polynucleotide kinase, restriction endonucleases and DNA polymerases were from New England Biolabs. Nucleoside triphosphates were from Sigma. [γ-^{32}P] ATP was from DuPont/New England Nuclear Corp. Ni-NTA resin was from Qiagen. SE-Sepharose resin and Superdex 75 size exclusion column were from Pharmacia. All other reagents were of the highest quality commercially available.

Construction, expression and purification of HIV-1 RT

The coding sequences for both p66 and p51 of RT (amplified by polymerase chain reaction (PCR) from plasmid $pRT_{66/51}$, a generous gift from R.S. Goody [30]), each linked to an independent T7 promoter, were introduced into the plasmid pLM1 [16,17]. Site-directed PCR mutagenesis was used to mutate Glu478 in the RNase H domain of p66 subunit to glutamine, with the aim of eliminating RNase H catalytic activity so that studies on RNA:DNA hybrids can be carried out [31]. Pro1 was also mutated to lysine to increase the expression level of RT in *E. coli*. For purposes of disulfide cross-linking studies, three residues on helix H of p66 – Gln258, Gly262 and Trp266 – were individually mutated to cysteine. In addition to the helix H mutation, each of the variant proteins used in this study contained the Cys280→Ser mutation introduced in prior crystallographic studies on RT [8]. Overall, each of the three variant RT proteins contains four cysteine residues – the one introduced in helix H, the remaining one present in wild-type p66 (C38), and both wild-type cysteine residues in p51 (Cys38 and Cys280). Recombinant RT enzymes were expressed in *E. coli* BL21 (DE3). The enzymes were purified to high homogeneity by a sequence of chromatography steps on Ni-NTA, SE-Sepharose, and Superdex 75. The fractions containing RT from the last step of purification were pooled, exchanged in a buffer containing 50 mM Tris–HCl (pH 7.4), 50 mM NaCl, 1 mM EDTA, 30% glycerol, flash frozen in liquid nitrogen, and stored at −78°C freezer. The protein concentration of purified RT was measured by UV absorbance at 280 nm using an extinction coefficient $\varepsilon_{280} = 266{,}800\ M^{-1}\ cm^{-1}$.

Synthetic oligonucleotides

DNA and RNA oligonucleotides were synthesized on an Applied Biosystems 392 DNA/RNA synthesizer. For modified oligonucleotides, the thiol tether was introduced into the template:primer through post-synthetic modification using the convertible nucleoside approach [25]. Cystamine deprotection gave the C2 tether protected as the mixed 2-aminoethyl disulfide, and 3-aminopropane disulfide deprotection gave the C3 tether. Both modified and unmodified oligonucleotides were purified using denaturing polyacrylamide gel electrophoresis (15% acrylamide, 8 M urea). Concentrations of the oligonucleotides were estimated by UV at 260 nm. For EMSAs, the template strand of DNA substrate was labeled at the 5′ end using T4 polynucleotide kinase and [γ-^{32}P] ATP according to the directions of the manufacturer. The duplex was formed by mixing complementary strands in a molar ratio of 1:1.05 (the radiolabeled or modified strand is the limiting one) or 1:1 (competitor DNA in EMSAs) in a buffer containing 10 mM Tris–HCl (pH 7.4), 10 mM NaCl. The mixed sample was heated to 70°C for 5 min, then cooled slowly to room temperature.

Electrophoretic mobility shift assay

DNA substrate was mixed with RT in Reaction buffer containing 50 mM Tris–HCl (pH 7.4), 25 mM NaCl, 25 mM KCl, 5 mM $MgCl_2$, 5% glycerol, with or without ddCTP. After 15 min, the competitor DNA was added to the mixture. The resulting sample was further incubated for an additional 50 min. The mixture was loaded onto a nondenaturing 5% polyacrylamide gel (20 × 16 cm). The gel was run at room temperature at 150 V until the bromophenol blue dye reached 4.2 cm from the bottom of the gel. The gel was dried and exposed to X-ray film for autoradiography.

Cross-linking reaction and analysis

For a typical cross-linking reaction, the mutant RT enzyme (4 μM) was mixed together with the thiol-tethered template:primer (10 μM) and a dNTP/ddNTP cocktail (100 μM each, refer to Figure 3a) in Reaction buffer (see above) and 2 mM β-mercaptoethanol, which was added to increase the specificity of cross-linking. Following incubation at 25°C, the cross-linking reaction was quenched by adding a thiol-capping reagent, methyl methanethiolsulfonate (20 mM). In experiments analyzing disulfide stability, measured amounts of DTT were added to the reaction mixture, and the concentration of thiol-capping reagent was elevated to 100 mM to ensure complete capping of all the free thiol groups in the reaction mixture. SDS loading buffer was added to the quenched reaction mixture. The sample was heated at 90°C for 5 min, and analyzed by SDS–PAGE (15% acryamide) under nonreducing conditions. The gel was stained in Coomassie blue solution overnight.

Purification of RT•template:primer complex

The cross-linking reaction was carried out on a preparative scale. Q258C-RT (133 μM in 75 μl storage buffer) was mixed together with an equimolar amount of the thiol-tethered template:primer in 50 μl TE buffer. Rather than elongate the primer by carrying out three cycles of polymerization as shown in Figure 3c, we included the first two bases in the primer strand, and carried out only one cycle of polymerization, adding a ddA residue to the 3′-end of the primer strand, (dTTP is included to bind noncovalently in the enzyme active site). The final product is identical to that produced by three cycles of polymerization (Figure 3c). Thus, to the Q258C-RT and template:primer mixture was added 50 μl of a nucleotide cocktail containing 2 mM of each ddATP and dTTP, 100 μl 5× Reaction buffer plus 2 mM β-mercaptoethanol. The final volume of the reaction was adjusted to 500 μl by the addition of TE buffer. Following 1.5 days of incubation at 25°C, the reaction mixture was diluted to 1.5 ml with elution buffer A (10 mM NaCl, 20 mM Tris.HCl, pH 8.0), and loaded into an anion-exchange column (Pharmacia, 0.5 × 5 cm). The covalent complex was eluted using NaCl gradient (10 mM to 1000 mM) in a buffer containing 20 mM Tris.HCl (pH 8.0) at the concentration of NaCl 350 mM. The sample was collected, exchanged twice with a exchanging buffer (10 mM MES, pH 6.5, 2 mM $MgCl_2$, 200 mM NaCl, 0.05 mM EDTA, 0.5 mM dTTP), and concentrated. More dTTP was added to the concentrated sample so that the final concentration of dTTP and complex is 2 mM and 8 mg/ml, respectively. The complex was used directly in crystallization trials, the details of which have been reported elsewhere [15].

Acknowledgements

We thank R. Chopra, B. Harris, D. Erlanson, R. Butcher and other members of Verdine and Harrison–Wiley research groups. The work was supported by NIH grants GM-44853 (to G.L.V.) and P01 GM-39589 (to S.C.H. and G.L.V.). H.H. acknowledges a NIH postdoctoral fellowship (GM-18621). S.C.H. is an investigator in the Howard Hughes Medical Institute.

References

1. Goff, S.P. (1990). Retroviral reverse transcriptase: synthesis, structure and function. *J. Acquir. Immune. Defic. Syndr.* **3**, 817-831.
2. Arts, E.J. & Le Grice, S.F. (1998). Interaction of retroviral reverse transcriptase with template-primer duplexes during replication. *Prog. Nucleic Acid. Res. Mol. Biol.* **58**, 339-393.

3. Mitsuya, H., Yarchoan, R. & Broder, S. (1990). Molecular targets for AIDS therapy. *Science* **249**, 1533-1544.
4. Havlir, D.V. & Richman, D.D. (1996). Viral dynamics of HIV: implications for drug development and therapeutic strategies. *Ann. Intern. Med.* **124**, 984-994.
5. Oxford, J.S., al-Jabri, A.A., Stein, C.A. & Levantis, P. (1996). Analysis of resistance mutants of viral polymerases. *Methods Enzymol.* **275**, 555-600.
6. Moyle, G.J. (1997). Current knowledge of HIV-1 reverse transcriptase mutations selected during nucleoside analogue therapy: the potential to use resistance data to guide clinical decisions. *J. Antimicrob. Chemother.* **40**, 765-777.
7. Kohlstaedt, L.A., Wang, J., Friedman, J.M., Rice, P.A. & Steitz, T.A. (1992). Crystal structure at 3.5 Å resolution of HIV-1 reverse transcriptase complexed with an inhibitor. *Science* **256**, 1783-1790.
8. Jacobo-Molina, A., *et al*., & Arnold, E. (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.
9. Jager, J., Smerdon, S.J., Wang, J., Boisvert, D.C. & Steitz, T.A. (1994). Comparison of three different crystal forms shows HIV-1 reverse transcriptase displays an internal swivel motion. *Structure* **2**, 869-876.
10. Ren, J., *et al*., & Stammers, D. (1995). High resolution structures of HIV-1 RT from four RT-inhibitor complexes. *Nat. Struct. Biol.* **2**, 293-302.
11. Ding, J., *et al*., & Arnold, E. (1995). Structure of HIV-1 reverse transcriptase in a complex with the non- nucleoside inhibitor alpha-APA R 95845 at 2.8 Å resolution. *Structure* **3**, 365-379.
12. Rodgers, D.W., *et al*., & 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.
13. Das, K., *et al*., & Arnold, E. (1996). Crystal structures of 8-Cl and 9-Cl TIBO complexed with wild-type HIV-1 RT and 8-Cl TIBO complexed with the Tyr181Cys HIV-1 RT drug-resistant mutant. *J. Mol. Biol.* **264**, 1085-1100.
14. Hsiou, Y., Ding, J., Das, K., Clark, A.D., Jr., Hughes, S.H. & Arnold, E. (1996). Structure of unliganded HIV-1 reverse transcriptase at 2.7 Å resolution: implications of conformational changes for polymerization and inhibition mechanisms. *Structure* **4**, 853-860.
15. Huang, H., Chopra, R., Verdine, G.L. & Harrison, S.C. (1998). Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. *Science* **282**, 1669-1675.
16. MacFerrin, K.D., Terranova, M.P., Schreiber, S.L. & Verdine, G.L. (1990). Overproduction and dissection of proteins by the expression-cassette polymerase chain reaction. *Proc. Natl Acad. Sci. USA* **87**, 1937-1941.
17. MacFerrin, K.D., Chen, L., Terranova, M.P., Schreiber, S.L. & Verdine, G.L. (1993). Overproduction of proteins using the expression-cassette polymerase chain reaction. *Methods Enzymol.* **217**, 79-102.
18. Houk, J., Singh, R. & Whitesides, G.M. (1987). Measurement of thiol-disulfide interchange reactions and thiol pK_a values. *Methods Enzymol.* **143**, 129-140.
19. Creighton, T.E., Zapun, A. & Darby, N.J. (1995). Mechanisms and catalysts of disulfide bond formation in proteins. *Trends Biotechnol.* **13**, 18-23.
20. MacMillan, A.M. & Verdine, G.L. (1990). Synthesis of functionally tethered oligonucleotides by the convertible nucleoside approach. *J. Org. Chem.* **55**, 5931-5933.
21. MacMillan, A.M. & Verdine, G.L. (1991). Engineering tethered DNA molecules by the convertible nucleoside approach. *Tetrahedron* **47**, 2603-2616.
22. Erlanson, D.A., Chen, L. & Verdine, G.L. (1993). DNA methylation through a locally unpaired intermediate. *J. Am. Chem. Soc.* **115**, 12583-12584.
23. Bebenek, K., *et al*., & Kunkel, T.A. (1997). A minor groove binding track in reverse transcriptase. *Nat. Struct. Biol.* **4**, 194-197.
24. Allerson, C.R. & Verdine, G.L. (1995). Synthesis and biochemical evaluation of RNA containing an intrahelical disulfide crosslink. *Chem. Biol.* **2**, 667-675.
25. Allerson, C.R., Chen, S.L. & Verdine, G.L. (1997). A chemical method for site-specific modification of RNA: the convertible nucleoside approach. *J. Am. Chem. Soc.* **119**, 7423-7433.
26. Tong, W., Lu, C.D., Sharma, S.K., Matsuura, S., So, A.G. & Scott, W.A. (1997). Nucleotide-induced stable complex formation by HIV-1 reverse transcriptase. *Biochemistry* **36**, 5749-5757.
27. Holmgren, A. (1995). Thioredoxin mechanism and structure: conformational changes on oxidation of the active-site sulfhydryls to a disulfide. *Structure* **3**, 239-243.
28. Ferentz, A.E., Keating, T.A. & Verdine, G.L. (1993). Synthesis and characterization of disulfide cross-linked oligonucleotides. *J. Am. Chem. Soc.* **115**, 9006-9014.
29. Uhlmann, E. & Peyman, A. (1990). Antisense oligonucleotides: a new therapeutic principle. *Chem. Rev.* **90**, 543-584.
30. Muller, B., Restle, T., Weiss, S., Gautel, M., Sczakiel, G. & Goody, R.S. (1989). Co-expression of the subunits of the heterodimer of HIV-1 reverse transcriptase in *Escherichia coli*. *J. Biol. Chem.* **264**, 13975-13978.
31. Schatz, O., Cromme, F.V., Gruninger-Leitch, F. & Le Grice, S.F. (1989). Point mutations in conserved amino acid residues within the carboxy-terminal domain of HIV-1 reverse transcriptase specifically repress RNase H function. *FEBS Lett.* **257**, 311-314.
32. Kraulis, P.J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946-950.
33. Carson, M. (1987). Ribbon models of macromolecules. *J. Mol. Graphics* **5**, 103-106.