above the following products were isolated and characterized. Hexadecane (**32**), octane (**33**), 1-octene (**34**), and (*E*)- and (*Z*)-2-octene (**36**) were obtained as colorless liquids and identified by comparison of their gas chromatographic retention times and ^{1}H NMR spectra with those of commercial samples.

Pentylcyclopropane (35) was isolated as a colorless liquid: ^{1}H NMR δ 1.30 (m, 8 H, 4 × CH_2), 0.87 (t, 3 H, J = 6 Hz, CH_3-), 0.62 (m, 1 H, CH-1), 0.50 (m, 2 H, CH-2α, -3α), 0.00 (m, 2 H, CH-2β, -3β); lit.[57] bp 128 °C; ^{1}H NMR δ 1.3 (m, 8 H, CH_2), 0.9 (t, 3 H, J = 6 Hz, CH_3), 0.45 (m, 3 H); IR 3100 cm^{-1}.

2-Methoxyoctane (37) was isolated as a colorless liquid: ^{1}H NMR δ 3.47 (s, 4 H, CH-2 and -OCH_3), 1.31 (m, 10 H, 5 × CH_2), 1.14 (d, 3 H, J = 6 Hz, CH_3-1), 0.93 (t, 3 H, J = 10 Hz, CH_3-8); lit.[4a] ^{1}H NMR (CCl_4) δ 3.47 (br s, 4 H), 1.31 (m, 10 H), 1.14 (d, 3 H, J = 6.5 Hz), 0.93 (m, 3 H).

Labeling Studies. A. Preparation of 1-Bromooctane-*1,1*-d_2 (25-*1,1*-d_2). In accordance with the general procedure of Kabalka,[58] 1.3 g (0.01 mol) of 1-octanol-*1,1*-d_2 (**38**-*1,1*-d_2)[1] was dissolved in 10 mL of $CHCl_3$ and cooled to 0 °C. To this solution, 1.6 g (0.02 mol) of pyridine was added. After 5 min of stirring, 2.9 g (0.02 mol) of 4-methylbenzenesulfonyl chloride was added in small portions. The solution was kept at 0 °C for 3 h and then extracted with 60 mL of diethyl ether. The ether layer was washed with three 30-mL portions of 1 N HCl, two 30-mL portions of 30% $NaHCO_3$ solution, and 50 mL of water and dried over saturated NaCl solution followed by anhydrous Na_2SO_4. Removal of the solvent by rotary evaporation gave 3.5 g of the tosylate as a colorless oil.

The tosylate was converted to bromide via a modification of the procedure of Wiberg.[59] To a 25-mL flask fitted with a condenser were added 3.50 g (0.012 mol) of the tosylate and 10 mL of dry acetone. Lithium bromide (3.0 g, 0.04 mol) was added and the acetone brought to reflux for 18 h. The flask was cooled, and 100 mL of water was added. The aqueous layer was extracted with three 50-mL portions of diethyl ether, 100 mL of 2-methylbutane, and 50 mL of diethyl ether. The combined organic fractions were washed with two 100-mL portions of water and dried over 50 mL of saturated NaCl solution followed by anhydrous Na_2SO_4. Removal of the solvent by rotary evaporation and chromatography of the resulting liquid on alumina gave 1.32 g (68% yield) of bromide **25**-*1,1*-d_2 as a colorless liquid: IR 2950, 2910, 2840, 2140, 1460, 1000, 980 cm^{-1}; ^{1}H NMR δ 1.82 (t, 2 H, J = 7 Hz, CH_2-2), 1.26 (m, 10 H, 5 × CH_2), 0.86 (t, 3 H, J = 6 Hz, CH_3-); ^{2}H NMR (CCl_4) δ 3.37 (s, CD_2-1). Analysis by ^{1}H and ^{2}H NMR showed 100% d_2 at C-1.

B. Preparation of 1-Bromooctane-*2,2*-d_2 (25-*2,2*-d_2). A 1.9-g sample of 1-octanol-*2,2*-d_2 (**38**-*2,2*-d_2)[1] was treated in a manner identical with that described for alcohol **38**-*1,1*-d_2. Chromatography on alumina and removal of the solvent by rotary evaporation gave 1.83 g (67% yield) of bromide **25**-*2,2*-d_2 as a colorless liquid: IR 2950, 2920, 2850, 2190, 2080, 1460, 1370, 1230, 840 cm^{-1}; ^{1}H NMR δ 3.36 (s, 2 H, CH_2-1), 1.25 (m, 10 H, 5 × CH_2), 0.86 (t, 3 H, J = 6 Hz, CH_3-); ^{2}H NMR (CCl_4) δ 1.81 (s, CD_2-2).

C. Irradiation. The results from irradiation of **25**-*1,1*-d_2 and **25**-*2,2*-d_2 are summarized in Table VI.

(4-Bromobutyl)- (40) and (4-Iodobutyl)benzene (41). The results from irradiation of bromide **39** and iodide **40**, prepared as described previously,[8h] are summarized in Table I. From preparative-scale irradiations conducted as outlined above, the following products were obtained as colorless liquids and identified by comparison of their gas chromatographic retention times, IR and ^{1}H NMR spectra with those of commercial samples: butylbenzene (**47**), 1,2,3,4-tetrahydronaphthalene (**48**), 3-butenylbenzene (**49**), (cyclopropylmethyl)benzene (**50**), and (*E*)- and (*Z*)-2-butenylbenzene (**51**).

Acknowledgment. Generous financial support by the National Science Foundation, the donors of the Petroleum Research Fund, administered by the American Chemical Society, and the University of North Carolina Research Council is gratefully acknowledged, as is helpful correspondence with Professor Martin Newcomb.

(57) Kell, D. R.; McQuillin, F. J. *J. Chem. Soc., Perkin Trans. 1* **1972**, 2096-2099.
(58) Kabalka, G. W.; Varma, M.; Varma, R. S. *J. Org. Chem.* **1986**, *51*, 2386-2388.
(59) Wiberg, K. B.; Lowry, B. R. *J. Am. Chem. Soc.* **1963**, *85*, 3188-3193.

Investigation of Domain Structure in Proteins via Molecular Dynamics Simulation: Application to HIV-1 Protease Dimer

S. Swaminathan,[†] W. E. Harte, Jr.,[†,‡] and D. L. Beveridge*[,†]

Contribution from the Chemistry Department, Hall-Atwater Laboratories, Wesleyan University, Middletown, Connecticut 06457, and the Pharmaceutical Research and Development Division, Bristol-Myers Squibb Company, 5 Research Parkway, Wallingford, Connecticut 06492.
Received March 21, 1990

Abstract: A general method for the investigation of domain structure in proteins based on the analysis of molecular dynamics simulations is described. The method is based on the idea that residues of a domain move in concert with one another, as determined by the analysis of the cross-correlation coefficients for atomic displacements computed from the simulation history. The correlation coefficients are displayed in the form of a two-dimensional "dynamical cross-correlation map" (DCCM) on which each type of protein domain has a signature appearance. Domain-domain correlations appear as off-diagonal elements in the DCCM and indicate novel aspects of action-at-a-distance and through-space communication in the structure which originate uniquely in the dynamical motions. The method is illustrated here with an analysis of a molecular dynamics simulation on HIV-1 protease (HIV-1 PR), a protein dimer that exhibits a diversity of secondary structural motifs.

I. Introduction

The general presence of domain substructures in globular proteins has long been recognized and used to systematize our understanding of protein structure[1] and folding.[2,3] We propose herein a new theoretical method for the investigation of domains in proteins based on the analysis of molecular dynamics simulations. The method also provides the possibility of new insight into the study of domain-domain communication in the dynamical structure and functional energetics of the protein. Our approach is illustrated here with an analysis of a molecular dynamics sim-

[†] Wesleyan University.
[‡] Bristol-Myers Squibb Co.

(1) Janin, J.; Chotia, C. *Methods Enzymol.* **1985**, *115*, 420-430.
(2) Lesk, A. M.; Rose, G. D. *Proc. Natl. Acad. Sci. U.S.A.* **1981**, *78*, 4304-4308.
(3) Kikuchi, T.; Nemethy, G.; Scheraga, H. A. *J. Protein Chem.* **1988**, *7*, 491-507.

0002-7863/91/1513-2717$02.50/0

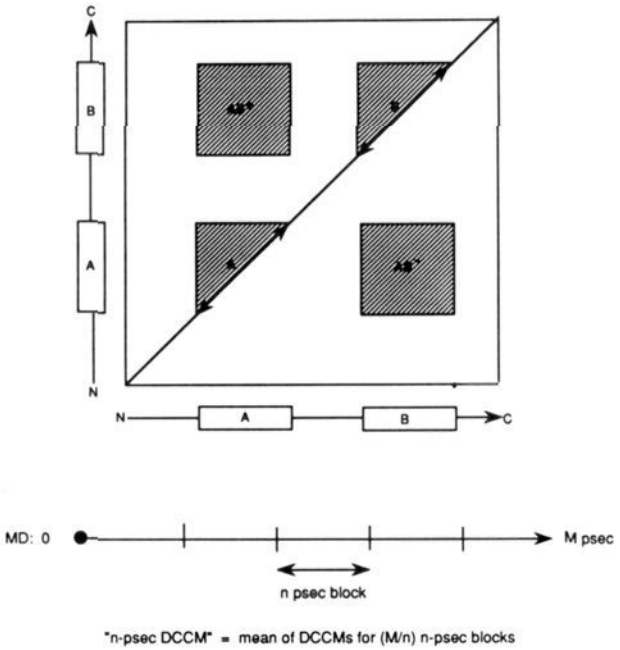

Figure 1. Schematic diagram of a dynamical cross-correlation map for a hypothetical protein with two distinct domains, A and B.

ulation on HIV-1 protease (HIV-1 PR), a protein dimer that exhibits a diversity of secondary structural motifs as evidenced in several recent X-ray crystal structures.[4-6]

II. Methods

The methods previously proposed for the detection of domain and domain boundaries have recently been reviewed by Janin and Chothia.[1] Domains can be defined in the context of either secondary or tertiary structure in proteins. Our use of the term refers to secondary structure such as helices and sheets. Methods for identification of domains include visual inspection,[7] distance maps,[8-12] automatic clustering and domain hierarchies,[2,13-15] surface areas,[16-19] and globularity indices.[20-23] Scheraga and co-workers[3,24,25] have recently developed a method based on statistically derived information about interresidue distance and discussed protein folding in terms of domains. Richards and Kundrot[26] recently proposed refinements of the distance matrix approach for secondary and supersecondary structure. All of these methods are based essentially on the time-averaged structure of the protein as obtained from X-ray crystallography.

From a dynamical point of view, the constituent atoms of a protein domain should execute concerted, well-correlated motions. Temperature factors determined for proteins in crystallography provide information on dynamic range, but as scalar quantities do not carry information on correlations. Experimental data related to correlations are accessible only at short range from 2D NOESY NMR analysis and at longer range only for local regions of structure in which a favorable spectroscopic probe is present. A molecular dynamics trajectory for a protein contains a description of the dynamical motions of all atoms included in the model, subject of course to the assumed force field. Concerted atomic motions can be identified by analyzing the cross-correlation coefficients for atomic displacements,[27,28] defined between two atoms i and j by the expression

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

where Δr_i is the displacement from the mean position of the ith atom. A contour plot of the matrix $[C_{ij}]$ constitutes a "dynamical cross-correlation map" (DCCM), in which atomic motions with strong correlations will have correspondingly large off-diagonal cross-peaks. A DCCM depends also on the time scale over which data on the correlations were gathered, and different types of interactions emerge on different time scales. This type of matrix was used to study residue displacements in cytochrome c by Ransom-Wright and McCammon[29] and the active-site dynamics of ribonuclease by Brunger et al.[30] The determinant of the C_{ij} matrix is related to the configurational entropy.[31]

The DCCM analysis of domain structure for the hypothetical case of a protein with two domains, A and B, is illustrated schematically in Figure 1. Positive correlations are collected in the upper triangle of the map and negative correlations in the lower triangle. Domains of contiguous residues should give rise to regions of positive correlations along the diagonal to the DCCM (the regions labeled A and B in Figure 2), indicating strong correlations among backbone atoms situated sequentially along the backbone. In the simplest of cases, such as an α helix, these regions will be essentially triangular in shape. Regions of anti-

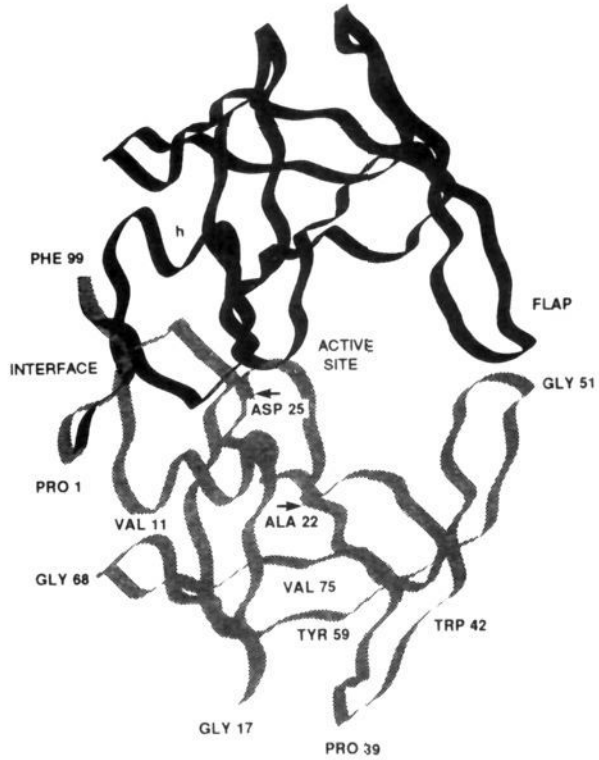

Figure 2. Ribbon tracing of the X-ray crystal structure of HIV-1 PR dimer as reported by Wlodawer et al.[5] Names of regions and amino acid sequence numbers important for the discussion of dynamical structure in the protein are indicated.

(4) Navia, M. A., et al. *Nature (London)* **1989**, *337*, 615–620.
(5) Wlodawer, A., et al. *Science* **1989**, *245*, 616–621.
(6) Lipatto, R., et al. *Nature (London)* **1989**, *342*, 299–301.
(7) Wetlaufer, D. B. *Proc. Natl. Acad. Sci. U.S.A.* **1973**, *70*, 697–701.
(8) Phillips, D. C. In *British Biochemistry Past and Present*; Goodwin, T. W., Ed.; Academic Press: London, 1970.
(9) Ooi, T.; Nishikawa, K. In *Conformation of Biological Molecules and Polymers*; Bergmann, E., Pullman, B. Eds.; Academic Press: New York, 1973.
(10) Rao, S. T.; Rossmann, M. G. *J. Mol. Biol.* **1973**, *85*, 177–181.
(11) Rossmann, M. G.; Liljas, A. *J. Mol. Biol.* **1974**, *76*, 241–256.
(12) Go, M. *Nature (London)* **1981**, *291*, 90–92.
(13) Crippen, G. M. *J. Mol. Biol.* **1978**, *126*, 315–322.
(14) Rose, G. D. *J. Mol. Biol.* **1979**, *134*, 447–470.
(15) Rose, G. D. *Methods Enzymol.* **1985**, *115*, 430–440.
(16) Lee, B. K.; Richards, F. M. *J. Mol. Biol.* **1971**, *55*, 379–406.
(17) Richards, F. M. *Annu. Rev. Biophys. Bioeng.* **1977**, *6*, 151–176.
(18) Wodak, S.; Janin, J. *Proc. Natl. Acad. Sci. U.S.A.* **1980**, *77*, 1736–1740.
(19) Wodak, S.; Janin, J. *Biochemistry* **1981**, *20*, 6544–6552.
(20) Janin, J. *J. Mol. Biol.* **1976**, *105*, 13–14.
(21) Teller, D. C. *Nature (London)* **1976**, *260*, 729–731.
(22) Rashin, A. A. *Stud. Biophys.* **1979**, *77*, 177.
(23) Rashin, A. A. *Nature (London)* **1981**, *291*, 85–87.
(24) Kikuchi, T.; Nemethy, G.; Scheraga, H. A. *J. Protein Chem.* **1988**, *7*, 427–471.
(25) Kikuchi, T.; Nemethy, G.; Scheraga, H. A. *J. Protein Chem.* **1988**, *7*, 473–489.
(26) Richards, F. M.; Kundrot, C. E. *Proteins: Struct., Funct. Genet.* **1988**, *3*, 71–84.
(27) McCammon, A. J.; Harvey, S. C. *Dynamics of Proteins and Nucleic Acids*; Cambridge University Press: Cambridge, U.K., 1986.
(28) Brooks, C. L., III; Karplus, M.; Pettitt, B. M. *Proteins: A Theoretical Perspective of Dynamics, Structure And Thermodynamics*; Wiley: New York, 1988.
(29) McCammon, J. A. *Rep. Prog. Phys.* **1984**, *47*, 1–46.
(30) Brunger, A. T.; Brooks, C. L.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 8458–8462.
(31) Karplus, M.; Kushick, J. N. *Macromolecules* **1981**, *14*, 325–332.

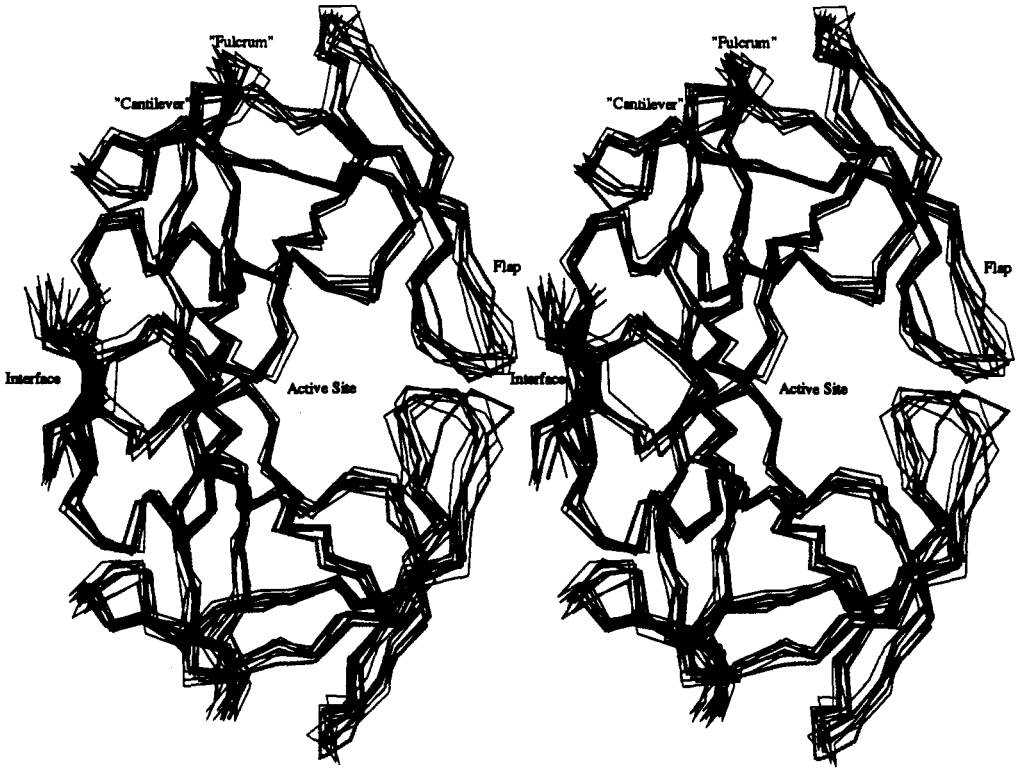

Figure 3. Dynamical range of motion subsumed by the HIV-1 PR dimer over the interval of 40–100 ps in the molecular dynamics simulation. The superposition of α-carbon traces for nine snapshots equally spaced along the MD trajectory is shown. The bolder solid line is the α-carbon trace of the crystal geometry as reported by Wlodawer et al.[5]

parallel β sheet will give rise to "plumes" of positive correlations emanating form the diagonal, and each distinct type of protein domain will likely have a signature appearance on a DCCM. Intradomain correlations between residues not contiguous in the primary sequence appear as off-diagonal matrix elements, also expected to be positive since the residues move in concert.

The remaining off-diagonal cross-peaks can be identified with domain–domain correlations in proteins and may be positive or negative, AB^+ and AB^-, respectively, in Figure 1. This information is expected to be particularly useful and interesting, as it conveys aspects of structure originating uniquely in the dynamical motions of the protein and not apparent at all from a visual inspection of the average structure.

III. Calculations and Results

To illustrate this idea, we consider a molecular dynamics simulation recently carried out by Harte et al.[32] on HIV-1 protease (HIV-1 PR), a 99-residue protein important in polyprotein processing in the life cycle of the AIDS virus. The crystal structure of HIV-1 PR molecules has recently been reported from several laboratories.[4-6] This simulation took as a point of departure the crystal structure of the synthetic [$Aba^{67,95}$] HIV-1 protease, recently solved by Wlodawer et al.[5] and depicted schematically in Figure 2.

Analysis of the crystal structure revealed the following regions of secondary structure (cf. Figure 1 of ref 5): the N-terminal and C-terminal residues, 1–5 and 95–99, respectively, form a dovetailed β-pleated sheet at the dimer interface. Residues 9–24 make up part of the b and c strands of β sheet. The c strand terminates at the active-site Asp-Thr-Gly triad. The c strand, d (30–35), c′ (69–78), and d′ strands of β sheet form the ψ region of the protein, so named because of its characteristic shape in the customary presentation of the 2° structure. The β strands a′ (43–49) and b′ (52–56) form a flap region guarding the active site. A short stretch of h helix, intermediate between a canonical α and 3_{10} helix in structure, was found for residues 86–94. Inspection and further analysis[32] of the MD trajectory suggested that residues 59–75 can be thought of as a "cantilever" to the flap, with residues 12–21 of the β sheet forming the "fulcrum".

The molecular dynamics simulation was performed on the protein dimer, the active form of the enzyme, together with 6990 water molecules in a hexagonal prism cell treated under periodic boundary conditions to model aqueous hydration.[32] The trajectory calculations were carried out on the program WESDYN[33] using the GROMOS 86 force field[34] and the SPC model for water. All charged groups were treated as neutral. Switching functions were used to make the long-range nonbonded interactions go smoothly to zero between 7.5 and 8.5 Å and applied on a group by group basis to avoid artifically splitting dipoles. The simulation involved 50 steps of conjugate gradient minimization of the crystal geometry, followed by 1.5 ps of heating to 300 K, 18.5 ps of equilibration, and 69.5 ps of free dynamics simulation under a temperature window of ±5 K. No rescaling was required beyond 33 ps into the run. Analysis was based on the interval between 40 and 100 ps in the simulation. The calculation was carried out on the Cray Y-MP/832 at the Pittsburgh Supercomputer Center and required ca. 100 h of machine time.

Examination of the calculated temperature and total energy indicates that the model system stabilizes in the realm of the crystal structure (ca. 1.3 Å rms deviation). An indication of the dynamical range of motion subsumed by the protein in the calculated trajectory, on the basis of superposition of α-carbon traces for nine structures taken at 10-ps intervals, is shown together with the crystal geometry in Figure 3. The dynamical structure is reasonably close to the crystal structure in all regions except for the flap, which is observed to close down. The structure of this region in the crystal is known to be influenced by intermolecular contacts in the crystal and adjusts in the simulation to a form we predict to be favored by aqueous hydration.

A calculated DCCM averaged over the two monomers of HIV-1 PR is shown in Figure 4. The elements of the map are the mean of two 40-ps block averages of calculated cross-correlation

(32) Harte, W. E., Jr., et al. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 8864–8868.

(33) Swaminathan, S. *WESDYN*; Wesleyan University: Middletown, CT, 1990.

(34) van Gunsteren, W. F.; Berendsen, H. J. C. *GROMOS86: Groningen Molecular Simulation System*; University of Groningen: Groningen, The Netherlands, 1986.

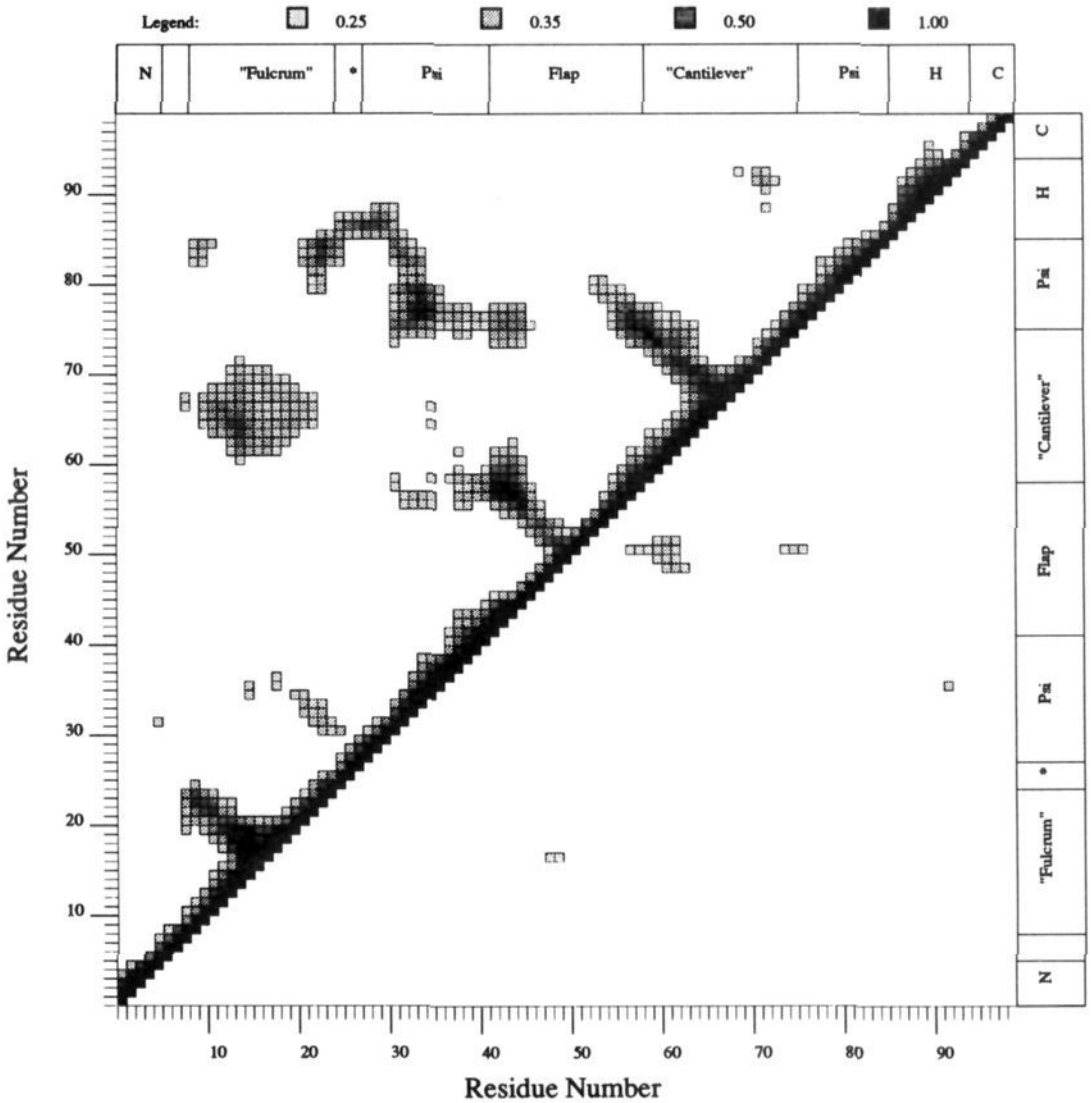

Figure 4. Calculated dynamical cross-correlation map for a monomeric unit of the HIV-1 PR dimer, averaged over the 40–100-ps interval of the simulation. Only correlations >0.25 are shown, and the intensity of shading is proportional to the size of the cross-correlation for each element.

coefficients, averaged over successive peptide N, C^{α}, and carbonyl C atoms to give one entry per amino acid residue. Only correlations above a threshold value of 0.25 are included. The intensity of the shading of the various elements is proportional to the magnitudes of the correlation coefficients. As in the schematic map of Figure 2, positive correlations are given in the upper triangle and negative correlations in the lower triangle.

The 40-ps DCCM in Figure 4 shows three distinct plumes of correlations emanating from the diagonal (backbone) and evidence in the off-diagonal region of considerable correlations between nonadjacent sequences. Each of the three plumes along the diagonal corresponds to regions of antiparallel β-pleated sheets and defines the fulcrum, the flap region, and the cantilever, respectively, as domains in the dynamical sense. The N-terminal, C-terminal, and h-helix regions also show cross-correlation patterns, but these domains are so short that they do not stand out on the map.

A major region of intradomain correlation between nonadjacent stretches of sequence is indicated in Figure 4 by off-diagonal elements between residues 22–25 and 81–85 and between residues 28–42 and 76–78. These correlations correspond to short stretches of parallel β sheets which arise as a consequence of the folding pattern of the tertiary structure. The cross-correlations help to define the ψ region of the protein as a domain, even though there is little triangular or other pattern of correlation along the diagonal for the constituent sequences. Collectively, these examples from HIV-1 PR illustrate some of the various ways in which domain structure can be expressed in DCCM as calculated from molecular dynamics simulation.

The monomer DCCM of Figure 4 also reveals interesting details about domain–domain communication. The flap motion shows both regions of positive and negative correlation with the motion of the cantilever. The b strand of the β-sheet region is positively correlated to the cantilever and negatively correlated with the flap, consistent with a functional role as a fulcrum. The domain structure and interactions thus provide support for the presence of a molecular cantilever to the flap domain. The active-site triad shows a distinct positive correlation to the h helix and to adjacent residues in the ψ region. The implications of this observation with respect to the functional energetics of HIV-1 PR enzyme action and inhibition are discussed in preliminary form elsewhere[32] and are currently the subject of continuations of this study. 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.[35]

The calculated 40-ps DCCM between residues of the two different monomeric units of the HIV-1 PR dimer is shown in Figure 5. The N-terminal and C-terminal ends of the two mo-

(35) Loeb, D. D., et al. *Nature (London)* **1989**, *340*, 397–400.

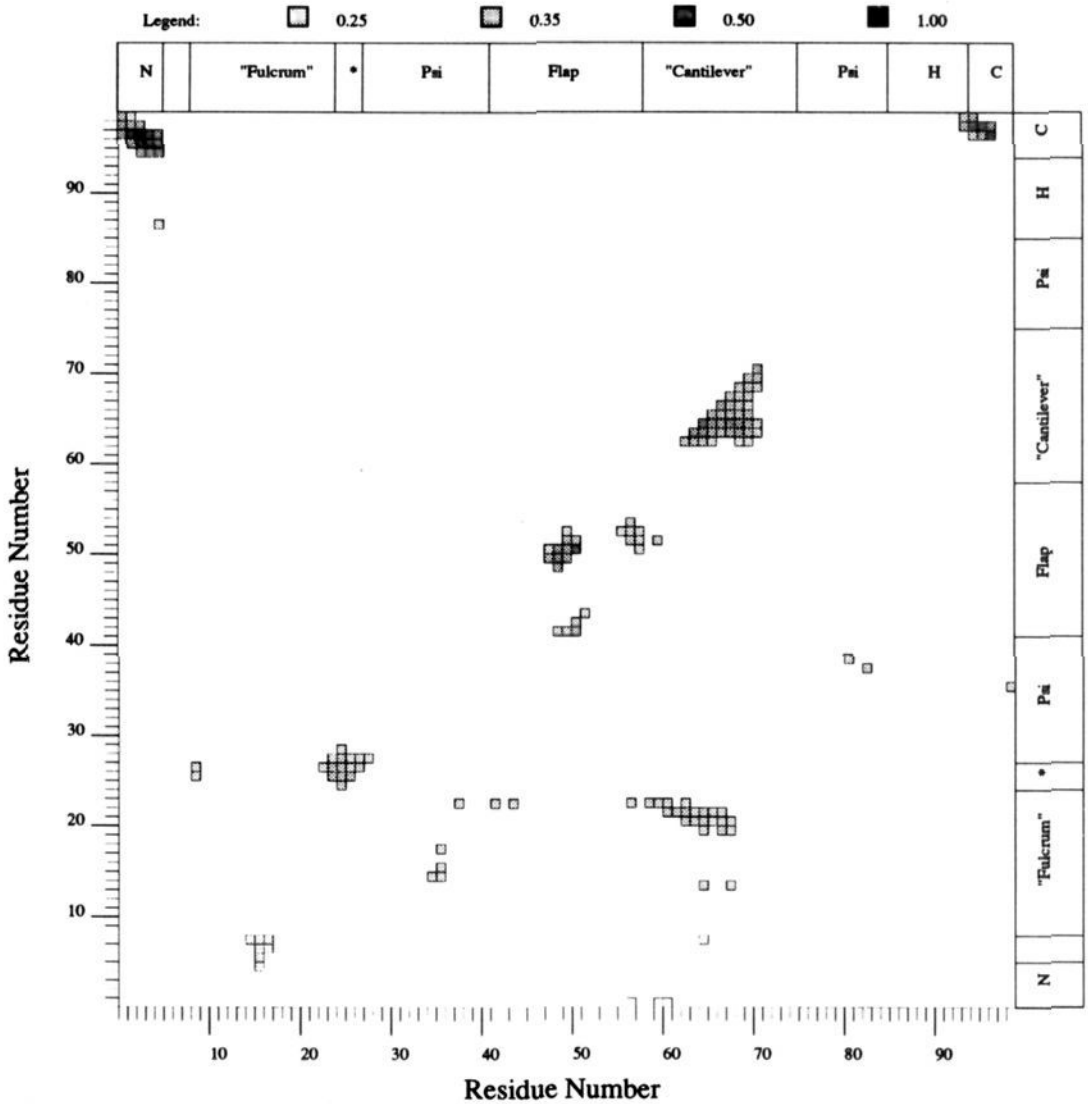

Figure 5. Calculated dynamical cross-correlation map between monomeric units of the HIV-1 PR dimer, averaged over the 40–100-ps interval of the simulation. Only correlations >0.25 are shown, and the intensity of shading is proportional to the size of the cross-correlation for each element.

nomeric units dovetail together to form the dimer interface. The correlated motions of these residues are clearly indicated in Figure 5. Intermonomeric correlations are indicated between the two active-site triads and between the two flap regions, which are indicated by the calculation to be hydrogen bonded in aqueous solution. This map also reveals that the cantilever domains of the two monomers, separated in space by some 30 Å, are strongly correlated. A distinct correlation between the β sheet of one monomer and the cantilever of the other is also evident.

IV. Summary and Conclusions

A unique role for theoretical molecular dynamics simulations in analyzing the dynamical structure of proteins and examining the correlations between dynamical structure and protein function is thus indicated. The cross-correlations calculated for the HIV-1 PR dimer in aqueous solution and presented as a DCCM provide a distinct delineation of domains in the protein. In addition, long-range cross-terms of the map provide a theoretical basis for the exploration of domain–domain communication. Theoretical evidence for the correlated motion of the flap and the cantilever domains of the structure via the fulcrum within a monomeric unit and for correlations between the cantilever domains of the two monomers are the most novel results of this analysis. The next step in this project will be to involve a similar analysis of domain correlations in HIV-1 PR inhibitor complexes. DCCM analysis of MD simulations may ultimately aid in the evaluation of potential specific inhibitors of HIV-1 PR and prove to be a useful technique in the development of an AIDS therapy.

In concluding, we emphasize that DCCM analysis requires a simulation carried out on the protein together with the corresponding aqueous hydration as illustrated here. Simulations on a protein in vaccuum would be expected to show spurious cross-correlations due to the characteristic unphysical contraction of the structure. The aqueous environment may itself mediate interesting interdomain correlations, which may be investigated by using a DCCM expanded to include solvent.

Note Added in Proof. Additional references on this type of methodology applied to BPTI have come to our attention: (a) T. Ichiye, Ph.D. Thesis, Harvard University, 1985. (b) T. Ichiye and M. Karplus, *Proteins: Struct. Funct. Genet.* **1991**, in press. (c) T. Nishikawa and N. Go *Proteins: Struct. Funct. Genet.* **1987**, *2*, 308.

Acknowledgment. This research was supported by Grant GM 37909 from the National Institutes of Health and Cooperative High Technology Research and Development Grant G02-90 from the State of Connecticut. Computer facilities were provided by Wesleyan University and the Pittsburgh Supercomputer Center. The atomic coordinates of the crystal geometry of HIV-1 PR (3HVP) were generously provided by Dr. Alex Wlodawer.

Registry No. Protease, 9001-92-7.