

Collective Motions in HIV-1 Reverse Transcriptase: Examination of Flexibility and Enzyme Function

Ivet Bahar^{1,2}, Burak Erman⁴, Robert L. Jernigan¹, Ali Rana Atilgan² and David G. Covell^{3*}

¹*Molecular Structure Section
Laboratory of Experimental and
Computational Biology
Division of Basic Sciences
National Cancer Institute
National Institutes of Health
Bethesda, MD 20892-5677
USA*

²*Polymer Research Center and
School of Engineering, Bogazici
University, and TUBITAK
Advanced Polymeric Materials
Research Center, Istanbul
Bebek 80815, Turkey*

³*Frederick Cancer Research and
Development Center, National
Cancer Institute, Science
Applications International
Corporation, Frederick
MD 21702, USA*

⁴*Sabanci University, Sabanci
Center, Istanbul 80745, Turkey*

In order to study the inferences of structure for mechanism, the collective motions of the retroviral reverse transcriptase HIV-1 RT (RT) are examined using the Gaussian network model (GNM) of proteins. This model is particularly suitable for elucidating the global dynamic characteristics of large proteins such as the presently investigated heterodimeric RT comprising a total of 982 residues. Local packing density and coordination order of amino acid residues is inspected by the GNM to determine the type and range of motions, both at the residue level and on a global scale, such as the correlated movements of entire subdomains. Of the two subunits, p66 and p51, forming the RT, only p66 has a DNA-binding cleft and a functional polymerase active site. This difference in the structure of the two subunits is shown here to be reflected in their dynamic characteristics: only p66 has the potential to undergo large-scale cooperative motions in the heterodimer, while p51 is essentially rigid. Taken together, the global motion of the RT heterodimer is comprised of movements of the p66 thumb subdomain perpendicular to those of the p66 fingers, accompanied by anticorrelated fluctuations of the RNase H domain and p51 thumb, thus providing information about the details of one processivity mechanism. A few clusters of residues, generally distant in sequence but close in space, are identified in the p66 palm and connection subdomains, which form the hinge-bending regions that control the highly concerted motion of the subdomains. These regions include the catalytically active site and the non-nucleoside inhibitor binding pocket of p66 polymerase, as well as sites whose mutations have been shown to impair enzyme activity. It is easily conceivable that this hinge region, indicated by GNM analysis to play a critical role in modulating the global motion, is locked into an inactive conformation upon binding of an inhibitor. Comparative analysis of the dynamic characteristics of the unliganded and liganded dimers indicates severe repression of the mobility of the p66 thumb in RT's global mode, upon binding of non-nucleoside inhibitors.

© 1999 Academic Press

*Corresponding author

Keywords: reverse transcriptase; flexibility; enzyme function; mutations

Introduction

Retroviral reverse transcriptases (RT) function to convert the single-stranded viral RNA genome into

Abbreviations used: RT, reverse transcriptase; GNM, Gaussian network model; RNase H, ribonuclease H; HIV-1, human immunodeficiency virus type 1; NNRTIs, non-nucleotide RT inhibitors; MD, molecular dynamics; ms, mean-square.

E-mail address of the corresponding author: covell@faxpdc.ncifcrf.gov

double-stranded DNA (Varmus, 1988; Weiss *et al.*, 1993). Two domains on the RT molecule are necessary for this process: a DNA polymerase domain that copies either an RNA or DNA template, and a ribonuclease H (RNase H) domain that cleaves RNA (Skalka & Goff, 1993). A competent RT of human immunodeficiency virus type-1 (HIV-1) is vital for viral replication, and thus remains an important target for anti-AIDS therapy (De Clercq, 1995a,b). Current strategies for the management of HIV-1 infection include agents that inhibit the function of RT (De Clercq, 1995a,b).

Clinical data indicate the emergence of resistant viral strains which limit the effectiveness of currently available drugs targeted against RT (Cohen, 1997). An understanding of the intrinsic conformational stability and flexibility of the HIV-1 RT molecule, both on a global and a local scale, may be essential for an assessment of the origin and effect of drug-selected amino acid mutations and for devising more efficient anti-AIDS drugs.

The determination of multiple structures of HIV-1 RT by X-ray crystallography has provided valuable information about the relationship between the structural components of RT and its enzymatic function (Kohlstaedt *et al.*, 1992; Jacobo-Molina *et al.*, 1993; Skalka & Goff, 1993; Jäger *et al.*, 1994; Ding *et al.*, 1995a,b; Esnouf *et al.*, 1995; Ren *et al.*, 1995; Hsiou *et al.*, 1996; Das *et al.*, 1996). HIV-1 RT is an asymmetric heterodimer comprised of two subunits of molecular mass 66 kDa (p66) and 51 kDa (p51), with identical residues in their first 428 amino acid positions (Figure 1). The p66 sub-

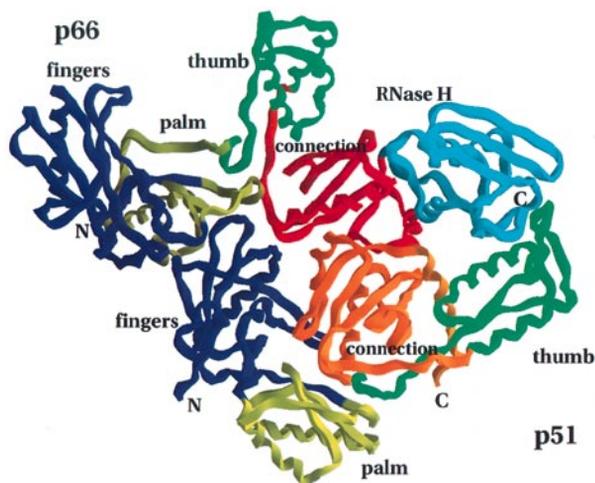


Figure 1. Ribbon diagram of the HIV-1 RT heterodimer in the complex with the NNRTI nevirapine (Ren *et al.*, 1995). The corresponding PDB code is 1rth. The p66 subunit is composed of two domains, the polymerase and the RNase H. The polymerase is comprised of the following subdomains: fingers (purple, residues, 1-88 and 121-146), thumb (green, 243-311), palm (dark yellow, 89-120 and 147-242), and connection (red, 312-425). The RNase H domain (426-560) is shown in blue. The p51 subunit has the same subdomains, fingers, palm, thumb and connection, as p66, shown in the same colors as their counterparts in p66. The connection subdomain of p51 is shown in orange, so as to be distinguishable from the adjacent p66, connection subdomain. The fingers and thumb subdomains of p51 are closely superimposable on their respective counterparts (fingers and thumb) in p66; whereas parts of the palm and connection differ. The tertiary packing of the subdomains is entirely different in the two subunits. The α -helices and β -strands referred to in the text (see Table 1) are indicated by the labels α_2 (78-83), β_4 (105-110), β_7 (179-183), β_8 (186-191), α_6 (195-211), β_9 (227-229), β_{10} (232-234), β_{11} (239-241), α_7 (254-267), α_8 (277-281), α_{10} (364-383), α_{14} (500-508), and α_{15} (516-527), using the notation of Ren *et al.* (1995).

unit is 560 residues long, with its DNA polymerase and RNase H domains in the amino and carboxyl-terminal portions, respectively. Based on an examination of crystal forms of HIV-1 RT, the anthropomorphic shape of a hand has been used to describe the polymerase domain, with subdomains made up of the fingers, palm, thumb, and connection to the RNase H domain (Figure 1). The p51 subunit comprises the same subdomains, but lacks the RNase H portion. The fingers, thumb, palm, and connection of the two subunits can be approximately superpositioned, pairwise, between p66 and p51. However, the tertiary packing of the subdomains within the subunits differs: p66 is described as an open hand, with a large cleft for binding double-stranded nucleic acids between the thumb and fingers subdomains, while p51 is considerably more compact, with no nucleic acid binding cleft (Nanni *et al.*, 1993).

Considerable effort has been directed at understanding the effect of mutations on RT polymerase function (Boyer *et al.*, 1992, 1994; Tantillo *et al.*, 1994). The polymerase active site (Asp110, Asp185 and Asp186) is located in the palm of p66. HIV-1 RT polymerase activity requires a primer and a template DNA strand. The precise positioning of the double-stranded portion of the template-primer is thought to be controlled by the fingers, palm and thumb of p66 (Gao *et al.*, 1998). Direct structural support for the role of various subdomains in template-primer positioning does not yet exist. Indirect evidence for flexible motions of the fingers, palm and thumb subdomains was found in a comparative study of three crystal forms of HIV-1 RT (Jäger *et al.*, 1994). Differences in local structure were observed that suggested a hinge-bending motion between the fingers and palm subdomains of p66, and the remainder of the molecule, described qualitatively as a swivel twist of the thumb subdomain. Additional crystallographic studies of HIV-1 RT, either bound with nucleic acid template-primers (Jacobo-Molina *et al.*, 1993) or non-nucleoside inhibitors (Ding *et al.*, 1995a,b; Ren *et al.*, 1995; Das *et al.*, 1996), or in the absence of ligands (Rodgers *et al.*, 1995; Hsiou *et al.*, 1996) support the flexibility of the p66 subunit. The p66 thumb, in particular, is suggested to be extremely flexible, based on the different conformations assumed by this subdomain in the DNA-bound, inhibitor-bound and unliganded structures (Hsiou *et al.*, 1996). The subunit p51, on the other hand, appears to act only as a support in the dimer, its internal flexibility being severely restricted.

A broad range of strategies has been proposed for therapeutic intervention in the treatment of AIDS. To date, HIV-1 RT inhibitors can be divided into two groups: those agents that directly interfere with DNA template-primer elongation (nucleoside inhibitors), and those that do not (non-nucleoside RT inhibitors, NNRTIs; De Clercq, 1994, 1995a,b). Although considerable evidence exists to suggest

that the intrinsic flexibility of RT may play an important role in its function, computational strategies for examining the global motions of HIV-1 RT at an atomic scale have been prevented by the large size of the molecule (a total of 982 residues in the two subunits p66 and p51) and the limits of molecular dynamics (MD) simulations to extremely short times; consequently, information regarding large-scale motions, processing mechanisms, and their applications towards the discovery of improved inhibitors has not been possible.

A new dynamics model was recently proposed (Bahar *et al.*, 1997; Haliloglu *et al.*, 1997), which permits an extremely efficient exploration of the collective motions for large systems. This coarse-grained approach, referred to as the Gaussian network model (GNM), is based on an analysis of the local packing density and bonded and non-bonded contact topology in a given structure. Both the slowest (global) motions and the fastest (local) motions are elucidated therein by a mode decomposition of the vibrational dynamics. The slowest modes provide information on the cooperative movements of large domains, relevant to binding and recognition, as well as identifying regions that are stationary during the collective motions of the overall molecule. The latter regions act as hinges inasmuch as they delimit the larger collective motions undergone by domains to which they are connected. Such regions potentially play an active role in modulating and monitoring the catalytic activity (Bahar *et al.*, 1998a; Bahar & Jernigan, 1998). The fastest modes, on the other hand, indicate residue positions severely constrained in the specific tertiary structure; these are generally correlated with the folding nuclei, and/or conserved residues, and therefore their mutation can have a disruptive effect on stability and function (Bahar *et al.*, 1998a; Demirel *et al.*, 1998).

Here, we provide an analysis of the range of motions available to HIV-1 RT. The crystal structures of HIV-1 RT, either complexed with NNRTIs or double-stranded DNA, or unliganded, will be examined using the GNM, to determine which portions of the molecule are the most flexible in different forms. The analysis will be performed both on the individual subunits (p66 and p51) of HIV-1 RT, as well as on the intact dimer. Regions predicted to have great flexibility in the dominant modes of motion will be compared with those experimentally determined to be important for binding and recognition. Regions exhibiting little flexibility will be contrasted to catalytically important sites, and other regions implicated in previous studies to play a critical role in the global conformational rearrangement of the enzyme. Results will be analyzed with regard to sites where drug-induced resistance mutations are known to occur, as well as to sites where panels of mutations have been constructed to examine HIV-1 RT polymerase function (Boyer *et al.*, 1992, 1994).

Results and Discussion

Description of protein dynamics using the Gaussian network model

The Gaussian network model inspects the local packing density and coordination order of each amino acid to determine its range of motions available in the folded state (Bahar *et al.*, 1997; Haliloglu *et al.*, 1997). Packing density is expressed in terms of the number of bonded or non-bonded α -carbon atoms located within a spherical shell of 7 Å about a central α -carbon. This cutoff separation is characteristic of the inter-residue first coordination shells in globular proteins (Miyazawa & Jernigan, 1985, 1996; Jernigan & Bahar, 1996; Bahar & Jernigan, 1997). The pairs of amino acids within this interaction range are referred to as "contacting" residues. Coordination order, on the other hand, refers to the separation, along the backbone, between two contacting residues.

These two properties, local packing density and coordination order, are the basic ingredients for constructing the Kirchhoff matrix of contacts, Γ , characteristic of the internal Hamiltonian of the investigated structure (Flory, 1976; Pearson, 1977; Kloczkowski *et al.*, 1989; Bahar *et al.*, 1997, 1998a; Haliloglu *et al.*, 1997). Γ is the counterpart of the stiffness matrix used in the analysis of elastic bodies. This is an $N \times N$ symmetric matrix for a protein of N residues, the ij th element of which is 1 if residues i and j are in contact, and zero otherwise. The diagonal elements are found from the negative sum of the off-diagonal elements, using $\Gamma_{ii} = -\sum_j \Gamma_{ij}$, where $j \neq i$, similar to transition rate matrices of stochastic processes.

In the GNM, motions are expressed as a superposition of modes of different frequencies (Haliloglu *et al.*, 1997). The procedure is similar to a normal mode analysis, with two major attendant advantages; simplicity and computational efficiency (Bahar *et al.*, 1997). These benefits are derived from adoption of a single parameter harmonic potential for all contacting pairs, instead of detailed, atom-specific or residue-specific potentials, following the original proposition by Tirion (1996). Computationally expensive energy minimization and MD simulations are thus avoided. Instead, the problem reduces to the inversion of the Kirchhoff matrix Γ . The simplicity of this approach offers a powerful advantage over alternative MD methods by readily allowing the exploration of large proteins, such as the ~ 1000 residue HIV-1 RT heterodimer. The cooperative motions in proteins of this size cannot be efficiently characterized with conventional MD simulations using present day computational facilities, due to excessive time and memory requirements for a thorough sampling of the conformational space.

The suitability of the GNM as a simple, yet physically reliable, tool for characterizing the structure-dynamics relationships in proteins has been established in two recent studies: the crystal-

lographic temperature factors of a series of proteins were accurately obtained (Bahar *et al.*, 1997); and the free energy changes observed in hydrogen exchange experiments under native, or only mildly denaturing, conditions were closely reproduced (Bahar *et al.*, 1998b). In both cases, correspondence between theory and experiments was excellent, based on this model for an accurate evaluation of the mean-square (ms) fluctuations of residues from the diagonal elements of the inverse Kirchhoff matrix, as $\langle \Delta R_i^2 \rangle \sim [\Gamma^{-1}]_{ii}$.

Here, the dynamics of HIV-1 RT will be decomposed, using the GNM, into a collection of $N - 1$ internal modes, with frequencies λ_k , $2 \leq k \leq N$, λ_1 being identically zero. We will concentrate on the subset of slowest modes. The slowest modes refer to the most cooperative motions. These are responsible for the global movements of entire subunits, or subdomains. We will examine the fluctuations in residue positions associated with these modes. The normalized distribution curve for the ms fluctuations of residues, driven by a given mode k , is referred to as the k th mode shape. The peaks and minima appearing in the slowest mode shapes are indicative of the structural domains or regions, critically important for binding and catalytic activity (Bahar *et al.*, 1998a; Bahar & Jernigan, 1998). Examination of the slowest mode shape of HIV-1 RT will thus give insight into the intrinsic flexibility of the structure, and serve to identify the hinge regions effectively controlling subdomain movements relevant to processivity.

First, calculations are performed for the subunits p66 and p51, in both monomeric and dimeric forms, using the nevirapine-bound HIV-1 RT structure determined at 2.2 Å resolution by Ren *et al.* (1995). These results will provide an assessment of the intrinsic conformational flexibility of the subunits in isolated form, and their perturbations in the heterodimer. Then, the dynamic characteristics of the nevirapine-bound heterodimer will be compared with those of the unliganded dimer (Hsiou *et al.*, 1996), and dimers complexed with other NNRTIs (Das *et al.*, 1996) and with double-stranded DNA (Jacobo-Molina *et al.*, 1993), for assessing the effect of NNRTI and/or DNA binding on the global dynamics of the enzyme. The blocks of residues participating in local hinge-bending mechanisms will be identified, along with the correlations between subdomains. The implications regarding the interaction between the activities of the polymerase and the RNase H portions of HIV-1 RT will be discussed.

Global motions of HIV-1 RT subunits: effects of dimerization

The mode shapes calculated for subunits p66 and p51 are displayed in Figure 2(a) and (b), respectively. The ordinate represents the normalized ms fluctuations of residues driven by the slowest two ($2 \leq k \leq 3$) modes of motion. The broken curves in Figure 2(a) and (b) represent the

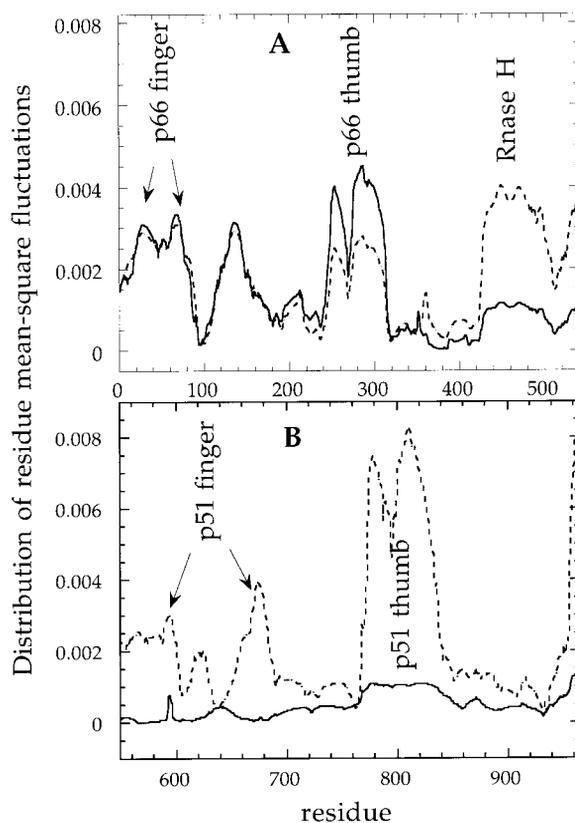


Figure 2. The normalized ms fluctuations of residues obtained with the slowest two modes using the Gaussian network model (GNM). Results are shown for residues in the complex with nevirapine, indexed along the abscissa. The broken and continuous curves represent the behavior of the subunits in isolated monomers and in the heterodimer, respectively. (a) p66; (b) and p51. The maxima in (a) coincide with the fingers, thumb and RNase H subdomains (see the legend to Figure 1 for residue ranges.) The palm and connection subdomains are severely restricted, as seen from the depressed portions of the curves in (a). Likewise, the p51 monomer exhibits broad peaks at the fingers and thumb subdomains, as shown in (b). The continuous curve therein reveals the striking suppression of the mobility of the fingers and thumb of p51 in the heterodimer.

behavior of the subunits p66 and p51, respectively, in isolated forms; the continuous curves display their counterparts in the heterodimer. Since crystal forms of monomeric p51 or p66 do not exist, the coordinates for each subunit are simply those taken from the heterodimeric form. The rationale for separately examining each subunit will be apparent in the next section. The α -carbon coordinates reported (Ren *et al.*, 1995) for the RT-nevirapine complex (PDB code 1rth) are used here for constructing the Kirchhoff matrix. Seventeen C-terminal residues of p66 are not reported (Ren *et al.*, 1995), as well as two terminal residues at both ends of p51, leading to Γ matrices of respective sizes 962, 526 and 436

for the HIV-1 RT dimer, and its monomeric subunits, p66 and p51, respectively.

Let us first consider the broken curves in Figure 2. These provide information about the intrinsic global flexibility of the subunits in the absence of the other subunit. A distinct feature is the occurrence of the largest amplitude fluctuations at regions associated with subdomains; the peaks in Figure 2(a) and (b) coincide with the fingers and thumb for p51 and p66, as well as the RNase H domain of p66. The exact residue positions for each subdomain are specified in the legend to Figure 1. Based on these observations, the p66 and p51 subunits possess an intrinsic flexibility at the finger and thumb subdomains, regions corresponding with the template-primer recognition sites. Flexibility is also associated with the RNase H domain. Of equal interest are the minima in each curve; regions where motion is restricted. These are expected to be act as hinges controlling the collective motions of the subunits. The palm and connection subdomains of both subunits are the loci of these minima.

The continuous curves in Figure 2 display the ms fluctuations in each subunit, based on the heterodimeric form of RT. This analysis reveals the global flexibility associated with the biologically functional form of the enzyme, and permits comparison with results obtained for each monomeric form. Striking differences are observed, e.g. the suppressed mobility within the finger and thumb subdomains of p51 in the heterodimer. As seen in Figure 2(b), nearly all of the fluctuations found in the monomeric form of the p51 subunit are absent, leaving a nearly flat fluctuation curve for p51 when analyzed in the context of the complete heterodimeric form. Thus distinctly different observations are made when comparing the fluctuations for the monomeric and heterodimeric forms of p66. The mobilities of the p66 fingers are unaffected by dimerization, while that of the thumb is actually enhanced over that of the monomeric form. A decrease in flexibility is found in the

RNase H domain of the dimeric form of p66 when compared with its monomeric form. This reduction in flexibility may be attributed to the stabilizing interactions between the RNase H domain and the p51 thumb subdomain.

The regions severely constrained in the collective motions of p66 in the monomeric form are almost identically preserved in the dimer. A list of residues exhibiting the lowest amplitude motions (normalized ms fluctuations below 10^{-3} in Figure 2(b)) is given in Table 1, along with the corresponding structural elements. The secondary structures named therein conform with those given by Ren *et al.* (1995); see also the legend to Figure 1. These regions are expected to play a critical role in monitoring the global motions of HIV-1 RT. The deepest minima of the mode shape, indicated in the last column of the Table, emerge as the potential hinge-bending centers of the dominant modes.

Comparison of the dynamic characteristics of unliganded dimer and those complexed with NNRTIs and DNA

In this section we concentrate on the dynamics of the p66 subunit, as this subunit exhibits the vital dynamic characteristics of the RT dimer, as explained above. Figure 3(a) compares the slow mode shapes of the p66 subunits in the unliganded HIV-1 RT (dotted), in the complexes with NNRTIs such as nevirapine (continuous, thin) and 9-CI TIBO (broken), and in a complex with a double-stranded DNA (continuous, thick). The respective codes in the Protein Data Bank (PDB) are 1dlo (Hsiou *et al.*, 1996), 1rth (Ren *et al.*, 1995), 1tvr (Das *et al.*, 1996), and 1hmi (Jacobo-Molina *et al.*, 1993). Calculations repeated for the C280S mutant HIV-1 RT complexed with 8-CI TIBO (PDB code 1hnv) yield results (data not shown) almost identical with those of 9-CI TIBO-bound RT. The curves in Figure 3 are normalized on the basis of the observed *B*-factors for each structures. The names of the p66 subdomains are explicitly given, and

Table 1. Regions severely constrained in the collective motion of RT, and corresponding to hinge-bending centers

Residues	Subdomain/domain ^b	Centers
87-110	p66 palm (α 2- β 4 loop, β 4)	Pro95- Pro97
177-192	p66 palm (β 7, β 8)	Tyr181
217-240	p66 palm (α 6- β 9 loop, β 9- β 11)	Pro236-Lys238
316-429	p66 connection (almost all)	Tyr319, Trp410-Pro412 (α 10)
499-531	RNase H (α 14, α 15)	Pro510-Glu514
266-274 ^a	p66 thumb (α 7, α 7- α 8 loop)	Gln269-Ile270

Except for the last row, these are residues with normalized ms fluctuations below 10^{-3} in the global motions of p66 (see Figure 2). See Figure 1 and its legend for the location of indicated helices and strands.

^a This is the sharp minimum at the central part of the p66 thumb.

^b The secondary structure identifications given in parentheses conform with the nomenclature used by Ren *et al.* (1996). Using instead that of Jacobo-Molina *et al.* (1993), the identifications become as follows: row 1, β 5a- β 6; row 2, β 9 β 10; row 3, β 11a- β 14; row5, α B'- β 5'; row 6, α H and loop α H- α I. And the centers listed in the third column refer to: row 1, β 5b; row 2, β 9; row 3, loop β 13- β 14; row 4, β 15 and β 20; row5, loop α B'- α D'; row 6, loop α H- α I.

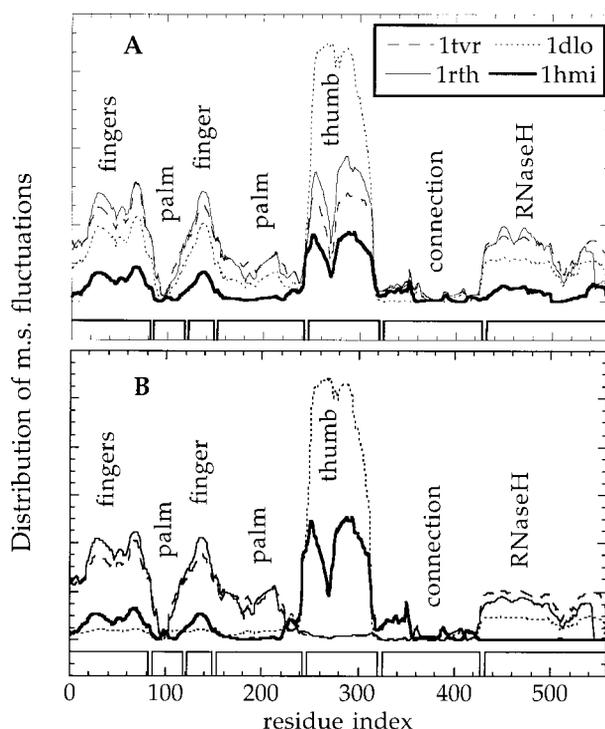


Figure 3. Comparison of the distributions of ms fluctuations for p66 residues in the unliganded RT dimer (1dlo), in the complex with DNA double-strands (1hmi), and in the complexes with the NNRTIs nevirapine (1rth) and 9-Cl TIBO (1tvr). (a) The cumulative distribution obtained from the superposition of the dominant three modes in the GNM; (b) the contribution of the slowest mode of motion, alone. Therein we note the largest changes to the mobility of the p66 thumb. See Figure 4.

their residue intervals indicated by the boxes along the lower abscissae.

The curves displayed in Figure 3(a) represent the fluctuations driven by the three slowest, most cooperative (or dominant) modes of motion. The close similarity of the curves demonstrates that the p66 subunit preserves the same mode shape, qualitatively, in all of these different forms of the dimer. The individual modes, when observed separately, show some differences, however, among the structures.

The differences essentially reside in the shape of the slowest single mode of motion for each structure, as illustrated in Figure 3(b). In this global mode, the most striking result is the extreme sensitivity of the p66 thumb's mobility to ligand-binding. In fact, in the unliganded form, the p66 thumb is distinguished by an extremely pronounced

flexibility, which is retained to some significant extent in the DNA-bound form; whereas in all of the NNRTI-bound structures, the mobility of the p66 thumb is almost completely suppressed. Thus, the differences in the structures of the different forms of RT, unliganded or complexed (Figure 4), are directly evidenced in their global dynamics, as well. A second interesting observation common to all examined structures was the strong coupling between the RNase H and p51 thumb. These were discerned to form a coherent pair of structural elements, whose mobilities are simultaneously suppressed (in the DNA-bound form) or enhanced (in the NNRTI-bound forms). The close coupling between the RNaseH and the p51 thumb is further apparent in the analysis of cross-correlations between subdomain motions (see below).

We note in Figure 3 that the two NNRTI-bound structures exhibit similar dynamic characteristics. Similarities in the binding of diverse NNRTIs were already pointed out in X-ray crystallographic studies (Ding *et al.*, 1995a), and here these structural similarities are shown to extend to their dynamic characteristics as well. The unliganded form, on the other hand, is distinguished by the pronounced mobility of the thumb subdomain. Interestingly, this property is valid despite the fact that the conformation of the thumb in the unliganded structure has been qualified as a "closed" one (Rodgers *et al.*, 1995), given its close proximity to the fingers, partly filling the DNA-binding cleft of the hand-shaped polymerase domain of p66 (Figure 4). Figure 4(a), (b), (c) and (d), in which the 9Cl-TIBO-bound, nevirapine-bound, unliganded and DNA-bound RTs are displayed respectively, shows that the enhancement in the flexibility of the p66 thumb in the unliganded form (1dlo) originates in its decoupling from the RNase H domain. Figure 4(d) illustrates the contraction of the enzyme upon binding to DNA, which explains the observed (Figure 3) generally more restricted mobility of the corresponding structure (PDB code 1hmi).

Suppression of mobility at the interface between p66 and p51

An additional comparison of the fluctuations in the heterodimer with those of each monomeric subunit considered separately, may be obtained by examining the ms fluctuations as displacements, $\langle \Delta R_i \Delta R_i \rangle$ for all modes of motion. This analysis identifies residues that exhibit large reductions in fluctuations upon dimerization across the complete spectrum of available frequencies. Figure 5(a) displays the ms fluctuations for the p66 residues in 1rth. The results for the monomeric and dimeric forms of the subunit are shown as the broken and the continuous curve, respectively. The differences between the two curves of Figure 5(a) are shown in Figure 5(b). The results[†] for the subunit p51 are presented in Figure 6.

In general, the ms fluctuations are either unchanged or slightly reduced in the heterodimeric

[†] The parameter γ , common to all residue pairs in the monomers and/or the dimer, is fixed at $0.8 \text{ RT}/\text{\AA}^2$ from a comparison of the GNM results for RT with the crystallographic B -factors, $B_i = 8\pi^2 \langle \Delta R_i \Delta R_i \rangle / 3$. This value scales the absolute heights of the curves, without affecting their shape.



Figure 4. Ribbon diagrams of the four RT structures considered in Figure 3. The PDB codes of the structures are indicated. The thumb, palm, fingers and connection subdomains of the p66 polymerase are colored in red, yellow, cyan, and green, respectively. The RNase H domain is shown in blue, and the p51 subunit is white. The NNRTIs in 1tvr and 1rth and the DNA double-strands in 1hmi are shown in magenta.

form when compared with each monomeric subunit, as revealed by the comparison of the continuous and broken curves in Figures 5(a) and 6(a). Selected clusters of residues exhibit reduced mobility in the heterodimeric form. Portions of the connection subdomain and RNase H domain of p66, as well as residues within the finger and thumb subdomains of p51 display the greatest

reduction in mobility. Seven stretches of residues are found to be the most severely constrained upon dimerization; Phe87-Trp88, Thr403-Trp410 and Lys540-Gly543 in p66, and Leu26-Glu28, Pro52-Pro55, Ser134-Gly141, and Cys280-Glu291 in p51. These are all located at the interface between the monomers.

Three clearly visible clusters are formed between these interfacial residues. Two of these involve

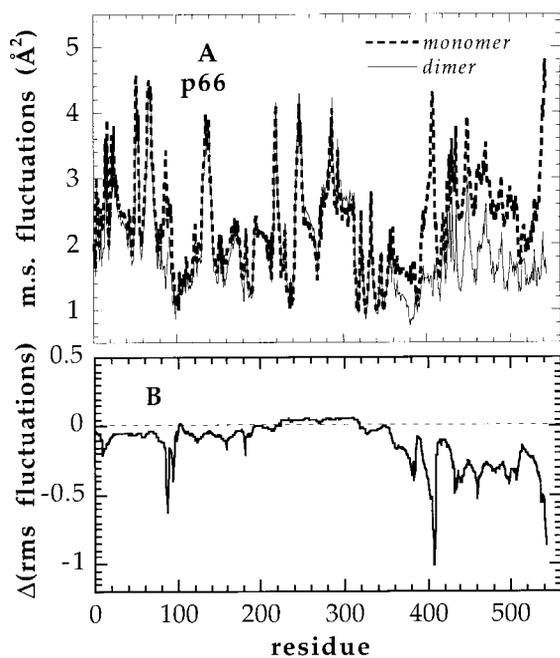


Figure 5. (a) Comparison of the ms residue fluctuations for all of the modes of motion for subunit p66, as a monomer (broken line) and in the RT dimer (continuous line). (b) Changes in the fluctuations of p66 residues upon dimerization found from the difference between the solid and broken curves in (a). Minima refer to residues most severely constrained upon dimerization: Phe87-Trp88, Thr403-Trp410 and Lys540-Gly543.

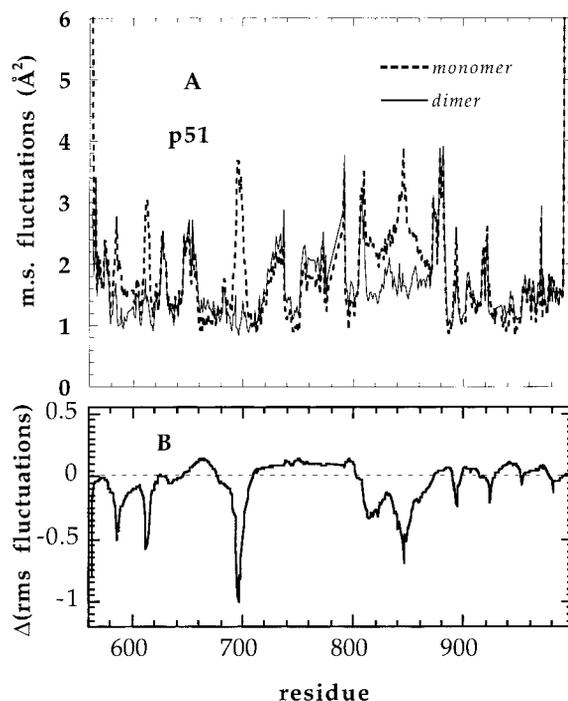


Figure 6. Change in residue fluctuations for all modes of motion upon dimerization for subunit p51. See the legend to Figure 5 for the descriptions of (a) and (b). Minima are observed at residues Leu26-Glu28, Pro52-Pro55, Ser134-Gly141, and Cys280-Glu291.

interactions between the same subdomains in each monomer, forming an interface between the finger-finger subdomains and the connection-connection subdomains of p51 and p66. This apparent symmetry is not conserved within the residues of each subdomain; a single region on the fingers of p66 (Phe87-Trp88) interacts with two regions on the fingers of p51 (Pro52-Pro55 and Ser134-Gly141). Thus, although the sequence identity of the residues is high between the monomeric subunits of RT, different residues within the subdomains of each subunit have their motion constrained upon dimerization. Additional asymmetric interactions between the monomeric subunits are also observed. The RNase H residues 540-543 interact with the p51 thumb residues 280-291, while the p66 connection residues Thr403-Trp410 interact with the p51 connection, but do not appear to constrain specific residues therein.

An independent analysis of the residues buried upon dimerization finds a strong correspondence between the most constrained residues, described above, and those residues calculated to make the largest contribution to the dimerization binding strength. This analysis treats dimerization as a binding process, and calculates each residue's contribution to the total binding free energy (Covell *et al.*, 1994, 1998; Wallqvist *et al.*, 1995). Accordingly, a total of 3697 Å² of solvent-accessible surface area is buried in the heterodimeric RT complex. An extraordinarily large binding strength of 34 kcal/mol is calculated for this interaction, a result consistent with the observed difficulty associated with chemically separating the monomeric species (S. Hughes, personal communication). The residues in p66 that make the largest contribution to total binding strength are Ile542, Trp88, Trp406 and Lys540 in p66, and Leu289, Pro52 and Glu138 in p51. The surface buried by these seven residues represents nearly 30% of the total interfacial area of the heterodimer. Each of these residues lies within the clusters identified above with the GNM to be severely constrained upon dimerization.

Comparisons with crystallographic flexibility

Although protein models resulting from crystallographic structure determination are a set of static coordinates, the temperature factors included in these datasets provide indications about the extent to which electron density is smeared out (Frauenfelder *et al.*, 1979). Thus, while the values of the Debye-Waller temperature factors (or *B*-factors) may be affected by disorder within a crystal, it is generally thought that the *B*-factors contain information about the intrinsic flexibility of residues in the folded state. For harmonic thermal vibrations, each atom's *B*-factor is directly proportional to its ms amplitude of vibration. Despite the occurrence of anharmonic modes and localized rotameric jumps near the folded state of proteins, the Debye-Waller factors are dominated by unimodal

quasi-harmonic motions, rather than anharmonic, multimodal jumps between conformational isomers, as elegantly demonstrated by Garcia *et al.* (1997).

Figure 7(a) displays the reported crystallographic *B*-factor (Ren *et al.*, 1996) for each p66 residue as a spectrally colored ribbon. Regions exhibiting the largest amplitude thermal fluctuations are shown in blue, whereas the least flexible regions are shown in red. A diffuse distribution of residue flexibility, as determined from the crystallographic *B*-factors, is observed. The largest *B*-factors are found at four locations scattered throughout the subdomains: (1) the tips of the fingers; (2) a loop in the palm; (3) a portion of the connection domain; and (4) a section of β -sheet within the thumb. The lowest *B*-factors are located primarily within the connection subdomain, while the palm and RNase H also exhibit relatively low thermal motions. These measures of flexibility correspond with fluctuations resulting from the combined effect of various frequency modes. A strong correlation is observed between the *B*-factors of RT and those calculated with the GNM method based on the superposition of fluctuations across all frequencies. This is consistent with earlier observations on other proteins (Bahar *et al.*, 1997, 1998a; Haliloglu *et al.*, 1997), and for a tRNA-synthetase complex (Bahar & Jernigan, 1998).

The flexibility of the thumb and fingers of p66 in the global motion of RT cannot, however, be

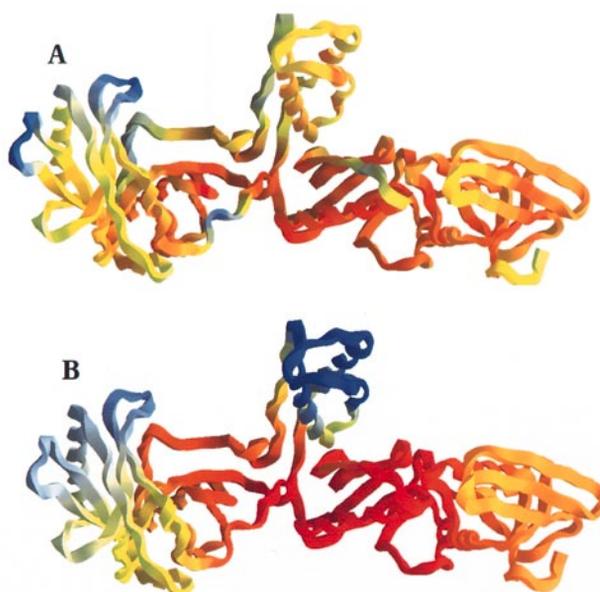


Figure 7. Ribbon diagram of the p66 subunit of HIV-1 RT, colored according to the mean-square fluctuation amplitudes of residues in the dimer. (a) Fluctuations from all modes determined from X-ray crystallographic *B*-factors (Ren *et al.*, 1995). (b) Fluctuations in the two slowest modes extracted by the GNM approach (see Figure 2(a)). Color code: spectrally from blue, the largest amplitude thermal fluctuations to red, the least flexible regions.

elucidated by a direct examination of the crystallographic B -factors. The GNM procedure readily permits decomposition of the protein dynamics into a series of different frequency modes, and provides a powerful means of separately examining the individual components, and in particular the dominant slowest modes, of the complete vibrational spectrum. Figure 7(b) shows a ribbon diagram of p66, color coded according to the fluctuations in the slowest two dominant modes of motion extracted by GNM analysis. This Figure simply illustrates the amplitudes of fluctuations for the heterodimeric form of p66, as presented in the continuous curve in Figure 2(a).

Comparison of the two ribbon diagrams in Figure 7(a) and (b) reveals agreement by jointly identifying the connection subdomain as the least flexible portion of RT. In contrast with the measures of flexibility determined from the analysis of B -factors, the GNM identifies the thumb region as the subdomain exhibiting the largest

range of motion for the slowest two modes, suggesting that it may be the most active element in the processing motion driving nucleic acids through the enzyme. The GNM analysis indicates a cooperative motion over nearly the entire thumb subdomain, in strong contrast with the B -factors that specify motions restricted largely to one strand within the thumb. A second region of large motion is observed at the tips of the fingers, in close agreement with the results obtained from thermal fluctuations. Based on the regions of flexibility associated with the two slowest modes of motion, the global flexibility of RT strongly suggests a grasping motion of the thumb and finger subdomains, with the connection and part of the palm subdomains providing a rigid support for this motion. An intermediate flexibility is found for the RNase H domain, which, as discussed below, may play an important role in orchestrating the proper positioning of the template-primer at the site of the RNase H function.

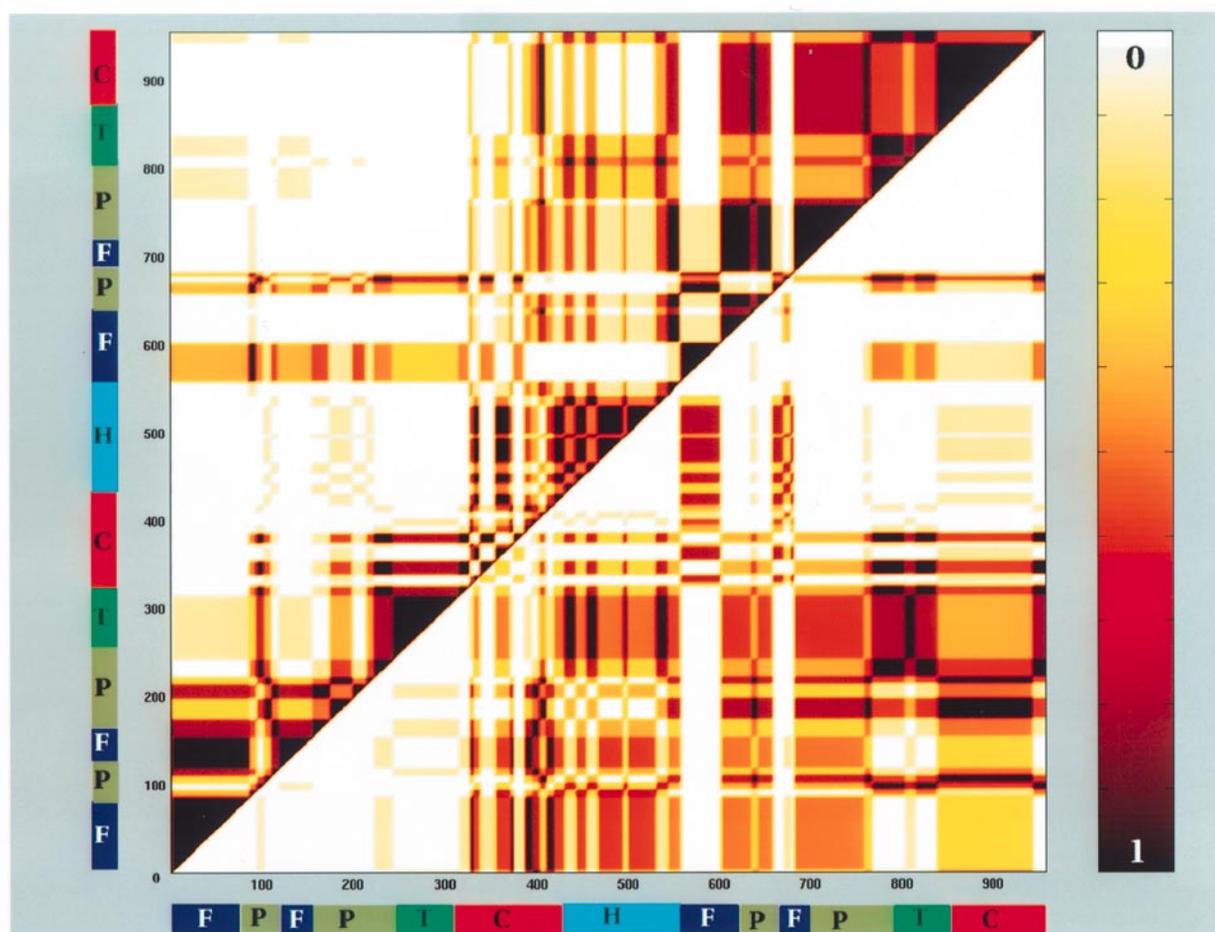


Figure 8. Correlation map for residue fluctuations during the global motions of HIV-1 RT. The correlations $C(i, j)$ between the fluctuations of residues i (abscissa) and j (ordinate) are calculated for the dominant slow modes ($2 \leq k \leq 5$) using equation (3), and then normalized (see the text) in the range $-1 \leq C(i, j) \leq 1$. The positive correlations are shown on the diagonal and upper triangular part of the map, and the negative correlations in the lower triangular part. The correlation levels shown in different colors are indicated on the right-hand side of the map. The different structural subdomains of the subunits are indicated by single-letter names (F, fingers; P, palm; T, thumb; C, connection; H, RNase H) along the vertical and horizontal bars, at the left and the bottom of the map.

Correlations between subdomain motions

The correlation map in Figure 8 describes the types and strengths of couplings between residue fluctuations during the global motions of HIV-1 RT. The ordinate and abscissa refer to residue numbers, including both subunits p66 and p51. The different subdomains are indicated by the corresponding capital letters (F for fingers, P for palm, etc.) along the left ordinate and lower abscissa. The map displays the orientational correlations between residue fluctuations, $\langle \Delta R_i \Delta R_j \rangle_k$, determined for a representative subset ($2 \leq k \leq 5$) of slow modes. The correlations are calculated for the nevirapine-bound form 1rth using equation (3) described in Materials and Methods, and then normalized with respect to autocorrelations, as described (Haliloglu *et al.*, 1997; Bahar *et al.*, 1998a). The results thus vary in the range $-1 \leq C(i, j) \leq 1$. The lower and upper limits refer to anticorrelated (coupled but in an opposite direction) and correlated (coupled, in the same direction) motions; $C(i, j) = 0$ refers to uncorrelated, no motion or orthogonal motions. For clarity, the regions exhibiting positive correlations are shown on the diagonal and upper triangular parts of the map; whereas the lower triangular part refers to anticorrelated motions. The color code shown on the right describes the absolute values of correlation functions: dark regions refer to residue pairs whose fluctuations are fully correlated, positively or negatively; red regions indicate relatively strong correlations; yellow and white refer to weakly correlated and uncorrelated pairs, respectively.

Let us first consider the diagonal and upper triangular portions of the map. Sequential residues exhibit strong positive correlations, in general, as indicated by the dark regions on and near the diagonal. This observation reflects the fact that subdomains are often comprised of sequential residues, and most of the subdomains exhibit strong positively correlated intradomain motions. We note in particular, that the residue pairs participating in the p66 fingers, which consist of two discontinuous segments 1-88 and 121-146, are almost fully correlated despite the interference of a part

(89-120) of the intervening palm subdomain. Likewise, the thumb (243-311) forms a highly coherent block. Residues in RNase H also tend to move together, while those of the p66 palm (89-120 and 147-242) and connection (312-425) exhibit a more complex behavior. In the latter case, their "blocked" correlation patterns suggest subregions of positively correlated motion within these subdomains. Positively correlated motions within subdomains generally involve residues belonging to the same strands or helices.

Numerous coupled motions are observed between subdomains. These interdomain couplings are indicated by the dark colored regions away from the diagonal. Within the p51 subunit, strong positively correlated motions are observed between the palm, thumb and connection subdomains. Thus, almost all residue pairs in p51, except those in the fingers, appear to move in concert. This supports the observation that the internal motions of p51 are damped by heterodimerization. Such positively correlated interdomain couplings are largely absent in the p66 subunit; this confirms that even with dimerization, the p66 monomer exhibits a diverse range of subdomain motions.

The lower portion of the correlation map indicates motions that are coupled, but in the opposite sense. The strongly positively correlated regions near and above the diagonal that were discussed earlier, necessarily have zero correlated motions of this type. Examples of the strongest cross-correlations of opposite sense are between the fingers and connection of p66, and between the p66 thumb and RNase H. Also, orientational anticorrelations between subdomains of the two different subunits are discernible, such as that between the thumb subdomains of p66 and p51. Similar coupled motions were inferred from X-ray crystallographic studies of different structures (Jäger *et al.*, 1994; Ding *et al.*, 1995b; Hsiou *et al.*, 1996).

Our results are summarized in Table 2. Here, the orientational correlations between the global motions of all subdomains are presented. The values listed in the Table refer to averages taken over all residue pairs belonging to a given pair of subdomains. The diagonal elements therein give information on the coherence of the motion of

Table 2. Correlations between subdomain motions in RT

	Fingers	Palm	p66 Thumb	Connection	RNaseH	Fingers	Palm	p51 Thumb	Connection
Fingers	<u>0.99</u>	0.46	0.09	-0.47	-0.43	-0.02	-0.37	-0.03	-0.25
Palm	0.46	<u>0.57</u>	0.28	-0.08	-0.29	0.05	-0.46	-0.46	-0.61
Thumb	0.09	0.28	<u>0.99</u>	0.16	-0.66	-0.05	-0.57	-0.75	-0.45
Connection	-0.47	-0.08	0.16	0.29	0.12	-0.03	-0.02	-0.22	-0.11
RNaseH	-0.43	-0.29	-0.66	0.12	<u>0.68</u>	-0.15	0.32	0.43	0.19
Fingers	-0.02	0.05	-0.05	-0.03	-0.15	0.49	0.35	-0.04	0.20
Palm	-0.37	-0.46	-0.57	-0.02	0.32	0.35	<u>0.79</u>	<u>0.55</u>	<u>0.74</u>
Thumb	-0.03	-0.46	-0.75	-0.22	0.43	-0.04	<u>0.55</u>	<u>0.81</u>	<u>0.67</u>
Connection	-0.25	-0.61	-0.45	-0.11	0.19	0.20	<u>0.74</u>	<u>0.67</u>	<u>0.91</u>

Determined from low frequency/large amplitude GNM fluctuations (subset of four modes). Underlined numbers have absolute values larger than 0.5.

residues within each subdomain. Except for the p66 connection subdomain, these are relatively large and positive, indicating that residues in a given subdomain undergo concerted motions in the same direction, in general. The off-diagonal terms provide a quantitative estimate of the extent of interdomain couplings. The strongest intersubunit cross-correlation occurs between the RNase H domain of p66 and the thumb subdomain of p51.

The off-diagonal entries in Table 2 that are small in magnitude may arise for two different reasons: (1) residue pairs belonging to the respective subunits may be all uncorrelated or weakly correlated; or (2) the individual pairs may have strong pairwise correlations, but not necessarily in the same sense (+ or -), such that (large) positive and negative correlations may counterbalance each other. In order to distinguish between the two cases, we consider averages over the "absolute" values of correlations. The case (2) was observed to hold, in general, except for two pairs: the first between the fingers and thumb of p66, and the second between the p66 fingers and p51 thumb. These three subdomains (p66 fingers, p66 thumb and p51 thumb) exhibit strongly coherent intradomain motions when observed individually. See the corresponding auto-correlations (diagonal elements) in Table 1. The high autocorrelations and low cross-correlations for a given pair of subdomains is simply indicative of the tendency of the subdomains to undergo motions perpendicular to each other. Thus, the p66 fingers are subject to fluctuations perpendicular to those of the p66 and p51 thumbs.

In sum, insofar as the collective dynamics of the p66 subunit are concerned (or of the entire HIV-1 RT dimer, since the motions of the p51 subunit are severely hindered), the final picture that arises is the following. (1) The motions of the fingers and the thumb are highly concerted (autocorrelation above 0.90), but along directions perpendicular to each other (cross-correlations below 0.1), similar to the directions of curling of the thumb and fingers in a hand. Thus, the analogy to the fingers and thumb of a hand appears to be appropriate on the basis of the dynamic directional preferences of the two subunits. (2) The RNase H domain exhibits coupled but anticorrelated motions with respect to both p66 fingers and p66 thumb (cross-correlations of -0.43 and -0.68, respectively; see Table 2). The movements of the RNase H domain in the global mode, generally accompanied by those of the p51 thumb, thus complement those of these two most flexible subdomains to control the closing/opening of the p66 cavity for DNA binding, which also contains the active site cleft. This motion probably corresponds to the processing motion in which the nucleic acid is moved along the surface of HIV-1 RT, and exhibits the coordinated release by the hand and pull by the RNase H domain, or alternatively a push by the closed hand followed by its release and a pull by the RNase H domain. While it is interesting to uncover all details of this motion, this requires further calculations in which

explicit directions of the fluctuations and their correlations are considered. Such a study should reveal the active members in this motion. Is this motion a push by the hand, with the fingers and thumb closed, or a grab by the RNase H subdomain, with the fingers and thumb opened? Learning these details now becomes a relatively straightforward extension of the present study.

Implications of domain motion on RT function

HIV-1 RT must precisely cleave the RNA template-primer to successfully copy the viral RNA genome (Gao *et al.*, 1998). Numerous structural studies on the behavior of mutants support the view that the polymerase region of p66 plays a role in the proper positioning of the template-primer both at the polymerase active site and at the RNase H active site. Evidence for this view can be found in studies by Hughes and co-workers, where selected mutations in the p66 polymerase domain were demonstrated to alter the efficiency and specificity of RNase H cleavage (Boyer *et al.*, 1992, 1994; Tantillo & Arnold, 1994). Alternative explanations for the role of mutations in RT function raise the possibility that rather than affecting nucleic acid positioning, mutations affect the structure of the polymerase and RNase H active sites and reduce their enzymatic efficiency and specificity.

The results presented herein offer some insight into the location of residues that may affect the global motions of RT, and by doing so, limit its ability to process the RNA template-primer. Figure 2 shows that the global motions of RT involve large amplitude distortions of the fingers and thumb subdomains of p66. As noted earlier, these motions are in direct contrast to those of the connection and of a considerable part of the palm, which exhibit a limited conformational mobility, much like the p51 subunit of the heterodimer.

Regions with comparably low ms fluctuations are listed in Table 1. These comprise three discontinuous stretches in the p66 palm, almost the entire p66 connection, and the C-terminal portion of the RNase H domain. The central parts of these regions apparently act as hinges for the large-scale movements of the p66 polymerase subdomains, and the RNase H domain. A closer examination of these centers reveals: (1) a loop in the p66 palm extending between residues 92 and 105, adjacent to the p66 connection; (2) the strand $\beta 7$ at the interface, comprised of residues 178-183; (3) the $\beta 10$ - $\beta 11$ hairpin (235-238) at the "knuckle" of the p66 thumb, complemented by a loop around Tyr319; (4) the helix $\alpha 10$ in the p66 connection, again on the interface between the two subunits; and (5) the loop (509-515) between helices $\alpha 14$ and $\alpha 15$ of RNase H, next to the p66 connection. Finally, the p66 thumb residues 266-274 near the C terminus of helix $\alpha 7$ are distinguished by their restricted mobility (see Figure 2), suggesting a joint flexure at that location.

Let us first consider the severely constrained regions located in the p66 palm, listed in the first three rows of Table 1. Those in the first two rows include the triad of catalytically active conserved aspartic acid residues, Asp110, Asp185 and Asp186, as well as the NNRTI binding site at the β 7 and β 8-hairpin, the strand β 4, and the loop containing L100 (Esnouf *et al.*, 1995). It is easily conceivable that this region, indicated by GNM analysis to play a critical role in modulating the cooperative motion of the entire molecule, is locked into an inactive conformation upon binding to the inhibitor. In particular, NNRTI-binding has a drastic effect on the mobility of the p66 thumb, as evidenced by the complete suppression of its fluctuations in the global (slowest) mode of motion (see Figure 3(b)). The p66 palm residues Tyr181, Tyr183 and Tyr188 move so as to accommodate the repositioning of the catalytic aspartic acids during inhibition (Esnouf *et al.*, 1995). Among other residues participating in the hinge-bending region of the p66 palm, we note that Leu187 is severely constrained due to its close interaction with the three aspartic acid residues. A Leu187 mutation to serine was indeed found to lead to a decrease in polymerase activity below 5% of the wild-type RT level. Finally, considering the third row of Table 1, we note that a rigid body rotation of the p66 thumb near the thumb's knuckle residues W239 and V317 was pointed out by Hsiou *et al.* (1996) based on the differences in the structures of unliganded and liganded RT; and these two residues lie within the collective motion hinges identified here.

Another calculation of residue burial energies, using methods described by Covell *et al.* (1994) and Wallqvist *et al.* (1995), indicates the sequence positions around 100, 220-240 and 266 to be highly constrained. These are all located within the hinge-bending regions presented in Table 1. Sequence analysis finds that for the set of retroviral reverse transcriptase sequences, these positions are always occupied by hydrophobic residues. More specifically, positions 100 and 234 always contain a branched aliphatic amino acid (I, L or V) while position 229 is always occupied by an aromatic residue (W, Y or F). This cluster of hydrophobic amino acids may specifically provide a "universal joint" for RT flexure. Inspection of the RT-nevirapine complex finds this inhibitor to be buried by the surfaces of L100, W229 and L234. Thus, nucleoside inhibitors of RT may directly affect the capacity of this region to function as a universal joint, simply by binding to this site. Among these three residue positions, drug-induced resistant mutations are known to occur only at position 100, where isoleucine appears to be tolerated (<http://igs-server.cnrs-mrs.fr/anrs/los-alamos/96compendium.html>). Site-directed mutation of position 100 to serine completely eliminates polymerase function (Boyer *et al.*, 1994). Substitution of serine for either tryptophan 229 or leucine 234 also inactivates RT (S. Hughes, unpublished observations).

The p66 thumb is composed of three helices α 7, α 8 and α 9 (Ren *et al.*, 1995; also designated as helices α H, α i and α J; Jacobo-Molina *et al.*, 1993). Tryptophan 266 lies towards the C cap of helix α 7. Recent studies find that the mutation W266T has a profound effect on both RT polymerase and RNase H activities, showing no polymerase activity and a nearly 90% reduction in RNase H activity. Substitution of tyrosine at position 266 resulted in only a 15% reduction of polymerase activity, indicating tolerance for an alternative hydrophobic amino acid at this site. The side-chain of W266 makes contacts with residues adjacent to the L100, W229 and L234 triad, and most likely functions, together with the near neighbors Gln269 and Ile270, and the knuckle residues in the β 10- β 11-hairpin, as a hinge for motions of the thumb.

Conclusion

The inhibition of RT activity might be introduced by two means: (1) destabilization of the overall structure upon mutating the residues key to the stability of the dimer. Residues at the interface, involved in the binding of the two monomers, could be conceived as sites of potentially disruptive mutations, inasmuch as the enzyme fulfils its function only in the dimeric form. (2) Suppression of conformational motions critical to function. Structural evidence suggests that the intrinsic conformational flexibility of HIV-1 RT is essential in controlling the mechanisms of polymerase and RNase H catalysis, binding of nucleoside and non-nucleoside inhibitors, and drug resistance. In the present study, we have identified residues, both at the interface (Figures 5 to 7), or at regions critically important for the global conformational fluctuations of the protein (Table 1), which might be suitable candidates for testing in mutation studies as potential sites for destabilizing the dimer, or impairing the function, of the enzyme.

Theoretical approaches using MD simulations and normal mode analysis are methods commonly used for understanding the collective motions and correlations in globular proteins. Applications of these methods to proteins are restricted by the size of many proteins of interest. Analyses of proteins greater than about 300 residues in length become prohibitively time consuming. The present method circumvents the shortcomings of these other two simulation methods. Computationally, only the inversion of the Kirchhoff matrix of spatially close residues is required for obtaining the correlations, and requires only about five minutes on a Silicon Graphics R8000 Workstation for molecules of around 1000 residues.

The results of our application of this method towards understanding the global motions of RT suggest a rather precise set of correlated and anti-correlated motions. Strong positively correlated intradomain motions are observed within most of

the subdomains of the p66 and p51 subunits. Strong interdomain correlations are observed in the p51 subunit, suggesting that the p51 subunit lacks much internal mobility; instead it moves as a fairly rigid block, presumably acting as a support for the motion of the p66 subunit, which on the contrary, exhibits diverse internal motions. The mode shape plotted in Figure 2 and illustrated in Figure 7(b) demonstrates that the two regions most active in the global dynamics of this enzyme are the p66 thumb and fingers subdomains. These undergo strongly coupled fluctuations along directions perpendicular to one another, being accompanied by an approximately opposite sense movement of the RNase H domain. Our analysis of the global motions of RT shows that the mechanism for the motion of heterodimeric RT involves "hinge pins" at mutationally resistant sites in the palm and connection subdomains of p66, and that the restriction of the mobility of the p66 thumb upon binding of NNRTIs can cooperatively affect the global dynamics of the enzyme.

When the first protein structures were being determined the widely held assumption was that these structures would directly inform us about all details of molecular functions and dynamics. At the present time, when we begin to have a representative large set of protein structures, it is clear just how naive this original viewpoint was. MD has been held out as an approach that would let us calculate function from structure. This too has proven to be disappointing. It now becomes obvious that new approaches are required to permit us to infer from structure the details of function. The present approach extends in this direction. By reducing most of the structural details, considerations of large-scale modes of motion can now be used to explore how these dominant fluctuations are related to molecular function.

Materials and Methods

The basic postulate in the GNM approach is that the protein in the folded state is equivalent to a three-dimensional elastic network. The C^α atoms are identified with the junctions of the network. They fluctuate under the joint effect of their interactions with all contacting residues. Each interaction is represented by a single parameter harmonic potential. The inter-residue interaction potential substitutes for the harmonic potential constraining the end-to-end separation of such network chains. The internal Hamiltonian of the protein is expressed, by analogy to the theory of random Gaussian networks as:

$$\mathcal{H} = 1/2\gamma \text{tr}\{\Delta R^T \Gamma \Delta R\} \quad (1)$$

where γ is the single parameter (force constant) of the Hookean potential originally proposed by Tirion (1996), ΔR represents the N -dimensional hypervector of the fluctuation vectors $\Delta R_1, \Delta R_2, \dots, \Delta R_N$ of the C^α atoms, the superscript T denotes the transpose, and tr designates the trace of the matrix enclosed in braces. The equilibrium correlation between the fluctuations of residues i

and j is found from (Kloczkowski *et al.*, 1989; Bahar *et al.*, 1997):

$$\begin{aligned} \langle \Delta R_i \Delta R_j \rangle &= (1/Z_N) \int \Delta R_i \Delta R_j \exp(H/kT) d\{\Delta R\} \\ &= (3kT/\gamma) [\Gamma^{-1}]_{ij} \end{aligned} \quad (2)$$

where Z_N is the configurational integral given by $Z_N = \int \exp\{-\mathcal{H}/kT\} d\{\Delta R\}$, k is the Boltzmann constant, T is the absolute temperature, and the integration is carried out over all residue fluctuations $\{\Delta R\} \equiv d\Delta R_1 d\Delta R_2 \dots d\Delta R_N$.

The dynamic characteristics of the protein are described in terms of (1) its natural frequencies, and (2) the shapes of the corresponding modes of motion. The former is given by the eigenvalues λ_i , $2 \leq i \leq N$ of Γ , excluding the zero eigenvalue λ_1 , and the latter by the eigenvectors u_i , $2 \leq i \leq N$, of Γ . The cross-correlations $\langle \Delta R_i \Delta R_j \rangle_k$ associated with the k th mode of motion are found from (Haliloglu *et al.*, 1997):

$$\begin{aligned} \langle \Delta R_i \Delta R_j \rangle_k &= (3kT/\gamma) [\lambda_k^{-1} u_k u_k^T]_{ij} \\ &= (3kT/\gamma) \lambda_k^{-1} [u_k]_i [u_k]_j \end{aligned} \quad (3)$$

Here, the subscripts designate the elements of the matrices (or vectors) enclosed in square brackets. Correlations driven by multiple modes are estimated from the weighted average of the above cross-correlations, using λ_k^{-1} as the weighting factor for the contribution of each mode k .

Acknowledgments

The authors thank S. H. Hughes and P.L. Boyer for thought-provoking discussions about RT and for generously sharing their mutation data. NATO support through CRG Project #951420 is gratefully acknowledged.

References

- Bahar, I. & Jernigan, R. L. (1997). Inter-residue potentials in globular proteins and the dominance of highly specific hydrophilic interactions at close separation. *J. Mol. Biol.* **266**, 195-214.
- Bahar, I. & Jernigan, R. L. (1998). Vibrational dynamics of transfer RNAs: comparison of the free and synthetase bound forms. *J. Mol. Biol.* **281**, 871-884.
- Bahar, I., Atilgan, A. R. & Erman, B. (1997). Direct evaluation of thermal fluctuations in proteins using a single parameter harmonic potential. *Fold. Des.* **2**, 173-181.
- Bahar, I., Atilgan, A. R., Demirel, M. C. & Erman, B. (1998a). Vibrational dynamics of folded proteins: significance of slow and fast modes in relation to function and stability. *Phys. Rev. Letters*, **80**, 2733-2736.
- Bahar, I., Wallqvist, A., Covell, D. G. & Jernigan, R. L. (1998b). Correlation between native state hydrogen exchange and cooperative residue fluctuations from a simple model. *Biochemistry*, **37**, 1067-1075.
- Boyer, P. L., Ferris, A. L. & Hughes, S. H. (1992). Cassette mutagenesis of the reverse transcriptase of the human immunodeficiency virus type 1. *J. Virol.* **66**, 7533-7537.

- Boyer, P. L., Ferris, A. L., Clark, P., Whitmer, J., Frank, P., Tantillo, C., Arnold, E. A. & Hughes, S. H. (1994). Mutational analysis of the fingers and palm subdomains of human immunodeficiency virus type-1 (HIV-1) reverse transcriptase. *J. Mol. Biol.* **243**, 472-483.
- Cohen, J. (1997). The daunting challenge of keeping HIV suppressed. *Science*, **277**, 32-33.
- Covell, D. G., Smythers, G. W., Gronenborn, A. M. & Clore, A. M. (1994). Analysis of hydrophobicity in the alpha and beta chemokine families and its relevance to dimerization. *Protein Sci.* **3**, 2064-2072.
- Covell, D. G., Jernigan, R. L. & Wallqvist, A. (1998). Structural analysis of inhibitor binding to HIV-1 protease: Identification of a common binding motif. *Theoret. Chem.* **423**, 93-100.
- Das, K., Ding, J., Hsiou, Y., Clark, A. D., Jr., Moereels, H., Koymans, L., Andries, K., Pauwels, R., Janssen, P. A. J., Boyer, P. L., Clark, P., Smith, R. H., Jr, Smith, M. B. K., Michejda, C. J., Hughes, S. H. & Arnold, E. (1996). Crystal structures of 8-CI and 9-CI TIBO complexed with wild type HIV-1 RT and 8-CI TIBO complexed with the Tyr181Cys HIV-1 RT drug-resistant mutant. *J. Mol. Biol.* **264**, 1085-1100.
- De Clercq, E. (1994). HIV resistance to reverse transcriptase inhibitors. *Biochem. Pharmacol.* **47**, 155-169.
- De Clercq, E. (1995a). Antiviral therapy for human immunodeficiency virus infections. *Clin. Microbiol. Rev.* **8**, 200-239.
- De Clercq, E. (1995b). Toward improved anti-HIV chemotherapy: therapeutic strategies for intervention with HIV-1 infections. *J. Med. Chem.* **38**, 2491-2517.
- Demirel, M. C., Atilgan, A. R., Jernigan, R. L., Erman, B. & Bahar, I. (1998). Identification of kinetically hot residues in proteins. *Protein Sci.* **7**, 2522-2532.
- Ding, J., Das, K., Moereels, H., Koymans, L., Andries, K., Janssen, P. A. J., Hughes, S. H. & Arnold, E. (1995a). Structure of HIV-1 RT/TIBO R 86183 reveals similarity in the binding of diverse non-nucleoside inhibitors. *Nature Struct. Biol.* **2**, 407-415.
- Ding, J., Das, K., Tantillo, C., Zhang, W., Clark, A. D., Jr, Jessen, S., Lu, X., Hsiou, Y., Jacobo-Molina, A., Andries, K., Pauwels, R., Moereels, H., Koymans, L., Janssen, P. A. J. & Smith, R. H., Jr *et al.* (1995b). Structure of HIV-1 reverse transcriptase in a complex with the nonnucleoside inhibitor α -APA R 95845 at 2.8 Å resolution. *Structure*, **3**, 365-379.
- Esnouf, R., Ren, J., Ross, R., Jones, Y., Stammers, D. & Stuart, D. (1995). Mechanism of inhibition of HIV-1 reverse transcriptase by nonnucleoside inhibitors. *Nature Struct. Biol.* **2**, 303-308.
- Flory, P. J. (1976). Statistical thermodynamics of random networks. *Proc. Roy. Soc. ser. A*, **351**, 351-380.
- Frauenfelder, H., Petsko, G. A. & Tsernoglou, D. (1979). Temperature-dependent X-Ray diffraction as a probe of protein structural dynamics. *Nature*, **280**, 558-563.
- Gao, H., Boyer, P. L., Arnold, E. & Hughes, S. H. (1998). Effects of mutations in the polymerase domain on the polymerase, RNase H, and strand transfer activities of human immunodeficiency virus type 1 reverse transcriptase. *J. Mol. Biol.* **277**, 559-572.
- Garcia, A. E., Krumhansl, J. A. & Frauenfelder, H. (1997). Variations on a theme by Debye & Waller: from simple crystals to proteins. *Proteins: Struct. Funct. Genet.* **29**, 153-160.
- Haliloglu, T., Bahar, I. & Erman, B. (1997). Gaussian dynamics of folded proteins. *Phys. Rev. Letters*, **79**, 3090-3093.
- Hsiou, Y., Ding, J., Das, K., Clark, A. D. J., 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.
- Jacobo-Molina, A., Ding, J., Nanni, R. G., Clark, A. D. J., Lu, X., Tantillo, C., Williams, R., Kamer, G., Ferris, A. L., Clark, P., Hizi, A., Hughes, H. S. & 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.
- Jäger, 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.
- Jernigan, R. L. & Bahar, I. (1996). Structure-derived potentials and protein simulations. *Curr. Opin. Struct. Biol.* **6**, 195-209.
- Kloczkowski, A., Mark, J. E. & Erman, B. (1989). Chain dimensions and fluctuations in random elastomeric networks. 1. Phantom Gaussian networks in the undeformed state. *Macromolecules*, **22**, 1423-1432.
- 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.
- Miyazawa, S. & Jernigan, R. L. (1985). Estimation of effective interresidue contact energies from protein crystal structures: Quasi-chemical approximation. *Macromolecules*, **18**, 534-552.
- Miyazawa, S. & Jernigan, R. L. (1996). Residue-residue potentials with a favorable contact pair term and an unfavorable high packing density term, for simulation and threading. *J. Mol. Biol.* **256**, 623-644.
- Nanni, R. G., Ding, J., Jacobo-Molina, A., Arnold, E. & Hughes, S. H. (1993). Review of HIV reverse transcriptase three-dimensional structure: implications for drug design. *Pers. Drug Dis. Des.* **1**, 129-150.
- Pearson, D. S. (1977). Scattered intensity from a chain in a rubber network. *Macromolecules*, **10**, 696-701.
- Ren, J., Esnouf, R., Garman, E., Somers, D., Ross, C., Kirby, I., Keeling, J., Darby, G., Jones, Y., Stuart, D. & Stammers, D. (1995). High resolution structures of HIV-1 RT from four RT-inhibitor complexes. *Nature Struct. Biol.* **2**, 293-308.
- Rodgers, D. W., Gamblin, S. J., Harris, B. A., Ray, S., Culp, J. S., Hellmig, B., Woolf, D. J., Debouck, C. & 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.
- Skalka, A. M. & Goff, S. P. (1993). *Reverse Transcriptase*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Tantillo, C., Ding, J., Jacobo-Molina, A., Nanni, R. G., Boyer, P. L., Hughes, S. H., Pauwels, R., Andries, K., Janssen, P. A. J. & Arnold, E. (1994). Location of anti-AIDS drug binding sites and resistance mutations in the three-dimensional structure of HIV-1 reverse transcriptase: implications for mechanisms of drug inhibition and resistance. *J. Mol. Biol.* **243**, 369-387.

- Tirion, M. M. (1996). Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Phys. Rev. Letters*, **77**, 1905-1908.
- Varmus, H. (1988). Retroviruses. *Science*, **240**, 1427-1435.
- Wallqvist, A., Jernigan, R. L. & Covell, D. G. (1995). A preference-based free energy parameterization of enzyme-inhibitor binding. Applications to HIV-1-protease inhibitor design. *Protein Sci.* **4**, 1881-1903.
- Weiss, R., Teich, N., Varmus, H. E. & Coffin, J. (1993). *RNA Tumor Viruses*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Edited by F. Cohen

(Received 21 May 1998; received in revised form 20 October 1998; accepted 21 October 1998)