Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Structural Biology

High anisotropy and frustration: the keys to regulating protein function efficiently in crowded environments

Akio Kitao and Kazuhiro Takemura

Highly anisotropic protein dynamics in equilibrium can be observed experimentally or through structural bioinformatics and molecular simulations. This anisotropic nature causes a response, to an external perturbation, along a small number of intrinsic large-amplitude directions as expected from the fluctuation–dissipation theorem. It is also key for controlling specific reactions as stochastic processes in macromolecular crowded environments. Protein anisotropy can be exploited for the calculation of physical properties, such as entropy, which can be employed for binding affinity studies. Energy frustration along soft modes including both global large-amplitude and localized small-amplitude movements is another key feature, as conformational transitions along soft modes, triggered by external perturbations such as the binding of other molecules, can act as a switch to control function.

Address
Institute of Molecular and Cellular Biosciences, The University of Tokyo, 1-1-1 Yayoi, Bunkyo, Tokyo 113-0032, Japan

Corresponding author: Kitao, Akio (kitao@iam.u-tokyo.ac.jp)

Current Opinion in Structural Biology 2017, **42**:50–58

This review comes from a themed issue on **Proteins: bridging theory and experiment**

Edited by **Igor N Berezovsky** and **Ugo Bastolla**

For a complete overview see the Issue and the Editorial

Available online 2nd November 2016

http://dx.doi.org/10.1016/j.sbi.2016.10.014

Introduction

What properties have native proteins acquired during evolution? How can native proteins conduct a specific regulated function in macromolecular crowded environments? Classically, this ability is attributed to the structure of the protein in the native state. Proteins in the native state typically adopt compact structures compared to that in the denatured state [1]. The packing density of the interior of a native protein is high and uniform if surface water molecules are included [2]. The radii of gyration of native proteins are almost proportional to (molecular weight)$^{1/3}$, the power law of a sphere [3]. Consistently, protein radius of gyration normalized by the radius of gyration of a sphere with the same volume is independent of the size of the protein [4]. On the other hand, artificial proteins with random amino acid sequences tend to be larger in size and do not form stable secondary structure [3,5], suggesting that compactness and higher-order structures are properties of native proteins acquired through molecular evolution.

The highly anisotropic nature of protein dynamics

A compact protein structure in the native state is closely related to the highly anisotropic nature of protein dynamics which utilizes compact and relatively rigid structural elements (such as domains), or flexible elements exposed to solvent (such as loops and linkers) as moving units. Systematic analysis of protein structure variations in crystal structures has revealed typical protein conformational changes. For example, pioneering work by Gerstein *et al.* described protein domain movements as hinge and shear motions [6,7]. Recently, Taylor *et al.* classified domain movements into five types: free, open-closed, anchored, sliding-twist and see-saw [8•] and applied to the assignment of hinge and shear movements, showing that a relative translation of domains is rare and that there is no difference between hinge and shear mechanisms [9••]. Significant domain movements are observed in many proteins. The analysis of a comprehensive and non-redundant dataset of structures differing by more than 0.5 Å indicated that more than half of the proteins in the dataset exhibit domain motions [10]. Proteins also conduct conformational transitions with smaller amplitude. Analysis of an equivalent database showed that main-chain dihedral angle transitions occur in 82% of the proteins [11]. Many of these dihedral angle transitions are responsible for global and local hinge motions and the flap motion of loops, but 24.3% of the transitions are involved in so-called 'path-preserving' motions, in which the localized collective dihedral transitions occur to preserve the main-chain path and which correlate with functional events such as ion bindings. It is difficult to detect this type of motion by analyzing atomic fluctuations because the amplitude of the fluctuations is very small. Therefore, the analysis of dihedral angles is also important. The high anisotropy of protein dynamics is also observed in structure ensembles determined by solution NMR. The conformational differences observed between the solution NMR structures and their crystal structure counterparts are consistent with the collective motion identified by principal component analysis (PCA) and the anisotropic network model (ANM) [12].

The high anisotropy of protein dynamics in equilibrium has been well characterized by collective coordinate sets determined by normal mode analysis (NMA), PCA and multidimensional scaling (MDS) [13–19]. In PCA, the collective coordinates are introduced from a variance-covariance matrix of a given coordinate systems (typically Cartesian coordinates of atoms) as:

$$\mathbf{A} = \langle \Delta\mathbf{q}\Delta\mathbf{q}^t \rangle \quad (1)$$

where $\Delta\mathbf{q}$ represents the column vector of the displacement of coordinates from the average, and $\langle \cdots \rangle$ shows the ensemble average. The superscript 't' indicates the matrix transpose. The axes of the collective coordinates in PCA (principal axes) are determined as the eigenvectors of $\mathbf{A}$:

$$\mathbf{AV} = \mathbf{V}\lambda \quad (2)$$

with the orthonormalized condition,

$$\mathbf{V}\mathbf{V}^t = \mathbf{V}^t\mathbf{V} = \mathbf{I} \quad (3)$$

where $\mathbf{V}$ and $\boldsymbol{\lambda}$ are the matrices of the eigenvectors and eigenvalues, and $\mathbf{I}$ is a unit matrix. The ith column vector of $\mathbf{V}$, $\mathbf{v}_i$, indicates the ith principal axis. Since the ith diagonal element of $\boldsymbol{\lambda}$, λ_i, is the variance of the ith principal component, its contribution to the total variance,

$$s_i = \frac{\lambda_i}{tr\lambda}, \quad (4)$$

shows the anisotropy of the system. If s_i is much larger than the others, the component is considered a 'soft mode' because a larger fluctuation occurs compared to other components. Proteins intrinsically have a small number of large-amplitude modes. For example [20], s_1 is equal to 0.35 (35%) among 1002 internal degrees of freedom and the accumulated value for the first ten components is 0.89 (89%) in the case of C^{α}-atom PCA of a 20 ns molecular dynamics (MD) simulation of FlhAc protein (Figure 1a). Consistent with this, the s_1 and accumulated values for the first ten components were 32 and 81% in a recent unpublished 1 μs MD simulation. These values are typical for proteins. Another good measure to understand the anisotropy is 'anharmonicity factor', which is defined as the root-mean-square-fluctuation along a PC axis divided by that expected from normal mode along the same axis [21]. It should be also noted that the anhamonicity factor also reflects the effect of multiple minima, which will be discussed later. This factor is typically more than two for large-amplitude modes [15] and can be more than 10 for the largest-amplitude PC mode [22]. The dominancy of a small number of collective degrees of freedom clearly indicates the high anisotropic nature of protein dynamics. The important concept here is that a subset spanned by a small number of collective coordinates is robust, and thus useful for investigating both simulation data and experimental data [13–19].

A recent trend is the consideration of time dependence in the analysis of MD simulations [23,24,25•,26]. Time-structure based independent component analysis (tICA) determines statistically independent components from a time-lagged covariance matrix [23,25•,27], and these independent components were also applied to build Markov state model (MSM) [28,29]. In ICA, all the modes are conceptually uncoupled. However, to understand the mechanisms of protein function, a more important goal is the investigation of the relationship between trigger and response. Independent subspace analysis (ISA) determines a set of subspaces as follows: The collective variables in each subspace are significantly correlated and correlation between the variables from distinct subspaces is insignificant [30]. Interestingly, only five subspaces were identified and all other collective variables are independent in T4 lysozyme. Cross correlation function analysis of the modes in the same subspace quantified the time delay and advance among the correlated modes, and showed that only small number of movements can have the relationship of trigger and response. The largest block consists of six modes and clearly showed the propagation of movements from a global motion mode to a local mode, and then on to other global modes. ISA is useful for identifying a series of correlated events including domain and local motions.

Fluctuation–dissipation theorem and protein function

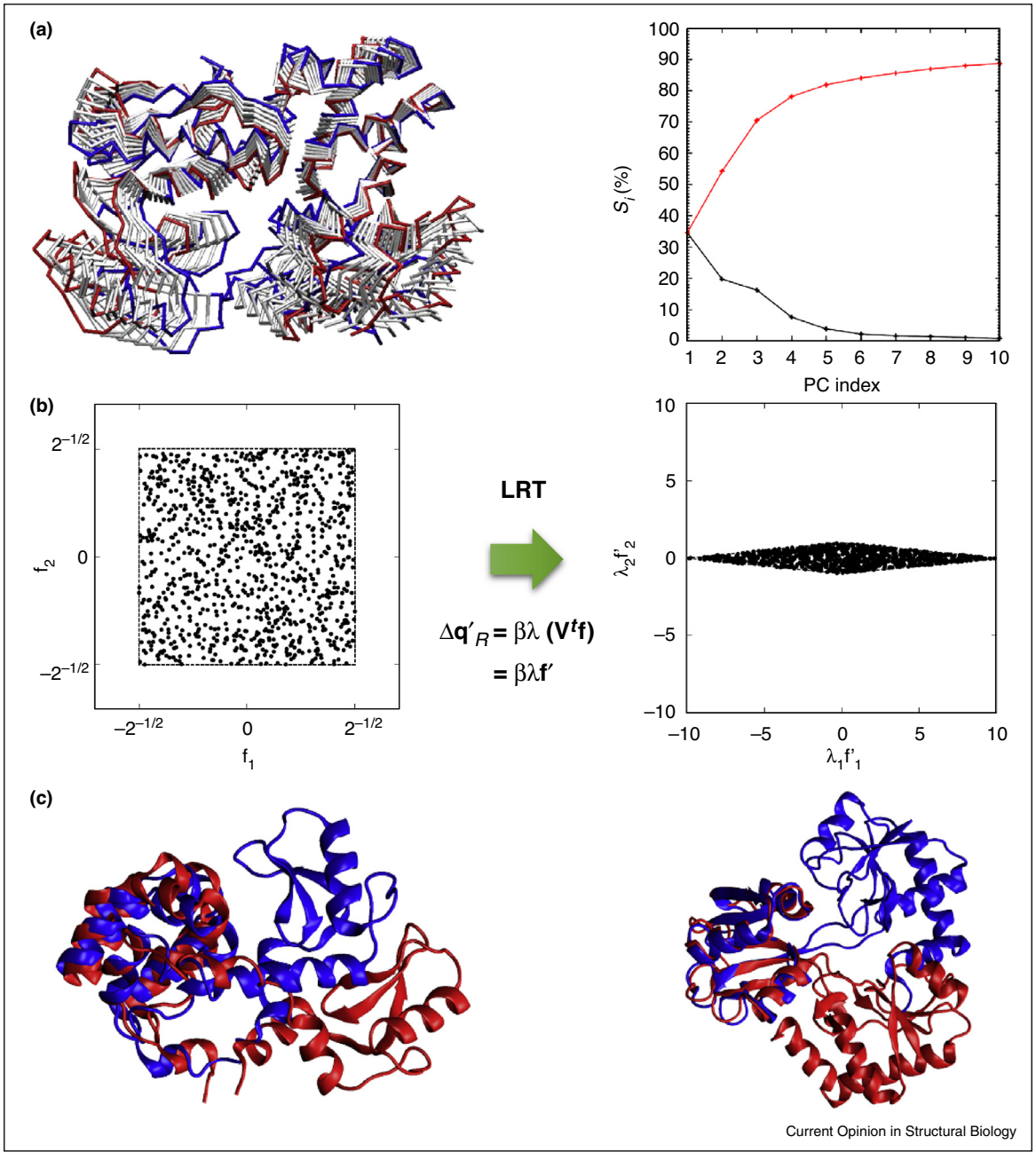
The high anisotropy of protein dynamics in equilibrium is closely correlated with specific protein response to a weak external perturbation as predicted by the fluctuation–dissipation theorem [31]. This statement is clearer if the response $\Delta\mathbf{q}_R$ to the perturbation force $\mathbf{f}$ is described by the time-independent linear response theory (ti-LRT):

$$\Delta\mathbf{q}_R = \beta\mathbf{A}\mathbf{f} \quad (5)$$

where β is the thermodynamic beta. The concept of LRT is applied to investigate protein dynamics [32,33,34•,35]. Ikeguchi *et al*. clearly demonstrated that ti-LRT explains and predicts structural changes in some proteins upon ligand binding [32]. In that work, they determined $\mathbf{A}$ from MD simulations of unliganded protein and reproduced the response of the liganded protein induced by $\mathbf{f}$ mimicking the protein–ligand interaction. Recently Yang and coworkers used ti-LRT and time-dependent (td-) LRT to investigate the response of myoglobin upon CO binding and showed agreement of the time response between LRT, ultraviolet resonance Raman spectroscopy, and time-resolved X-ray crystallography, suggesting that the primary response can be described by LRT [34•].

If the response is observed in the principal component space, we obtain

Figure 1



The highly anisotropic nature of protein dynamics and the fluctuation–dissipation theorem. **(a)** An example of (left) the largest amplitude motion determined by principal component analysis (PCA) shown by representative structures along the axis and (right) the contribution s_i (black line) and the accumulated contribution (red line) as a function of the principal component index determined by MD of bacterial flagellar protein FlhAc [20]. **(b)** (left) Random perturbations and (right) induced response in two-dimensional space for the case $\lambda_1/\lambda_2 = 10$. **(c)** Simulated domain movements of (left) T4 lysozyme and (right) glutamine binding protein. Both cases utilized random perturbation forces in the transform and relax sampling (TRS) without assuming the fluctuation–dissipation theorem [33]. Two extreme snapshots for open (red) and close (blue) structures are shown. *Source*: The plots in (b) are reprinted from [33], with the permission of AIP Publishing.

$$\Delta\mathbf{q}'_R = \mathbf{V}^t\Delta\mathbf{q}_R = \beta\boldsymbol{\lambda}(\mathbf{V}^t\mathbf{f}) \\ = \beta\boldsymbol{\lambda}\mathbf{f}' \quad , \tag{6}$$

by projection onto the principal axes. If $\mathbf{f}$ is applied as isotropic random perturbations, the perturbation along the principal axes $\mathbf{f}'$ is also isotropic; however, the responses $\Delta\mathbf{q}'_R$ are highly anisotropic because they are scaled by λ_i (Figure 1b) [33]. Eq. (6) indicates that

random perturbations cause highly anisotropic responses because proteins fluctuate in a highly anisotropic manner in equilibrium.

Interestingly, the highly anisotropic response of proteins can be simulated without assuming the validity of the fluctuation–dissipation theorem and LRT. In transform and relax sampling (TRS), we applied random external perturbations to proteins and demonstrated that the domain motions of multi-domain proteins (Figure 1c) and the folding of a 'mini-protein' chigolin are highly enhanced [33]. The results of the aforementioned analysis between NMR structures and their counterparts obtained by X-ray crystallography [12] can be also understood by appreciating that proteins in a crystal are perturbed more by their surroundings compared to those in solution and thus shift along intrinsically soft directions. Although proteins operate in highly 'noisy' environments perturbed by many other biomolecules, they tend to respond largely along soft directions under random perturbations as a stochastic process, depending on the intrinsic nature of the protein in equilibrium.

Calculation of protein entropy from anisotropy and its relation to binding free energy estimation

The entropy of a protein is an essential measure to quantify the contribution of protein dynamics; however, the calculation of entropy from molecular simulation is not straightforward and counting all the significant microstates within a limited simulation time is a challenging problem. The entropy of a molecule *in vacuo* can be decomposed into the contributions from translational, rotational and internal motions,

$$S = S^{trans} + S^{rot} + S^{\mathrm{int}}. \tag{7}$$

The first two terms are rigorously given as:

$$S^{trans} = k_B \ln\left\{\frac{(2\pi M k_B T)^{3/2}}{h^3} e^{5/2} V\right\} \tag{8}$$

$$S^{rot} = k_B \ln\left\{\frac{(2\pi k_B T)^{3/2}(I_X I_Y I_Z)^{1/2}}{h^3} e^{3/2} 8\pi^2\right\} \tag{9}$$

where k_B, h, T, M and V are Boltzmann constant, Planck constant, absolute temperature, mass of the solute and volume (1 l/mol), respectively. I_X, I_Y and I_Z represent the principal moments of inertia. S^{int} is calculated by considering the anisotropy of the target molecule. In normal mode vibrations, S^{int} can be given as:

$$S^{\mathrm{int}} = \sum_i k_B \left\{\frac{1}{\alpha_i(\exp\alpha_i^{-1} - 1)} - \ln(1 - \exp(-\alpha_i^{-1}))\right\} \tag{10}$$

where $\alpha_i = 2\pi k_B T/\hbar\omega_i$ and ω_i is the angular frequency of each normal mode. Eq. (10) shows a semi-quantum mechanical treatment to avoid non-physical negative entropy. MD results allow estimation of ω_i from the quasi-harmonic (QH) relation $\omega_i = \sqrt{k_B T/\lambda_i}$ if the unit of mass is ignored. The QH approach by Karplus and Kushick [36] and the empirical formula by Schlitter [37] are widely used to calculate S^{int}. These methods can give a reasonable value, which is considered as the upper bound [19]. Beyond the QH approaches, more sophisticated methods were proposed. For example, entropy calculation methods using ICA [38] and full correlation analysis (FCA) [39] also take advantage of considering anharmonic distributions along collective modes. Hensen *et al.* recently proposed a method to calculate entropy from force [40,41]. Other entropy calculation methods beyond the QH are reviewed in detail by Kassen *et al.* [42•].

Entropy calculations are useful to investigate protein binding affinity. For instance, we can consider a thermodynamic cycle involving the binding free energy of two proteins A and B in ΔG_s solution as shown in Figure 2. ΔG_s can be directly calculated from dissociation simulations but it tends to be overestimated in protein–protein complexes [43]. ΔG_s is also calculated through the detour shown by the yellow arrow in Figure 2:

$$\Delta G_s = \Delta G_v + \Delta\Delta\mu \tag{11}$$

where ΔG_v is the binding energy *in vacuo* and $\Delta\Delta\mu$ is the difference in the solvation free energies of the complex $\Delta\mu_{AB}$ and the monomers $\Delta\mu_A$ and $\Delta\mu_B$,

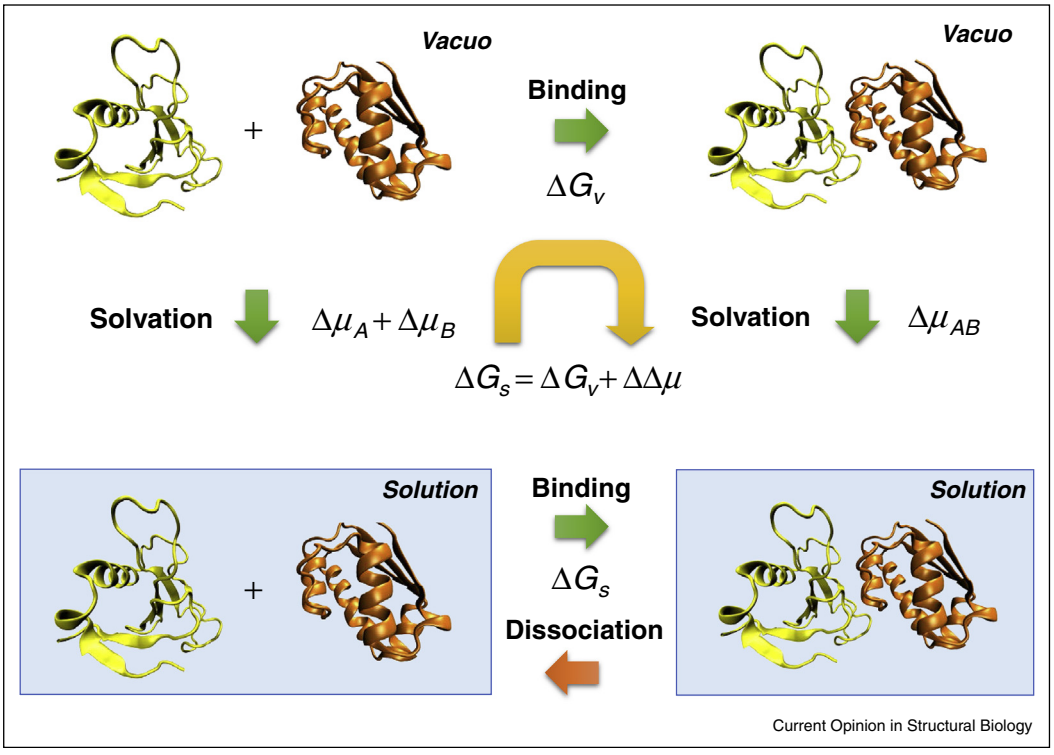
$$\Delta G_v = \Delta E - T\Delta S \tag{12}$$

$$\Delta\Delta\mu = \Delta\mu_{AB} - (\Delta\mu_A + \Delta\mu_B) \tag{13}$$

where ΔE and ΔS are obtained as differences in the average potential energies and entropies *in vacuo* as $\Delta E = E_{AB} - (E_A + E_B)$ and $\Delta S = S_{AB} - (S_A + S_B)$. It remains debatable if consideration of the entropy significantly improves the estimation of binding free energy [44], however, the inclusion of entropy is shown to improve the correlation with experimental results in protein–ligand binding [45,46]. In principle, contribution of protein entropy should be considered for more accurate calculation of binding free energy.

The all-atom calculation of solvation free energy with free energy perturbation and thermodynamic integration is straightforward for small molecules but very difficult for large molecules. Solution theory in the energy representation recently enabled the calculation of solvation free energy using relatively short MD simulation in the energy representation module (ERmod) [47••]. Free energy calculation of lysozyme–triNAG complex using this scheme showed that the method distinguishes the most plausible binding mode as the lowest binding energy mode [48]. The same

Figure 2



Thermodynamics cycle of protein–protein binding and binding free energy calculation. The green arrows from left to right indicate binding of two proteins A and B. The vertical arrows show solvation of A, B and complex AB from gas phase. The yellow arrow represents the cycle to calculate binding free energy in solution.

procedure was also examined for protein–protein complexes without the entropy correction and was found to be useful for selecting low energy complex models similar to the crystal structure from a set of generated models [49]. The entropy correction in binding free energy calculation with all-atom calculation has not been well examined yet, but we will show in the near future that refinement of the method with entropy corrections further improves the free energy calculation. The alternative method to calculate solvation free energy with explicit solvent model is 3D-RISM (reference interaction site model) in which distribution functions are obtained from the integral equations without conducting MD simulation [50,51].

It should be also noted that protein anisotropic motion also correlated with hydration structure and vice versa. Therefore, the protein motion should also affect solvation free energy. Recently, hydration structure changes were shown to be necessary for the domain motion of glutamate dehydrogenase [52••], which implies that the solvation free energy depends on large-amplitude conformational change and that consideration of the protein anisotropy is also essential for free energy calculation.

Inconsistency and frustration in the native state of proteins

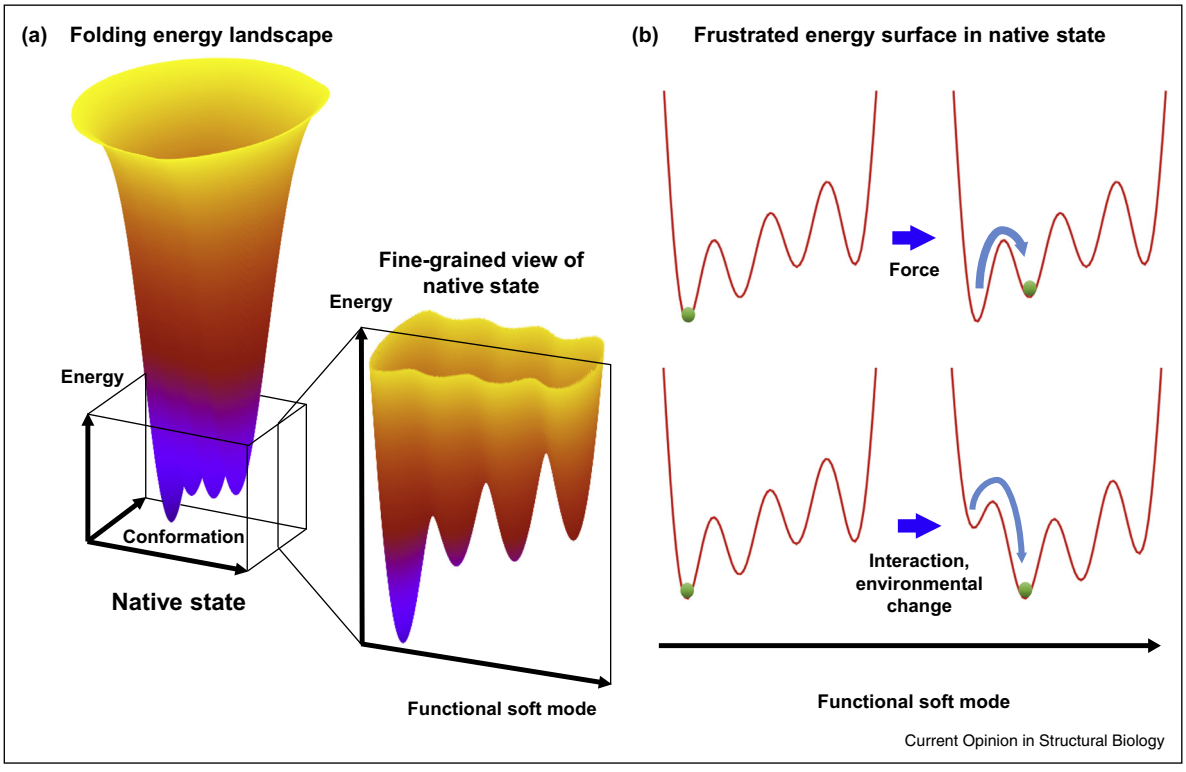
The highly anisotropic nature of native proteins assures a tendency to enhance specific protein responses in crowded environments, but it is not sufficient to regulate function. It would be problematic if protein functional responses were induced only by random perturbations in the absence of specific control mechanisms. Therefore, proteins are equipped with mechanisms that trigger the switch between functional and nonfunctional modes. For example, the catalytic domain of horse liver alcohol dehydrogenase (LADH) tends to be open in the unliganded state because domain closure movement is blocked by a proline-rich loop; however, the loop structure is altered upon the binding of the coenzyme NAD^+ and the domain closure occurred within 10 ns [53]. The unbound state of LADH is primed for binding NAD^+ and NAD^+ binding triggers domain closure and enzymatic function. Ligand-triggered collective motion is also observed in glutamine binding protein [54]. Once triggered, this type of motion is expected to occur promptly because of the high anisotropy of the protein and because of the mechanisms predicted by the fluctuation–dissipation theorem. Recently dehydration of a cleft in glutamate

dehydrogenase was shown to induce rearrangements of hydrogen bond networks and acts as a switch for domain motion [52••]. Considering these results, to understand such mechanisms, the protein energy landscape and its change upon triggering should be investigated.

A rugged energy landscape is a common feature of multicomponent systems. The multiple-minima feature of the energy surface can be interpreted as being due to inconsistency and frustration. Energy frustration arises from inconsistencies in multiple interactions. These concepts were introduced in early protein folding studies: the various energy terms responsible for folding are consistent [55,56] and thus the native state is minimally frustrated [57]. In contrast to folding studies, we focus on the native state and take a finer-grained view in both spatial and energy scales and thus inconsistencies and frustration are more evident (Figure 3a). The Jumping-among-minima (JAM) model provides a concept for investigating the distribution of energy minima (which can be considered as conformational substates or microstates) in the high dimensional space of proteins [15,22]. If a proper set of collective coordinates is selected, the energy surface along the most of the collective axes (e.g., 95%) is essentially harmonic with a single-minimum feature. The analysis of MD result by JAM model also showed that the energy surface along a small number of large-amplitude principal axes has a hierarchical multiple-minima feature and rugged smaller-amplitude motions on the nearly parabolic energy surface are more localized [15,22]. This means that soft anisotropic modes of proteins are frustrated and external perturbations can trigger the protein to switch from one state to another along the intrinsic collective directions (Figure 3b). Furthermore, given the aforementioned path-preserving motion [11], the concept of soft modes should not be limited to the large-amplitude modes but should be further extended to localized frustrated degrees of freedom.

Since the inconsistencies are involved with the balance among various interactions, their origins are often not obvious. One exception is bacterial flagellar filament [58]. The flagellar filament of *Salmonella typhimurium* is composed of a single protein that undergoes polymorphic

Figure 3



A schematic view of energy frustration on protein free energy surface and its role on the regulation of function. **(a)** Folding energy landscape viewed at (left) a relatively coarse-grained resolution in spatial and energy scales and (right) a fine-grained view of the energy landscape more focused at higher spatial and energy resolutions. **(b)** Conformational change along a functional soft mode induced by external perturbation. Top panel shows the case in which a force is applied from outside and induced a transition from the global energy minimum to another minimum. The effect of the force is not reflected in the surface and thus the energy surface does not change. Lower panel shows the case where a change in interaction or environment is reflected on the energy surface, which also changes in depths of energy minima and thus causes conformational change from one minimum to the other.

transitions between left-handed and right-handed supercoils induced by the reversal of motor rotation, pH, ionic strength, and mutations. The existence of the multiple supercoil structures is well understood by the spontaneous coexistence of two states in the polymer [59]. The inconsistency between the intrasubunit interactions preferring one state and the intersubunit interactions stabilizing polymerization to another state results in energy frustration [58]. To quantify localized energy frustration for a given protein structure, a protein frustratometer has been proposed to examine energy frustration [60,61,62•]. Using the frustratometer, Gianni *et al.* investigated the frustrated interactions in frataxin, iron binding protein that involves with the assembly of iron–sulfur cluster, and showed that the frustrated regions are correlated to binding sites of metals and ferrochelatase [63••].

Changes in the protein energy landscape during binding events are conceptually well investigated. The paradigm shifts from traditional 'lock-and-key' to 'induced-fit' and 'conformational selection and population shift' now provide more detailed understanding of molecular recognition in proteins, as reviewed recently [64–69]. The schematic free energy profiles appearing in these reviews should be understood as being equivalent to representative soft modes in a multidimensional space.

Conclusions

The highly anisotropic nature of protein dynamics induces a protein response to external perturbations along a small number of intrinsic large-amplitude directions as expected from the fluctuation–dissipation theorem. This anisotropy is key for inducing specific protein reactions as stochastic processes in macromolecular crowded environments. We also showed that protein anisotropy is useful for calculating the entropy and these calculations can be used for binding affinity studies. Energy frustration occurs along large-amplitude atomic motions as well as collective dihedral transitions such as path-preserving motions. Here we called the movements along these frustrated degrees of freedom 'soft modes'. Conformational transition along the soft modes can act as a switch to regulate protein function and can be triggered by external perturbations such as the binding of other molecules.

Conflict of interest statement

Nothing declared.

Acknowledgements

This research was supported by Grants-in-Aid for Scientific Research in Innovative Areas (No. 25104002) and Grants-in-Aid for Scientific Research B (No. 15H04357) from the Japan Society for The Promotion of Science (JSPS) and Ministry of Education, Culture, Sports, Science and Technology of Japan (MEXT) to AK, and by Innovative Drug Discovery Infrastructure through Functional Control of Biomolecular Systems, Priority Issue 1 in Post-K Supercomputer Development (Project ID: hp150270), to AK from MEXT. The computations were partly performed using the supercomputers at RCCS, The National Institute of Natural Science and ISSP, The University of Tokyo. This research also used computational resources of the K computer provided by the RIKEN Advanced Institute for Computational Science through the HPCI System Research project (Project ID: hp120223 and hp140030).

References and recommended reading

Papers of particular interest, published within the period of review,

have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Wilkins DK, Grimshaw SB, Receveur V, Dobson CM, Jones JA, Smith LJ: **Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques**. *Biochemistry* 1999, **38**:16424-16431.

2. Tsai J, Taylor R, Chothia C, Gerstein M: **The packing density in proteins: standard radii and volumes**. *J Mol Biol* 1999, **290**:253-266.

3. Yamauchi A, Yomo T, Tanaka F, Prijambada ID, Ohhashi S, Yamamoto K, Shima Y, Ogasahara K, Yutani K, Kataoka M *et al.*: **Characterization of soluble artificial proteins with random sequences**. *FEBS Lett* 1998, **421**:147-151.

4. Lobanov MY, Bogatyreva NS, Galzitskaya OV: **Radius of gyration as an indicator of protein structure compactness**. *Mol Biol* 2008, **42**:623-628.

5. Tanaka J, Doi N, Takashima H, Yanagawa H: **Comparative characterization of random-sequence proteins consisting of 5, 12, and 20 kinds of amino acids**. *Protein Sci* 2010, **19**:786-795.

6. Gerstein M, Lesk AM, Chothia C: **Structural mechanisms for domain movements in proteins**. *Biochemistry* 1994, **33**:6739-6749.

7. Gerstein M, Krebs W: **A database of macromolecular motions**. *Nucleic Acids Res* 1998, **26**:4280-4290.

8. • Taylor D, Cawley G, Hayward S: **Classification of domain movements in proteins using dynamic contact graphs**. *PLoS One* 2013, **8**:e81224.
Systematic classification of domain movements of two-domain protein structures in the Protein Data Bank (PDB) using so-called 'dynamics contact graphs (DCG)'. Visualization of domain movement with DCG is shown to be useful for analyzing the detailed mechanisms of domain movements.

9. •• Taylor D, Cawley G, Hayward S: **Quantitative method for the assignment of hinge and shear mechanism in protein domain movements**. *Bioinformatics* 2014, **30**:3189-3196.
The application of DCG for the assignment of hinge and shear movements. The authors showed that a relative translation of domains is rare and proposes to call 'shear movement' as 'interface-preserving movement', and 'hinge' as 'interface-creating movement'.

10. Qi G, Lee R, Hayward S: **A comprehensive and non-redundant database of protein domain movements**. *Bioinformatics* 2005, **21**:2832-2838.

11. Nishima W, Qi G, Hayward S, Kitao A: **DTA: dihedral transition analysis for characterization of the effects of large main-chain dihedral changes in proteins**. *Bioinformatics* 2009, **25**:628-635.

12. Yang LW, Eyal E, Bahar I, Kitao A: **Principal component analysis of native ensembles of biomolecular structures (PCA_NEST): insights into functional dynamics**. *Bioinformatics* 2009, **25**:606-614.

13. Case DA: **Normal-mode analysis of protein dynamics**. *Curr Opin Struc Biol* 1994, **4**:285-290.

14. Hayward S, Go N: **Collective variable description of native protein dynamics**. *Annu Rev Phys Chem* 1995, **46**:223-250.

15. Kitao A, Go N: **Investigating protein dynamics in collective coordinate space**. *Curr Opin Struct Biol* 1999, **9**:164-169.

16. Berendsen HJC, Hayward S: **Collective protein dynamics in relation to function**. *Curr Opin Struct Biol* 2000, **10**:165-169.

17. Bahar I, Rader AJ: **Coarse-grained normal mode analysis in structural biology**. *Curr Opin Struct Biol* 2005, **15**:586-592.

18. Ma J: **Usefulness and limitations of normal mode analysis in modeling dynamics of biomolecular complexes**. *Structure* 2005, **13**:373-380.

19. Cui Q, Bahar I (Eds): *Normal Mode Analysis: Theory and Applications to Biological and Chemical Systems*. Chapman & Hall/CRC; 2006.

20. Saijo-Hamano Y, Imada K, Minamino T, Kihara M, Shimada M, Kitao A, Namba K: **Structure of the cytoplasmic domain of FlhA and implication for flagellar type III protein export**. *Mol Microbiol* 2010, **76**:260-268.

21. Hayward S, Kitao A, Go N: **Harmonicity and anharmonicity in protein dynamics — a normal-mode analysis and principal component analysis**. *Proteins-Struct Funct Genet* 1995, **23**:177-186.

22. Kitao A, Hayward S, Go N: **Energy landscape of a native protein: jumping-among-minima model**. *Proteins-Struct Funct Genet* 1998, **33**:496-517.

23. Naritomi Y, Fuchigami S: **Slow dynamics in protein fluctuations revealed by time-structure based independent component analysis: the case of domain motions**. *J Chem Phys* 2011, **134**:065101.

24. Mitsutake A, Iijima H, Takano H: **Relaxation mode analysis of a peptide system: comparison with principal component analysis**. *J Chem Phys* 2011, **135**:164102.

25. • Naritomi Y, Fuchigami S: **Slow dynamics of a protein backbone in molecular dynamics simulation revealed by time-structure based independent component analysis**. *J Chem Phys* 2013, **139**:215102.

The application of tICA to 1 μs molecular dynamics trajectory of lysine-, arginine-, ornithine-binding protein (LAO). tICA was successful in characterizing both 'slow time scale' global and local motions as the first five independent components.

26. Mori T, Saito S: **Dynamic heterogeneity in the folding/unfolding transitions of FiP35**. *J Chem Phys* 2015, **142**:135101.

27. Molgedey L, Schuster HG: **Separation of a mixture of independent signals using time-delayed correlations**. *Phys Rev Lett* 1994, **72**:3634-3637.

28. Schwantes CR, Pande VS: **Improvements in Markov state model construction reveal many non-native interactions in the folding of NTL9**. *J Chem Theory Comput* 2013, **9**:2000-2009.

29. Perez-Hernandez G, Paul F, Giorgino T, De Fabritiis G, Noe F: **Identification of slow molecular order parameters for Markov model construction**. *J Chem Phys* 2013, **139**:015102.

30. Sakuraba S, Joti Y, Kitao A: **Detecting coupled collective motions in protein by independent subspace analysis**. *J Chem Phys* 2010, **133**:185102.

31. Kubo R: **The fluctuation–dissipation theorem**. *Rep Prog Phys* 1966, **29**:255-284.

32. Ikeguchi M, Ueno J, Sato M, Kidera A: **Protein structural change upon ligand binding: linear response theory**. *Phys Rev Lett* 2005, **94**:078102.

33. Kitao A: **Transform and relax sampling for highly anisotropic systems: application to protein domain motion and folding**. *J Chem Phys* 2011, **135**:045101.

34. • Yang LW, Kitao A, Huang BC, Go N: **Ligand-induced protein responses and mechanical signal propagation described by linear response theories**. *Biophys J* 2014, **107**:1415-1425.

Study of the response of myoglobin upon CO binding using ti-LRT and td-LRT. The observed relaxation behavior is compared to the results obtained using UV resonance Raman spectroscopy and time resolved X-ray crystallography. Agreement between the computational and experimental approaches implies that the LRT assumption can reproduce the primary response of myoglobin.

35. Tamura K, Hayashi S: **Linear response path following: a molecular dynamics method to simulate global conformational changes of protein upon ligand binding**. *J Chem Theory Comput* 2015, **11**:2900-2917.

36. Karplus M, Kushick JN: **Method for estimating the configurational entropy of macromolecules**. *Macromolecules* 1981, **14**:325-332.

37. Schlitter J: **Estimation of absolute and relative entropies of macromolecules using the covariance-matrix**. *Chem Phys Lett* 1993, **215**:617-621.

38. Nguyen PH: **Estimating configurational entropy of complex molecules: a novel variable transformation approach**. *Chem Phys Lett* 2009, **468**:90-93.

39. Hensen U, Lange OF, Grubmuller H: **Estimating absolute configurational entropies of macromolecules: the minimally coupled subspace approach**. *PLoS One* 2010, **5**:e9179.

40. Hensen U, Grater F, Henchman RH: **Macromolecular entropy can be accurately computed from force**. *J Chem Theory Comput* 2014, **10**:4777-4781.

41. Louet M, Seifert C, Hensen U, Grater F: **Dynamic allostery of the catabolite activator protein revealed by interatomic forces**. *PLoS Comput Biol* 2015, **11**:e1004358.

42. • Kassem S, Ahmed M, El-Sheikh S, Barakat KH: **Entropy in bimolecular simulations: a comprehensive review of atomic fluctuations-based methods**. *J Mol Graph Model* 2015, **62**:105-117.

Good review of entropy calculation methods from molecular simulations. This review is suitable for readers aiming to understand representative methods: normal mode analysis (NMA), quasi-harmonics (QH), mining minima, nearest neighbor (NN), mutual information expansion (MIE), maximum information spanning trees (MIST), and hypothetical scanning molecular dynamics (HSMD).

43. Yamashita T, Fujitani H: **On accurate calculation of the potential of mean force between antigen and antibody: a case of the HyHEL-10-hen egg white lysozyme system**. *Chem Phys Lett* 2014, **609**:50-53.

44. Genheden S, Ryde U: **The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities**. *Expert Opin Drug Discov* 2015, **10**:449-461.

45. Oehme DP, Brownlee RT, Wilson DJ: **Effect of atomic charge, solvation, entropy, and ligand protonation state on MM-PB(GB)SA binding energies of HIV protease**. *J Comput Chem* 2012, **33**:2566-2580.

46. Genheden S, Kuhn O, Mikulskis P, Hoffmann D, Ryde U: **The normal-mode entropy in the MM/GBSA method: effect of system truncation, buffer region, and dielectric constant**. *J Chem Inf Model* 2012, **52**:2079-2088.

47. •• Sakuraba S, Matubayasi N: **Ermod: fast and versatile computation software for solvation free energy with approximate theory of solutions**. *J Comput Chem* 2014, **35**:1592-1608.

This paper introduces a highly efficient solvation free energy calculation method by the energy representation module (ERmod). ERmod calculates the solvation free energy of solute molecules including large molecules such as proteins. Solvation into mixed solvent, for example, solvent and other solutes and solvent and membrane, can be easily handled. The calculation can be conducted from the relatively short trajectory of standard MD programs. The program is distributed by SourceForge.

48. Takemura K, Burri RR, Ishikawa T, Ishikura T, Sakuraba S, Matubayasi N, Kuwata K, Kitao A: **Free-energy analysis of lysozyme-triNAG binding modes with all-atom molecular dynamics simulation combined with the solution theory in the energy representation**. *Chem Phys Lett* 2013, **559**:94-98.

49. Takemura K, Guo H, Sakuraba S, Matubayasi N, Kitao A: **Evaluation of protein-protein docking model structures using all-atom molecular dynamics simulations combined with the solution theory in the energy representation**. *J Chem Phys* 2012, **137**:215105.

50. Kovalenko A, Ten-No S, Hirata F: **Solution of three-dimensional reference interaction site model and hypernetted chain equations for simple point charge water by modified method of direct inversion in iterative subspace**. *J Comput Chem* 1999, **20**:928-936.

51. Kovalenko A, Hirata F: **Potential of mean force between two molecular ions in a polar molecular solvent: a study by the**

three-dimensional reference interaction site model. *J Phys Chem B* 1999, **103**:7942-7957.

52. •• Oroguchi T, Nakasako M: **Changes in hydration structure are necessary for collective motions of a multi-domain protein**. *Sci Rep* 2016, **6**:26302.
In this paper, the authors investigated the correlation between local hydration structure change and collective motions using X-ray crystallography, atomic force microscopy and MD. They showed that 'drying' and 'wetting' of a hydrophobic pocket of glutamate dehydrogenase is amplified to domain motion.

53. Hayward S, Kitao A: **Molecular dynamics simulations of NAD+-induced domain closure in horse liver alcohol dehydrogenase**. *Biophys J* 2006, **91**:1823-1831.

54. Loeffler HH, Kitao A: **Collective dynamics of periplasmic glutamine binding protein upon domain closure**. *Biophys J* 2009, **97**:2541-2549.

55. Go N: **Theoretical studies of protein folding**. *Annu Rev Biophys Bioeng* 1983, **12**:183-210.

56. Go N: **The consistency principle in protein structure and pathways of folding**. *Adv Biophys* 1984, **18**:149-164.

57. Bryngelson JD, Wolynes PG: **Spin-glasses and the statistical-mechanics of protein folding**. *Proc Natl Acad Sci U S A* 1987, **84**:7524-7528.

58. Kitao A, Yonekura K, Maki-Yonekura S, Samatey FA, Imada K, Namba K, Go N: **Switch interactions control energy frustration and multiple flagellar filament structures**. *Proc Natl Acad Sci U S A* 2006, **103**:4894-4899.

59. Asakura S: **Polymerization of flagellin and polymorphism of flagella**. *Adv Biophys* 1970, **1**:99-155.

60. Ferreiro DU, Hegler JA, Komives EA, Wolynes PG: **Localizing frustration in native proteins and protein assemblies**. *Proc Natl Acad Sci U S A* 2007, **104**:19819-19824.

61. Jenik M, Parra RG, Radusky LG, Turjanski A, Wolynes PG, Ferreiro DU: **Protein frustratometer: a tool to localize energetic frustration in protein molecules**. *Nucleic Acids Res* 2012, **40**:W348-W351.

62. • Parra RG, Schafer NP, Radusky LG, Tsai MY, Guzovsky AB, Wolynes PG, Ferreiro DU: **Protein Frustratometer 2: a tool to localize energetic frustration in protein molecules, now with electrostatics**. *Nucleic Acids Res* 2016, **44**:W356-W360.
This paper describes a unique webserver to quantify localized energy frustration for a given PDB structure using the AWSEM-MD Frustratometer. This version can include electrostatic interactions.

63. •• Gianni S, Camilloni C, Giri R, Toto A, Bonetti D, Morrone A, Sormanni P, Brunori M, Vendruscolo M: **Understanding the frustration arising from the competition between function, misfolding, and aggregation in a globular protein**. *Proc Natl Acad Sci U S A* 2014, **111**:14141-14146.
The authors investigated the competition between function, misfolding and aggregation of frataxin based on the concept of frustration, and showed that the regions involved in misfolding is optimized to avoid aggregation.

64. Csermely P, Palotai R, Nussinov R: **Induced fit, conformational selection and independent dynamic segments: an extended view of binding events**. *Trends Biochem Sci* 2010, **35**:539-546.

65. Nussinov R, Ma B, Tsai CJ: **Multiple conformational selection and induced fit events take place in allosteric propagation**. *Biophys Chem* 2014, **186**:22-30.

66. Hatzakis NS: **Single molecule insights on conformational selection and induced fit mechanism**. *Biophys Chem* 2014, **186**:46-54.

67. Clore GM: **Interplay between conformational selection and induced fit in multidomain protein–ligand binding probed by paramagnetic relaxation enhancement**. *Biophys Chem* 2014, **186**:3-12.

68. Vogt AD, Pozzi N, Chen Z, Di Cera E: **Essential role of conformational selection in ligand binding**. *Biophys Chem* 2014, **186**:13-21.

69. Weikl TR, Paul F: **Conformational selection in protein binding and function**. *Protein Sci* 2014, **23**:1508-1518.