Regular Article

Residue-based quadratic free energy relationship is a mathematical formulation of the consistency principle of protein folding

Daisuke Kohda [1], Seiichiro Hayashi [1], Daisuke Fujinami [1,2]

[1] *Division of Structural Biology, Medical Institute of Bioregulation, Kyushu University, Higashi-ku, Fukuoka 812-8582, Japan*

[2] *Graduate School of Integrated Pharmaceutical and Nutritional Sciences, University of Shizuoka, Suruga-ku, Shizuoka 422-8526, Japan*

ABSTRACT

The consistency principle represents a physicochemical condition requisite for ideal protein folding. The consistency principle assumes that any pair of amino acid residues in partially folded structures has an attractive short-range interaction *only if* the two residues are in contact in the final native structure. This native-centric model is supported by theoretical and simulation studies, but no experimental evidence has been presented. NMR spectroscopy can determine the residue-specific equilibrium constant, K, and the residue-specific rate constant, k (forward and backward). Linear free energy relationships (LFER) in the rate equilibrium free energy relationship (REFER, log k vs. log K) plots are widely seen in protein-related phenomena. Here, we examined the theoretical basis of the residue-based LFER. First, we derived a basic expression, $\rho_{ij} = ½(\phi_i + \phi_j)$, from the consistency principle, where ρ_{ij} is the slope of the line segment that connects residues *i* and *j* in the REFER plot, and ϕ_i and ϕ_j are the local fractions of the native structure in the transient state ensemble (TSE). Next, we showed that the general solution is the alignment of the (log K, log k) data points on a parabolic curve in the REFER plot. Importantly, unlike LFER, the quadratic free energy relationship (QFER) is compatible with the precedent formation of local structures in TSE. QFER provides a new insight: A foldable polypeptide chain consists of several folding units, which are loosely but *consistently* coupled to undergo smooth structural changes. QFER also enables a new ϕ-value analysis without the need for mutations.

Key words: ϕ-value analysis, LFER, QFER, transition state ensemble, two-state exchange

Significance

The physicochemical basis of smooth protein folding has been theoretically explained by the consistency principle of protein folding. Here we propose that the consistency principle is experimentally formulated by the quadratic relationship in the double logarithm plot of residue-specific equilibrium and rate constants of a polypeptide chain. One application is a new-type ϕ-value analysis, free from the adverse effects of mutations. This study will trigger the development of innovative experimental techniques that enable the measurement of accurate residue-specific thermodynamic and kinetic parameters of foldable polypeptides.

Corresponding author: Daisuke Kohda, Division of Structural Biology, Medical Institute of Bioregulation, Kyushu University, Maidashi 3-1-1, Higashi-ku, Fukuoka 812-8582, Japan. ORCID iD: https://orcid.org/0000-0001-8234-3776, e-mail: kohda@bioreg.kyushu-u.ac.jp

Introduction

Protein molecules fold into their thermodynamically stable and biologically active structures[1]. This special property is acquired by evolution and optimized for each amino acid sequence. Particularly, small globular proteins fold quickly, typically in less than milliseconds, from unstructured states to a compact native structure. The entire process of protein folding is frequently a two-state transition without passing through any metastable intermediate states[2,3]. By increasing temporal and spatial resolutions, folding intermediates can be detected[4,5]. Nevertheless, the coarse-graining of the folding process is useful for understanding the physicochemical basis of quick and smooth protein folding.

Protein folding can be described as an ensemble of the downslope motions of partially folded structures on the funnel-like energy landscape (Fig. 1A)[6–8]. The vertical axis represents the stability (free energy G^*, which excludes the contributions from conformational entropy) of a partially folded structure, and the horizontal plane represents the conformational diversity (entropy S^* originating from conformational entropy) of the ensemble. If the funnel-like energy landscape has a perfectly smooth surface (i.e., the range of microscopic roughness, $\delta E = 0$), then it materializes the maximally rapid folding of a polypeptide chain from any unfolded state. This situation is ideal. By contrast, real proteins have rugged energy surfaces. Through evolutionary selections, the rugged surface of the funnel-like landscape is minimized ($\delta E < 3k_BT$) by balancing between rapid folding rates and functionality requirements, such as in catalytic and allosteric proteins[9,10].

In the funnel view, proteins fold *via* multiple pathways. By contrast, the bimodal mass distributions obtained from hydrogen-deuterium exchange (HX) studies suggested that proteins fold along a defined pathway[11]. In the defined-pathway model, a protein molecule consists of smaller units, called 'foldons'[12,13]. A foldon is a cooperative native-like structural element, and a folded foldon serves as a template for the folding of the next foldon, in a hierarchical manner. The folding pathway is predetermined by the same cooperative interactions that determine the final native structure. A recent reanalysis using the maximum entropy principle suggested that the multiple pathway model could explain the results of the HX study of cytochrome c. However, the original experiment clearly showed the formation of one foldon ahead of the other foldons, indicating the predefined formation of a foldon, at least in an early step of cytochrome c folding[14]. The funnel view *versus* foldon view has been a matter of debate for decades[15,16]. Is it possible to resolve the conflict between the microscopic view (the smooth surface of a folding funnel) and the macroscopic view (the sequential formation of foldons)? Modeling of protein folding at a mesoscale is expected to be effective for this purpose[17]; for instance, at the residue level adopted in this study.

What is the physicochemical basis for the conditions for a perfectly smooth energy landscape? Nobuhiro Gō proposed the 'consistency principle' theory in 1983[18]. He used a lattice model for simulation, in which each residue of a 'protein' is a point on a two-dimensional lattice and all pairs of the residues in any partially folded structures have negative (i.e., preferable) short-range interaction energy *only if* the two residues are in contact within a native structure[19]. In other words, the consistency principle requests the absence of non-native structures or interactions at any moment on the folding path from any random initial structure. Gō and his colleagues found their simple lattice model can explain various physical properties of real proteins [18]. The 2D lattice model was subsequently extended to the cubic lattice model[20,21] and then to the coarse-grained Cα model[22]. These 3D version models had excellent predictive powers of folding transitions and pathways[23,24]. Although the absolute bias to the final folded state appears odd, it makes sense that the state realized by the consistency principle is equivalent to the ideal state of an actual gas. The ideal gas approximation is extremely useful as a reference for understanding real gases.

Bryngelson and Wolynes proposed a more realistic version of the protein folding principle in 1987, based on the theory of spin glasses[25]. Frustration occurs owing to the impossibility of satisfying all favorable energetic interactions simultaneously and results in a rough surface of the energy landscape[26]. To be kinetically foldable, the funnel must have a sufficiently steep slope so it can overcome local traps created by the frustration. The minimum frustration principle has the advantage of the mathematical expression: $T_g < T$, where T_g is the glass-transition temperature and T is the physiological temperature. Compared to the minimum frustration principle, the consistency principle has been criticized as narrative, due to a lack of mathematical expression[27].

NMR spectroscopy can probe the local environments of a polypeptide chain. This feature enables NMR to determine the residue-specific equilibrium constant, K, and residue-specific rate constants, k (forward) and k' (backward), of the two-state exchange of a polypeptide chain. Unexpectedly, we found that these observed constants varied from one residue to another in the two-state topological isomerization of a 27-residue peptide, nukacin ISK-1[28,29]. Such variations in the thermodynamic and kinetic parameters of a polypeptide chain have occasionally been reported over the years[30,31], but their significance has remained enigmatic. This is because the possibility of artifacts caused by systematic measurement biases or random measurement errors cannot be ruled out. Intriguingly, we found a linear relationship between the residue-specific log K values and the residue-specific log k values for nukacin ISK-1, even after eliminating biases and errors from the measurement and the analytical procedures for determining these parameters[32]. This linear relationship must

be more than a coincidence.

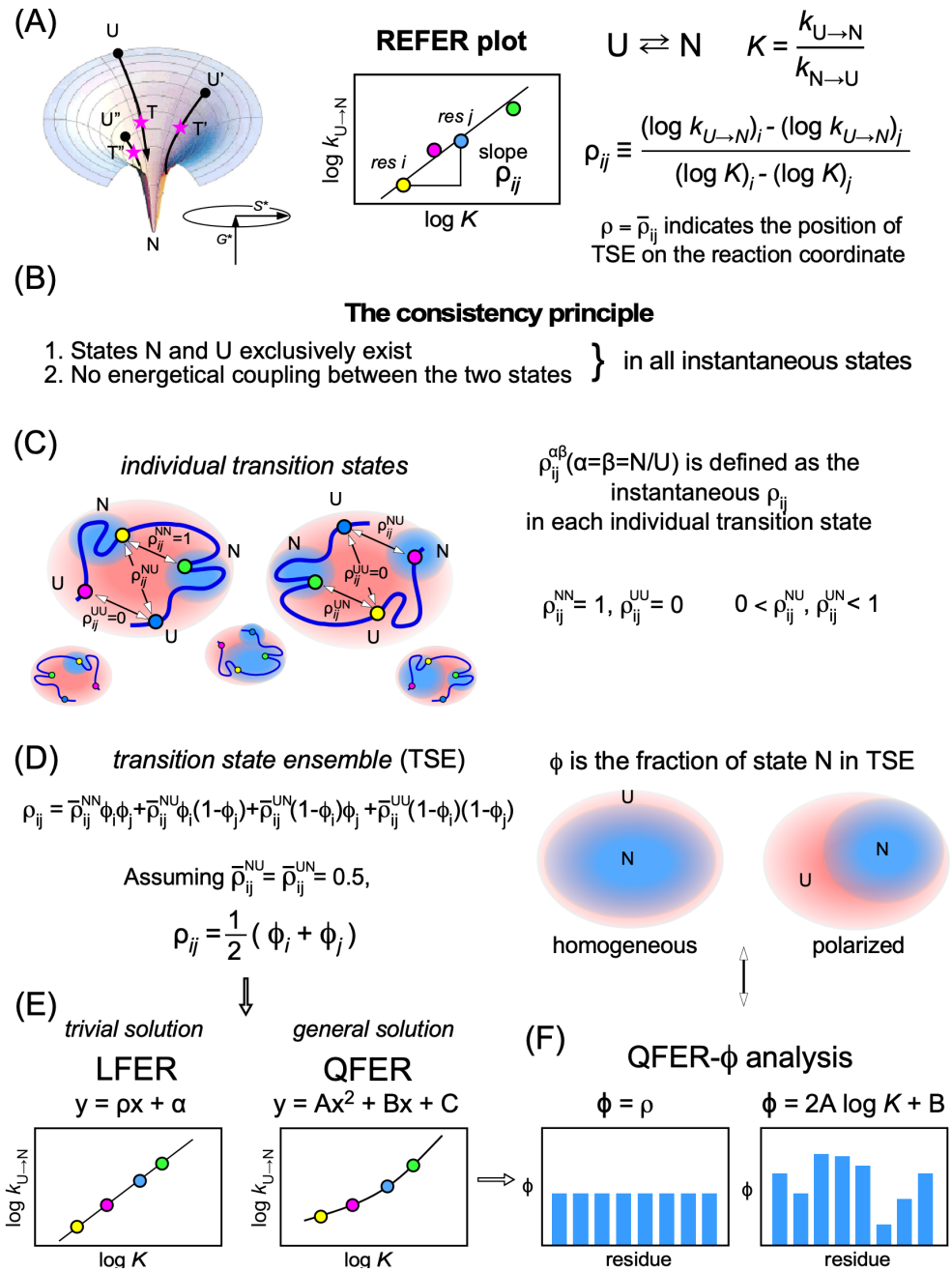

Figure 1 Outline of the transition-state analysis based on the consistency principle of protein folding. (A) Funnel-like energy landscape and definition of the slope ρ_{ij} between residue *i* and residue *j* in a rate-equilibrium free energy relationship (REFER) plot. (B) Two assumptions in the consistency principle of protein folding. (C) Individual transition states derived from a diverse set of initial structures. $\rho_{ij}^{\alpha\beta}$ ($\alpha = \beta = N/U$) is the instantaneous ρ_{ij} between residue *i* and residue *j* in a transition state and has four values, $\rho_{ij}^{NN} = 1$, $\rho_{ij}^{UU} = 0$, and $0 < \rho_{ij}^{NU}, \rho_{ij}^{UN} < 1$. (D) In the transition state ensemble (TSE), we assumed $\bar{\rho}_{ij}^{NU} = \bar{\rho}_{ij}^{UN} = 0.5$. Then, ρ_{ij} is the average of ϕ_i and ϕ_j, where the ϕ value is a fraction of the N(-like) state on a per-residue basis in TSE. (E) Homogeneous TSE is formulated by residue-based LFER (linear free energy relationship), and polarized TSE is formulated by residue-based QFER (quadratic free energy relationship). (F) An application of QFER. The QFER-ϕ analysis provides per-residue ϕ values without the need for mutations.

The double logarithmic relation of the two parameters, K and k, is generally called REFER (rate-equilibrium free energy relationship) (Fig. 1A), and the linear relationship in the REFER plot is referred to as LFER (linear free energy relationship), as the two axes, log K and log k, are proportional to the changes in the corresponding free energy terms. In particular, the LFER we found is referred to as residue-based LFER (rbLFER) because it is derived from amino acid residues in one polypeptide chain[29]. To test the general applicability of rbLEFR, we performed a retrospective study by collecting residue-specific equilibrium and rate constants in exchange processes, including protein conformational changes, the coupled folding and binding of intrinsically disordered polypeptides (IDP), and the structural fluctuations of folded globular proteins. The good linearity in a substantial number of the REFER plots indicated that the rbLFER holds in a wide variety of protein-related phenomena[33].

The triadic relationship, K = k/k' or log K = log k – log k', could generate an artificial linear relationship in the REFER plot if a careless experimenter generated a REFER plot with unignorable measurement biases and/or large measurement errors. Let the rbLFER be expressed in a mathematical form, log k = ρ log K + α. The second relationship in the REFER plot between log k' and log K is not independent, and has the form of log k' = (ρ-1) log K + α [29]. To identify a true rbLFER, the third correlation between log k and log k' must be observed[33]. In addition, the three correlations must have special properties to make rbLFER physically meaningful. The slope ρ represents the energetical similarity between the initial ground state and the transition state, and the position of the transition state on the reaction coordinate[34,35]. Thus, the slope ρ must be between 0 and 1. This condition is equivalent to a negative correlation in the log k vs. log k' plot[33].

As mentioned above, the consistency principle is a physical model of ideal foldable polypeptide chains. Although the consistency principle has been widely supported by molecular simulations, including the trajectory analysis of small proteins obtained by all-atom MD simulations with explicit water solvent[36], no experimental evidence has been presented. We suggested that the rbLFER was an experimental consequence of the consistency principle[29].

Here, we examined the relation between log K and log k under the consistency principle. We found that the alignment of data points on a parabolic curve in the REFER plot is the general solution (Fig. 1E). The mathematical expression, referred to as the quadratic free energy relationship (QFER), enabled us to devise a new method for analyzing structural changes of proteins (Fig. 1F). We propose that the QFER-ϕ analysis is an alternate and superior method to the mutational ϕ analysis, which has been widely used for many years[37–39].

Methods

Quadratic OLS (ordinary least squares) fitting was performed with the standard polynomial fitting in Excel, and quadratic TLS (total least squares) fitting was performed by solving cubic equations (Appendix I) with the add-in program, Solver, in Excel. The protein cartoons were generated with the program PyMOL, version 2.4.2 (Schrödinger).

Results

Arithmetic relation between ρ and ϕ based on the consistency principle

Let ρ_{ij} denote the slope of the line segment connecting two points corresponding to residues *i* and *j* in a REFER plot (Fig. 1A).

$$\rho_{ij} \equiv \frac{(\log k_{U\to N})_i - (\log k_{U\to N})_j}{(\log K)_i - (\log K)_j} \tag{1}$$

We reword the consistency principle of protein folding as follows: in any instantaneous structure on any folding path, (Rule 1) states that N and U exclusively exist and (Rule 2) affirms that there is no energetical coupling between the two states (Fig. 1B). We start by considering a particular transition state, *t*, derived from an arbitrary initial state (Fig. 1C). Based on the first rule, the transition state *t* contains residues in either state N or state U. $\rho_{ij}^{\alpha\beta}$ (α = β = N/U) is defined as the instantaneous ρ_{ij} between residues *i* and *j* in the transition state *t*. If residues *i* and *j* are both in state N, then $\rho_{ij}^{NN} = 1$, because the regions around the two residues in *t* are considered to be very close to state N. Here, recall that the slope ρ in the REFER plot represents the position of the transition state on the reaction coordinate from U to N. Similarly, if residues *i* and *j* are both in state U, then $\rho_{ij}^{UU} = 0$, because the regions around the two residues in *t* are considered to be very close to state U. In contrast, if residues *i* and *j* are in different states, then ρ_{ij}^{NU} and ρ_{ij}^{UN} are unknowable in each *t* state.

We consider the ensemble of transition states starting from all possible initial states (Fig. 1D). In the transition state ensemble (TSE), $\bar{\rho}_{ij}^{\alpha\beta}$ is defined as the ensemble average of the individual $\rho_{ij}^{\alpha\beta}$ values. Clearly, $\bar{\rho}_{ij}^{NN} = 1$ and $\bar{\rho}_{ij}^{UU} = 0$, because they are constant in TSE. Next, ϕ_i is defined as the fraction of state N around a residue in the TSE on a scale of 0 to 1, so that ρ_{ij} can be represented by the weighted average of $\bar{\rho}_{ij}^{\alpha\beta}$ (α = β = N/U).

$$\rho_{ij} = \bar{\rho}_{ij}^{NN}\phi_i\phi_j + \bar{\rho}_{ij}^{NU}\phi_i(1-\phi_j) + \bar{\rho}_{ij}^{UN}(1-\phi_i)\phi_j + \bar{\rho}_{ij}^{UU}(1-\phi_i)(1-\phi_j) \tag{2}$$

where $\phi_i\phi_j$ represents the fraction of the coexisting N states around residues *i* and *j*, and so on. As mentioned above, the individual ρ_{ij}^{NU} and ρ_{ij}^{UN} values are undetermined in each transition state, but they may vary symmetrically around 0.5 between 0 and 1 in TSE in ideal conditions, considering that the second rule requires no special interactions at the interface between state N and state U. Thus, we assume the best estimates of the values as $\bar{\rho}_{ij}^{NU} = \bar{\rho}_{ij}^{UN} = 0.5$, considering the ensemble averaging taken over TSE.

$$\rho_{ij} = 1\phi_i\phi_j + 0.5\,\phi_i(1-\phi_j) + 0.5\,(1-\phi_i)\phi_j + 0\,(1-\phi_i)(1-\phi_j) \tag{3}$$

This formula is simplified to the following:

$$\rho_{ij} = \frac{1}{2}\,(\phi_i + \phi_j) \tag{4}$$

which means that the slope of the line segment connecting residues *i* and *j* in a REFER plot is equal to the average of the ϕ values of the same pair of residues.

Let *n* be the number of observed residues. The number of ϕ is *n*, and the number of ρ is ½*n*(*n*-1). If equation 4 is valid, then the average and variance of ρ_{ij} are calculated to be

$$\overline{\rho} = \overline{\phi} \tag{5}$$

$$var\,\rho = \frac{1}{2}\overline{\phi^2} + \frac{1}{2}\overline{\phi_i\phi_j} - \overline{\phi}^2. \tag{6}$$

Refer to Appendix II for the derivation.

If all data points are perfectly on the linear line in a REFER plot, then all ρ_{ij} have the same value. Let this value be ρ.

$$\rho = \frac{1}{2}\,(\phi_i + \phi_j) \tag{7}$$

The solution of equation 7 is the single value ϕ for all residues,

$$\rho = \phi \tag{8}$$

which means that all residues share a single ϕ value in TSE and thus TSE is homogenous in the case of rbLFER (Fig. 1D).

QFER is the general solution

We seek a general solution that allows different ϕ values in TSE. Let the shape of the curve be *f*(x) in a REFER plot. The *f*(x) is transformed to *u*(x) = {*f*(x)-C}/x, where C is a constant. The *u* (x) is the solution of the following differential equation. Refer to Appendix III for the derivation.

$$\frac{du}{dx} = \frac{u-B}{x} \tag{9}$$

where B is another constant. In integrating equation 9, yet another constant A is introduced. The solution is a parabolic curve, $f(x) = Ax^2 + Bx + C$. The assignment of log k and log K to this formula provides the following quadratic free energy relationship (QFER) (Fig. 1E).

$$log\,k = A\,(\log\,K)^2 + B\log\,K + C \tag{10}$$

Here, ϕ is equal to the slope of the tangent at the data point (see (*) in Appendix III).

$$\phi = 2A\log\,K + B \tag{11}$$

Equation 11 indicates that the consistency principle is compatible with a polarized TSE, in which ϕ differs from residue to residue (Fig. 1D).

Comparison between the OLS and TLS fittings

Curve fitting is generally performed with the ordinary least squares (OLS) method. OLS assumes errors in the measurement of log k (i.e., y-axis) and no errors in that of log K (x-axis). OLS minimizes the sum of the squares of the vertical offsets to determine the fitting curve shape (Fig. 2). However, errors are also present in the measurement of log K. The total least squares (TLS) method is appropriate in this case. TLS minimizes the sum of the squares of the perpendicular offsets, instead of the vertical offsets. Refer to Appendix I for the details of the procedure. Practically, however, the resultant quadratic curves are not substantially different (Fig. 2). Thus, we use the results of the OLS fitting hereafter.

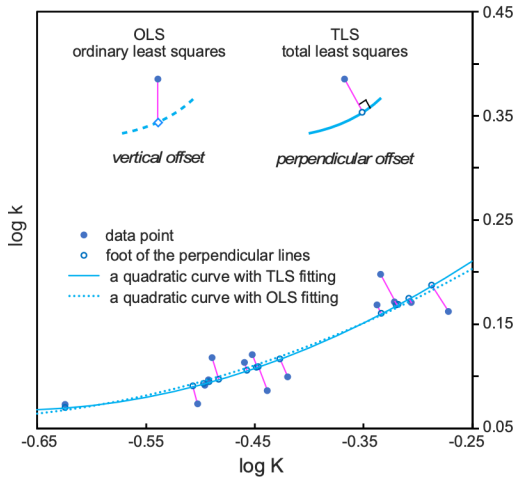

Figure 2 Comparison of two regression methods, OLS (ordinary least squares) and TLS (total least squares). The *insets* show the differences in the definition of an offset (i.e., the lengths of the magenta line segments). OLS and TLS seek parameters of the regression curves to minimize the sum of the squares of the offsets.

Nukacin ISK-1 and pKID-KIX systems as illustrations

In the previous retrospective analysis, we interpreted the REFER plots on the assumption of LFER[33]. Here, we reinterpreted the same REFER plots using QFER. In most instances, the fitting did not improve, indicating that the deviations of data points were within acceptable error ranges and the TSE should be regarded as homogeneous. As exceptions, curve fitting improved for two cases: the two-state topological interconversion of nukacin ISK-1 and the coupled folding and binding of the pKID peptide to the KIX domain. Nukacin ISK-1 is a 27-residue lantibiotic peptide (Fig. 3A)[40]. The two sets of $^{1}H-^{15}N$ amide cross peaks in the $^{1}H-^{15}N$ HSQC (heteronuclear single quantum coherence) spectrum indicated the existence of two states in solution, termed states A and B[41]. The interconversion rate was slow on the sub-second time scale, reflecting the high energy barrier, 16 k_BT[28,42]. We found residue-specific variations of the equilibrium constant, K, and the interconversion rates, $k_{A\to B}$ and $k_{B\to A}$, and linear relationships between log $k_{A\to B}$ and log K and between log $k_{B\to A}$ and log K in the REFER plot[29,32]. Because the structures of the two states differ in their backbone topologies, some side chains thread through the main chain ring formed by one of the three mono-sulfide linkages. Next, the 28-residue pKID peptide is an intrinsically disordered polypeptide (IDP) from the rat CREB protein (Fig. 3B). The pKID peptide contains a phosphoserine residue, which is critical for the high-affinity binding to a physiological partner, the 81-residue KIX domain of the mouse CBP (CREB binding protein)[43]. The coupled folding and binding of the pKID peptide to the KIX domain was analyzed by ^{15}N relaxation dispersion NMR, using a three-state model[44]. The present QFER analysis is focused on the first half of the pKID-KIX interaction, between the free state and the encounter complex. Note that the encounter complex state is referred to as an "intermediate" in the original report[44].

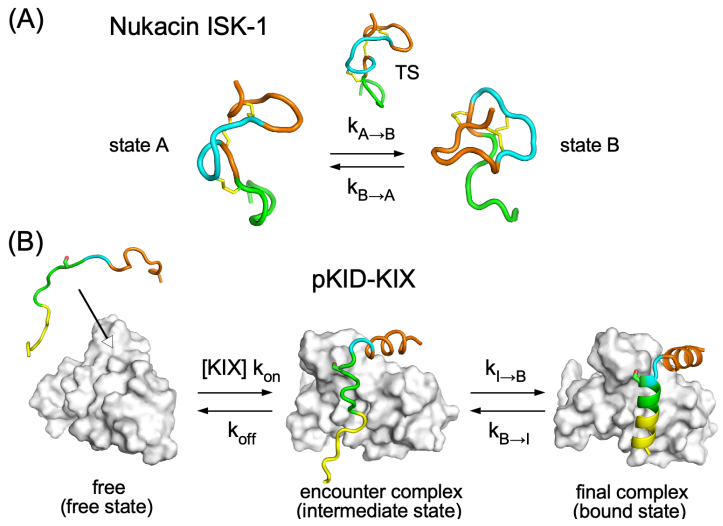

Figure 3 The two analyzed systems. (A) The two-state topological interconversion between states A and B of nukacin ISK-1. The amino acid sequence is color-coded according to the N-tail (green), hinge (cyan), and core (orange) regions. Nukacin ISK-1 contains three mono-sulfide linkages (yellow). The 3D structures of states A and B were determined by NMR [PDB IDs: 5Z5Q and 5Z5R], respectively. The 3D structure of the transition state (TS) was inferred from the targeted MD simulation[28]. (B) The coupled folding and binding of the pKID peptide to the KIX domain. The pKID peptide contains a phosphoserine residue (green stick side chain with red terminus). The interaction was analyzed by a three-state model in the original report[44]. The present REFER analysis is focused on the first half, between the free state and the encounter complex. The 3D structure of the final complex is the NMR structure of the pKID-KIX complex [PDB ID: 1KDX][43], and the other pKID peptide structures are imaginary, generated from the 1KDX structure using the sculpting function in the program PyMOL. The amino acid sequence is color-coded according to the αA-I (orange), kink (cyan), αB-I (green), and αB-II (yellow) structures.

Comparison between LFER and QFER

We now compare the fittings with LFER and QFER. The R^2 values are higher in QFER than in LFER (Fig. 4). As discussed previously[33], the triadic relationship, log K = log k – log k', could generate a fake rbLFER, and the third correlation in the log k vs. log k' plot must be examined to judge the significance of the rbLFER (Fig. 4, *insets*). In the case of LFER, log k vs. log k' has a linear relationship. In the case of QFER, log k vs. log k' has a tilted parabolic relationship (Appendix IV). The comparisons confirmed the improvement of fitting in the log k vs. log k' plots when QFER is applied.

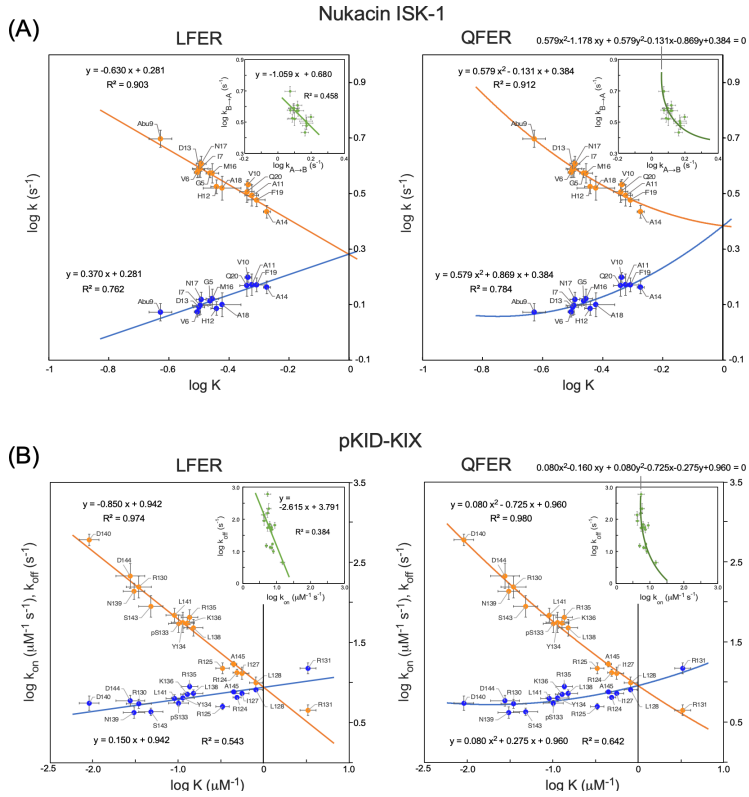

Figure 4 Comparison of LFER and QFER analyses. (A) Linear and quadratic OLS regression analyses of nukacin ISK-1. (B) Linear and quadratic OLS regression analyses of pKID-KIX. The *insets* show the log k vs. log k' plots with a fitted line or a quadratic curve.

We now describe the relationships among the coefficients of the regression equations. In the case of LFER, we obtain a linear formula, log k = ρ log K + α. Then, the assignment of the triadic relation, log k'= log k – log K, results in another linear formula, log k'= (ρ-1) log K + α. Thus, the difference between the slopes of the two linear formulas is always 1, and the coefficients of the constant terms are the same. In the case of QFER, we obtain a quadratic formula, log k =A (log K)^2 + B log K + C. Then, the assignment of log k'= log k – log K results in log k' = A(log K)^2 + (B-1) log K + C. Thus, the difference between the coefficients of the linear term is always 1, and the coefficients of the squared term, as well as the constant term, are the same. Lastly, the sign of the squared term, A, must be positive to make ϕ physically meaningful, because the fraction of state N in TSE (i.e., ϕ) should increase simultaneously with the fraction of state N in the ground state ensemble (i.e., increased state N fraction in GSE means increased log K), as clarified by equation 11.

Polarity and polar pattern

We formally define an apparent equilibrium constant $K^{\ddagger}$ as follows,

$$\mathrm{Log}\, K^{\ddagger} \equiv \log \frac{N^{\neq}}{U^{\neq}} = \log \frac{\phi}{1-\phi}\,. \tag{12}$$

By substituting equation 11 into equation 12, we obtain the following.

$$\log K^{\ddagger} = \log \frac{2A \log K + B}{-2A \log K + (1-B)} \tag{13}$$

In the graphs of equation 13 (Fig. 5), the line segments (black lines) approximate equation 13 in the ϕ range of 0.1 to 0.9. The line segments are expressed as follows.

$$\log K^{\ddagger} = \mu \log K + \beta \tag{14}$$

This is LFER between log $K^{\ddagger}$ and log K. The slope can be approximated as μ ≈ 4A (Appendix V).

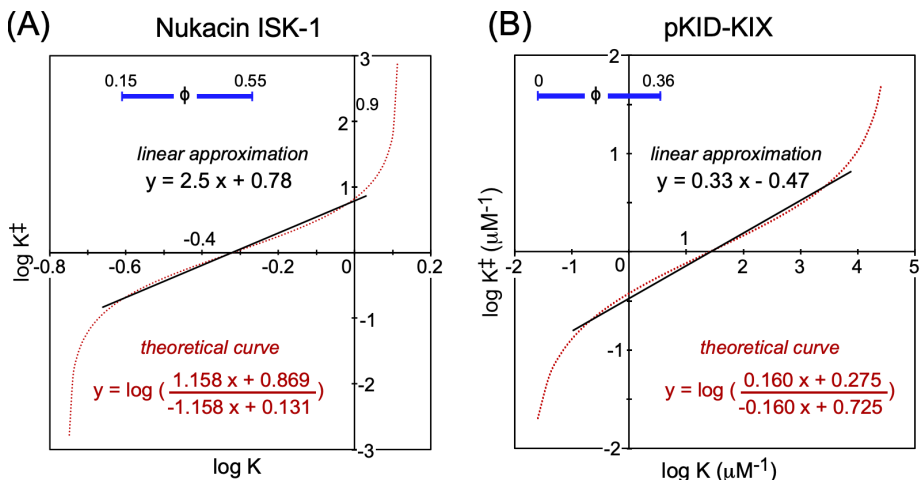

Figure 5 Log K‡ vs. log K plot. (a) Nukacin ISK-1 and (b) pKID-KIX. The red dotted lines are the theoretical curves defined by equation 13. The black solid lines are the linear approximation defined by equation 14. The blue bars indicate the range of ϕ values covered by the experimental data points.

We define 'polarity' as the uneven K‡ values in the TSE state, and 'polar pattern' as the spatial distribution of the K‡ values in the TSE structure. Similarly, we define the polarity and polar pattern of the ground state ensemble (GSE) using the K values. Equation 14 indicates that the polar patterns of TSE and GSE are closely interrelated, under the assumption of the consistency principle. The slope μ is an indicator for the polarization degree of TSE relative to GSE. If μ is zero, then TSE has no connection to GSE. The REFER plot is well described by LFER (A = 0), and the TSE is homogenous. Next, if $0 < \mu < 1$, then TSE is slightly polar. In the case of pKID-KIX, μ is equal to 0.33 (Fig. 5B). When the pKID peptide binds to the KIX protein (Fig. 3B), the degree of TSE polarization is one-third of that in the encounter complex. If $1 < \mu$, then TSE is more polar than GSE. In the case of nukacin ISK-1, μ is equal to 2.5 (Fig. 5A). In the transition state, some side chains thread through the main chain ring formed by one of the three mono-sulfide linkages (Fig. 3A). This may be the reason for the much higher polarization degree of TSE than that of GSE.

Structural insights into the transition states

Nukacin ISK-1 has a small hydrophobic core around residue Phe21 in state B[28], and hence the state B structure is more compact than the state A structure (Fig. 3A). The change in the direction of A→B can be regarded as a protein folding process, U→N, despite the small extent of structure formation. Similarly, the pKID binding to the KIX protein can also be regarded as a protein folding process, U→N, because the pKID peptide has a higher level of structural formation in the encounter complex than in the free state. Thus, the plots in Fig. 6A provide information about the transition state in the direction of structural formation ($k_{A\rightarrow B}$ and k_{on}). In both cases, visual inspection reveals some clusters of the (log K, log k) data points. For clarification, the data points are color-coded according to the structures of the destination states; i.e., state B for nukacin ISK-1 and the encounter complex state for pKID. As shown in Fig. 3, the amino acid sequence of nukacin ISK-1 is divided into three regions, N-tail, hinge, and core, by reference to the state B structure[28], and that of the pKID peptide is divided into four regions, αA-I, kink, αB-II, and αB-III, by reference to the bound conformation of pKID[44]. Residues with the same color tend to form a cluster in the REFER plots (Fig. 6A). This suggests that the adjacent residues in the destination structures change states in a coordinated manner. Thus, the clusters identified in the REFER plots agree with the foldon concept. Interestingly, the application of the consistency principle to the experimental results verifies the existence of foldons in TSE.

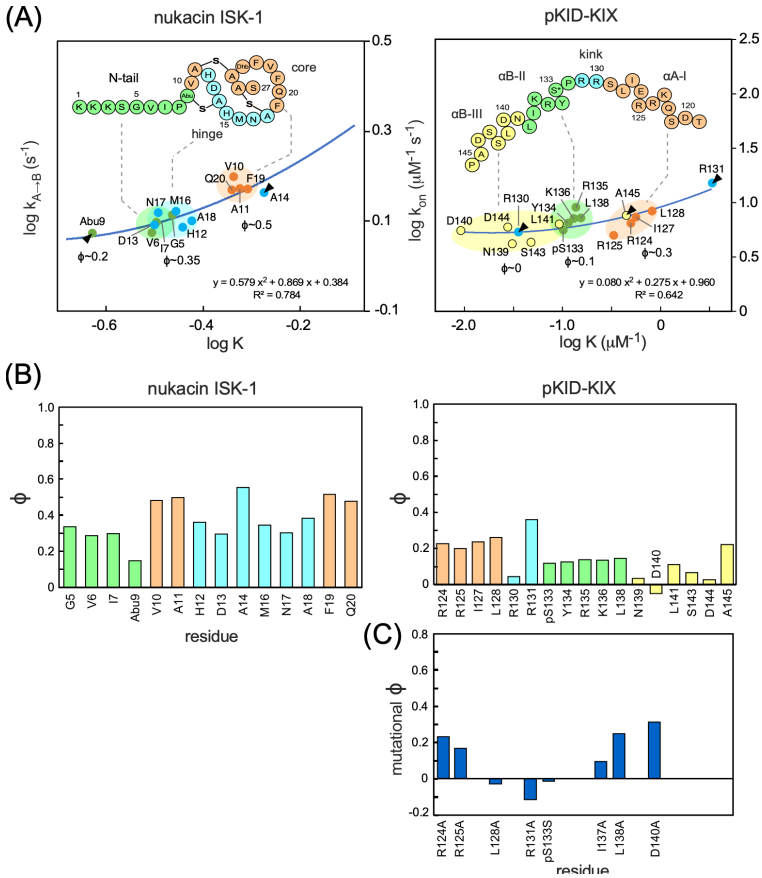

Figure 6 Interpretation of the QFER in the REFER plots of nukacin ISK-1 and pKID-KIX. (A) The connection between the energetic view (the cluster formation of the (log K, log k) data points) and the structural view (the segmentation of the amino acid sequences by reference to the three-dimensional structures). The arrowheads mark the exceptional residues, which are on the curves but away from the clusters to which they belong. (B) ϕ-values calculated according to equation 11. Equation 11 indicates that the ϕ value is the slope of the tangent at the data point corresponding to the residue. Because the data point is not always on the curve due to measurement errors, it is better to calculate the slope of the tangent at the nearest point on the curve by the TLS fitting, but we used the OLS fitting as a convenient alternative (Fig. 2, *insets*). The color coding is the same as in (A). (C) ϕ-values retrieved from the previous mutational ϕ-value analysis[45].

There are exceptional residues: Abu9 and Ala14 in nukacin ISK-1, and Arg130, Arg131, and Ala145 in pKID-KIX (arrowheads, Fig. 6A). The side chains of Abu9 and Ala14 of nukacin ISK-1 are linked by a mono-sulfide bond and involved in the 6-membered main-chain ring (the left *inset* of Fig. 6A). As discussed in detail previously[32], Abu9 and Ala14 probably became exceptions by adopting special states in the TSE to enable the threading of the His12 and Asp13 side chains through another main-chain ring. In the case of pKID-KIX, Arg130 and Arg131 are located at the kinked position between the two α-helices, αA and αB. Like the nukacin case, the two residues probably adopt special states in TSE. Ala145 was also an exception, but the mechanism is unknown.

QFER-ϕ analysis

The ϕ-value analysis based on equation 11 is referred to as a QFER-ϕ analysis. The QFER-ϕ analysis provides information about the preferential structure formation in the TSE state (Fig. 6B). The results of a mutational ϕ analysis are available for the pKID-KIX interaction[45]. The kinetic experiments using the N-terminally FITC-labeled pKID peptide showed two phases. The faster phase is assumed to correspond to the formation of the encounter complex of the pKID-KIX interaction (Fig. 3B). The authors stated that the overall ϕ-value pattern was only interpretable, due to the large errors in individual ϕ-values (Fig. 6C), and they discussed it as follows: First, all of the ϕ values were low (0.1-0.2), suggesting that the transition state in the formation of the encounter complex was largely unstructured. Secondly, the

negative ϕ-values suggested non-native interactions in the kinked region. By comparison, the small ϕ-values over the entire length and the abnormal pattern (though non-negative) in the kinked region (Arg130 and Arg131) were also observed in the QFER-ϕ analyses (Fig. 6B).

We compared the original mutational ϕ-value analysis developed by Alan Fersht[37–39] and our QFER-ϕ analysis. The original analysis uses the combination of thermodynamic and kinetic experiments of single-point mutants to provide information on the localized structure formation around the mutated side chain in TSE. In contrast, the QFER-ϕ analysis uses the data derived from the backbone amide cross peaks, and thus provides information about the main chain. The original ϕ-value analysis has limitations due to its assumptions: no changes in folding paths and the absence of U state perturbations by mutations[46], whereas the QFER-ϕ analysis does not require these assumptions.

From a practical perspective, one must perform one or multiple amino acid mutations per residue and repeat experiments for many residual positions in the mutational ϕ-value analysis, but only one amino acid sequence and a one-shot experiment suffice in the QFER-ϕ analysis. From an applied standpoint, experimentally derived ϕ values are useful for protein folding simulations[47]. Due to a common framework based on a single quadratic curve, the QFER-ϕ analysis is expected to generate mutually consistent ϕ values. This feature offers a great advantage for molecular simulation studies. The ϕ-ϕ plot is also a beneficial experimental application[48].

A different type of ϕ-value analysis was proposed, using the NMR relaxation dispersion method[49,50]. The NMR method should be useful, but requires amino acid mutations and thus the same assumptions as the original mutational ϕ-value analysis. The ψ-value analysis is another mutational method to probe the transition state[51]. A pair of histidine residues (biHis) is introduced into a protein molecule to create a metal ion binding site, and the stabilizing effects caused by increasing concentrations of metal ions are analyzed. The ψ-value analysis has technical advantages over the mutational ϕ-value analysis, but the incorporation of the biHis probes has the risk of changing the folding paths. In sum, the QFER-ϕ method covers the shortcomings of the mutational ϕ-value and ψ-value analyses. The only downside of the QFER-ϕ method is the requirement for highly accurate per-residue measurements.

Discussion

The consistency principle was originally proposed for the smooth folding of foldable proteins (i.e., naturally optimized amino acid sequences) from unstructured states[18]. The consistency principle is now broadly applicable to conformational changes, native-state dynamics, ligand binding of proteins, and even RNA[27]. Here, we derived a mathematical presentation of the consistency principle. Importantly, our formulas are verifiable by real experimental data. First, we adopted two simple and reasonable assumptions: $\bar{\rho}_{ij}^{NU} = \bar{\rho}_{ij}^{UN} = 0.5$ and ϕ is a function of log K. We then derived a basic equation, $\rho_{ij} = ½(\phi_i + \phi_j)$ (equation 4). In the log k vs. log K plot, a set of points satisfying the equation is on a parabolic curve defined by $\log k = A(\log K)^2 + B\log K + C$ (equation 10). This quadratic free energy relationship (QFER) is an extended version of the residue-based LFER, which we reported previously [29,33]. The residue-based QFER (rbQFER) is a mathematical formulation of the protein folding and conformational changes of *ideal* proteins. From a practical viewpoint, rbQFER provides a framework for analyzing the transition state of two-state structural exchanges. The first application is the elucidation of a new linear relationship between $\log K^{\ddagger}$ and log K (Fig. 5). The slope μ is an indicator for the polarization degree of TSE relative to GSE and can be approximated as $\mu \approx 4A$. If A = 0, then QFER reduces to LFER, and the TSE state is homogenous. The second application is a new ϕ-value analysis (Figs. 1F and 6B). The ϕ value is obtainable as the slope of the tangent nearest to the data point, according to equation 11. As we have already discussed, the greatest advantage of the QFER-ϕ analysis is that mutated proteins are not needed.

We summarize the new insights into the dynamic nature of the conformational changes of protein molecules obtained from the residue-specific thermodynamic and kinetic analyses (Fig. 7). Conventionally, assuming perfect cooperativity among residues (Fig. 7A), the two-state exchange of a protein and the two-state binding of an IDP ligand are described by a single set of K, k, and k' values. The transition states of the two-state exchanges are uniform and characterized by a single ϕ value. By contrast, the residue-based LFER/QFER provides a new view (Fig. 7B). Protein molecules and IDP ligands are divided into independent structural units, which are conveniently referred to as 'foldons' and 'interactons', respectively. Each foldon and interacton is characterized by a specific set of K, k, and k' values. This situation represents a mosaic of different degrees of cooperativity among residues[28]. The transition states of foldons and interactons may have different ϕ values, indicating their heterogeneity and the precedent local structure formation of one foldon/interacton ahead of the others. What is important is that all foldons and interactons are *consistently* coupled to achieve smooth conformational changes and interactions, judging from the fact that the residues are on the line/curves in the REFER plots. Exceptionally, some residues are located apart from the cluster to which to belong, but these exceptional residues are still on the same line/curve (Fig. 6A), indicating they are still consistent and do not disturb the smooth folding/interaction. They may play special roles in the transition state and could serve as hot spots for manipulating the dynamic properties of the protein molecules and IDP ligands.

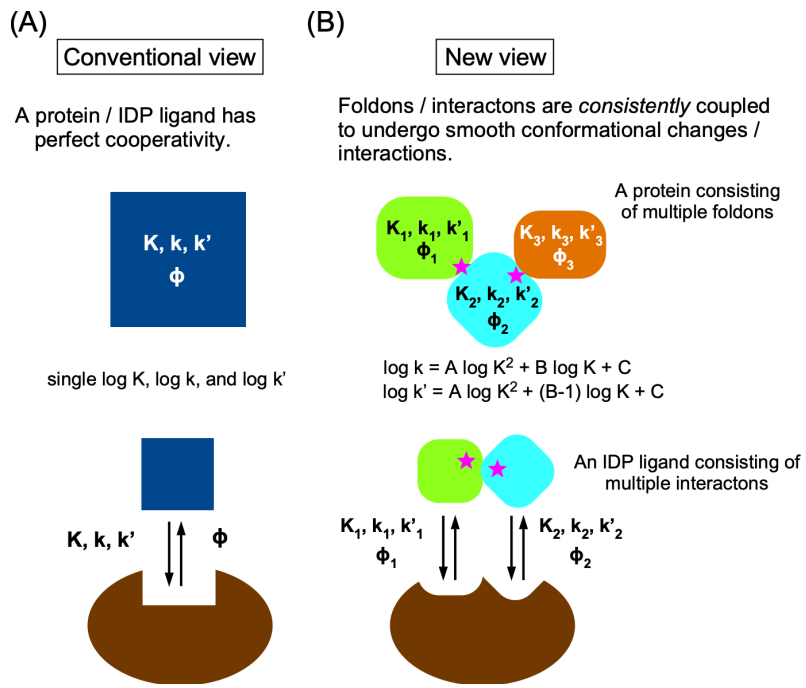

Figure 7 Two views on the dynamic nature of proteins. (A) Conventionally, a protein/IDP ligand has perfect cooperativity. The two-state exchanges are described with single thermodynamic and kinetic parameters. (B) Residue-based LFER/QFER provides a new view. A protein/IDP ligand consists of multiple foldons/interactions, which are loosely but *consistently* coupled to each other. The magenta stars mark the exceptional residues, which are on the lines/curves but away from the clusters to which to belong (Fig. 6A). They are located at the interfaces between foldons/interactions and may play special roles in the transition state.

Appendices

Appendix I

The OLS fitting to a quadratic curve, $y = Ax^2 + Bx + C$, was performed by the standard polynomial fitting in Excel. The TLS fitting was performed as follows: Let a data point be (x_0, y_0) and the foot of the perpendicular be (x, y) from the data point on a quadratic curve, $y = Ax^2 + Bx + C$ (Fig. 2). When a vector, $\mathbf{u} = (x-x_0, y-y_0)$, and a tangent vector, $\mathbf{v} = (1, 2A+B)$, bisect each other at right angles, then the inner product, $\mathbf{u} \cdot \mathbf{v} = (x - x_0) + (y - y_0)(2A + B) = 0$, which is a cubic equation in terms of x with real coefficients. The foot point (x, y) was obtained by solving the cubic equation with Cardano's method. Then, we used the Solver in Excel to find optimal A, B, and C values by minimizing the sum of the square of the orthogonal distance, $|\mathbf{u}|$.

Appendix II

Let n be the number of observed residues. The combinatorial number of two residues, N, is equal to $\frac{1}{2}n(n-1)$. First, we compute the average of ρ_{ij}.

$$\bar{\rho} = \frac{1}{N}\sum_{i,j}^{i\neq j}\frac{1}{2}(\phi_i + \phi_j) = \frac{1}{2N}\left\{\sum_j^{j\neq i}\sum_i \phi_i + \sum_i^{i\neq j}\sum_j \phi_j\right\} = \frac{1}{n-1}\left\{\sum_j^{j\neq i}\sum_i \frac{\phi_i}{n} + \sum_i^{i\neq j}\sum_j \frac{\phi_j}{n}\right\}$$

$$= \frac{1}{n-1}\left\{\bar{\phi}\sum_j^{j\neq i} 1 + \bar{\phi}\sum_i^{i\neq j} 1\right\} = \frac{1}{n-1}\left\{\frac{1}{2}(n-1)\bar{\phi} + \frac{1}{2}(n-1)\bar{\phi}\right\} = \bar{\phi}$$

Then, we compute the variance of ρ_{ij}.

$$var\ \rho = \frac{1}{N}\sum_{i,j}^{i\neq j}(\rho_{ij}-\overline{\rho})^2 = \frac{1}{N}\sum_{i,j}^{i\neq j}(\rho_{ij}^2 - 2\rho_{ij}\overline{\rho} + \overline{\rho}^2)$$

$$= \frac{1}{N}\sum_{i,j}^{i\neq j}\left\{\frac{1}{4}(\phi_i+\phi_j)^2 - (\phi_i+\phi_j)\overline{\rho} + \overline{\rho}^2\right\}$$

$$= \frac{1}{N}\left\{\frac{1}{4}\sum_{i,j}^{i\neq j}\phi_i^2 + \frac{1}{2}\sum_{i,j}^{i\neq j}\phi_i\phi_j + \frac{1}{4}\sum_{i,j}^{i\neq j}\phi_j^2 - \overline{\rho}\sum_{i,j}^{i\neq j}\phi_i - \overline{\rho}\sum_{i,j}^{i\neq j}\phi_j + \overline{\rho}^2\sum_{i,j}^{i\neq j}1\right\}$$

$$= \frac{1}{4}\overline{\phi^2} + \frac{1}{2}\overline{\phi_i\phi_j} + \frac{1}{4}\overline{\phi^2} - \overline{\phi}^2 - \overline{\phi}^2 + \overline{\phi}^2$$

$$= \frac{1}{2}\overline{\phi^2} + \frac{1}{2}\overline{\phi_i\phi_j} - \overline{\phi}^2$$

Note that we used the following equalities.

$$\textstyle\sum_{i,j}^{i\neq j} 1 = \frac{1}{2}n(n-1) \text{ and } \sum_{i}^{i\neq j} 1 = \frac{1}{2}(n-1)$$

Appendix III

Let the equation of the solution curve be *f*(x). We assume that ϕ is a function of x. The next equation is obtained from equation 4.

$$\rho_{12} = \frac{f(x_2)-f(x_1)}{x_2-x_1} = \frac{1}{2}\{\phi(x_2)+\phi(x_1)\}$$

Given the limit of $x_2 \to x_1$, the left-hand side and the right-hand side are transformed as follows:

$$\lim_{x_2\to x_1}\frac{f(x_2)-f(x_1)}{x_2-x_1} = f'(x_1) \text{ and } \lim_{x_2\to x_1}\frac{1}{2}\{\phi(x_2)+\phi(x_1)\} = \phi(x_1).$$

Thus, ϕ is the derivative of *f*(x). (*)

By substitution, we obtain the following.

$$\frac{f(x_2)-f(x_1)}{x_2-x_1} = \frac{1}{2}\{f'(x_2)+f'(x_1)\}.$$

Assume that $x_1 = 0$ and $x = x_2$ without loss of generality.

$$\frac{f(x)-f(0)}{x} = \frac{1}{2}\{f'(x)+f'(0)\}$$

Specify initial conditions as C = *f*(0) and B = *f*'(0) and set y = *f*(x) - C.

$$\frac{y}{x} = \frac{1}{2}\{y'+B\}$$

$$y' = 2\frac{y}{x} - B \qquad (**)$$

By setting *u* = y/x, we have

$$\frac{du}{dx} = \frac{u-B}{x}$$

and variable separation gives

$$\frac{du}{u-B} = \frac{dx}{x}$$

Integrate both sides

$$log|u-B| = log|x| + a \quad (a \text{ is a constant of integration}).$$

By setting $A = \pm e^a$, then we have

$$\frac{u - B}{x} = A \quad (A \neq 0)$$

Here, if A is assumed to be 0, then y = Bx is the solution of the differential equation (**). Thus, A is an arbitrary constant. Using u = y/x = {f(x) – C}/x, we obtain the following.

$$f(x) = A\,x^2 + B\,x + C$$

Appendix IV

The assignment of log K = log k – log k' to log k = A(log K)^2 + B log K + C and the simplification produces a quadratic formula, A(log k)^2 - 2A(log k)(log k') + A(log k')^2 + (B-1)log k - Blog k' + C = 0. This is an equation of a conic curve, $ax^2 + bxy + cy^2 + dx + ey + f = 0$. The discriminant of the equation is $D = b^2 - 4ac$. In our case, $D = (-2A)^2 - 4AA = 0$, and thus the quadratic formula represents a tilted parabola.

Appendix V

The slope of the curve defined by equation 13 is (log K‡)' = (2A/ln10) / {0.5(1-0.5)} = 3.47A at $\phi = 0.5$. This equation provides the slope of 2.0 for nukacin ISK-1 and that of 0.28 for pKID-KIX at $\phi = 0.5$. If we apply linear approximation in the range of $0.1 \leq \phi \leq 0.9$, then the slope μ is 2.5 for nukacin ISK-1 and 0.33 for pKID-KIX. These μ values are about 20-25% larger than the slope of the curve at $\phi = 0.5$. Thus, we can use an empirical value, $\mu \approx 4A$.

Conflict of Interest

The authors note no conflict of interest concerning the research presented here.

Author Contributions

Conceptualization and Investigation, D.K., H.S., and D.F.; Writing – Original Draft, D.K.; Writing – Review & Editing, D.K., H.S., and D.F.; Funding Acquisition, D.K.

Acknowledgements

This work was partly performed in the Medical Research Center Initiative for High Depth Omics, at the Medical Institute of Bioregulation, Kyushu University. This work was supported by the Japan Society for the Promotion of Science (JSPS, Japan), KAKENHI Grant Number JP21H02448, and by the Mitsubishi Foundation (Japan) Research Grants in the Natural Sciences, Grant Number 202110017, to D.K.

References

[1] Anfinsen, C. B. Principles that Govern the Folding of Protein Chains. Science (80-.). 181, 223–230 (1973). https://doi.org/10.1126/science.181.4096.223

[2] Jackson, S. E., Fersht, A. R. Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition. Biochemistry 30, 10428–10435 (1991). https://doi.org/10.1021/bi00107a010

[3] Bai, Y., Zhou, H., Zhou, Y. Critical nucleation size in the folding of small apparently two-state proteins. Protein Sci. 13, 1173–1181 (2004). https://doi.org/10.1110/ps.03587604

[4] Malhotra, P., Udgaonkar, J. B. How cooperative are protein folding and unfolding transitions? Protein Sci. 25, 1924–1941 (2016). https://doi.org/10.1002/pro.3015

[5] Vu, D. M., Brewer, S. H., Dyer, R. B. Early turn formation and chain collapse drive fast folding of the major cold shock protein CspA of *Escherichia coli*. Biochemistry 51, 9104–9111 (2012). https://doi.org/10.1021/bi301296y

[6] Bhatia, S., Udgaonkar, J. B. Heterogeneity in Protein Folding and Unfolding Reactions. Chem. Rev. 122, 8911–8935 (2022). https://doi.org/10.1021/acs.chemrev.1c00704

[7] Wolynes, P. G., Onuchic, J. N., Thirumalai, D. Navigating the Folding Routes. Science (80-.). 267, 1619–1620 (1995). https://doi.org/10.1126/science.7886447

[8] Frauenfelder, H., Sligar, S. G., Wolynes, P. G. The Energy Landscapes and Motions of Proteins. Science (80-.). 254, 1598–1603 (1991). https://doi.org/10.1126/science.1749933

[9] Ferreiro, D. U., Komives, E. A., Wolynes, P. G. Frustration in biomolecules. Q. Rev. Biophys. 47, 285–363 (2014). https://doi.org/10.1017/S0033583514000092

[10] Ferreiro, D. U., Komives, E. A., Wolynes, P. G. Frustration, function and folding. Curr. Opin. Struct. Biol. 48, 68–73 (2018). https://doi.org/10.1016/j.sbi.2017.09.006

[11] Hu, W., Kan, Z. Y., Mayne, L., Englander, S. W. Cytochrome c folds through foldon-dependent native-like intermediates in an ordered pathway. Proc. Natl. Acad. Sci. U. S. A. 113, 3809–3814 (2016). https://doi.org/10.1073/pnas.1522674113

[12] Englander, S. W., Mayne, L. The nature of protein folding pathways. Proc. Natl. Acad. Sci. U. S. A. 111, 15873–15880 (2014). https://doi.org/10.1073/pnas.1411798111

[13] Englander, S. W., Mayne, L. The case for defined protein folding pathways. Proc. Natl. Acad. Sci. U. S. A. 114, 8253–8258 (2017). https://doi.org/10.1073/pnas.1706196114

[14] Johnson, E. A Maximum Caliber analysis of the Foldon Hypothesis. Proteins Struct. Funct. Bioinforma. 90, 1170–1178 (2022). https://doi.org/10.1002/prot.26299

[15] Eaton, W. A., Wolynes, P. G. Theory, simulations, and experiments show that proteins fold by multiple pathways. Proc. Natl. Acad. Sci. U. S. A. 114, E9759–E9760 (2017). https://doi.org/10.1073/pnas.1716444114

[16] Englander, S. W., Mayne, L. Reply to Eaton and Wolynes: How do proteins fold? Proc. Natl. Acad. Sci. U. S. A. 114, E9761–E9762 (2017). https://doi.org/10.1073/pnas.1716929114

[17] Rollins, G. C., Dill, K. A. General mechanism of two-state protein folding kinetics. J. Am. Chem. Soc. 136, 11420–11427 (2014). https://doi.org/10.1021/ja5049434

[18] Go, N. Theoretical Studies of Protein Folding. Annu. Rev. Biophys. Bioeng. 12, 183–210 (1983). https://doi.org/10.1146/annurev.bb.12.060183.001151

[19] Taketomi, H., Ueda, Y., Gō, N. STUDIES ON PROTEIN FOLDING, UNFOLDING AND FLUCTUATIONS BY COMPUTER SIMULATION. Int. J. Pept. Protein Res. 7, 445–459 (2009). https://doi.org/10.1111/j.1399-3011.1975.tb02465.x

[20] Shakhnovich, E. I., Gutin, A. M. Implications of thermodynamics of protein folding for evolution of primary sequences. Nature 346, 773–775 (1990). https://doi.org/10.1038/346773a0

[21] Go, N. Snake cube puzzle and protein folding. Biophys. Physicobiology 16, 256–263 (2019). https://doi.org/10.2142/biophysico.16.0_256

[22] Nymeyer, H., García, A. E., Onuchic, J. N. Folding funnels and frustration in off-lattice minimalist protein landscapes. Proc. Natl. Acad. Sci. U. S. A. 95, 5921–5928 (1998). https://doi.org/10.1073/pnas.95.11.5921

[23] Koga, N., Takada, S. Roles of native topology and chain-length scaling in protein folding: A simulation study with a Gō-like model. J. Mol. Biol. 313, 171–180 (2001). https://doi.org/10.1006/jmbi.2001.5037

[24] Hoang, T. X., Cieplak, M. Sequencing of folding events in Go-type proteins. J. Chem. Phys. 113, 8319–8328 (2000). https://doi.org/10.1063/1.1314868

[25] Bryngelson, J. D., Wolynes, P. G. Spin glasses and the statistical mechanics of protein folding. Proc. Natl. Acad. Sci. U. S. A. 84, 7524–7528 (1987). https://doi.org/10.1073/pnas.84.21.7524

[26] Bryngelson, J. D., Onuchic, J. N., Socci, N. D., Wolynes, P. G. Funnels, pathways, and the energy landscape of protein folding: A synthesis. Proteins Struct. Funct. Genet. 21, 167–195 (1995).

https://doi.org/10.1002/prot.340210302
[27] Takada, S. Gō model revisited. Biophys. Physicobiology 16, 248–255 (2019). https://doi.org/10.2142/biophysico.16.0_248
[28] Fujinami, D., Motomura, H., Oshima, H., Mahin, A.-A., Elsayed, K. M., Zendo, T., et al. Mosaic Cooperativity in Slow Polypeptide Topological Isomerization Revealed by Residue-Specific NMR Thermodynamic Analysis. J. Phys. Chem. Lett. 11, 1934–1939 (2020). https://doi.org/10.1021/acs.jpclett.9b03591
[29] Fujinami, D., Hayashi, S., Kohda, D. Residue-Specific Kinetic Insights into the Transition State in Slow Polypeptide Topological Isomerization by NMR Exchange Spectroscopy. J. Phys. Chem. Lett. 12, 10551–10557 (2021). https://doi.org/10.1021/acs.jpclett.1c02387
[30] Farrow, N. A., Zhang, O., Forman-Kay, J. D., Kay, L. E. Comparison of the Backbone Dynamics of a Folded and an Unfolded SH3 Domain Existing in Equilibrium in Aqueous Buffer. Biochemistry 34, 868–878 (1995). https://doi.org/10.1021/bi00003a021
[31] Puglisi, R., Brylski, O., Alfano, C., Martin, S. R., Pastore, A., Temussi, P. A. Quantifying the thermodynamics of protein unfolding using 2D NMR spectroscopy. Commun. Chem. 3, 1–7 (2020). https://doi.org/10.1038/s42004-020-00358-1
[32] Hayashi, S., Kohda, D. The time-zero HSQC method improves the linear free energy relationship of a polypeptide chain through the accurate measurement of residue-specific equilibrium constants. J. Biomol. NMR 76, 87–94 (2022). https://doi.org/10.1007/s10858-022-00396-y
[33] Fujinami, D., Hayashi, S., Kohda, D. Retrospective study for the universal applicability of the residue-based linear free energy relationship in the two-state exchange of protein molecules. Sci. Rep. 12, 16843 (2022). https://doi.org/10.1038/s41598-022-21226-z
[34] Settanni, G., Rao, F., Caflisch, A. Φ-Value analysis by molecular dynamics simulations of reversible folding. Proc. Natl. Acad. Sci. U. S. A. 102, 628–633 (2005). https://doi.org/10.1073/pnas.0406754102
[35] Matouschek, A., Fersht, A. R. Application of physical organic chemistry to engineered mutants of proteins: Hammond postulate behavior in the transition state of protein folding. Proc. Natl. Acad. Sci. U. S. A. 90, 7814–7818 (1993). https://doi.org/10.1073/pnas.90.16.7814
[36] Best, R. B., Hummer, G., Eaton, W. A. Native contacts determine protein folding mechanisms in atomistic simulations. Proc. Natl. Acad. Sci. U. S. A. 110, 17874–17879 (2013). https://doi.org/10.1073/pnas.1311599110
[37] Matouschek, A., Kellis, J. T., Serrano, L., Fersht, A. R. Mapping the transition state and pathway of protein folding by protein engineering. Nature 340, 122–126 (1989). https://doi.org/10.1038/340122a0
[38] Fersht, A. R., Matouschek, A., Serrano, L. The folding of an enzyme. I. Theory of protein engineering analysis of stability and pathway of protein folding. J. Mol. Biol. 224, 771–782 (1992). https://doi.org/10.1016/0022-2836(92)90561-W
[39] Fersht, A. R., Sato, S. Φ-Value Analysis and the Nature of Protein-Folding Transition States. Proc. Natl. Acad. Sci. U. S. A. 101, 7976–7981 (2004). https://doi.org/10.1073/pnas.0402684101
[40] Sashihara, T., Kimura, H., Higuchi, T., Adachi, A., Matsusaki, H., Sonomoto, K., et al. A Novel Lantibiotic, Nukacin ISK-1, of *Staphylococcus warneri* ISK-1: Cloning of the Structural Gene and Identification of the Structure. Biosci. Biotechnol. Biochem. 64, 2420–2428 (2000). https://doi.org/10.1271/bbb.64.2420
[41] Fujinami, D., -Mahin, A.-A., Elsayed, K. M., Islam, M. R., Nagao, J., Roy, U., et al. The lantibiotic nukacin ISK-1 exists in an equilibrium between active and inactive lipid-II binding states. Commun. Biol. 1, 150 (2018). https://doi.org/10.1038/s42003-018-0150-3
[42] Fujinami, D., Motomura, H., Oshima, H., Mahin, A.-A., Elsayed, K. M., Zendo, T., et al. Correction to "Mosaic Cooperativity in Slow Polypeptide Topological Isomerization Revealed by Residue-Specific NMR Thermodynamic Analysis". J. Phys. Chem. Lett. 12, 10766–10767 (2021). https://doi.org/10.1021/acs.jpclett.1c03473
[43] Radhakrishnan, I., Pérez-Alvarado, G. C., Parker, D., Dyson, H. J., Montminy, M. R., Wright, P. E. Solution structure of the KIX domain of CBP bound to the transactivation domain of CREB: A model for activator:coactivator interactions. Cell 91, 741–752 (1997). https://doi.org/10.1016/S0092-8674(00)80463-8
[44] Sugase, K., Dyson, H. J., Wright, P. E. Mechanism of coupled folding and binding of an intrinsically disordered protein. Nature 447, 1021–1025 (2007). https://doi.org/10.1038/nature05858
[45] Dahal, L., Kwan, T. O. C., Shammas, S. L., Clarke, J. pKID Binds to KIX via an Unstructured Transition State with Nonnative Interactions. Biophys. J. 113, 2713–2722 (2017). https://doi.org/10.1016/j.bpj.2017.10.016
[46] Sosnick, T. R., Dothager, R. S., Krantz, B. A. Differences in the folding transition state of ubiquitin indicated by φ and ψ analyses. Proc. Natl. Acad. Sci. U. S. A. 101, 17377–17382 (2004). https://doi.org/10.1073/pnas.0407683101
[47] Salvatella, X., Dobson, C. M., Fersht, A. R., Vendruscolo, M. Determination of the folding transition states of

barnase by using Φ I -value-restrained simulations validated by double mutant Φ IJ -values. Proc. Natl. Acad. Sci. U. S. A. 102, 12389–12394 (2005). https://doi.org/10.1073/pnas.0408226102

[48] Gianni, S., Jemth, P. Conserved nucleation sites reinforce the significance of Phi value analysis in protein-folding studies. IUBMB Life 66, 449–452 (2014). https://doi.org/10.1002/iub.1287

[49] Cho, J. H., O'Connell, N., Raleigh, D. P., Palmer, A. G. φ-Value analysis for ultrafast folding proteins by NMR relaxation dispersion. J. Am. Chem. Soc. 132, 450–451 (2010). https://doi.org/10.1021/ja909052h

[50] Neudecker, P., Zarrine-Afsar, A., Davidson, A. R., Kay, L. E. Φ-value analysis of a three-state protein folding pathway by NMR relaxation dispersion spectroscopy. Proc. Natl. Acad. Sci. U. S. A. 104, 15717–15722 (2007). https://doi.org/10.1073/pnas.0705097104

[51] Sosnick, T. R., Krantz, B. A., Dothager, R. S., Baxa, M. Characterizing the Protein Folding Transition State Using ψ Analysis. Chem. Rev. 106, 1862–1876 (2006). https://doi.org/10.1021/cr040431q