JCTC Journal of Chemical Theory and Computation

Computing the Amino Acid Specificity of Fluctuations in Biomolecular Systems

K. Hamacher*,†,‡ and J. A. McCammon†,‡,§

Center for Theoretical Biological Physics, Department of Chemistry and Biochemistry, Department of Pharmacology, and Howard Hughes Medical Institute, University of California at San Diego, La Jolla, California 92093-0365

Received October 7, 2005

Abstract: We developed a new amino acid specific method for the computation of spatial fluctuations of proteins around their native structures. We show the consistency with experimental values and the increased performance in comparison to an established model, based on statistical estimates for a set of test proteins. We apply the new method to HIV-1 protease in its wild-type form and to a V82F−I84V mutant that shows resistance to protease inhibitors. We further show how the method can be successfully used to explain the molecular biophysics of drug resistance of the mutant.

1. Introduction

Fluctuations of biomolecular complexes around their native states are of great importance in functional studies in molecular biophysics. One can deduce several features such as entropy changes upon binding,[1] probable binding sites for drugs,[2] or the overall stability and function of complexes[3] from the detailed analysis of those fluctuations.

Traditional normal-mode analysis (NMA)[4−6] utilizes the carefully developed force fields of molecular dynamics (MD) packages. Here, one minimizes the energy of the structure and diagonalizes the Hessian of the energy function in the minimum to obtain normal modes. One has, however, to bear in mind two shortcomings: first, NMA requires a lot of computational time in the minimization step, and second, the MD force fields are optimized to mimic several static and dynamical properties at the same time and, therefore, are not necessarily optimal for low-energy fluctuations around the native state.[7]

To tackle the first problem, one-parameter potentials were introduced: the Tirion potential[8] and the related Gaussian network model (GNM).[9] The sole parameter is here adjusted to reproduce experimental B factors, and one starts from an experimental protein structure without any minimization. The anisotropic network model (ANM)[10] and the GNM were thoroughly investigated with respect to molecular dynamics simulation[11] and proved to be in good agreement despite their simplicity.

NMA, GNM, and ANM were successfully applied to a variety of molecules such as tRNAs,[12] tryptophan synthase,[13] cowpea chlorotic mottle virus,[14] and tubulin.[3]

The success of the Tirion potential and the GNM stems from the observation that the large number of atomic contacts distributed in the protein makes the central limit theorem valid—leading to the mean strength γ in, for example, the Tirion potential $V = \sum_{(ij)\in\mathcal{C}}(\gamma/2)(s_{ij} - s_{ij}^{o})^2$ with $s_{ij} = |\vec{r}_{ij}|$, where $\vec{r}_{ij}$ is the distance vector between the C_α atoms of contacting residues i and j. Members of the set of contacting residues are defined by $\mathcal{C} = \{(i, j)|s_{ij}^{o} < R_c\}$ with some given R_c, and the superscript o indicates the native conformation of the N residues throughout this paper. These methods were extended in several aspects: larger systems,[15,16] anisotropy,[10] or side-chain geometries (βGM).[17] Additionally, investigations, for example, ref 18, included anharmonic effects.

These methods, however, still lack—because of the intrinsic averaging procedure—the ability to incorporate the chemical nature of the residues constituting the protein. Therefore, there is a gap between NMA based on empirical MD force fields, which provide for specificity, and the one-

* Corresponding author phone: 1-(858)-534-7356; e-mail: hamacher@ctbp.ucsd.edu; URL: http://www.kay-hamacher.de.

† Center for Theoretical Biological Physics.

‡ Department of Chemistry and Biochemistry.

§ Department of Pharmacology and Howard Hughes Medical Institute.

10.1021/ct050247s CCC: $33.50
Published on Web 04/15/2006

parameter models that provide for fast computations and, therefore, large-scale screening capabilities important for studies of stability and function as well as drug design. We fill this gap with the proposed method.

2. The Model

Let the potential consist of two terms: a Hookean pairwise potential for covalent bonds $(K/2)\sum_i(s_{i,i+1} - s^{o}_{i,i+1})^2$ and a contact potential that assigns to contacting residues i and j an amino acid type-specific energy κ_{ij} in the native structure, which is at the same time assumed to be the global minimum. Fluctuations can break contacts if they get larger than some value a. These constraints can be modeled (see ref 19 for a similar approach) by a harmonic term inside a well with width a: $(\kappa_{ij}/a^2)[((s_{ij} - s^{o}_{ij})^2 - a^2)\Theta(a^2 - (s_{ij} - s^{o}_{ij})^2)]$, where $\Theta(x)$ is the step function [$\Theta(x) = 1$ for $x \geq 0$ and is 0 otherwise].

In the native configuration $\vec{s}^{o}_{ij}$, this results in the value κ_{ij}. We will use this particular property later on.

If the typical displacement is smaller than the well's width, we can omit the step function and obtain

$$V = \alpha a^{-2}\left[\frac{a^2K}{2}\sum_i(s_{i,i+1} - s^{o}_{i,i+1})^2 + \sum_{(i,j)\in\mathcal{C}'}\kappa_{ij}(s_{ij} - s^{o}_{ij})^2\right] \quad (1)$$

where we dropped an unimportant constant and introduced the scaling factor α to fit experimental B factors later on. Here, $\mathcal{C}' = \mathcal{C}\backslash\{(i,j)||i-j| = 1\}$ is the set of noncovalent-bonded contacts.

We note in passing that starting with the bond-breaking model and deriving eq 1 makes the whole approach consistent with the ANM and the knowledge-based potentials we will use later on to parametrize κ_{ij}: the prefactor αa^{-2} is an overall fit parameter that the ANM chooses so that the average B factor equals the one of the experiment and the various κ_{ij}'s *weight* the different contacts against each other, while we do not have to demand that the κ_{ij}'s are exact energy values (an overall scaling factor might be included).

Because we are interested in fluctuations, we develop V to the second order around the native state, which is the global minimum of eq 1 generically. The constant and the linear term vanish, and we obtain $V \approx {}^1/_2\Delta\vec{R}^{T}\mathbf{H}\Delta\vec{R}$, where $\Delta\vec{R}$ is the $3N$-dimensional displacement vector of all residues. The matrix $\mathbf{H}$ is the Hessian of eq 1 with respect to all $3N$ coordinates. We omit the somewhat clumsy expressions[20] for brevity here.

By means of singular value decomposition,[21] we obtain the eigenvalues λ_k and the eigendirections $\vec{u}_k$ of the movements. We effectively extend the anisotropic network model[10] to describe amino acid specific interactions. We refer to this method as extended ANM (EANM) in this paper because we would regain the ANM potential by setting uniformly $(2\alpha\kappa_{ij}/a^2) = \alpha K = \gamma$. Spatial fluctuations can afterward be computed from the inverse of $\mathbf{H}$ by

$$\langle(R^{\alpha}_i)^2\rangle = 3k_{B}T\sum_k\frac{[\vec{u}_k\vec{u}^{T}_k]^{\alpha\alpha}_{ii}}{\lambda_k} \quad (2)$$

for $\alpha = x, y$, and z, and the B factor of residue i can be computed from $B_i = (8\pi^2/3)\sum_{\alpha=x,y,z}\langle(R^{\alpha}_i)^2\rangle$.

Here, we weight the interactions κ_{ij} inside one protein (hereafter called "intrachain") according to the absolute value of the established contact potential of Miyazawa and Jernigan,[22] while the κ_{ij}'s for interchain contact between residues in multiple-chain proteins (see below for the HIV-1 protease) are taken as the absolute values as suggested by Keskin et al.[23] The rationale behind this is the property of our model (see above) that, in the native state, the potential is just the sum of the contact energies κ_{ij} of all the contacts of the native state. The intra- and interchain contacts are weighted differently because there are subtle differences in the contact potentials for these two classes of contacts. The major physical difference between the two types of contacts is illustrated: (a) The packing density inside a single protein is higher; therefore, the statistics of contacts is different. (b) The "hydrophobic force" acts differently on residues of one chain during folding and residues in protein−protein-binding events.

The values of these knowledge-based potentials take various effects into account that are responsible for the observed frequencies of contacts: electrostatics, solvent effects, hydrogen bonds, and so on. Although, this choice of κ_{ij} is strictly valid only for ensembles in the Bethe approximation in which the formation of pair contacts is thought to be equivalent by a chemical reaction and thus governed by a reaction law for the association/dissociation of pairs of residues individually. The underlying principles of minimal frustration and self-averaging were first discussed in the works of Wolynes and others and Onuchic and others.[24,25] It turns out that this approximation is precise enough for our purposes (see section "Performance" below for a justification).

The hard-core repulsion of Miyazawa and Jernigan's contact potential[22] is not taken into account as we are dealing with the native fold of the protein that cannot show any clashes. This proved to be a reasonable choice in other applications of the contact potential values,[22] too.[26] The low-frequency movements we are interested in are also the ones with large-scale changes in the structure, thus tending to involve solvent effects that are included in κ_{ij}. The strength along the peptide bond K is a fitting parameter, while we adjust αa^{-2} to agree with experimental B factors in the spirit of the GNM. The fitting allows one to fix the energy scale, avoiding the setting of a "folding" temperature a priori as was discussed to be crucial for contact potentials.[22,27] Additionally, we compensate with the fitted K for the various temperatures used in experiments and the unknown well width a.

Because we start with an experimental structure and the κ_{ij}'s were derived also from those structures, we are consistent. Because of the experimental structure, we also do not have to worry about finding a "problematic" minimum as might happen in NMA. The necessary structure minimization step in NMA might lead to different configurations for different mutants—making the comparison of eigenvectors and the implied principal directions of motion difficult or impossible.

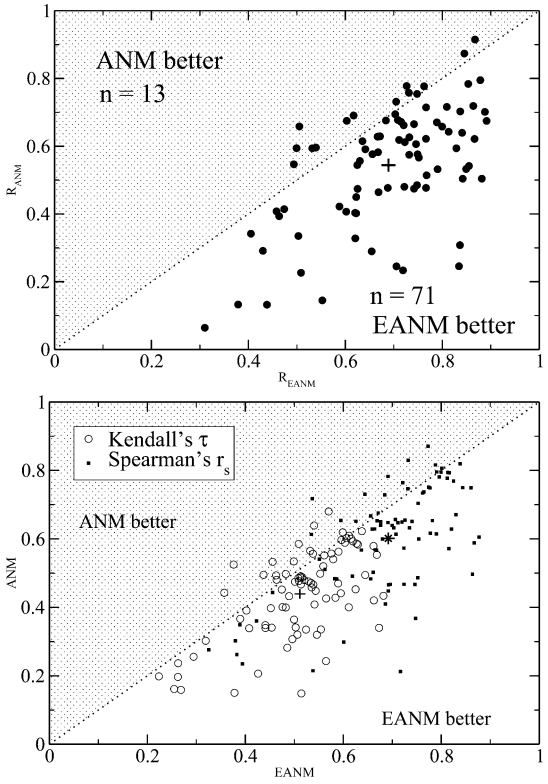


Figure 1. Graphs comparing the performance of EANM and ANM with respect to the set of test proteins. For every point, we computed the B factors and measured the agreement with three different data analysis methods. The upper graph shows the linear regression coefficient of the predicted vs the experimental values, while the lower uses two ranking measures. The crosses and the asterisk indicate the average over the set. The dashed lines are guides to the eye to distinguish the "winning method". The data points, for which ANM was better than EANM, are close to this line—indicating that EANM is even then not much off.

3. Performance

We applied EANM to 84 proteins[28] taken from the PDB database and set the contact defining distance to $R_c = 13$ Å. To reproduce the experimental B factors, we scanned for K of eq 1 in steps of five for every protein individually. We optimized the linear correlation between the experimental values and the computed ones.

In a recent coarse-grained study, a value of 83.333 RT/Å^2 was derived for the peptide bond.[29] In another study[30] on allostery of trypsiogen, Wall and Ming found $K = 53.4$ kcal/(mole Å^2) $= 89$ RT/Å^2 to be optimal. The obtained best K values here were found to be 82 ± 2.86 RT/a^2.

Because the all-atom average B factor of the proteins is 18.8 Å^2 (17.5 Å^2 for the C_α alone), an average displacement/resolution of 0.85 Å results. The K value and the displacement are both consistent with the above-mentioned value of Trylska et al. and Wall and Ming for $a \approx 1$ Å. The small standard deviation of K is in accordance with the small temperature range of experimental B factors.

The quality of every computation is measured by the linear correlation R, Spearman's r_s, and Kendel's τ[21] between the predicted and the experimental B factors. In Figure 1, we plotted for every protein the performance of the best K EANM computation in the x direction and the respective ANM result in the y direction.[31] Note that EANM is variational in the sense that increasing the accuracy of the K scan can only decrease the error with respect to the experimental values. Therefore, the results become, in general, better for smaller ΔK values, and EANM will perform even better than ANM.

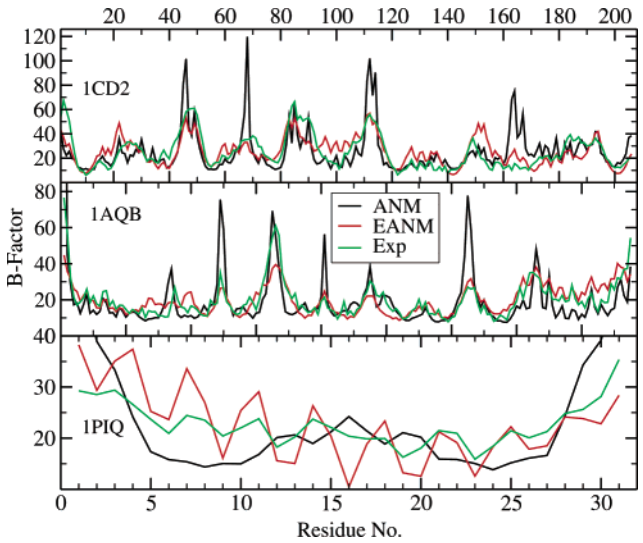


Figure 2. B factors for three proteins. Upper part: Note the failure of ANM for residues 162−168 and 68; both the experimental and the EANM values are much smaller. This is not just a common factor between those values and ANM, as can be seen from the areas around 20−35, 120−140, and 175−200. Middle: the ANM magnitudes are, at certain locations, good (e.g., 1−45 and 66). EANM is, however, capable of giving more-detailed insight over the whole range of the backbone. Lower part: EANM is not just lowering excessively high B factors but also reflects more of the underlying topology.

Figure 2 shows the B factors for three arbitrarily chosen proteins from the test set. Even under this crude K scan, not only is the agreement better but, additionally, structural features are reproduced in more detail with EANM.[20]

On the basis of the top two proteins of Figure 2, one might be tempted to attribute the increased performance of EANM most of all to lowering excessively high B values. To account for this, we first note that the ranking measure by Kendel and Spearman does not take into the account the overall magnitude of the B values, only their respective ranking—therefore, neglecting the order of magnitude of the values. Nevertheless, the performance of EANM is much better, giving evidence of the fact that EANM is indeed capable of giving a higher level of detail of the *shape* of the B values of the molecular system. A potential lowering of the B values can only result from the strengthening of the covalent bonds by larger K values (an order of magnitude larger than the κ_{ij} values). To test whether the increased performance of the EANM can be attributed to the larger K values alone, we repeated the computation in the test set with the best $K_{best} = 82RT$ and a uniform contact interaction $\kappa_{ij}^{uniform} = 3.563\ RT$, which is the average of the contact potential by Miyazawa

and Jernigan.[22] We found that, under all three approaches to measure the performance (linear correlation coefficient, Spearman's r_s, and Kendel's τ), the results were worse than those with the amino acid specific κ_{ij}. We thus conclude that the increased performance cannot be attributed to the lowering of excessively high B values or the stronger covalent bond alone.

4. Application to HIV-1 Protease

HIV-1 protease[32] is a crucial protein during HIV infection. Protease inhibitors bind to a "pocket" of this dimer and prevent further activity, thus reducing the spread of HIV viruses in an organism.

GNM was used by Kurt et al.[33] to investigate the effect of a ligand bound to the protein complex. Because the effects are only due to the change of the contact map for the complex upon binding the ligands, no amino acid specificity could be investigated.

Drug resistance due to mutation of the protease has, however, become a major problem[34] and prompts for sequence-specific studies. The dissociation rate of protease inhibitors is greatly influenced by the dynamics of the flaps of the pocket. A recent and comprehensive investigation[35] on several mutants of HIV-1 protease used normal-mode analysis. This, however, involves time-consuming minimization, energy function evaluation, and so forth, as discussed in the Introduction.

We want to show here how our model closes the gap between traditional NMA and GNM/ANM/βGM for this molecular complex. We used the experimentally determined wild-type structure[32] and applied the mutations V82F and I84V to be able to focus on those residues solely. We set the distance for which contacts are obtained to $R_c = 20$ Å because we found a more pronounced effect for this value.

Our setting is justified by the fact that, in the pocket, the small number of water molecules eventually bridges the spatial gap of size ~22 Å, extending the range of interaction partially throughout the pocket of HIV-1 protease as a water-molecule network, as recently discussed by Papoian et al.[36] Our choice of $R_c = 20$ Å, therefore, does not introduce contacts between the mutation sites and the flaps. All of the influence of the mutations is therefore carried by a mechanical mechanism encoded in the structure of the HIV-1 protease homodimer (see below for results on the mechanism). To confirm our choice further, we note that this R_c leads to the correct shape of the frequency distribution. The frequency distribution of a general protein motif was first obtained by Elber and Karplus[38] with the effective medium approximation. For the particular case of the HIV-1 protease, Zoete et al.[35] showed the contribution to the RMSD, which is, by means of eq 2, proportional to the square of the frequency—our results are in agreement with these RMSD contributions.

Our main results are depicted in Figures 3 and 4. The mutation of just two residues in each molecule of the protease dimer results in an increased movement of the flaps.

One can now use the eigenvectors to investigate the change in the motion of the residues upon mutation. One approach—the one we pursue in the following—is to look for orthogonality between modes in the different mutants for particular residues of interest. This allows an investigation of whether some residues are subject to larger effects than others.

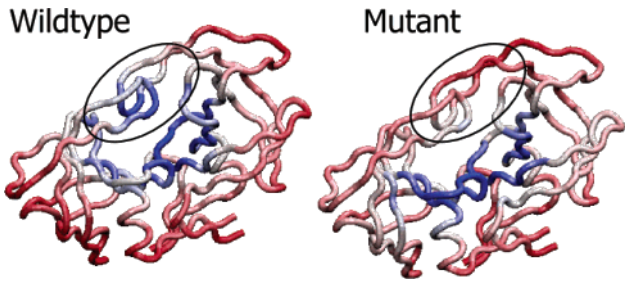


Figure 3. Structures of HIV-1 protease, its mutant, and the respective B factors. Overlaying the predicted B factors onto the structure of HIV-1 protease shows the increased flexibility of the "flaps". While the base of the pocket shows small fluctuations (blue), the mutant provides for the possibility of larger movements (red) of the "flaps". Note that the mutation occurs *on the base* of the pocket.

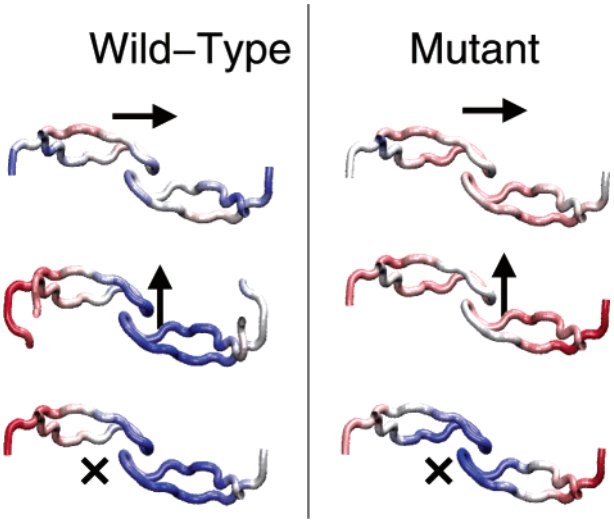
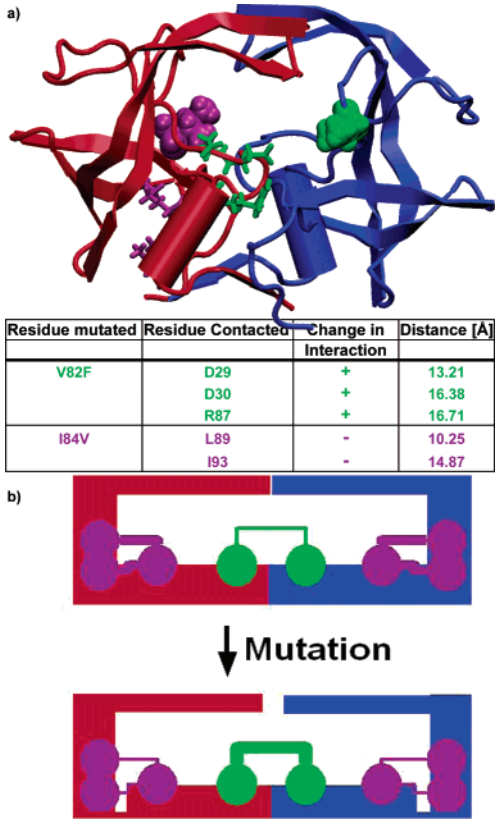


Figure 4. Anisotropic fluctuations for the HIV-1 protease in the directions indicated by the arrows. The mutation causes larger movements to open/curl the flaps, while the other two spatial directions are not affected on a larger scale (red, higher fluctuations; blue, lower fluctuations).

The residues 43−58 constitute the flaps in the homodimer of HIV-1 protease. Analyzing the overlap of all 594 obtained modes in the subspace of those residues between the wild-type and the mutant, we found that the 21st modes are the slowest mutual orthogonal ones.

These modes are $\sim\sqrt{4.1}$ faster than the overall slowest mode. The small ratio guarantees that this mode is substantially realized.[39] From eq 2, we can immediately conclude that the contribution of this mode to the overall B factor is $\approx 1/4.1 \approx 24\%$ of the first mode. The mode provides for the curling of the tips as a precursor of opening.[20] The curling was observed to be a prerequisite to opening the flaps.[37]

To elucidate the mechanism that results in increased floppiness of the flaps, we performed some thousand mutations studies: we changed the interaction strength between the contacts of residues 82 and 84 pairwise while leaving the rest alone. We therefore computed a "hybrid"



Residue mutated	Residue Contacted	Change in Interaction	Distance [Å]
V82F	D29	+	13.21
	D30	+	16.38
	R87	+	16.71
I84V	L89	-	10.25
	I93	-	14.87

Figure 5. (a) Residues that are mutated (spheres) and the respective residues they contact (sticks). These provided for most of the change in the spectrum. The table shows the residue numbers and the respective change in the interaction strengths upon mutation. In addition, we show the distances between residues which supports our choice of $R_c = 20$ Å here. (b) The emerging picture: upon mutation, the "base" becomes more stiff because of the increased strength of the bonds between the two monomers; at the same time, the flaps and their "anchors" are weaker bonded and, therefore, can move more freely in comparison to the base.

system, which contained the original contacts of residues 82 and 84 in the wild type to most of their contacting neighbors while changing κ_{ij} to the mutant values for pairs of contacting residues. This allows for a systematic investigation of the origin of the effect of the mutation. A *global* measure of the influence of such a hybrid mutation is the difference in the spectrum of the eigenvalues induced by the mutation. We extracted up to a noticeable gap in the above-mentioned difference the top of these pairs, which consisted remarkably of the same residues in different combinations. The results are shown in Figure 5.

We clearly see that the mutation stiffens the base while at the same time loosening its grip on the "anchors" of the flaps through which the loosened stiffness is then transferred to the flaps through the connections of contacts between residues in the anchor and the flaps. Note that the stiffening of the base occurs because of interaction between the two chains in the homodimer, while the weakening occurs in each monomer alone. Note also that this finding explains why our choice of $R_c = 20$ Å is sound here: the most important residues, D30 and R87, are farther apart than the generic distance of $R_c = 13$ Å.

In patients, the V82F−I84V mutant shows some additional mutations that seem, on the basis of our investigation, not necessary for the adaptation/evolution of HIV toward protease inhibitor resistance. They might, however, be necessary for protein stability or folding dynamics, issues that are beyond the scope of the present study.

5. Conclusions

In this study, we extended the widely used ANM method to allow for amino acid specificity. We justified the extended model, discussed its parametrization, and showed its superior performance on a set of proteins in comparison to the ANM. After fitting appropriate couplings for the bonded interactions, we applied the method to a problem most relevant to pharmacology and molecular biology: the drug resistance of a V82F−I84V mutant of HIV-1 protease. We showed the effect of this mutation on the movements of the "flaps" of this dimer. Time-consuming molecular dynamics were employed to reveal this mechanism in other studies,[37] while our approach for the protease took merely an hour of CPU time, including the refinement of *K*. We hope that our work facilitates further investigations into the molecular biophysics of HIV and other pharmacological problems.

Acknowledgment. K.H. is supported through a Liebig Fellowship of the Fonds der Chemischen Industrie. Other support has been provided by NSF, NIH, HHMI, CTBP, NBCR, W.M. Keck Foundation, and Accelrys, Inc. Stimulating discussions with A. Perryman and T. Hwa are gratefully acknowledged. Figures 3 and 4 were generated using VMD.[40] We are grateful to anonymous reviewers' comments and suggestions, which helped us to improve the presentation of our results and steered us to elucidate more on the mechanism of the mutation in the HIV-1 protease mutant.

References

(1) Lazaridis, T. *Curr. Org. Chem.* **2002**, *6*, 1319.

(2) Micheletti, C.; Cecconi, F.; Flammini, A.; Maritan, A. *Protein Sci.* **2002**, *11*, 1878.

(3) Keskin, O.; Durell, S.; Bahar, I.; Jernigan, R.; Covell, D. *Biophys. J.* **2002**, *83*, 663.

(4) Brooks, B.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **1983**, *80*, 6571.

(5) Brooks, B.; Janezic, D.; Karplus, M. *J. Comput. Chem.* **1995**, *16*, 1522.

(6) Kitao, A.; Go, N. *Curr. Opin. Struct. Biol.* **1999**, *9*, 164.

(7) Novotny, J.; Bruccoleri, R.; Karplus, M. *J. Mol. Biol.* **1984**, *177*, 787.

(8) Tirion, M. *Phys. Rev. Lett.* **1996**, *77*, 1905.

(9) Bahar, I.; Atilgan, A. R.; Erman, B. *Folding Des.* **1997**, *2*, 173.

(10) Atilgan, A.; Durrell, S.; Jernigan, R.; Demirel, M.; Keskin, O.; Bahar, I. *Biophys. J.* **2001**, *80*, 505.

(11) Doruker, P.; Atilgan, A. R.; Bahar, I. *Proteins: Struct., Funct., Genet.* **2000**, *40*, 512.

(12) Bahar, I.; Jernigan, R. L. *J. Mol. Biol.* **1998**, *281*, 871.

(13) Bahar, I.; Jernigan, R. *Biochemistry* **1999**, *38*, 3478.

(14) Tama, F.; Brooks, I. *J. Mol. Biol.* **2002**, *318*, 733.

(15) Tama, F.; Gadea, F.; Marques, O. *Proteins: Struct., Funct., Genet.* **2000**, *41*, 1.

(16) Doruker, P.; Jernigan, R.; Bahar, I. *J. Comput. Chem.* **2002**, *23*, 119.

(17) Micheletti, C.; Carloni, P.; Maritan, A. *Proteins* **2004**, *55*, 635.

(18) Miyashita, O.; Onuchic, J. N.; Wolynes, P. G. *PNAS* **2003**, *100*, 12570.

(19) This approximation is similar to Micheletti's self-consistent pair contact probability approximation.[41]

(20) Details to be published elsewhere.

(21) Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; Vetterling, W. T. *Numerical Recipies in C*; Cambridge University Press: Cambridge, England, 1995.

(22) Miyazawa, S.; Jernigan, R. L. *J. Mol. Biol.* **1996**, *256*, 623.

(23) Keskin, O.; Bahar, I.; Badretdinov, A.; Ptitsyn, O.; Jernigan, R. *Protein Sci.* **1998**, *7*, 2578.

(24) Bryngelson, J.; Wolynes, P. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *16*, 7524.

(25) Bryngelson, J.; Onuchic, J.; Socci, N.; Wolynes, P. *Proteins: Struct., Funct., Genet.* **1995**, *21*, 167.

(26) Shen, T.; Canino, L.; McCammon, J. *Phys. Rev. Lett.* **2002**, *89*.

(27) Miyazawa, S.; Jernigan, J. *Macromolecules* **1985**, *18*, 534. Miyazawa, S.; Jernigan, J. *Protein Eng.* **1994**, *7*, 1209. Gutin, A. M.; Badretdinov, A. Y.; Finkelstein, A. V. *Mol. Biol. (USSR)* **1992**, *26*, 94. Sali, A.; Shakhnovich, E.; Karplus, M. *J. Mol. Biol.* **1994**, *235*, 1614.

(28) We have chosen subsets of the well-known WHAT-IF and PDBSelect lists, namely, these proteins: 1a55, 1a7g (2 models), 1aba, 1abe, 1amp, 1aqb, 1aru, 1byb, 1c53, 1cd2, 1chd, 1cnv, 1coo, 1cot, 1ctf, 1ctt, 1d3z (10 models), 1dhr, 1eca, 1ede, 1enh, 1es6, 1esc, 1ezm, 1fkj, 1flp, 1fxd, 1gca, 1gof, 1hfc, 1hyp, 1iae, 1lid, 1mba, 1mjc, 1mml, 1npk, 1nxb, 1oyc, 1paz, 1pea, 1php, 1pht, 1piq, 1poa, 1ptq, 1r69, 1rcb, 1ris, 1sbp, 1tca, 1tde, 1ten, 1tig, 1ubq, 1xnb, 2cmd, 2cpl, 2ctc, 2ebn, 2end, 2i1b, 2liv, 2mcm, 2mnr, 2rn2, 2sas, 2sn3, 3dfr, 3il8, 3lzm, 451c, 4fgf, and 5p21. The sizes of these homomers range from 32 to 639 residues. We set $R_c = 13$ Å, as was determined to be optimal in general ANM applications.[10]

(29) Trylska, J.; Tozzini, V.; McCammon, J. Exploring Global Motions and Correlations in the Ribosome. *Biophys. J.* **2005**, *89*, 1455−1463.

(30) Preprint: qbio.BM/0506031. Ming, D.; Wall, M. E. *Phys. Rev. Lett.* **2005**, *95*, 198103.

(31) Note that by focusing on the linear correlation coefficient and the two ranking measures we avoid the adjustment of the overall magnitude completely.

(32) Reiling, K.; Endres, N.; Dauber, D.; Craik, C.; Stroud, R. *Biochemistry* **2002**, *41*, 4582.

(33) Kurt, N.; Scott, W.; Schiffer, C.; Haliloglu, T. *Proteins: Struct., Funct., Genet.* **2003**, *51*, 409.

(34) Little, S. J.; Holte, S.; Routy, J. P.; Daar, E. S.; Makrowitz, M.; Collier, A. C.; Koup, R. A.; Mellors, J. W.; Connick, E.; Conway, B.; Kilby, M.; Want, L.; Whitcomb, J. M.; Hellmann, N. S.; Richman, D. D. *N. Engl. J. Med.* **2002**, *347*, 385.

(35) Zoete, V.; Michielin, O.; Karplus, M. *J. Mol. Biol.* **2002**, *315*, 21.

(36) Papoian, G.; Ulander, J.; Wolynes, P. *J. Am. Chem. Soc.* **2003**, *125*, 9170.

(37) Perryman, A.; Lin, J.-H.; McCammon, J. *Protein Sci.* **2004**, *13*, 1108.

(38) Elber, R.; Karplus, M. *Phys. Rev. Lett.* **1986**, *56*, 394.

(39) A movie of this mode is available from the following URL: http://McCammon.ucsd.edu/khamache/EANM.html.

(40) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33.

(41) Micheletti, C.; Banavar, J. R.; Maritan, A. *Phys. Rev. Lett.* **2001**, *87*, 088102.

CT050247S