http://www.jstage.jst.go.jp/browse/biophysics

BIOPHYSICS

Vol. 7, pp. 1–10 (2011)
doi: 10.2142/biophysics.7.1

Systematic alanine insertion reveals the essential regions that encode structure formation and activity of dihydrofolate reductase

Rumi Shiba[1], Mika Umeyama[1], Sayaka Tsukasa[1], Hironari Kamikubo[1], Yoichi Yamazaki[1], Mariko Yamaguchi[1], Masahiro Iwakura[2] and Mikio Kataoka[1]

[1]*Graduate School of Materials Science, Nara Institute of Science and Technology (NAIST), Ikoma, Nara 630-0192, Japan*
[2]*Protein Design Research Group, Institute for Biological Resources and Functions, National Institute of Advanced Industrial Science and Technology (AIST), 1-1-1 Higashi, Tsukuba, Ibaraki 305-8566, Japan*

Received November 19, 2010; accepted December 7, 2010

Decoding sequence information is equivalent to elucidating the design principles of proteins. For this purpose, we conducted systematic alanine insertion analysis to reveal the regions in the primary structure where the sequence continuity cannot be disrupted. We applied this method to dihydrofolate reductase (DHFR), and examined the effects of alanine insertion on structure and the enzymatic activity by solubility assay and trimethoprim resistance, respectively. We revealed that DHFR is composed of "Structure Elements", "Function Elements" and linkers connecting these elements. The "Elements" are defined as regions where the alanine insertion caused DHFR to become unstructured or inactive. Some "Structure Elements" overlap with "Function Elements", indicating that loss of structure leads to loss of function. However, other "Structure Elements" are not "Function Elements", in that alanine insertion mutants of these regions exhibit substrate- or inhibitor-induced folding. There are also some "Function Elements" which are not "Structure Elements"; alanine insertion into these elements deforms the catalytic site topology without the loss of tertiary structure. We hypothesize that these elements are involved essential interactions for structure formation and functional expression. The "Elements" are closely related to the module structure of DHFR. An "Element" belongs to a single module, and a single module is composed of some number of "Elements." We propose that properties of a module are determined by the "Elements" it contains. Systematic alanine insertion analysis is an effective and unique method for deriving the regions of a sequence that are essential for structure formation and functional expression.

Key words: function element, structure element, design principle of protein architecture, module, sequence consecutivity

Correspondence author: Mikio Kataoka, Graduate School of Materials Science, Nara Institute of Science and Technology (NAIST), Ikoma, Nara 630-0192, Japan.
e-mail: kataoka@ms.naist.jp

The amino acid sequence of a protein dictates its tertiary structure and function[1]. Decoding the information in the sequence is equivalent to elucidating the design principles of proteins; this is known as the protein folding problem. Many experimental and theoretical studies on this problem have been conducted to deepen our understanding[1]. The consistency principle[2] and the folding funnel[3,4] concept are the leading theories for the understanding of protein folding. However, these theories never tell us which sequences are foldable. The module hypothesis[5] teaches us that a protein is composed of building blocks. Since a module is assumed to be corresponding to an exon, each building block is predictable[5]. Although fascinating, this hypothesis cannot tell us how the building blocks interact with each other in the tertiary structure.

The idea of building blocks introduced by the module

hypothesis suggests that either the contiguity or connectivity of amino acid residues is the essential property of a building block. The simplification of an amino acid sequence has been adopted to identify the regions containing the information that is essential for the structure and function of the protein[6–8]. Consequently, it has been suggested that the residues that cannot accept simplification will be consecutive in the primary sequence, forming contiguous regions. Another approach to examine the effect of sequence connectivity is circular permutation analysis[9–11]. Circular permutation of a protein consists of connecting the native N- and C-termini covalently with an artificial peptide linker and cleaving the peptide backbone at one specific site. In principle, the cleavage site can be selected anywhere between the N-terminus and the C-terminus. In some cases, the mutant will be active and in native conformation; in other cases, the mutant will be inactive or have a denatured structure. In the latter cases, we can assume that the site should be connected and is therefore a part of a building block.

Iwakura et al. systematically applied circular permutation to dihydrofolate reductase (DHFR)[12–14]. Consequently, they revealed that DHFR is composed of several building blocks called folding element (FE) and linkers connecting FE. FE is defined as regions where cleavage and generation of new N- and C-termini are not tolerated. The work clearly indicates the effectiveness of circular permutation for the identification of the building blocks of a protein that are responsible for structure formation. In the case of DHFR, the two termini are very close to each other. For such a protein, circular permutation can be easily applied with a short, reasonably designed linker. However, if the two termini are far away, circular permutation would be difficult to apply because of a long linker required.

We hypothesized that insertion of an additional amino acid into a sequence brings about a separation of the preceding and the following sequences at the inserted point. The method can be applied to all proteins, regardless of the relative positions of their termini. Alanine insertion has been successfully applied to staphylococcal nuclease[15,16]. We expect that systematic alanine insertion can identify building blocks as effectively as systematic circular permutation. We applied the method to DHFR and compared the results with those of the circular permutation analysis.

DHFR (EC 1.5. 1.3) is a monomeric, two-domain protein with 159 amino acids that catalyzes the reduction of dihydrofolate (DHF) to tetrahydrofolate (THF), using NADPH as the reducing cofactor[17,18]. DHFR is a clinically important enzyme: it is the target of a number of antifolate drugs, such as trimethoprim (TMP) and methotrexate (MTX), and is involved in the production of an anticancer drug, 1-leucovorin[19]. DHFR is also a widely utilized model protein in the study of folding.

In this study, we constructed all possible insertion mutants of 159-residue DHFR. When alanine (A) is inserted between i-th and (i+1)-th residues, we call it iA(i+1). The functions and structures of all insertion mutants were qualitatively examined by TMP resistance assay and solubility assay, respectively. Structures and enzymatic activities of some purified mutants were also evaluated. We identified some essential building blocks for structure formation and enzymatic activity as regions where alanine insertion brings either the loss of foldability or the loss of enzymatic activity. The relationship between the module structure of DHFR and the building blocks gives further insight into the design principles of DHFR architecture.

Materials and Methods

Construction of alanine-inserted variants

Construction of the alanine-inserted genes was performed by polymerase chain reaction (PCR) using oligonucleotide primers including an alanine codon at the inserted position, and the genes were cloned into the vector pUC18. The plasmid containing DHFR wild-type gene is pTZwt1-3[20]. *Escherichia coli* JM109 was used as a host for cloning and expression. The sequences of the genes of the inserted mutants were confirmed using an ABI PRIZM 3100-Avant Genetic Analyzer (Applied Biosystems). The *E. coli* DHFR contains 13 alanines; if the n-th residue is A, (n−1)An and nA(n+1) have the same sequence. Therefore, we constructed 145 possible alanine insertion mutants.

Trimethoprim (TMP) resistance assay

TMP is a competitive inhibitor of DHFR. *E. coli* JM109 cells transformed with either wild type or alanine insertion mutants were streaked on agar plates containing 50 μg/ml ampicillin with and without TMP. TMP concentration was 1 μg/ml, if present. After 12-hour incubation, we measured colony formation (Fig. 1)[12,21,22].

Solubility assay

E. coli JM109 transformant cells were cultured for 12 hours at 37°C with 2×YT medium. Cells were fractured by sonication. After centrifugation, both the supernatant and precipitant were subjected on SDS-PAGE. The resultant gels were stained with Coomassie Fluor™ Orange Protein Gel Stain (Invitrogen). The fluorescence intensity of the DHFR band in each lane (excitation 300 nm, emission 570 nm) was analyzed using Image Master (BIO-RAD Molecular Imager Chem Dox™ XRS). The ratio of the precipitant intensity to the total intensity was obtained for all mutants.

Protein purification

Alanine insertion mutants and wild-type protein were expressed in *E. coli* JM109. Proteins were purified using MTX affinity column chromatography as described by Iwakura et al[14]. The resulting eluate was further purified via DEAE sepharose column chromatography[23,24].

Some alanine-inserted variants were accumulated as

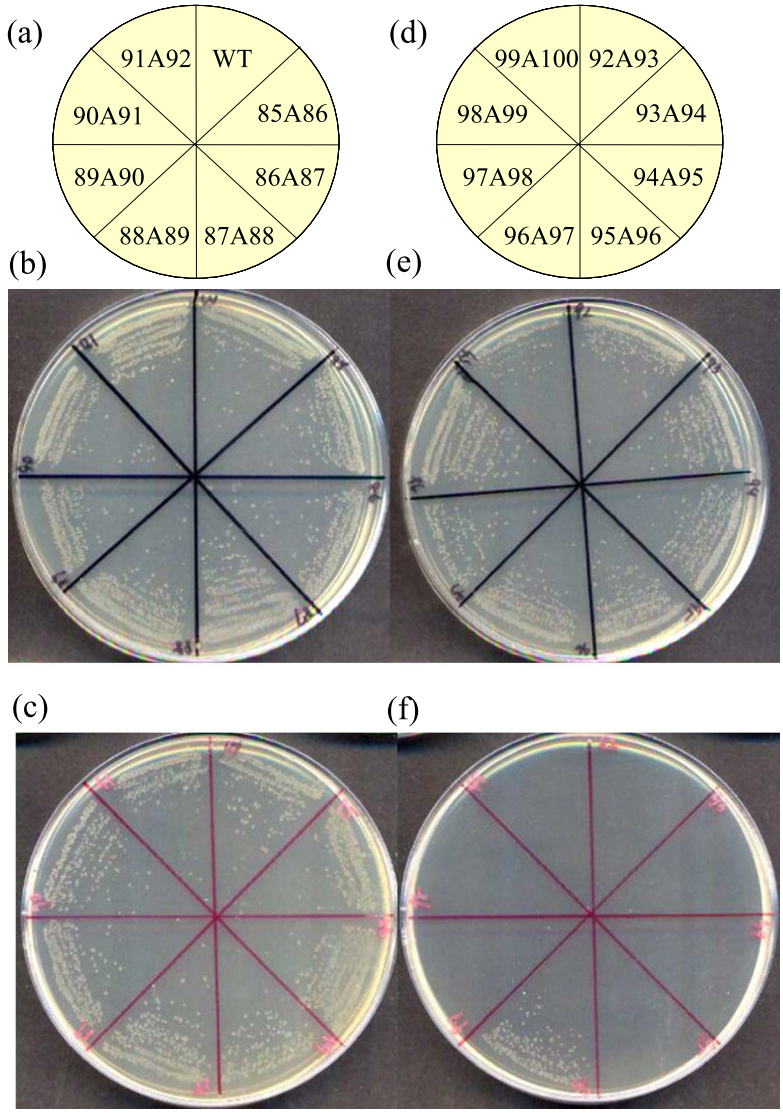

Figure 1 Examples of trimethoprim (TMP) resistance assay. The type of mutant contained in each transformant is shown in (a) and (d). Panels (b) and (e) show the colony formation of each transformant without TMP. Panels (c) and (f) indicate the colony formation of each transformant in media containing 1 μg/ml TMP.

inclusion bodies. Inclusion bodies were collected by centrifugation after cell disruption and dissolved in 10 mM potassium phosphate buffer A (10 mM potassium phosphate, 0.25 mM EDTA, 1 mM DTT) containing 8 M urea (pH 7.0)[14]. After 1-hour incubation, the urea concentration was lowered to 1 M by dilution with 10 mM potassium phosphate buffer A (pH 7.0). The solubilized inclusion bodies were subjected to MTX affinity chromatography. The elution buffer was 10 mM potassium phosphate buffer B (10 mM potassium phosphate, 1 mM EDTA, 10 mM β-mercaptoethanol) containing 5 mM folate and 1 M KCl (pH 9.8). The resultant eluate was dialyzed against 10 mM potassium phosphate buffer C (10 mM potassium phosphate, 0.2 mM EDTA, 1 mM β-mercaptoethanol) containing 1 M KCl (pH 9.8) and further dialyzed against 10 mM potassium phosphate buffer C (pH 7.8). Protein concentration was determined by the absorbance at 280 nm using the extinction coefficient (ε_{280} = 31100 $M^{-1}cm^{-1}$)[25].

Far-UV CD spectra

Far-UV circular dichroism (CD) spectra were measured with a JASCO model J-820 spectropolarimeter. The sample concentration was adjusted to 0.2 mg/ml in 10 mM potassium phosphate buffer C (pH 7.8). The sample was put into a quartz cell with a light path length of 1 mm. The temperature of the sample was kept at 20°C using circulating temperature-controlled water[14].

Activity assay

Enzymatic activity was measured as described by Iwakura

et al[20,21]. Briefly, the disappearance of NADPH and DHF was monitored by measuring the absorbance at 340 nm for 10 minutes on a JASCO V-650 spectrophotometer at 20°C. The reaction solution contained 5.56 nM DHFR, 50 μM DHF and 60 μM NADPH, in 10 mM potassium phosphate buffer C (pH 7.8).

Results and Discussion

Effect of alanine insertion on TMP resistance

The effect of each alanine insertion on DHFR activity was screened using the TMP resistance assay. This screening method is based on overproduction of active DHFR transformed cells[21,22]. Since DHFR is essential for growth in *E. coli*, TMP, a competitive inhibitor of DHFR, suppresses colony formation. If a plasmid encoding wild-type DHFR or an active DHFR mutant is transformed to the *E. coli* cell, the transformant will be TMP-resistant because of overproduction of active enzyme to overcome TMP inhibition. On the other hand, if the transformant contains inactive DHFR, it will never form colonies. Figure 1 shows examples of the TMP resistance assay for alanine insertion mutants. A plate was divided into eight compartments, and one transformant was streaked in each compartment, as shown in Figure 1. The type of mutant expressed by each transformant is described in Figures 1(a) and (d). Figures 1(b) and (e) show plates without TMP. All transformants form colonies, confirming that cells are growing normally. Figures 1(c) and (f) show plates containing 1 μg/ml TMP. It is clear that some transformants as well as the wild type form colonies, whereas the other transformants are TMP-sensitive. DHFR mutants produced in TMP-resistant transformants are active, while the TMP-sensitive transformants produce inactive mutants. We can conclude that 92A93, 93A94, 94A95, 95A96, 97A98, 98A99 and 99A100 are inactive. The result strongly indicates that alanine cannot be inserted into these positions in order to produce an active DHFR. Another interesting and significant finding is that the insertion effect appears in a series of consecutive points in the primary structure. If the mutant iA(i+1) creates a TMP-sensitive transformant, the i-th and the (i+1)-th residues need to be connected and cannot accept alanine insertion or any breakage.

All the insertion mutants were examined by the TMP resistance assay. We found that 36 transformants among the possible 145 were TMP sensitive. The insertion points appeared in 12 different regions on the primary structure (Fig. 7): region a, L4-I5; region b, D27-L28; region c, F31-L36; region d, D37-W47; region e, L54-P55; region f, G56-R57; region g, K58-N59; region h, M92-G96; region i, G97-P105; region j, Q108-H114; region k, D122-T123 and region l, F125-P126. These regions are considered to be essential for enzymatic activity and/or DHF/TMP binding.

Effect of alanine insertion on structure formation of DHFR

The loss of enzymatic activity suggests the loss of foldability. The solution structures of alanine insertion mutants should be examined in order to confirm this assumption. To screen the structures of the numerous mutants, we applied the solubility assay. The idea is that the foldable mutants are soluble, but the unstructured mutants are not soluble and therefore form inclusion bodies. After centrifugation of the fractured cells, the supernatant and precipitant were subjected to SDS-PAGE. The gels were stained with Coomassie Fluor™ Orange Gel Stain.

Figure 2 shows typical examples of the results of SDS-PAGE followed by fluorescent stain. The deeply stained band is due to DHFR. Whole-cell lysate was also subjected to PAGE to confirm expression of the mutant. For wild type and 1A2, the fluorescent intensity of the DHFR band of

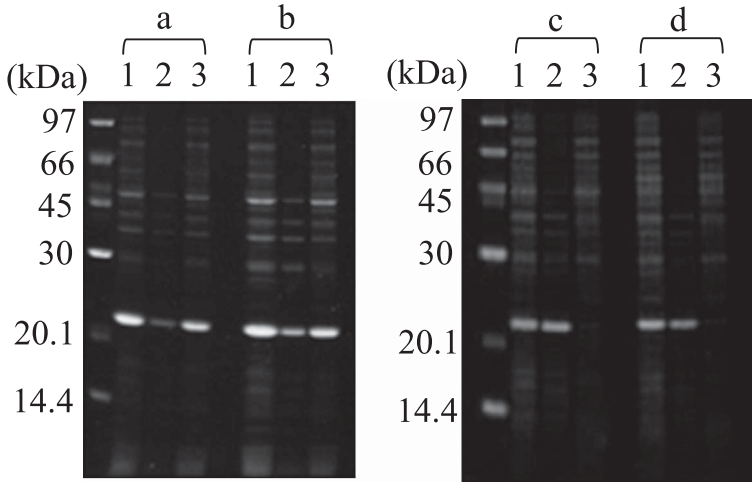

Figure 2 Example of the solubility assay with fluorescent staining of SDS-PAGE. a, wild type, b, 1A2, c, 112A113 and d, 113A114. Lane 1, whole-cell; 2, precipitant; 3, supernatant. The left lane of each panel is the molecular weight marker, and the corresponding molecular weights are given.

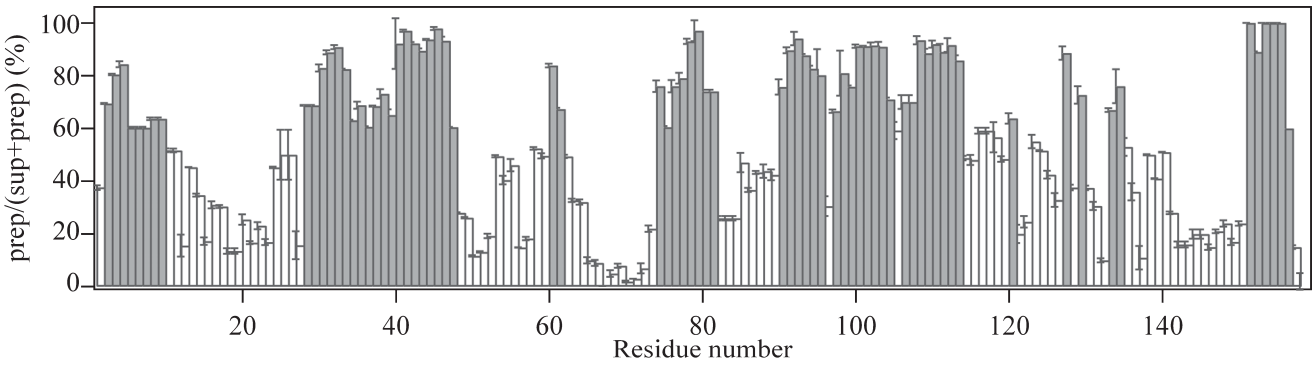

Figure 3 The results of the solubility assay. The ratio of the fluorescent intensity of the precipitant band to the sum of the intensities of the precipitant and supernatant bands is represented as a vertical bar at the mutation site. "Structure elements" are shaded in this figure.

supernatant is much higher than that of the precipitant, indicating that these are soluble. On the other hand, the precipitant DHFR band only shows fluorescence for 112A113 and 113A114, indicating that these are insoluble. The ratio of the fluorescence intensity of the precipitant to that of the total fluorescent intensity was examined for all possible alanine insertion mutants. Figure 3 shows a characteristic pattern. There are several regions on the primary structure that exhibit a high precipitant ratio, similar to the result of TMP resistance assay. We assume that a region with a high precipitant ratio is essential for structure formation, and that alanine cannot be inserted in these regions.

In order to confirm that a mutant with high precipitant ratio is unstructured, we randomly selected mutants to examine their solution structures by CD. Figure 4(a) shows examples of CD spectra of wild type and some alanine insertion mutants. Wild type, 73A74 and 85A86 have low precipitant ratios (lower than 40%). 47A48 and 74A75 have precipitant ratios higher than 60%, and 25A26 has a medium ratio (50%). We can conclude that 73A74 and 85A86 take native conformation, although the spectral shapes are slightly different from that of wild type. On the other hand, the CD spectra of 47A48 and 74A75 are typical spectra for denatured structures. The difference may be due to the loss of the exciton coupling of two tryptophans observed in the wild type[26]. The CD spectrum of 25A26 possesses both structured and unstructured properties. The CD values at 203 nm are plotted against the precipitant ratio in Figure 4(b). From the figure, we set the boundary between being foldable and unfoldable at a precipitant ratio of 60%. The alanine-insertion sites that break protein tertiary structure form contiguous regions in the primary sequence. We assumed that these regions undergo interactions that are essential in order to form the tertiary structure. We can derive 12 such regions: region 1, I2-V10; region 2, L28-E48; region 3, I60-L62; region 4, W74-I82; region 5, E90-G96; region 6, G97-P105; region 7, K106-H114; region 8, E120-G121; region 9, D127-Y128; region 10, E129-P130; region 11, W133-S135; and region 12, Y151-E157. These regions are shaded in Figure 3.

"Structure Element" and "Function Element"

Some of the regions a–l derived by the TMP assay are included in the regions 1–12 derived by solubility assay (Fig. 7). The similarity of the regions derived by the two different methods confirms the aforementioned assumption that the loss of enzymatic activity is due to the loss of foldability. Thus, we termed the regions obtained by solubility assay "Structure Elements."

However, there are also small differences between the two. For example, no corresponding regions are observed by TMP assay for the regions 3–4 or 8–12 derived by solubility assay. Thus, the alanine insertion mutants in these regions are unstructured under physiological conditions but have enzymatic activity or TMP binding ability. The discrepancy can be explained by a model in which TMP binding induces folding into a native conformation. Such inhibitor-induced folding is actually observed for staphylococcal nuclease mutants[27–31]. There is another type of discrepancy: no corresponding regions derived by solubility assay are observed for regions e–g and k–l derived by the TMP resistance assay; these mutants have undergone loss of activity without loss of structure. Therefore, we hypothesize that regions derived by TMP resistance assay are responsible for the expression of enzymatic activity. Thus, we termed the regions obtained by TMP resistance assay "Function Elements."

We measured the enzymatic activity of several mutants to assess whether the TMP resistance assay is an adequate functional assay of DHFR. Figure 5 shows the reaction curves of enzymatic activity for wild type and the insertion mutants. It is clear that all the mutants expressed in the TMP-sensitive transformants have little or no activity, as shown in Figure 5(a), confirming that in these cases, the "Function Elements" determined by TMP resistance assay are valid. The reaction curves for the mutants expressed in the TMP-resistant transformants are shown in Figures 5(b) and (c). Some mutants are active (Fig. 5(b)), supporting the validity of the TMP resistance assay. However, the other mutants are essentially inactive (Fig. 5(c)). We hypothesize that these mutants maintain TMP binding ability but lose either catalytic activity or NADPH binding ability. There-

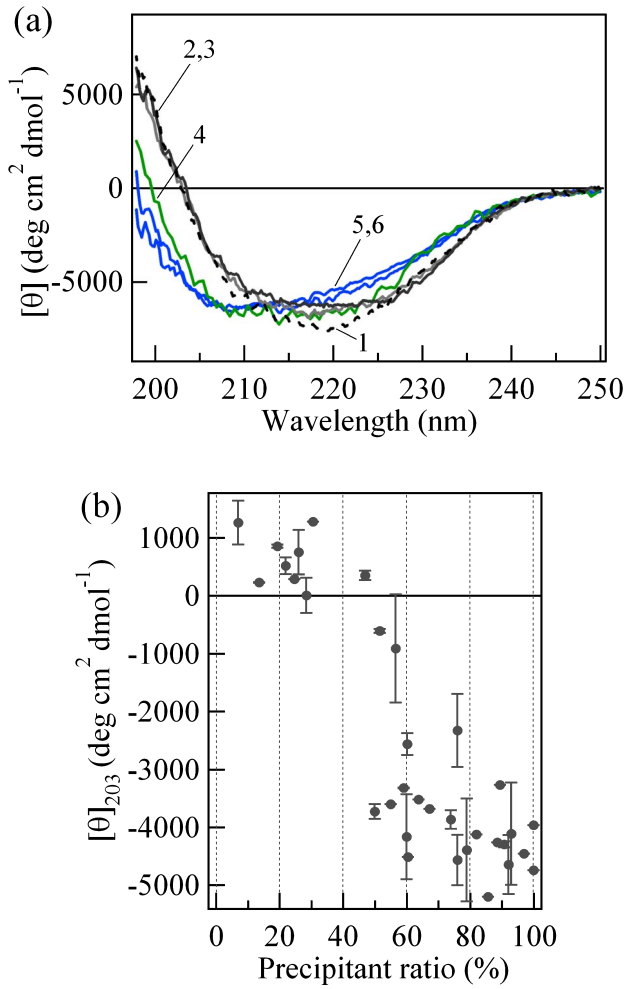

Figure 4 Solution structures of some alanine insertion mutants and the comparison with the solubility assay. (a) Far-UV CD spectra of wild type and alanine insertion mutants. Curves 1–6 represent wild type, 73A74, 85A86, 25A26, 47A48 and 74A75, respectively. DHFR concentration is 0.2 mg/ml in 10 mM potassium phosphate buffer C (pH 7.8) at 20°C. (b) Ellipticity at 203 nm is plotted against the precipitant ratio shown in Figure 4. The ellipticity is almost 0 for the folded DHFR, whereas it is large and negative for unstructured DHFR.

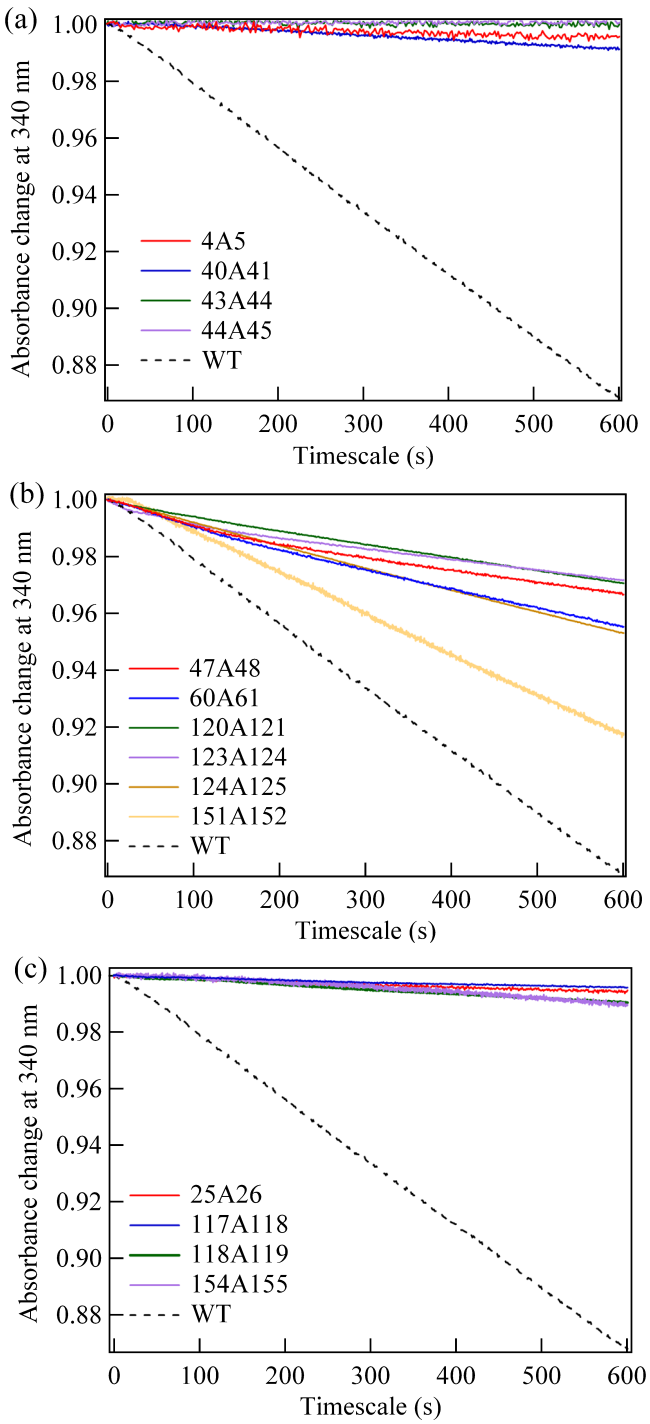

Figure 5 The enzymatic activity of wild type and some alanine insertion mutants. Enzymatic activity was monitored by absorbance change at 340 nm. (a) The reaction curves of TMP-sensitive mutants. (b) The reaction curves of some TMP-resistant mutants. (c) The reaction curves of the other TMP-resistant mutants.The reaction solution contains 5.56 nM DHFR, 50 μM DHF and 60 μM NADPH, in 10 mM potassium phosphate buffer C (pH 7.8). Reactions were performed at 20°C.

fore, the TMP resistance assay can only evaluate the TMP binding ability, and is insufficient to identify the complete "Function Elements."

Based on these analyses, we propose that DHFR is composed of essential regions for structure formation, function expression and linkers connecting the elements. The "Structure Elements" and "Function Elements" are mapped onto the DHFR structure in Figure 6. These elements are the building blocks of protein architecture.

Properties of two types of elements

The "Structure" and "Function Elements" are defined as regions where alanine insertion cannot be tolerated. If the perturbation brought about by the alanine insertion is locally relaxed at the inserted point, the overall structure is

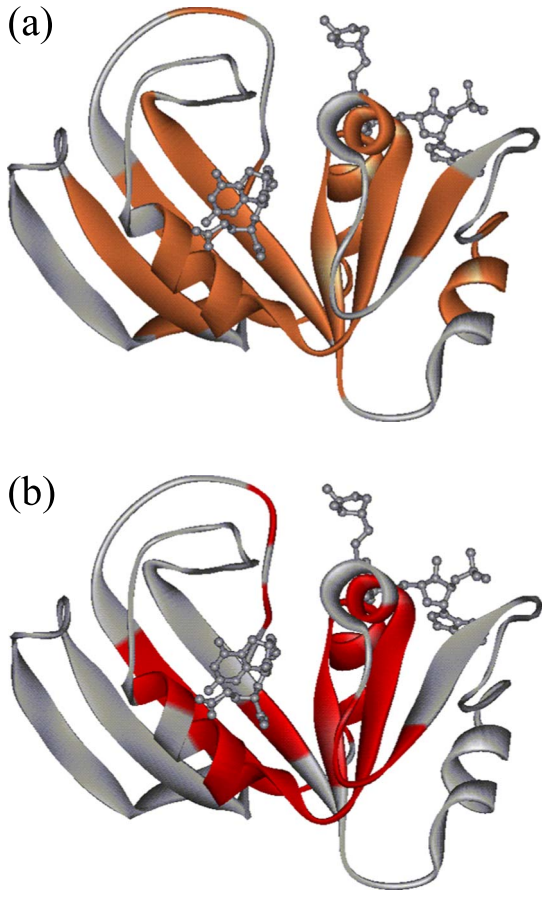

Figure 6 Mapping of the "Structure Elements" (a) and "Function Elements" (b) onto the structure of DHFR (PDB code: 1RX4). This figure was prepared using the program Weblab Viewer Pro[18].

not affected seriously. However, if the perturbation or the distortion by alanine insertion into these elements cannot be tolerated locally at the inserted point, then the perturbation propagates to the overall proteins. For example, if the distortion cannot be relaxed locally, insertion of a single residue into an α-helix causes a 100° rotation and 1.5 Å shift for the residues following or preceding the insertion point. Single-residue insertion into a β-strand will completely reverse the orientation of the residues following the inserted position. The structural properties of DHFR are compared with the "Structure Elements." "Structure Elements" do not necessarily correspond to secondary structure components, e.g., α-helices and β-strands. For example, the N-terminal part of the first helix (αB) and the C-terminal half of the second helix (αC) do not belong to the "Structure Elements." The N-terminal parts of βC and βD and the C-terminal half of βG also do not fall within the "Structure Elements." The "Structure Elements" 8–10 contain neither α-helix nor β-sheet. These observations suggest that the relaxation of the distortion brought about by alanine insertion does not depend on secondary structure components.

Pulse-labeling hydrogen exchange experiments revealed that 13 amide hydrogens were protected within 6 msec of the initiation of the refolding reaction of wild-type DHFR[32]. These early folding sites are all involved in the "Structure Elements" (Fig. 7). Therefore, some of the "Structure Elements" play important roles in the formation of the folding core. However, the "Structure Element" 3 and 8–11 do not contain initiation sites. These "Structure Elements" would be responsible for the stability of the native conformation. We assume that the latter "Structure Elements" are responsible for the non-local interactions generated between different elements.

The "Function Element" a, b–d, and h–j overlap with the "Structure Elements." Therefore, alanine insertion into these elements causes loss of the activity due to the loss of structure formation. The "Function Element" e and f, which do not correspond to the "Structural Elements", contain residues responsible for the substrate (DHF) binding sites. Alanine insertion into these regions may destroy the proper configuration of the binding site without the loss of the tertiary structure. The "Function Elements" k, which is not a "Structure Element", does not contain the DHF-binding site, but contains two residues responsible for NADPH-binding. On the other hand, the "Function Element" g and l contain neither DHF- nor NADPH-binding sites, suggesting that these "Function Elements" are responsible for catalytic activity. In order to define the "Function Elements" completely, we need a more quantitative analysis of the effect of alanine insertion on enzymatic activity.

Iwakura and coworkers proposed the concept of the folding elements (FE's) by systematic circular permutation analysis of DHFR[14]. The FE's were assigned to contiguous regions where cleavage abolishes enzymatic function[14]. The assigned FE's essentially coincide with the "Structure Elements" and the "Function Elements" identified in this study. In particular, the coincidence between the "Structure Elements" and FE's are remarkable. The connectivity and contiguity of amino acid sequences of these elements are essential for maintenance of the native tertiary structure. The properties of the "Structure Elements" share the common properties with the FE's. Therefore, the alanine insertion analysis gives essentially identical information to that obtained from circular permutation analysis.

There are some trivial differences between "Structure Element" and FE (Fig. 7). For example, "Structure Element" 2 covers both FE2 and FE3. "Structure Element" 3 is narrower than FE4, while "Structure Element" 4 is wider than FE5. By circular permutation, the cleavage of the amino acid sequence is rigorously performed[14]. Alanine insertion is milder than circular permutation, because new N- and C-termini do not appear at the cleavage site[15,16]. However, the degrees of freedom at the insertion point are more constrained than when the backbone is completely cleaved in circular permutation. These differences will doubtlessly lead to subtle differences in the determination of the elemental boundaries.

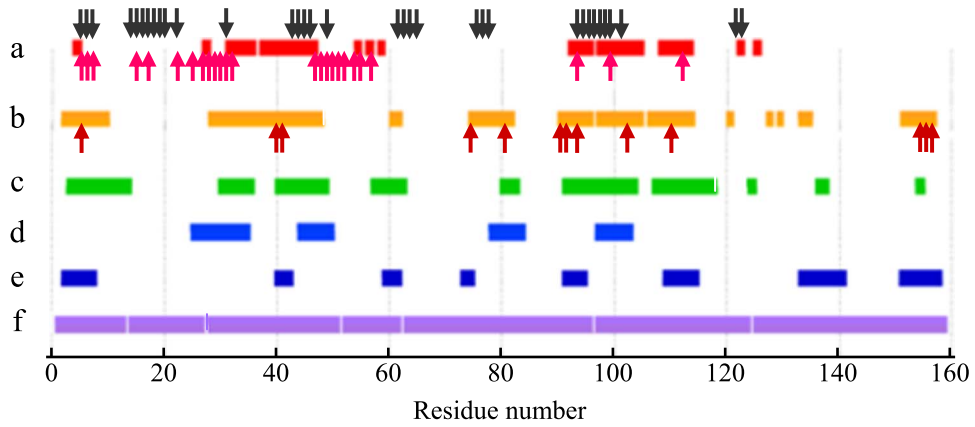

Figure 7 Comparison of the "Structure Elements" and "Function Elements" with some structural properties of DHFR. (a) "Function Elements": a, L4-I5; b, D27-L28; c, F31-L36; d, D37-W47; e, L54-P55; f, G56-R57; g, K58-N59; h, M92-G96; i, G97-P105; j, Q108-H114; k, D122-T123; l, F125-P126.(b) "Structure Elements": 1, I2-V10; 2, L28-E48; 3, I60-L62, 4, W74-I82; 5, E90-G96; 6, G97-P105; 7, K106-H114; 8, E120-G121; 9, D127-Y128; 10, E129-P130; 11, W133-S135; 12, Y151-E157. (c) Folding elements[14]: 1, S3-I14; 2, W30-L36; 3, V40-S49; 4, R57-S63; 5, E80-A83; 6, I91-L104; 7, A107-E118; 8, H124-F125; 9, V136-S138; 10, E154-I155. (d) α-helices: αB, P25-T35; αC, R44-I50; αD, V78-A84; αE, G97-F103. (e) β-sheets: βA, I2-L8; βB, V40-G43; βC, N59-L62; βD, T73-V75; βE, I91-G95; βF, K109-I115; βG, W133-H141; βH, Y151-R158. (f) Modules: M1, M1-V13; M2, I14-D27; M3, L28-G51; M4, R52-L62; M5, S63-G96; M6,G97-H124; M7, F125-R159. DHF-binding sites (pink arrow), NADPH-binding sites (black arrow) and early-folding sites[32] (brown arrow) are also shown.

"Structure and Function Elements" as constituents of a module

A "module" is a small compact structural unit of a globular protein[3,33]. A globular protein is composed of several modules[3,33,34]. Since the module boundaries are closely correlated to the intron positions of the corresponding genes, the module structure is one of the key factors in evolution of proteins by exon shuffling[3]. DHFR is composed of 7 modules (Fig. 7), as determined by distance map, which is the original method introduced by Go[3].

We compared the "Structure and Function Elements" with the modules of DHFR. All the "Structure Elements" are contained within a single module, i.e., no "Structure Elements" are shared by two modules. Module 2 contains no "Structure Elements". "Structure Elements" 5 and 6 are consecutive. Alanine insertion can be only accepted at the position between the 96th and the 97th residues, defining the boundary between "Structure Elements" 5 and 6. The position is exactly the same as the module boundary. On the other hand, each module contains one or multiple "Structure Elements" except for module 2. The "Structure Elements" are closely related to the module structure. We hypothesize that the "Structure Elements" are responsible for inter-module interaction.

Each "Function Element" belongs to a single module as well, except for "Function Element" b. Each module contains one or more "Function Elements" except for modules 2. Module 2 contains no "Function Elements". In some proteins, each module can take a stable conformation[35]. Sometimes, some modules also possess substantial function[36]. These modules can be referred to as "structural modules" or "functional modules"[37]. We hypothesize that the properties of modules would be defined by the elements involved. If a module contains either "Structure Elements" or "Function Elements", the module serves as either a structural or functional module. "Function Element" b contains only 2 residues, D27 and L28, which are the boundary between modules 2 and 3. This is an exception for the observation that each "Function Element" never spans two modules. One possible explanation for the exception would be the ambiguity in the determination of module boundaries. In the present case, the boundary can also be located between L28 and A29. Another possibility is that the combination of two modules during evolution creates a new function that appears at the new functional site at the module boundary. In the case of barnase, the key residues for catalysis are located at the module boundaries[36]. Alanine insertion at the module boundary may not be tolerated in such a case.

We demonstrate an interesting relationship between the elements and the modules. If modules are the essential building blocks, "Structure Elements" are the connectors between two modules or the stabilizers of a module. "Function Elements" would confer functional properties onto a module.

The effectiveness of alanine insertion

The circular permutation method provides useful information about protein architecture and folding[14]. We showed that systematic alanine insertion gives essentially identical information to that obtained from circular permutation analysis. In the case of circular permutation, the original N-terminus and the original C-terminus must be connected with an appropriate linker[12]. Therefore, the method can only be applied to proteins whose N- and C-termini are close to one another. The design of an appropriate linker is also difficult. On the other hand, alanine insertion never has such

difficulties. The method can be applied to every protein of interest. Therefore, systematic alanine insertion is a unique and useful method for the dissection of protein architecture. In the case of DHFR, circular permutants that were produced in TMP-sensitive cells could not be purified[13,14]. However, all alanine insertion mutants, even those produced in the TMP-sensitive cells, could be overexpressed and purified. Therefore, the alanine insertion is also advantageous over the circular permutation in this regard. We are now applying the method to other functional proteins to reveal the design principles of protein architecture.

Acknowledgements

We thank Ms. Satomi Kayamura for her technical assistance. This work is partly supported by the Grant-in-Aid for Scientific Research (B) to MK (20370062) and the Grant-in-Aid for Scientific Research on Innovative Areas to MK (20107006). RS is grateful to the financial support by the Program for Enhancing Systematic Education in Graduate Schools (H006) from the Graduate School of Materials Science, Nara Institute of Science and Technology.

REFERENCES

1. Anfinsen, C. B. & Scheraga, H. A. Experimental and theoretical aspects of protein folding. *Adv. Protein Chem.* **29**, 205–300 (1975).
2. Go, N. Theoretical studies of protein folding. *Ann. Rev. Biophys. Bioeng.* **12**, 183–210 (1983).
3. Onuchic, J. N., Luthey-Schulten, Z. & Wolynes, P. G. Theory of protein folding: The energy landscape perspective. *Ann. Rev. Phys. Chem.* **48**, 545–600 (1997).
4. Dill, K. A. & Chan, H. S. From levinthal to pathways to funnels. *Nat. Struct. Biol.* **4**, 10–19 (1997).
5. Go, M. Correlation of DNA exonic regions with protein structural units in haemoglobin. *Nature* **291**, 90–92 (1981).
6. Riddle, D. S., Santiago, J. V., Bray-Hall, S. T., Doshi, N., Grantcharova, V. P., Yi, Q. & Baker, D. Functional rapidly folding proteins from simplified amino acid sequences. *Nat. Struct. Biol.* **4**, 805–809 (1997).
7. Kuroda, Y. & Kim, P. S. Folding of bovine pancreatic trypsin inhibitor (BPTI) variants in which almost half the residues are alanine. *J. Mol. Biol.* **298**, 493–501 (2000).
8. Shirai, K., Yamazaki, Y., Kamikubo, H., Imamoto, Y. & Kataoka, M. Attempt to simplify the amino-acid sequence of photoactive yellow protein with a set of simple rules. *Proteins: Struct. Funct. Bioinformat.* **67**, 821–833 (2007).
9. Goldenberg, D. P. & Creighton, T. E. Circular and circularly permuted forms of bovine pancreatic trypsin inhibitor. *J. Mol. Biol.* **165**, 407–413 (1983).
10. Goldenberg, D. P. Circularly permuted proteins. *Protein Eng.* **2**, 493–5 (1989).
11. Luger, K., Hommel, U., Herold, M., Hofsteenge, J. & Kirschner, K. Correct folding of circularly permuted variants of a beta alpha ballel enzyme in vivo. *Science* **243**, 206–210 (1989).
12. Iwakura, M. & Nakamura, T. Effects of the length of glycine linker connecting the N- and C-termini of a circiularly permuted dihydrofolate reductase. *Protein Eng.* **11**, 707–713 (1998).
13. Nakamura, T. & Iwakura, M. Circular permutation analysis as a method for distinction of functinal elements in the M20 loop of *Escherichia coli* dihydrofolate reductase. *J. Biol. Chem.* **274**, 19041–19047 (1999).
14. Iwakura, M., Nakamura, T., Yamane, C. & Maki, K. Systematic circular permutation of an entire protein reveals essential folding elements. *Nat. Struct. Biol.* **7**, 580–585 (2000).
15. Sondek, J. & Shortle, D. Accommodation of single amino acid insertions by the native state of staphylococcal nuclease. *Proteins: Struct. Funct. Bioinformat.* **7**, 299–305 (1990).
16. Kataoka, M., Yunoki, J. & Takahara, H. Effects of deletions or insertions on the structure, stability and function of staphylococcal nuclease. In "Old and New Views of Protein Folding" (eds. by K. Kuwajima and M. Arai) pp. 205–214, Elsevier, Amsterdam (1999).
17. Bystroff, C. & Kraut, J. Crystal structure of unliganded *Escherichia coli* dihydrofolate reductase. Ligand-induced conformational changes and cooperativity in binding. *Biochemistry* **30**, 2227–2239 (1991).
18. Sawaya, M. R. & Kraut, J. Loop and subdomain movements in the mechanism of *Escherichia coli* dihydrofolate reductase: crystallographic evidence. *Biochemistry* **36**, 586–603 (1997).
19. Ohshiro, T., Kuge, Y., Igarashi, A., Mochida, K., Iwakura, M. & Uwajima, T. Production of dihydrofolate reductase by cloned *Escherichia coli* and its application to asymmetric synthesis of l-leucovorin. *Biosci. Biotechnol. Biochem.* **56**, 437–440 (1992).
20. Iwakura, M., Jones, B. E., Luo, J. & Matthews, C. R. A strategy for testing the suitability of cysteine replacements in dihydrofolate reductase from *Escherichia coli*. *J. Biochem.* **117**, 480–488 (1995).
21. Iwakura, M., Shimura, Y. & Tsuda, K. Cloning of dihydrofolate reductase gene of *Escherichia coli* K12. *J. Biochem.* **91**, 1205–1512 (1982).
22. Iwakura, M., Shimura, Y. & Tsuda, K. Construction of plasmid vectors for cloning of promoters using the dihydrofolate reductase gene of *Escherichia coli* K-12. *J. Biochem.* **93**, 927–930 (1983).
23. Iwakura, M., Furusawa, K., Kokubu, T., Ohashi, S., Tanaka, Y., Shimura, Y. & Tsuda, K. Dihydrofolate reductase as a new "Affinity Handle". *J. Biochem.* **111**, 37–45 (1992).
24. Arai, M. & Iwakura, M. Probing the interactions between the folding elements early in the folding of *Escherichia coli* dihydrofolate reductase by systematic sequence perturbation. *J. Mol. Biol.* **347**, 337–353 (2005).
25. Touchette, N. A., Perry, K. M. & Matthews, C. R. Folding of dihydrofolate reductase from *Escherichia coli*. *Biochemistry* **25**, 5445–5452 (1986).
26. Kuwajima, K., Garvey, E. P., Finn, B. E., Matthews, C. R. & Sugai, S. Transient intermediates in the folding of dihydrofolate reductase as detected by far-ultraviolet circular dichroism spectroscopy. *Biochemistry* **30**, 7693–7703 (1991).
27. Anfinsen, C. B. The formation and stabilization of protein structure. *Biochem. J.* **128**, 737–749 (1972).
28. Flanagan, J. M., Kataoka, M., Shortle, D. & Engelman, D. M. Truncated staphylococcal nuclease is compact but disordered. *Proc. Natl. Acad. Sci.* **89**, 748–752 (1992).
29. Hirano, S., Mihara, K., Yamazaki, Y., Kamikubo, H., Imamoto, Y. & Kataoka, M. Role of C-terminal region of staphylococcal nuclease for foldability, stability, and activity. *Proteins: Struct. Funct. Bioinformat.* **49**, 255–265 (2002).
30. Hirano, S., Kamikubo, H., Yamazaki, Y. & Kataoka, M. Elucidation of information encoded in tryptophan 140 of staphylococcal nuclease. *Proteins: Struct. Funct. Bioinformat.* **58**, 271–277 (2005).
31. Onitsuka, M., Kamikubo, H., Yamazaki, Y. & Kataoka, M. Mechanism of induced folding: Both folding before binding

and binding before folding can be realized in staphylococcal nuclease mutants. *Proteins: Struct. Funct. Bioinformat.* **72**, 837–847 (2008).
32. Bryan, B. E. & Matthews, C. R. Early intermediates in the folding of dihydrofolate reductase from *Escherichia coli* detected by hydrogen exchange and NMR. *Protein Sci.* **4**, 167–177 (1995).
33. Go, M. Modular structural units, exons, and function in chicken lysozyme. *Proc. Natl. Acad. Sci.* **80**, 1964–1968 (1983).
34. Noguti, T., Sakakibara, H. & Go, M. Localization of hydrogen-bonds within modules in barnase. *Proteins: Struct. Funct. Genet.* **16**, 357–363 (1993).
35. Ikura, T., Go, N., Kohda, D., Inagaki, F., Yanagawa, H., Kawabata, M., Kawabata, S., Iwanaga, S., Noguti, T. & Go, M. Secondary structural features of modules M2 and M3 of barnase in solution by NMR experiment and distance geometry calculation. *Proteins: Struct. Funct. Genet.* **16**, 341–356 (1993).
36. Yanagawa, H., Yoshida, K., Torigoe, C., Park, J., Sato, K., Shirai, T. & Go, M. Protein anatomy: Functional roles of barnase module. *J. Biol. Chem.* **268**, 5861–5865 (1993).
37. Yura, K., Shionyu, M., Kawatani, K. & Go, M. Repetitive use of a phosphate-binding module in DNA polymerase β, Oct-1 POU domain and phage repressor. *Cell Mol. Life Sci.* **55**, 472–486 (1999).