Kobe University Repository : Kernel

PDF issue: 2025-12-05

Derivation of extremely slow dynamics of protein motion based on entropy invariance

Tanaka, Shigenori

(Citation)
Physics Letters A,372(8):1280-1282

(Issue Date)
2008-02

(Resource Type)
journal article

(Version)
Accepted Manuscript

(URL)
https://hdl.handle.net/20.500.14094/90000874

KOBE UNIVERSITY

Derivation of extremely slow dynamics of protein motion based on entropy invariance

Shigenori Tanaka [a,b,*]

[a] *Graduate School of Human Development and Environment, Kobe University, 3-11 Tsurukabuto, Nada, Kobe 657-8501, Japan*

[b] *CREST, Japan Science and Technology Agency, 4-1-8 Honcho, Kawaguchi, Saitama 332-0012, Japan*

Abstract

It is a mysterious fact that protein systems often show an extremely slow dynamics of their molecular motions with time scales much longer than nanosecond order, although their characteristic frequencies obtained by the normal mode analysis fall in much shorter temporal regions. This Letter provides a heuristic account for why and how such extremely slow modes of protein motions naturally emerge from fast molecular modes on the basis of an idea of entropy invariance in the principal component analysis.

PACS: 87.10.+e; 87.14.Ee; 87.15.He
Keywords: Protein; Conformational change; Slow dynamics; Principal component analysis; Entropy

It is well known [1–3] that a lot of dynamical modes of molecular motions coexist in proteins whose characteristic time scales vary widely from femtosecond to second regions. Among them, slow modes with time scales longer than nanoseconds often play essential roles for the functions of proteins. However, the origin of such slow dynamics has not yet been elucidated well. In this Letter I provide a heuristic account for why and how such extremely slow modes in protein system emerge from fast molecular modes on the bases of the principal component analysis (PCA) or essential dynamics (ED) description [1–5] and the entropy invariance principle.

Let us consider the coordinates $q = \{q_i\}$, the momenta $p = \{p_i\}$ ($i = 1, 2, \ldots, N$) and Hamiltonian $H(p, q)$ characterizing the dynamics of a pro-

* Corresponding author. Fax: +81-78-803-7761
Email address: tanaka2@kobe-u.ac.jp (Shigenori Tanaka).

tein motion. In the following we consider a classical-mechanical system and employ a partition function in the canonical ensemble as

$$Z = \frac{1}{(2\pi\hbar)^N} \iint \cdots \int \exp\left[-\frac{H(p,q)}{k_B T}\right] \prod_{i=1}^{N} dp_i dq_i, \tag{1}$$

where T, k_B and $\hbar$ are the temperature, the Boltzmann constant and the Planck constant.

It is assumed that we have obtained the optimized structure of the system expressed by the coordinate $q = \bar{q}$, where the minimized potential energy of the system is given by $U(\bar{q})$. Then, performing the normal mode analysis around the optimized structure, the Hamiltonian is approximately given by

$$H = U(\bar{q}) + \sum_{i=1}^{N} \left(\frac{p_i^2}{2m_i} + \frac{1}{2} m_i \omega_i^2 x_i^2\right), \tag{2}$$

where m_i and ω_i are the mass and the normal frequency of the mode i; $x = \{x_i = q_i - \bar{q}_i\}$ is the displaced coordinate around $\bar{q}$. The normal mode i is thus characterized by the set of coordinates $\{x_i\}$ and momenta $\{p_i\}$. Substituting Eq. (2) into Eq. (1), we obtain

$$Z = \exp\left[-\frac{U(\bar{q})}{k_B T}\right] \prod_{i=1}^{N} \left(\frac{k_B T}{\hbar\omega_i}\right). \tag{3}$$

The free energy F and the entropy S are then expressed as

$$F = -k_B T \ln Z = U(\bar{q}) - k_B T \sum_{i=1}^{N} \ln\left(\frac{k_B T}{\hbar\omega_i}\right), \tag{4}$$

$$S = -\left(\frac{\partial F}{\partial T}\right)_V = k_B \sum_{i=1}^{N} \left[\ln\left(\frac{k_B T}{\hbar\omega_i}\right) + 1\right]. \tag{5}$$

We can thus evaluate the entropy of the system by calculating the normal-mode frequencies $\{\omega_i\}$.

When we consider the system at $T = 300$ K, we find $k_B T = 209$ cm^{-1}. Usually, the normal-mode frequencies of a protein widely range from more than 3000 cm^{-1} of high-frequency stretching vibrational modes to less than 1 cm^{-1} of collective modes. Therefore, the natural logarithms found in Eq. (5) take values ranging from about -3 to 5 and greater, thus contributing to the

entropy extensively.

The methods of principal component analysis (PCA) and essential dynamics (ED) are well known as effective tools for eliciting long-time motions of proteins through molecular dynamics simulations [1–5]. Let us consider mass-weighted displacements of time-dependent atomic coordinates $r_i(t)$ ($i = 1, 2, \ldots, N$) as

$$R_i(t) = \sqrt{m_i}\,[r_i(t) - \langle r_i \rangle]\,, \tag{6}$$

where m_i and $\langle r_i \rangle$ refer to the atomic mass and the average of atomic coordinates, respectively. We here perform an orthogonal coordinate transformation as

$$R_i(t) = \sum_{j=1}^{N} A_{ij} Q_j(t) \tag{7}$$

with

$$\sum_{j=1}^{N} A_{ij}^2 = 1 \tag{8}$$

so that

$$\langle Q_i(t) Q_j(t) \rangle = \lambda_i \delta_{ij}, \tag{9}$$

where $\langle\ \rangle$ means the statistical average. We have thus obtained a diagonal matrix,

$$\Lambda = (\lambda_{ij}) = (\lambda_i \delta_{ij}) = \left(\langle Q_i^2 \rangle \delta_{ij} \right), \tag{10}$$

which represents the variance of the normal coordinates $\{Q_i\}$. An effective frequency,

$$\Omega_i = \sqrt{\frac{k_B T}{\langle Q_i^2 \rangle}} = \sqrt{\frac{k_B T}{\lambda_i}} \tag{11}$$

is then associated with each normal coordinate Q_i.

The PCA and ED can extract " essential " slow modes of protein motion by arranging the modes above in the descending order of the variance λ_i.

By keeping only a small number of modes with small Ω_i, we can reproduce the essential parts of large-scale slow modes of protein motions, which are often associated with their important functions [1–3]. This methodology is also employed for a very effective compression of trajectory data obtained by molecular dynamics simulations [4,5].

We here remark that the entropy of protein system can be expressed as

$$S = k_B \sum_{i=1}^{N} \left[\ln \left(\frac{k_B T}{\hbar \Omega_i} \right) + 1 \right], \tag{12}$$

using the frequencies $\{\Omega_i\}$ given by Eq. (11) as well. In contrast to Eq. (5), the expression (12) takes into account a nonlinear effect of protein motions through molecular dynamics simulations, while the functional form of Eq. (12) has arisen from the harmonic approximation (quasi-harmonic description). The key idea here is that the values of Ω_i corresponding to the larger values of λ_i are usually much smaller than all the ω_i's, reflecting the capture of slow modes. For example, Ω_i^{-1} corresponding to the largest value of λ_i often takes a value of nanosecond or longer range, describing a very slow motion compared with the slowest dynamics captured by the usual normal-mode analysis.

We then assume that the two ways of expressions for the entropy, Eqs. (5) and (12), should give the same value for the identical system (entropy invariance ansatz). Recalling that the smallest Ω_i often falls in the range of GHz order or lower (note that 10^{12} Hz corresponds to 33.36 cm^{-1}), the leading logarithmic terms in Eq. (12) would take values much larger than those in Eq. (5). However, the summations in the two expressions would give the same value via the coalescence of modes and the rearrangement of the distribution of frequencies. Actually, it is known that a limited number of slow modes (say, $5 - 10$ % [1,5]) dominate and govern the protein dynamics in the case of PCA or ED description [1–3], which is the consequence of the data compression or the information reduction [4,5] addressed above.

Considering the discussions above, we could here provide a heuristic account for why and how the extremely slow modes emerge in protein systems. The essential idea of PCA or ED is that a very few number of extremely slow modes are produced through a combination of a number of vibrational modes described by the usual normal mode analysis. We here provide a simple mathematical model to express this relation so that the entropy given by n normal modes with the frequency ω is equal to that of an extremely slow mode with

the frequency Ω. Recalling Eqs. (5) and (12), we thus set

$$n\left[\ln\left(\frac{k_BT}{\hbar\omega}\right)+1\right]=\ln\left(\frac{k_BT}{\hbar\Omega}\right)+1. \tag{13}$$

After simple algebra this then leads to

$$\Omega=\omega\left(\frac{\hbar\omega}{ek_BT}\right)^{n-1}. \tag{14}$$

If we express the ratio between Ω and ω as

$$\frac{\Omega}{\omega}=10^{-p}, \tag{15}$$

we finally find

$$p=(n-1)\log\left(\frac{ek_BT}{\hbar\omega}\right). \tag{16}$$

Here, we will evaluate the " slow-down " index p according to a typical case, as suggested by actual molecular dynamics simulations [4,5]. To obtain a positive value of p, $n>1$ and $\hbar\omega<ek_BT$ are required. Let us consider a case in which $T=300$ K, $\hbar\omega=0.1\ k_BT$ and $n=10$ (see, e.g., Fig. 1 in Ref. [5]). We then find $p=12.9$. Thus, in contrast to the fast modes with the ω^{-1} falling in the range of picosecond order, the emergent slow mode is delayed by about thirteen digits, whose Ω^{-1} falls in the temporal range of seconds. This simple model thus gives an account for the mechanism producing an extremely slow mode from molecular vibrational modes.

The model addressed above relies on a couple of assumptions and approximations. One of the most significant assumptions is concerned with the validity of the quasi-harmonic approximation. It is known [1–3] that protein motions often show a diffusive behavior in nanosecond molecular dynamics simulations. It therefore seems that such a strongly anharmonic behavior cannot be described in terms of the (quasi)harmonic approximation. However, this seemingly diffusive mode may demonstrate a recursive (oscillatory) behavior when one would perform a much longer simulation, which could be described in the framework of the quasi-harmonic approximation. Good performances of PCA and ED methodologies in nanosecond molecular dynamics simulations concerning the effective data compression, which were confirmed through the comparison with sub-microsecond simulations [4,5], suggest a justification for

this viewpoint.

Acknowledgements

The author would like to thank Dr. Tatsuya Nakano for useful discussions. This work was partly supported by the CREST project of Japan Science and Technology Agency (JST).

References

[1] A. Amadei, A.B.M. Linssen, H.J.C. Berendsen, Proteins 17 (1993) 412.

[2] S. Hayward, N. Go, Annu. Rev. Phys. Chem. 46 (1995) 223.

[3] A. Kitao, N. Go, Curr. Opin. Struct. Biol. 9 (1999) 164.

[4] T. Meyer, C. Ferrer-Costa, A. Pérez, M. Rueda, A. Bidon-Chanal, F.J. Luque, C.A. Laughton, M. Orozco, J. Chem. Theory Comput. 2 (2006) 251.

[5] O. Lange, H. Grubmüller, J. Phys. Chem. B 110 (2006) 22842.