

A space-time structure determination of human CD2 reveals the CD58-binding mode

Akio Kitao* and Gerhard Wagner†‡

*Department of Chemistry, Graduate School of Science, Kyoto University, Kitashirakawa Oiwake-cho, Sakyo-ku, Kyoto 606-8502, Japan; and †Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115

Communicated by Christopher T. Walsh, Harvard Medical School, Boston, MA, December 13, 1999 (received for review November 5, 1999)

We describe a procedure for a space-time description of protein structures. The method is capable of determining populations of conformational substates, and amplitudes and directions of internal protein motions. This is achieved by fitting static and dynamic NMR data. The approach is based on the jumping-among-minima concept. First, a wide conformational space compatible with structural NMR data is sampled to find a large set of substates. Subsequently, intrasubstate motions are sampled by using molecular dynamics calculations with force field energy terms. Next, the populations of substates are fitted to NMR relaxation data. By diagonalizing a second moment matrix, directions and amplitudes of motions are identified. The method was applied to the adhesion domain of human CD2. We found that very few substates can account for most of the experimental data. Furthermore, only two types of collective motions have high amplitudes. They represent transitions between a concave (closed) and flat (open) binding face and resemble the change upon counter-receptor (CD58) binding.

Many experimental observations indicate that proteins continuously fluctuate, and many proteins change their conformation when performing their function, most significantly when binding ligands (1). NMR data contain implicitly information about the details of internal motions. However, dynamics studies with this technique primarily have been used to obtain information about the time scales of internal motions and order parameters S^2 as rough estimates of the extents of motions (2, 3). Few attempts have been made to move beyond this. For example, Brüschweiler and Case (4) have developed a collective NMR relaxation model. They used a set of low-frequency normal mode coordinates to fit order parameter S^2 as defined by the model free approach of Lipari and Szabo (5). Ulyanov *et al.* (6) considered multiple substates in the PARSE (probability assessment via relaxation rates of a structural ensemble) procedure. They attempted to reproduce two-dimensional nuclear Overhauser data by assuming multiple conformers, and probabilities of conformers were determined. Nilges and coworkers (7, 8) have studied protein motions with principle component analysis and observed that relatively few concerted motions can explain the relaxation properties the proteins studied.

Here we describe our attempts to make a more complete use of dynamic data and to develop a strategy for obtaining a space-time description of protein structures by fitting both structural nuclear Overhauser constraints and dynamic ^{15}N relaxation data. The goal was to obtain the coordinates of the most populated conformational states and the amplitudes and directions of the dominant internal motions. We assume that dynamic protein structures can be described by a distribution of conformational substates, and motions within and between the substates. Intrasubstate motions are fast and can be simulated reasonably well with molecular dynamics (MD) approaches. Intersubstate motions are much slower and not readily accessible to MD calculations. However, they represent the dominant motions with the largest amplitudes. Therefore, we based our analysis on the jumping-among-minima (JAM) concept (9). Intrasubstate motions are simulated with MD calculations using common force fields, and intersubstate motions are simulated

with a procedure that averages the contributions of all the substates. The weights (populations) of the substates that reproduce best the experimental data are fitted to the experimental dynamics data. With this achieved, a second-moment matrix for the deviations of the coordinates from the average positions is defined. The matrix is diagonalized, and eigenvalues and eigenvectors define the amplitudes and directions of internal collective motions (JAM modes).

We apply this approach to the adhesion domain of human CD2 (hCD2), a transmembrane glycoprotein found on T lymphocytes and natural killer cells (10). CD2 mediates cell-cell recognition by binding to the counter-receptor CD58 that is found on the surface of a variety of target cells, such as antigen-presenting cells. CD2 contains two Ig superfamily domains. Solution and crystal structures have been determined for the adhesion domains or both extracellular domains of rat CD2 (11, 12) and hCD2 (13–15). Recently, the structures of the adhesion domain of the counter-receptor hCD58 (16) and the complex of the adhesion domains of hCD2 and hCD58 also have been solved (17). In the crystal structures of rat and human CD2, a crystallographic dimer is observed with a conformation similar to the CD2/CD58 heterodimer (12, 14, 17) and more different from free CD2 in solution. Thus, the conformational changes of either protein upon counter-receptor binding have become known. The adhesion domain of hCD2 forms a nine-stranded Ig superfamily V-set fold. The dynamics of hCD2 have been characterized with ^{15}N NMR relaxation experiments (18). Here, we have present a space-time description of the adhesion domain of CD2 based on the JAM approach. Surprisingly, a very small number of JAM modes dominate the dynamics of the adhesion domain of hCD2, and these appear to be directly related to the main function of the protein, interaction with the counter-receptor, hCD58.

Methods

Principle of Approach. In the JAM approach (9), the ensemble average of an arbitrary dynamical variable A , $\langle A \rangle$, is given by $\langle A \rangle = \sum_{k=1}^M f_k \langle A \rangle_k$, where k is the index for the conformational substate. $\langle A \rangle_k$ represents the ensemble average of variable A within the substate k , and f_k is the statistical weight of the substate. The average $\langle A \rangle_k$ is obtained by molecular simulation sampling motions within the substate k . The statistical weights f_k are determined to reproduce the experimentally determined value of $\langle A \rangle$. This equation can be easily extended to matching several dynamic parameters simultaneously. In this paper, the statistical weights are fitted to reproduce NMR order parameters, S^2 , as an initial application (5). Order parameters were calculated by using the averaging scheme of our JAM model,

Abbreviations: MD, molecular dynamics; JAM, jumping among minima; hCD, human CD.

‡To whom reprint requests should be addressed.

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.

Article published online before print: *Proc. Natl. Acad. Sci. USA*, 10.1073/pnas.0305403997.
Article and publication date are at www.pnas.org/cgi/doi/10.1073/pnas.0305403997

which is based on the method by Henry and Szabo (19) and is shortly described below. The order parameter, S , is given by,

$$S^2 = \langle r^{-6} \rangle^{-1} \left[\frac{3}{2} \text{tr}(\langle \Phi \rangle^2) - \frac{1}{2} (\text{tr} \langle \Phi \rangle)^2 \right]. \quad [1]$$

The elements of the tensor $\Phi(t)$ are defined as,

$$\Phi_{\alpha\beta}(t) = \frac{1}{r^3(t)} \frac{r_\alpha(t)r_\beta(t)}{r^2(t)}, \quad [2]$$

where $r(t) = |\mathbf{r}(t)|$. A vector $\mathbf{r}(t)$ is a bond vector between two nuclei, such as proton and nitrogen. $r_\alpha(t)$ and $r_\beta(t)$ denote the x , y , or z components of $\mathbf{r}(t)$. In the JAM model, Eq. 1 can be rewritten as,

$$S^2 = \langle r^{-6} \rangle^{-1} \left[\frac{3}{2} \text{tr} \left(\sum_{k=1}^M f_k \langle \Phi \rangle_k \right)^2 - \frac{1}{2} \left(\text{tr} \sum_{k=1}^M f_k \langle \Phi \rangle_k \right)^2 \right]. \quad [3]$$

The average of tensor Φ within the k th substate, $\langle \Phi \rangle_k$, is determined from a molecular simulation. Once a set of f_k is determined, we can calculate various quantities, such as directions and amplitudes of atomic fluctuations. This is achieved by defining a second moment matrix \mathbf{C} with the elements C_{ij} . It describes the fluctuations of all the \mathbf{x}_i and is defined as,

$$C_{ij} \equiv \langle (x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle) \rangle. \quad [4]$$

The i th diagonal element of \mathbf{C} is a mean-square fluctuation along the i th coordinate from the average conformation. Therefore, the matrix \mathbf{C} gives the complete information on the magnitudes of the atomic fluctuations. In the JAM model (9), \mathbf{C} is expressed as:

$$\begin{aligned} C_{ij} &= \sum_{k=1}^M f_k (\langle x_i \rangle_k - \langle x_i \rangle) (\langle x_j \rangle_k - \langle x_j \rangle) \\ &+ \sum_{k=1}^M f_k \langle (x_i - \langle x_i \rangle_k)(x_j - \langle x_j \rangle_k) \rangle_k \\ &= D_{ij} + \sum_{k=1}^M f_k E_{ij}^k. \end{aligned} \quad [5]$$

The matrix $\mathbf{E}^k = (E_{ij}^k)$, the second moment matrix within the k th substate, is obtained directly from molecular simulation in each conformational substate, and the matrix $\mathbf{D} = (D_{ij})$, which represents the distribution of conformational substates, is determined from f_k and the average positions of the substates. To extract large-amplitude motions, collective coordinates are determined by solving the standard eigenvalue problem of either \mathbf{C} or \mathbf{D} . The former is called principal component analysis (20), the latter JAM analysis (9).

Sequence of Refinement Steps. The space-time refinement consists of three stages. In stage 1, an ensemble of structures is generated that satisfy the spatial restraints. This is achieved with five independent MD runs that include restraint energy terms, covalent energy terms, and quadratic repulsive energy terms between nonbonded atom pairs. To sample a large conformational space, relatively weak constraints were used. We checked the efficiency of the sampling in these MD simulations by using principal component analysis. We found that the conformational space highly overlapped between the different simulations. This

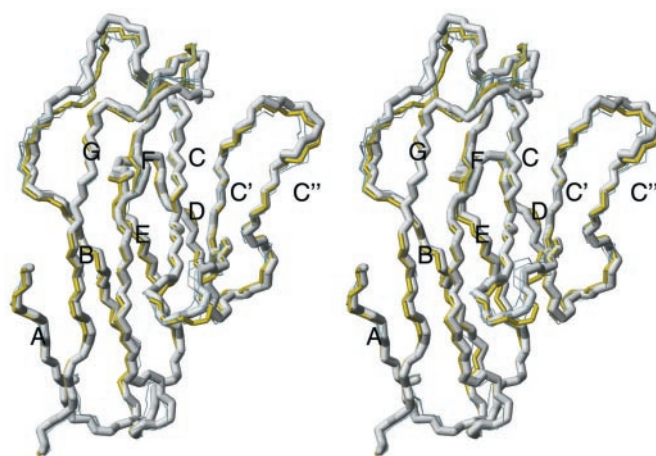


Fig. 1. Stereo diagram of the dominant substates of the adhesion domain of hCD2 in solution. The six most populated substates are shown. The relative population of each substate is indicated with the thickness of the lines. Substates 1 and 2 are colored white and yellow, respectively, the others are colored light blue. The instantaneous structures closest to the average structures of each substate are drawn. The β -strands are labeled A–G. The figure was prepared with the program MOLMOL (27).

indicates that these simulations sample rather completely the conformational space compatible with the spatial NMR constraints (data not shown). In stage 2, intrastate fluctuations are simulated by using MD calculations with full force field. In the first cycle, 50 conformations were sampled from the former five MD trajectories at every 50 psec, and the restraint energy was minimized. Starting from these minimum-energy conformations, MD simulations are carried out *in vacuo* by using the full force fields without restraint energy terms. It should be noted that intrastate fluctuations in solution occurring on a short time scale (mostly consisting of vibrational or damped oscillation modes) are compatible with those determined by the simulation *in vacuo* (9). However, the longer time-scale behavior is rather different. In each MD simulation, after 10-psec equilibration, 15-psec trajectory was stored and used to calculate time corre-

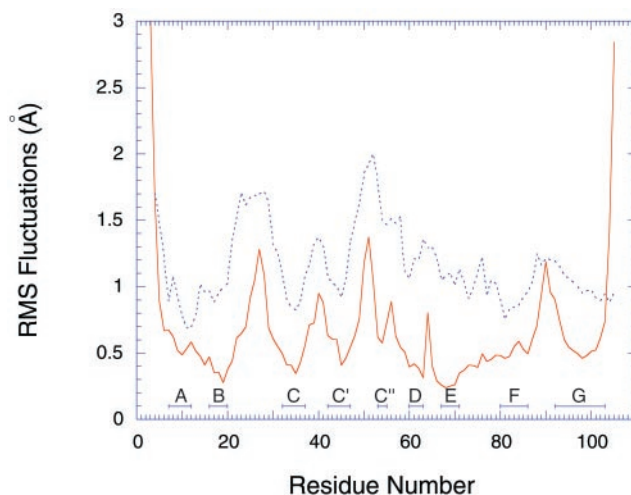


Fig. 2. Comparison of rms fluctuations of backbone atoms, obtained from the diagonal elements of the second moment matrix \mathbf{C} (solid line) and B factor of the x-ray structure of hCD2 (dotted line) (14). The latter was obtained by using the relation that mean square fluctuations of the i th atom are given by $\langle \Delta x_i^2 \rangle = \langle (x_i - \langle x_i \rangle)^2 + (x_{i+1} - \langle x_{i+1} \rangle)^2 + (x_{i+2} - \langle x_{i+2} \rangle)^2 \rangle = (3/8\pi^2) B_i$ (dashed line).

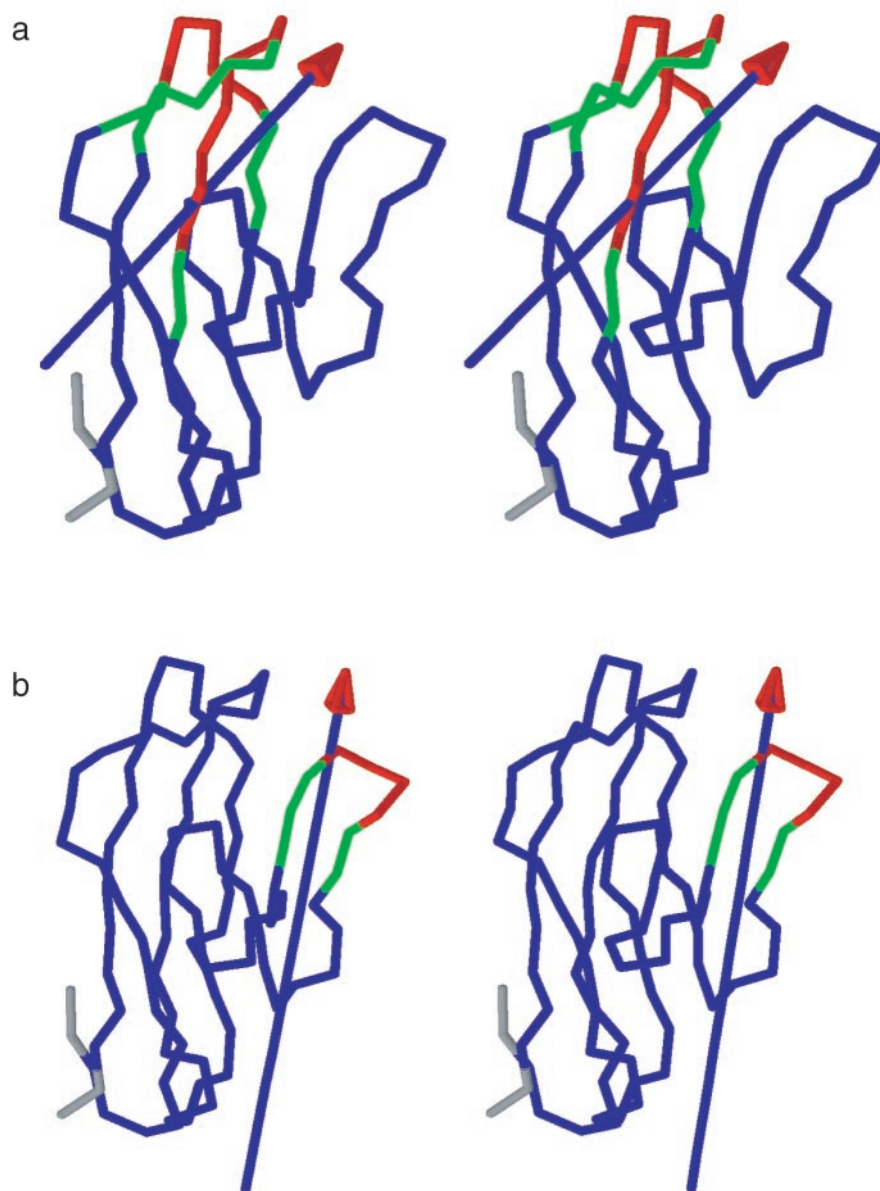


Fig. 3. Dynamic domain motions in (a) the first and (b) second JAM mode. The residues shown in blue, red, and green represent fixed-domain, moving-domain, and bending residues, respectively, as determined by DYNDOM (24). The arrows represent the axes of domain motions. The moving-domains rotate around these axes. The residues in the fixed domain are (a) Gly-11–Gln-25, Lys-34–Tyr-81, Lys-91–Asn-92, Leu-94–Lys-101, and (b) Gly-11–Phe-47, Phe-54–Lys-101. Those in moving domain are (a) Met-26–Ile-33, Lys-82–Gly-90, Val-93 and (b) Arg-48–Thr-53. Bending residues are (a) Gln-25–Met-26, Asp-32–Lys-34, Tyr-81–Lys-82, Gly-90–Leu-94 and (b) Gln-46–Arg-48, Thr-53–Phe-54. The residues that act as a mechanical hinge are (a) Asp-32, Ile-33, Lys-34, Lys-82 and (b) Gln-46, Phe-47, Arg-48, Thr-53, and Phe-54. The figure was created by using the programs DYNDOM (24) and RASMOL (28).

lation functions for the intrastate motions. In stage 3, statistical weights f_k are determined to minimize the difference between experimental and calculated order parameters. Stages 2 and 3 are repeated until the fitting of the order parameters converges. In work described here, this cycle was repeated twice. For the second cycle of stages 2 and 3, the 30 substates with the highest statistical weights f_k were selected. In addition, 35 substates, which are relatively close in conformational space to the substates with relatively high values of f_k (larger than 0.01), were added to the ensemble. We examined how the refinement results depend on the initial values of f_c and the number of the order parameters used in the refinement. We found that the final set of f_k values does not depend on the initial set. Furthermore, a limited variation of the number of order parameters used had only small effect on the final f_k values.

Results and Discussion

Data Used for Space-Time Refinement of hCD2 Adhesion Domain. The space-time refinement of hCD2 was based on the following experimental data. We started out randomly choosing one of an ensemble of 18 structures previously determined with static NMR constraints. The spatial constraints used for the space-time refinement described here were adapted from Wyss *et al.* (15). They consisted of 50 distance restraints for hydrogen bonding pairs, 945 nuclear Overhauser restraints (860 intrapolypeptide, 46 intraglycan, and 39 polypeptideglycan), and 192 dihedral restraints (92 ϕ , 19 χ_1 , 13 β -methylene, and 68 glycan). The dynamic constraints were adapted from previously performed NMR relaxation parameters of NH groups (18). Using relaxation rates, $R_N(N_z)$, $R_N(N_{x,y})$, and $R_N(H_z \rightarrow N_z)$, spectral densities, $J_{\text{eff}}(0)$, $J(\omega_N)$, and $J_{\text{ave}}(\omega_H)$ were determined by using spectral

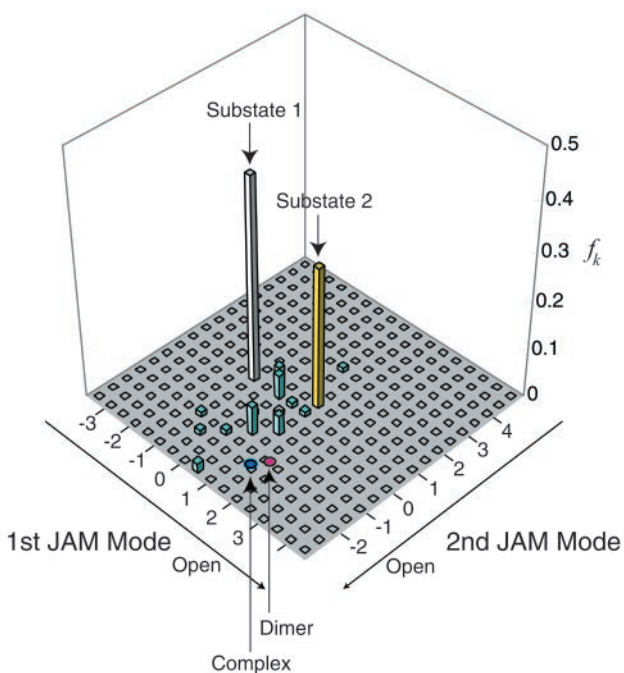


Fig. 4. Statistical weights f_k of each substate projected onto the two-dimensional space spanned by the first and second JAM coordinates. The bars shown by white and yellow represent f_k of substates 1 and 2, respectively. The magenta and blue marks represent the positions of the crystal structure of the hCD2 dimer (14) and the complex with hCD58 (17), in this JAM mode subspace. The axes of the first and second JAM coordinates are scaled by the SDs of 0.90 Å and 0.42 Å, respectively.

density mapping methods (21, 22). First, the three spectral densities are fitted to the original model free approach of Lipari and Szabo (5) assuming isotropic tumbling with a single overall correlation time and a single exponential decay for internal motion. If the fitting results are not satisfactory, an extended model with a chemical exchange term λR_{ex} was applied. Of the 102 order parameters obtained from the experimental data Wyss *et al.* (18), 66 main-chain order parameters with the lowest experimental errors were used for the refinement.

Two Substates Are Dominant. The application of this procedure to the adhesion domain of hCD2 showed the rather surprising results that very few substates can account for the mobility data in hCD2. In the final stage of this space-time refinement, 65 substates were used.

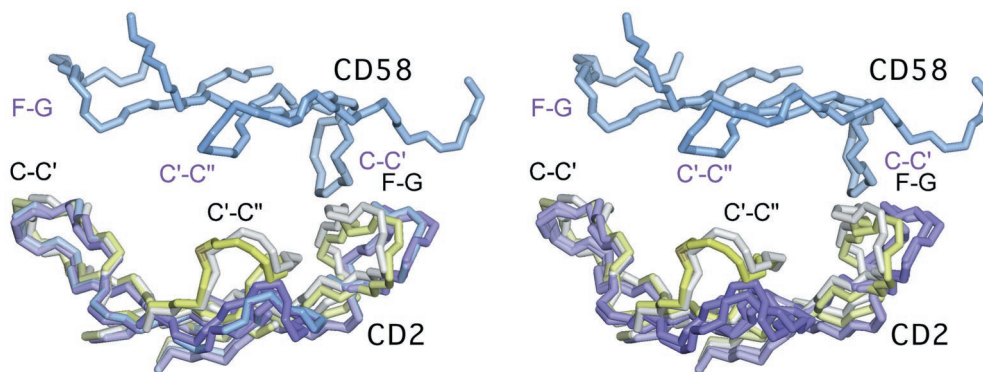


Fig. 5. The CD2 surface flattens upon CD58 binding (in stereo). Substates 1 and 2 of free CD2 are shown in white and yellow, respectively. The crystal structure of the hCD2, which is a crystallographic homodimer (14) and the CD2/CD58 heterodimer (17) are shown in dark and light blue, respectively. The figure was prepared with INSIGHT II (Molecular Simulations, Waltham, MA).

Among them, only two substates have dominantly large statistical weights. The most populated substate 1 has the statistical weight $f_1 = 0.42$ and substate 2 has a weight $f_2 = 0.24$. Thus, the probability of the state point staying in either substate 1 or 2 is 0.66. The structures of the six most populated substates are shown in Fig. 1. To illustrate the populations, the thickness of the bonds is proportional to the values of f_k .

CD58-Binding Face Is Very Flexible. With the f_k values determined, we calculated the second-moment matrix **C** that contains the mean square atomic fluctuations as the diagonal elements. These can be compared with experimentally measured values from NMR and x-ray data. In Fig. 2, the rms fluctuation (RMSF) for each residue averaged over the main-chain atoms is compared with values derived from crystallographic *B* factor. The curves are in good agreement although the values calculated from the NMR data are generally smaller. Six loops, the B–C, C–C', C'–C'', C''–D, and D–E loops are more flexible, whereas the A–B and E–F loops are rigid. This may be functionally important because the loops near the binding site are flexible whereas two loops relatively far from it are not. This suggests that the regions near the binding site are designed to be flexible even in the free CD2 prepared for the conformational change upon binding, whereas other regions are designed to be more rigid. It also should be noted that the crystal structure of hCD2 contains the N-terminal V-set adhesion domain and a second C-type domain. Therefore, the RMSF near the C terminus in the crystal is relatively low. Except for this region, the RMSF in crystal is much larger than that in present refinement. It is reasonable because the RMSF in the refinement originated only from internal motion, whereas *B* factor consists of internal contribution, external contribution, and crystal defects (23).

The Motions in Free hCD2 Resemble the Conformational Change upon CD58 Binding. The eigenvectors and eigenvalues of the **D** component in the second moment matrix (Eq. 5) yield collective coordinates and amplitudes of motions along these coordinates. The collective coordinates are linear combinations of the Cartesian atomic coordinates and form an orthogonal set (9). Pictorially, the axis of the first JAM mode is defined as the direction along which conformational substates are most broadly distributed. That of the second JAM mode is orthogonal to the first one, and along this axis, substates are distributed second most. Among 65 JAM modes, the first and second one dominantly contributed to the total intersubstate mean square fluctuation (MSF) by 77.1%. In other words, conformational substates are mostly distributed in the two-dimensional space spanned by the first and second JAM modes. The first JAM

mode, which contributes to the total MSF by 63.5%, is directly involved with the conformational transition between substate 1 and 2. The second JAM mode contributes to total MSF by 13.6%. The motions along these two axes are typical dynamic domain motions. We further analyzed the first two modes with the program DYNDOM (24), which identifies rigid substructures and linker regions. The results are shown in Fig. 3. The moving domain in the first JAM mode consists of a part of the B–C loop, a part of the F strand, and the F–G loop. The moving domain of the second JAM mode is on the C'–C'' loop. The CD58 binding site consists of the C–C', C'–C'', and F–G loops and the regions between (25, 26), which includes most of the moving domains of both JAM modes. In Fig. 4, projections of the positions of the significantly populated conformational substates onto the two-dimensional space spanned by the first and second JAM modes are shown. The collective coordinates of the two crystal structures of the homodimer and the heterodimer (14, 25, 26) also are shown. If CD2 moves along the positive direction in the first JAM mode, or along the negative direction in the second JAM mode (Fig. 4), the binding site is more opened. Therefore, substate 2 is more opened than substate 1 with respect to mode 1, and both crystal structures have a more open binding site than all substates of the free protein in solution. This is shown in close detail in Fig. 5. Thus, it seems that the directions of the collective motions as described above are closely related to the conformational changes occurring upon binding the counter-receptor CD58. It seems that the dominant change is along the first JAM mode. There are significant changes along the second and several other JAM modes as well, however. The motions in the free protein resemble the conformational change upon CD58 binding although the conformational fluctuations in the free

state have smaller amplitudes than the changes occurring upon CD58 binding. In the first cycle of the refinement, the conformations relatively close to the CD58-binding form were included in stage 3; however, their f_k values were negligibly small. The binding process may use conformations that are populated in free CD2 and sufficient to make an energetically favorable initial complex, followed by additional small changes to form the final complex structure.

Conclusion

We have developed a method that can determine the average structure, the distribution of conformational substates, as well as directions and amplitudes of internal motions in a protein. We have applied this method to the adhesion domain of hCD2 and have shown that the fluctuations in the free protein resemble the onset of the counter-receptor-binding event. Here, we have fitted the dynamic structures to the order parameters, S^2 , derived from ^{15}N relaxation experiments. Extension of this approach to fitting in addition nuclear Overhauser intensities to a dynamic model is straight forward, although computationally challenging. The time scale of the motions analyzed here is in the ps to ns range, which is faster than the overall motions of the protein. Experimentally, we have observed slower motions in the μs to ms range as well, in particular for the CD58-binding site (18). To rationalize motions at this slow time scale is still beyond the approach presented here.

From April 1998 to March 1999, A.K. was a Monbusho-Sponsored Japanese Overseas Research Fellow at Harvard Medical School. This work was supported by National Science Foundation Grant MCB 931 6938 and National Institutes of Health Grant GM47467.

- Gerstein, M., Lesk, A. M. & Chothia, C. (1994) *Biochemistry* **33**, 6739–6749.
- Wagner, G. (1994) *Curr. Opin. Struct. Biol.* **3**, 748–754.
- Kay, L. E. (1998) *Biochem. Cell. Biol.* **76**, 145–152.
- Brüschweiler, R. & Case, D. A. (1994) *Phys. Rev. Lett.* **72**, 940–943.
- Lipari, G. & Szabo, A. (1982) *J. Am. Chem. Soc.* **104**, 4546–4559.
- Ulyanov, N. B., Schmitz, U., Kumar, A. & James, T. L. (1995) *Biophys. J.* **68**, 13–24.
- Abseher, R., Horstink, L., Hilbers, C. W. & Nilges, M. (1998) *Proteins* **31**, 370–382.
- Horstink, L., Abseher, R., Nilges, M. & Hilbers, C. W. (1999) *J. Mol. Biol.* **287**, 569–577.
- Kitao, A., Hayward, S. & Go, N. (1998) *Proteins* **33**, 496–517.
- Davis, S. J., Ikemizu, S., Wild, M. K. & van der Merwe, P. A. (1998) *Immunol. Rev.* **163**, 217–236.
- Driscoll, P. C., Cyster, J. G., Campbell, I. D. & Williams, A. F. (1991) *Nature (London)* **353**, 762–765.
- Jones, E. Y., Davis, S. J., Williams, A. F., Harlos, K. & Stuart, D. I. (1992) *Nature (London)* **360**, 232–239.
- Wyss, D. F., Withka, J. M., Knoppers, M. H., Sterne, K. A., Recny, M. A. & Wagner, G. (1993) *Biochemistry* **32**, 10995–11006.
- Bodian, D. L., Jones, E. Y., Harlos, K., Stuard, D. I. & Davis, S. J. (1994) *Structure (London)* **2**, 755–766.
- Wyss, D. F., Choi, J. S., Li, J., Knoppers, M. H., Willis, K. J., Arulanandam, A. R. N., Smolyar, A., Reinherz, E. L. & Wagner, G. (1995) *Science* **269**, 1273–1278.
- Sun, Z., Dotsch, V., Kim, M., Li, J., Reinherz, E. & Wagner, G. (1999) *EMBO J.* **18**, 2941–2949.
- Wang, J., Smolyar, A., Tan, K., Liu, J., Kim, M., Sun, Z., Wagner, G. & Reinherz, E. (1999) *Cell* **97**, 791–803.
- Wyss, D. F., Dayie, K. T. & Wagner, G. (1997) *Protein Sci.* **6**, 534–542.
- Henry, E. R. & Szabo, A. (1985) *J. Chem. Phys.* **82**, 4753–4761.
- Kitao, A., Hirata, F. & Go, N. (1991) *Chem. Phys.* **158**, 447–472.
- Peng, J. W. & Wagner, G. (1992) *Biochemistry* **31**, 8571–8586.
- Lefevre, J.-F., Dayie, K. T., Peng, J. W. & Wagner, W. (1996) *Biochemistry* **35**, 2674–2686.
- Kidera, A., Inaka, K., Matsushima, M. & Go, N. (1992) *J. Mol. Biol.* **225**, 477–486.
- Hayward, S. & Berendsen, H. J. C. (1998) *Proteins* **30**, 144–154.
- Arulanandam, A. R. N., Withka, J. M., Wyss, D. F., Wagner, G., Kister, A., Pallai, P., Recny, M. A. & Reinherz, E. L. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 11613–11617.
- Arulanandam, A. R. N., Kister, A., McGregor, M. J., Wyss, D. F., Wagner, G. & Reinherz, E. L. (1994) *J. Exp. Med.* **180**, 1861–1871.
- Koradi, R., Billeter, M. & Wuthrich, K. (1996) *J. Mol. Graphics* **14**, 51–55.
- Sayle, R. & Milner-White, E. J. (1995) *Trends Biochem. Sci.* **20**, 374–376.