Theoretical model and simulation of protein conformational change coupled with ligand binding

Okazaki, Kei-ichi

(Degree)
博士（理学）

(Date of Degree)
2009-03-25

(Date of Publication)
2011-11-01

(Resource Type)
doctoral thesis

(Report Number)
甲4512

(URL)
https://hdl.handle.net/20.500.14094/D1004512

Doctoral Dissertation

Theoretical model and simulation of protein conformational change coupled with ligand binding

（リガンド結合と共役したタンパク質の構造変化の理論モデル構築とシミュレーション）

January 2009

Graduate School of Science and Technology,

Kobe University

Kei-ichi Okazaki

Contents

General introduction 4

References 7

Chapter 1: Multiple-basin energy landscape for large-amplitude conformational motions of proteins: Structure-based molecular dynamics simulations 9

Abstract 9

1.1 Introduction 10

1.1.1 Experimental approach 10

1.1.2 Theoretical approach 11

1.1.2.1 Quasi-harmonic regime 11

1.1.2.2 Large-amplitude regime 13

1.1.3 Energy landscape theory 13

1.2 Materials and methods 16

1.2.1 Proteins studied 16

1.2.2 Single Gō model 17

1.2.3 Double-basin model: Basic ideas 18

1.2.4 Double-basin model: Some reformations 20

1.2.5 Molecular dynamics simulation 22

1.2.5 “Multiple”-basin model 23

1.3 Results 25

1.3.1 Simulation of conformational changes 25

1.3.2 Dominant pathways of the transition 26

1.3.3 Temperature dependence of the rate 29

1.3.4 The transition rate coefficient vs the driving force; Tafel plot 30

1.4 Discussion 31

1.5 Conclusion 32

References 33

Chapter 2: Dynamic energy landscape view of coupled binding and protein conformational change: Induced-fit versus population-shift mechanisms ... 37

Abstract 37

2.1 Introduction 38

2.1.1 Experimental works: induced-fit or population-shift 38
2.1.2 Dynamic energy landscape 39
2.2 Materials and Methods 41
2.2.1 Model protein 41
2.2.2 Modeling ligand binding and protein conformational transition 41
2.2.3 Ligand-binding model 42
2.2.4 The MD-MC coupling simulations 44
2.2.5 Definition of the four states and transition rate analysis 44
2.2.6 Steady-state analysis 45
2.3 Results 46
2.3.1 Binding and conformational transitions are stochastically coupled 46
2.3.2 Free energy surface of conformational change and ligand binding 47
2.3.3 Induced-fit versus population-shift: Four states analysis 49
2.3.4 Transition rates and equilibrium constants in four states representation 50
2.4 Discussion 52
2.5 Conclusions 55
References 56

Chapter 3: Allostery in working mechanism of molecular motor; F_1-ATPase 59
Abstract 59
3.1 Introduction 60
3.2 Materials and Methods 64
3.2.1 Crystal structures of F_1-ATPase used 64
3.2.2 Simulating substep rotation of γ-subunit 64
3.2.3 Simulating allosteric interaction between ADP site and γ-subunit 67
3.2.4 Simulating allosteric coupling between ATP-binding and ADP-release 68
3.2.5 Simulation protocols 70
3.3 Results 71
3.3.1 αβ-structure corresponding to each nucleotide state 71
3.3.2 Substep rotation of γ-subunit using the identified binding dwell structure 73
3.3.3 Allosteric regulation of conformational change of the ADP site by γ-angle 75
3.3.4 γ-truncation simulation reveals allosteric coupling between ATP-binding and ADP-release 77
3.4 Conclusion 80
References 81

Conclusions of the thesis 83

Appendix 1: Methods for molecular dynamics simulation 85
Newtonian dynamics with Berendsen thermostat 85
Langevin dynamics 85
Unit system 86
References 87

Appendix 2: Eigenvalue equation for combining energy surfaces 88
Reference 89

Acknowledgements 91

General introduction

Proteins are major constituents in living cells. They constitute most of a cell's dry mass. In cells, they work as cytoskeletons, channels, pumps, catalysts, molecular motors, antibodies, hormones, and so on. The protein is a chain of connected amino acids and its sequence is determined by genetic information in deoxyribonucleic acid (DNA). The amino acid sequence has been evolutionally designed for the protein to play specific roll in cells. To function, the protein must interact with other molecules, that is, ligands (from the Latin word *ligare*, meaning "to bind"(1)). The ligands can be ions, small molecules or macromolecules like DNA or proteins. For example, enzymes bind reactants at the active site, catalyze the reaction, and release products. In cells, the sequence of the enzymatic reactions forms network of pathways, like the metabolic pathway. In the pathways, the amounts of the products are often controlled by feedback controls (1). It is necessary for cells to control the amounts of certain substances at certain timing. At molecular level, enzymes are controlled by a mechanism known as allostery (from the Greek words *allos*, meaning "other", and *stereos*, meaning "space" or "shape"). In allostery, ligands, called as activator (activating the enzyme) or inhibitor (inhibiting it), regulate the efficiency of the enzyme by binding to regulatory site that is distant from active site. Allostery is not a property limited to enzymes, but can be wildly defined as the binding of a ligand at one site causing a change in affinity or enzymatic efficiency at a distance site (2). The allosteric properties have been successfully engineered into proteins for drug design or novel biosensors (3). The allosteric effect is a major interest throughout this thesis.

Many examples have shown that the allosteric effect is caused by conformational change. The binding of the allosteric ligand to the regulatory site induces conformational change and it affects the distant site. In chapter 1, we construct a structure-based coarse-grained model that describes large-amplitude conformational changes of proteins (4). Then we analyze mechanism of conformational change depending on topology of the protein or temperature. We observed "cracking" or local unfolding during conformational changes, which is the outstanding property of proteins compared to macroscopic machines.

The allosteric effect is ultimately a coupling between ligand-binding and conformational change. The coupling mechanism has been discussed for half a century. Traditionally, this coupling was explained by the 50-years-old induced-fit model of Koshland (5), in which proteins are in their apo conformations in the unbound state, and binding to the apo forms induces conformational transition to the holo conformations. But recently, supported by NMR experiments(6, 7) and computer simulations(8), an emerging view is that proteins in the unbound state exist in many conformers that include open and closed ones and that, upon binding, the dominant population shifts from the open form to the closed one. This

"population-shift" model (known also as conformational selection or pre-existing equilibrium model), originated from the Monod-Wyman-Changeux model of allostery (9), receives more popularity in recent years (10, 11). So, there are two representative models that are mutually exclusive and try to explain the coupling mechanism; the induced-fit and the population-shift model. In chapter 2, we clarified for what systems one model fits better than the other by performing molecular simulations of coupled binding and conformational change (12).

Among various kinds of protein, motor proteins have an outstanding property to transform free energy of nucleotide triphosphate (ATP or GTP) hydrolysis to mechanical work. They often consist of several subunits and have several moving parts and nucleotide binding sites. So, the moving parts should control each other by allosteric effect and move coordinately. In chapter 3, as a target application of the allosteric model in chapter 2, we investigate working mechanism of a rotary motor, F_1-ATPase. A minimal active subcomplex for F_1-ATPase is $\alpha_3\beta_3\gamma$-subunits. The central γ-subunit rotates inside a cylinder made of $\alpha_3\beta_3$-subunits. Three β-subunits mainly host catalytic sites and change conformations depending on nucleotide state. These conformational changes apparently drive the rotation of γ-subunit. We first see coupling between γ-angle and conformational change of ADP-bound β-subunit. Then we see coupling among conformational change of Empty β-subunit, that of ADP-bound β-subunit and rotation of γ.

Lastly, we want to clarify what kind of dynamics we are interested in, and what time and length scale they have. In Fig. 0.1, time and length scale of motions involved in protein and water are summarized. There is a wide range of motions that span many orders of time and length scales. This hierarchy of protein dynamics makes complexity of this molecule. Our focus is biologically important, functional protein dynamics. In the dynamics, proteins often change backbone conformations largely. This backbone movement often becomes the rate limiting process for function. As seen in Fig. 0.1, the time scale for the large backbone movement is nano- to micro-seconds. This large movement arises from accumulation of several 10 dihedral angles' rotation ($\approx 1ns$). And the rotation of each dihedral angle is driven by thermal fluctuations of proteins or random kicks from water molecules (femto- to pico-seconds) (13). Because of this hierarchic dynamics, we have to choose appropriate methods to the phenomenon we are interested in, depending on time or length scale of it. Roughly speaking, there are two kinds of approach; bottom-up and top-down approaches.

In the bottom-up approach, we build up protein behavior from detailed atomic interactions that are based on quantum chemical calculation in first principle manner. This bottom-up approach has been the spirit of conventional all atom force fields. But, as we mentioned above, proteins show complex behavior that involves the hierarchic dynamics. Thus, it is not easy to extract essence of the complicated dynamics in this way. On the other hand, in the top-down approach,

at first, we assume minimal, empirical model (interactions) that can describe protein behavior qualitatively. At this point, the model may be too idealized to describe real protein behaviors. Thus, we introduce more detailed terms step by step if they are needed. In this step, we can understand essence of the behavior more directly. There are many examples that belong to this approach. For example, extension of ideal gas equation of state to virial equation of state to describe real gasses or elastic model of solid to describe specific heat, etc. Apparently, there is certain amount of arbitrariness in this top-down approach, but, even so, it helps to create new concepts or way of understanding nature. We take the top-down approach and assume that energy landscape of proteins is funnel-shaped. Thus, we use an idealized model (interactions) that realizes the funnel-shaped energy landscape (14), in which the native structure takes the most stable energy and interactions that destabilize the native state are neglected. This kind of model is called "Gō-like model"(15), and has been shown to reproduce folding process well (See chapter 1). We also use a coarse-grained model in which each amino acid is regarded as one bead, rather than conventional all atom models (16). In the coarse-grained model, we focus on large-amplitude motions like backbone rearrangement, rather than detailed atomic scale mechanism like side chain packing, because functional protein dynamics often involve the large-amplitude motions.

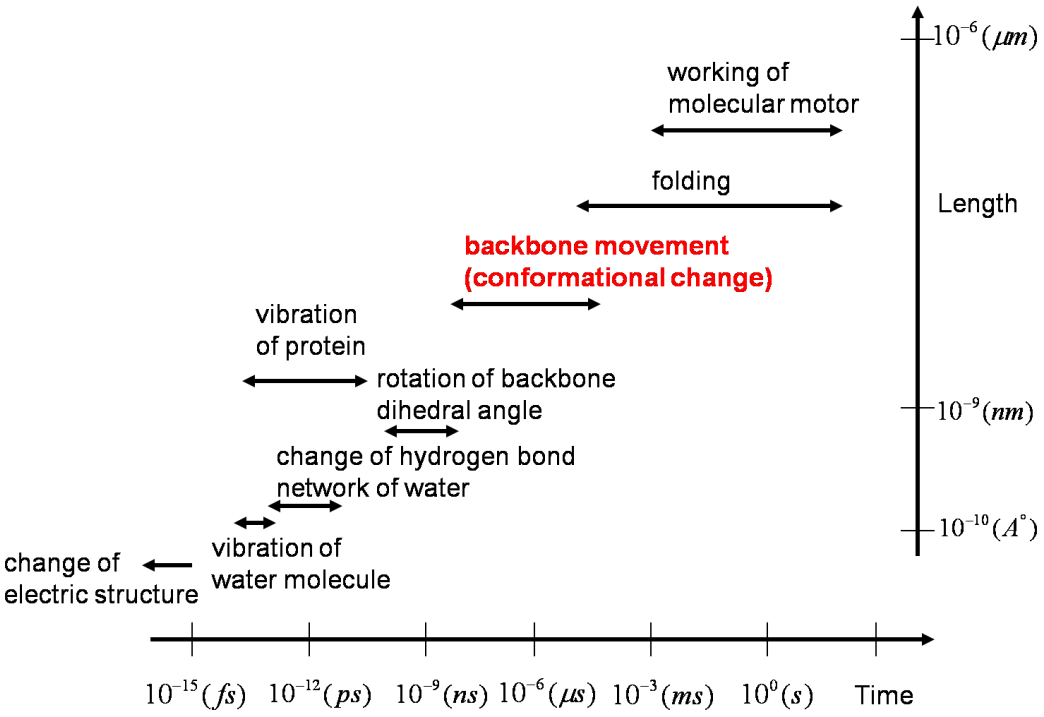

Fig. 0.1 Time and length scale of hierarchic dynamics of protein and water

References

1. Bruce Alberts AJ, Julian Lewis, Martin Raff, Keith Roberts, Peter Walter (2002) *MOLECULAR BIOLOGY OF THE CELL* (Garland Science, Taylor & Francis Group, New York).
2. Swain JF, Gierasch LM (2006) The changing landscape of protein allostery. *Curr Opin Struct Biol* 16:102-108.
3. Lee J*, et al.* (2008) Surface sites for engineering allosteric control in proteins. *Science* 322:438-442.
4. Okazaki K, Koga N, Takada S, Onuchic JN, Wolynes PG (2006) Multiple-basin energy landscapes for large-amplitude conformational motions of proteins: Structure-based molecular dynamics simulations. *Proc Natl Acad Sci U S A* 103:11844-11849.
5. Koshland DE (1958) Application of a Theory of Enzyme Specificity to Protein Synthesis. *Proc Natl Acad Sci U S A* 44:98-104.
6. Volkman BF, Lipson D, Wemmer DE, Kern D (2001) Two-state allosteric behavior in a single-domain signaling protein. *Science* 291:2429-2433.
7. Henzler-Wildman KA*, et al.* (2007) Intrinsic motions along an enzymatic reaction trajectory. *Nature* 450:838-844.
8. Ikeguchi M, Ueno J, Sato M, Kidera A (2005) Protein structural change upon ligand binding: linear response theory. *Phys Rev Lett* 94:078102.
9. Monod J, Wyman J, Changeux JP (1965) On the Nature of Allosteric Transitions: a Plausible Model. *J Mol Biol* 12:88-118.
10. Bosshard HR (2001) Molecular recognition by induced fit: how fit is the concept? *News Physiol Sci* 16:171-173.
11. James LC, Tawfik DS (2003) Conformational diversity and protein evolution--a 60-year-old hypothesis revisited. *Trends Biochem Sci* 28:361-368.
12. Okazaki K, Takada S (2008) Dynamic energy landscape view of coupled binding and protein conformational change: induced-fit versus population-shift mechanisms. *Proc Natl Acad Sci U S A* 105:11182-11187.
13. Sasai M (2008) *蛋白質の柔らかなダイナミクス (Soft dynamics of proteins)* (Baifukan, Tokyo).
14. Onuchic JN, Wolynes PG (2004) Theory of protein folding. *Curr Opin Struct Biol* 14:70-75.
15. Go N (1983) Theoretical studies of protein folding. *Annu Rev Biophys Bioeng* 12:183-210.

16. Tozzini V (2005) Coarse-grained models for proteins. *Curr Opin Struct Biol* 15:144-150.

Chapter 1: Multiple-basin energy landscape for large-amplitude conformational motions of proteins: Structure-based molecular dynamics simulations

Abstract

Biomolecules often undergo large-amplitude motions when they bind or release other molecules. Unlike macroscopic machines, these biomolecular machines can partially disassemble (unfold) and then reassemble (fold) during such transitions. Here we put forward a minimal structure-based model, the "multiple-basin model", that can directly be used for molecular dynamics simulation of even very large biomolecular systems so long as the endpoints of the conformational change are known. We investigate the model by simulating large scale motions of four proteins: glutamine binding protein, S100A6, dihydrofolate reductase and HIV-1 protease. The mechanisms of conformational transition depend on the protein basin topologies and change with temperature near the folding transition. The conformational transition rate varies linearly with driving force over a fairly large range. This linearity appears to be a consequence of partial unfolding during the conformational transition.

1.1 Introduction

To function, biomolecules often undergo large-amplitude structural changes, upon binding or releasing ligands. These structural changes organize the workings of biomolecular machines such as the ribosome, molecular chaperones and molecular motors. Structural information on the conformational ensembles before and after the conformation changes is often available through X-ray crystallography or NMR. These experiments, however, provide primarily quasi-static information. They reveal directly less about the transition dynamics between two end structures. Thus, we see global time-dependent structural information at high resolution is rarely obtained directly by experiments. Simulations can potentially provide full time-dependent structural information on biomolecular machines. Yet, conventional atomistic simulations currently only reach times up to microseconds. This time scale falls orders of magnitude short of the typical physiologically important time scales of milliseconds to seconds. To overcome this limitation, one approach is to coarse-grain the molecular representation (1). Reduction in complexity allows one to simulate much longer times. This so-called "minimalist approach" has been quite successful for studying protein folding (2-5). The purpose of this chapter is to investigate a minimal structure-based model which we call the "multiple-basin model", to simulate large-scale conformational changes when structures for the endpoints of the transition are available (6). This approach can be used for simulations of even very large biomolecular complexes.

In the following sections, first, we briefly review some previous experimental and theoretical approaches for conformational change of proteins. Then, we introduce energy landscape theory to give some basis to our model.

1.1.1 Experimental approach

Currently, one of the most powerful experimental methods to detect conformational dynamics of proteins at residue level is nuclear magnetic resonance (NMR) relaxation of backbone amide. It allows the characterization of complex internal dynamics over a broad range of time scales. Henzler-Wildman *et al.* used order parameter S^2 to see fast pico- to nanosecond motions of adenylate kinase (AdK), ranging from 0 (unrestricted) to 1 (completely restricted) (7). Volkman *et al.* used exchange value R_{ex} to characterize slow micro- to millisecond functional motions of nitrogen regulatory protein C (NtrC) (8). Lange *et al.* refined structural ensemble of ubiquitin against residual dipolar couplings (RDCs) to detect motions slower than overall protein rotational tumbling time, which is up to microseconds, and found that most of the complex with other proteins are formed through conformational selection rather than induced-fit (9).

Single-molecule fluorescence resonance energy transfer (FRET) is another important tool to

detect conformational change of proteins. This method is very informative because it detects motions of individual molecule that are not ensemble-averaged. In this method, protein is fluorescently labeled with donor at one site and acceptor at the other site. The fluorescence energy transfer obeys Foster mechanism and the efficiency decreases with -6th order of the distance between donor and acceptor. The efficiency is calculated from the fluorescence emission from donor or acceptor, that is, when the distance is short, the emission from acceptor is strong, and when the distance is long, the emission from donor becomes strong. The efficiencies are transformed to distance information (10). Henzler-Wildman *et al.* and Hanson *et al.* both applied this method to adenylate kinase (AdK) and found that this protein samples both open and closed conformations without ligand, and the population shifts to closed with ligand, indicating population-shift mechanism (11, 12).

Although NMR relaxation and FRET experiments are powerful tools for investigating dynamics, these experiments provide limited information about dynamics. For example, you cannot obtain time-dependent structural information from these experiments. In this sense, theoretical approaches that will be described below have advantage.

1.1.2 Theoretical approach

1.1.2.1 Quasi-harmonic regime

We first consider a regime in which proteins show only small deviations from the fiducial native structure. In this regime, protein is in one basin and the basin can be well approximated by quasi-harmonic. This kind of approach includes normal mode analysis and elastic network representation of proteins. The linear response approach, which is proposed recently (13), does not require harmonic potential approximation, but it describe conformational change as a perturbation to an equilibrium in single basin, so it can be referred in this context.

We start with normal mode analysis in which the potential energy is approximated as harmonic around a minimum where gradient becomes zero.

$$V(\vec{x}) \approx \frac{1}{2}\sum_{i,j}^{3N} h_{ij}\left(x_i - x_{i0}\right)\left(x_j - x_{j0}\right)$$

where Hessian matrix $H = \{h_{ij}\}$ is defined as $h_{ij} = \left.\frac{\partial^2 V}{\partial x_i \partial x_j}\right|_{\vec{x}=\vec{x}_0}$. We can solve the eigenvalue equation,

$$(M^{-1/2} H M^{-1/2})\vec{a}_i = \omega_i^2 \vec{a}_i$$
$$\vec{a}_i \cdot \vec{a}_j = \delta_{ij}$$

where $\vec{a}_i$ is normal mode direction and ω_i is frequency of the mode. The matrix M contains atomic masses on its diagonal (14). It has been shown that the directions of low-frequency modes often provide useful information for description of correlated motions (15). Normal mode analysis conventionally had been performed to atomic force fields after energy minimization.

On the other hand, the elastic network model is a simplified topology-based model often coarse-grained to one bead per residue, and its potential is a sum of harmonic potentials,

$$E = C \sum_{d_{ij}^0 < R_C} \left(d_{ij} - d_{ij}^0 \right)^2$$

where d_{ij} is the distance between atom i and atom j , d_{ij}^0 is the native distance, R_C is cutoff for connecting two beads, C is constant. Tirion performed normal mode analysis to elastic network model and showed that this simple model is enough to reproduce slow, low frequency dynamics (16). Tama and Sanejouand found that structural changes upon ligand binding usually occur predominantly in directions that correspond to combinations of a few low frequency modes (17).

Ikeguchi *et al*. have shown using their linear response theory that the direction of structural change can be predicted from the simple formula as follows (13).

$$\vec{r}_i^{\,1} - \vec{r}_i^{\,0} \cong \beta \sum_j \left\langle \Delta \vec{r}_i \Delta \vec{r}_j \right\rangle_0 \vec{f}_j$$

where the left part of the equation represents the displacement of atom i after the perturbation, $\left\langle \Delta \vec{r}_i \Delta \vec{r}_j \right\rangle_0$ is the variance-covariance matrix of atomic fluctuations in ligand-free state, and $\vec{f}_j$ is the force from ligand acting on atom j . If we diagonalize the variance-covariance matrix and rewrite the equation above,

$$\vec{X}^{\,1} - \vec{X}^{\,0} \cong \beta \sum_k \left(\vec{U}_k \cdot \vec{F} \right) \lambda_k \vec{U}_k$$

where $\vec{X}$ is collective coordinate of atoms (3N dimension vector), $\lambda_k, \vec{U}_k$ are eigenvalue and eigenvector of the variance-covariance matrix, and $\vec{F}$ is collective force from ligand. As seen in the equation, the role of the force from ligand is to give weights to principal component modes. In their results, the predicted conformational change is not dependent on the direction or position of the applied force. This suggests that the low frequency modes themselves are robust and important for functional motions.

In contrast to the quasi-harmonic approximations described here, conformational transitions must involve rearrangement of non-local contacts of amino acid pairs. Such notions clearly

require going beyond the quasi-harmonic picture: The protein breaks some contacts specific to the initial conformation and forms new contacts that are specific to the final conformation.

1.1.2.2 Large-amplitude regime

Functional dynamics of proteins often relates to large-amplitude conformational changes. Protein dynamics in this large-amplitude regime has been termed a "proteinquake"(18), and may involve "cracking"(19, 20), or local unfolding. In this regime, concepts or models that have been used in folding simulations may be useful. There has been fewer studies in this regime because conventional all atom force field simulations fail as a result of huge computational costs. Thus, coarse-grained models may have advantages. Our focus throughout the thesis is in this regime. Miyashita *et al*. proposed "nonlinear elastic model" to follow large-amplitude movements adiabatically (19). In this nonlinear elastic model, they performed normal mode analysis to the elastic network potential defined at the initial conformation and move the conformation slightly to the direction of modes that overlaps mostly with the direction to the final conformation. Then, they redefine the elastic network potential at the conformation and again move the conformation in the same way. They iteratively performed the process and generated the nonlinear conformational change path. They calculated strain energy along the nonlinear conformational change path. They also applied the same model to the final conformation and the crossing point of reactant and product energy surfaces becomes the transition state. They showed that cracking has to occur to give sufficiently low free energy barrier. They also showed that cracking leads to linear behavior of the rate versus driving force over a large range of reaction driving forces (19).

1.1.3 Energy landscape theory

As we described above, large-amplitude conformational changes often involve cracking, or local unfolding. Thus, it may be useful to review concepts or models that have been used in protein folding studies. The energy landscape theory has been a central concept describing protein folding (21, 22), and simulation models that are based on this concept have reproduced experimental results (23, 24). The energy landscape theory gives us an answer to a fundamental question; why proteins fold into unique native structures so fast?

The schematic energy landscape of random poly-peptide is shown in Fig. 1.1(a). If we call a structure that takes minimum effective energy (sum of protein conformational energy and solvation free energy) "native" structure, there are many "miss-fold" structures that have equally low energies as the native structure in the energy landscape. In this type of energy

landscape, it is difficult for proteins to refold into the native structure and highly probable to fall into miss-fold structures. Therefore, spontaneous folding to the native structure is impossible in this type of energy landscape.

Then, what type of energy landscape realizes the spontaneous folding to the native structure? In the 80s-90s, theoreticians came up with the idea that the energy landscape is like Fig. 1.1(b) (22). In this energy landscape, the native structure is at the bottom of the energy landscape, and global shape of the energy landscape is funnel-like. In this energy landscape, it is clear that the protein once denatured by some conditions can fold back to the native structure when the solvent condition is set to the physiological one. Therefore, proteins evolve from random poly-peptide chains adjusting their amino acid sequences to shape up the energy landscapes to funnel-like. In other words, proteins minimized frustrations (interactions that destabilize the native structure) to fold fast (the minimal frustration principle) (25).

Gō model is a standard model that is based on the energy landscape theory and realizes funnel-like energy landscape. In Gō model, residue pairs that interact in the native structure are called “native contacts”, and attractive interaction is considered to native contacts only. The minimal frustration principle or the consistency principle that was originally proposed by Gō (26) states that we can disregard non-native interactions as a zero-th order of approximation (24). Physicochemical character of amino acid sequence is not considered in this model. Surprisingly, this simple model reproduces transition state ensemble for some proteins and explains folding rate dependency on relative contact order (RCO) (27) that represents complexity of native structures quantitatively for small two state proteins (28). Moreover, Gō model also reproduces the quasi-harmonic motion in the limit of the weak fluctuations, as well as folding reaction (29). So, it can have the ability to describe large-amplitude conformational transitions, modeling both the quasi-harmonic fluctuations around each basin and the transient and partial unfolding near the transition region.

In the standard energy landscape for protein folding, however, only a single dominant minimum, which corresponds to the native structure, is assumed. Studying conformational changes of functional proteins requires more than one basin to be taken into account. Each basin should correspond to the structure with or without ligands. The standard structure based models are not directly applicable to this case of multiple basins. In this chapter, we explicitly build up an energy landscape encoding multiple near degenerate basins.

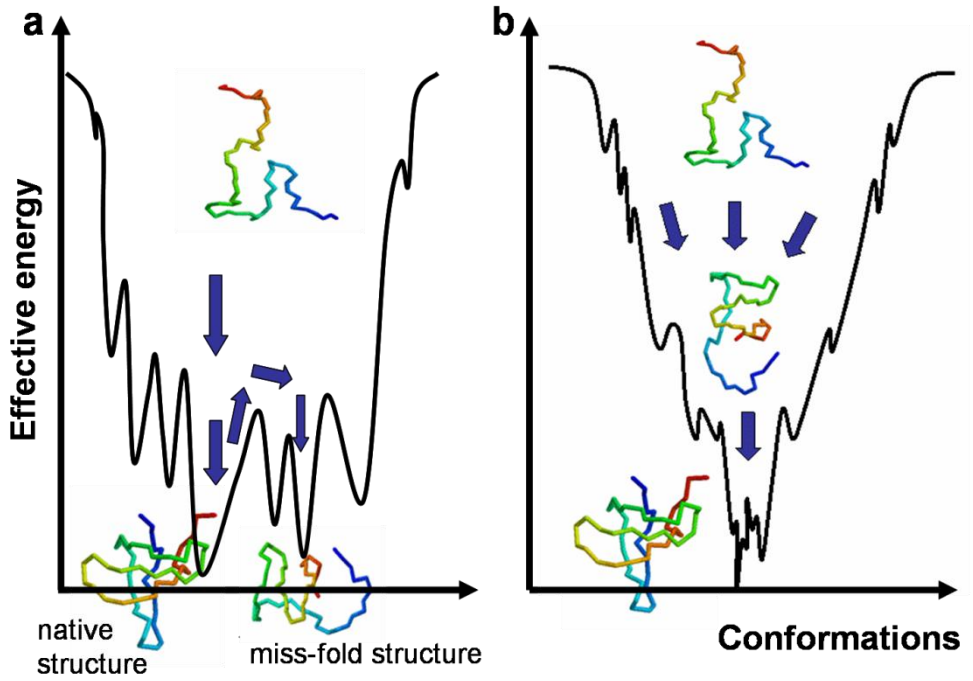

Fig. 1.1 Energy landscapes (a) rugged energy surface of random poly-peptide **(b)** funnel-like energy surface of protein

1.2 Materials and methods

1.2.1 Proteins studied

We studied four proteins, glutamine binding protein (GBP), S100A6, dehydrofolate reductase (DHFR) and HIV-1 protease. For each, two PDB structures were used to construct the multiple-basin model. GBP is composed of two domains. Without substrates, GBP is found in an open form (1ggg, red in Fig. 1.2(a)), and upon glutamine binding, the hinge between domains swings forming the closed structure (1wdn, green). S100A6 is a structural analog of calmodulin. Its conformational change, a shear motion, occurs between the apo state (1k9p, red in Fig. 1.2(b)) and the calcium-bound holo state (1k9k, green). HIV-1 protease has a β-hairpin loop (the flap) that adopts an open conformation (3hvp, red in Fig. 1.2(c)) without an inhibitor but acquires a closed conformation (4hvp, green) with an inhibitor. The conformational change is relatively small and is of the hinge type. DHFR changes its active site loops via a shear motion between the occluded state (1rx6, red in Fig. 1.2(d)) and the closed state (1rx2, green).

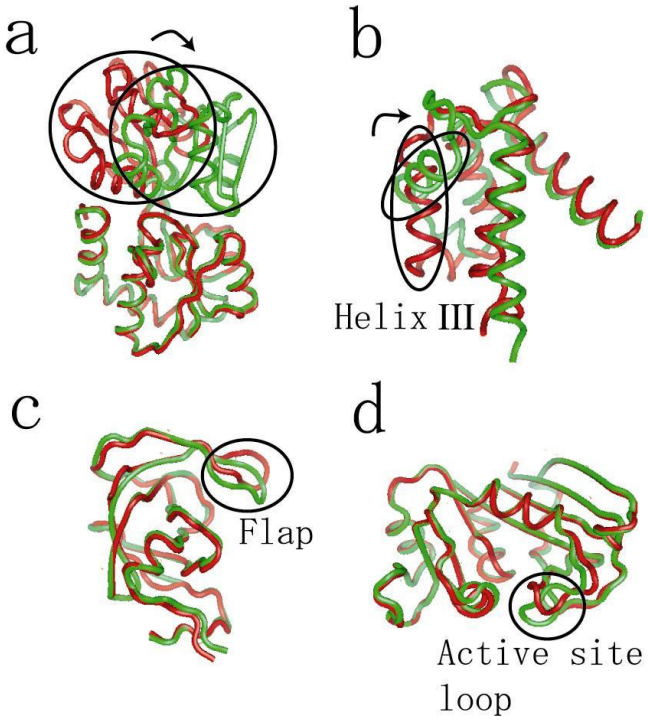

Fig. 1.2 Two PDB structures that are used in simulation (a) GBP **(b)** S100A6 **(c)** HIV-1 protease **(d)** DHFR. Red corresponds to unbound structure, though for DHFR, occluded state. And green corresponds to bound structure. Figures are created by PyMOL.(30)

Protein	PDB code 1 (unbound)	PDB code 2 (bound)	Residue number	Number of contacts per residue	Mechanism of Motion*
GBP	1ggg	1wdn	220	6.90	Hinge
S100A6	1k9p	1k9k	89	5.24	Shear
DHFR	1rx6	1rx2	159	6.42	Shear
HIV-1 protease	3hvp	4hvp	99	5.61	Hinge

Table 1.1 Characteristics of proteins used

*****Classification from "Database of Macromolecular Movement"(31)

1.2.2 Single Gō model

There are many versions of Gō models and their variants, and here we describe the version of Clementi *et al.*(4) that has been extensively used for folding simulation and more importantly will be the starting point of my work in the subsequent sections. In this model, amino acid is represented by a single bead that is located at Cα atom. The potential energy function can be divided into two types, one is local interactions, $V_{bond}, V_{\theta}, V_{\phi}$ and the other is non-local interactions, $V_{contact}, V_{repulsion}$. "Local" interaction means that each term of interactions involves amino acids that are close in sequence. The other interactions are called "non-local". Among local interactions, V_{bond} is the bond energy, V_{θ} is the bond-bond angle energy, V_{ϕ} is dihedral angle energy. All of these potentials are biased toward the native structure. Non-local interactions are different between native contact pairs and non-native pairs. Native contact here is defined as the amino acid pair that has non-hydrogen atoms of one amino acid within a distance of 6.5Å from any non-hydrogen atom of the other amino acid. $V_{contact}$ is the attractive interaction of native contacts, which biases towards the native conformation. $V_{repulsion}$ is the repulsive interaction of non-native contacts. Except for $V_{repulsion}$, all of energy terms take the minimum values at the native structure. In practice, V_{bond} does not change so much between

random coil and the native structure, and thus actually, $V_\theta, V_\phi, V_{contact}$ are the terms that makes funnel-like energy landscape toward the native structure.

If we write the potential function that based on the native conformation R_0, at conformation R, explicitly,

$$\begin{aligned} V(R \mid R_0) &= V_{bond} + V_\theta + V_\phi + V_{contact} + V_{repulsion} \\ &= \sum_{bonds} K_r (r_i - r_{0i})^2 \\ &+ \sum_{angles} K_\theta (\theta_i - \theta_{0i})^2 + \sum_{dihedral} \{K_\phi^{(1)}[1-\cos(\phi_i - \phi_{0i})] + K_\phi^{(3)}[1-\cos 3(\phi_i - \phi_{0i})]\} \\ &+ \sum_{\substack{native \\ contact \\ i>j-3}} \varepsilon_1 \left[5\left(\frac{r_{0ij}}{r_{ij}}\right)^{12} - 6\left(\frac{r_{0ij}}{r_{ij}}\right)^{10} \right] + \sum_{\substack{non-native \\ contact \\ i>j-3}} \varepsilon_2 \left(\frac{C}{r_{ij}}\right)^{12} \end{aligned}$$

where r_i is ith bond length that is the distance between ith and $i+1$th Cα atoms, and θ_i is ith bond angle that is the angle ith, $i+1$th and $i+2$th Cα atoms make, and ϕ_i is dihedral angle that is the angle ith, $i+1$th, $i+2$th and $i+3$th Cα atoms make. r_{ij} is the distance between ith and jth amino acid residues. All the parameters with subscription "0" mean the values of the corresponding variables at the native structure. For other parameters, $K_r = 100.0$, $K_\theta = 20.0$, $K_\phi^{(1)} = 1.0$, $K_\phi^{(3)} = 0.5$, $\varepsilon_1 = 0.18$, $\varepsilon_2 = 0.18$, $C = 4.0$ Å.

Energy unit is arbitrary, but relative value compared to thermal energy $k_B T$ has meaning (See Appendix 1).

1.2.3 Double-basin model: Basic ideas

Here, in order to build up an energy landscape encoding a number of funnel-shaped basins, we first apply Gō potential to each structure of functional protein, and then merge the Gō potentials to make the energy surface that allows transitions between different basins. When two conformations of the functional protein are obtained, first, Gō potential that based on the native conformation $R_\mu (\mu = 1,2)$ at a given protein conformation R, $V(R \mid R_\mu)$ is calculated for each conformation. Then we merge the two Gō potentials to make the potential, V_{MB}, that has two funnel-shaped minima. The way we merge the two potential is analogous to what has often been used to describe the quantum mechanical coupling of two potential energy surfaces (empirical valence bond-like approach (32-34)). That is, we define V_{MB} by the eigenvalue of

characteristic equation as follows (See Appendix 2).

$$\begin{pmatrix} V(R \mid R_1) & \Delta \\ \Delta & V(R \mid R_2) + \Delta V \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = V_{MB}(R \mid R_1, R_2) \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$$

where Δ is the coupling constant that is now constant for each protein, and ΔV is constant that change relative stability of two states. The condition that eigenvector is non-zero gives us the secular equation as follows.

$$\begin{vmatrix} V(R \mid R_1) - V_{MB}(R \mid R_1, R_2) & \Delta \\ \Delta & V(R \mid R_2) + \Delta V - V_{MB}(R \mid R_1, R_2) \end{vmatrix} = 0$$

By solving this equation, we obtain as the lowest energy solution,

$$V_{MB}(R \mid R_1, R_2) = \frac{V(R \mid R_1) + V(R \mid R_2) + \Delta V}{2} - \sqrt{\left(\frac{V(R \mid R_1) - V(R \mid R_2) - \Delta V}{2} \right)^2 + \Delta^2}$$

The eigenvector $\begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$ indicates whether the system is in the state 1 or state 2, and so we can use $\ln(c_2/c_1) = \chi$ as a reaction coordinate for conformational change. When the protein is in the basin of conformation 1, i.e. $c_1 > c_2$, χ takes negative value, and when the protein is in the basin of conformation 2, i.e. $c_1 < c_2$, χ takes positive value.

We have two parameters, Δ and ΔV. Δ is the coupling constant that determines the energy barrier height. ΔV is the constant that is added to $V(R \mid R_2)$, to change relative energy height of two minima.

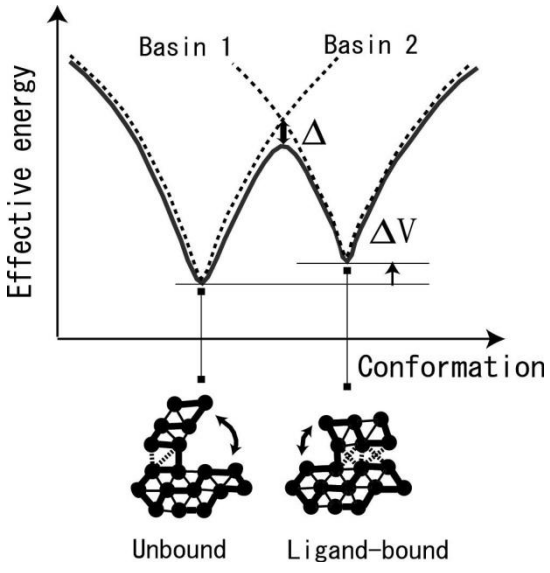

Fig. 1.3 Multiple-basin energy landscape
Two single funnels used for model construction are depicted by dashed lines. Conformational change is associated with the rearrangement of some contacts. Contacts that are specific to the unbound conformation are broken, and new contacts are formed in the bound conformation.

1.2.4 Double-basin model: Some reformations

We make some reformations to original Gō model that was described before. I described that $V_\theta, V_\phi, V_{contact}$ are the terms that are biased toward native structure. In the context of multiple-basin model, the native structure should be called as "reference structure" because there is more than one structure. The reformations are related to these energy terms. The problem is that these energy terms are too strongly biased toward each of the reference structures. Because of this, when we consider the merged potential of two funnel-shaped potentials, the transition regions have unnaturally high energy and transition takes almost forever.

First, we improve on the local energy terms, V_θ, V_ϕ. We change energy coefficients, $K_\theta, K_\phi^{(1)}$ and $K_\phi^{(3)}$. We calculate strain energies between one reference conformation and the other for each bond angle and dihedral angle, i.e. $K_\theta\left(\theta_{oi}^1 - \theta_{oi}^2\right)^2$ and $K_\phi^{(1)}[1-\cos(\phi_{0i}^1 - \phi_{0i}^2)] + K_\phi^{(3)}[1-\cos 3(\phi_{0i}^1 - \phi_{0i}^2)]$ for each i, where $\theta_{0i}^1, \theta_{0i}^2$ is i th bond angle in reference conformation 1,2 and ϕ_{0i}^1, ϕ_{0i}^2 is i th dihedral angle in reference conformation 1,2, with $K_\theta = 20.0, K_\phi^{(1)} = 1.0, K_\phi^{(3)} = 0.5$. Therefore, the strain energy at the i th bond indicates how much the angle changes between two conformations. There must be a tendency that bond angles (dihedral angles) that take highly different values in two conformations are flexible. Otherwise, angles could not change that large. So, we set the threshold for bond (dihedral) angle strain energy to 1.0 (0.5), and if the strain energy of a bond angle(dihedral angle) between two native conformations exceed this threshold, we reduce the constant K_θ ($K_\phi^{(1)}, K_\phi^{(3)}$) so that the strain energy becomes equal to this threshold.

Second, we change the non-local energy term, $V_{contact}$. Amino acid pairs of the protein are divided into three types:

Type 1: a pair that makes a contact in both conformations
Type 2: a pair that makes contact in one conformation but not in the other
Type 3: a pair that does not make any contact in either conformation

In the type 3, there is no problem because of the exact same repulsive interaction in both conformations. In the type 1 and 2, we have to do some reformations to overcome the high energy barrier problem.

In the type 1, the amino acid pair is native contact in both conformations, and has different native contact distances in conformation 1 and 2, $r_{0ij}^{(1)}, r_{0ij}^{(2)}$. If $r_{0ij}^{(1)} < r_{0ij}^{(2)}$, when the protein tries to transit from conformation 2 to 1, in the original Gō model, it needs a high energy for r_{ij} to go from $r_{0ij}^{(2)}$ to $r_{0ij}^{(1)}$ as shown in Fig. 1.4(a), left. The problem is the native contact energy function that has larger contact distance. So we change the energy form as shown in Fig. 1.4(a), right. That is, for the repulsive part (energy takes positive value), we use the same repulsive interaction. For the attractive part (energy takes negative value), we use the original energy form. Between them, we set the energy equal to zero.

In the type 2, the amino acid pair makes contact in one conformation but non-native contact in the other conformation. The problem is the native contact energy increases more sharply than the non-native contact energy when r_{ij} goes smaller. So we change energy form of the non-native contact energy so that the repulsive energy same as that of native contact energy is utilized. See Fig. 1.4(b).

Mathematically, we can express this reformulation as follows. We divide the non-local interaction into two terms $V_{native-attr}$ and V_{repul}, $V_{native-attr}$ is native contact attraction interaction, and given as,

$$V_{native-attr}\left(\Gamma \mid \Gamma_{\mu}\right) = \varepsilon_1 \sum_{i>j-3}^{native-contact} \min\left\{1{,}5\left(\frac{r_{0ij}}{r_{ij}}\right)^{12} - 6\left(\frac{r_{0ij}}{r_{ij}}\right)^{10} + 1\right\}$$

where plus 1 represents offset to set potential minimum values to zero.

V_{repul} consists of two sums, one is a sum of pairs that at least in one structure are native contacts (Type1&2), and another is a sum of pairs that is not native contacts in both structures (Type3). Now V_{repul} is given as,

$$V_{repul}\left(\Gamma \mid \Gamma_{\mu}\right) = \sum_{i>j-3}^{Type1\&2} \max\left\{0, \varepsilon_1\left[5\left(\frac{r_{0ij}^{\min}}{r_{ij}}\right)^{12} - 6\left(\frac{r_{oij}^{\min}}{r_{ij}}\right)^{10}\right]\right\} + \sum_{i>j-3}^{Type3} \varepsilon_2\left(\frac{C}{r_{ij}}\right)^{12}$$

where $r_{0ij}^{\min} = \min_{\{\mu\}} \left\{ r_{0ij}^{(\mu)} \right\}$.

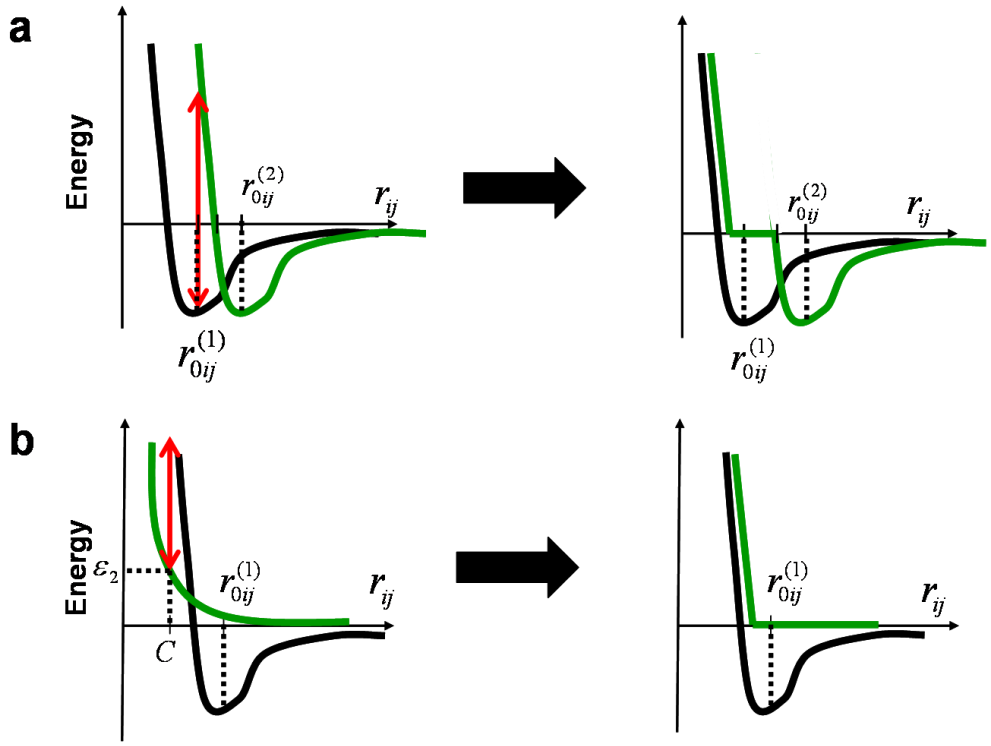

Fig. 1.4 Reformations of non-local energy (a) type 1 contact pair **(b)** type 2 contact pair

1.2.5 Molecular dynamics simulation

In this work, molecular dynamics (MD) simulation was carried out using the constant temperature Newtonian dynamics. The velocity Verlet algorithm was used for time propagation and the temperature was controlled by the simple Berendesen thermostat (See Appendix 1). The mass for all amino acids are set to be the same.

The force of i th bead is given as,

$$\mathbf{F}_i = -\frac{\partial V_{MB}}{\partial \mathbf{r}_i} = -\left[\frac{\partial V_{MB}}{\partial V(R \mid R_1)} \frac{\partial V(R \mid R_1)}{\partial \mathbf{r}_i} + \frac{\partial V_{MB}}{\partial V(R \mid R_2)} \frac{\partial V(R \mid R_2)}{\partial \mathbf{r}_i} \right]$$

$$= \frac{\partial V_{MB}}{\partial V(R \mid R_1)} \mathbf{F}_i^{(1)} + \frac{\partial V_{MB}}{\partial V(R \mid R_2)} \mathbf{F}_i^{(2)}$$

$\mathbf{F}_i^{(1)}, \mathbf{F}_i^{(2)}$ is the force from the potential energy that is based on conformation 1,2, respectively.

The simulation temperatures were determined using a crude estimate of the folding transition temperature T_F for each reference structure of studied proteins. For S100A6 and DHFR, T_F was calculated by the protocol of Koga *et al.*(28). Following this, we assumed that T_F scales with the number of contacts per residue. Thus, using the data on S100A6 and DHFR, we estimated the T_F's of other proteins. Default simulation temperatures of conformational change were set as $0.8T_F^{(min)}$, where $T_F^{(min)}$ is the smaller of the two T_F's associated with two reference structures. Preliminary tests indicated that at higher temperatures, proteins would globally unfold while at lower temperatures, the conformational transitions were too slow to sample in reasonable simulation times.

It was necessary to fix the values of the two parameters introduced for the multiple-basin model, the coupling term Δ and the relative stability ΔV. The former controls the energy barrier between two states; the larger negatively is Δ , the smaller is the barrier. ΔV modulates the relative stability of two states. Both parameters can be determined using experimental input. For convenience, we adjusted these parameters so that reversible transitions between two conformations were realized for each of the proteins studied. Starting with a fairly small value, Δ is gradually increased until the transitions take place within acceptable computation times. Second, ΔV was tuned so that transitions from each conformation take place with equal frequency. The resulting value is ΔV_{eq}. The parameters so obtained are (in units of k_BT), Δ=-59 and ΔV_{eq} = –4.4 for GBP, Δ=-66 and ΔV_{eq} =11 for S100A6, Δ=-6 and ΔV_{eq} = –2 for DHFR, and Δ=-20 and ΔV_{eq} =2 for HIV-1 protease.

1.2.5 "Multiple"-basin model

When we generally consider n conformations of functional protein, we should solve $n \times n$ determinant to obtain merged potential energy, that is,

$$\begin{vmatrix} V(R \mid R_1) - V_{MB}(R \mid R_1, \cdots, R_n) & \Delta & \cdots & \Delta \\ \Delta & V(R \mid R_2) - V_{MB}(R \mid R_1, \cdots, R_n) & \Delta \cdots & \Delta \\ \vdots & \Delta & \ddots & \vdots \\ & \vdots & & \Delta \\ \Delta & \Delta & & V(R \mid R_n) - V_{MB}(R \mid R_1, \cdots, R_n) \end{vmatrix} = 0$$

Practically, unless more than three $V(R|R_v)$ are near degenerate, we can solve 2x2 equations with two smallest V's. In case of the degeneracy, calculating the force on each residue needs a slightly involved procedure. We express above $n \times n$ determinant as D and differentiate the equation $D = 0$ by the coordinate of i th bead,

$$\frac{\partial D}{\partial \mathbf{r}_i} = \frac{\partial D}{\partial d_{11}} \frac{\partial (V(\Gamma \mid \Gamma_1) - V_{MultiGo})}{\partial \mathbf{r}_i} + \frac{\partial D}{\partial d_{22}} \frac{\partial (V(\Gamma \mid \Gamma_2) - V_{MultiGo})}{\partial \mathbf{r}_i} + \cdots + \frac{\partial D}{\partial d_{nn}} \frac{\partial (V(\Gamma \mid \Gamma_n) - V_{MultiGo})}{\partial \mathbf{r}_i}$$

$$= \tilde{d}_{11}\left(-\mathbf{F}_i^{(1)} + \mathbf{F}_i\right) + \tilde{d}_{22}\left(-\mathbf{F}_i^{(2)} + \mathbf{F}_i\right) + \cdots + \tilde{d}_{nn}\left(-\mathbf{F}_i^{(n)} + \mathbf{F}_i\right) = 0$$

$$\mathbf{F}_i = \sum_{i=1}^{n} \tilde{d}_{ii} \mathbf{F}_i^{(i)} \Big/ \sum_{i=1}^{n} \tilde{d}_{ii}$$

where $\tilde{d}_{ij}$ is minor determinant.

1.3 Results

1.3.1 Simulation of conformational changes

We first illustrate the results using glutamine binding protein (GBP). GBP is composed of two domains connected by a hinge (See Fig. 1.2(a)). Without substrates, GBP is found in an open form (red in the figure). Upon binding to glutamine, the hinge swings to make the closed form (green). Using the open form for basin 1 and the closed form for basin 2 as fiducial structures, we tuned ΔV so that the protein spends equal time in each basin. At this tuning ligand/protein concentration coincides with the dissociation constant. The simulation temperature T was set to 0.8 times the folding transition temperature (See Materials and Methods). One finds $\Delta V_{eq} = -4.4 k_B T$. We obtained a reversible transition between two basins as in Fig. 1.5(b). Here, we see that the protein resides in each basin for reasonably long times. The transition occurs infrequently but, very rapidly without any detectable intermediate state.

We plot the free energy profile $F(\chi)$ in Fig. 1.5(d). The open and closed conformations $\chi \sim \pm 1.3$ have roughly the equal free energies separated by a single free energy barrier of modest height $\sim 7 k_B T$. In contrast, the average energy profile $\langle E(\chi) \rangle$ shown in Fig. 1.5(e) suddenly increases at $\chi \sim \pm 0.7$ and the purely energetic contribution to the barrier becomes $\sim 32 k_B T$. This large increase in energy is compensated by an entropic contribution $-TS = F - \langle E \rangle \sim -25 k_B T$ (Fig. 1.5(e)). The sudden increase in conformational entropy at $\chi \sim \pm 0.7$ is the hallmark of cracking(19). Importantly, the ability to crack drastically lowers the free energy barrier. We also simulated the conformational change with $\Delta V = 0 k_B T$, where the open conformation is more stable (Fig. 1.5(a) and (d)), and with $\Delta V = -8.9 k_B T$, where the closed conformation is more stable (Fig. 1.5(c) and (d)).

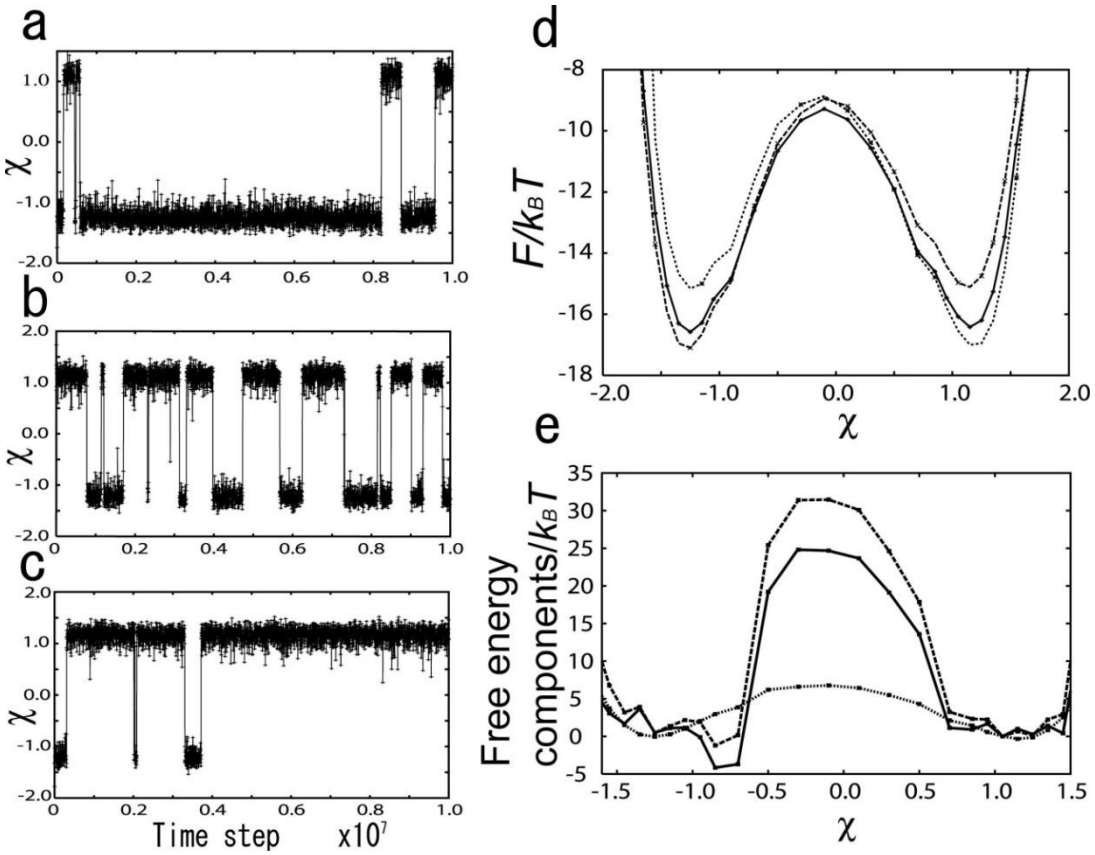

Fig. 1.5 Trajectories and free energy profiles of conformational changes of GBP (a) A trajectory with $\Delta V = 0 k_B T$ **(b)** with $\Delta V = -4.4 k_B T$ **(c)** with $\Delta V = -8.9 k_B T$ **(d)** The free energy profiles with three different values of ΔV. The dashed curve corresponds to $\Delta V = 0 k_B T$, the solid curve corresponds to $\Delta V = -4.4 k_B T$ and the dotted curve corresponds to $\Delta V = -8.9 k_B T$. **(e)** Energetic $\langle E \rangle$ (dashed line) and entropic TS (solid line) contributions to the free energy profile (dotted line) for the case of $\Delta V = -4.4 k_B T$.

1.3.2 Dominant pathways of the transition

How are contacts specific to the initial basin broken and how are the new contacts specific to the final basin formed? To quantitatively answer this question, we used three types of *Q* scores (i.e., the fraction of formed contacts). The contacts of the two reference structures 1 and 2 are classified into three types, **1)** those that are unique to structure 1, **2)** those that are unique to structure 2, and **3)** those contacts that are common to both structures 1 and 2. For each, we define the fraction of those contacts actually formed for any given structure; namely, *Q*(struct 1) for the type 1 contact set, *Q*(struct 2) for the type 2, and *Q* (common) for the type 3. A contact is defined as "formed" when its pair distance falls within a distance 1.15 times that of the reference structure.

The free energy surface for GBP is drawn on the *Q*(closed)- *Q*(open) plane in Fig. 1.6(a). A representative trajectory, superimposed on this surface, illustrates the typical transition dynamics. There are two free energy minima corresponding to the open (top left basin) and closed (right basin) states. These two minima are connected by a straight valley, indicating that breaking of contacts specific to the initial basin and formation of contacts specific to the final basin occur simultaneously. The simultaneity of the transitions is characteristic of the GBP topology change.

The corresponding free energy surface for S100A6 is drawn in Fig. 1.6(b). S100A6 is the calcium binding domain, a structural analog of a domain of calmoduin (structures depicted in Fig. 1.2(b)). The conformational change from the apo (unbound) to holo (bound) states involves an 86^{o} reorientation of helix III leading to a relatively large-scale shear motion. In contrast to the GBP case, the free energy surface for S100A6 suggests that, upon changing from apo to holo, first the contacts specific to apo are broken, followed by forming contacts specific to holo. The transition is sequential.

The different characteristics of the two free energy surfaces reflecting distinct mechanisms of conformational change may in part be attributed to the difference in the type of motion; a hinge type motion for GBP and a shear type motion for S100A6. In the inter-domain hinge motion of GBP, the residues that lose contacts upon conformational change are different from those that gain new contacts. Thus, disrupting some contacts and forming new contacts can proceed concomitantly. In contrast, the shear motion in S100A6 requires the same residues to exchange contact partners and thus has to be inherently sequential.

$\alpha(\text{hinge}) = \left| n_{A\cap\bar{B}}(i) - \quad n_{\bar{A}\cap B}(i) \right|$ is plotted over residues *i* for GBP (green) and S100A6 (red) in Fig. 1.7(a). Here A (B) stands for the contact set of the structure 1 (2). $n_{\mathrm{A}\cap\bar{\mathrm{B}}}(i)$ is the number of contacts that involves the residue *i* in the set of $\mathrm{A}\cap\bar{\mathrm{B}}$. The large α implies that, upon conformational change from structure 1 to 2, residues that lose contacts and those that gain contacts tend to be separated, which is a characteristics of the hinge type motion.

$\beta(\text{shear}) = \text{Min} \quad \left[n_{A\cap\bar{B}}(i), \quad n_{\bar{A}\cap B}(i) \right]$ is plotted over residues *i* for GBP (green) and S100A6 (red) in Fig. 1.7(b). The meanings of symbols are the same as above. Here, the large β implies that, upon conformational change, many residues tend to exchange their contact partners, i.e., the same residues lose and gain their contacts, which may be the characteristics of the shear motion.

For DHFR and HIV-1 protease, in which the conformational changes are rather smaller than these systems, no remarkable features are apparent in the free energy surface in *Q* (Fig. 1.6(c),(d)).

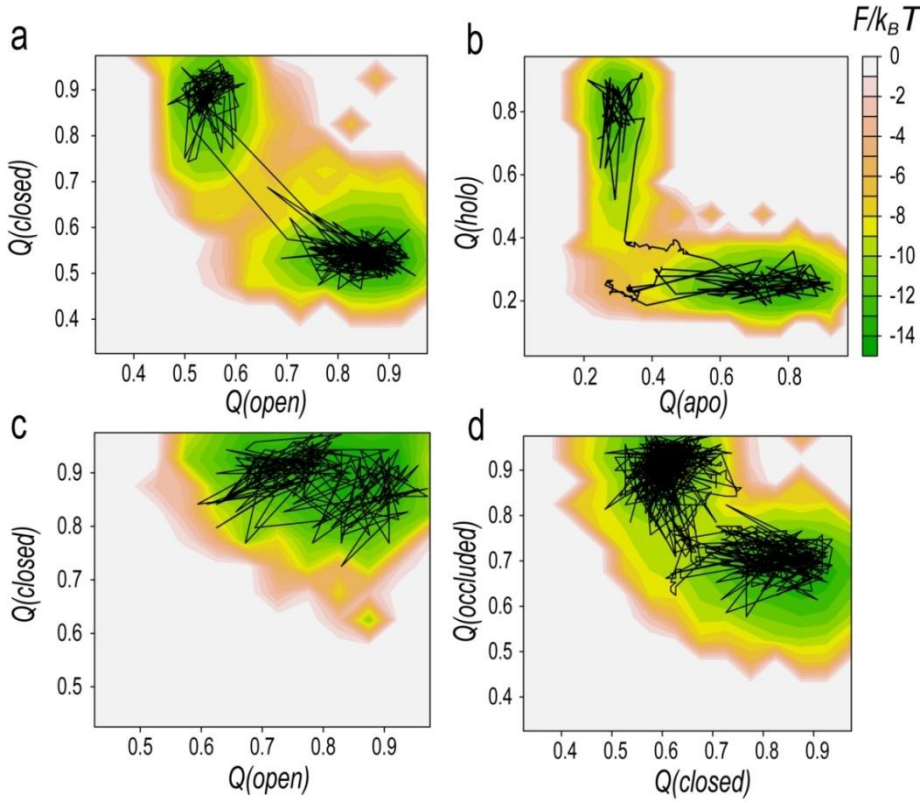

Fig. 1.6 Free energy surfaces of conformational changes of four proteins (a) Conformational change of GBP, **(b)** S100A6, **(c)** HIV-1 protease, **(d)** DHFR. Representative trajectories are superimposed on the surfaces.

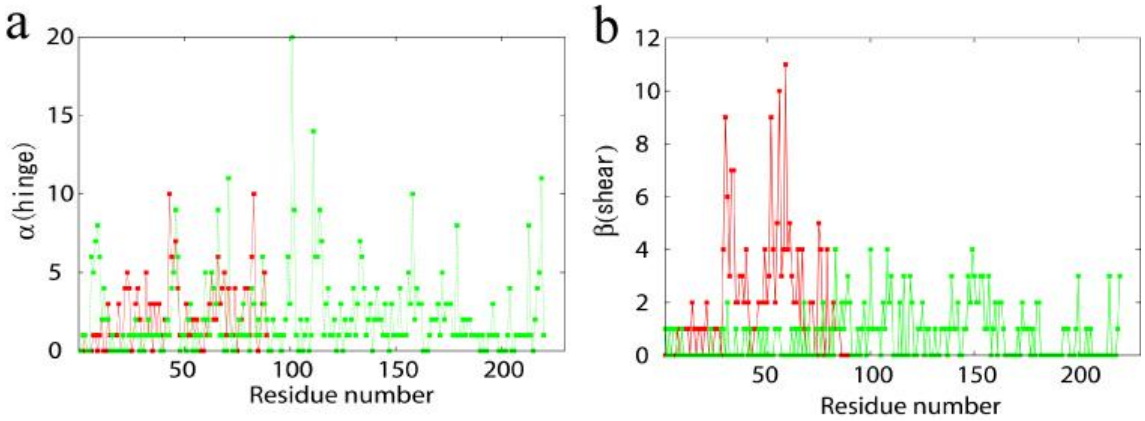

Fig. 1.7 Characteristics of contacts for hinge or shear motion (a) α(hinge) characterizes hinge type motion. Green line is for GBP, red line is for S100A6. **(b)** β(shear) characterizes shear type motion. Green line is for GBP, red line is for S100A6.

1.3.3 Temperature dependence of the rate

At low temperature, proteins do not have enough thermal energy to break contacts, and transitions will be very slow. When the temperature reaches nearly the folding temperature T_F, the conformational change becomes coupled with transient but global unfolding.

We compare the free energy surfaces of GBP at T=0.8T_F (Fig. 1.8(a)) with that at T=0.88T_F (Fig. 1.8(b))(Here, T_F is that of the open form). Using Q(open)-Q(close) and Q(common) the former functions as the conformational reaction coordinate and the latter monitors local and global unfolding of the core. There was no qualitative change of the surface between two temperatures, but the population shifts to the open form at the elevated temperature.

In contrast, the same analysis for S100A6 reveals a change of mechanism. At 0.80T_F, the protein proceeds directly from one basin to the other (Fig. 1.8(c)), but when the temperature is increased to 0.88T_F, an additional free energy minimum emerges (Fig. 1.8(d)). Conformations in the minimum are somewhat extended because of loose packing between helices. At this higher temperature there are two possible paths. One direct, the other path via an extended intermediate conformation.

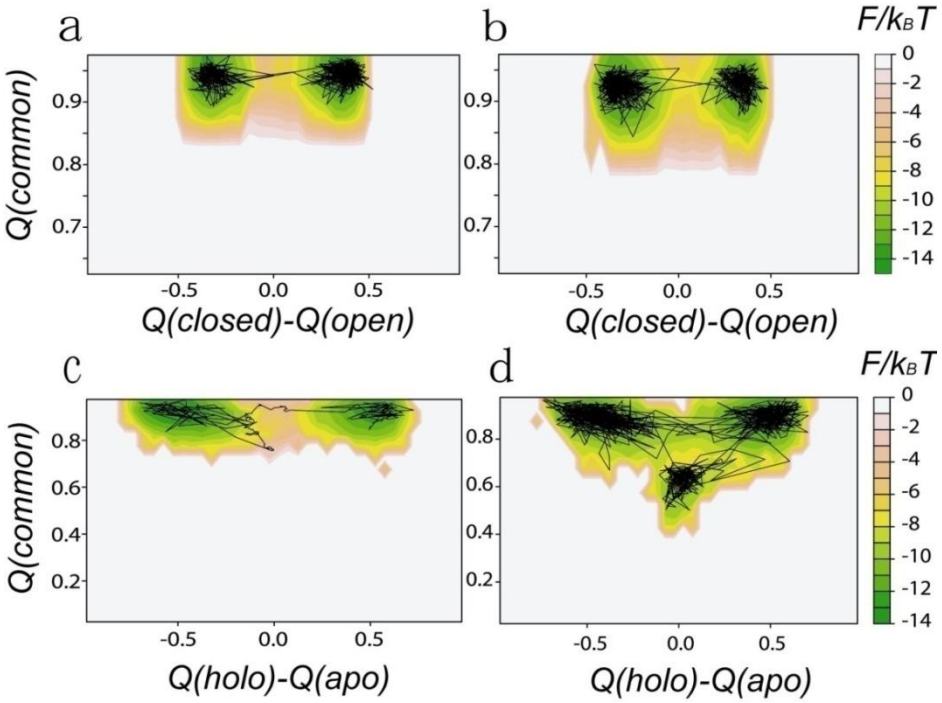

Fig. 1.8 Temperature dependence of conformational change dynamics of GBP and S100A6 **(a)** GBP with $T = 0.8T_F$(open) **(b)** GBP with $T = 0.88T_F$(open) **(c)** S100A6 with $T = 0.8T_F$(apo) **(d)** S100A6 with $T = 0.88T_F$(apo) . Representative trajectories are superimposed on the surfaces.

1.3.4 The transition rate coefficient vs the driving force; Tafel plot

How does the transition rate depend on the driving force of the conformation change, i.e., ΔV ? This driving force depends on the ligand concentration if binding/unbinding is sufficiently fast. Although the transition rate may be estimated from the barrier in free energy profile (like Fig. 1.5(d)), this estimate of the rate may depend on the specific choice of the reaction coordinate (χ in Fig. 1.5(d)) reflecting re-crossing effects. Using this estimate the dependence of the transition rate on the driving force can be determined. In the electron transfer processes, a quadratic dependence of the barrier on stability is predicted(35). This has been experimentally proven to be fairly accurate. For the conformational changes of proteins, Miyashita *et al.* also argued that a fully elastic model would also give such a curved dependence. On the other hand, they suggested that local unfolding, or cracking, will lead to a linear dependence over a large range of the driving force(19, 20). We now examine this notion using multi-basin model MD simulations.

The conformational transition rate coefficient k_{change} was estimated as the inverse of the first passage time as in (28). The transition is considered complete when χ first reaches the value at the minimum of the final basin. For very large driving forces, the conformational transition becomes barrier-less and limited only by diffusion, as expected. In this regime the rate becomes saturated. Here, we limit ourselves to thermodynamic conditions having significant barriers: The trajectories reside in the initial basin at least 100 MD steps (on average) before making the first transition.

For the four proteins studied, we calculated both the rates going from open to closed and in the other direction, over as large a range of the driving force without reaching the barrier-less regime. The resulting Tafel plots (Fig. 1.9) are surprisingly linear (with the exception of the binding reaction for S100A6, which is curved). The linear dependence is the consequence of the local unfolding as described in Miyashita *et al.* (19, 20) (See Fig. 1.10).

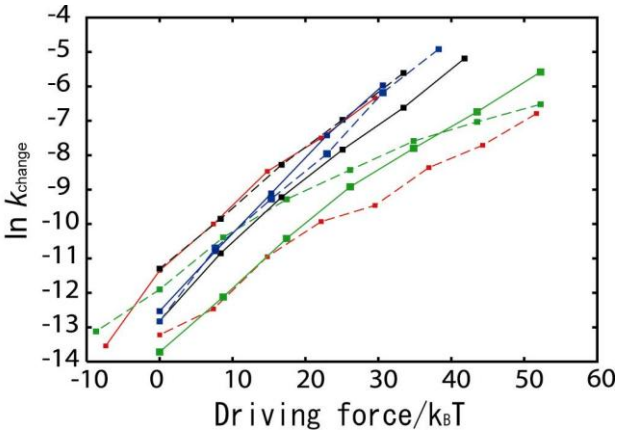

Fig. 1.9 The transition rate constant as a function of the driving force (Tafel plot) GBP is in red, S100A6 is in green, DHFR is in black, and HIV-1 protease is in blue. The dashed line is transition from unbound to bound, the solid line is from bound to unbound. The sign of the driving force is taken so that the increase corresponds to stabilization of the final state.

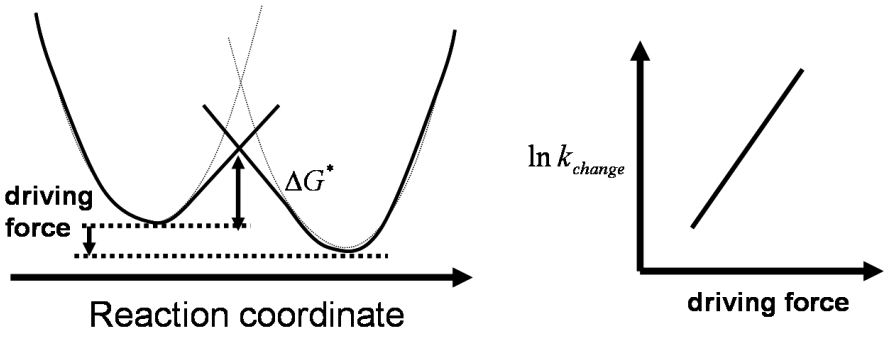

Fig. 1.10 Relation between the character of energy surface and Tafel plot Energy surfaces become linear around transition region by cracking. As a result of the energy surface, Tafel plot becomes linear.

1.4 Discussion

Building on the elastic picture of Miyashita and collaborators, the multiple-basin models somewhat similar to the present one have recently been proposed. Maragakis and Karplus put forward a plastic network model, in which individual basins are approximated by the Tirion harmonic model which are then smoothly connected by the secular equation formulation, as we did.(36) Being locally pure harmonic, local unfolding is not taken into account in this model. This plastic network model should work best for small-amplitude conformational changes. The model developed by Best *et al.* is close in spirit to the present multi-basin Hamiltonian because it also connects single-basin Gō potentials in a smooth function.(37) Their interpolation was achieved by the analogy to Boltzmann averaging, instead of the secular equation formula. Their model should give similar results to ours when there are two basins. When there are more than two basins, the present model allows the possibility to modulate the barrier heights of each basin-hopping motion individually and thus is somewhat flexible.

Directly modeling the multiple-basin energy landscapes is at an early stage. There is considerable room for creativity and improvement at this low resolution scale. An ingredient that is missing in the present model as well as in the others is precisely how to correctly account for the interaction with ligands. In the current model, the effects of ligand are implicit in the value of ΔV, which modulates the overall stability. But such a term does not account for the local nature of the interactions with the ligand. For example, S100A6 binds to calcium ions at the EF-hand loops and the interactions with calcium ions make the EF-hand loop more rigid. This locality may play an important role in allostery. Further work in this direction is done in the next chapter.

The present multiple-basin model can directly be used for MD simulation of very large biomolecular systems, such as molecular motors. Recently, the molecular mechanisms of the

rotary motor F_1-ATPase were studied using the switching Gō model.(38) In that work, the change in the nucleotide state was modeled as a "vertical excitation", resulting in switching between single basin models. The multiple-basin model proposed here provides a natural framework for realizing thermally activated conformational motions coupling ligand binding and release.

1.5 Conclusion

Proteins often undergo large-amplitude conformational changes by allostery, substrate binding or product release to take the appropriate conformations to fulfill their functions. We proposed the "multiple-basin model" for such large-amplitude conformational changes that involve partial unfolding and applied the model to the conformational changes of four proteins, glutamine binding protein (GBP), S100A6, dihydrofolate reductase (DHFR) and HIV-1 protease, to simulate the processes. As results, first, the different free energy surfaces that show the different mechanisms of conformational change were obtained. Conformational change consists of two processes that are breaking contacts of initial structure and making contacts of final structure. In the case of the conformational change of GBP, two processes were cooperative. In contrast, in the case of the conformational change of S100A6, two processes were sequential. Second, the free energy surfaces at different simulation temperature showed that how conformational change pathway is altered as a function of temperature. In the case of S100A6, at the high temperature, the globally extended intermediate emerged on the free energy surface. The pathway through this intermediate occurs, although usual direct pathway from initial to final conformation still occurs. Third, the transition-state barrier height dependence on relative stability of the two conformations was shown to be linear, which implies partial unfolding.

References

1. Tozzini V (2005) Coarse-grained models for proteins. *Curr Opin Struct Biol* 15:144-150.
2. Takada S, Luthy-Schulten, Z. & Wolynes, P. (1999) Folding dynamics with nonadditive forces: A simulation study of a designed helical protein and a random heteropolymer. *J. Chem. Phys.* 110:11616-11629.
3. Clementi C, Jennings PA, Onuchic JN (2000) How native-state topology affects the folding of dihydrofolate reductase and interleukin-1beta. *Proc Natl Acad Sci U S A* 97:5871-5876.
4. Clementi C, Nymeyer H, Onuchic JN (2000) Topological and energetic factors: what determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? An investigation for small globular proteins. *J Mol Biol* 298:937-953.
5. Honycutt JD, & Thirumalai, D. (1992) The Nature of Folded States of Globular Proteins. *Biopolymers* 32:695-709.
6. Okazaki K, Koga N, Takada S, Onuchic JN, Wolynes PG (2006) Multiple-basin energy landscapes for large-amplitude conformational motions of proteins: Structure-based molecular dynamics simulations. *Proc Natl Acad Sci U S A* 103:11844-11849.
7. Henzler-Wildman KA*, et al.* (2007) A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. *Nature* 450:913-916.
8. Volkman BF, Lipson D, Wemmer DE, Kern D (2001) Two-state allosteric behavior in a single-domain signaling protein. *Science* 291:2429-2433.
9. Lange OF*, et al.* (2008) Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution. *Science* 320:1471-1475.
10. Goto Y, Kuwajima K, Tanizawa, K (2005) *タンパク質科学　構造・物性・機能 (Protein science: structure, property and function)* (Kagaku-dojin, Kyoto).
11. Henzler-Wildman KA*, et al.* (2007) Intrinsic motions along an enzymatic reaction trajectory. *Nature* 450:838-844.
12. Hanson JA*, et al.* (2007) Illuminating the mechanistic roles of enzyme conformational dynamics. *Proc Natl Acad Sci U S A* 104:18055-18060.
13. Ikeguchi M, Ueno J, Sato M, Kidera A (2005) Protein structural change upon ligand binding: linear response theory. *Phys Rev Lett* 94:078102.
14. Case D (1994) Normal mode analysis of protein dynamics. *Curr Opin Struct Biol*

4:285-290.

15. Go N, Noguti T, Nishikawa T (1983) Dynamics of a small globular protein in terms of low-frequency vibrational modes. *Proc Natl Acad Sci U S A* 80:3696-3700.
16. Tirion MM (1996) Large Amplitude Elastic Motions in Proteins from a Single-Parameter, Atomic Analysis. *Physical Review Letters* 77:1905-1908.
17. Tama F, Sanejouand YH (2001) Conformational change of proteins arising from normal mode calculations. *Protein Eng* 14:1-6.
18. Ansari A*, et al.* (1985) Protein states and proteinquakes. *Proc Natl Acad Sci U S A* 82:5000-5004.
19. Miyashita O, Onuchic JN, Wolynes PG (2003) Nonlinear elasticity, proteinquakes, and the energy landscapes of functional transitions in proteins. *Proc Natl Acad Sci U S A* 100:12570-12575.
20. Miyashita O, Wolynes, P.G. & Onichic J.N. (2005) Simple Energy Landscape Model for the Kinetics of Functional Transitions in Proteins. *J. Phys. Chem.* 109:1959-1969.
21. Frauenfelder H, Sligar SG, Wolynes PG (1991) The energy landscapes and motions of proteins. *Science* 254:1598-1603.
22. Onuchic JN, Luthey-Schulten Z, Wolynes PG (1997) Theory of protein folding: the energy landscape perspective. *Annu Rev Phys Chem* 48:545-600.
23. Onuchic JN, Wolynes PG (2004) Theory of protein folding. *Curr Opin Struct Biol* 14:70-75.
24. Clementi C (2008) Coarse-grained models of protein folding: toy models or predictive tools? *Curr Opin Struct Biol* 18:10-15.
25. Bryngelson JD, Wolynes PG (1987) Spin glasses and the statistical mechanics of protein folding. *Proc Natl Acad Sci U S A* 84:7524-7528.
26. Go N (1983) Theoretical studies of protein folding. *Annu Rev Biophys Bioeng* 12:183-210.
27. Baker D (2000) A surprising simplicity to protein folding. *Nature* 405:39-42.
28. Koga N, Takada S (2001) Roles of native topology and chain-length scaling in protein folding: a simulation study with a Go-like model. *J Mol Biol* 313:171-180.
29. Takano M, Higo, J., Nakamura, H.K. & Sasai, M. (2004) On the model granularity to simulate protein dynamics: A biological physics view on biomolecular computing. *Natural Computing* 3:377-393.
30. DeLano WL (2002) (DeLano Scientific, San Carlos, CA, USA.).
31. Gerstein M, Krebs W (1998) A database of macromolecular motions. *Nucleic Acids Res* 26:4280-4290.

32. Chu JW, Voth GA (2007) Coarse-grained free energy functions for studying protein conformational changes: a double-well network model. *Biophys J* 93:3860-3871.
33. Warshel A, Weiss RM (1981) Empirical valence bond calculations of enzyme catalysis. *Ann N Y Acad Sci* 367:370-382.
34. Chang Y, Miller, WH (1990) An empirical valence bond model for constructing global potential-energy surfaces for chemical reactions of polyatomic molecular systems. *J. Phys. Chem.* 94:5884-5888.
35. Marcus RA (1992) Electron transfer reactions in chemistry: theory and experiment. *Nobel Lecture*.
36. Maragakis P, Karplus M (2005) Large amplitude conformational change in proteins explored with a plastic network model: adenylate kinase. *J Mol Biol* 352:807-822.
37. Best RB, Chen YG, Hummer G (2005) Slow protein conformational dynamics from multiple experimental structures: the helix/sheet transition of arc repressor. *Structure* 13:1755-1763.
38. Koga N, Takada S (2006) Folding-based molecular simulations reveal mechanisms of the rotary motor F1-ATPase. *Proc Natl Acad Sci U S A* 103:5367-5372.

Chapter 2: Dynamic energy landscape view of coupled binding and protein conformational change: Induced-fit versus population-shift mechanisms

Abstract

Allostery, the coupling between ligand binding and protein conformational change, is the heart of biological network and it has often been explained by two representative models, the induced-fit and the population-shift models. Here, we clarified for what systems one model fits better than the other by performing molecular simulations of coupled binding and conformational change. Based on the dynamic energy landscape view, we developed an implicit ligand binding model combined with the double-basin Hamiltonian that describes conformational change. From model simulations performed for a broad range of parameters, we uncovered that each of the two models has its own range of applicability, stronger and longer-ranged interaction between ligand and protein favors the induced-fit model, and weaker and shorter-ranged interaction leads to the population-shift model. We further postulate that the protein binding to small ligand tends to proceed via the population-shift model, whereas the protein docking to macromolecules such as DNA tends to fit the induced-fit model.

2.1 Introduction

Biological network is constructed by series of molecular recognitions and responses, which are ultimately attributed to conformational change of biomolecules upon binding to their partners. Traditionally, the coupling between the binding and the conformational transition was explained by the 50-years-old induced-fit model of Koshland (1), in which proteins are in their apo conformations in the unbound state, and binding to the apo forms induces conformational transition to the holo conformations. This model is apparently supported by accumulated examples of X-ray structures of the same protein in apo (open) form without the ligand and in holo (closed) form with the ligand (2, 3). Recently, however, growing evidence primarily by NMR and computer simulations suggests that protein is dynamic, and that intrinsic dynamics of a protein in the unbound state involves transient motion towards closed and functional conformation (4-6). An emerging view thus is that proteins in the unbound state exist in many conformers that include open and closed ones and that, upon binding, the dominant population shifts from the open form to the closed one. This "population-shift" model (known also as conformational selection or pre-existing equilibrium model), originated from the Monod-Wyman-Changeux model of allostery (7), receives more popularity in recent years (8, 9).

In this chapter, we investigated what kinds of ligand-binding processes employ the induced-fit route and *vice versa*, by performing a series of molecular simulations of coupled binding and conformational transition with a minimalist model of proteins (10). Technically, the multiple-basin energy landscape is realized by the structure-based multiple-basin model Hamiltonian recently proposed (11-13), and the ligand-binding model is proposed here as the stochastic jump between unbound and bound surfaces. We used glutamine binding protein (GBP) as a model protein for illustration of the idea (Fig. 2.4(a)). The relative simplicity of the model made it possible to simulate binding and unbinding processes coupled with conformational change so many times for a variety of parameter sets. Instead of concentrating on one set of parameters that could be closest to the protein, here we emphasize the advantage of exploring broad range of parameters. By this way, we can get better insights on the allostery.

In the following sections, we fist review some experimental works which characterize the mechanism of ligand-binding: induced-fit or population-shift. Then we introduce the idea of dynamic energy landscape to illustrate our model and define induced-fit and population-shift pathways clearly.

2.1.1 Experimental works: induced-fit or population-shift

There is little experimental evidence that distinguishes between the induced-fit and

population-shift mechanisms of coupled binding and conformational change. Although there are many examples that show conformational change upon ligand binding, mainly from X-ray structures, the mechanism cannot be identified from only static structures.

James *et al.* used pre-steady-state kinetics analysis, and together with x-ray structures, they revealed ligand-binding mechanism of an antibody, SPE7 (14). Pre-steady-state binding kinetics were analyzed by monitoring changes in SPE7's intrinsic fluorescence upon rapid mixing with ligand. The fluorescence quenching showed a complex pattern that can only be described by three exponentials, each decaying at different rate. The rate of the fast phase showed linear dependency on ligand concentration. So, this phase corresponds to simple association of antibody with ligand. The rate of the intermediate phase decreased with ligand concentration. This pattern is consistent with equilibrium between two pre-existing antibody conformers, only one of which binds the ligand. This phase apparently shows population-shift mechanism. The slow phase is unimolecular and first-order in its kinetics and is kinetically separated from the previous phases.

Estabrook *et al.* used a fluorescence technique to observe the loop movement of M.HhaI, a sequence-specific DNA cytosine C^5 methyltransferase, upon DNA binding and proved that this enzyme takes the induced-fit pathway (15).

Nevo et al. used a single-molecule technique, dynamic force spectroscopy, to discriminate the two mechanisms by comparing the distributions of forces required to unbind the complexes in the presence and in the absence of the effector (16-18). The two complexes, RanGppNHp-imp β (GppNHp is a GTP analog) and RanGDP-imp β were investigated in the presence or absence of the effector RanBP1. As results, RanGppNHp-imp β showed bimodal force distributions both in the presence or absence of the effector, only changing their relative populations. This suggests the population-shift mechanism. On the other hand, RanGDP-imp β showed unimodal force distributions both in the presence and absence of the effector, changing means of distributions. This suggests the induced-fit mechanism. The important suggestion by this experiment is that, depending on ligand (or binding partner), the binding mechanism can change between the two alternative models (16).

2.1.2 Dynamic energy landscape

To address ligand-binding mechanism, we need to deal with ensemble aspects of protein dynamics, which may be best realized by global energy landscape perspective of proteins (19, 20): The protein has globally funnel-like energy landscape, and when the bottom of the funnel is magnified, multiple minima exist among which protein dynamically transits in functional motion. On top of this standard landscape view, the energy landscape shape needs to be altered upon binding or releasing its interacting partner. This led us to a view of dynamic energy

landscape (21-23). A protein in the unbound state resides primarily in an open form with the lowest energy, and a less-stable minimum may exist near a closed form (Fig. 2.1). Binding to the ligand changes the energy landscape so that the closed form becomes the lowest energy (Fig. 2.1). By modeling binding and release as jumps between two energy landscapes, we can naturally represent the induced-fit and the population-shift scenarios as specific routes on the two landscapes. We note that it tends to be thought that the dynamic energy landscape theory is directly linked to the population-shift model, but it is not true. In the induced-fit scenario, a protein sitting at the open conformer on the unbound surface (UO in Fig. 2.1) binds the ligand to jump onto the bound surface (BO), which is followed by the conformational change to the closed conformation (BC). Conversely, in the population-shift scenario, the protein pre-exists in the closed form without ligand (UC in Fig. 2.1) at some small probability and this fraction can bind the ligand directly to reach the BC. Majority of the molecules resides in the UO state before binding. Part of this fraction promptly makes the conformational transition to UC state so that the system achieves the equilibrium on the unbound surface. Thus, for majority, the protein transits from UO to UC, and to BC. The question is thus which is the on-pathway intermediate from UO to BC, BO (which suggests the induced-fit) or UC (suggesting the population-shift)?

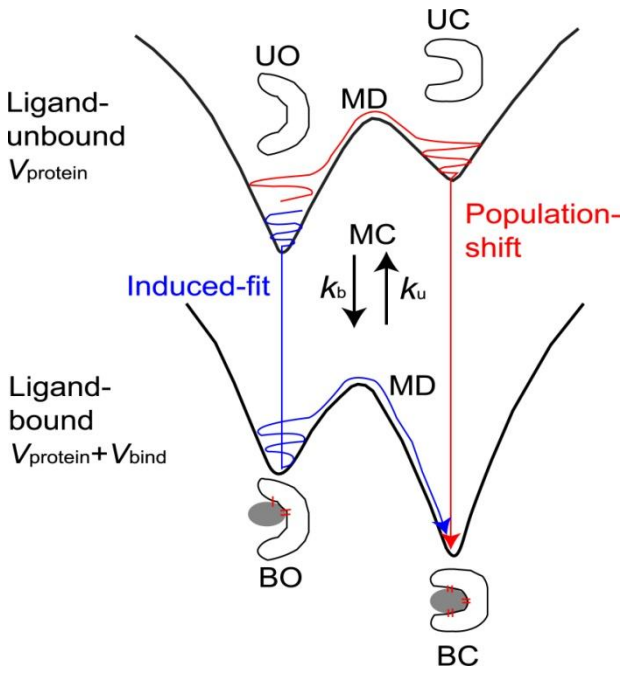

Fig. 2.1 A schematic view of the dynamic energy landscape

Protein can reside on one of two energy landscapes, one landscape for ligand-unbound (U) and the other for ligand-bound (B) states. Each landscape has (at least) two minima, one for open (O) and the other for closed (C) conformations. Thus, there are four states, UO, UC, BO, and BC. Protein can jump between two energy landscapes by the ligand binding/unbinding.

2.2 Materials and Methods

2.2.1 Model protein

We chose glutamine binding protein (GBP) as a model protein throughout the paper. Upon binding to a glutamine its conformation changes from open to closed form, which is an archetypical hinge-bending motion (Fig. 2.4(a)). We used two X-ray structures, ligand-free open form (PDB code: 1GGG) and ligand-bound closed form (PDB code: 1WDN). In simulations, we used 5-224 residues, discarding those missing in the PDB structures. We note that we used this protein for illustrative purpose: Parameters in the model do not necessarily faithfully correspond to those in the real glutamine binding protein.

2.2.2 Modeling ligand binding and protein conformational transition

We described that a protein has two distinct ligand-binding states; bound (B) and unbound (U) states. In the unbound state, the protein has just its intra-energy, $V_{protein}$, whereas the protein in the bound state has the intra-energy plus the ligand binding energy, $V_{protein} + V_{bind}$. The protein intra-energy $V_{protein}$ is expressed by all of Cα-atoms, and is set up so that it has the global minimum at the open (O) conformation and a less-stable minimum at the closed (C) conformation by multiple-basin model (See chapter 1 for explicit expressions). The ligand binding energy V_{bind} does not contain the explicit coordinates of the ligand atoms, but is a function of the Cartesian coordinates of ligand-binding sites of the protein (See section 2.2.3 for details). V_{bind} takes negative and large absolute value when the local environment around the binding pocket is close to that of the closed (C) conformation, and $V_{protein} + V_{bind}$ has its global minimum at the closed conformation. We note that in the reference X-ray structures the open (O) conformation corresponds to the ligand-unbound (U) state and the closed (C) corresponds to the bound (B) state, but this is not explicitly imposed in the simulation so that the protein experiences BO and UC states at some, albeit small, probabilities.

In time propagation, the protein conformation is moved by the standard molecular dynamics (MD) simulation, whereas the ligand-binding state (B or U) is stochastically changed by the rates k_b (binding) and k_u (unbinding) implemented as the Metropolis Monte Calro (MC) scheme (Fig. 2.1). While in the unbound state, a ligand molecule reaches the binding pocket at every Δt_b time with probability $p = k_b \Delta t_b = k_D [L] \Delta t_b$. With this probability, the state changes to the bound (B) one. Here k_b is the apparent first-order rate for binding, k_D is the

diffusion-controlled second-order rate constant for binding, and $[L]$ is the ligand concentration. While in the bound state, at every Δt_u time, the bound ligand has chance to dissociate at a probability that is dependent on V_{bind}. The mixed MD-MC scheme proposed here is a convenient way of simulating protein conformational dynamics coupled with ligand binding (See section 2.2.4 for details).

2.2.3 Ligand-binding model

The ligand binding energy V_{bind} does not contain the explicit coordinates of the ligand atoms, but is a function of the Cartesian coordinates of protein amino acids that are directly involved in ligand-binding, which is somewhat similar to (24). We first identified the residues involved to ligand binding by using LIGPLOT (25) (Fig. 2.2). Among these residues, we then defined "ligand-mediated contact pairs" as pairs of which heavy atom distance is below 10Å and that is not included in the native contact pair. For each of these ligand-mediated pairs, we adopted a Gaussian function so that we can separately control the minimum and the width of the potential,

$$V_{bind} = \sum_{\substack{ligand-mediated \\ contact \quad pairs}} -c_{lig}\,\varepsilon_1 \exp\left[-\frac{\left(r_{ij}/r_{0ij}-1.0\right)^2}{2\left(\sigma/r_{0ij}\right)^2}\right]$$

where ε_1 is the depth for the native contact, c_{lig} is a parameter that changes strength of ligand-mediated contact. r_{0ij} is the distance between *i*- and *j*-th residues at the closed form and thus V_{bind} takes the minimum value when the local environment around the binding pocket is close to that of the closed form. σ / r_{0ij} defines the interaction range. We changed $\sigma/\mathrm{r_{0ij}}$ from 0.01 to 0.5, which covers wide range of interactions including hydrogen bonding that can be expressed by Lennard-Jones (LJ) 12-10 function (26, 27), van der Waals interaction that is expressed by LJ 12-6 function, and coulomb interaction (Fig. 2.3).

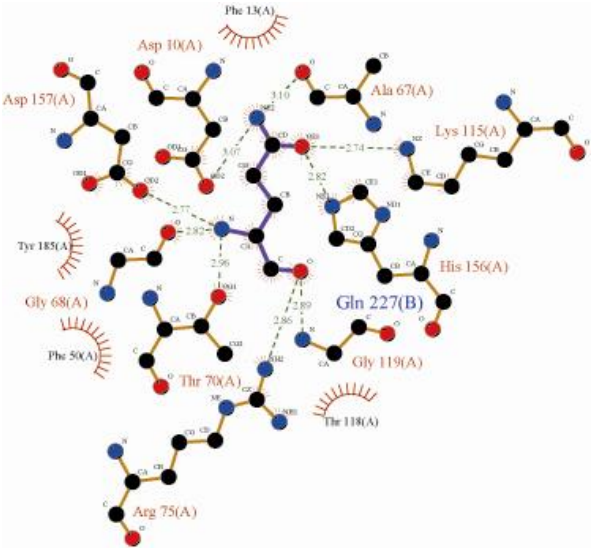

Fig. 2.2 Ligand binding site of GBP identified by LIGPLOT(25) Purple bonds belong to the ligand, orange bonds belong to the hydrogen bonded residues from protein, and dashed lines represent hydrogen bonds between ligand and GBP. Hydrophobic contacts made with GBP are indicated by spoked arcs pointing toward ligand.

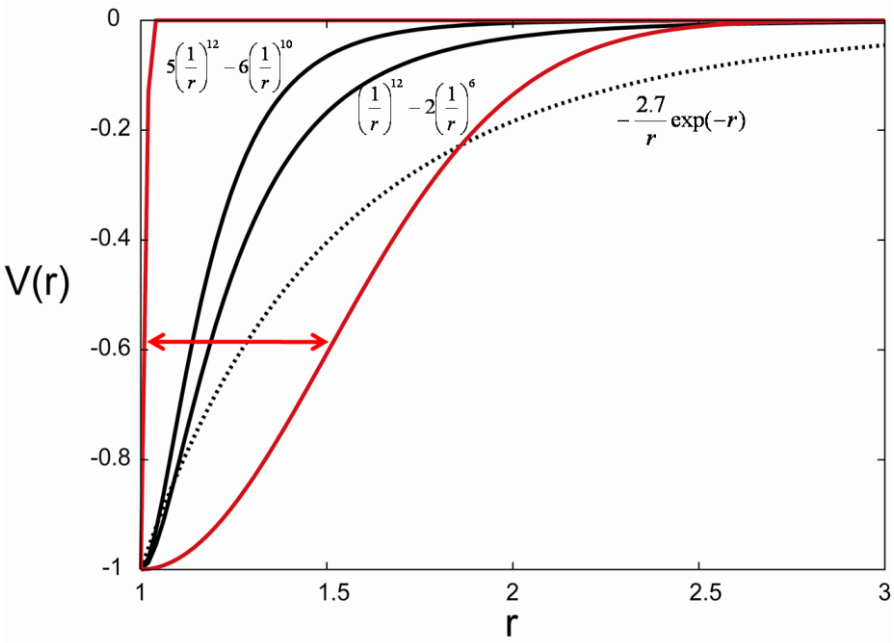

Fig. 2.3 Interaction range of ligand binding energy function Different types of energy function are plotted. Black solid lines represent two Lennard-Jones (LJ) interactions, black dotted line represents coulomb interaction with Debye-Huckel type screening. Red lines represent Gaussian function used here with different parameters; left one is the shortest range $\sigma/r_{0ij} = 0.01$ and right one is the longest range $\sigma/r_{0ij} = 0.5$.

2.2.4 The MD-MC coupling simulations

The protein structure was propagated by the standard MD method, while the ligand-binding state was updated via a MC method. MD simulation was carried out using the constant temperature Newtonian dynamics, where the mass of all residues were set to identical. The velocity Verlet algorithm was used for time propagation with a simple Berendsen thermostat (28) (See Appendix 1). We estimated the folding temperatures of single-basin Go models for the open and the closed forms using the Weighted Histogram Analysis Method (29), and used 0.8 times the lower of two folding temperatures as the simulation temperature.

The MC transition between ligand bound and unbound states was characterized by binding and unbinding rate constants k_b and k_u. At a given $[L]$, the apparent ligand-binding rate is given by $k_b = k_D[L]$. The binding is assumed to be diffusion-controlled, and so the second-order rate constant k_D is given as $k_D = 4DS(R)$(30). D is the diffusion constant for a glutamine. Based on comparison with all-atom simulation, we estimated a MD step of the current model as $100\,fs$. This, together with an experimental estimate of the diffusion constant, we set $D = 1.0 \times 10^{-2}$ [Å^2/MD step]. $S(R)$ is the length-scale of the binding site, for which we adopted the distance between Phe50 and Lys115 that locate edge of the binding site. This depends on protein conformation; $S(R) \approx 22.7$ Å in the open form, and $S(R) \approx 8.6$ Å in the closed form. So, the ligand-access is easier in the open form. The ligand concentration $[L]$ is set as 0.1M unless otherwise stated. The ligand unbinding rate is given by $k_u = \Delta t_u^{-1} \exp(-|V_{bind}|/k_B T)$ where Δt_u corresponds to a period of fluctuations of a residue in the binding site, and is now set to 100 MD time steps.

2.2.5 Definition of the four states and transition rate analysis

First, we defined the four states. Ligand-binding state is either unbound (U) or bound (B). The conformation is classified as open (O) if $\chi \le 0.0$ and as closed (C) if $\chi > 0.0$. So, we have the four states; UO, UC, BO, and BC.

For the transition rates, under the condition that the U and B are equally probable, we performed long-time simulations. In the trajectory, we observed reversible and multiple transitions among the four states, and counted the number of transitions between neighboring states n_{ij}. Here, transitions in both directions should be equal, in principle, and so we take an average of numbers in two-way transitions. To avoid over-counting the re-crossing, we defined the completion of one transition when χ changed its sign *and* reached its absolute value larger than unity. Under the Markovian approximation, using $k_{i\to j}\tau_i = k_{j\to i}\tau_j = n_{ij}$, we obtained the

transition rates between states, where $k_{i\to j}$ ($k_{j\to i}$) is the rate constant from state i (j) to state j (i), τ_i (τ_j) is the residential time of state i (j) along the trajectory.

2.2.6 Steady-state analysis

Within the four-state representation, we performed a simple steady-state analysis. For example, when we calculate the rate of induced-fit pathway, that is, from UO via BO to BC, the reaction scheme is UO⇄BO→BC. For the steady-state, we assumed that the molecular source is at UO state and the molecular sink is at BC. Thus, the reverse transition from BC to BO is zero. We further assume that the concentration of the intermediate BO is in steady-state, $\frac{d[BO]}{dt} = k_1[UO] - (k_{-1} + k_2)[BO] = 0$ where k_1 is apparent first-order rate from UO to BO at the fixed ligand concentration, k_{-1} is the dissociation rate from BO to UO and k_2 is the rate from BO to BC. From this, we directly get the rate of induced-fit k_{if}, $\frac{d[BC]}{dt} = \frac{k_1 k_2}{k_{-1} + k_2}[UO] = k_{if}[UO]$. The same sort of analysis about the population-shift pathway leads to the rate of population-shift k_{ps}.

2.3 Results

2.3.1 Binding and conformational transitions are stochastically coupled

Using the glutamine binding protein (GBP) as a model system (Fig. 2.4(a)), we simulated binding-coupled conformational dynamics at various parameters. Fig. 2.4(b) is a representative time course when about half of the time the protein was in the bound state. In the top panel, the reaction coordinate χ that monitors the conformation is negative (positive) at the open (closed) conformations. The conformation changed reversibly and quite abruptly (Each transition completed in ~1000 MD steps) many times in the trajectory. The ligand-binding energy V_{bind} of the same trajectory in the bottom panel of Fig. 2.4(b) also showed sudden transitions, but their timings were not identical to those of conformational transitions (More clearly, see Fig. 2.4(c), which is a close-up of a particular time-window). V_{bind} took roughly three energy levels; 0 (unbound), $\sim -3k_BT$ (weak-binding energy) and $\sim -8k_BT$ (strong-binding energy). Comparing to the time course of χ, we see that the closed conformation tended to have strong-binding energy, while the open-form protein took weak-binding energy or unbound state.

Notably, conformational transitions and bindings/releases did not occur simultaneously. In the trajectory of Fig. 2.4(c), we see that the protein tended to change its conformation from open to closed one slightly before ligand-binding suggesting the population-shift mechanism. Conversely, when the protein transited from closed to open conformations, the ligand-unbinding was followed by the conformational transition. It is interesting to note that the conformational change to the closed form initiated the binding, i.e., the population-shift scenario, while the unbinding initiated the conformational change to the open form: Thus the sequences of events are reversed. Below, we investigated these coupling mechanisms in more detail depending on the nature of binding interactions.

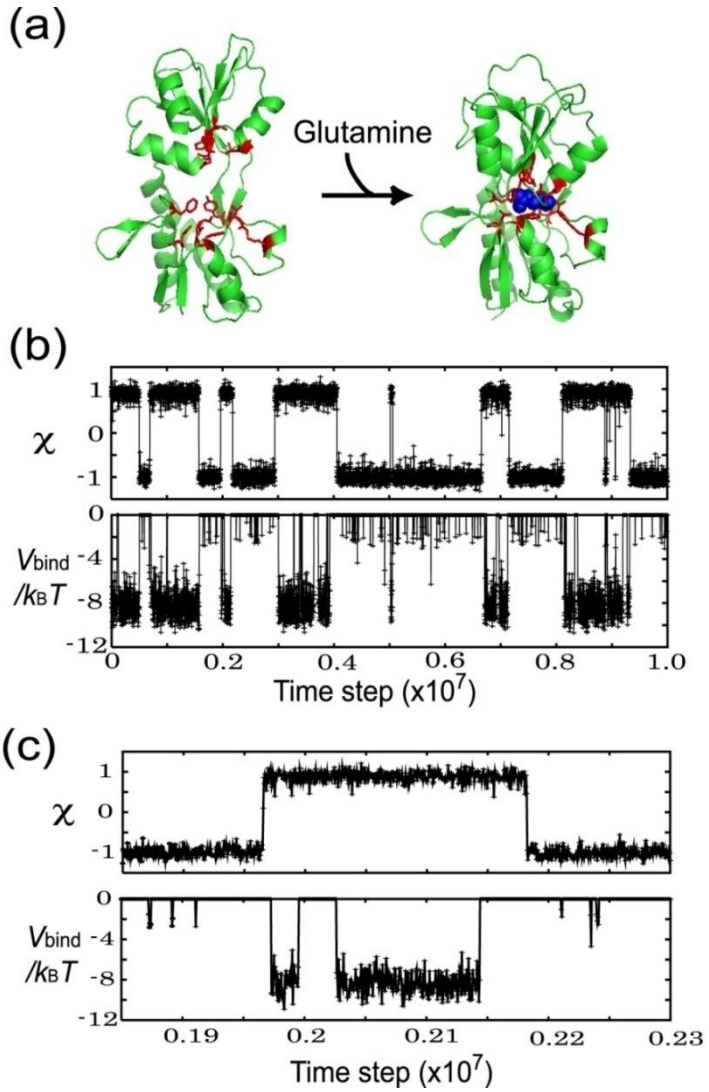

Fig. 2.4 A representative trajectory of coupled conformational change and ligand binding **(a)** The open and closed structures of GBP and its ligand, glutamine (blue) are shown. The ligand-binding residues are represented in red sticks. **(b, upper)** A representative time course of the conformational change coordinate χ. Its negative (positive) value corresponds to the open (closed) conformations. **(b, lower)** A time course of ligand binding energy, V_{bind}, for the same trajectory as upper. **(c)** A small time window in the trajectory of (b) is magnified.

2.3.2 Free energy surface of conformational change and ligand binding

To see the coupling between protein conformational change (monitored by χ) and ligand binding (monitored by V_{bind}) more directly, we computed the free-energy surface on the $\chi - V_{bind}$ plane (Fig. 2.5(a), (b) where typical trajectories are superimposed). Four states,

unbound open (UO), unbound closed (UC), bound open (BO), and bound closed (BC) states, are seen on the free energy surfaces. We note that the ligand binding state here is discrete, i.e., either bound or unbound, and thus the free energy surfaces and trajectories are discontinuous between $V_{bind} = 0$ and $V_{bind} < 0$. Results for two different ranges of interactions are shown in Fig. 2.5(a) and (b); the former (latter) corresponds to shorter (longer) range interactions. We note that the strength coefficient c_{lig} of interaction was tuned for each case so that about half of the time the protein was in the bound state.

With a shorter-range interaction ($\sigma = 0.05 r_{0ij}$), the dominant pathway from UO to BC passed through UC, suggesting that it corresponded to the population-shift pathway: Conformational change was followed by the ligand binding (Fig. 2.5(a)). Contrary, with a longer-range interaction ($\sigma = 0.15 r_{0ij}$) (Fig. 2.5(b)), in addition to the population-shift pathway, we saw the alternative induced-fit pathway, where the binding was followed by the conformational change (Fig. 2.5(b)). The range of ligand binding interaction directly altered the coupling between the binding and the conformational change.

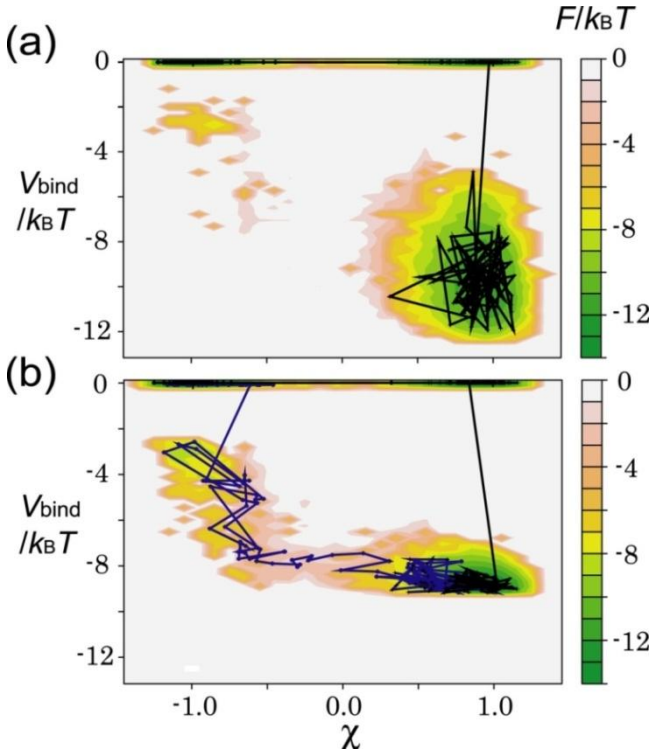

Fig. 2.5 Two-dimensional free-energy surfaces as a function of protein conformational change (χ) and the ligand-binding energy (V_{bind}) (a) The case of the short-ranged interaction $\sigma/r_{0ij} = 0.05$. The black line is a representative trajectory. **(b)** The case of the long-ranged interaction $\sigma/r_{0ij} = 0.15$. The black and blue lines are two representative trajectories.

2.3.3 Induced-fit versus population-shift: Four states analysis

For more quantitative argument, we conducted the first-passage analysis. Starting from the UO state, we performed simulations until the protein reached the BC state for the first time. Once the protein reached the BC state, we observed from which it reached, either BO or UC. We simulated 50 runs for each parameter set.

First, we looked into the effect of the interaction strength c_{lig} (Fig. 2.6(a)) onto the pathway. We found that, as the ligand interaction became stronger, the ratio of the induced-fit pathway increased. This suggests that stronger interaction between the ligand and the protein tends to favor the induced-fit mechanism.

Next, we investigated the effect of the range of interactions σ. Because, as in the previous paragraph, we know that the shift in equilibrium between unbound and bound drastically changes the pathway, for each value of σ, we tuned the parameter c_{lig} so that the time spent in the bound state is 50%. The result showed clear dependence of the ratio of the population-shift pathway on the range of ligand interaction (Fig. 2.6(b)): The shorter-range ligand interaction favors the population-shift pathway and *vice versa*.

In summary, as ligand-binding interaction becomes stronger, the dominant pathway shifts from population–shift to induced-fit. When the ligand affinity is the same, shorter-range ligand interaction favors the population-shift pathway, and *vice versa*.

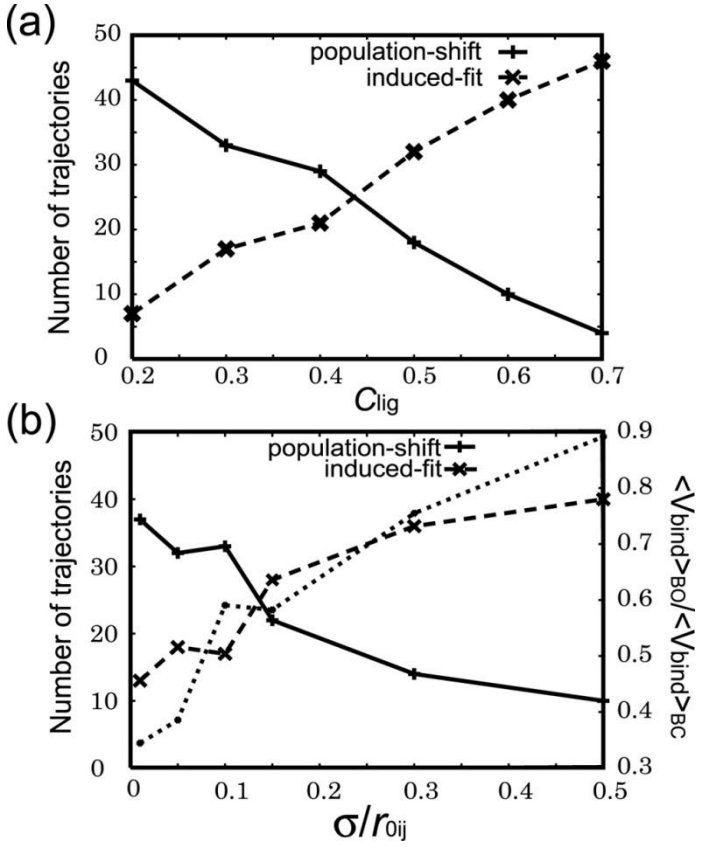

Fig. 2.6 The first-passage analysis from UO to BC Of 50 trajectories, the number of trajectories that went through BO (dashed line, denoted as induced-fit) and UC (solid line, denoted as population-shift) is plotted. **(a)** The ligand-binding strength (c_{lig}) is altered. **(b)** The interaction range σ is altered, whereas the strength of ligand interaction (c_{lig}) is tuned, in rach point, so that the ligand is bound for half of the simulation time in equilibrium condition. The ratio of the average ligand-binding energy in the BO state relative to that in the BC state, $\langle V_{bind} \rangle_{BO} / \langle V_{bind} \rangle_{BC}$, is also plotted with the dotted line, for which the scale on the right applies.

2.3.4 Transition rates and equilibrium constants in four states representation

For further dissecting the transitions among four states, we calculated every transition rates between two-adjacent states in the four states representation, which also gave us the equilibrium constants (See Materials and Methods for details). We note that the binding rate constant given here is an apparent first-order rate for a given concentration of the ligand. The resulting rates are shown in Fig. 2.7(a) (in unit of 10^{-5}/MD step) for a short-ranged interaction $\sigma = 0.05 r_{0ij}$

and in Fig. 2.7(b) for a long-ranged one $\sigma = 0.3r_{0ij}$. In either case, the strength coefficient c_{lig} was tuned up so that ligand was bound for half of the time.

In the case of short-ranged interaction (Fig. 2.7(a)), the rate constant from BO to UO is markedly large. The short-ranged interaction is sensitive to the conformation of the binding site. Since the binding is assumed to be optimized at the closed conformation, the binding energy at the open form (BO) is small in absolute value, which led to the low equilibrium population in BO. A simple steady-state analysis from UO to BC resulted in that the population-shift pathway is twice as probable as the induced-fit pathway, that is, $k_{if} = 0.054$ and $k_{ps} = 0.11$ (See Materials and Methods for details).

Contrary, for the long-ranged interaction case (Fig. 2.7(b)), the same steady-state analysis gave that the induced-fit pathway is three times as probable as the population-shift ($k_{if} = 0.37$ and $k_{ps} = 0.12$). The longer-range interaction is less sensitive to the conformation of the binding site, so that the BO state has more stability. The transition from UO to BO is fast. Thus, the population of BO is relatively large and is possible to be promptly refilled, which makes the induced-fit pathway favored.

Since the apparent first-order rate constants of the ligand-binding are proportional to the ligand concentration, we can put the ligand-concentration dependence back in the diagrams Fig. 2.7(a) and (b). From these, we calculated the equilibrium titration curves of four states, as depicted in Fig. 2.7(c) and (d). $[L]$ is scaled as unit of the ligand concentration used in simulations. For both short- and long-ranged interaction cases, as expected, the concentration of BC increased, and that of UO decreased with the ligand concentration. When we focus on the regime where the concentrations of UO and BC are comparable ($[L] \sim 1$), the short-ranged case (Fig. 2.7(c)) results in UC more favored than BO, whereas the long-ranged interaction (Fig. 2.7(d)) gives comparable concentrations of BO and UC.

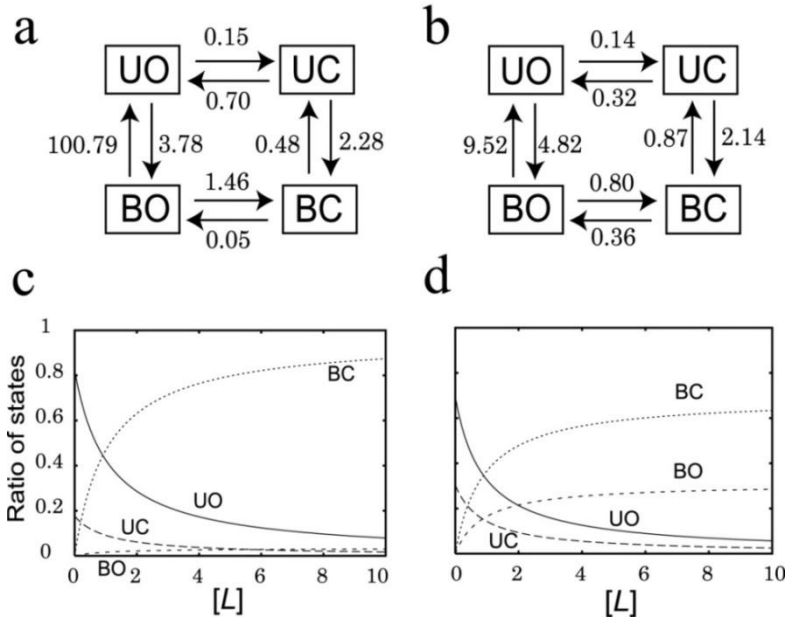

Fig. 2.7 The rate constants between the four states and the equilibrium titration curves The binding rate constants here are apparent first-order rates at a given ligand concentration, and unit is 10^{-5} per MD step. **(a)** The case of the short-ranged ligand interaction ($\sigma = 0.05r_{0ij}$). **(b)** The case of the long-ranged ligand interaction ($\sigma = 0.3r_{0ij}$). **(c), (d)** The equilibrium titration curves of the four states against the ligand concentration. (c) was derived from the equilibrium constants in (a), and (d) was derived from the equilibrium constants in (b).

2.4 Discussion

We elucidated that strong and long-ranged ligand interaction favors the induced-fit mechanism, whereas weak and short-ranged ligand interaction leads to the population-shift mechanism in simulation study. Now, we look into some examples for which the coupling mechanisms were relatively well established experimentally. Binding of an antigen, hapten DNP-Ser, to an antibody, SPE7, was characterized to take the population-shift pathway (14). SPE7 in its unbound state is in equilibrium between two pre-existing conformers (Ab^1 and Ab^2) and one of which (Ab^2) corresponds to the ligand-bound conformation. Looking into the binding site characteristics, we noticed that the binding site of SPE7 is hydrophobic surrounded by many bulky aromatic rings. Its partner, hapten, is also hydrophobic having an aromatic ring (Fig. 2.8(a)). Thus their interaction is primarily hydrophobic interaction as well as specific hydrogen bonding, both of which are short-ranged. This is consistent with our finding.

The induced-fit mechanism has been suggested for various DNA-protein binding. Here we picked up M.HhaI, DNA cytosine C^5 metyltransferase, which binds to DNA substrates (15). A flexible loop within M.HhaI (residues 80-100) recognizes the cognate DNA. It has been shown

that the loop is reorganized to the closed conformer when the enzyme binds to cognate DNA, but not closed when it binds to nonspecific DNA, suggesting that the recognition process of the loop is induced-fit. Not surprisingly for the DNA-binding protein, the binding loop of the protein is positively charged, and DNA is negatively charged (Fig. 2.8(b)). Thus their interaction is mainly electrostatic and thus is long-ranged. On top, since DNA is a relatively large ligand, the interaction energy between M.HhaI and DNA is likely to be very large. All these are consistent with our finding.

Although the allostery is inherently the dynamic problem, its kinetics is constrained by the underlying energetics. Energetically, putting aside kinetics, the induced-fit pathways would be favored when both of the following conditions hold: 1) With a high concentration of ligand in solution, the BO state is significantly more stable than the UO, and 2) the BC is more stable than the BO. We assume that the ligand binding is strongest at the closed conformation and that fraction of the ligand-protein interaction in the BC state is formed in the BO state. We directly see that the stronger ligand binding energy tends to satisfy the above two conditions, thus guiding to the induced-fit pathways. Long-ranged interaction between protein and ligand tends to satisfy the first condition above because long-ranged interaction makes the interaction energy less sensitive to the conformation of protein and thus the ligand-protein interaction at the open conformer is increased i.e., the first condition above, which thus support the induced-fit mechanism. We tested this interpretation by computing $\langle V_{bind} \rangle_{BO} / \langle V_{bind} \rangle_{BC}$, and found that this ratio correlates very well with the fraction of the induced-fit pathway (The dotted line in Fig. 2.6(b)). If the interaction range is too large, however, the interaction energy becomes nearly the same between the open and closed conformers, which is opposed to the second condition above. In the limit, the conformation would not change to the closed form. In this context, probably, having a large interface between protein and ligand would be beneficial to have the induced-fit pathway partly because the large interface implies strong interaction and partly because the large interface makes it easy to have partial, yet significant, interaction in the BO state and the rest of interaction achieved in the BC state. Therefore, we postulate that protein interaction with small ligand tends to favor the population-shift pathway as in the cases of antibody-antigen and enzyme-substrate, whereas protein docking to DNA, RNA, or protein with large interface favors the induced-fit.

While we shed light on relatively simple cases, real molecular systems must contain more complexity. First, conformational change does not proceed uniformly, but may be hierarchic. For example, a sub-structure near the binding site may move differently from the rest of the protein. In this two-mode case, one mode may correspond to the population-shift model, whereas the other mode may fit the induced-fit (31). A specific role of a loop near an active site

was characterized for horse liver alcohol dehydrogenase (32). Another very plausible scenario is the "population-shift followed by induced-fit" pathway. Unbound protein has a transient local minimum, which is not identical but is directed towards the closed conformation. Ligand-binding shifts the population to this partially closed conformation, which further induces conformational change up to the closed conformation. Substrate-binding and product-release in enzymes turns on another type of complexity because the substrate and the product are not the same, opening a possibility that binding and release have different pathways.

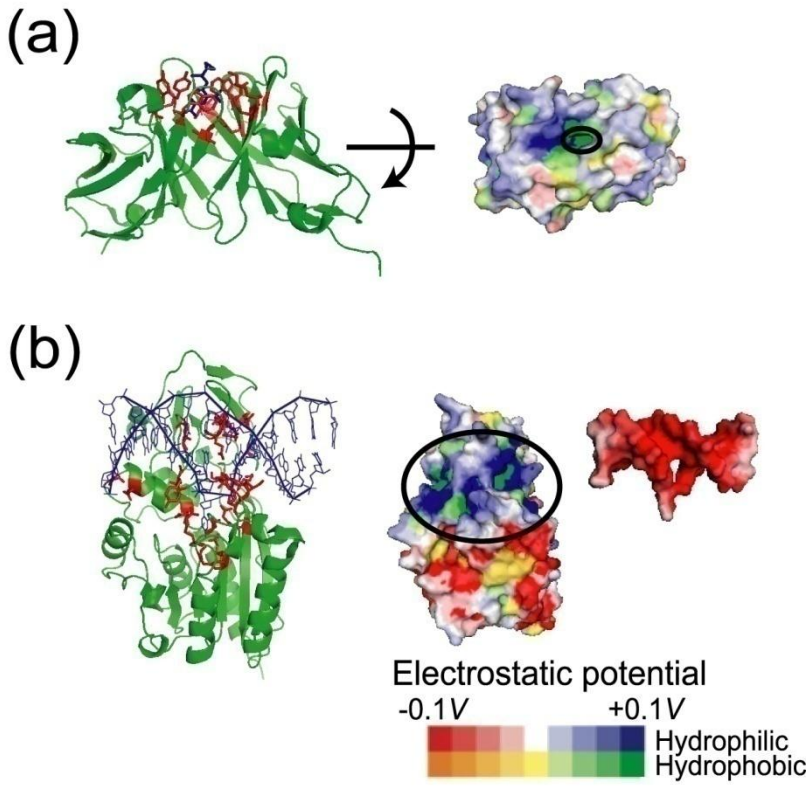

Fig. 2.8 Molecular examples that were suggested to show the population-shift and induced-fit mechanisms (a) An antibody, SPE7, and its ligand hapten DNP-ser are suggested to show the population-shift mechanism. (*left*) SPE7 dimer is shown in cartoon with its binding sites in red sticks. Hapten DNP-Ser is shown in blue sticks. (*right*) Hydrophobicity/hydrophilicity and the electrostatic potential of SPE7 are shown by color. The ligand-binding site is highly hydrophobic. **(b)** M. Hhal (DNA cytosine C^5 methyltransferase) binds to DNA via the induced-fit mechanism. (*left*) M.Hhal is shown in cartoon with its binding sites in red-stick representation, and DNA is shown in blue. (*right*) Hydrophobicity/hydrophilicity and the electrostatic potential of both molecules are shown. The binding sites of M. Hhal (DNA) have high (low) electrostatic potential, indicating that their binding is via electrostatic interactions. The electrostatic potentials were drawn by eF-site and PDBjViewer (33).

2.5 Conclusions

Using the multiple-basin model of protein conformational change and a newly developed implicit ligand binding model, we investigated the interplay between the induced-fit and the population-shift mechanisms of coupled binding and conformational transition process. We found that strong and long-ranged interaction favors the induced-fit pathway, whereas weak and short-ranged interaction leads to the population-shift pathway. These are consistent with many available experimental data.

The coarse-grained model of coupled binding and conformational change developed here is quite general and can be applied to simulate broad-range of large-amplitude motion in biomolecular complex, such as ATPase and other macromolecular machines (34-36).

References

1. Koshland D (1958) Application of a theory of enzyme specificity to protein sythesis. *Proc Natl Acad Sci USA* 44:98-104.
2. Gerstein M, Krebs W (1998) A database of macromolecular motions. *Nucleic Acids Res* 26:4280-4290.
3. Qi G, Lee R, Hayward S (2005) A comprehensive and non-redundant database of protein domain movements. *Bioinformatics* 21:2832-2838.
4. Volkman BF, Lipson D, Wemmer DE, Kern D (2001) Two-state allosteric behavior in a single-domain signaling protein. *Science* 291:2429-2433.
5. Henzler-Wildman KA*, et al.* (2007) Intrinsic motions along an enzymatic reaction trajectory. *Nature* 450:838-844.
6. Bahar I, Chennubhotla C, Tobi D (2007) Intrinsic dynamics of enzymes in the unbound state and relation to allosteric regulation. *Curr Opin Struct Biol* 17:633-640.
7. Monod J, Wyman J, Changeux JP (1965) On the Nature of Allosteric Transitions: a Plausible Model. *J Mol Biol* 12:88-118.
8. Bosshard HR (2001) Molecular recognition by induced fit: how fit is the concept? *News Physiol Sci* 16:171-173.
9. James LC, Tawfik DS (2003) Conformational diversity and protein evolution--a 60-year-old hypothesis revisited. *Trends Biochem Sci* 28:361-368.
10. Okazaki K, Takada S (2008) Dynamic energy landscape view of coupled binding and protein conformational change: induced-fit versus population-shift mechanisms. *Proc Natl Acad Sci U S A* 105:11182-11187.
11. Maragakis P, Karplus M (2005) Large amplitude conformational change in proteins explored with a plastic network model: adenylate kinase. *J Mol Biol* 352:807-822.
12. Best RB, Chen YG, Hummer G (2005) Slow protein conformational dynamics from multiple experimental structures: the helix/sheet transition of arc repressor. *Structure* 13:1755-1763.
13. Okazaki K, Koga N, Takada S, Onuchic JN, Wolynes PG (2006) Multiple-basin energy landscapes for large-amplitude conformational motions of proteins: Structure-based molecular dynamics simulations. *Proc Natl Acad Sci USA* 103:11844-11849.
14. James LC, Roversi P, Tawfik DS (2003) Antibody multispecificity mediated by conformational diversity. *Science* 299:1362-1367.

15. Estabrook RA, Reich N (2006) Observing an induced-fit mechanism during sequence-specific DNA methylation. *J Biol Chem* 281:37205-37214.
16. Nevo R, Brumfeld V, Elbaum M, Hinterdorfer P, Reich Z (2004) Direct discrimination between models of protein activation by single-molecule force measurements. *Biophys J* 87:2630-2634.
17. Nevo R, Brumfeld V, Kapon R, Hinterdorfer P, Reich Z (2005) Direct measurement of protein energy landscape roughness. *EMBO Rep* 6:482-486.
18. Nevo R*, et al.* (2003) A molecular switch between alternative conformational states in the complex of Ran and importin beta1. *Nat Struct Biol* 10:553-557.
19. Onuchic JN, Luthey-Schulten Z, Wolynes PG (1997) Theory of protein folding: the energy landscape perspective. *Annu Rev Phys Chem* 48:545-600.
20. Frauenfelder H, Sligar SG, Wolynes PG (1991) The energy landscapes and motions of proteins. *Science* 254:1598-1603.
21. Tsai CJ, Ma B, Nussinov R (1999) Folding and binding cascades: shifts in energy landscapes. *Proc Natl Acad Sci USA* 96:9970-9972.
22. Kumar S, Ma B, Tsai CJ, Sinha N, Nussinov R (2000) Folding and binding cascades: dynamic landscapes and population shifts. *Protein Sci* 9:10-19.
23. Gulukota K, Wolynes PG (1994) Statistical mechanics of kinetic proofreading in protein folding in vivo. *Proc Natl Acad Sci USA* 91:9292-9296.
24. Whitford PC, Miyashita O, Levy Y, Onuchic JN (2007) Conformational transitions of adenylate kinase: switching by cracking. *J Mol Biol* 366:1661-1671.
25. Wallace AC, Laskowski RA, Thornton JM (1995) LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng* 8:127-134.
26. Brooks B, Bruccoleri, RE, Olafson, BD, States, DJ, Swaminathan, S, Karplus, M (1983) CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comput. Chem.* 4:187.
27. Takada S, Luthey-Schulten, Z, Wolynes, PG (1999) Folding dynamics with nonadditive forces: A simulation study of a designed helical protein and a random heteropolymer. *J. Chem. Phys.* 110:11616.
28. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR (1984) Molecular dynamics with coupling to an external bath. *J Chem Phys* 81:3684-3690.
29. Kumar S, Bouzida D, Swendsen RH, Kollman PA, Rosenberg JM (1992) The Weighted Histogram Analysis Method for Free-Energy Caluculations on Biomolecules. I. The Method. *J Comput Chem* 13:1011.
30. Shoup D, Lipari G, Szabo A (1981) Diffusion-controlled bimolecular reaction rates. The effect of rotational diffusion and orientation constraints. *Biophys J* 36:697-714.

31. Formaneck MS, Ma L, Cui Q (2006) Reconciling the "old" and "new" views of protein allostery: a molecular simulation study of chemotaxis Y protein (CheY). *Proteins* 63:846-867.
32. Hayward S, Kitao A (2006) Molecular dynamics simulations of NAD+-induced domain closure in horse liver alcohol dehydrogenase. *Biophys J* 91:1823-1831.
33. Kinoshita K, Nakamura H (2004) eF-site and PDBjViewer: database and viewer for protein functional sites. *Bioinformatics* 20:1329-1330.
34. Koga N, Takada S (2006) Folding-based molecular simulations reveal mechanisms of the rotary motor F1-ATPase. *Proc Natl Acad Sci USA* 103:5367-5372.
35. Hyeon C, Onuchic JN (2007) Mechanical control of the directional stepping dynamics of the kinesin motor. *Proc Natl Acad Sci U S A* 104:17382-17387.
36. Hyeon C, Onuchic JN (2007) Internal strain regulates the nucleotide binding site of the kinesin leading head. *Proc Natl Acad Sci U S A* 104:2175-2180.

Chapter 3: Allostery in working mechanism of molecular motor; F_1-ATPase

Abstract

F_1-ATPase is water soluble portion of F_1F_o-adenosine triphosphate (ATP) synthase. F_1-ATPase by itself rotates γ-subunit with certain torque using free energy drop of ATP hydrolysis. How free energy of ATP hydrolysis is used to rotate γ-subunit mechanically? The mechano-chemical coupling has been a central theme of this motor protein. Some experiments suggest that γ-angle regulates three nucleotide states, which is sometimes referred as the γ-dictator mechanism. We examine the possibility that γ-angle regulates the ADP release as experimentally suggested. As results, we found that the more γ-angle rotates forward, the faster the ADP site conformational change, thus ADP-release, and interaction between γ-subunit and β-subunit of the ADP site was important for the conformational change. We further analyzed allosteric coupling between ATP-binding and ADP-release, which accompanies 80° rotation of γ, and found that, for the tight coupling between ATP-binding and ADP-release, the existence of γ-axle is necessary, supporting γ-dictator mechanism.

3.1 Introduction

F_1-ATPase (F_1) is water soluble portion of F_1F_o-adenosine triphosphate (ATP) synthase. Both F_1 and F_o (membrane-embedded portion) are rotary motors and have a common rotary shaft. F_1 being powered by ATP hydrolysis and F_o by proton flow, their rotary directions are opposite to each other. F_1 by itself rotates counterclockwise (viewed from membrane side) driven by ATP hydrolysis, but under physiological condition, F_1F_o-ATPase synthesizes ATP from adenosine diphosphate (ADP) and inorganic phosphate (P_i) driven by proton flow through F_o.

A minimal active subcomplex for F_1-ATPase is $\alpha_3\beta_3\gamma$-subunits. The central γ-subunit rotates inside a cylinder made of $\alpha_3\beta_3$-subunits. α- and β-subunits are arranged alternatively and three β-subunits mainly host catalytic sites (Fig. 3.1(a)). Single molecule experiments have visualized the counterclockwise rotations (1) and the discrete 120° stepwise rotations (2). Each step is driven by hydrolysis of one ATP molecule at high efficiency. Input energy is free energy drop of one ATP hydrolysis, $\approx 20k_BT$ under physiological condition (3), and output is to rotate γ-subunit with a certain torque. Thus, the central question has been how free energy of ATP hydrolysis is used to rotate γ-subunit mechanically.

This mechano-chemical coupling mechanism has been studied over years. Yasuda *et al.* found that the 120° step is further resolved into 80° and 40° substeps (4). The 80° substep is driven by ATP binding, and the 40° substep by product release, ADP or P_i. The dwell time before a 40° substep consists of two $\approx 1ms$ reactions. One of the reactions is ATP hydrolysis in a site that bound ATP one step ago (5). Adachi *et al.* almost identified the coupling mechanism finding that the other reaction before a 40° substep is P_i release and ADP is released at 240° after it is bound as ATP at 0° (6). The overall coupling scheme is summarized in Fig. 3.2.

It was proved that the mechano-chemical coupling is reversible. That is, mechanical rotation of γ-subunit in the synthesis direction by an external force leads to ATP synthesis like in the physiological condition (7). Thus, it has been believed that γ-angle controls the kinetics of chemical reactions in the three catalytic sites, which is sometimes referred to as γ-dictator mechanism (3, 8). Hirono-Hara *et al.* applied external forces on γ by magnetic tweezer and made γ-angle forward from the ADP-inhibited form, which fails to release ADP and poses after the 80° rotation (9). They found that the rate constant of ADP dissociation exponentially increases with γ-angle in the forward direction, which indicates γ-angle induces conformational change of the ADP-bound β-subunit to open. In contrast to the γ-dictator mechanism, Furuike *et al.* truncated γ-subunit from the bottom step by step until the remaining rotor simply sat on the concave entrance of $\alpha_3\beta_3$-orifice (8). Without firm pivoting at the bottom, all truncation mutants rotated in the correct direction, although the average rotary

speeds were low and irregular motions were observed. This may suggest another possibility that the $\alpha_3\beta_3$-cylinder by itself undergoes nucleotide-dependent circular conformational changes without supervision by γ, and the rotation of γ is simply driven by them.

Although the coupling scheme between γ-rotation and nucleotide state at three catalytic sites was basically clarified, physicochemical mechanism at molecular level is still unknown. X-ray structures revealed conformational changes of β-subunit and $\alpha\beta$-interface depending on nucleotide states (10) (Fig. 3.1(b)). Without nucleotide, β-subunit takes open conformation, which is denoted as β_E (α-subunit is also denoted as α_E). Upon ATP binding, β-subunit changes its conformation to closed one, which is denoted as β_{TP} (α-subunit is also denoted as α_{TP}) (Fig. 3.1(b)). Hydrolysis of ATP to ADP does not change β-subunit conformation so much, but change $\alpha\beta$-interface tightly packed, which is denoted as α_{DP}, β_{DP}. F_1-ATPase apparently uses these conformational changes to produce torque to rotate γ-subunit.

Molecular dynamics simulation is potentially powerful to visualize structural dynamics of the motor machine. Conventional all atom simulations have been used to observe the dynamics when a torque or a biasing force was applied to rotate γ-subunit in the synthesis direction (11, 12). But in the hydrolysis direction, it takes so much time to observe conformational changes, usually micro-milliseconds, that is beyond the reach of the all atom simulations (13). In this situation, coarse-grained (CG) model that neglects atomic details can be more useful. Koga and Takada proposed one of CG models, "switching Gō model", to describe conformational change of $\alpha\beta$-subunits and assumed only short-ranged repulsive interactions between γ and $\alpha_3\beta_3$-subunits (14). First, starting from the structure solved by Abrahams *et al.*(10) (denoted as the 1994 structure), they switched three $\alpha\beta$-subunit's conformations one step forward all at once, that is, from $\{\mathrm{DP, E, TP}\}_{\alpha\beta 1,\alpha\beta 2,\alpha\beta 3}$ (Fig. 3.2(a)) to $\{\mathrm{E, TP, DP}\}_{\alpha\beta 1,\alpha\beta 2,\alpha\beta 3}$ (Fig. 3.2(e)), and observed $\approx 120°$ rotations of γ. This result indicates that structural complementarities are of primary importance for the rotation rather than detailed hydrophobic or side chain packing interactions. Second, they switched only αβ2 (corresponding to bottom right circle in Fig. 3.2(a)) from $\alpha\beta_E^{1994}$ to $\alpha\beta_{TP}^{1994}$ (See Materials and Methods for the notation) remaining other two the same. This situation corresponds to generally called, tri-site model (all three catalytic sites occupied at a certain time). The simulation results did not show any unidirectional rotations indicating the tri-site model is impossible. Third, from experimental suggestions (15, 16) they redefined the nucleotide states of the 1994 structure as $\{\mathrm{DP + P_i, E, TP}\}_{\alpha\beta 1,\alpha\beta 2,\alpha\beta 3}$ and assumed that upon phosphate release, αβ1 changes its conformation from $\alpha\beta_{DP}^{1994}$ to $\alpha\beta_{HC}^{2001}$ (so called "half closed" form). They also assumed simultaneous reactions of ATP binding at site 2 and ADP release at site1. Thus, the overall catalysis scheme is, starting from $\{\mathrm{DP + P_i, E, TP}\}_{\alpha\beta 1,\alpha\beta 2,\alpha\beta 3}$ with the 1994 structure, then the phosphate is released at site 1, which is

followed by the simultaneous ATP binding at site 2 and ADP release at site 1. In this scheme, the site occupancy of nucleotides is always two (always-bi-site model). As simulation results, they found the $\approx 30°$ substep upon phosphate release and the $\approx 80°$ substep upon simultaneous reactions of ATP binding and ADP release and $\approx 10°$ rotation upon ATP hydrolysis. Surprisingly, the proposed always-bi-site model turned out to fit perfectly well with the completed experimental scheme published later (6).

In this chapter, we applied the multiple-basin model that we developed in chapter 1 and the implicit ligand binding model that we developed in chapter 2 to working mechanism of F_1-ATPase. Comparing to the switching Gō model developed by Koga and Takada (14), the multiple-basin model can describe spontaneous conformational transitions driven by thermal fluctuations rather than sudden potential switch in the switching Gō model. Applying these methods, we want to clarify allosteric regulations in working mechanism of this motor. First, we examine correlation between nucleotide state and x-ray structure. The $\alpha\beta$-structure that corresponds to the DP state (after P_i release) and the structure before the 80° substep (Fig. 3.2(a)) have not been known clearly. We carried out structural comparison of $\alpha\beta$-structures to identify the structure, and using the structure, we simulate substep rotation of γ to validate the result. Second, we focus on allosteric regulation of ADP release from the ADP site (upper circle in Fig. 3.2(a)) by γ-rotation, which indicates γ-dictator mechanism. We set up a simulation model that corresponds to the experiment by Hirono-Hara *et al.*(9). Third, we focus on allosteric coupling among ATP-binding on the Empty site (bottom right circle in Fig. 3.2(a)), ADP-release from the ADP site (upper circle in Fig. 3.2(a)), and γ-rotation, which should induce the 80° substep.

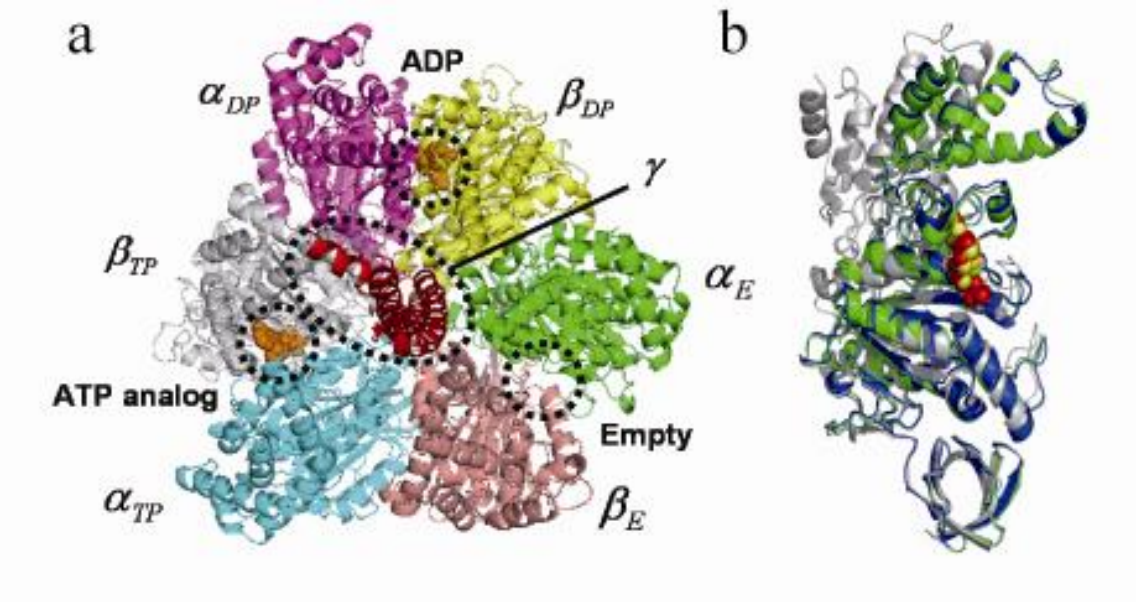

Fig. 3.1 X-ray structure of F_1-ATPase (a) $\alpha_3\beta_3\gamma$-complex of bovine F_1-ATPase(10). γ-subunit and three catalytic sites are circled with dotted line. **(b)** Three β-subunits are fitted using N terminal parts (bottom parts in the figure). β_E is shown in gray, β_{DP} is shown in green, and β_{TP} is shown in blue. ATP analog is shown in red spheres and ADP is shown in yellow spheres. C terminal parts (upper parts in the figure) show considerable conformational change upon nucleotide binding, although β_{DP} and β_{TP} show almost the same structure.

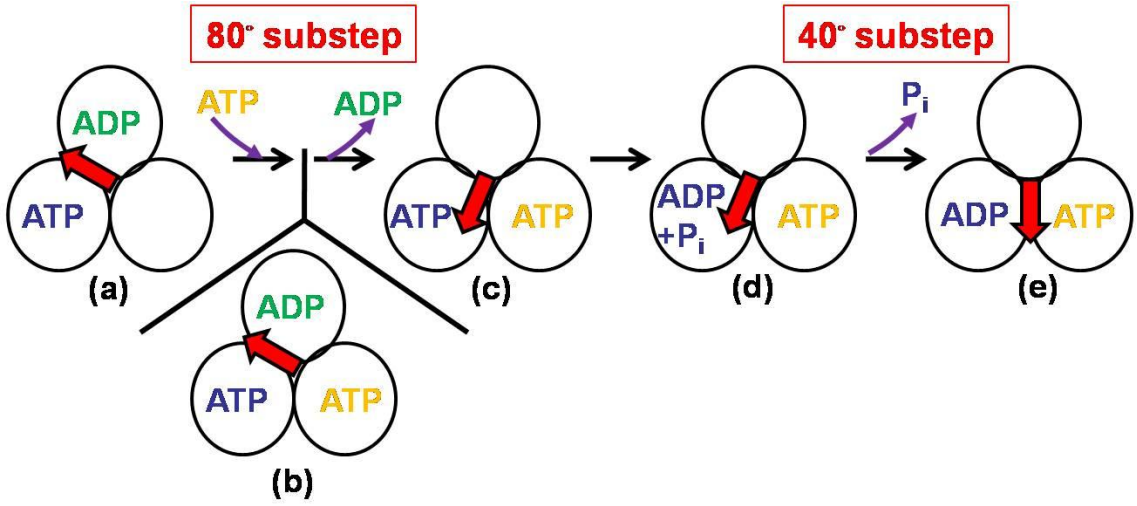

Fig. 3.2 Experimentally identified coupling scheme of γ-rotation and nucleotide states(6) In this figure, adjacent αβ-subunits are grouped together and represented by circles. The colors of nucleotide represent the specific sites in which the nucleotide belongs to. The 80° substep is accompanied by ATP-binding and ADP-release. On the other hand, the 40° substep is accompanied by P_i-release.

3.2 Materials and Methods

3.2.1 Crystal structures of F_1-ATPase used

We used three x-ray structures of F_1-ATPase. One is bovine F_1-ATPase determined in 1994 by Abrahams *et al.* (PDB number; 1bmf, described as the 1994 structure) (10). The three catalytic sites are at the interfaces between α- and β-subunits and predominantly hosted by β-subunits. So, these adjacent $\alpha\beta$-subunits are grouped together. The three catalytic sites in the 1994 structure were filled with AMP-PNP (ATP analog), ADP and nothing, which are denoted as $\alpha\beta_{TP}^{1994}$, $\alpha\beta_{DP}^{1994}$ and $\alpha\beta_{E}^{1994}$ respectively. P_i was not present in the ADP site, so Abrahams *et al.* suggested that the structure is in the ADP-inhibited form (10). Later, it was suggested that the ADP site might bind azide and it prevented P_i from binding (13, 17). Another structure is bovine mitochondrial F_1-ATPase determined in 2001 by Menz *et al*. (PDB number; 1h8e, described as the 2001 structure) (18). In this structure, all three catalytic sites bind nucleotides. $\alpha\beta_{TP}$ and $\alpha\beta_{DP}$ bind ADP and fluoroaluminate (AlF_4^-), and $\alpha\beta_{HC}$ binds ADP and sulfate, which are denoted as $\alpha\beta_{TP}^{2001}$, $\alpha\beta_{DP}^{2001}$, $\alpha\beta_{HC}^{2001}$. $\alpha\beta_{HC}^{2001}$ takes a novel structure in which $\alpha\beta$-interface is closed although β-subunit takes open structure, so this structure is called “half closed” (HC). The third structure is yeast F_1-ATPase determined in 2006 by Kabaleeswaran *et al*. (PDB number; 2hld, described as the 2006 structure) (19). The three catalytic sites in the 2006 structure were filled with AMP-PNP, AMP-PNP and nothing, which are denoted as $\alpha\beta_{TP}^{2006}$, $\alpha\beta_{HO}^{2006}$ and $\alpha\beta_{E}^{2006}$ respectively. The interface between α_{HO}^{2006} and β_{HO}^{2006} is more open than that of α_{DP}^{1994} and β_{DP}^{1994}, so it is called “half open” (HO) structure (13). The other two, $\alpha\beta_{TP}^{2006}$ and $\alpha\beta_{E}^{2006}$ are essentially the same to that of the 1994 structure.

3.2.2 Simulating substep rotation of γ-subunit

First, $\alpha_3\beta_3$-subunits are grouped into three. As described above, α-and β-subunits that host the same catalytic site are grouped together; chain C&D, A&E and B&F of the x-ray structures are grouped together, which are denoted as $\alpha\beta 1$ (ADP site), $\alpha\beta 2$ (Empty site) and $\alpha\beta 3$ (ATP site) respectively (Fig. 3.3). When we consider interactions, we group the interactions in the same way. That is, $\alpha\beta 1$ contains intra-chain interactions of chain C and D, inter-chain interactions of chain C-D and chain D-A. The last inter-chain interaction of chain D-A is not trivial. It is not clear that into which the interactions between the three grouped $\alpha\beta$-subunits should be grouped. We focus on the fact that α-subunits do not change conformations. So, it is natural to group these interactions into β-subunit’s sides that mainly contribute conformational changes. Therefore, $\alpha\beta 2$ contains intra-chain interactions of chain A and E, inter-chain interactions of chain A-E and E-B, and $\alpha\beta 3$ contains intra-chain interactions of chain B and F,

inter-chain interactions of chain B-F and F-C (Fig. 3.3).

We prepare three structures; the 1994 structure, the structure in which αβ1 of the 1994 structure is replaced to $\alpha\beta_{\mathrm{HO}}^{2006}$ (the near 2006 structure, See Results), and the structure in which three αβs are replaced to the structures of the 1994 structure after 120° rotation (the 120° rotated 1994 structure). We performed two steps of simulation; at step 1, starting with the 1994 structure, we change conformation of αβ1 to the near 2006 structure, and at step 2, starting with the last structure of step 1, we change three αβs to the 120° rotated 1994 structure. We describe conformational changes by multiple-basin model that we developed in chapter 1. We assume that the interaction between $\alpha_3\beta_3$ and γ is only repulsive interaction and constrain the coordinates of N terminal of three β-subunits analogous to the His-tag used in single molecule experiments to fix the molecule on a plate and C terminal of γ-subunit otherwise it flies away. Thus, the energy function becomes,

$$\mathrm{V} = \mathrm{V}_{\alpha\beta\,1} + \mathrm{V}_{\alpha\beta\,2} + \mathrm{V}_{\alpha\beta\,3} + \mathrm{V}_{\gamma} + \mathrm{V}_{\alpha 3\beta 3-\gamma} + \mathrm{V}_{\mathrm{spring}}$$

At step 1,

$$\mathrm{V}_{\alpha\beta\,1} = \frac{\mathrm{V_G}\left(\mathrm{R}_{\alpha\beta\,1}|\alpha\beta_{\mathrm{DP}}^{1994}\right) + \mathrm{V_G}\left(\mathrm{R}_{\alpha\beta\,1}|\alpha\beta_{\mathrm{HO}}^{2006}\right) + \Delta\mathrm{V}_{\alpha\beta\,1}}{2} - \sqrt{\frac{\mathrm{V_G}\left(\mathrm{R}_{\alpha\beta\,1}|\alpha\beta_{\mathrm{DP}}^{1994}\right) - \mathrm{V_G}\left(\mathrm{R}_{\alpha\beta\,1}|\alpha\beta_{\mathrm{HO}}^{2006}\right) - \Delta\mathrm{V}_{\alpha\beta\,1}}{2} + \Delta^2}$$

$$\mathrm{V}_{\alpha\beta\,2} = \mathrm{V_G}\left(\mathrm{R}_{\alpha\beta\,2}|\alpha\beta_{\mathrm{E}}^{1994}\right)$$

$$\mathrm{V}_{\alpha\beta\,3} = \mathrm{V_G}\left(\mathrm{R}_{\alpha\beta\,3}|\alpha\beta_{\mathrm{TP}}^{1994}\right)$$

At step 2,

$$\mathrm{V}_{\alpha\beta\,1} = \frac{\mathrm{V_G}\left(\mathrm{R}_{\alpha\beta\,1}|\alpha\beta_{\mathrm{HO}}^{2006}\right) + \mathrm{V_G}\left(\mathrm{R}_{\alpha\beta\,1}|\alpha\beta_{\mathrm{E}}^{1994}\right) + \Delta\mathrm{V}_{\alpha\beta\,1}}{2} - \sqrt{\frac{\mathrm{V_G}\left(\mathrm{R}_{\alpha\beta\,1}|\alpha\beta_{\mathrm{HO}}^{2006}\right) - \mathrm{V_G}\left(\mathrm{R}_{\alpha\beta\,1}|\alpha\beta_{\mathrm{E}}^{1994}\right) - \Delta\mathrm{V}_{\alpha\beta\,1}}{2} + \Delta^2}$$

$$\mathrm{V}_{\alpha\beta\,2} = \frac{\mathrm{V_G}\left(\mathrm{R}_{\alpha\beta\,2}|\alpha\beta_{\mathrm{E}}^{1994}\right) + \mathrm{V_G}\left(\mathrm{R}_{\alpha\beta\,2}|\alpha\beta_{\mathrm{TP}}^{1994}\right) + \Delta\mathrm{V}_{\alpha\beta\,2}}{2} - \sqrt{\frac{\mathrm{V_G}\left(\mathrm{R}_{\alpha\beta\,2}|\alpha\beta_{\mathrm{E}}^{1994}\right) - \mathrm{V_G}\left(\mathrm{R}_{\alpha\beta\,2}|\alpha\beta_{\mathrm{TP}}^{1994}\right) - \Delta\mathrm{V}_{\alpha\beta\,2}}{2} + \Delta^2}$$

$$V_{\alpha\beta 3} = \frac{V_G\left(R_{\alpha\beta 3}|\alpha\beta_{TP}^{1994}\right) + V_G\left(R_{\alpha\beta 3}|\alpha\beta_{DP}^{1994}\right) + \Delta V_{\alpha\beta 3}}{2}$$

$$- \sqrt{\frac{V_G\left(R_{\alpha\beta 3}|\alpha\beta_{TP}^{1994}\right) - V_G\left(R_{\alpha\beta 3}|\alpha\beta_{DP}^{1994}\right) - \Delta V_{\alpha\beta 3}}{2} + \Delta^2}$$

Common for step 1 and 2,

$$V_\gamma = V_G\left(R_\gamma|\gamma^{1994}\right)$$

$$V_{\alpha 3\beta 3-\gamma} = \sum_{\substack{i\in\alpha 3\beta 3,\\ j\in\gamma}} \varepsilon_{rep}\left(\frac{D}{r_{ij}}\right)^{12}$$

$$V_{spring} = \begin{cases} 0.0 & (d_i < d_i^C) \\ 0.18\left(d_i - d_i^C\right)^2 & (d_i \geq d_i^C) \end{cases}$$

where $V_G(R_X|Y)$ is single Gō potential of X, Y is its reference structure (we changed two parameters from the original one used in chapter 1, $\varepsilon_1 = 0.36$, $\varepsilon_2 = 0.36$), $\varepsilon_{rep} = 0.36$, $D = 6.0$A, d_i is displacement length of residue i from initial coordinate, i is three $\beta 1$ and $\gamma 272$ (tail, see Fig. 3.4), d_i^C is set to 2.0Å for $\beta 1$ and 1.5Å for $\gamma 272$. $\Delta = 200.0$ throughout the simulations, at step 1, $\Delta V_{\alpha\beta 1}$ is changed from 100.0 to -500.0 at 10^6th MD step. At step 2, we tried two ways; in one way, $\Delta V_{\alpha\beta 1}$ and $\Delta V_{\alpha\beta 2}$ are changed from 100.0 to -500.0 at 10^5th MD step and then $\Delta V_{\alpha\beta 3}$ is changed from 100.0 to -500.0 at 1.1×10^6th MD step, in the other way, we changed all at 10^5th MD step. Energy unit is arbitrary, but relative value compared to thermal energy k_BT has meaning (See section 3.2.5 for the value of thermal energy).

For the definition of the rotation angle of γ, we first transform the coordinate system so that the vertical direction that penetrates the $\alpha_3\beta_3$-cylinder-like structure corresponds to z-axis. We use the vector Λ that represents the direction from $\gamma 78$ (neck) to $\gamma 90$ (top) as below,

$$\Lambda = \sum_{i\in\gamma 79-90} \frac{\vec{r}_i - \vec{r}_{\gamma 78}}{\left|\vec{r}_i - \vec{r}_{\gamma 78}\right|}$$

We use x-y components of this vector to define the rotation angle from the initial one,

$$\theta_\gamma(t) = \cos^{-1}\left(\frac{\Lambda_{xy}(t)\cdot\Lambda_{xy}(0)}{\left|\Lambda_{xy}(t)\right|\left|\Lambda_{xy}(0)\right|}\right)$$

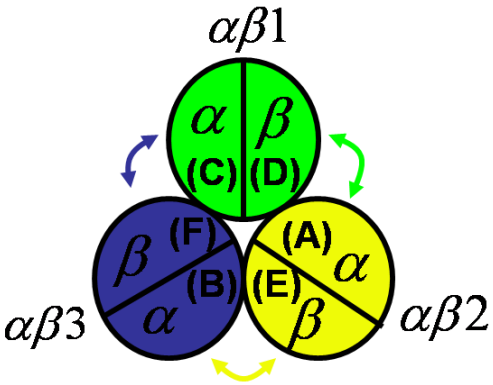

Fig. 3.3 Three interaction groups of the model
The colors; green, yellow, and blue, represent the three interaction groups. The interactions between paired αβ are grouped into β-side group.

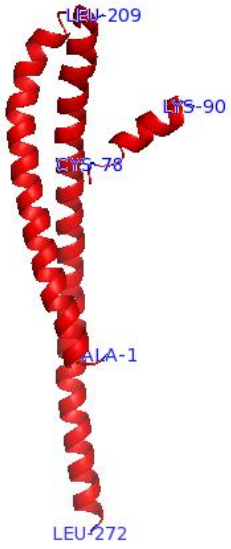

Fig. 3.4 X-ray structure of γ-subunit (the 1994 structure)
Amino acid sequences and residue numbers are described. Leu-272 is called "tail", Cys-78 is called "neck", Lys-90 is called "top", and Leu-209 is called "head".

3.2.3 Simulating allosteric interaction between ADP site and γ-subunit

In this simulation, we focus on conformational change of the ADP site, that is, $\alpha\beta 1$ under γ -angle constraints. So, $\alpha\beta 1$ is set to double-basin potential energy that we proposed recently (See chapter 1). $\alpha\beta 2$, $\alpha\beta 3$ and γ are set to single-basin potential using single Gō model. We consider ADP binding effect in the ADP site by implicit ligand model that is described in chapter 2. Then we add constraint potential to fix γ -angle at arbitrary angles. Thus, the potential energy function becomes,

$$V = V_{\alpha\beta 1} + V_{\alpha\beta 2,\alpha\beta 3} + V_{\text{bind}-\text{ADP}} + V_{\gamma-\text{angle}} + V_{\gamma} + V_{\alpha 3\beta 3-\gamma} + V_{\text{spring}}$$

$$V_{\alpha\beta\,1} = \frac{V_G\left(R_{\alpha\beta\,1}|\alpha\beta_{DP}^{1994}\right) + V_G\left(R_{\alpha\beta\,1}|\alpha\beta_{E}^{1994}\right) + \Delta V}{2} - \sqrt{\left(\frac{V_G\left(R_{\alpha\beta\,1}|\alpha\beta_{DP}^{1994}\right) - V_G\left(R_{\alpha\beta\,1}|\alpha\beta_{E}^{1994}\right) - \Delta V}{2}\right)^2 + \Delta^2}$$

$$V_{\alpha\beta\,2,\alpha\beta\,3} = V_G\left(R_{\alpha\beta\,2}|\alpha\beta_{E}^{1994}\right) + V_G\left(R_{\alpha\beta\,3}|\alpha\beta_{TP}^{1994}\right)$$

$$V_{bind\,-ADP} = \begin{cases} 0.0 \text{ (if ADP is unbound)} \\ \sum_{\substack{\text{ligand} \\ \text{contacts}}} -c_{lig}\,\varepsilon_1 \exp\left[-\frac{\left(r_{ij}/r_{0ij} - 1.0\right)^2}{2\left(\sigma/r_{0ij}\right)^2}\right] \text{ (if ADP is bound)} \end{cases}$$

$$V_{\gamma-angle} = K_{\gamma-angle}\left(\cos\theta_\gamma - \cos\theta_\gamma^C\right)^2 + K_{\gamma-angle}\left(\sin\theta_\gamma - \sin\theta_\gamma^C\right)^2$$

where $V_G(R_X|Y)$ is single Gō potential of X, Y is its reference structure, $\Delta = 190.0$, $\Delta V = -27.5$, $\varepsilon_{rep} = 0.2$, $D = 4.0$ Å, $K_{\gamma-angle} = 10.0$, θ_γ is the rotation angle of γ, which is measured at $\gamma 90$ (top, see Fig. 3.4), θ_γ^C is the constraint angle, $C_{lig} = 1.7$, $\sigma/r_{0ij} = 0.05$. The last three terms, V_γ, $V_{\alpha 3\beta 3-\gamma}$, V_{spring} are the same as the section 3.2.2. Energy unit is arbitrary, but relative value compared to thermal energy k_BT has meaning (See section 3.2.5 for the value of thermal energy). For energy function of the single Gō potential, see chapter 1.

3.2.4 Simulating allosteric coupling between ATP-binding and ADP-release

In this simulation, we focus on allosteric coupling between ATP-binding and ADP-release, which accompany the 80° substep (Fig. 3.2(a)-(c)). As described above, we grouped interactions into three groups (Fig. 3.3). All three αβ-subunits are set to have double-basin potential energy by the model we proposed recently (See chapter 1). We mainly used the 2006 structure and defined its nucleotide state as $\{DP, E, TP\}_{\alpha\beta\,1,\alpha\beta\,2,\alpha\beta\,3}$ (See Results), which corresponds to before the 80° substep (Fig. 3.2(a)). We used the 2006 structure as the initial structure and set the potential so that one basin is at the 2006 structure and the other is the structure after 80° rotation. Thus, $\alpha\beta 1$ has two potential basins at $\alpha\beta_{HO}^{2006}$ and $\alpha\beta_{E}^{2006}$, $\alpha\beta 2$ at $\alpha\beta_{E}^{2006}$ and $\alpha\beta_{TP}^{2006}$, and $\alpha\beta 3$ at $\alpha\beta_{TP}^{2006}$ and $\alpha\beta_{DP}^{1994}$. We used $\alpha\beta_{DP}^{1994}$ because there is no $DP + P_i$ state in the 2006 structure ($\alpha\beta_{HO}^{2006}$ is in DP state (after P_i release)). In this model, we analyze the coupling between conformational change of $\alpha\beta 2$ upon ATP-binding and $\alpha\beta 1$ upon ADP-release. Thus, the energy function becomes,

$$V = V_{\alpha\beta\,1} + V_{\alpha\beta\,2} + V_{\alpha\beta\,3} + V_\gamma + V_{\alpha 3\beta 3-\gamma} + V_{spring}$$

$$V_{\alpha\beta 1} = \frac{V_G\left(R_{\alpha\beta 1}|\alpha\beta_{HO}^{2006}\right) + V_G\left(R_{\alpha\beta 1}|\alpha\beta_{E}^{2006}\right) + \Delta V_{\alpha\beta 1}}{2} - \sqrt{\left(\frac{V_G\left(R_{\alpha\beta 1}|\alpha\beta_{HO}^{2006}\right) - V_G\left(R_{\alpha\beta 1}|\alpha\beta_{E}^{2006}\right) - \Delta V_{\alpha\beta 1}}{2}\right)^2 + \Delta^2}$$

$$V_{\alpha\beta 2} = \frac{V_G\left(R_{\alpha\beta 2}|\alpha\beta_{E}^{2006}\right) + V_G\left(R_{\alpha\beta 2}|\alpha\beta_{TP}^{2006}\right) + \Delta V_{\alpha\beta 2}}{2} - \sqrt{\left(\frac{V_G\left(R_{\alpha\beta 2}|\alpha\beta_{E}^{2006}\right) - V_G\left(R_{\alpha\beta 2}|\alpha\beta_{TP}^{2006}\right) - \Delta V_{\alpha\beta 2}}{2}\right)^2 + \Delta^2}$$

$$V_{\alpha\beta 3} = \frac{V_G\left(R_{\alpha\beta 3}|\alpha\beta_{TP}^{2006}\right) + V_G\left(R_{\alpha\beta 3}|\alpha\beta_{DP}^{1994}\right) + \Delta V_{\alpha\beta 3}}{2} - \sqrt{\left(\frac{V_G\left(R_{\alpha\beta 3}|\alpha\beta_{TP}^{2006}\right) - V_G\left(R_{\alpha\beta 3}|\alpha\beta_{DP}^{1994}\right) - \Delta V_{\alpha\beta 3}}{2}\right)^2 + \Delta^2}$$

$$V_{\gamma} = V_G\left(R_{\gamma}|\gamma^{2006}\right)$$

$$V_{\alpha 3\beta 3-\gamma} = \sum_{\substack{i\in\alpha 3\beta 3,\\ j\in\gamma}} \varepsilon_{rep}\left(\frac{D}{r_{ij}}\right)^{12}$$

$$V_{spring} = \begin{cases} 0.0 & (d_i < d_i^C) \\ 0.18\left(d_i - d_i^C\right)^2 & (d_i \geq d_i^C) \end{cases}$$

where $V_G(R_X|Y)$ is single Gō potential of X, Y is its reference structure (we changed two parameters from the original one used in chapter 1, $\varepsilon_1 = 0.36$, $\varepsilon_2 = 0.36$), $\varepsilon_{rep} = 0.36$, $D = 6.0$A, d_i is displacement length of residue i from initial coordinate, i is three β1 and γ276 (tail), d_i^C is set to 2.0Å for β1 and 1.5Å for γ276. $\Delta = 200.0$, $\Delta V_{\alpha\beta 1} = \Delta V_{\alpha\beta 3} = -25.0$. For $\Delta V_{\alpha\beta 2}$, we change the value during the simulations, which mimics a ATP binding event. We use $\Delta V_{\alpha\beta 2} = 0.0$ for first 100,000 MD step, then switch it to $\Delta V_{\alpha\beta 2} = -100.0$. Energy unit is arbitrary, but relative value compared to thermal energy k_BT has meaning (See section 3.2.5 for the value of thermal energy).

Inspired by the experiment truncating γ-axle step by step (8), we set up a similar situation on the model above. For realizing truncation of γ-axle, we modified interactions between $\alpha_3\beta_3$ and γ. We deleted the interactions between truncated γ residues and $\alpha_3\beta_3$-subunits, although internal interactions of γ-subunit was not modified. We fixed the C-terminal side residue after the truncation, instead of the tail residue, for which only z-axis direction was fixed (z-axis is in

line with axle direction).

3.2.5 Simulation protocols

We perform molecular dynamics simulations using Langevin equation (See Appendix 1) (20, 21). The masses for all beads are set to identical. The simulation temperature is controlled through the random force term in Langevin equation, and set so that thermal energy becomes $k_BT = 0.4$ (section 3.2.3), or $k_BT = 0.3$ (section 3.2.2, 3.2.4) (Energy unit is arbitrary, see Appendix 1). The friction constant is set to $\gamma = 0.25$ (section 3.2.3) or $\gamma = 0.05$ (section 3.2.2, 3.2.4), which means the relaxation time becomes 20 MD step or 100 MD step, respectively.

3.3 Results

3.3.1 αβ-structure corresponding to each nucleotide state

The coupling scheme between γ rotation and three nucleotide states was basically identified recently (6). So, the next step is to assign the structure to each nucleotide state. Single molecule studies have revealed that F_1 has two stable conformational states before 80° or 40° substeps, which are referred to as the "binding dwell state" (Fig. 3.2 (a) and (e)) and "catalytic dwell state" (Fig. 3.2 (c) and (d)) respectively. The 1994 structure was identified as catalytic dwell state by simulations (14) and experiments (10, 22). But, the structure that corresponds to binding dwell state is still unclear, especially, αβ-structure corresponding to the DP state (after P_i release) has been unclear. According to the coupling scheme experimentally completed, P_i release induces 40° substep. So, there should be some conformational change from the DP+P_i state (16). Koga and Takada used $\alpha\beta_{HC}^{2001}$ as the DP state and observed $\approx 30^\circ$ substep, but it was somewhat arbitrary (14). In this section, we compare all αβ-structures of the three x-ray structures; the 1994, 2001, 2006 structures, and try to specify the αβ-structure of the DP state.

We calculated root mean square deviation (RMSD) using only C_α atoms for all pairs of αβ-structures. When we compare structures from different species, we ignored the regions identified as gap or insertion by sequence alignment (blast2sequence(23) was used). The results are summarized in Table 3.1. Then, based on pairwise RMSD values, we mapped all αβ-structures on two-dimensional surface using multi-dimensional scaling (one of multivariate analysis) (Fig. 3.5). From the mapping, we can see relative distances among αβ-structures. On the map, there are three groups and two isolated structures. The three groups; $\alpha\beta_{TP}^{1994,2001,2006}$, $\alpha\beta_{DP}^{1994,2001}$ and $\alpha\beta_{E}^{1994,2006}$, correspond to the TP, DP+P_i and E state, respectively. One isolated structure, $\alpha\beta_{HO}^{2006}$, is the half open, and the other, $\alpha\beta_{HC}^{2001}$, is the half closed. We can assume that the DP state is in between the DP+P_i and E states. Therefore, we identified the half open structure ($\alpha\beta_{HO}^{2006}$) as the DP state rather than the half closed ($\alpha\beta_{HC}^{2001}$). From this result, we can identify the 2006 structure as the binding dwell state. This result is supported by the x-ray structure of the 2006 structure. We fitted three independent structures in the 2006 structure using only α3β3-subunits, and observed tendency for γ to rotate in the hydrolysis direction (data not shown).

RMSD	1bmf_ AE	1bmf_ BF	1bmf_ CD	1h8e_ AE	1h8e_ BF	1h8e_ CD	2hld_ AE	2hld_ BF	2hld_ CD
1bmf_AE		5.3	5.1	2.2	5	4.9	1	5.1	4.1
1bmf_BF			2	4.5	0.4	1.9	5.3	1	2
1bmf_CD				3.6	2	0.3	5.3	1.9	1.6
1h8e_AE					4.5	3.8	2.7	4.3	3.1
1h8e_BF						2	5.1	1.1	2.1
1h8e_CD							5.1	1.8	1.7
2hld_AE								5.1	4.1
2hld_BF									1.6
2hld_CD									

Table 3.1 Pairwise RMSD of αβ-structures 1bmf is PDB number of the 1994 structure, 1h8e is PDB number of the 2001 structure, and 2hld is PDB number of the 2006 structure. "AE" represents chain A&E, "BF" represents chain B&F, and "CD" represents chain C&D.

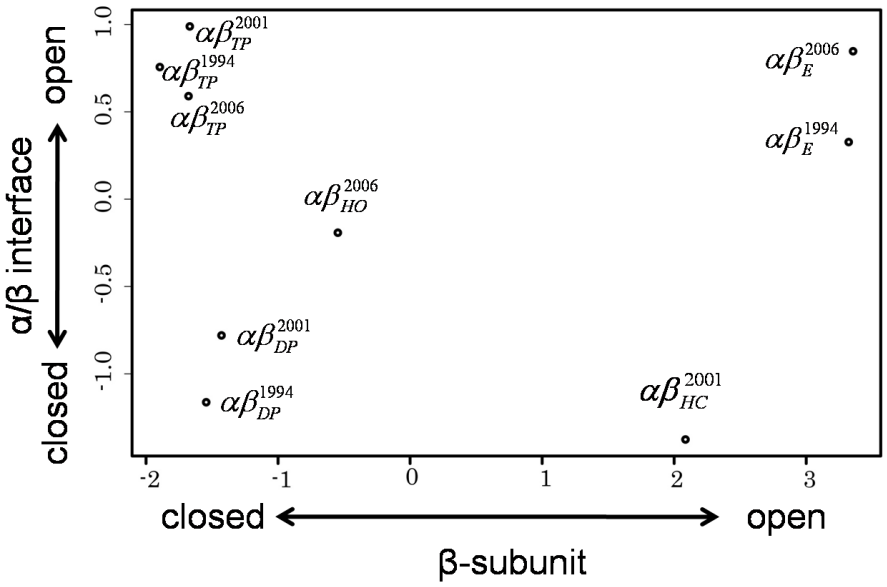

Fig. 3.5 Two-dimensional mapping of αβ-structures x-axis can be interpreted as conformational change of β-subunits and y-axis can be interpreted as conformational change of α/β interfaces.

3.3.2 Substep rotation of γ-subunit using the identified binding dwell structure

In the previous section, we assigned $\alpha\beta_{HO}^{2006}$ as the αβ-structure of DP state, thus, assigned the 2006 structure as the binding dwell structure. But the result was based on only structural comparison of αβ-subunits. Because single molecule experiments have shown that F_1 takes two conformational states, binding dwell and catalytic dwell states, to produce 80°+40° substep rotations of γ, more direct way to validate the structure of the binding dwell state is to show the substep rotations using the identified structures; the 1994 structure as the catalytic dwell state and the 2006 structure as the binding dwell state.

Using the two structures, we change conformations in two ways; as in Fig. 3.6(a), one way is (i)→(ii)→(iii)→(iv), and the other way is (i)→(ii)→(iv). The structure (ii) is almost the same as the 2006 structure, because we can see $\alpha\beta_{E}^{1994} \approx \alpha\beta_{E}^{2006}$, $\alpha\beta_{TP}^{1994} \approx \alpha\beta_{TP}^{2006}$ in Fig. 3.5. Thus, (i)→(ii) corresponds to the process from the catalytic dwell state to the binding dwell state, which is expected to produce 40° substep, and (ii)→(iv) corresponds to the process that is expected to produce 80° substep. (iii) is an intermediate state of the process (ii)→(iv) as assumed by Koga and Takada (14). In Fig. 3.6(b), 10 trajectories for the process (i)→(ii) are shown. We changed conformation from (i) to (ii) at 1.0×10^6th MD step. Starting with the last structures in Fig.3.6(b), trajectories for the process (ii)→(iii)→(iv) are shown in Fig. 3.6(c), and for the process (ii)→(iv) are shown in Fig. 3.6(d). We plotted histogram of rotation angle of γ using only last 5.0×10^5 MD steps for each state in Fig. 3.6(e) for the process (i)→(ii)→(iii)→(iv), in Fig. 3.6(f) for the process (i)→(ii)→(iv). For the process (i)→(ii)→(iii)→(iv), we observed 35° shift of peak position at (i)→(ii), 60° shift at (ii)→(iii), 30° shift at (iii)→(iv). In the study of Koga and Takada, they used the same structures for (iii)→(iv) and observed only $\approx 10^\circ$ rotation. In contrast to our Langevin dynamics simulation, they carried out Newtonian dynamics in low temperature (See Appendix 1). There is room for discussion on this inconsistency. For the process (i)→(ii)→(iv), we observed 35° shift of peak position at (i)→(ii), 85° shift at (ii)→(iv). This result is consistent with experiments. Therefore, we succeeded in producing substep rotations using the 2006 structure as the binding dwell state. Because the intermediate state (iii) produces additional substep that is not observed in experiments, this state may be not stable, that is, rotation of γ may induce the conformational change at bottom left site from TP state to DP+P_i state although ATP is not hydrolyzed simultaneously and will be hydrolyzed in the catalytic dwell. There is also room for discussion on this point.

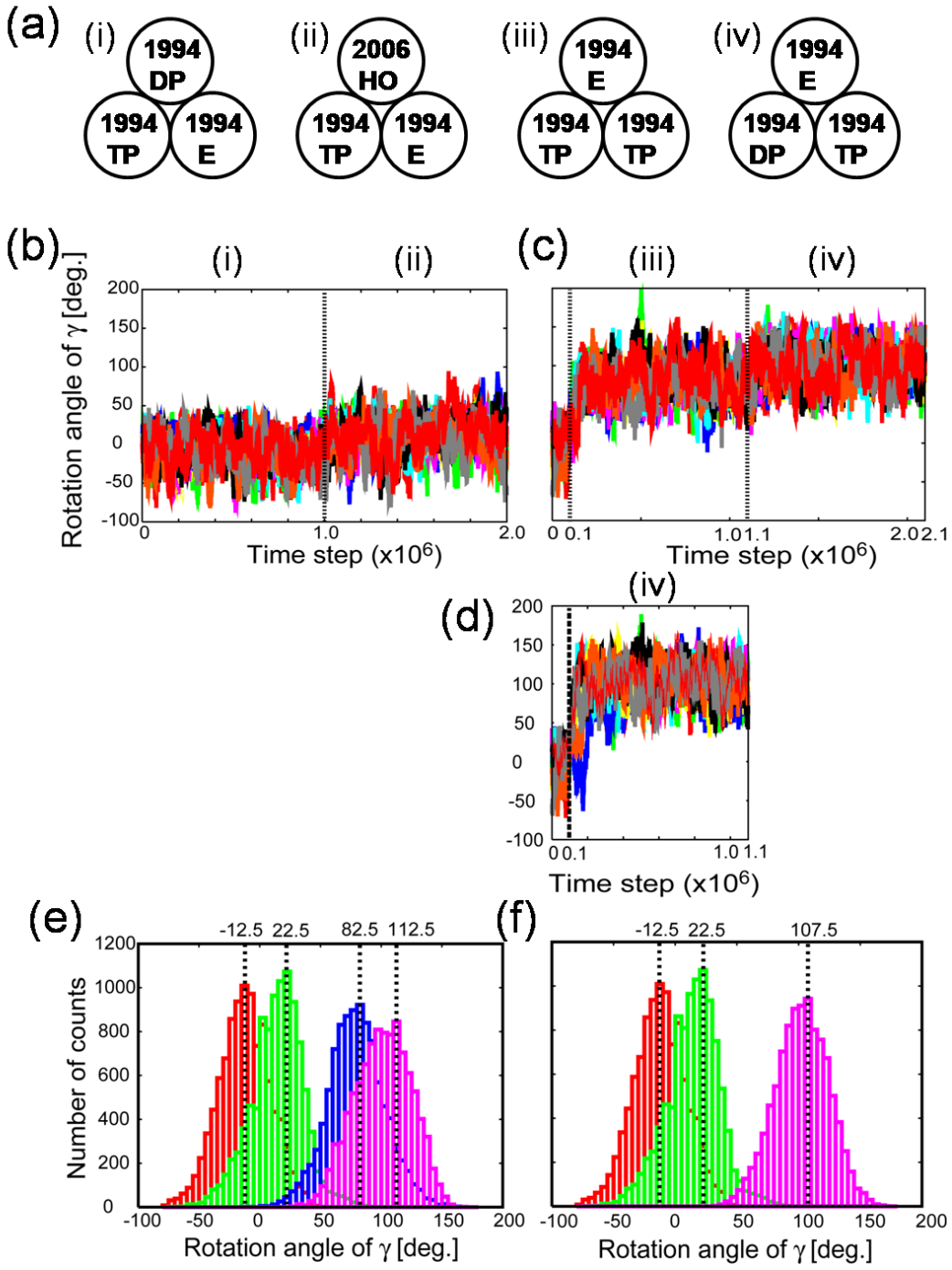

Fig. 3.6 Trajectories and histograms of substep rotations of γ (a) Structures we used for simulations of substep rotations. **(b)** 10 trajectories for the process (i)→(ii) **(c)** the process (ii)→(iii)→(iv) **(d)** the process (ii)→(iv) **(e)** The histogram of rotation angle of γ in (b)+(c). Red is for (i), green is for (ii), blue is for (iii), and purple is for (iv). **(f)** The histogram of rotation angle of γ in (b)+(d). The color representations are same as in (e).

3.3.3 Allosteric regulation of conformational change of the ADP site by γ-angle

To see a possibility of the γ-dictator mechanism in which γ-angle controls the nucleotide states at three catalytic sites, we focus on ADP release from the ADP site. There is an experiment by Hirono-Hara *et al.* in which ADP-inhibited form of F_1 were mechanically activated by pushing the γ-angle with magnetic tweezer (9). They showed the rate constant of the mechanical activation exponentially increased with the pushed angle, implying that γ-angle controls the ADP release from the ADP site by lowering the affinity. Here, we set up the simulation model that mimics the above experiment. In this model, γ-angle is controlled by the constraint potential $V_{\gamma-\mathrm{angle}}$, and conformational change at the ADP site is modeled by multiple-basin model that we proposed recently (See Materials and Methods for details). The initial conformation is the 1994 structure that has been suggested as the ADP-inhibited form (10). We focus on conformational change at the ADP site from closed ($\alpha\beta_{\mathrm{DP}}^{1994}$) to open ($\alpha\beta_{\mathrm{E}}^{1994}$).

The representative trajectories are shown in Fig. 3.7. The γ-angle was fixed at 0° or 80° in each trajectory. The ADP site changed its conformation to open when the γ-angle was fixed at 80°, but did not change the conformation when the γ-angle fixed at 0°. We plotted repulsive energy between chain D (β-subunit of the ADP site) and γ, and found that high repulsive energies were observed more frequently when fixed at 80° before the conformational change. When we look at the region around 1.0×10^5th time step, both trajectories had an intrinsic fluctuation to open conformation (Fig. 3.7(a)) (interaction energy with γ was low for both trajectories (Fig. 3.7(b))). But, after the fluctuation, only the trajectory with the γ-angle fixed at 80° was pushed strongly by γ (Fig. 3.7(b)) and changed its conformation to open. This implies that γ-subunit pushes β-subunit of the ADP site to open when it rotates in the forward direction. We also plotted ADP binding energy, and found that the conformational change to open weakened the binding energy and then ADP was released.

We further analyzed the γ-angle dependency of conformational change rate. We simulated with fixed γ-angle from 0° to 80° at 20° intervals, and plotted survival rate of the closed conformation for each fixed angle. For each fixed γ-angle, about 15 trajectories of 10^7 MD steps were analyzed. Then we fitted the plots by single exponential and estimated the rate constant of conformational change to open for each fixed angle. In Fig. 3.8(b), the rate constants along the fixed angle are plotted. There is a clear tendency that more the γ-angle rotates, faster the conformational change occurs. This tendency becomes much clearer after 40° rotation. Therefore, we have clearly proved that the γ-angle controls the conformational change of the ADP site by pushing β-subunit of the ADP site. In the experiment by Hirono-Hara *et al.*, they estimated the angle dependency of the activation energy as $-1.3\mathrm{k_BT}/10^\circ$, which means $\ln\left(\mathrm{k}_{+10^\circ}/\mathrm{k}_0\right) = 1.3$. Compared to the experiment, the angle dependency of the rate constant

obtained here is about 5 times weaker although the tendency agrees. It may be limitation of the coarse-grained model we used (there is no side-chain interaction).

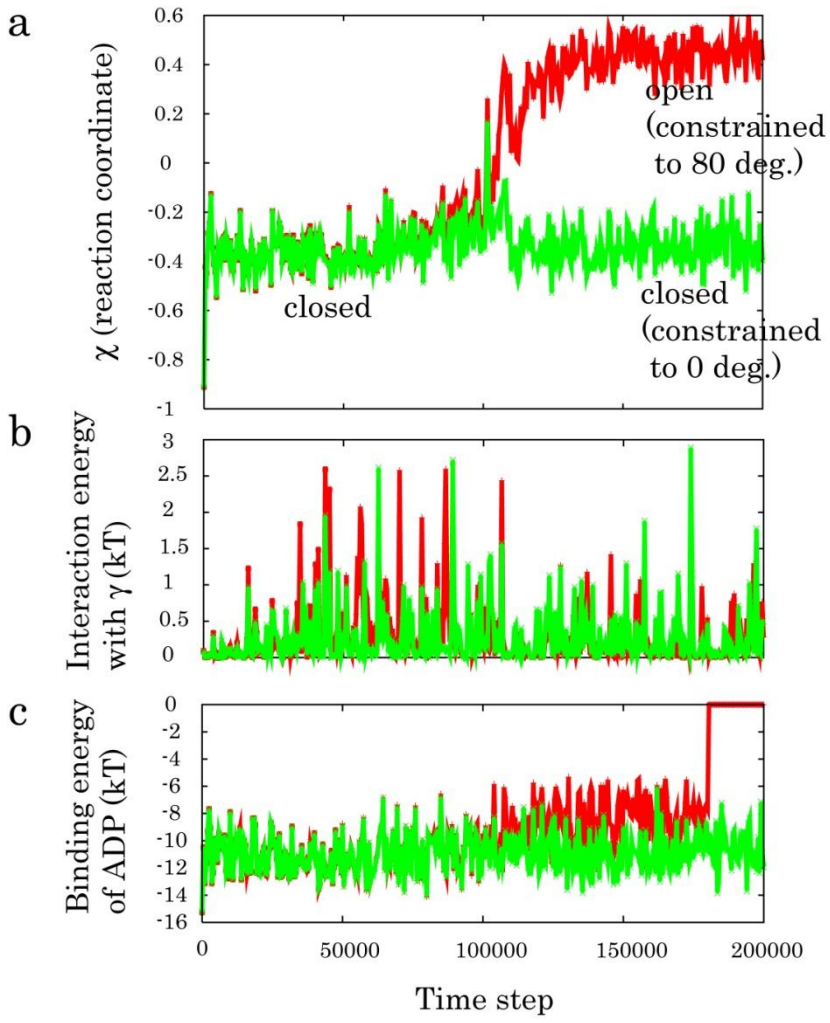

Fig. 3.7 Representative trajectories of allosteric regulation of conformational change of the ADP site by γ-angle (a) Reaction coordinate χ of the ADP site conformational change is plotted for the case of the constraint angle 0°(green) or 80°(red). **(b)** Interaction energy between γ-subunit and β-subunit of the ADP site is plotted for the case of the constraint angle 0°(green) or 80°(red). **(c)** Binding energy of ADP is plotted for the case of the constraint angle 0°(green) or 80°(red).

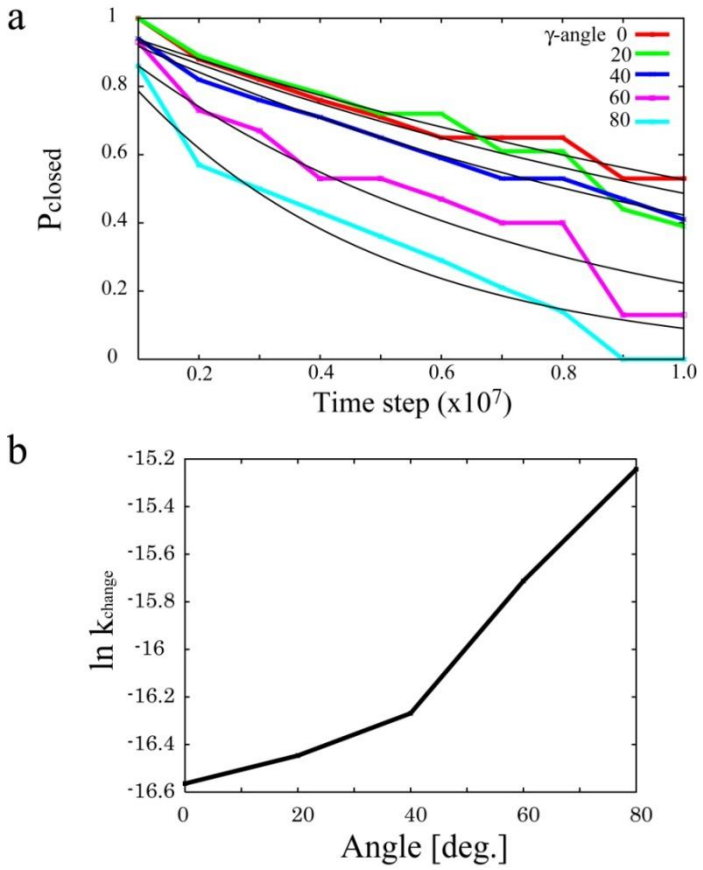

Fig. 3.8 Rate constant for the ADP site to open increases with forward γ-angle (a) The survival rate of the closed state is plotted along MD time step for each fixed γ-angle, and black lines are fitted curves by exponential. **(b)** The logarithm of the rate constants are plotted along fixed γ-angle.

3.3.4 γ-truncation simulation reveals allosteric coupling between ATP-binding and ADP-release

Although we showed above that simultaneous reactions of ATP binding to the Empty site and ADP release from the ADP site leads to 85° substep (section 3.3.2), these reactions are basically different elemental reactions. So, there should be a tightly coupled allosteric effect between the conformational change of the Empty site upon ATP binding and that of the ADP site upon ADP release. We especially focus on the role of γ-subunit to the allosteric effect because the recent experiment suggested that functional rotation in $\alpha_3\beta_3$-subunits may be enough for the rotation and γ-subunit does not play main role in the rotation (8). To observe this situation, we construct the model in which we used the 2006 structure as the initial structure because we

identified that the 2006 structure seems to be on pathway and the ATP-waiting, before 80° state (Fig. 3.2(a)). In simulation, we forced the Empty site (αβ2) to change conformation to closed, mimicking the ATP-binding induced conformational change, at certain MD step, then observed conformational changes of the ADP site (αβ1) and γ-rotation (See Materials and Methods).

In Fig. 3.9(a)-(e), 10 trajectories of 1×10^7 MD step are plotted on χ (reaction coordinate of the ADP site conformational change)-γ-angle surface. We classified trajectories into five types according to the direction of γ-rotation, the occurrence of transition of the ADP site, and the sequence of the events; type 1: transition → forward rotation, type 2: forward rotation→transition→forward rotation, type 3: backward rotation→transition→forward rotation, type 4: backward rotation→transition→backward rotation, type 5: no transition. In type 1-3 trajectories, we can see the transitions of the ADP site to open state lead to forward γ-angle, although in a type 4 trajectory, γ rotated reverse direction. We consider type 1 and 2 trajectories show normal behavior although timing of the ADP site transition is different, and type 3 and 4 trajectories show abnormal behavior. In Fig. 3.9(f) and (g), we plotted histogram of γ-angle before and after the ADP site conformational change. When we compare two distributions, we can observe the peak position shifts $\approx 85^\circ$ upon the ADP site conformational change. This result is consistent with experimental fact that ATP-binding and ADP-release leads to 80° substep.

Then, we truncated γ-axle step by step inspired by the experiment (8), and observed effect of the truncation to the 80° substep. We analyzed 10 trajectories of 1×10^7 MD step for each truncation. Statistics of trajectories are summarized in Fig. 3.10(a). We classified trajectories into five types in the same way as above. In this classification, type 1 and 2 correspond to normal behavior, in which no abnormal backward rotation is observed. Trajectories in which the ADP site did not change its conformation in simulation time correspond to type 5, and it increases as γ is truncated more. This result means that the truncation of γ strongly affects allostery between ATP-binding to the Empty site and ADP-release from the ADP site. Therefore, for the tight coupling between ATP-binding and ADP-release, the existence of γ-axle is necessary, which supports the γ-dictator mechanism. Then, the fraction of type 1 and 2 trajectories for each truncation was plotted in Fig. 3.10(b). The tendency of the plot along truncations is similar to the experiment (8), that is, it decreases until dC21 (deletion of 21 residues from C-terminal) and becomes relatively flat between dC21 and dN7C29 (deletion of 7 residues from N-terminal and 29 residues from C-terminal) and decreases again, although the experimentalists observed rotation rate instead of the fraction of normal behavior. But we could not observe the rotation from dN11C32 truncation, although the experimentalists observed that all truncations up to dN22C43 rotated in the correct direction. Because the rotation rate for dN22C43 becomes 100 times slower than the wild in the experiment, it may be impossible to

observe the rotation as a result of limited simulation time.

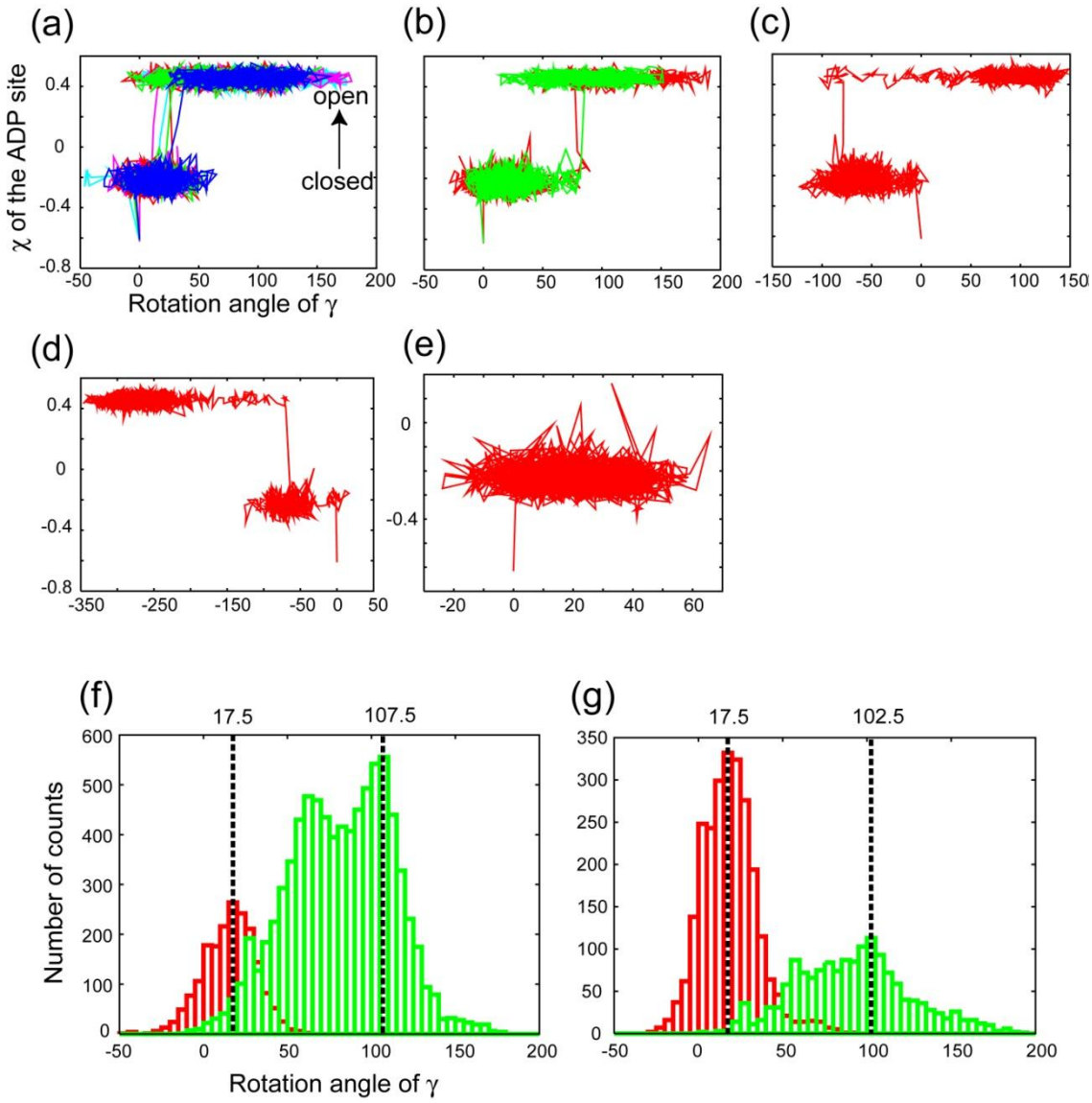

Fig. 3.9 Coupling between ATP-binding and ADP-release leads to the 80° substep (a)-(e) Trajectories are plotted on χ (reaction coordinate of the ADP site conformational change)-γ-angle surface. (a)-(e) correspond to type 1-5 trajectories (See text for details). **(f) and (g)** Histograms of γ-angle before (red) and after (green) the ADP site conformational change are plotted for type 1 trajectories in (f) and for type 2 trajectories in (g).

(a)

type	1	2	3	4	5
Wild	5	2	1	1	1
dC17	5	0	0	1	4
dC21	1	1	0	3	5
dN4C25	3	0	2	1	4
dN7C29	0	2	2	0	6
dN11CC32	0	0	0	0	10
dN22C43	0	0	0	0	10

(b)

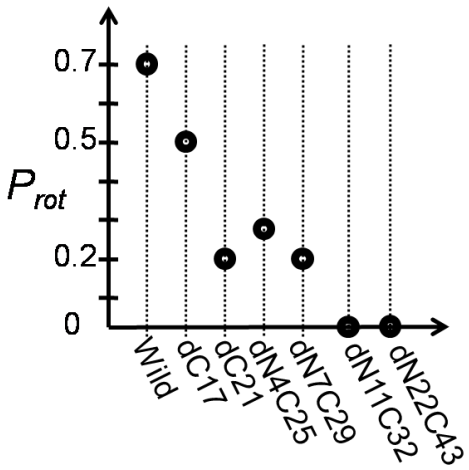

Fig. 3.10 The effect of γ-truncation to the 80° substep (a) Statistics of 10 trajectories for each truncation. We classified trajectories into five types (See text for details). **(b)** We plotted the fraction of type 1 and 2 trajectories for each truncation (P_{rot}).

3.4 Conclusion

We applied the multiple-basin model that describe protein conformational change and the implicit ligand binding model that describe ligand binding/unbinding to allostery in working mechanism of F_1-ATPase. As results, first, by analyzing αβ-structures of the three x-ray structures, we concluded that the DP state possibly corresponds to $\alpha\beta_{DP}^{2006}$, and the 2006 structure is the state before 80° substep. Second, we observed that rate constant of the ADP site conformational change is regulated by γ-angle, which suggests the γ-dictator mechanism. Interaction between γ-subunit and β-subunit of the ADP site was important for the conformational change. Third, we examined coupling between ATP-binding and ADP-release utilizing the 2006 structure. We confirmed $\approx 80^\circ$ rotation after the ADP site conformational change. We also truncated γ-axle step by step inspired by the experiment, and found that, for the tight coupling between ATP-binding and ADP-release, the existence of γ-axle is necessary, supporting γ-dictator mechanism.

References

1. Noji H, Yasuda R, Yoshida M, Kinosita K, Jr. (1997) Direct observation of the rotation of F1-ATPase. *Nature* 386:299-302.
2. Yasuda R, Noji H, Kinosita K, Jr., Yoshida M (1998) F1-ATPase is a highly efficient molecular motor that rotates with discrete 120 degree steps. *Cell* 93:1117-1124.
3. Kinosita K, Jr., Adachi K, Itoh H (2004) Rotation of F1-ATPase: how an ATP-driven molecular machine may work. *Annu Rev Biophys Biomol Struct* 33:245-268.
4. Yasuda R, Noji H, Yoshida M, Kinosita K, Jr., Itoh H (2001) Resolution of distinct rotational substeps by submillisecond kinetic analysis of F1-ATPase. *Nature* 410:898-904.
5. Shimabukuro K*, et al.* (2003) Catalysis and rotation of F1 motor: cleavage of ATP at the catalytic site occurs in 1 ms before 40 degree substep rotation. *Proc Natl Acad Sci U S A* 100:14731-14736.
6. Adachi K*, et al.* (2007) Coupling of rotation and catalysis in F(1)-ATPase revealed by single-molecule imaging and manipulation. *Cell* 130:309-321.
7. Itoh H*, et al.* (2004) Mechanically driven ATP synthesis by F1-ATPase. *Nature* 427:465-468.
8. Furuike S*, et al.* (2008) Axle-less F1-ATPase rotates in the correct direction. *Science* 319:955-958.
9. Hirono-Hara Y, Ishizuka K, Kinosita K, Jr., Yoshida M, Noji H (2005) Activation of pausing F1 motor by external force. *Proc Natl Acad Sci U S A* 102:4288-4293.
10. Abrahams JP, Leslie AG, Lutter R, Walker JE (1994) Structure at 2.8 A resolution of F1-ATPase from bovine heart mitochondria. *Nature* 370:621-628.
11. Bockmann RA, Grubmuller H (2002) Nanoseconds molecular dynamics simulation of primary mechanical energy transfer steps in F1-ATP synthase. *Nat Struct Biol* 9:198-202.
12. Ma J*, et al.* (2002) A dynamic analysis of the rotation mechanism for conformational change in F(1)-ATPase. *Structure* 10:921-931.
13. Pu J, Karplus M (2008) How subunit coupling produces the gamma-subunit rotary motion in F1-ATPase. *Proc Natl Acad Sci U S A* 105:1192-1197.
14. Koga N, Takada S (2006) Folding-based molecular simulations reveal mechanisms of the rotary motor F1-ATPase. *Proc Natl Acad Sci U S A* 103:5367-5372.
15. Yasuda R*, et al.* (2003) The ATP-waiting conformation of rotating F1-ATPase revealed by single-pair fluorescence resonance energy transfer. *Proc Natl Acad Sci U*

S A 100:9314-9318.

16. Masaike T, Muneyuki E, Noji H, Kinosita K, Jr., Yoshida M (2002) F1-ATPase changes its conformations upon phosphate release. *J Biol Chem* 277:21643-21649.
17. Bowler MW, Montgomery MG, Leslie AG, Walker JE (2006) How azide inhibits ATP hydrolysis by the F-ATPases. *Proc Natl Acad Sci U S A* 103:8646-8649.
18. Menz RI, Walker JE, Leslie AG (2001) Structure of bovine mitochondrial F(1)-ATPase with nucleotide bound to all three catalytic sites: implications for the mechanism of rotary catalysis. *Cell* 106:331-341.
19. Kabaleeswaran V, Puri N, Walker JE, Leslie AG, Mueller DM (2006) Novel features of the rotary catalytic mechanism revealed in the structure of yeast F1 ATPase. *Embo J* 25:5433-5442.
20. Honeycutt JD, Thirumalai D (1992) The nature of folded states of globular proteins. *Biopolymers* 32:695-709.
21. Veitshans T, Klimov, D, Thirumalai, D (1996) Protein folding kinetics: timescales, pathways and energy landscapes in terms of sequence-dependent properties. *Folding & Design* 2:1-22.
22. Okuno D*, et al.* (2008) Correlation between the conformational states of F1-ATPase as determined from its crystal structure and single-molecule rotation. *Proc Natl Acad Sci U S A* 105:20722-20727.
23. Tatusova TA, Madden TL (1999) BLAST 2 Sequences, a new tool for comparing protein and nucleotide sequences. *FEMS Microbiol Lett* 174:247-250.

Conclusions of the thesis

Throughout the thesis, we studied conformational change of proteins, ligand binding, and relationship between conformational change and ligand binding, that is, allostery.

In chapter 1, we construct a structure-based coarse-grained model that describes large-amplitude conformational changes of proteins, so called "multiple-basin model". As results of simulating conformational changes of four proteins; glutamine binding protein (GBP), S100A6, dihydrofolate reductase (DHFR) and HIV-1 protease, first, the different free energy surfaces that show the different mechanisms of conformational change (that is, hinge or shear motion) were obtained, processes of breaking contacts and making contacts are cooperative in hinge-type motion, whereas these process are sequential in shear-type motion. Second, the free energy surfaces at different simulation temperature showed that how conformational change pathway is altered as a function of temperature. In the case of S100A6, at the high temperature, the globally extended intermediate emerged on the free energy surface. The pathway through this intermediate occurs, although usual direct pathway from initial to final conformation still occurs. Third, the transition-state barrier height dependence on relative stability of the two conformations was shown to be linear, which implies partial unfolding.

In chapter 2, we investigated what kinds of ligand-binding processes prefer the induced-fit or population-shift route, constructing a simulation model of protein conformational change coupled with ligand-binding, in which the multiple-basin energy landscape is realized by the structure-based multiple-basin model Hamiltonian and the ligand-binding is modeled as the stochastic jump between unbound and bound surfaces with MD-MC coupling simulation scheme. We found that strong and long-ranged interaction favors the induced-fit pathway, whereas weak and short-ranged interaction leads to the population-shift pathway. These are consistent with many available experimental data.

In chapter 3, we applied the multiple-basin model that we developed in chapter 1 and the implicit ligand binding model that we developed in chapter 2 to working mechanism of F_1-ATPase. As results, first, by analyzing αβ-structures of the three x-ray structures, we concluded that the DP state possibly corresponds to $\alpha\beta_{\mathrm{DP}}^{2006}$, and the 2006 structure is the state before 80° substep. Second, we observed that rate constant of the ADP site conformational change is regulated by γ-angle, which suggests the γ-dictator mechanism. Third, we examined coupling between ATP-binding and ADP-release utilizing the 2006 structure. We confirmed $\approx 80^\circ$ rotation after the ADP site conformational change. We also truncated γ-axle step by step inspired by the experiment, and found that, for the tight coupling between ATP-binding and ADP-release, the existence of γ-axle is necessary, supporting γ-dictator mechanism.

Appendix 1: Methods for molecular dynamics simulation

We used two different methods for molecular dynamics simulation with constant temperature. In chapter 1 and 2, Newtonian dynamics with Berendsen thermostat was used, on the other hand, in chapter 3, Langevin dynamics was used. These two methods basically reproduce similar results, but have some differences. One difference is about conservation of momentum. In the Newtonian dynamics, momentum conserves. Thus, unless solvent molecules are considered explicitly, we cannot describe diffusional motions of molecular motors. In chapter 3, considering this point, we use Langevin dynamics instead of Newtonian dynamics. In the following sections, we describe algorithms of Newtonian dynamics and Langevin dynamics we used and refer to unit system we used.

Newtonian dynamics with Berendsen thermostat

In order to solve the Newton equation of motion, velocity Verlet algorithm has been widely used (1),

$$\vec{r}(t+h) = \vec{r}(t) + h\vec{v}(t) + \frac{1}{2}h^2\vec{a}(t)$$

$$\vec{v}(t+h) = \vec{v}(t) + \frac{1}{2}h[\vec{a}(t) + \vec{a}(t+h)]$$

In this algorithm, the next position $\vec{r}(t+h)$ is calculated from the present position $\vec{r}(t)$, velocity $\vec{v}(t)$, and acceleration $\vec{a}(t)$. The acceleration $\vec{a}(t)$ is calculated from force of the position $\vec{r}(t)$. Then, the next velocity is calculated from the present velocity $\vec{v}(t)$, acceleration $\vec{a}(t)$, and the next acceleration $\vec{a}(t+h)$.

To control temperature during integrations, we used the method by Berendsen *et al*. (2) in which the velocity is scaled at every time step so that temperature $T(t) = 2/(3Nk_B)\sum_i m_i/2\,|\vec{v_i}|^2$ becomes T^0,

$$\vec{v_i} \leftarrow \left[1 + \frac{h}{\tau}\left(\frac{T^0}{T(t)} - 1\right)\right]^{1/2}\vec{v_i}$$

τ is the constant that determines the strength of coupling to the heat bath. If $h/\tau = 1$, it becomes a simple form, $\vec{v_i} \leftarrow [T^0/T(t)]^{1/2}\vec{v_i}$.

Langevin dynamics

In Langevin dynamics (3), we add damping term and Gaussian random force term to the

Newton equation of motion. Thus, the equation of motion becomes,

$m\vec{a} = \overrightarrow{F_c} - m\gamma\vec{v} + \vec{\Gamma}$

where $\overrightarrow{F_c}$ is conformational force, γ is friction coefficient, and $\vec{\Gamma}$ is random force. We again use velocity Verlet algorithm to solve the equation. We assign the equation of motion above to velocity Verlet equations and get,

$$\vec{r}(t+h) = \vec{r}(t) + h\left(1 - \frac{h\gamma}{2}\right)\vec{v}(t) + \frac{h^2}{2m}\left[\overrightarrow{F_c}(t) + \vec{\Gamma}(t)\right]$$

$$\vec{v}(t+h) = \left(1 - \frac{h\gamma}{2}\right)\left[1 - \frac{h\gamma}{2} + \left(\frac{h\gamma}{2}\right)^2\right]\vec{v}(t) + \frac{h}{2m}\left[1 - \frac{h\gamma}{2} + \left(\frac{h\gamma}{2}\right)^2\right]\left[\overrightarrow{F_c}(t) + \overrightarrow{F_c}(t+h) + \vec{\Gamma}(t) + \vec{\Gamma}(t+h)\right]$$

Radom force $\vec{\Gamma}(t)$ should satisfy,

$\langle\vec{\Gamma}\rangle = 0$

$\langle\vec{\Gamma}(t)\cdot\vec{\Gamma}(t')\rangle = 2m\gamma k_B T\delta(t-t')$

Since the equation of motion is discretized and solved numerically, the latter condition becomes,

$$\langle\vec{\Gamma}(t)\cdot\vec{\Gamma}(t+nh)\rangle = \frac{2m\gamma k_B T}{h}\delta_{0,n}$$

where $\delta_{0,n}$ is Kronecker delta, n is integer number. Thus, in the Langevin dynamics, temperature is related to the standard variance in the Gaussian distribution of the random force.

Unit system

When we solve the equation of motion such as $m\frac{d^2r}{dt^2} = -\frac{\partial E}{\partial r}$, we need certain units for length, time, weight, energy. We used $\text{Å} = 10^{-10}\text{m}$ for length unit, set mass of each bead (amino acid) to 10.0, which means unit for weight is $\approx 10\text{g/mol}$ because average mass for an amino acid is $\approx 100\text{g/mol}$. Units for time and energy corresponding to real system become unclear because of coarse-graining and uncertainty of interaction parameter sets. Energy unit was introduced through Boltzmann constant, $k_B = N_A k_B = 1.98 \times 10^{-3}\text{kcal/(mol}\cdot\text{K)}$ (we consider the value for 1mol just because it is useful). Now, we can derive time unit τ assigning other values to the Newton equation of motion above, $\tau \approx 150\text{fs}$. But, we note that this value is only correct when the interaction parameters are appropriately set so that temperature in simulation corresponds to Kelvin. Ideally, we should set interaction parameters to reproduce folding temperature correctly for each protein, but practically, it is difficult. Thus, we usually scale energy by thermal energy, $k_B T$. In this way, we do not have to care about certain units.

References

1. Allen M, Tildesley, DJ (1987) *Computer Simulation of Liquids* (Oxford University Press, Oxford).
2. Berendsen H, Postma, JPM, van Gunsteren, WF, DiNola, A, Haak, JR (1984) Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81:3684-3690.
3. Veitshans T, Klimov, D, Thirumalai, D (1996) Protein folding kinetics: timescales, pathways and energy landscapes in terms of sequence-dependent properties. *Folding & Design* 2:1-22.

Appendix 2: Eigenvalue equation for combining energy surfaces

In chapter 1, we constructed multiple-basin model in which two (or more) Gō potentials are combined into multiple-basin potential by eigenvalue characteristic equation. This way is analogous to what has often been used to describe the quantum mechanical coupling of two (or more) potential energy surfaces. We derive the eigenvalue equation starting from Schrödinger equation here.

Schrödinger equation is an eigenvalue equation, $\mathcal{H}|\Phi\rangle = \varepsilon|\Phi\rangle$, where $\mathcal{H}$ is Hamiltonian, ε is energy, and $|\Phi\rangle$ is wave function. We approximate wave function by trial function, $|\tilde{\Phi}\rangle = \sum_{i=1}^{N} c_i|\psi_i\rangle$, where $\{|\psi_i\rangle\}$ is a set of basis function. We can get minimum energy and optimal set of coefficient $\{c_i\}$ by variation method. This problem can be reduced to diagonalization of matrix representation of Hamiltonian $(H)_{ij} = H_{ij} = \langle\psi_i|\mathcal{H}|\psi_j\rangle$ (1). This matrix representation of Hamiltonian, H, corresponds to the matrix we make when combining two (or more) potentials. That is, diagonal element $(H)_{ii} = \langle\psi_i|\mathcal{H}|\psi_i\rangle$ corresponds to potential of state i.

We show a derivation to obtain the eigenvalue equation of matrix representation of Hamiltonian, H. We assume that a set of basis function satisfies $\langle\psi_i|\psi_j\rangle = \delta_{ij}$, and trial function is normalized, $\langle\tilde{\Phi}|\tilde{\Phi}\rangle = \left(\sum_{i=1}^{N} c_i\langle\psi_i|\right)\left(\sum_{j=1}^{N}|\psi_j\rangle c_j\right) = \sum_{i,j} c_i c_j\langle\psi_i|\psi_j\rangle = \sum_i c_i^2 = 1$. We minimize energy, that is, expectation value of Hamiltonian $\langle\tilde{\Phi}|\mathcal{H}|\tilde{\Phi}\rangle$ under condition of $\langle\tilde{\Phi}|\tilde{\Phi}\rangle = 1$. Apparently, we can use Lagrange's method of undetermined multipliers,

$$\mathcal{L}(c_1,\cdots,c_N,\lambda) = \langle\tilde{\Phi}|\mathcal{H}|\tilde{\Phi}\rangle - \lambda\left(\langle\tilde{\Phi}|\tilde{\Phi}\rangle - 1\right) = \sum_{i,j} c_i c_j H_{ij} - \lambda\left(\sum_k c_k^2 - 1\right)$$

To obtain minimum value, we differentiate $\mathcal{L}$ partially by c_ℓ ($\ell = 1,\cdots,N$)

$$\frac{\partial\mathcal{L}}{\partial c_\ell} = \sum_j c_j H_{\ell j} + \sum_i c_i H_{i\ell} - 2\lambda c_\ell = 0$$

Because Hamiltonian is Hermite operator, H becomes a symmetric matrix $H_{\ell j} = H_{i\ell}$,

$$\frac{\partial\mathcal{L}}{\partial c_\ell} = 2\sum_i c_i H_{i\ell} - 2\lambda c_\ell = 0$$

Therefore, we get

$$H\vec{c} = \lambda\vec{c}, \qquad \vec{c} = \begin{pmatrix} c_1 \\ \vdots \\ c_N \end{pmatrix}$$

This eigenvalue equation is the one we want, but meaning of eigenvalue λ is still unclear. If we diagonalize H, we obtain N number of eigenvector $\vec{c}^{(\alpha)}$ $(\alpha = 1,\cdots,N)$ and corresponding eigenvalue $\lambda^{(\alpha)}$. Because H is symmetric matrix, eigenvectors should satisfy $\left(\vec{c}^{(\alpha)}\right)^\dagger\vec{c}^{(\beta)} = \delta_{\alpha\beta}$. We prepare two wave functions corresponding to two eigenvectors we get,

$|\tilde{\Phi}^{(\alpha)}\rangle = \sum_{i=1}^{N} c_i^{(\alpha)} |\psi_i\rangle,\ |\tilde{\Phi}^{(\beta)}\rangle = \sum_{j=1}^{N} c_j^{(\beta)} |\psi_j\rangle$

Then, we calculate,

$$\langle \tilde{\Phi}^{(\alpha)} | \mathcal{H} | \tilde{\Phi}^{(\beta)} \rangle = \left(\sum_i c_i^{(\alpha)} \langle \psi_i | \right) \mathcal{H} \left(\sum_j |\psi_j\rangle c_j^{(\beta)} \right) = \sum_i c_i^{(\alpha)} \left(\sum_j H_{ij}\, c_j^{(\beta)} \right) = \left(\vec{c}^{(\alpha)} \right)^{\dagger} H\, \vec{c}^{(\beta)} =$$

$$\lambda^{(\beta)} \left(\vec{c}^{(\alpha)} \right)^{\dagger} \vec{c}^{(\beta)} = \lambda^{(\beta)} \delta_{\alpha\beta}$$

This equation shows that eigenvalue $\lambda^{(\alpha)}$ is expectation value of Hamiltonian by $|\tilde{\Phi}^{(\alpha)}\rangle$, thus, eigenvalue $\lambda^{(\alpha)}$ means energy of $|\tilde{\Phi}^{(\alpha)}\rangle$. Therefore, we get,

$$H\vec{c} = \varepsilon\vec{c}, \qquad \vec{c} = \begin{pmatrix} c_1 \\ \vdots \\ c_N \end{pmatrix}$$

If we consider N=2 case, the eigenvalue equation becomes,

$$\begin{pmatrix} \langle \psi_1 | \mathcal{H} | \psi_1 \rangle & \langle \psi_1 | \mathcal{H} | \psi_2 \rangle \\ \langle \psi_2 | \mathcal{H} | \psi_1 \rangle & \langle \psi_2 | \mathcal{H} | \psi_2 \rangle \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \varepsilon \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$$

When we use this equation to combine two Gō potentials, we replace $\langle \psi_1 | \mathcal{H} | \psi_1 \rangle \rightarrow V(R|R_1)$, $\langle \psi_2 | \mathcal{H} | \psi_2 \rangle \rightarrow V(R|R_2)$, $\langle \psi_1 | \mathcal{H} | \psi_2 \rangle = \langle \psi_2 | \mathcal{H} | \psi_1 \rangle \rightarrow \Delta$.

Reference

1. Szabo A, Ostlund, NS (1982) *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory* (Macmillan Publishing Co., Inc.)

Acknowledgements

First of all, I thank my advisor, Prof. Shoji Takada. He provided me a lot of novel, insightful, and practical ideas throughout the studies. I learned from him many important things as a researcher, for example, what kinds of question are important and interesting in this field, experience of international communication, and so on. I also thank members of Takada lab., especially, Dr. Nobuyasu Koga for providing me a program “CaFold” from which I developed my method, and more importantly, teaching me attitude as a researcher from everyday conversations, Dr. Sung-Joon Park for teaching me computer skills and knowledge, and encouraging me to join CASP7 (Critical Assessment of Techniques for Protein Structure Prediction) contest in 2006, Dr. Rie-Tatsumi Koga and Naoto Hori for discussing research and scientific things at voluntary seminars, Dr. Ryo Kanada for discussing application of methods to molecular motors, Dr. Hiroo Kenzaki for discussing source codes of CaFold and detailed methods, Dr. Wenfei Li and Dr. Xin-Qiu Yao for discussing detailed methods and conversations in English. I am very proud of graduating from Takada Lab. and starting my career as a researcher. I also thank some former members of Takada lab. for fruitful advises at seminars or meetings.

I also thank my official advisor for the last one and half year, Prof. Motonari Tsubaki for kindly accepting me after the laboratory moved to Kyoto university and I studied there practically. I thank Naoki Yamamoto and Ikuo Kurisaki for discussions at voluntary seminars and some suggestions as a beginning researcher or friend. I have been financially supported by JSPS (Japan Society for the Promotion of Science) research fellowships for young scientists from Apr. 2007 to Mar. 2009. Lastly, I want to thank my parents for various kinds of support.

Kei-ichi Okazaki

2009/1/8 Kyoto, Japan