Protein Structural Change Upon Ligand Binding: Linear Response Theory

Mitsunori Ikeguchi, Jiro Ueno, Miwa Sato, and Akinori Kidera*

Graduate School of Integrated Science, Yokohama City University, Tsurumi-ku, Yokohama 230-0045, Japan

(Received 25 October 2004; published 24 February 2005)

A simple formula based on linear response theory is proposed to explain and predict the structural change of proteins upon ligand binding. By regarding ligand binding as an external perturbation, the structural change as a response is described by atomic fluctuations in the ligand-free form and the protein-ligand interactions. The results for three protein systems of various sizes are consistent with the observations in the crystal structures, confirming the validity of the linear relationship between the equilibrium fluctuations and the structural change upon ligand binding.

DOI: 10.1103/PhysRevLett.94.078102 PACS numbers: 87.15.He

Proteins undergo structural changes upon ligand binding and control the succeeding steps of protein functions [1]. Recently, various experimental observations have illuminated the correlation between the equilibrium fluctuations in the ligand-free state and the structural change upon ligand binding [2]. Computational studies have also found that the low-frequency normal modes calculated for the free state can well explain the observed structural changes occurring upon ligand binding [3–5]. As the low-frequency normal modes describe the global pattern of the equilibrium fluctuations [6], both experimental and theoretical studies indicate that the structural change upon ligand binding occurs mostly within the conformational ensemble in the ligand-free state [7,8]. This explanation leads to linear response theory (LRT), which states that the response behavior is related to the equilibrium fluctuations in the unperturbed state [9]. In this study, the response of protein structures to ligand binding is formulated based on LRT and the structural changes are predicted from the structural information of the ligand-free state.

The structural change upon ligand binding is described as the response of the protein atoms to the perturbing interactions of the ligand molecule, as in the model of the induced fit [10,11]. The Hamiltonian of the bound state, H_1, can be described in terms of that of the free state, H_0, and the perturbation from the ligand molecule as follows:

$$H_1 = H_0 + \sum_i \int d\mathbf{r} V_i(\mathbf{r}) \phi_i(\mathbf{r}), \tag{1}$$

where $\phi_i(\mathbf{r}) \equiv \delta(\mathbf{r} - \mathbf{r}_i)$, $\mathbf{r}_i$ is the position of protein atom i, and V_i is the perturbation potential at atom i. The response, $\delta\rho_i(\mathbf{r})$ [$\equiv \langle\phi_i(\mathbf{r})\rangle_1 - \langle\phi_i(\mathbf{r})\rangle_0$], is evaluated to the first order as

$$\delta\rho_i(\mathbf{r}) \simeq -\beta \sum_j \int d\mathbf{r}' \langle \delta\phi_i(\mathbf{r}) \delta\phi_j(\mathbf{r}')\rangle_0 V_j(\mathbf{r}'), \tag{2}$$

where $\langle x\rangle_\lambda = \int d\mathbf{r} x e^{-\beta H_\lambda} / \int d\mathbf{r} e^{-\beta H_\lambda}$ [$\lambda = 0$ (free) or 1 (bound)], and $\beta = 1/k_B T$ with Boltzmann factor k_B and absolute temperature T. Equation (2) is simply a static formulation of the LRT, indicating that the structural change is related to the perturbation, V_j, and the response function, $\langle \delta\phi_i(\mathbf{r}) \delta\phi_j(\mathbf{r}')\rangle_0$, defined in the free state.

In the process of ligand binding, however, the response frequently extends over wide spatial and temporal ranges such that the actual time course of the response follows a complicated pathway on the rugged potential surface, exceeding the linear regime [12,13]. To describe such large complicated fluctuations in proteins, the quasiharmonic description in protein dynamics [6,14] is adopted in the present scheme. Rather than defining the response function for an instantaneous structure, the function is taken as an average along a long trajectory of a molecular dynamics (MD) simulation in a quasiharmonic manner. This means that the response, $\delta\rho_i(\mathbf{r})$, will also be predicted at the resolution of the quasiharmonic approximation. Equation (2) is rewritten in the form of the expectation values as follows:

$$\Delta\mathbf{r}_i \simeq \beta \sum_j \langle \Delta\mathbf{r}_i \Delta\mathbf{r}_j\rangle_0 \mathbf{f}_j, \tag{3}$$

where $\Delta\mathbf{r}_i$ ($= \int d\mathbf{r} \mathbf{r} \delta\rho_i(\mathbf{r})$) is the expectation of the coordinate shift of atom i, $\langle \Delta\mathbf{r}_i \Delta\mathbf{r}_j\rangle_0$ is the variance-covariance matrix of the atomic fluctuations in the ligand-free state, and $\mathbf{f}_j$ is the external force acting on atom j. The details of the above equations are found in the online supplementary material [15]. Equation (3) allows the structural change to be predicted using the quasiharmonic response function derived from an MD simulation in combination with an appropriate model of the external force mimicking ligand binding. Structural changes upon ligand binding were calculated for three protein systems of various sizes: ferric binding protein (FBP) [16,17], citrate synthase [18], and F_1-ATPase [19,20], the ligand-free and ligand-bound crystal structures which are available from the Protein Data Bank. The variance-covariance matrix in the free state, $\langle \Delta\mathbf{r}_i \Delta\mathbf{r}_j\rangle_0$, was calculated from a 10-ns MD simulation in explicit solvent for FBP and citrate synthase and from inversion of the Hessian calculated using the elastic network model [21] for F_1-ATPase. As a first test of the LRT formula, the

perturbing force, $\mathbf{f}_j$, representing ligand-protein interactions, was modeled by a single vector with reference to the correct binding site in the bound form [17,18,20].

The MD simulations for FBP and the citrate synthase in the free state were carried out using the program MARBLE [22] with the AMBER96 force field for proteins and the TIP3P model for water [23,24]. In the case of F_1-ATPase, a 10-ns MD simulation was not sufficient to derive the variance-covariance matrix suitable for representing the structural changes, due to the large size of the system and the long relaxation time. Therefore, the elastic network model [21] was employed, in which the harmonic constraints are imposed only on C_α atoms. Details of the methods are found in the online supplementary material [15].

In the calculation of the variance-covariance matrix, each structure was divided into a smaller domain and a larger domain using the program DynDom [25]. The external motions of the latter were eliminated by superposition such that the structure change could be identified as the motion of the smaller domain (moving domain) relative to the larger domain (fixed domain). This superposition method avoids the complicated domain motions that arise when the external motions of the whole molecule are removed. Although the elastic network model used for F_1-ATPase is not compatible with this superposition method, superposition of the whole molecule effectively removed the external motion of the fixed domain, because the fixed domains are much larger than the moving domains. In the case of citrate synthase, due to the symmetrical nature of the homodimeric form, the variance-covariance matrix was symmetrized by superimposing the original trajectory onto the trajectory for the monomer-swapped form.

The ligand interaction was mimicked by an attractive force between an atom in the moving domain and the center of the ligand. The ligand position was determined by the superposition of the free and bound structures by the method described above. The interacting atom was selected as the atom interacting with the ligand in the bound form. In the case of the elastic network model, an appropriate C_α atom was chosen as the interacting atom. As the displacement calculated by Eq. (3) is proportional to the imposed force, the magnitude of the force was adjusted so as to make the average displacement the same as that of the experimental data. It should be noted that this single-vector model was adopted for the sake of simplicity and is not necessarily indicative of the real kinetic process of ligand binding.

The results for FBP are summarized in Fig. 1. The structure of FBP changes upon binding of a ferric ion, the influence of which was modeled by an attractive force between Oε1 of Glu57 in the moving domain and the ion in the fixed domain [the yellow arrow in Fig. 1(a)]. The displacement of C_α atoms predicted by Eq. (3) is well correlated with the observations of the crystal structures (Fig. 1), with a correlation coefficient between the predicted and experimental displacement vectors of 0.95.

The predicted result is not sensitive to the force model employed, or the direction and position of the applied force [Fig. 1(c)]. The robustness originates from the nature of protein dynamics. When the variance-covariance matrix is diagonalized, Eq. (3) is rewritten by

$$\Delta\mathbf{r}_i \simeq \beta \sum_k \mathbf{v}_i^{(k)} \lambda^{(k)} \sum_j \mathbf{v}_j^{(k)} \cdot \mathbf{f}_j, \tag{4}$$

where $\lambda^{(k)}$ and $\mathbf{v}_i^{(k)}$ are the eigenvalue and the eigenvector for atom i and the kth normal mode, respectively. It is

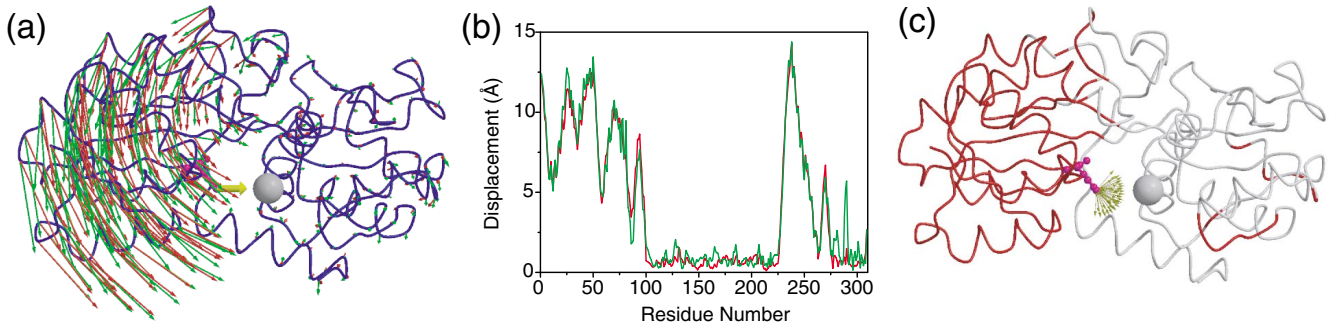

FIG. 1 (color). Structural change of FBP upon binding of Fe^{3+}. (a) Experimental (red) and predicted (green) structural changes of C_α atoms, represented by arrows on the backbone trace of the average structure in the simulation (blue) (correlation coefficient, 0.95). The experimental displacement is the difference between the ligand-free form (PDB ID: 1D9V) and the bound form (1MRP). Upon binding of Fe^{3+} (gray sphere), the moving domain (residues 1-82, 88-101, 226-276, 308-309; left part) undergoes a closure motion against the fixed domain (residues 83-87, 102-225, 277-307; right part). The model force applied to Oε1 of Glu58 (magenta, ball and stick) is depicted by a yellow arrow. (b) The magnitudes of the experimental (red) and predicted (green) displacements of C_α atoms are in agreement (correlation coefficient, 0.98). The scale was adjusted so as to make the average the same as that of the observation. (c) The prediction is robust with respect to the direction and position of the applied force. Correlation coefficients of >0.8 are obtained for force vectors (yellow arrows) on Oε1 of Glu58 and the attractive forces toward the ligand from the C_α atoms (red).

known that only a small number of $\lambda^{(k)}$, corresponding to low-frequency normal modes, are large enough to make significant contributions to the fluctuations [6]. Thus, proteins change their structures in the limited space spanned by the highly collective low-frequency normal modes. The role of protein-ligand interactions is to select and merge the low-frequency normal modes with signs and weights determined by the inner product between the eigenvector and the force vector. Therefore, a single normal mode is not sufficient to predict the displacement. When using the response function calculated using each of the five lowest frequency modes separately, the correlation coefficients between the predicted and crystal structures decrease from 0.95 to 0.79, 0.54, 0.20, 0.04, and 0.12.

Citrate synthase forms a homodimer under native conditions, and the monomer has an active site that catalyzes the reaction of oxaloacetate + acetyl-coenzyme A ↔ citrate + coenzyme A. Oxaloacetate induces the closure of the moving domain to form the binding pocket for acetyl-coenzyme A (Fig. 2) [18]. The external force, indicated by yellow arrows in Fig. 2(a), was applied to both monomers at the same time to predict the crystal structure of the bound form, in which both monomers bind citrate [18]. The force was modeled between Nε2 of His320 in the moving domain and the center of the reaction product, citrate, in the fixed domain based on the fact that citrate binds at the same position in the fixed domain as oxaloacetate [26]. The conformational change predicted by Eq. (3) well reproduces the experimental observation (Fig. 2) with a correlation coefficient of 0.78. When using the response function calculated by each of the five lowest frequency modes separately, the correlation coefficients are 0.00, 0.66, 0.01, 0.35, and 0.10. The near-zero coefficients for the first and third-lowest frequency modes arise from the antisymmetric motions of the homodimer in these modes, opposing the symmetric motion in the response. Calculation of the variance-covariance matrix from another 10-ns MD simulation for the isolated monomer provides very poor agreement with the observed crystal structures, with a correlation coefficient of −0.05. This means that the functional motion occurs in the conformational ensemble in the dimeric form and differs significantly from that in the monomer form.

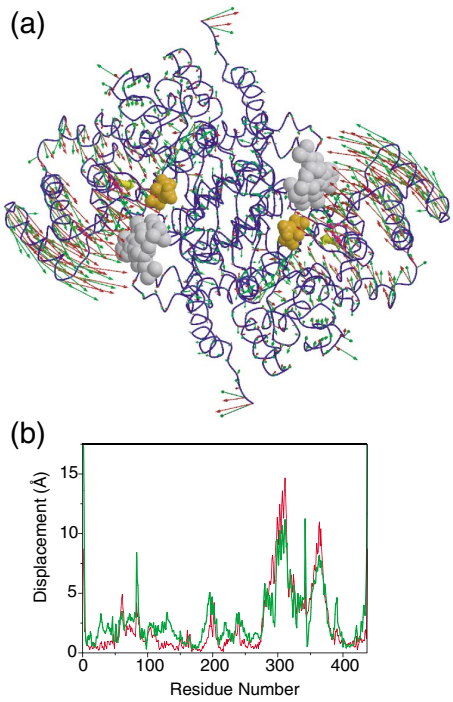

FIG. 2 (color). Structural change of citrate synthase dimer. (a) Experimental (red) and predicted (green) structural changes, represented by arrows on the backbone trace of the average structure in the simulation (blue) (correlation coefficient, 0.78). The experimental displacement is the difference between the ligand-free form (PDB ID: 1CTS) and the bound form (2CTS). The moving domain (residues 1-4, 56-65, 273-379, 436-437) undergoes motion against the fixed domain (residues 5-55, 66-272, 380-435). Citrate and coenzyme A are represented by orange and gray spheres, respectively. The force applied to Nε of His58 (magenta, ball and stick) is denoted by a yellow arrow. (b) The magnitudes of the experimental (red) and predicted (green) displacements of C_α atoms are well correlated (correlation coefficient, 0.83).

The last example is F_1-ATPase. ATP binding to the open β_E subunit in F_1-ATPase causes a closure motion, which in turn induces rotation of the central γ stalk. The crystal structure of the bound form contains ADP and SO_4 in the corresponding β subunit [20]. The external force mimicking nucleotide binding was directed from C_α of Gly161 in the P loop of the β_E subunit toward the center of the phosphate of ADP in the fixed domain [Fig. 3(a)]. The predicted motion of the β_E subunit correlates well with the difference between the two crystal structures, with a correlation coefficient of 0.84 (Fig. 3). However, the predicted rotation of the γ stalk occurs in the opposite direction to that of the crystal structure [close-up in Fig. 3(a)]. This difference is considered to arise from the fact that the nucleotide bound form of the crystal structure represents the posthydrolysis and preproduct-release step on the ATP synthesis pathway [20]. ATP binding to the β_E subunit should cause rotation of the γ stalk in the direction of the ATP hydrolysis pathway, opposite to the ATP synthesis pathway. Therefore, the present prediction indicates the correct direction of the rotation.

The three examples above confirmed the linear relationship between the equilibrium fluctuations in the ligand-free form and the structural change upon ligand binding, as expressed in the simplistic LRT formula of Eq. (3). This formula successfully described the major structural changes as responses to weak perturbations without requiring strong interactions. This predictability originates from the nature of protein dynamics in that the large equilibrium fluctuations intrinsically occur mostly in a small number of collective low-frequency normal modes [6], which may give rise to motions large and specific enough to attain the ligand-bound form. The ligand interactions therefore

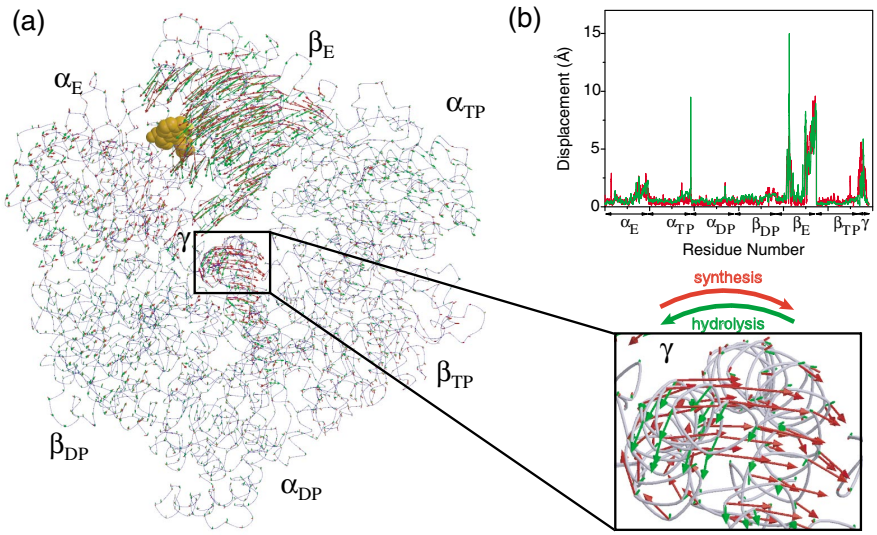

FIG. 3 (color). Structural change of F_1-ATPase. (a) Experimental (red) and predicted (green) structural changes, represented by arrows on the backbone trace of the free form (blue) (correlation coefficient of the two displacement vectors in β_E, 0.84). The experimental displacement is the difference between the nucleotide-free form (PDB ID: 1BMF) and the bound form (1H8E). Superposition of the two structures is shown for all C_α atoms. The nucleotide ligand is represented by the orange space-filling model. The force applied to C_α of Gly161 is denoted by a yellow arrow. Close-up of γ shows the direction of rotation. The experimental displacement is in the direction of ATP synthesis, while the predicted displacement is in the direction of ATP hydrolysis. (b) The magnitudes of the experimental (red) and predicted (green) displacements are well correlated (correlation coefficient, 0.85).

appear simply to bias the potential surface to initiate relaxation toward the bound form. Even in the large rotational motions in F_1-ATPase, the same model is applicable.

The authors thank Dr. J. Tame for valuable suggestions. This work was supported by Grants-in-Aid from the Ministry of Education, Culture, Sports, Science, and Technology of Japan (M. I. and A. K.). All computations were carried out at the division of Science of Biological Supramolecular Systems at Yokohama City University.

*E-mail address: kidera@tsurumi.yokohama-cu.ac.jp

[1] N. Echols, D. Milburn, and M. Gerstein, Nucleic Acids Res. **31**, 478 (2003).
[2] C.-S. Goh, D. Milburn, and M. Gerstein, Curr. Opin. Struct. Biol. **14**, 104 (2004).
[3] W. G. Krebs *et al.*, Proteins: Struct., Funct., Genet. **48**, 682 (2002).
[4] H. Valadié, J. J. Lacapčre, Y. H. Sanejouand, and C. Etchebest, J. Mol. Biol. **332**, 657 (2003).
[5] Q. Cui, G. Li, J. Ma, and M. Karplus, J. Mol. Biol. **340**, 345 (2004).
[6] S. Hayward and N. Go, Annu. Rev. Phys. Chem. **46**, 223 (1995).
[7] S. Kumar *et al.*, Protein Science **9**, 10 (2000).
[8] H. R. Bosshard, News Physiol. Sci. **16**, 171 (2001).
[9] J. P. Hansen and I. McDonald, *Theory of Simple Liquids* (Academic, New York, 1986).
[10] D. E. Koshland, Jr., Proc. Natl. Acad. Sci. U.S.A. **44**, 98 (1958).
[11] S. Hayward, J. Mol. Biol. **339**, 1001 (2004).
[12] B. Isralewitz, M. Gao, and K. Schulten, Curr. Opin. Struct. Biol. **11**, 224 (2001).
[13] J. R. Schnell, H. J. Dyson, and P. E. Wright, Annu. Rev. Biophys. Biomol. Struct. **33**, 119 (2004).
[14] C. L. Brooks III, M. Karplus, and B. M. Pettitt, *Proteins: A Theoretical Perspective of Dynamics, Structure, and Thermodynamics* (Wiley Interscience, New York, 1988).
[15] See EPAPS Document No. E-PRLTAO-94-094509 for a 5-p. supplement providing details of equations. A direct link to this document may be found in the online article's HTML reference section. The document may also be reached via the EPAPS homepage (http://www.aip.org/pubservs/epaps.html) or ftp.aip.org in the directory /epaps/. See the EPAPS homepage for more information.
[16] C. M. Bruns *et al.*, Biochemistry **40**, 15631 (2001).
[17] C. M. Bruns *et al.*, Nat. Struct. Biol. **4**, 919 (1997).
[18] S. Remington, G. Wiegand, and R. Huber, J. Mol. Biol. **158**, 111 (1982).
[19] J. P. Abrahams *et al.*, Nature (London) **370**, 621 (1994).
[20] R. I. Menz, J. E. Walker, and A. G. Leslie, Cell **106**, 331 (2001).
[21] M. M. Tirion, Phys. Rev. Lett. **77**, 1905 (1996).
[22] M. Ikeguchi, J. Comput. Chem. **25**, 529 (2004).
[23] P. A. Kollman *et al.*, *Computer Simulation of Biomolecular Systems: Theoretical and Experimental Applications*, edited by W. F. Van Gunsteren, P. K. Weiner, and A. J. Wilkinson, (Kluwer, Dordrecht, 1997), Vol. 3, pp. 83–96.
[24] W. L. Jorgensen *et al.*, J. Chem. Phys. **79**, 926 (1983).
[25] S. Hayward and R. A. Lee, J. Mol. Graphics Modell. **21**, 181 (2002).
[26] M. Karpusas, B. Branchaud, and S. J. Remington, Biochemistry **29**, 2213 (1990).