Proc. Natl. Acad. Sci. USA
Vol. 91, pp. 7242–7246, July 1994
Biochemistry

Structural basis of asymmetry in the human immunodeficiency virus type 1 reverse transcriptase heterodimer

(DNA polymerase/protein folding/protein structure/acquired immunodeficiency syndrome/drug design)

J. WANG*, S. J. SMERDON*, J. JÄGER*, L. A. KOHLSTAEDT*†, P. A. RICE*‡, J. M. FRIEDMAN*§, AND T. A. STEITZ*¶||

Departments of *Molecular Biophysics and Biochemistry, and ¶Chemistry, Howard Hughes Medical Institute, Yale University New Haven, CT 06520-8114

Contributed by T. A. Steitz, April 11, 1994

ABSTRACT **The reverse transcriptase from human immunodeficiency virus type 1 is a heterodimer consisting of one 66-kDa and one 51-kDa subunit. The p66 subunit contains both a polymerase and an RNase H domain; proteolytic cleavage of p66 removes the RNase H domain to yield the p51 subunit. Although the polymerase domain of p66 folds into an open, extended structure containing a large active-site cleft, that of p51 is closed and compact. The connection subdomain, which lies between the polymerase and RNase H active sites in p66, plays a central role in the formation of the reverse transcriptase heterodimer. Extensive and very different intra- and intersubunit contacts are made by the connection subdomains of each of the subunits. Together, contacts between the two connection domains constitute approximately one-third of the total contacts between subunits of the heterodimer. Conversion of an open p66 polymerase domain structure to a closed p51-like structure results in a reduction in solvent-accessible surface area by 1600 Å^2 and the burying of an extensive hydrophobic surface. Thus, the monomeric forms of both p66 and p51 are proposed to have the same closed structure as seen in the p51 subunit of the heterodimer. The free energy required to convert p66 from a closed p51-like structure to the observed open p66 polymerase domain structure is generated by the burying of a large, predominantly hydrophobic surface area upon formation of the heterodimer. It is likely that the only kind of dimer that can form is an asymmetric one like that seen in the heterodimer structure, since one dimer interaction surface exists only in p51 and the other only in p66. We suggest that both p51 and p66 form asymmetric homodimers that are assembled from one subunit that has assumed the open conformation and one that has the closed structure.**

Perhaps the most surprising aspect of the structure of human immunodeficiency virus type 1 (HIV-1) reverse transcriptase (RT) is the observation that the polymerase domain assumes a different structure in the two subunits in spite of having the same polypeptide chain sequence (1). HIV-1 RT consists of one 66-kDa polypeptide chain (p66) consisting of a polymerase domain and an RNase H domain and one 51-kDa polypeptide chain (p51) containing only the polymerase domain. These two subunits interact asymmetrically to generate only one polymerase cleft that binds one primer–template, one dNTP, one noncompetitive inhibitor, and one tRNA (2–5). The polymerase domains of p51 and p66 differ by having an alternative arrangement of four subdomains (1). The three subdomains that form the large polymerase active-site cleft in p66 are called "fingers," "thumb," and "palm" by analogy of this polymerase structure to that of a right hand. The fourth subdomain is called "connection" because it lies between the polymerase and RNase H active sites in p66. Although the heterodimer is the most stable dimer, with an equilibrium dissociation constant (K_d) of $\approx 1 \times 10^{-9}$ M (6, 7), both p66 and p51 homodimers have been observed *in vitro* but are much less tightly associated (7). The major question addressed here is how a single amino acid sequence can form two quite different structures and result in such an asymmetric subunit interaction. Furthermore, and in the light of the structural observations, it is of interest to consider the likely conformation of p51 or p66 monomers as well as the possible structures of homodimers of these subunits.

The crystal structure of the HIV-1 RT heterodimer complexed with a noncompetitive inhibitor, Nevirapine, was initially derived from a 3.5-Å resolution electron density map (1) and has now been partially refined at 2.9-Å resolution (8, 9). The structure of HIV-1 RT complexed with the Fab portion of a monoclonal antibody and duplex DNA determined at 3-Å resolution shows the same structure for the RT and provides experimental evidence for the primer–template location (10).

RESULTS AND DISCUSSION

The Asymmetric Dimer Structure. The four polymerase subdomains of HIV-1 RT have very different relative orientations in the two subunits of the RT heterodimer (Fig. 1). The p51 subunit has a compact structure that we can refer to as "closed," while the p66 subunit has a more extended structure and a large cleft that can be referred to as "open." With the connection domains oriented identically, the different sets of interactions made by the fingers, palm, and thumb subdomains of each of the two subunits are clearly seen (Fig. 1). Changes in the contacts between the connection and the fingers subdomains are more modest.

Interactions between the two subunits are completely asymmetric in that the subunit interface on p51 involves different amino acid residues than the interface on p66 (Fig. 2 and below). Contacts between connection subdomains form the only interactions between equivalent subdomains from each subunit. However, even in this case, a different set of surfaces is utilized (Figs. 1–3). Therefore, contacts between the two connection subdomains are intrinsically asymmetric and many equivalent residues make different protein–protein interactions. Those side chains that pack in alternative ways at different interfaces in the two subunits are predominately hydrophobic.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: RT, reverse transcriptase; HIV-1, human immunodeficiency virus type 1.

†Present address: Department of Chemistry, University of California, Santa Barbara, CA 93106.
‡Present address: National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Building 5, Room 237, Bethesda, MD 20892.
§Present address: Department of Chemistry, University of Houston, Houston, TX 77204-5641.
||To whom reprint requests should be addressed.

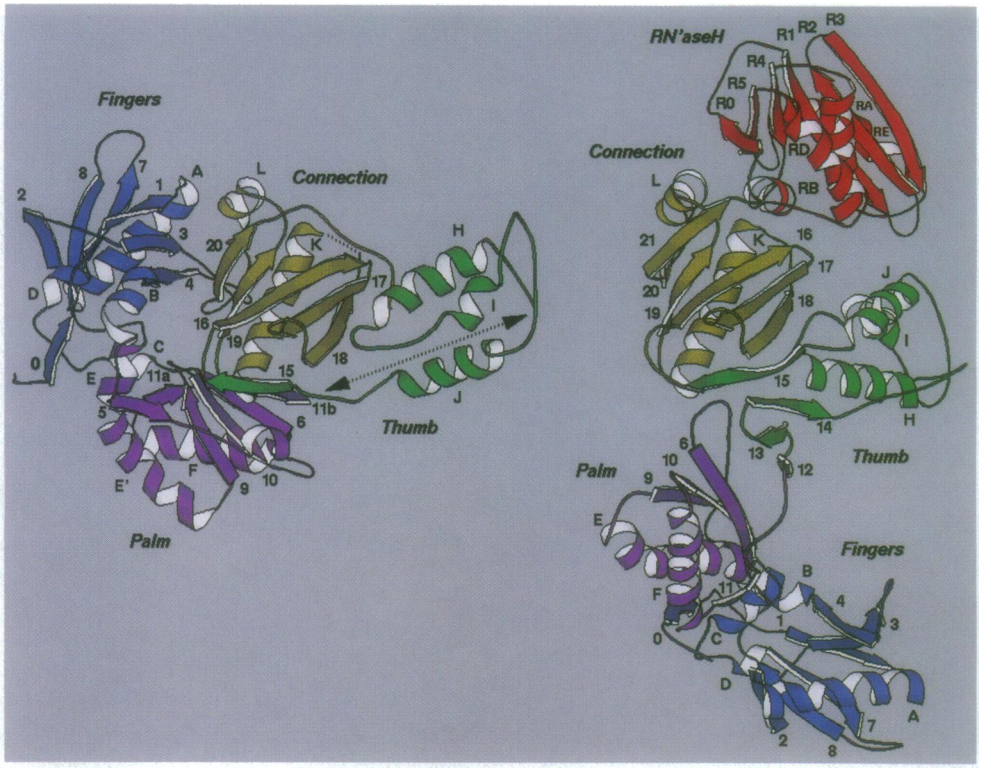

FIG. 1. Schematic drawing (MOLSCRIPT; ref. 11) of the polypeptide backbone of the p51 (*Left*) and p66 (*Right*) subunits of HIV-1 RT with the connection subdomain of the two subunits oriented identically. The α-helices are lettered sequentially from the N terminus and the β-strands are numbered. RNase H secondary structure is labeled as described by Davies *et al.* (12). The most dramatic difference in the structures of these two subunits is the positions of the fingers and palm subdomains relative to that of the connection subdomain. In p66 the connection subdomain makes no contact with the p66 fingers and almost none with the palm, whereas the p51 connection subdomain has extensive contacts with both. Part of the connection subdomain surface formed by helices A and K and β-strands 16, 19, and 20 interacts with the fingers and palm in p51 but makes intersubunit contacts in the p66 subunit. The correspondence between the secondary structure elements and the amino acid sequence numbers based on the partially refined heterodimer model (9) is given in Table 1.

The overall structure of the connection subdomain is not identical in the two subunits, presumably to accommodate the different environments in which it is located. The major alteration includes β-strand 20 and α-helix L, which are differently positioned. The loop between these two elements in p66 interacts with the connection subdomain of p51, while the equivalent loop in p51 is in contact with the p51 fingers and palm subdomains. Additionally, β-strand 21, which leads to the RNase H domain in p66, shows a different conformation in p51, where it folds back to form part of the surface involved in binding the primer–template (Fig. 4; refs. 1, 10, and 15).

The palm of p51 makes very limited interactions with p66, and the fingers and thumb of p66 are not involved in dimer formation at all. In all probability, the extensive contacts observed between the thumb of p51 and the RNase H domain of p66 (about one-third of the surface area buried upon dimer formation) contribute substantially to the greater stability of the heterodimer as compared with the p51 homodimer which has been observed experimentally (6, 7).

Solvent-Accessible Surface Area Favors p51 Conformation. To assess factors that stabilize these two very different polymerase domain structures, we calculated the solvent-accessible surface area (16) of the polymerase domains of p66 and p51 (Table 2). As expected from its more compact structure, the solvent-accessible surface of the p51 polymerase domain is ≈1600 Å^2 less than that of the p66 polymerase domain. Upon formation of the heterodimer from p51 and p66, 4560 Å^2 becomes buried and inaccessible to solvent.

The major factor leading to the compactness of protein structures is the energetic benefit of reducing the surface area

Table 1. Correspondence between secondary structural elements and amino acid sequence numbers

	p66	p51		p66	p51		p66	p51		p66 (RNase H)
β1	19–23	20–24	β8	140–146	141–146	αJ	298–310	298–310	β_R1	438–447
αA	27–44	28–44	αE	154–174	154–175	β15	315–319	318–322	β_R2	452–459
β2	47–51	49–52	β9	178–183	178–183	β16	326–331	325–333	β_R3	462–470
β3	57–63	58–64	β10	186–191	186–191	β17	336–342	336–342	α_RA	474–488
β4	71–76	71–76	αF	195–212	198–212	β18	349–355	349–355	β_R4	492–497
αB	78–85	78–85	β11	214–218	215–217	αK	364–383	364–383	α_RB	500–508
β5	93–97	93–97	β12	226–230	—*	β19	387–391	387–391	α_RD	516–527
β6	104–112	105–110	β13	232–235	—*	αL	395–404	395–401	β_R5	530–536
αC	113–118	112–116	β14	238–242	—*	β20	408–412	—‡	α_RE	544–555
αD	124–128	123–128	αH†	253–268	254–270	β21	413–418	412–418		
β7	128–133	129–132	αI	276–281	277–282					

*This region is disordered in p51 and has been deleted from the model.
†In our current model, β20 does not exist in p51.
‡Helix G, tentatively assigned in the unrefined model (1), has now been reassigned as extended structure.

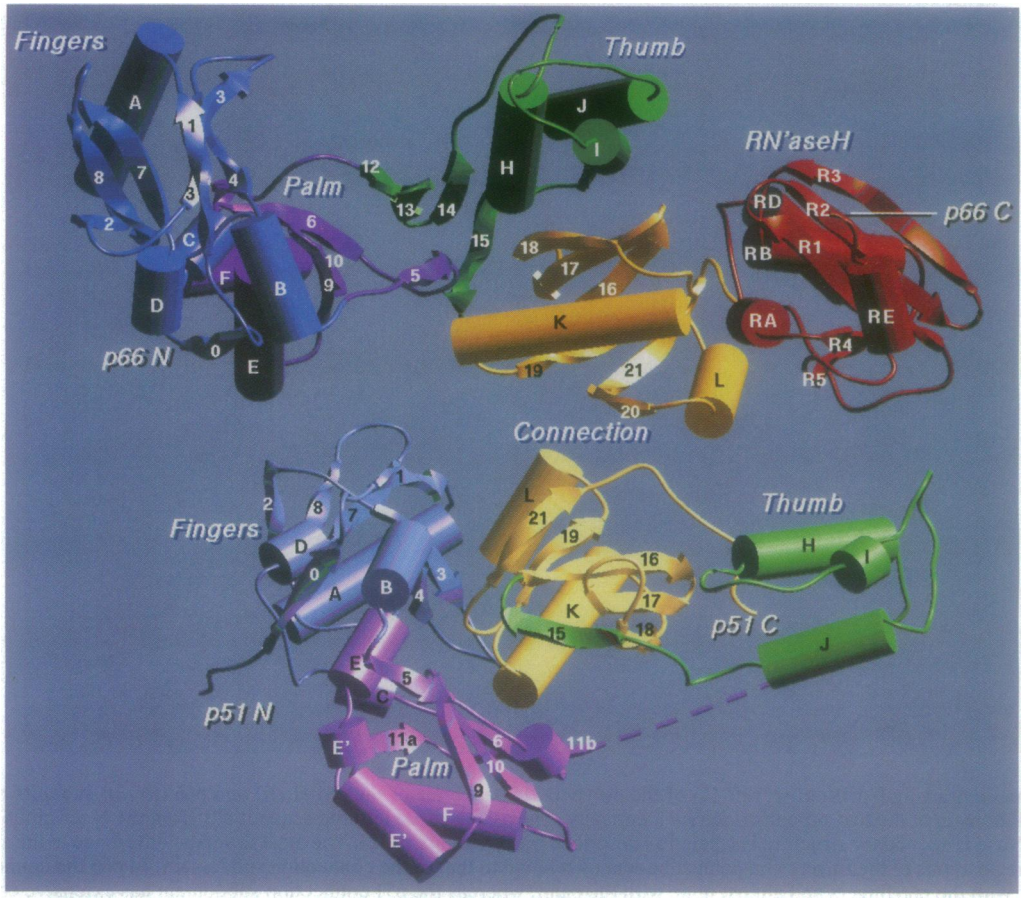

FIG. 2. Schematic drawing of the polypeptide backbone of the RT heterodimer, created with the program RIBBONS (13). α-Helices and β-strands are represented by tubes and arrows, respectively. Subdomains are named and colored as in Fig. 1. The p66 (*Upper*) and p51 (*Lower*) subunits are pulled apart in the vertical direction to make the interaction surfaces clear.

in contact with solvent, particularly if the surface is hydrophobic (17). Not only is the surface area of the p51 polymerase domain smaller than that of the p66 polymerase domain, but also the atoms that are buried in the structural transition from the p66 to the p51 conformation involve mostly aromatic residues and are thus virtually all hydrophobic in nature.

The size of the surface area that becomes buried when the open polymerase domain structure of p66 changes to the closed structure of p51 can be converted into the expected contribution of the hydrophobic effect to the stabilization of the closed structure. Chothia (18) showed a linear correlation between the surface areas of hydrophobic side chains and their free energy of transfer from a polar to a nonpolar solvent. Using the conversion factor of 20 cal/Å^2 derived from this correlation (17, 18), we calculate that the hydrophobic effect which stabilizes the closed p51 conformation as compared with the open p66 polymerase structure could be as much as 32 kcal/mol. Other factors such as the increased interdomain flexibility of the extended and open p66 polymerase domain conformation will doubtless be more favorable for the open structure and thereby reduce the energy difference between the open and closed conformations. Nevertheless, the more favorable conformational entropy of the open p66 is likely to be energetically smaller than the unfavorable hydrophobic effect of exposing a large, nonpolar surface to water. We conclude that both p66 and p51, when they exist as monomers in solution, will be in the closed p51 conformation as observed in the heterodimer. Furthermore, since the p51 conformation does not exhibit an active-site polymerase cleft in which the trio of catalytic carboxyl groups is exposed, the monomeric state of this enzyme would be inactive except under conditions where primer–template binding promotes formation of an "open" p66-like conformation.

Formation of the Asymmetric Dimer Structure. If the most stable monomeric conformation of the polymerase domain is the closed p51 form, then how is the open p66 conformation generated? Since formation of the heterodimer buries more than twice the surface area that is exposed by converting one subunit from the closed to the open conformation, it appears that formation of the dimer can generate the free energy required to stabilize the open p66 conformation. That is, converting monomeric closed p66 and closed p51 subunits to heterodimers of one open p66 and one closed p51 subunit results in the net burying of 2936 Å per dimer, which could be as much as 58 kcal/mol if hydrophobic effects were the sole factors. Overcoming the entropy of immobilization that accompanies dimerization may cost 20–30 kcal/mol (19–21); furthermore, it is likely that the five subdomains of the open p66 monomer exhibit substantial flexibility with respect to each other and that dimer formation would reduce this conformational variation. In fact, significant interdomain flexibility remains in the heterodimer as seen by comparing structures in different crystal packing environments (J.J., S.J.S., J.W., and T.A.S., unpublished work). The experimentally determined equilibrium constant for heterodimer dissociation of $\approx 1 \times 10^{-9}$ M, which is equivalent to a ΔG_{diss} of 12.3 kcal/mol (6, 7), requires that some or all of these entropic factors must be significant.

Formation of homodimers of either p51 or p66 with both subunits in either the open or the closed form is not possible using the same subunit interaction surfaces observed in the heterodimer. As can easily be appreciated by inspection of Fig. 5, the interaction surface that contacts p66 exists only in

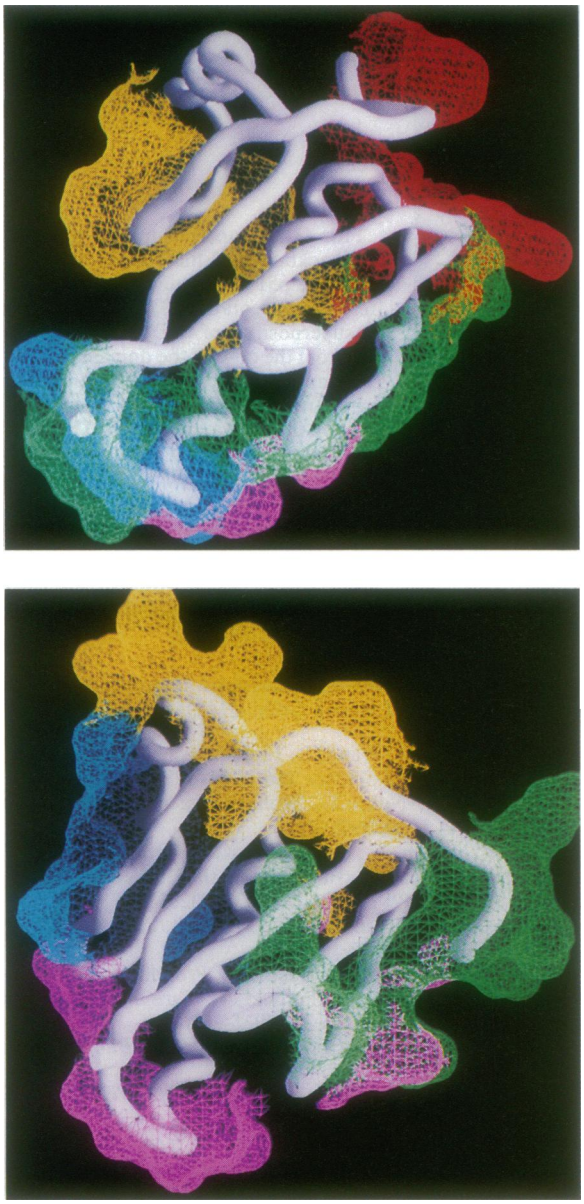

FIG. 3. A GRASP (14) representation of the van der Waals surfaces of the connection subdomains of p51 and p66 (oriented identically as in Fig. 1) showing, by color coding, the subdomains (from either subunit) with which they interact according to the coloring scheme of Figs. 1 and 2. Many regions interact with different subdomains in the two subunits, requiring side chains to accommodate to alternative sets of contacts. (*Upper*) p66 connection subdomain. Aside from surfaces contacting RNase H (red) and p66 thumb (green), all of its contacts are with p51 fingers, palm, and connection. (*Lower*) p51 connection. In this case, all contacts, aside from those with the p66 connection in yellow, are with subdomains of p51 (fingers, blue; thumb, green; palm, magenta). The change in contacts is accompanied by a change in the position of the C terminus relative to the equivalent residues in p66.

p51. Similarly, it is only in the p66 subunit that residues which form the interface with p51 are situated to form a contiguous surface. Those residues in p66 that contact p51 are found to be mostly buried in the p51 subunit and are unavailable to participate in homodimer formation. Likewise, those residues in the p51 subunit that contact p66 lie in scattered locations in the p66 subunit. Thus, only the asymmetric dimer observed in the crystal structure will occur, and higher aggregation states cannot form.

It is, however, possible to construct homodimers of either p66 or p51 subunits by making the assumptions that the polymerase domain of one subunit folds into an open conformation while the other remains in a closed conformation and, further, that these subunits in the open and closed conformation interact as they do in the heterodimer. In support of this notion, both homodimers of p66 and p51 have been shown to possess significant RT activity (22). Additionally, Restle *et al.* (6) have identified a panel of monoclonal antibodies which recognize a variety of epitopes clustered in three regions of the RT sequence and which have a variety of effects on both polymerization/RNase H activity and dimerization. In particular, monoclonal antibody 9, whose epitope was mapped to a region comprising amino acids 528–560 in the RNase H domain, was demonstrated to bind more tightly to p66 monomers than p66 homodimers. No binding to the heterodimer was detected. These observations are consistent with a model in which the epitope is fully accessible in the monomer and buried in the heterodimer as is observed in the crystal structure. Thus, weak binding to the p66 homodimer may be attributable to residual recognition of a partially unfolded RNase H domain of the "closed" p66 subunit (see below).

Table 2. Solvent-accessible surface area of the polymerase (pol) domains of p66 and p51

Surface	Total, Å²	% total	% nonpolar (atoms)
All atoms	46,007		62
Surface area buried			
p66 pol (open), p51 (closed)	1,623		98
p51/p66	4,559	100	65
p51/RNase H	1,579	35	67
p51/p66 connection	1,820	40	67
p51/p66 fingers and palm	1,233	27	57
p51 connection/p66 connection	1,267	28	67
p51/p66 thumb	0	0	—

Solvent-accessible surfaces were calculated with the algorithm of Lee and Richards (16). Coordinates of the RT heterodimer used in these calculations were those resulting from a partial refinement of a model against data to 2.9-Å resolution (8, 9). Currently, the crystallographic *R* factor of this model is 0.26 with rms bond deviations of 0.016 Å. The total solvent-accessible surface of the p66/p51 heterodimer was determined by rolling a 1.4-Å-radius probe over the van der Waals surface of the entire structure. The reduction in accessible surface area produced upon dimerization is the difference between the total surface area of the dimer and the summed surface areas of the two isolated subunits.

In the case of a p51 homodimer found in this asymmetric conformation, the surface area buried upon dimer formation is 2980 Å², which is 1579 Å² less than the surface buried upon formation of the heterodimer. This discrepancy results from the absence of the interaction surface between the p51 thumb and the RNase H domain. Formation of an asymmetric p66 homodimer would be expected to sequester as large a surface area as the formation of the heterodimer. However, β-strand 21 of the connection domain and β-strands R1 and R2 of the RNase H domain of the "closed" p66 subunit overlap with β-strands 19 and 20 of the connection domain from the other subunit if its position relative to the connection domain is maintained as that of p66 in the context of the heterodimer (data not shown). This steric overlap could be relieved through a straightforward rearrangement of the position of the RNase H domain. Observed differences in conformation between p66 and p51 in the region of residues 400–427 indicate that this is feasible (Fig. 4). Alternatively, the clash could be relieved by partially unraveling the RNase H domain. The location of the primary site of cleavage by the HIV protease between residues 440 and 441 lies within the core of RNase H and is relatively inaccessible in the crystal structures of both the heterodimer and the isolated domain (1, 10, 12). Such a sterically induced unfolding event may be significant in the process of formation of p51 from p66.

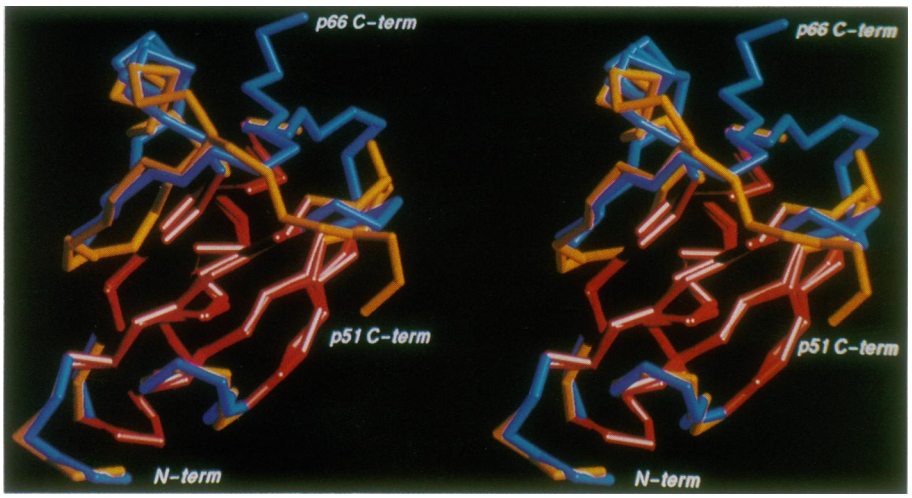

FIG. 4. A superposition of the α-carbon backbones of the connection subdomains of p51 (gold) and p66 (blue) in a stereo representation. The rms difference for superposition of 46 C^{α} atoms (shown in red) from the core of the domain is 0.69 Å. The rms difference for all C^{α} atoms in this region (residues 325–391) following superposition in this way is 1.65 Å. The largest differences in the two structures involve the C-terminal 7 residues.

A combination of these arguments can, in principle, explain the experimentally determined order of dissociation constants of RT hetero- and homodimers (6). The dissociation constant for the heterodimer at low ionic strength and neutral pH is $\approx 1 \times 10^{-9}$ M, whereas that for the p66 homodimer is 10^{-6} M and that for the p51 homodimer is 10^{-5} M. The reduced stability of the p66 homodimer as compared with the heterodimer may result from the use of dimerization free energy to destabilize and unravel the structure of RNase H. The p51 homodimer is less stable than the heterodimer because the interaction between the p51 thumb and RNase H is absent.

The necessity for rearrangement of the subdomains from the closed p51 conformation to the open p66 conformation as a prerequisite for heterodimer formation may afford another opportunity for drug intervention. If we are correct in our prediction that the open polymerase catalytic cleft observed in p66 occurs only in the heterodimer, then any molecule that interferes with the refolding of the p66 subunit or the formation of the heterodimer (6) will yield an inactive p51 monomer-like conformation.

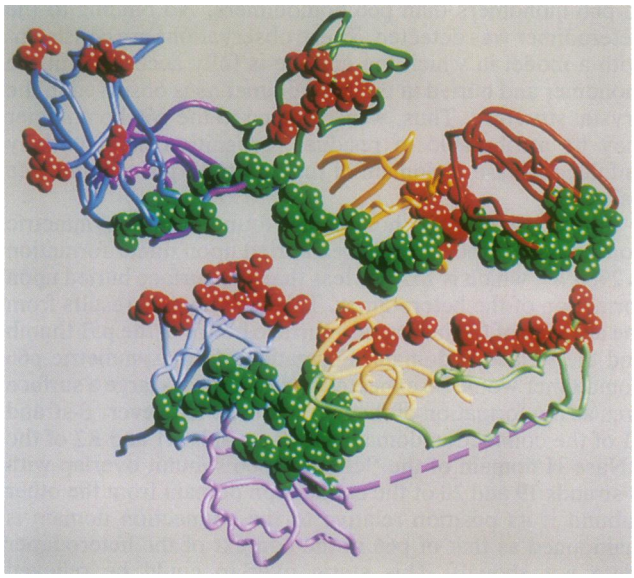

FIG. 5. A backbone representation of the HIV-1 RT heterodimer showing the surface residues (in red, space-filling representation) of p51 (*Lower*) that interact with p66 (*Upper*) and the residues (green, space-filling) of p66 that interact with p51, made using the program RIBBONS (13). The locations of the residues that form the subunit interaction surfaces of the heterodimer are also shown in the other subunit where they do not, in fact, form an interaction surface.

We thank F. M. Richards for discussions. Support was provided by National Institutes of Health Grant GM39546.

1. Kohlstaedt, L. A., Wang, J., Friedman, J. M., Rice, P. A. & Steitz, T. A. (1992) *Science* **256**, 1783–1790.
2. LeGrice, S. F., Naas, T., Wohlgensinger, B. & Schatz, O. (1991) *EMBO J.* **10**, 3905–3911.
3. Barat, C., Lullien, V., Schatz, O., Keith, G., Nugeyre, M. T., Grunigner-Leitich, F., Barre-Sinoussi, F., LeGrice, S. F. & Darlix, J. L. (1989) *EMBO J.* **8**, 3279–3285.
4. Hostomsky, Z., Hostomska, Z., Fu, T.-B. & Taylor, J. (1992) *J. Virol.* **66**, 3179–3182.
5. Wu, J. C., Warren, T. C., Adams, J., Proudfoot, J., Skiles, J., Raghavan, P., Perry, C., Potocki, I., Farina, P. R. & Grob, P. M. (1991) *Biochemistry* **30**, 2022–2026.
6. Restle, T., Muller, B. & Goody, R. S. (1990) *J. Biol. Chem.* **265**, 8986–8988.
7. Divita, G., Restle, T. & Goody, R. S. (1993) *FEBS Lett.* **324**, 153–158.
8. Steitz, T. A., Smerdon, S. J., Jäger, J., Wang, J., Kohlstaedt, L. A., Friedman, J. M., Beese, L. S. & Rice, P. A. (1993) *DNA and Chromosomes* (Cold Spring Harbor Lab. Press, Plainview, NY), pp. 495–504.
9. Smerdon, S. J., Jäger, J., Wang, J., Kohlstaedt, L. A., Chirino, A. J., Friedman, J. M., Rice, P. A. & Steitz, T. A. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 3911–3915.
10. Jacobo-Molina, A., Ding, J., Nanni, R., Clark, A. D., Lu, X., Tantillo, C., Williams, R. L., Kamer, G., Ferris, A. L., Clark, P., Hizi, A., Hughes, S. H. & Arnold, E. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 6320–6324.
11. Kraulis, P. (1990) *J. Appl. Crystallogr.* **24**, 946–950.
12. Davies, J. F., Hostomska, Z., Hostomsky, Z., Jordan, S. R. & Mattews, D. A. (1991) *Science* **252**, 88–95.
13. Carson, M. (1991) *J. Appl. Crystallogr.* **24**, 958–961.
14. Nicholls, A., Sharp, K. & Honig, B. (1991) *Proteins* **11**, 281–285.
15. Jaques, P. S., Wohrl, B. M., Howard, K. S. & LeGrice, S. L. J. (1994) *J. Biol. Chem.* **269**, 1388–1393.
16. Lee, B. & Richards, F. M. (1971) *J. Mol. Biol.* **55**, 379–400.
17. Richards, F. M. (1977) *Annu. Rev. Biophys. Bioeng.* **6**, 151–176.
18. Chothia, C. (1974) *Nature (London)* **248**, 338–339.
19. Page, M. L. & Jencks, W. P. (1971) *Proc. Natl. Acad. Sci. USA* **68**, 1678–1683.
20. Chothia, C. & Janin, J. (1975) *Nature (London)* **256**, 705–708.
21. Finkelstein, A. & Janin, J. (1989) *Protein Eng.* **3**, 1–4.
22. Bavand, J. R., Wagner, R. & Richmond, T. J. (1993) *Biochemistry* **32**, 10543–10552.