

## Relating the Structure of *HIV-1 Reverse Transcriptase* to Its Processing Step

<http://www.adeninepress.com/jbsd>

### Abstract

By treating an enzyme as a coarse-grained uniform block of material, utilizing only the  $\alpha$ -Carbon positions, the normal modes of motion can be obtained. For reverse transcriptase the slower of these motions are suggestive of being involved in the processing step, where the RNA or DNA strand is copied onto a new DNA strand at a polymerase site, and the RNA strand is subsequently cut up at the distant Ribonuclease H site. The slowest mode of motion involves hinge bending about a site midway between the polymerase and Ribonuclease H sites, suggesting that it can push or pull the RNA strand between these two sites. Pulling the nucleic acid strand would require tight binding to the RNase H site. The next slowest mode involves a hinge that opens and closes the protein like a clamp, which could facilitate the release of the nucleic acids for their step-wise progression. The third mode could rotate the substrate. An overall description of the step-wise processing step would involve close coordination among these steps. Results suggest that the smaller p51 subunit serves only as ballast to support the various modes of motion involving the different parts of the p66 subunit.

### Introduction

When the first protein structures were being determined, the common view was that these structures would inform us directly about all details of molecular function and dynamics. At the present time, when thousands of protein structures are known, it is clear that this earlier view was naïve. Although, in principle, molecular dynamics is a method that could provide the missing link between structure and function, it has not informed us well about the largest-scale motions. It is clear that there is still a serious need for new approaches to infer functions from structures, especially those involving large-scale motions. The results described here are an effort in this direction.

We follow a coarse-grained approach wherein atomic details are not directly accounted for, but instead we utilize only residue positions and their packing densities. It is a homogeneous model in the sense that all interactions between residues are treated identical in form and in value. Only the geometry represented by the positions of the residues is considered, and pairwise connections are formed between all residue pairs close to one another in space, whether sequentially connected or not.

We have been applying this simple model (1-3) to investigate fluctuations in proteins and the normal modes of motion with the aim of uncovering motions related to function (4-7). These fluctuations are quite small in magnitude, and consequently are not necessarily representative of all motions in the protein, but still they do provide a representative sample of motions that do include long-range effects. By investigating these normal modes of motion, we can observe long distance correlations between distant parts of the protein. These include even the extremely large-scale modes, such as the hinge motions in immunoglobulins (unpublished), topoi-

**R. L. Jernigan<sup>1\*</sup>,**  
**Ivet Bahar<sup>1,2,</sup>**  
**David G. Covell<sup>3,</sup>**  
**Ali Rana Atilgan<sup>2,</sup>**  
**Burak Erman<sup>4,</sup>**  
**and Daniel T. Flatow<sup>1</sup>**

<sup>1</sup>Molecular Structure Section,  
Laboratory of Experimental  
and Computational Biology,  
Division of Basic Sciences,  
National Cancer Institute,  
National Institutes of Health,  
MSC 5677,

Bethesda, MD 20892-5677

<sup>2</sup>Polymer Research Center  
and School of Engineering,  
Bogazici University,  
and TUBITAK Advanced Polymeric  
Materials Research Center,  
Bebek 80815, Istanbul, Turkey

<sup>3</sup>Frederick Cancer Research  
and Development Center,  
National Cancer Institute,  
Science Applications  
International Corporation,  
Frederick MD 21702, USA

<sup>4</sup>Sabanci University,  
Sabanci Center,  
Istanbul 80745, Turkey

\*Phone: (301) 496-4783;  
Fax: (301) 402-4724;  
E-mail: Robert\_Jernigan@nih.gov

somerase II (7), and reverse transcriptase (6). This simple calculation is feasible for extremely large proteins or molecular assemblages. As an example, we are routinely performing calculations on the motions of GroES/GroEL, having more than 8015 residues (unpublished).

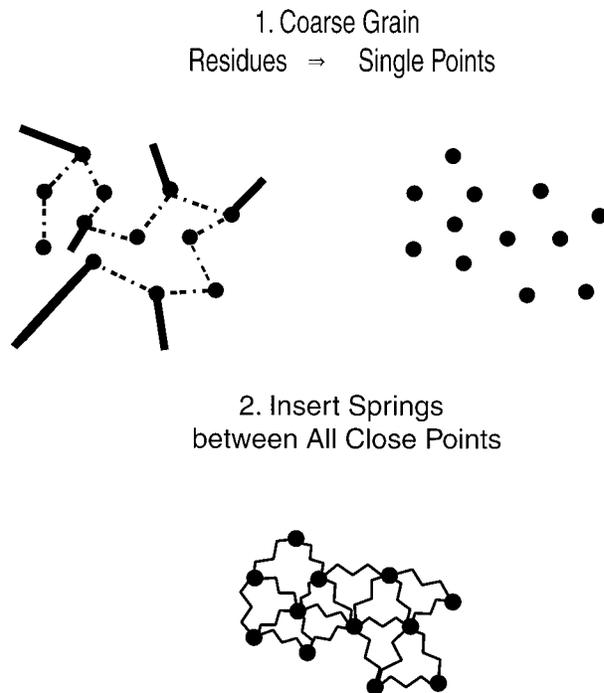
The approach utilized here has several premises: 1) the geometry manifested in residue packing in the native state is energetically optimal so that any residues' deviations from these positions are appropriately penalized with identical symmetric Gaussian energy functions, 2) the details of the packing of residues determines the large-scale, long-range hinge motions of proteins, 3) the motions can be decomposed with a normal mode-type of analysis to yield information about a full range of motions, from the small-scale to the largest-scale, and 4) correlations observable for the mean-square fluctuations represent the couplings between different parts of the protein structure in these motions. With this method it is possible to obtain useful information particularly about the large scale motions, and these do not depend on atomic details, but instead only on the coarse grained structure at the residue level. Consequently the large-scale motions should be more reliable than the small-scale ones which are naturally more dependent on the atomic details.

#### Description of Protein Dynamics Using the Gaussian Network Model

The most important underlying assumption with this GNM approach is that the protein in its native state, as determined by crystallography, is in its lowest energy form. The protein structure is modeled as being equivalent to a three-dimensional elastic network, where the  $C^\alpha$  atoms are taken as the junctions in the network. Residue positions fluctuate, in small magnitude motions, under the joint influence of all interactions between all close residues; thus the protein is reduced to a highly cooperative set of interconnected Gaussian springs. See Figure 1 for a diagram of the method. Each contact interaction is represented by an identical harmonic potential so that any deviations of the same magnitude from the original positions in the original structure (native state) will have identical energy penalties. The total internal energy of the protein then is given as

$$E = \frac{1}{2} \gamma \text{Tr} (\Delta \mathbf{R}^T \Gamma \Delta \mathbf{R}) \quad [1]$$

where  $\gamma$  is the single force constant of the Hookean spring between interacting residues (originally proposed by Tirion (8)),  $\Delta \mathbf{R}$  is the N-dimensional vector formed from the individual residue fluctuation vectors of all  $C^\alpha$  atoms  $\Delta \mathbf{R}_1, \Delta \mathbf{R}_2, \dots, \Delta \mathbf{R}_N$ , the superscript T indicates the matrix transpose, and  $\text{Tr}$  designates the matrix trace.  $\Gamma$  is like a contact matrix and is the counterpart of the *stiffness matrix* used in the analysis of elastic bodies (9). It is an  $N \times N$  symmetric matrix for a protein of N residues, whose  $ij^{\text{th}}$  element is 1 if residues  $i$  and  $j$  are in contact, and zero otherwise. The diagonal elements are taken to be the negative sum of the off-diagonal elements in the same column, i.e.  $\Gamma_{ii} = -\sum_j \Gamma_{ij}$  where  $j$  is not equal to  $i$ , similar to transition rate matrices for stochastic processes.



**Figure 1:** The model utilized for calculations of fluctuations. In the coarse graining,  $C^\alpha$  coordinate positions only are chosen, and then positions close in space ( $< 7\text{\AA}$ ), whether bonded or not are connected with Gaussian springs to penalize deviations from the crystal positions. The model thus consists of a highly cooperative set of coupled springs. The normal modes of this system are examined to investigate particularly the long-range correlations in motion.

The equilibrium correlations between the fluctuations of two residues  $i$  and  $j$  are then given by (1,9)

$$\langle \Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_j \rangle_k = (1/Z_N) \int \Delta \mathbf{R}_i \cdot \mathbf{R}_j \exp(-E/kT) d\{\Delta \mathbf{R}\} = (3kT/\gamma) [\Gamma^{-1}]_{ij} \quad [2]$$

where  $Z_N$  is the configuration integral. Note that the matrix  $\Gamma$  has a zero eigenvalue that must be removed (1-7). The inverse  $\Gamma^{-1}$  shown above is a partial inverse of this modified matrix rather than the inverse of  $\Gamma$  itself.

$$Z_N = \int \exp(-E/kT) d\{\Delta \mathbf{R}\}$$

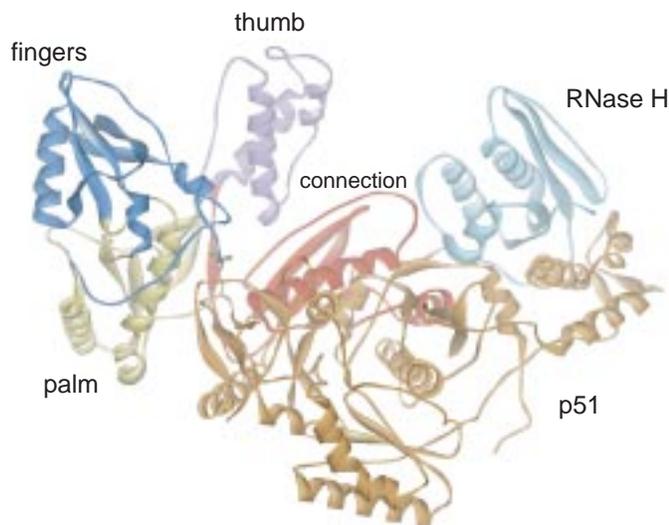
with  $k$  the Boltzmann constant,  $T$  the absolute temperature, and the integrations are carried out over all residue fluctuations in the set  $d\{\Delta \mathbf{R}\}$ , i.e.,  $d\Delta \mathbf{R}_1 d\Delta \mathbf{R}_2 \dots d\Delta \mathbf{R}_N$ .

The characteristic protein hinge motions are described in terms of frequencies and shapes of the corresponding modes of motion. The former are related to the eigenvalues  $\lambda_i$ ,  $2 \leq i \leq N$  of the connectivity matrix  $\Gamma$ , excluding the zero eigenvalue  $\lambda_1$ , and the latter to the eigenvectors  $\mathbf{u}_i$ ,  $2 \leq i \leq N$ . The cross-correlations  $\langle \Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_j \rangle_k$  for the  $k^{\text{th}}$  mode are found from (3)

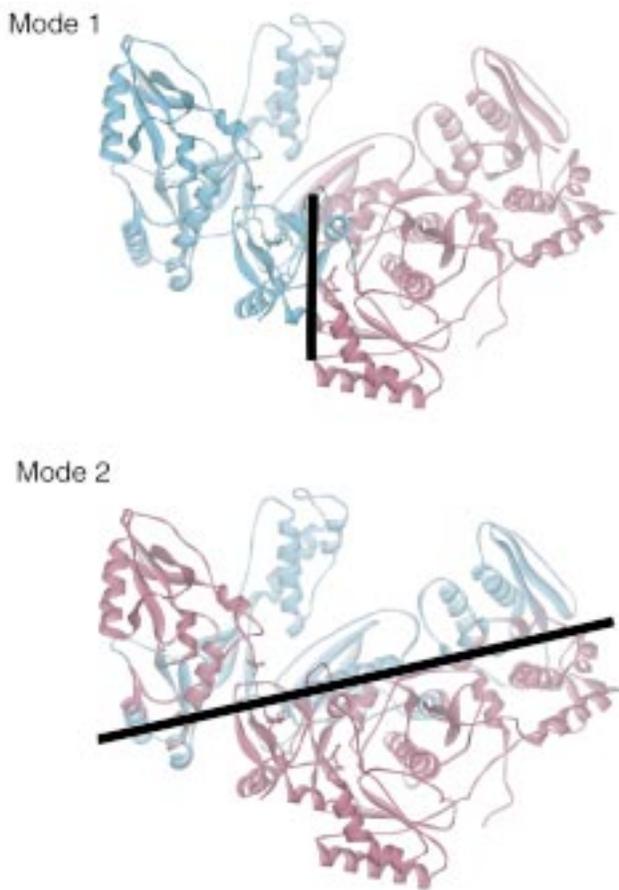
$$\langle \Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_j \rangle_k = (3kT/\gamma) [\lambda_k^{-1} \mathbf{u}_k \mathbf{u}_k^T]_{ij} = (3kT/\gamma) \lambda_k^{-1} [\mathbf{u}_k]_i [\mathbf{u}_k]_j \quad [3]$$

where all subscripts designate elements of the matrices and vectors.

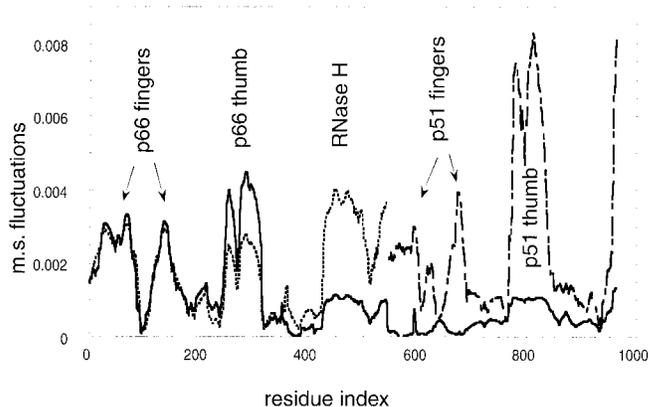
Because they are more likely to be related to the processing motions of reverse transcriptase, we consider here the motions



**Figure 2:** Ribbon diagram of HIV-1 RT heterodimer. The p66 subunit has two principal domains, the polymerase site including the fingers, palm and thumb and the distant RNase H site. The polymerase site of the p66 subunit has the following separate elements: fingers (dark blue, residues 1-88 & 121-146), thumb (purple, 243-311), palm (yellow), and connection (red, 312-425). The RNase H domain (426-560) is shown in green. The p51 subunit (orange-brown) has the same subdomains, fingers, palm, thumb and connection of the polymerase site of p66, but in a closed form without a nucleic acid binding cleft.



**Figure 4:** Ribbon diagram displaying the structural blocks moving as intact units in the two slowest modes of motion. The purple and green colors indicate the coherent blocks that move in anti-correlated directions by rotation about a single axis. Axes of rotation are drawn in approximate positions. Only modes 1 and 2 have a single rotation axis; higher modes have increasingly larger numbers of rotation axes.



**Figure 3:** The normalized m.s. fluctuations of residues for the slowest two modes found with the Gaussian network model (GNM). Results are for each residue. The dotted and dashed curves represent the behavior of the subunits as isolated monomers and the solid curve that of the heterodimer. The palm and connection subdomains are severely restricted, and the fingers and thumb are quite flexible as seen on the dotted curve. Likewise, the p51 monomer exhibits broad peaks at the fingers and thumb subdomains, as shown in dashes. The solid curve reveals the striking suppression of mobility in all parts of the p51 subunit in the heterodimer.

of the largest-scale, slowest, low frequency modes that we next discuss in some detail.

## Results

### The Reverse Transcriptase Structure

The structure of HIV-1 Reverse Transcriptase (here taken from pdb file 1rth (10,11)) is shown in Figure 2. The enzyme has two subunits, p66 shown at the top in blue, yellow, purple, red and green and p51 shown in orange-brown near the bottom. From an examination of crystal forms of HIV-1 RT, the polymerase domain has the anthropomorphic shape of a hand, having subdomains of fingers, palm, thumb, and connection to the RNase H domain (Figure 2). p51 has the same subdomains, except that it lacks the RNase H portion. The fingers, thumb, palm, and connection of the two subunits can be approximately superpositioned, between the p66 and the p51 subunits. However, the tertiary packing of the subdomains within the subunits differs: p66 is an open hand, with a large cleft for binding double-stranded nucleic acids between the thumb and fingers, whereas the p51 is like a closed hand and is more compact, having no nucleic acid binding cleft (12).

### Motions of Reverse Transcriptase

The position of the double-stranded portion of the template-primer is believed to be controlled by the fingers, palm and thumb of p66 (13). Only indirect evidence for the flexible motions of the fingers, palm and thumb subdomains exists from a comparative study of three crystal forms of HIV-1 RT (14). Differences in local structure suggested a hinge-bending motion between the fingers and palm subdomains of p66, and the remainder of the molecule, which was described qualitatively as

a swivel twist of the thumb subdomain. Additional crystallographic structures of HIV-1 RT, either bound to nucleic acid template-primers (15) or to non-nucleoside inhibitors (11, 16-18), or absent ligands (19,20) indicate the flexibility of the p66 subunit. The p66 thumb particularly is suggested to be extremely flexible, based on its different conformations in the DNA-bound, inhibitor-bound and unliganded structures (20). The subunit p51, on the other hand, appears to be rigid having its internal flexibility severely restricted.

#### *Modes of Motion of Reverse Transcriptase*

$\alpha$ -carbon coordinates of the RT-nevirapine complex crystal structure determined to 2.2Å resolution (11) are used for constructing the Kirchhoff matrix. 17 C-terminal residues of p66 were not reported, as well as two terminal residues at both ends of p51, leading to  $\Gamma$  matrices of respective sizes 963, 543 and 420 for RT dimer, p66 and p51 subunits, respectively. The parameter  $\gamma$ , common to all residue pairs in the monomers and/or the dimer, is determined from a comparison of the results for RT with its crystallographic B factors  $B_i = 8\pi^2 \langle \Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_i \rangle / 3$  equal (6) to 0.8 RT/Å<sup>2</sup>. This value specifies the height of the curve, without affecting its shape.

Results are shown for the fluctuations of the two slowest modes along the sequence in Figure 3. The large fluctuations of the p51 fingers and thumb, evident when it is in the monomer form, are suppressed in the dimer form of the enzyme. Thus, p66 shows an intrinsic flexibility at the finger and thumb subdomains of the polymerase domain, as well as at the RNase H domain. Likewise, the p51 monomer exhibits broad peaks at the fingers and thumb subdomains. The most striking observation is the suppression of the mobility of the fingers and thumb subdomains of p51 upon dimerization. This contrasts the behavior of the same subdomains in p66. Interestingly, the mobility of p66 fingers is unaffected by dimerization, and the thumb's motility is even further enhanced. The RNase H portion exhibits some reduced flexibility, apparently resulting from its interaction with the p51 thumb. We propose that the residues presently identified to exhibit a substantial decrease in mobility in the dimer, compared to that in the monomers, play an important role in the stabilization of the heterodimer.

The two slowest modes of motion of the enzyme are shown in Figure 4. The slowest motion has a vertical axis through the center of the molecule. And the second slowest has a nearly horizontal axis through the molecule. These motions are simple and easily related to the stepwise processing of the RNA strand engaged in by the enzyme. The enzyme copies the NA strand into DNA at the polymerase site and then cuts up the RNA strand at the distant RNase H site. It is easy to envision the role of these two slowest modes in this process. The 1<sup>st</sup> mode could pull the nucleic acid strand through the polymerase site and the 2<sup>nd</sup> mode could release the strand for this pulling step. In this way, properly coordinating these two modes of motion could account for much of the stepwise processing of the enzyme.

#### *Putative Pulling of the RNA by Reverse Transcriptase*

In this case the correlations and anti-correlations between the various parts of the protein for the slowest mode (Table I) indicate a clear mechanism for the step-wise pulling of the RNA strand through the polymerase site. Interestingly, it would imply the critical importance of the NA binding strongly to the RNase H region as an anchoring site to effect the step-wise pulling of the NA strand. In addition there would be the necessity that the NA not be attached strongly to the connection region of the protein between the two domains, because this might prevent the requisite motions. A lack of contacts in the connection region is required in order for both the NA and the protein to move freely. There are some additional implications about the NA conformation itself that can be drawn: it must be under sufficient tension so that it can transmit the pulling motion between the two sites. This means that it cannot be a relaxed single strand form and loosely deployed like a loose rope between the two sites. Is a hybrid DNA-RNA or a DNA-DNA double helix sufficiently rigid? There are additional questions that can be raised about the nucleic acid strands' conformations. If it has a helical form, then translocation by one base might require some rotational motion. A rotational motion for the nucleic acid chain could be obtained from another mode of motion. The third slowest mode (Table III) has the fingers and RNase H portions moving in the same direction opposing the remainder of the p66 subunit which could result in a rotation by rolling these elements over the cylindrical double helix surface. Alternatively fixed interactions between the NA and protein could facilitate the rotation by a correlated motion of the pair of tongs corresponding to fingers and RNase H. One relevant mechanism is the "inchworm" model proposed for transcription elongation by Chamberlin (21).

#### *Alternative Pushing Mechanism for Reverse Transcriptase*

If the double helix is pushed and compressed instead, one immediately thinks of the well-known strong correlation in nucleic acid helices between elongation and twist. Is it possible that, instead of being pulled, the NA is compressed by pushing the RNase H site towards the polymerase site in Mode 1 (Table I, Figure 4), and that this compression step is transmuted into a change in twist at the polymerase site? The possibility of a B-form to A-form transition immediately arises, if both strands were DNA, but is not likely when one is RNA. Each base pair step compressed in this way would correspond to a compression of  $3.4\text{Å} - 2.9\text{Å} = 0.5\text{Å}$  (the difference in rise between B and A forms) and a rotation of  $35.9 - 33.1^\circ = 2.8^\circ$  (the difference in helical twist between B and A forms). Consequently in order to achieve a helical twist change corresponding to a one base pair step, would require a change of  $30^\circ - 36^\circ$ , so 10-13 base pair steps would need to undergo the transition. This in turn would correspond to a translation of 5 to 6.5Å. This amount of compression would be required in order to rotate the NA, in a helical form, from one base to the next base in the polymerase site. A subsequent relaxation back from A-form to B-form could occur as the two sites pull away from each other in the opposite direction. The effect of elongation on nucleic acid conformations



**Table VIII**

Correlations for Mode 8. The fingers-1, palm-1 and p51 subunit are not coherent. The fingers-2, connection and RNaseH oppose the palm-2 and thumb.

	F1	P1	F2	P2	T	C	H	P51
<b>F1</b>	0							
<b>P1</b>		0						
<b>F2</b>			-				-	
<b>P2</b>				+	+			
<b>T</b>				+	+			
<b>C</b>						-	-	
<b>H</b>							-	
<b>P51</b>								0

**Table IX**

Correlations for Mode 9. Neither the fingers, palm-1 of p66 or the p51 subunit are coherent. The palm-2, thumb and RNaseH move opposite to the connection subdomain.

	F1	P1	F2	P2	T	C	H	P51
<b>F1</b>	0							
<b>P1</b>		0						
<b>F2</b>			0					
<b>P2</b>				+	+		+	
<b>T</b>				+	+		+	
<b>C</b>						-		
<b>H</b>				+	+		+	
<b>P51</b>								0

**Table X**

Correlations for Mode 10. Most of the subdomains and the p51 subunit are not coherent. Only the thumb and connection subdomains are coherent and move together in opposition to portions of other subdomains.

	F1	P1	F2	P2	T	C	H	P51
<b>F1</b>	0							
<b>P1</b>		0						
<b>F2</b>			0					
<b>P2</b>				0				
<b>T</b>					-	-		
<b>C</b>					-	-		
<b>H</b>							0	
<b>P51</b>								0

tributors to the overall motions, since the mode contributions decrease with increasing mode number. Increasing numbers of rotation hinges appear upon moving to higher modes, so the highest modes will correspond to higher frequency motions. Also, the individual subdomains become increasingly fragmented for the higher frequency motions. It is quite remarkable that the p51 subunit combines in various ways with subdomains of the p66 subunit in their individual motions. In this way the p51 subunit can be considered to serve as ballast added to the individual parts of the p66 domain in various combinations, for the various modes of motion. Thus it controls (and slows) the rates of motion of these various modes. This is a direct result of the way in which the structure of the p66 subunit is splayed out in a rather extended way across the top of the p51 subunit.

## Discussion

It is noteworthy that these mechanistic considerations would not be possible from a simple direct examination of the overall

residue fluctuations, but only by decomposing the protein dynamics into a series of different frequency modes, and by concentrating on the slowest, largest amplitude, modes.

Enzyme mechanisms operative on a more local basis could also be studied by a similar approach, but this in general would require the use of atoms instead of residues for successful calculations. An ideal goal of these studies of enzymes would be to obtain a direct connection between the individual steps in the enzyme's mechanisms and the individual modes of motion of the protein, but it remains to be seen whether this is a sustainable goal. In the present case we are trying to learn about these individual steps.

Molecular dynamics simulations and normal mode analyses are methods commonly used for understanding the collective motions and correlations in proteins. However, applications of both of these methods to proteins are usually restricted only to the smallest proteins. In cases where the protein has more than about 300 residues, these two methods are prohibitively time-consuming. The present method substantially overcomes this limitation. Computationally, only the inversion of the Kirchoff matrix, which follows from the contact matrix, is required for obtaining the correlations, which takes about 15 minutes on a Silicon Graphics R8000 Workstation for the present molecule of nearly 1,000 residues.

The mode shapes illustrated in Figure 3 clearly demonstrate that in HIV-1 RT, the two regions undergoing the largest spatial excursions in the global dynamics of the enzyme are the fingers and thumb regions of the p66 subunit, because of their locations distant from the hinges in the slowest modes. Actually restricting the mobility of the p66 thumb upon binding of non-nucleoside inhibitors has been postulated as a mechanism of inhibition of RT activity. A general approach for inhibiting enzymes is suggested by the present study wherein the residues involved in the hinge motions would be bound together and not permitted to move when an inhibitor is bound. The present method may also be an appropriate general way for defining structural domains in proteins.

In summary this approach can be utilized to provide a highly mechanical model for the large-scale steps for an enzyme's action. In this paper we have glimpses of how the static protein structure implies the positions of rotational hinges that could correspond to these steps. This is an interesting and potentially significant way to obtain information about the motions of a protein from its static structure.

## References and Footnotes

1. I. Bahar, A. R. Atilgan, and B. Erman, *Fold. & Des.* 2, 173-181 (1997).
2. I. Bahar, A. Wallqvist, D. G. Covell and R. L. Jernigan, *Biochemistry* 37, 1067-1075 (1998).
3. T. Haliloglu, I. Bahar and B. Erman, *Phys. Rev. Lett.* 79, 3090-3093 (1997).
4. I. Bahar, A. R. Atilgan, M. C. Demirel and B. Erman, *Phys. Rev. Lett.* 80, 2733-2736 (1998).
5. I. Bahar and R. L. Jernigan, *J. Mol. Biol.* 281, 871-884 (1998).
6. I. Bahar, B. Erman, R. L. Jernigan, A. R. Atilgan and D. G. Covell, *J. Mol.*

- Biol.* 285, 1023-1037 (1999).
7. R. L. Jernigan, M. C. Demirel and I. Bahar, *Int. J. Quant. Chem. (B. Pullman Memorial Volume)* 75, 301-312 (1999).
8. M. M. Tirion, *Phys. Rev. Lett.* 77, 1905-1908(1996).
9. A. Kloczkowski, J. E. Mark and B. Erman, *Macromolecules* 22, 1423-1432 (1989).
10. A. M. Skalka and S. P. Goff, Eds., *Reverse Transcriptase*. (Cold Spring Harbor Laboratory Press, New York, 1993).
11. J. Ren, R. Esnouf, E. Garman, D. Somers, C. Ross, I. Kirby, J. Keeling, G. Darby, Y. Jones, D. Stuart and D. Stammers, *Nature Struct. Biol.* 2, 293-308 (1995).
12. R. G. Nanni, J. Ding, A. Jacobo-Molina, E. Arnold and S. H. Hughes, *Pers. Drug Dis. & Des.* 1, 129-150 (1993).
13. H. Gao, P. L. Boyer, E. Arnold and S. H. Hughes, *J. Mol. Biol.* 277, 559-572 (1998).
14. J. Jäger, S. J. Smerdon, J. Wang, D. C. Boisvert and T. A. Steitz, *Structure* 2, 869-876 (1994).
15. A. Jacobo-Molina, J. Ding, R. G. Nanni, A. D. J. Clark, X. Lu, C. Tantillo, R. Williams, G. Kamer, A. L. Ferris, P. Clark, A. Hizi, S. H. Hughes and E. Arnold, *Proc. Natl. Acad. Sci. USA* 90, 6320-6324 (1993).
16. J. Ding, K. Das, H. Moereels, L. Koymans, K. Andries, P. A. J. Janssen, S. H. Hughes, and E. Arnold, *Nature Struct. Biol.* 2, 407-415 (1995).
17. J. Ding, K. Das, C. Tantillo, W. Zhang, A. D. Clark Jr., S. Jessen, X. Lu, Y. Hsiou, A. Jacobo-Molina, K. Andries, R. Pauwels, H. Moereels, L. Koymans, P. A. J. Janssen, R. H. Smith Jr., M. K. Koepke, C. J. Michejda, S. H. Hughes and E. Arnold, *Structure* 3, 365-379 (1995).
18. K. Das, J. Ding, Y. Hsiou, A. D. Clark Jr., H. Moereels, L. Koymans, K. Andries, R. Pauwels, P. A. J. Janssen, P. L. Boyer, P. Clark, R. H. Smith Jr., M. B. K. Smith, C. J. Michejda, S. H. Hughes and E. Arnold, *J. Mol. Biol.* 264, 1085-1100 (1996).
19. D. W. Rodgers, S. J. Gamblin, B. A. Harris, S. Ray, J. S. Culp, B. Hellmig, D. J. Woolf, C. Deboucq and S. C. Harrison, *Proc. Natl. Acad. Sci. USA* 92, 1222-1226 (1995).
20. Y. Hsiou, J. Ding, K. Das, A. D. J. Clark, S. H. Hughes and E. Arnold, *Structure* 4, 853-860 (1996).
21. M. J. Chamberlin, in *The Harvey Lectures, Ser. 88*, 1-21 (1995).
22. W. K. Olson, K. M. Kosikov, A. Colasanti, A. A. Gorin and V. B. Zhurkin, in "Biological Physics: Third International Symposium", Eds. H. Frauenfelder, G. Hummer and R. Garcia, 14-36 (American Institute of Physics, NY, 1999).
23. K. M. Kosikov, A. A. Gorin, V. B. Zhurkin and W. K. Olson, W.K., *J. Mol. Biol.* 289, 1301-1326 (1999).