Proc. Natl. Acad. Sci. USA
Vol. 87, pp. 8864–8868, November 1990
Biophysics

Domain communication in the dynamical structure of human immunodeficiency virus 1 protease

W. E. HARTE, JR.*†, S. SWAMINATHAN*, M. M. MANSURI†, J. C. MARTIN†, I. E. ROSENBERG†, AND D. L. BEVERIDGE*

*Chemistry Department, Hall–Atwater Laboratories, Wesleyan University, Middletown, CT 06457; and †Pharmaceutical Research and Development Division, Bristol-Myers Squibb Company, 5 Research Parkway, Wallingford, CT 06492

Communicated by Frederic M. Richards, August 14, 1990 (received for review April 10, 1990)

ABSTRACT **A dynamical model for the structure of the human immunodeficiency virus 1 (HIV-1) protease dimer in aqueous solution has been developed on the basis of molecular dynamics simulation. The model provides an accurate account of the crystal geometry and also a prediction of the structural reorganization expected to occur in the protein in aqueous solution compared to the crystalline environment. Analysis of the results by means of dynamical cross-correlation coefficients for atomic displacements indicates that domain–domain communication is present in the protein in the form of a molecular "cantilever" and is likely to be involved in enzyme function at the molecular level. The dynamical structure also suggests information that may ultimately be useful in understanding and further development of specific inhibitors of HIV-1 protease.**

Human immunodeficiency virus 1 (HIV-1) protease is a dimer of 99-residue proteins which is vitally important to polyprotein processing in the life cycle of the AIDS virus (1). This enzyme is currently the only significant macromolecular component of the AIDS virus for which detailed structural information is available (2–4) and is a target of considerable pharmaceutical interest in the quest for AIDS therapies (5). The current structural model for HIV-1 protease developed from crystallography (3), depicted in Fig. 1, has established the essential secondary and tertiary structure of the enzyme. Homology with other aspartyl proteases (6) indicates that the protein dimer formed in the crystal is the native enzyme. The observed isotropic temperature factors indicate certain dynamical elements in the structure, particularly the flexibility in the flap regions opposite the active site Asp-Thr-Gly triads. The recent crystal structure of an HIV-1 protease–inhibitor complex (7) displays the binding site in detail and also provides evidence of substantial changes in the enzyme remote from the binding site. These changes emphasize the need for a consideration of structural dynamics in understanding HIV-1 protease action.

We describe herein a theoretical model for HIV-1 protease dimer in dilute aqueous solution developed from molecular dynamics simulation. Analysis of the results provides knowledge about the structure and functional energetics that originates uniquely in the dynamical motions, and in particular evidence for through-space interactions between domains of well-defined secondary structure.

METHODS AND CALCULATIONS

Molecular dynamics (MD) calculations in this study were performed with the Monte Carlo (MC) and MD simulation program WESDYN 1.0 (8), using the GROMOS 86 force field (9) and the simple point charge (SPC) model for water (10). The point of departure for this study was the x-ray crystal structure

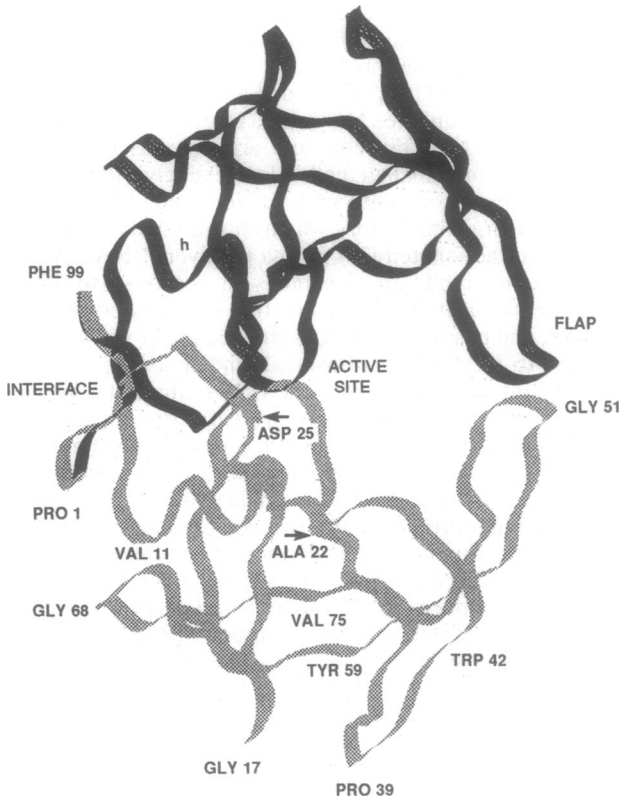

FIG. 1. Ribbon tracing of the x-ray crystal structure of HIV-1 PR dimer as reported by Wlodawer *et al.* (3). Names of regions and amino acid sequence numbers important for the discussion of dynamical structure in the protein are indicated.

of HIV-1 protease solved by Wlodawer *et al.* (3) and deposited in the Protein Data Bank (11) as 3HVP. The protein was solvated with 6990 molecules of water in a hexagonal prism cell and treated in the simulation under periodic boundary conditions. The object of study is thus a representation of a dilute aqueous solution of the protein dimer. The volume of the system was chosen to produce a density of 1 g/cm^3. The simulation protocol included an initial solvent relaxation of 3000 passes of MC (21×10^6 configurations), followed by 50 steps of conjugate gradient minimization on the total system. The MC process serves to equilibrate the solvent molecules in the potential field of the protein and obviates unrealistic fluctuations in kinetic energy at the onset of MD. The MD protocol involved a heating to 300 K over 1.5 psec, a Gaussian equilibration step of 2.5 psec, and a trajectory involving 96 psec of free MD.

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

Abbreviations: HIV-1, human immunodeficiency virus 1; MD, molecular dynamics; MC, Monte Carlo; DCCM, dynamical cross-correlation map.

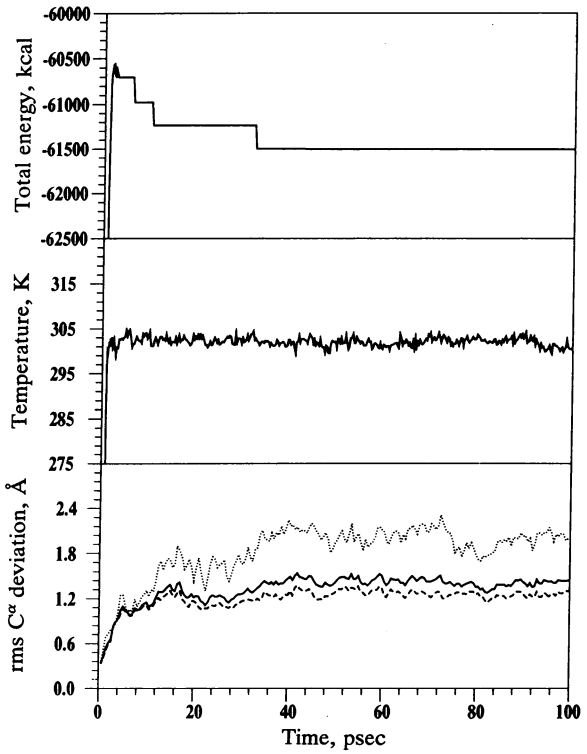

FIG. 2. Calculated total energy, temperature, and rms C^{α} deviation of dynamical structures from the 3HVP crystal geometry (3) vs. time from the MD simulation on HIV-1 PR. For the rms deviations: ——, all; ·····, flap; – – –, nonflap.

The convergence characteristics of the simulation are given in Fig. 2. The total energy and temperature were observed to be stable after equilibration, and velocity scaling beyond the 33-psec point of trajectory was unnecessary.

RESULTS AND DISCUSSION

Comparison of Calculated Results with the X-ray Crystal Structure. The rms deviation of the calculated MD structures from the crystal structure vs. time (Fig. 2 *Bottom*) stabilized at ≈1.3 Å. This relatively small deviation indicates that the dynamical structure of HIV-1 protease remained in the realm of the crystal geometry during the course of the simulation and is further evidence of the inherent stability of the model.

The overall motion and dynamical structure of the protein is presented in Fig. 3 as a superposition of nine snapshots obtained at equally spaced intervals from the MD trajectory. The dynamic range of the MD structures brackets the crystal geometry (bold broken line). The main discrepancy is found in the flap region. The position of the flap in the crystal structure was influenced by intermolecular contacts (3). In the simulation, the flap adjusts to a geometry predicted to be favored by aqueous hydration, forming hydrogen bonds to the flap of the other monomer unit.

A topological representation of the dynamical secondary structure is shown in Fig. 4. The backbone hydrogen bonds are indicated as solid lines, with a thickness proportional to the calculated average bond strength. Following the previously suggested descriptors of secondary structure in HIV-1 protease (3), the interface residues (1–5, 95–99) form a dovetailed β-pleated sheet, slightly tighter than that obtained by an energetic analysis of the crystal geometry in this region.

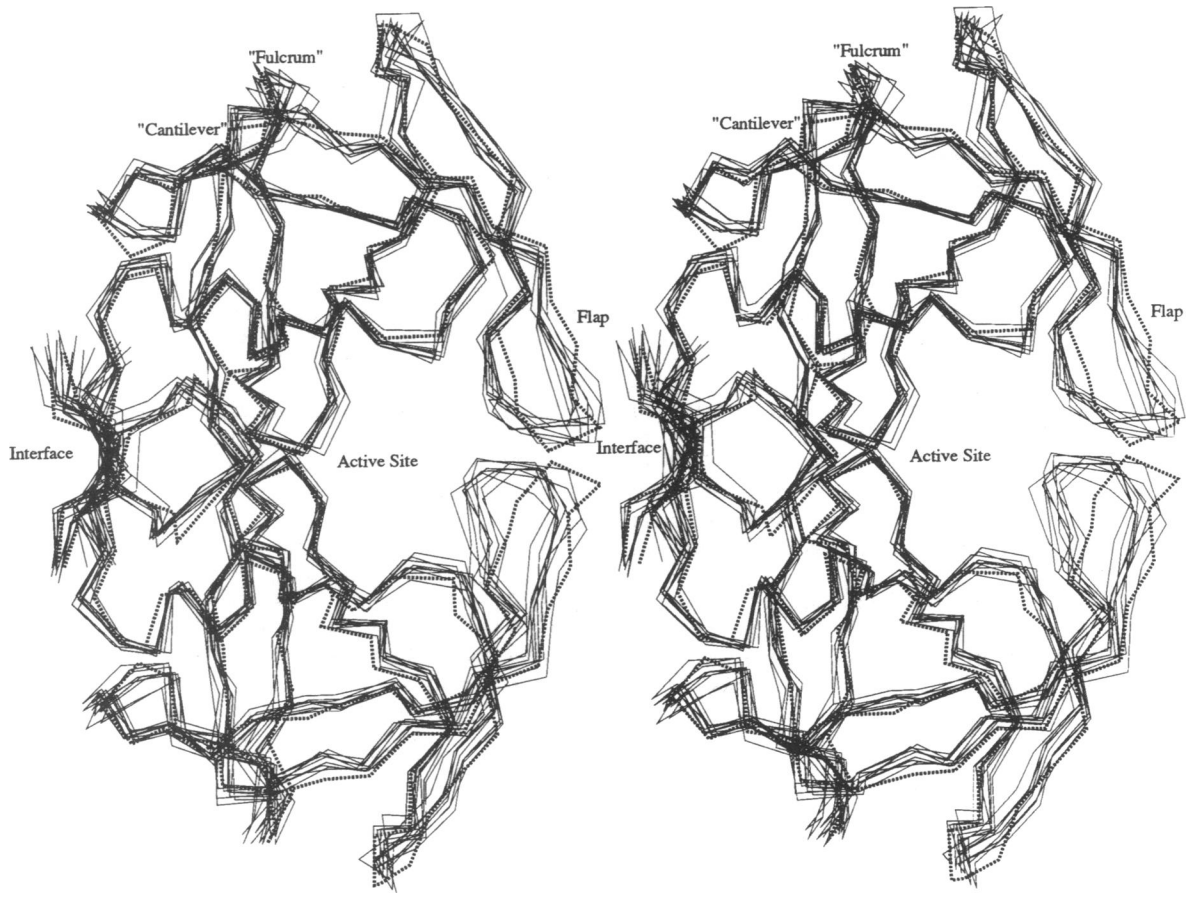

FIG. 3. Dynamical range of motion subsumed by the HIV-1 protease dimer over the interval of 40–100 psec in the MD simulation. This figure shows the superposition of C^{α} traces for nine snapshots equally spaced along the MD trajectory. The bold broken line is the C^{α} trace of the crystal geometry reported by Wlodawer *et al.* (3) and deposited in the Brookhaven Data Bank (11) as 3HVP.

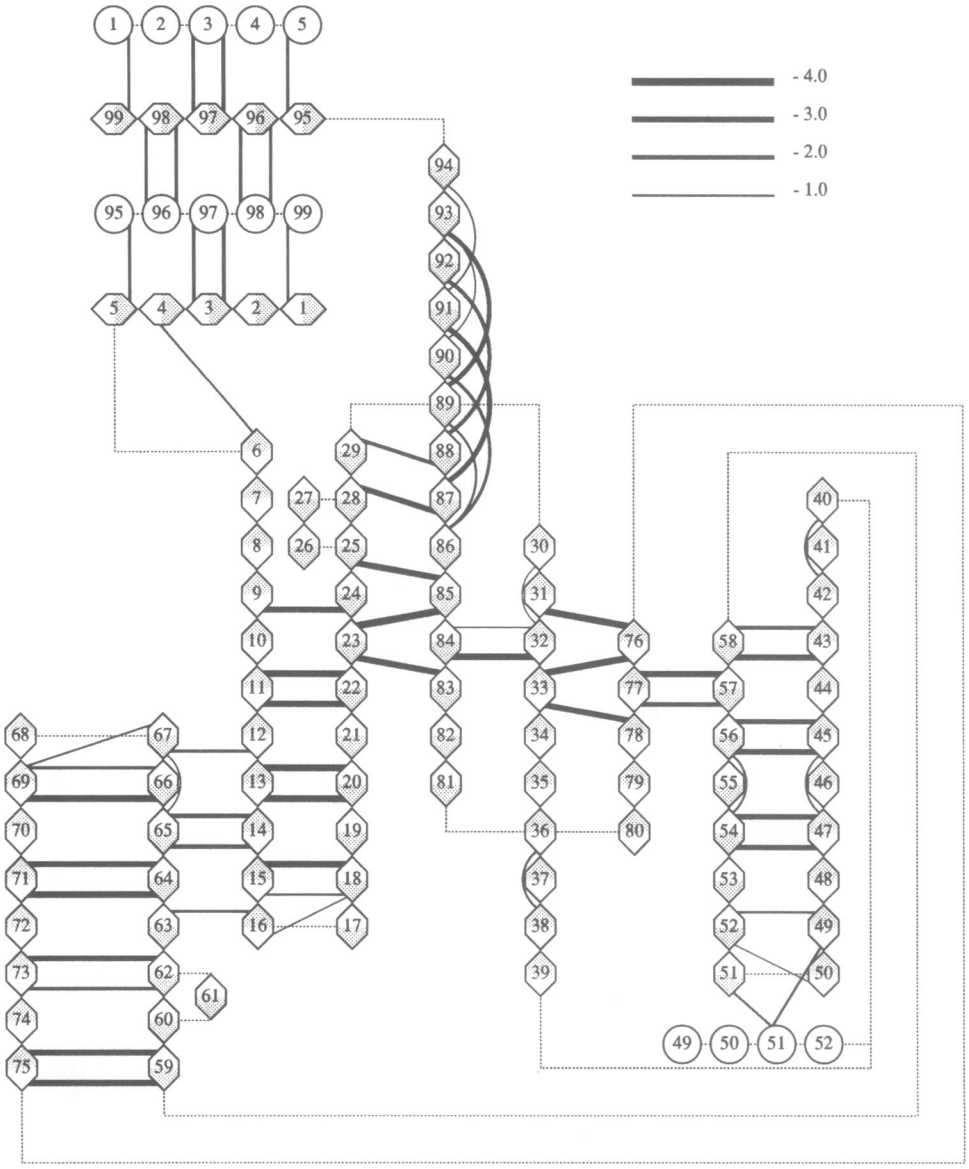

FIG. 4. Topology of the secondary structure and the calculated hydrogen bonding patterns for the backbone atoms of HIV-1 protease. Solid lines indicate the hydrogen bonds and their relative strengths; broken line indicates the backbone peptide bonds, with the residues in hexagons; residue numbers in circles are for the other chain in the dimer.

Residues 9–24 make up the b and c strands of an extended β-sheet. Note also the β-linkage between the active site (25–29) and the N terminus of the helix (86–94). The β-sheet in the flap region (42–58) remains regular, with lower end of the flap (49–52) forming weak but persistent intermolecular backbone hydrogen bonds with the analogous region of the other monomer.

An analysis of the calculated temperature factors (not shown) indicates regions of high and low thermal mobility. The flaps are flexible, but other regions of regular secondary structure show less mobility. The qualitative trends in the calculated vs. the observed temperature factors obtained from crystallography are in good accord but the latter are generally higher than the simulation values. The symmetry constraints in the solution of the crystal structure increase the reported thermal factors due to asymmetric motions.

Correlated Domain Motion and Molecular Communication. The dynamical characteristics of the protein in a MD simulation can be analyzed to yield information about correlated motions (12, 13). Correlated motions can occur among proximal residues composing well-defined domain regions of secondary structure and also between regions as in domain–domain communication. The extent of correlated motion is indicated by the magnitude of the corresponding correlation coefficient. The cross-correlation coefficient for the displacement of any two atoms i and j is given by:

$$C_{ij} = \langle \Delta r_i \Delta r_j \rangle / (\langle \Delta r_i^2 \rangle \langle \Delta r_j^2 \rangle)^{1/2},$$

where Δr_i is the displacement from the mean position of the ith atom. The elements C_{ij} can be collected in matrix form and displayed as a three-dimensional dynamical cross-correlation map (DCCM) (14). The C_{ij} are computed as averages over successive backbone N, C^{α}, and C atoms to give one entry per pair of amino acid residues. There is a time scale implicit in the C_{ij} as well. DCCMs for HIV-1 protease were calculated as block averages over intervals of time from 5 to 40 psec from the MD trajectory. The mean of two 40-psec block averages is represented in the DCCM shown in Fig. 5. Only correlations above a threshold value of 0.25 are included. The intensity of the shading is proportional to the magnitude of the coefficient. The positive correlations are given in the upper triangle, and the negative correlations in the lower triangle.

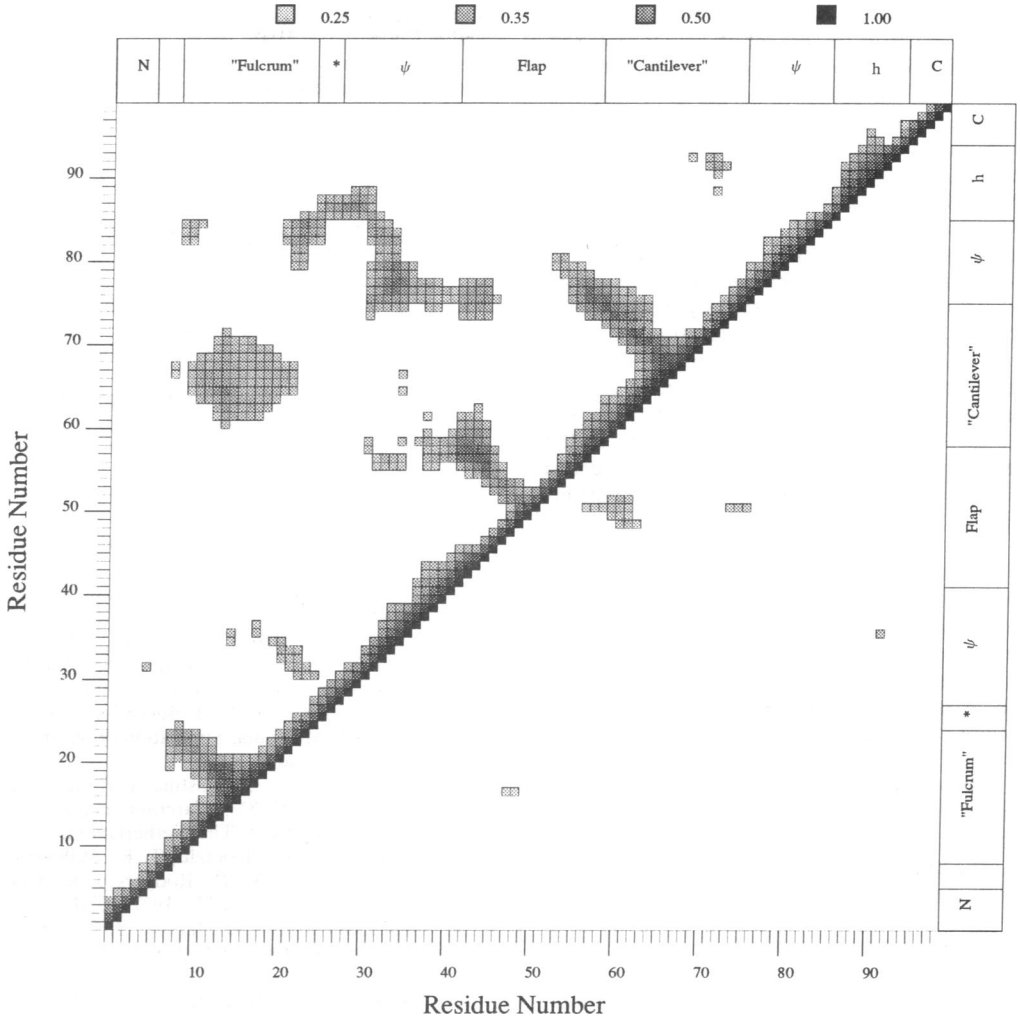

FIG. 5. Calculated DCCM for a monomeric unit of the HIV-1 protease dimer, averaged over the 40- to 100-psec interval of the simulation. Only correlations >0.25 are shown, and the intensity of shading is proportional to the size of the cross-correlation coefficient for each element, as indicated by the key at the top. Positive correlations are in the upper-left triangle, and negative correlations in the lower-right triangle.

Regions of regular secondary structure are expected to move in concert. Domains of contiguous residues as in an α-helix or β-sheet give rise to regions of significant positive correlations emanating from the diagonal of the DCCM. The three plumes emanating from the diagonal into the upper triangle of the DCCM of Fig. 5 are diagnostic of three distinct contiguous sequences of antiparallel β-sheet domains. The width of a plume represents correlation of proximal residues in the β-sheet rather than that of the whole domain. Moving up the diagonal, the first region is composed of residues 9–24, with correlated motion extending over five residues within the domain. The second region (residues 42–58) is the flap, with motions correlated over up to nine residues in the strongly pleated sheet section. Residues 59–75 form the longest and widest domain of β-sheet, with correlations extending over 10 residues. The N-terminal, C-terminal, and h-helix regions are characterized by triangular regions of positive correlation situated along the diagonal of the DCCM.

Specific off-diagonal peaks in the DCCM in Fig. 5 are indicative of a domain composed of residues noncontiguous in the amino acid sequence. The major cross-peaks found in the area of the DCCM between residues 22–25 and 81–85 and between residues 28–42 and 76–86 arise from the interaction of noncontiguous residues which fold to form parallel β-sheets and define the ψ region as a domain.

Other cross-peaks in the 40-psec DCCM are indicative of the presence of domain–domain interactions. The active site residues along with the flanking residues (22–29) have correlated motion with residues of the h-helix (86–90) with a width spanning up to five residues. Along with the flap (42–58) the two additional plume domains, residues 19–21 and 59–75, form an interesting ensemble. A major cross-peak of positive correlations is seen between residues 9–21 and 59–75. The onsets of the flap region (42–55 and 55–58) correlate with certain areas of β-sheet (59–65, 73–85) of the third plume domain. Negative domain–domain correlations are indicated between the flap and both other plume domains.

The observed interdomain correlations suggest that flap motion occurs with compensatory changes in residues 59–75 in the manner of a "cantilever." Pursuing this analogy, the β-sheet region composed of residues 9–21, correlated to both the flap and the cantilever, mediates as a "fulcrum." This correlated motion and domain communication is recognizable in the dynamical structure shown in Fig. 3, where the flap is observed to close down as the cantilever moves up. Negative compensatory correlations exist between the middle of the flaps and the center of the cantilever and reinforce this interpretation. Interdomain correlations are not evident from inspection of the average structure of the protein, but those linked to the active site are likely to play an important role in the functional energetics of the enzyme. It should be noted that site-specific mutagenesis of certain residues of the cantilever deactivates the protease and leads to noninfectious virions (15).

Examination of the structure of HIV-1 protease in the inhibitor complex (7) relative to that of the uncomplexed dimer shows a large difference in the flap positions but little change in the cantilever domain. Thus the force of the cantilever analogy needs further examination. The occurrence of domain–domain correlations in the dynamical structure of the protease is, however, clearly indicated. Information on this aspect of dynamical protein structure can be obtained only from MD simulation.

The DCCMs calculated over various intervals of time, to be described in detail elsewhere, indicate that the basic domain structure of the protein is well established on a 5-psec time scale, but the domain–domain correlations do not emerge until the 25- to 40-psec time scale. Thus structural or free energy determinations from protein dynamics simulations (16) on a time scale shorter than that reported here could miss important effects. Also, solvation is important in the theoretical model. The unphysical contraction of structure in an *in vacuo* model of the protein would introduce spurious domain correlations.

CONCLUSIONS

The proposed dynamical model for HIV-1 protease gives a quite reasonable account of the crystal geometry and predicts a reorganization of the flap region that should be favored in aqueous solution. The domain–domain structure and interactions were analyzed in detail and provide evidence for the presence of a molecular cantilever to the flap domain. Suggestions on the role of the flap in enzyme function include guarding the entrance to the active site cleft (17) and exclusion of water from the binding site (18). The active site triad characteristic of aspartyl proteases is correlated with a region of helix near the C terminus. The domain communications in HIV-1 protease uncovered in this study are thus likely to be involved in functional energy transfer and action of the enzyme.

The HIV-1 protease–inhibitor complex and the activated complex for substrate binding can be studied similarly, mapping out the dynamical structural changes and differential correlations expected on complexation and catalysis. Dynamical cross-correlation analysis of MD simulations may ultimately aid in the evaluation of potential specific inhibitors of HIV-1 protease and thus be a useful tool in the design of an AIDS therapy.

The atomic coordinates of the crystal geometry of HIV-1 protease were generously provided by Dr. Alex Wlodawer. This research was supported by Grant GM 37909 from the National Institutes of Health and a Cooperative High Technology Research and Development Grant from the State of Connecticut. Computer facilities were provided by Wesleyan University and the Pittsburgh Supercomputer Center.

1. Kohl, N. E., Emini, E. A., Schleif, W. A., Davis, L. J., Heimbach, J. C., Dixon, R. A. F., Scolnick, E. M. & Sigal, I. S. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 4686–4690.
2. Navia, M. A., Fitzgerald, P. M. D., McKeever, B. M., Leu, C.-T., Heimbach, J. C., Herber, W. K., Sigal, I. S., Darke, P. L. & Springer, J. P. (1989) *Nature (London)* **337,** 615–620.
3. Wlodawer, A., Miller, M., Jaskólski, M., Sathyanarayana, B. K., Baldwin, E., Weber, I. T., Selk, L. M., Clawson, L., Schneider, J. & Kent, S. B. H. (1989) *Science* **245,** 616–621.
4. Lipatto, R., Blundell, T., Hemmings, A., Overington, J., Wilderspin, A., Wood, S., Merson, J. R., Whittle, P. J., Danley, D. E., Geoghegan, K. F., Hawrylik, S. J., Lee, S. E., Scheld, K. G. & Hobart, P. M. (1989) *Nature (London)* **342,** 299–301.
5. Yarchoan, R., Mitsuya, H. & Broder, S. (1988) *Sci. Am.* **259** (4), 110–119.
6. Pearl, L. H. & Taylor, W. R. (1987) *Nature (London)* **329,** 351–354.
7. Miller, M., Schneider, J., Sathyanarayana, B. K., Toth, M. V., Marshall, G. R., Clawson, L., Selk, L., Kent, S. B. H. & Wlodawer, A. (1989) *Science* **246,** 1149–1151.
8. Swaminathan, S. (1990) WESDYN, Wesleyan Molecular Dynamics Program (Wesleyan Univ., Middletown, CT).
9. van Gunsteren, W. F. & Berendsen, H. J. C. (1986) GROMOS Groningen Molecular Simulation Program (Univ. of Groningen, The Netherlands).
10. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F. & Hermans, J. (1981) in *Intermolecular Forces*, ed. Pullman, B. (Reidel, Dordrecht, The Netherlands), pp. 331–342.
11. Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Jr., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tatsumi, M. (1977) *J. Mol. Biol.* **112,** 535.
12. McCammon, A. J. & Harvey, S. C. (1986) *Dynamics of Proteins and Nucleic Acids* (Cambridge Univ. Press, Cambridge, U.K.).
13. Brooks, C. L., III, Karplus, M. & Pettitt, B. M. (1988) *Proteins: A Theoretical Perspective of Dynamics, Structure and Thermodynamics* (Wiley, New York).
14. Swaminathan, S., Harte, W. E., Jr., & Beveridge, D. L. (1990) *J. Am. Chem. Soc.*, in press.
15. Loeb, D. D., Swanstrom, R., Everitt, L., Manchester, M., Stamper, S. E. & Hutchison, C. A., III (1989) *Nature (London)* **340,** 397–400.
16. Beveridge, D. L. & DiCapua, F. M. (1989) *Annu. Rev. Biophys. Biophys. Chem.* **18,** 431–492.
17. Skalka, A. M. (1989) *Cell* **56,** 911–913.
18. Blundell, T. & Pearl, L. (1989) *Nature (London)* **337,** 596–597.