J. Biochem. 2009;145(6)751–762
doi:10.1093/jb/mvp034

Sequence Perturbation Analysis: Addressing Amino Acid Indices to Elucidate the C-Terminal Role of *Escherichia Coli* Dihydrofolate Reductase

Hisashi Takahashi†, Akiko Yokota†, Tatsuyuki Takenawa and Masahiro Iwakura*

Protein Design Research Group, Institute for Biological Resources and Functions, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba Central 6, 1-1-1 Higashi, Tsukuba, Ibaraki 305-8566, Japan

Received January 19, 2009; accepted February 19, 2009; published online March 2, 2009

Because amino acid residues intrinsically possess many factors participating in protein structures and functions, to determine main (or unique) factors at a specific site in a protein sequence should be of great help for understanding how a protein obtains its structure and function. In this study, we proposed a means of sequence perturbation analysis to address the above concerns involving comprehensive AA indices. We constructed all 19 possible single mutant proteins as to the three sites in the C-terminal of *Escherichia coli* dihydrofolate reductase (DHFR), and measured the activity and thermal stability of each of all the single mutant proteins. The significantly perturbed properties with each systematic single mutation at each mutational site were examined in terms of the linear correlation with each AA index. As a result, at each of Arg158 and Arg159 of DHFR, the AA index for the isoelectric points of amino acids showed strong correlation with the transition temperature of thermal denatuation, suggesting that the electrostatic interaction is the main factor influencing the C-terminal role of the DHFR. The feasibility and general versatility of our sequence perturbation analysis were also examined by application to other sites of DHFR.

Key words: amino acid indices, C-terminal role, dihydrofolate reductase, mutational sensitivity, sequence perturbation analysis.

Abbreviations: CD, circular dichroism; DHF, dihydrofolate; DHFR, dihydrofolate reductase; MRE, mean residue ellipticity; MTX, methotrexate; SVD, singular value decomposition; THF, tetrahydrofolate; TMP, trimethoprim; UV, ultraviolet.

Understanding how an amino acid sequence determines the tertiary structure, function and stability of a protein is still one of the challenging problems in life science research. Owing to the huge number of studies aimed at solving these challenging problems, there has been substantial progress in the field of protein structure predictions, as exemplified by Critical Assessment of Techniques for Protein Structure Prediction (CASP) experiments (*1*).

According to Anfinsen's experiment, all of the information needed to realize a protein structure and function under specified solution conditions is encoded in a linear sequence (*2*) of 20 naturally occurring amino acid residues as building blocks. A wide variety of physicochemical and biochemical factors of amino acid residues have been extensively investigated and each factor is expressed as an amino acid index (AA index) that is represented by a set of 20 numerical values. More than 500 AA indices have been proposed in various literature, and they have been collected and released as an online database, named the 'AAindex' database (*3*, *4*) (http://www.genome.jp/aaindex/). Since protein structures and functions are defined by combinations of the properties of amino acid residues, an observed protein property, such as a structural parameter, a stability parameter and a functional parameter, can be expressed as a function of AA indices as versatile numerical variables.

As reported before (*5*), we proposed a means of sequence perturbation analysis which comprises constructing all possible single mutant proteins as to a specified site, measuring a specific parameter for a property of the mutant proteins, and evaluating the observed changes versus single amino acid replacements. Such study has been demonstrated to be useful for the assignment of structural formation during protein folding from a denatured state to the native structure (*5*). Usually, a single amino acid replacement has only small effects on entire protein properties such as the tertiary structure, the function, the stability and so on of the protein, therefore, the replacement can be treated as small perturbations of the protein properties.

Here, we propose a novel approach involving the sequence perturbation concept, extended sequence perturbation analysis, for identifying an AA index which is closely correlated with an observed protein property at a

†The first two authors are equal contributors to this work.
*To whom correspondence should be addressed. Tel: +81-29-861-6179, Fax: +81-29-856-4055, E-mail: masa-iwakura@aist.go.jp

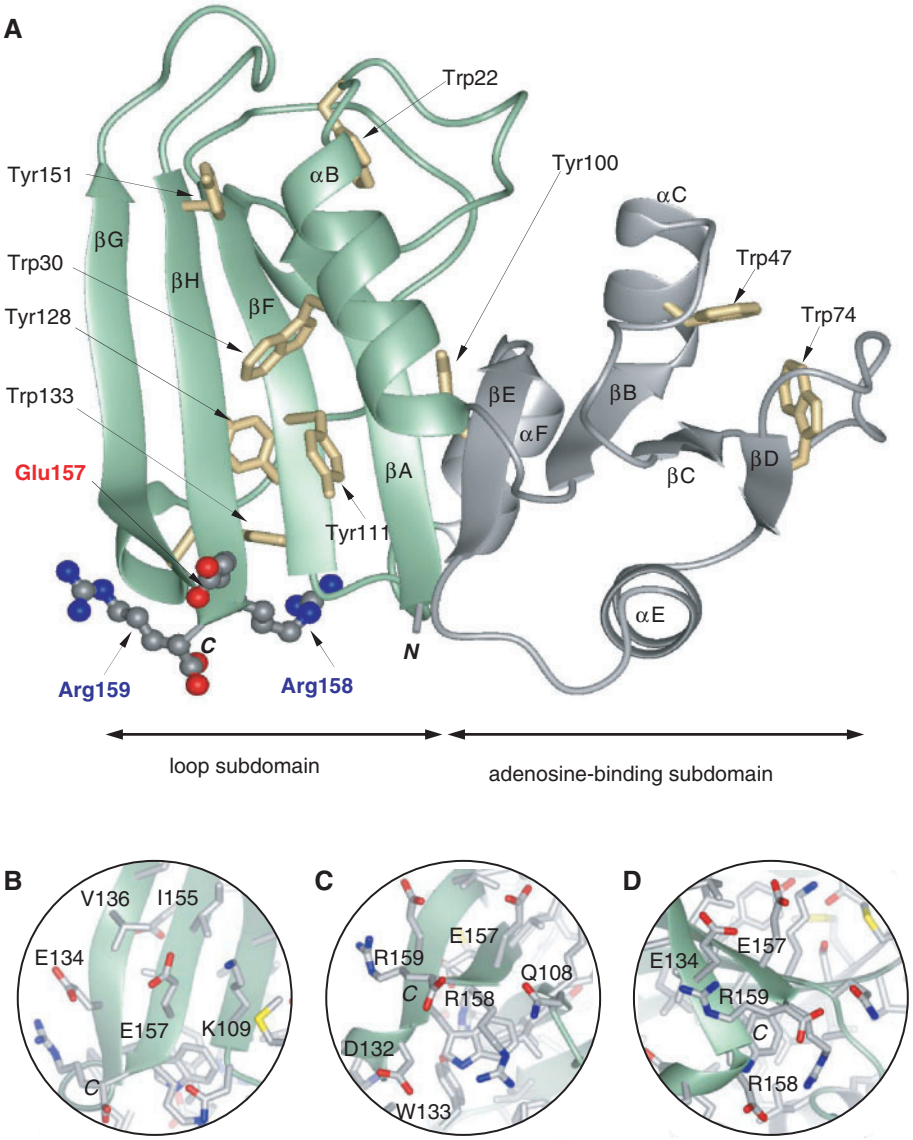

Fig. 1. **Schematic representation of the DHFR structure.** Overall structure (A) and local structures around the C-terminal regions of E157 (B), R158 (C) and R159 (D) of *E. coli* DHFR (PDB code: 1RX4) were prepared with the program MOLMOL (*39*). (A) C-terminal residues (E157, R158 and R159) are shown as ball-and-stick models. Five tryptophan and four tyrosine residues are shown as a stick model (in gold). The left half of the molecule corresponds to the loop subdomain (in light green) and the right half to the adenosine-binding subdomain (in grey). (B–D) The target amino acid residue of each of E157, R158 and R159, and each adjacent residue are represented as stick models in (B), (C) and (D), respectively.

specified site in a protein through extensive single amino acid replacements as sequence perturbation. The change in an observed property at a specified position of an amino acid sequence would be related linearly to a function of a certain AA index or to some combination of certain AA indices. Then, the role of a specified position of an amino acid sequence would be understood through the involved AA index in terms of physicochemical and biochemical properties. To test the feasibility of our idea, we used the C-terminal region of dihydrofolate reductase (DHFR, EC 1.5.1.3) from *Escherichia coli* (Fig. 1) as a model, considering that the usage of C-terminal amino acid residues of naturally occurring proteins is apparently biased (*6–8*), as is the case for DHFR.

Escherichia coli DHFR is a monomeric, 159 residue protein that has been well characterized in terms of structure, function and folding as a model enzyme (*9–11*). The tertiary structure of this enzyme is of the parallel α/β class, consisting of four α-helices and eight β-strands (*12*), and is divided into two parts, the loop subdomain (residues 1–37 and 107–159) and the adenosine-binding subdomain (residues 38–106) (*10*). DHFR catalyses the reduction of dihydrofolate (DHF) to tetrahydrofolate (THF) (*13*) using the reducing cofactor NADPH, and is

an essential enzyme required for normal folate metabolism in prokaryotes and eukaryotes, because it maintains the THF levels needed to support the biosynthesis of purines, pyrimidines and amino acids. DHFR is a clinically important enzyme not only because it is the target of a number of antifolate drugs, such as trimethoprim (TMP) and methotrexate (MTX), but also because it can be used to produce *l*-leucovorin, an anticancer drug, in a stereospecific manner (*14*). DHFR has been used as an affinity handle protein through C-terminal fusion with various polypeptides (*15*).

In this study, we tried to construct all possible 19 single mutants as to three sites in the C-terminus of DHFR, purified them, and measured their activity in terms of k_{cat}, K_m and k_{cat}/K_m, and their thermal stability in terms of the transition temperatures of thermal denaturation, T_{m1} and T_{m2}. Such observed properties of the single mutant proteins were analysed with the concept of our sequence perturbation analysis using the AAindex1 database (*4*). The roles of the C-terminal amino acid residues of DHFR were also discussed based on the results of our analysis.

EXPERIMENTAL PROCEDURES

Mutagenesis—Site-directed random mutagenesis to construct the expression vectors of all the single mutants was carried out by means of a standard PCR method using that of the wild type DHFR as a template DNA (*15*, *16*). The resultant mutant plasmids transformed into *E. coli* JM109 competent cells were screened as to ampicillin- and trimethoprim-resistance.

Protein Purification and Quality Test—DHFR variants were purified mainly by methotrexate affinity chromatography and anion-exchange chromatography after adequate culture and pre-purification steps (*15*, *16*). The purities and molecular weights of the prepared mutant proteins were check by SDS–PAGE and liquid chromatography mass spectroscopic (LC-MS) measurements (*17*), respectively.

Enzymatic Activity Measurements—The steady-state kinetics parameters (k_{cat}, K_m and k_{cat}/K_m) were determined spectrophotometrically at 15°C by following the disappearance of NADPH and DHF estimated from the time course of absorbance at 340 nm ($\varepsilon_{340} = 11{,}800\ M^{-1}\,cm^{-1}$) (*18*) and calculated. The standard assay mixture comprised various concentration of DHF, 100 μM NADPH, 14 mM 2-mercaptoethanol, MTEN buffer (50 mM morpholinoethanesulfonic acid (pH 7.0), 25 mM tris (hydroxymethyl) aminomethane, 25 mM ethanolamine, 100 mM NaCl) and enzymes. The concentration of the wild type was determined spectrophotometrically using the extinction coefficient of $\varepsilon_{280} = 31{,}100\ M^{-1}\,cm^{-1}$ (*19*) and those of mutants were estimated by method of Pace *et al*. (*20*), based on the extinction coefficient for the wild type (*19*).

Thermal Denaturation—Thermal denaturation was monitored using an Aviv 14 DS spectrophotometer by means of the scanning absorption spectrum from 310 to 260 nm with a 2-s integration time. The measurements were carried out in the presence of 10 mM potassium phosphate (pH 7.8), 0.2 mM EDTA and 0.1 mM dithiothreitol (DTT). The temperature was raised by 1°C from 5°C to 90°C with a 2-min equilibration time at each temperature. A series of temperature-dependent absorption spectra A (λ, T_i) taken for a set of n temperature slices {Ti} were analysed using the singular value decomposition (SVD) algorithm and the thermal transition curve fitted to the three-state transition between the native (N), intermediate (I) and unfolded states (U) (*21*). Midpoints of the transition temperatures from the N to I states, and from the I to U states, are defined as T_{m1} (°C) and T_{m2} (°C), respectively. The concentrations of the mutants were determined spectrophotometrically by method of Gill and von Hippel (*22*), based on the extinction coefficient for the wild type (*19*).

CD Measurement—Equilibrium far-ultraviolet circular dichroism (far-UV CD) spectra were obtained on an Aviv 62 DS spectropolarimeter at 15°C by scanning from 250 to 190 nm with a 20-s integration time, in the same buffer as thermal denatuation experiments. The path length of the sample cell was 1 mm. A typical protein concentration was 5.5 μM, and the protein concentration was estimated as described above (in the thermal denatuation experiments). The MRE (mean residue ellipticity) value was calculated using the following equation, $\mathrm{MRE} = \Theta/(C \times D \times NA)$, where Θ is the ellipticity in milli-degrees, C the molar protein concentration, D the path length of the sample cell in centimetre, and NA the number of residues in the protein (159), respectively.

Mutant Data Set—For each measured property of the mutant proteins including the wild-type protein, respective data sets for the specified mutation sites were set, respectively (Table 1). Naturally, the maximum number of the data points in each mutant data set is 20. The statistic parameters (average, variation and standard deviation) were derived by simple statistical analysis.

Reference Data Set—To obtain a reference data set, the enzyme activity measurements and thermal denaturation experiments involving the wild-type DHFR were carried out eight times (Table 1). The measurements were carried out on different days from each other. The statistic parameters (average, variation and standard deviation) were derived by simple statistical analysis. As for the standard deviation (σ) of the reference data set, the value of the 95% confidential coefficient from the t-distribution table was multiplied in consideration of the effect of the small number of data in the reference data set.

Accession and Normalization of AA Indices—From the AAindex1 database (http://www.genome.jp/aaindex/), all listed 544 AA indices (at present, *i.e.*, January, 2009) were downloaded. AA indices are basically composed of a set of 20 numerical values and represent various physicochemical and biochemical properties of amino acids. Each AA index was normalized between 0 and 1 using the following expression and was used for our sequence perturbation analysis;

$$\mathrm{AAindex}_{norm}(i) = \frac{\mathrm{AAindex}(i) - \mathrm{AAindex}_{min}}{\mathrm{AAindex}_{max} - \mathrm{AAindex}_{min}} \qquad (1)$$

where AA index(i), and AA index$_{norm}$(i) are, respectively, the original and normalized values of amino acid i for a

Table 1. **Statistics of the mutant data sets and reference data set.**

Content of data set					
Observed property	Target protein	Number of data in the data set	Average	Variance	SD
k_{cat}	E157X	19	5.26	2.56	1.60
	R158X	19	5.39	0.322	0.567
	R159X	20	5.60	0.871	0.933
	WT (reference)	8	6.18	0.651	1.91[a]
K_m	E157X	19	1.13	0.105	0.325
	R158X	19	1.19	0.106	0.325
	R159X	20	1.50	0.118	0.343
	WT (reference)	8	1.10	0.086	0.693[a]
k_{cat}/K_m	E157X	19	4.80	2.14	1.46
	R158X	19	4.80	1.28	1.13
	R159X	20	3.89	1.13	1.06
	WT (reference)	8	5.62	1.09	2.47[a]
T_{m1}	E157X	19	41.3	10.1	3.19
	R158X	19	24.9	47.0	6.85
	R159X	20	32.0	24.2	4.92
	WT (reference)	8	43.4	0.178	0.998[a]
T_{m2}	E157X	19	57.2	1.71	1.31
	R158X	19	56.1	1.33	1.15
	R159X	20	56.4	0.555	0.745
	WT (reference)	8	56.8	0.745	2.04[a]

[a]As a correction for the effect of the number of data in the reference data set, a coefficient of 2.3646 is used for the correction as the 95% confidential coefficient estimated from the *t*-distribution table, where the degree of freedom is 7.

particular property, and AA index$_{min}$ and AA index$_{max}$ are, the minimum and maximum values in the AA index, respectively. In this study, we used all indices; binary-mode like indices (for example, composed of only 0 and 1), and incomplete indices with a deficiency of information on a few specified amino acid residues (composed of <20 numerical values.)

Correlation Analysis of the Sensitive Data Set with AA Indices—544 normalized AA indices were used for correlation analysis in terms of linear regression with the least square method. To obtain a 'ranking plot', the least square residuals (R) was used as the measure of the correlation as follows.

The selected mutational sensitive data set was fitted to the following equation

$$y_j = aX_j + b \quad (2)$$

where y_j is each observed value in data set of mutant j, x_j is the numerical value of a AA index for the respective amino acid for mutation j, and a and b are coefficients, respectively.

The least square residuals (R) were defined as follows, using Eq. 2.

$$R = \frac{1}{n}\sum_{j=1}^{n}(y_j - \hat{y})^2 \quad (3)$$

where n is the number of data used for the linear regression, and $\hat{y}$ is the least-squares estimator for each mutant (j) of the linear regression using each AA index.

All statistical analyses were performed using program Statistica 6.0 (StatSoft, Inc.) and in-house software.

Others—All chemicals used in this study were basically of reagent grade.

Some detailed descriptions of the above methods are available under Supplementary Experimental Procedures.

RESULTS

Outline of Our Sequence Perturbation Analysis Procedure—Figure 2 shows our sequence perturbation analysis procedure which was proposed and tested in this study: The first step (step 1) is to generate all the possible mutant proteins with a single amino acid replacement at a specified site. This step includes site-directed mutagenesis or adequate methods for mutagenesis, protein expression and purification. The second step (step 2) is to check the purified mutant proteins in terms of purity and molecular weight. The third step (step 3) is to measure the protein properties focused on. In this study, we focused on Michaelis parameters (k_{cat}, K_m and k_{cat}/K_m) and transition temperatures of thermal denaturation (T_{m1} and T_{m2}). The fourth step (step 4) is to determine whether each observed property responds significantly to the amino acid replacement ('mutational sensitive'). Because all the data include experimental errors, the experimental error on repeated measurements using the wild-type protein was used as the measure to determine if observed mutational responses were significant or not. Then, only the 'mutational sensitive' properties are subjected to the next step. The fifth step (step 5) is to conduct 'ranking' in terms of the results of linear regression analysis of a 'mutational sensitive' property as to all the AA indices in the AAindex1 database. By obtaining a 'ranking plot' with sorted least square residuals, it can be easily figured out what type of factor contributes to the observed property.

Mutagenesis and Characterization of Single Mutant Proteins (Steps 1-3)—In this work, we tried to construct all possible single mutant proteins at three sites, E157, R158 and R159, at the C-terminus of *E. coli* DHFR. All the mutant proteins except E157P and R158D were obtained as highly purified and suitable ones for further experiments. The quality of each purified protein was checked by SDS–PAGE, LC-MS and size-exclusion gel chromatography on a SMART system, and confirmed to be essentially homogeneous and free of the wild-type DHFR. Regarding the E157P and R158D mutant proteins, only very small amounts of them could be expressed and purified, and the amounts were not sufficient for further study. As a result, a total of 56 proteins including the wild-type DHFR were subjected to measurement of k_{cat}, K_m and k_{cat}/K_m as the observed properties for enzymatic function, and T_{m1} and T_{m2} as those for thermal stability. The observed properties were plotted, as shown in Fig. 3A–C, as a function of the replaced amino acid residues at positions of E157, R158 and R159, respectively.

Mutational Sensitivity Check (Step 4)—To determine whether a change in the observed properties of a mutant protein from those of the wild-type protein significantly

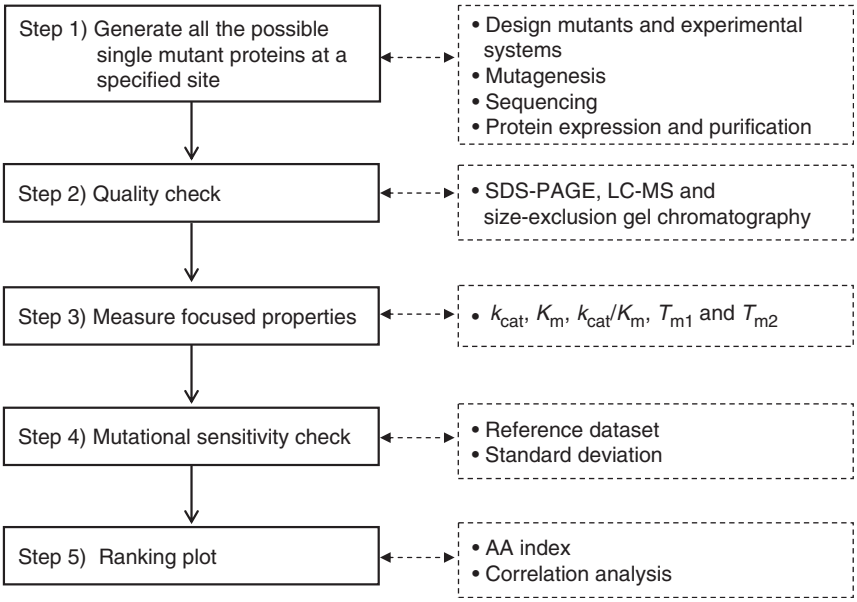

Fig. 2. **Schematic representation of the proposed sequence perturbation analysis.** The left half of the scheme (rectangles with solid lines) shows the concept for each procedure (steps), and the right half (rectangles with broken lines) shows the items which are related to the respective steps, respectively (see text).

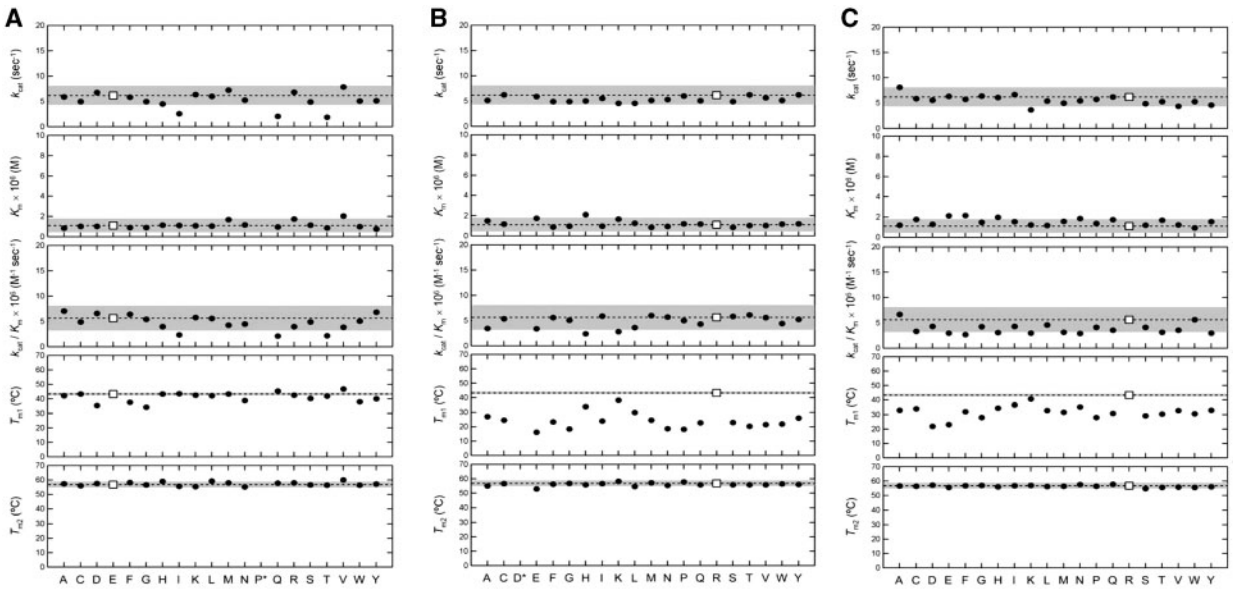

Fig. 3. **Plots of the observed properties.** Each panel shows k_{cat}, K_m, k_{cat}/K_m, T_{m1} and T_{m2} for each single mutant (black circles) as to E157 (A), R158 (B) and R159 (C), and the wild type (white square) from top to bottom, respectively. The amino acid residues of the respective mutants and the wild type are represented on the *x*-axis as a one-letter code in alphabetical order. The horizontal dashed line and grey zone represent the average and the range of the standard deviation multiplied by the correction coefficient of plus or minus (from the average used as a baseline) ($\pm$ 2.3646 $\times$ σ of reference data set) for the wild type, determined through eight repeated times measurements, respectively. P^* for E157 (in Fig. 3A) and D^* for R158 (in Fig. 3B) mean that the data for the E157P and R158D mutants are not included due to the expression levels being too low to measure, respectively.

reflects the amino acid replacement or not, we focused on the data distribution of the observed data set of a specified property of all the single mutant proteins as to the specified site, for example, the k_{cat} data set of E157X contains the data for 20 single mutant proteins including the wild-type protein at maximum, and so on.

To estimate experimental error accompanying measurement of a specified property, we made a reference data set containing the data for eight separate measurements using the wild-type protein (see EXPERIMENTAL PROCEDURES). In this study, we define 'mutational sensitive' as when the standard deviation, σ, of a mutant data

set is larger than that of the reference data set considering the difference in the number of data between the data sets. Table 1 shows the statistics for the data sets obtained in this study. As a result, the T_{m1} data sets of E157X, R158X and R159X were selected as being mutational sensitive. Other data sets, namely, the k_{cat}, K_m, k_{cat}/K_m and T_{m2} data sets, were insensitive as to mutations, demonstrating mutational robustness in the C-terminal region in terms of the enzymatic function at least. As also shown in Fig. 3, mutational sensitivity of the selected data sets was clearly demonstrated.

Figure 4 shows the far-UV CD spectra of all the single mutant proteins. The degree of the CD spectra changes caused by each of the systematic single amino acid replacements seemed to be in the following order: R158X > R159X > E157X. This order is consistent with the σ-values for the T_{m1} data sets.

Ranking Plot (Step 5)—Each data set selected in step 4 was subjected to correlation analysis with an AA index by mean of least square analysis. To carry out the analysis, all 544 AA indices from the AAindex1 database were normalized as described under SUPPLEMENTARY EXPERIMENTAL PROCEDURES and used in this study. We simply focused on the value of least square residuals for each linear regression analysis, and the residual values were sorted and plotted as a 'ranking plot' (Fig. 5).

As shown in Fig. 5, two types of ranking plots were obtained; one showed a gapped curve at a few initial points as in cases of the T_{m1} data sets of R158X and R159X ('type 1'), and the other showed a relatively smooth curve as in the case of the T_{m1} data set of E157X ('type 2').

As for type 1 ranking plots, it is clear that only a few AA indices are well correlated to the perturbed protein property. In the case of the T_{m1} data set of R158, the entry numbers of rank 1 and rank 2 AA indices are ZIMJ680104, which is related to the isoelectric point (*23*), and FAUJ880111, which is related to the positive charge (*24*), respectively (Table 2). In the case of the T_{m1} data set of R159X, those of the rank 1 and rank 2 AA indices are KLEP840101, which is related to the net charge (*25*), and ZIMJ680104, which is related to the isoelectric point (*23*), respectively. In both cases, it is noteworthy that the ZIMJ680104 AA index is ranked within the top 2.

As for the type 2 ranking plots, no particular AA index with good correlation was demonstrated. However, when we looked at the descriptions of AA indices within rank 10 in the plot of the T_{m1} data set of E157X, a few categories of descriptions were indicated. As listed in Table 2, five or six AA indices out of the 10 can be related to turn or bend formation properties, and a few AA indices related to charge effects are suggested even when the 157th amino acid residue is charged glutamic acid.

Examination of the Feasibility and General Versatility of our Sequence Perturbation Analysis—To date, many mutation studies on DHFR have been carried out and many data sets are available although each data set is not always perfect. In this study, we also tried to obtain ranking plots by using available data sets of sites other than E157X, R158X and R159X, both ones that have been published (*5*) and ones that have not been published yet but have already been studied in our laboratory.

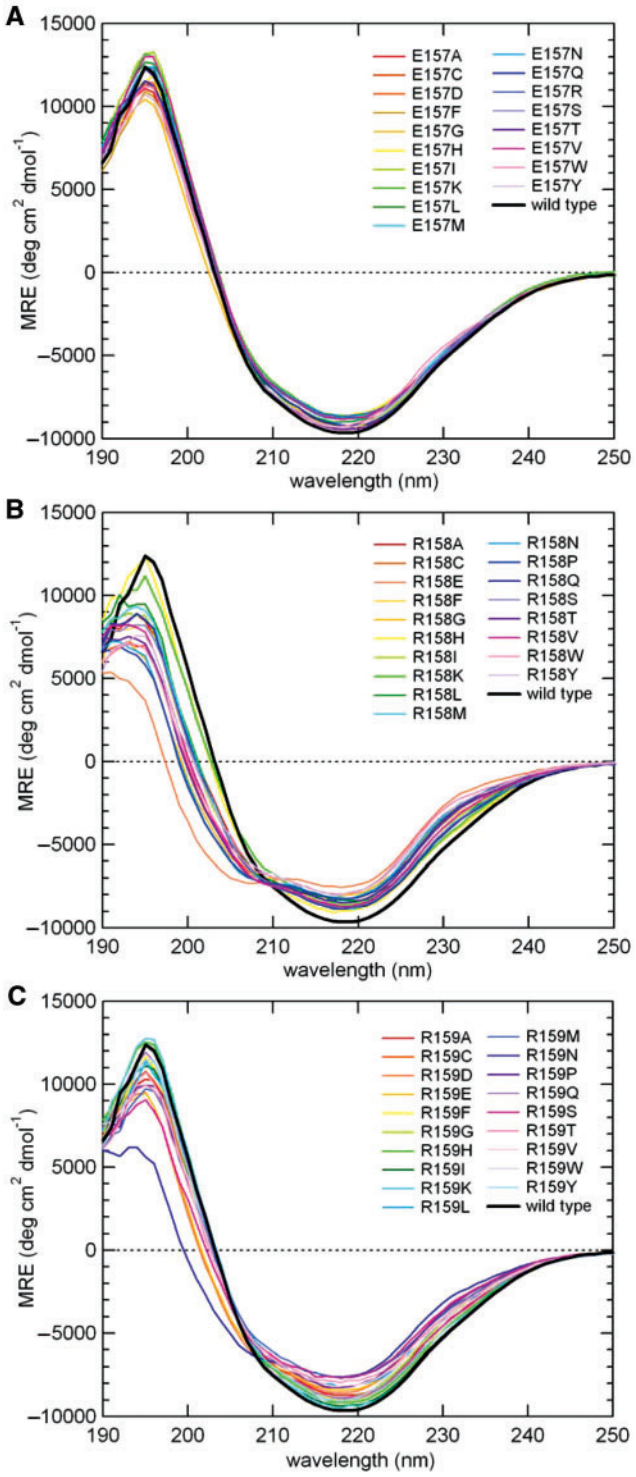

Fig. 4. **Far-UV CD spectra of the wild type and mutant DHFRs.** The panels show far-UV CD spectra of the respective single mutants as to E157 (A), R158 (B) and R159 (C). The colour code for each mutant protein is indicated in each panel and, in all panels, the wild-type protein is shown as a thick black line.

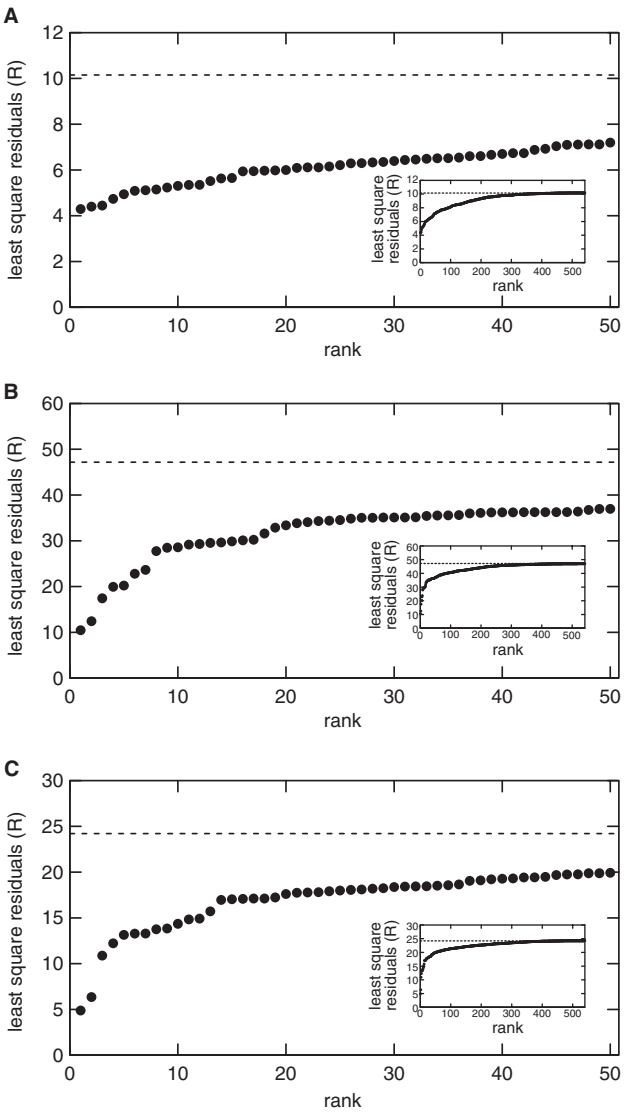

Fig. 5. **Ranking plots for the sensitive data sets for the C-terminus.** For the mutational sensitive data sets of T_{m1} for E157 (A), R158 (B) and R159 (C), the least square residuals (R) for each linear regression analysis (using each AA index) are sorted and plotted. In the larger panels, the minimum top 50 R-values, corresponding to best-fitting AA indices ranking in the top 50, are presented. In the insets, all R-values except inadequate data are plotted. The dashed line shows the variance of each data set (see Table 1).

As far as the available mutation sites of M16 (unpublished), M20 (unpublished), W30 (*5*), V40 (*5*), N59 (*5*) and I155 (*5*), the following data sets (total 30) were used for the feasibility and versatility tests; the k_{cat} data sets of M16X, M20X, W30X, V40X, N59X and I155X, the K_m data sets of M16X, M20X, W30X, V40X, N59X and I155X, the k_{cat}/K_m data sets of M16X, M20X, W30X, V40X, N59X and I155X, the ΔG data sets of W30X, V40X, N59X and I155X, the C_m data sets of W30X, V40X, N59X and I155X and the m-value data sets of W30X, V40X, N59X and I155X. From among these data sets, 18 were identified as mutational sensitive ones according to the step 4 procedure (Table 3).

Figure 6 shows the ranking plots of all the mutational sensitive properties. As in Fig. 5, the ranking plots can essentially be classified into two types. Among the above 18 data sets, 11 showed type 1 plots. As for the following 10 of 11 these data sets, the exception being the K_m data set of M20X, small numbers of AA indices showed good correlation with the respective data sets: k_{cat} data sets of M20X, W30X and V40X, K_m data set of N59X, k_{cat}/K_m data set of M20X, ΔG data sets of W30X and I155X, C_m value data sets of W30X and I155X, and m-value data set of I155X. For all the ranking plots, the descriptions of AA indices within rank 10 are presented in Supplementary Table 1.

DISCUSSION

Mutational Sensitivity—It is crucial to distinguish whether the effects of amino acid replacements are significant or not, because all measurements always involve experimental error and because repeated measurements of all the mutant proteins to reduce the experimental error are essentially hard to perform with an increased number of mutant proteins, such as in the systematic perturbation analysis as in this study. In the our analysis, we focused on a data set for all the single mutant proteins as to a specified site, and compared each data set with the reference data set for repeated measurements with the wild-type protein in terms of the standard deviation value, σ of the data set, which is a good measure of the data distribution. Apparently, when σ of a mutant data set is smaller than that of the reference data set, the mutation effects at the focused site are ignorable as a total. Even in this case, it might be possible that a certain amino acid replacement could cause a big real change not an experimental error by accident; however, such a mutation would be very particular, like a pin point, so it would not be useful for statistical analysis. On the other hand, when σ of a mutant data set is larger than that of the reference data set, the change in the observed property of the mutant protein is most likely significant even if the data include some or less experimental errors, therefore, we can say that the data set is sensitive as to the amino acid replacement or 'mutational sensitive' site in terms of the observed properties. Also, we can define mutational robustness by using the standard deviation of the single mutant data set for the observed parameter.

Significance of the 'Ranking Plot'—Because amino acid residues intrinsically possess many factors participating in protein structures and functions, determination of which factors contribute to them at a specific site of a protein sequence should be of great help for understanding how a protein obtains its structure and function. A wide variety of factors have been extensively investigated and each factor is referred to as an AA index with a set of 20 numerical values. In this study, we introduced a 'ranking plot' to visualize how each AA index is well correlated with the mutational sensitive data set. When one or two AA indices are typically correlated as shown in Fig. 5 (type 1 plot), a certain

Table 2. **Fitting parameters and AA index information on best linear-fitting indices for T_{m1} of E157, R158 and R159.**

Rank[a]	n[b]	Least square residuals, R[c]	a[d]	R[e]	(P-value)[f]	AA index[g]	Cluster[h]	Description[i]
T_{m1} of E157								
1	19	4.29 (42.3%)	12.08	0.760	(1.61E-04)	AVBF000102	U	Screening coefficients gamma, non-local (Avbelj, 2000)
2	19	4.40 (43.3%)	–8.52	0.753	(2.01E-04)	ISOY800105	O	Normalized relative frequency of bend S (Isogai *et al.*, 1980)
3	19	4.44 (43.7%)	–9.00	0.750	(2.18E-04)	RACS820109	O	Average relative fractional occurrence in AL(i–1) (Rackovsky-Scheraga, 1982)
4	19	4.73 (46.6%)	–8.45	0.730	(3.83E-04)	PALJ810105	A	Normalized frequency of turn from LG (Palau *et al.*, 1981)
5	19	4.94 (48.7%)	–9.52	0.716	(5.63E-04)	LEVM780106	A	Normalized frequency of reverse turn, unweighted (Levitt, 1978)
6	19	5.09 (50.1%)	–7.31	0.706	(7.26E-04)	CHOP780214	O	Frequency of the 3rd residue in turn (Chou-Fasman, 1978b)
7	19	5.11 (50.3%)	–9.02	0.704	(7.61E-04)	TANS770105	O	Normalized frequency of chain reversal S (Tanaka-Scheraga, 1977)
8	19	5.15 (50.7%)	18.59	0.702	(8.06 E-04)	ROBB760104	A	Information measure for C-terminal helix (Robson-Suzuki, 1976)
9	19	5.23 (51.5%)	–7.47	0.696	(9.37E-04)	CRAJ730103	A	Normalized frequency of turn (Crawford *et al.*, 1973)
10	19	5.31 (52.3%) (10.15)[j]	–8.15	0.691	(0.00106)	ISOY800103	A	Normalized relative frequency of bend (Isogai *et al.*, 1980)
T_{m1} of R158								
1	19	10.44 (22.1%)	30.29	0.882	(5.97E-07)	ZIMJ680104	H	Isoelectric point (Zimmerman *et al.*, 1968)
2	19	12.45 (26.4%)	16.11	0.857	(2.74E-06)	FAUJ880111	H	Positive charge (Fauchere *et al.*, 1988)
3	19	17.42 (36.9%)	27.61	0.793	(5.08E-05)	KLEP840101	H	Net charge (Klein *et al.*, 1984)
4	19	19.95 (42.3%)	18.98	0.759	(1.67E-04)	FINA910103	H	Helix termination parameter at position j-2,j-1,j (Finkelstein *et al.*, 1991)
5	19	20.22 (42.9%)	23.99	0.755	(1.89E-04)	HUTJ700103	P	Entropy of formation (Hutchens, 1970)
6	19	22.79 (48.3%)	21.32	0.718	(5.43E-04)	EISD860102	H	Atom-based hydrophobic moment (Eisenberg-McLachlan, 1986)
7	19	23.64 (50.1%)	18.37	0.705	(7.52E-04)	HUTJ700102	P	Absolute entropy (Hutchens, 1970)
8	19	27.74 (58.8%)	21.09	0.640	(0.00317)	FINA910104	H	Helix termination parameter at position j+1 (Finkelstein *et al.*, 1991)
9	17	28.48 (60.4%)	3.14	0.175	(0.502)	ROSM880104	U	Hydropathies of amino acid side chains, neutral form (Roseman, 1988)
10	19	28.59 (60.6%) (47.16)[j]	19.36	0.626	(0.00418)	QIAN880112	A	Weights for alpha-helix at the window position of 5 (Qian-Sejnowski, 1988)
T_{m1} of R159								
1	20	4.87 (20.1%)	19.67	0.894	(1.11E-07)	KLEP840101	H	Net charge (Klein *et al.*, 1984)
2	20	6.35 (26.2%)	19.59	0.859	(1.24E-06)	ZIMJ680104	H	Isoelectric point (Zimmerman *et al.*, 1968)
3	20	10.88 (44.9%)	–13.29	0.742	(1.79E-04)	CHOP780204	H	Normalized frequency of N-terminal helix (Chou-Fasman, 1978b)
4	20	12.22 (50.5%)	12.38	0.704	(5.34E-04)	FINA910103	H	Helix termination parameter at position j-2,j-1,j (Finkelstein *et al.*, 1991)
5	20	13.15 (54.3%)	–11.64	0.676	(0.00107)	ROBB760102	H	Information measure for N-terminal helix (Robson-Suzuki, 1976)
6	20	13.28 (54.8%)	15.63	0.672	(0.00117)	HUTJ700103	P	Entropy of formation (Hutchens, 1970)
7	20	13.30 (54.9%)	–14.90	0.672	(0.00119)	FINA910101	H	Helix initiation parameter at position i–1 (Finkelstein *et al.*, 1991)
8	20	13.76 (56.8%)	15.56	0.657	(0.00164)	FINA910104	H	Helix termination parameter at position j+1 (Finkelstein *et al.*, 1991)
9	20	13.84 (57.1%)	–10.74	0.655	(0.00174)	FAUJ880112	H	Negative charge (Fauchere *et al.*, 1988)
10	20	14.35 (59.2%) (24.22)[j]	8.80	0.638	(0.00245)	FAUJ880111	H	Positive charge (Fauchere *et al.*, 1988)

[a]The best linear-fitting AA indices for each data set of observed properties are ranked based on the respective least square residuals (R)[c]. [b]The number of available data points used in each linear fitting. [c]Least square residuals (R) in each linear fitting. See details in EXPERIMENTAL PROCEDURES section in the main text. The percentage in parentheses indicates the ratio of R relative to pre-fitting least square residuals [j] of each observed property data set. [d]Coefficient of change in slope in linear fitting. [e]Correlation coefficient in each linear fitting. [f]Reliability coefficient for correlation coefficient [e] in each linear fitting. [g]Accession number of AA index used and selected in each linear fitting. [h]The cluster of each AA index cited from Tomii and Kanehisa (*3*); *A*, α and turn propensities; *B*, β propensity; *C*, composition; *H*, hydrophobicity; *P*. physicochemical properties; *O*, other properties; *U*, unclassified indices. [i]The description of each index. [j]Pre-fitting least square residuals, equivalent to variance in Table 1.

Table 3. **Statistics of the mutant data sets and reference data set for other sites of DHFR.**

Content of data set					
Observed property[a]	Target protein	Number of data in the data set	Average	Variance	Standard deviation
k_{cat}	M16	18	11.5	34.3	5.86
	M20	15	6.67	35.3	5.94
	W30	10	12.6	6.36	2.52
	V40	10	6.79	9.75	3.12
	N59	17	5.08	3.36	1.83
	I155	13	6.02	1.18	1.08
	WT (reference)[b]	8	6.18	0.651	1.91
K_m	M16	18	1.80	0.277	0.527
	M20	15	2.69	1.71	1.31
	W30	10	2.03	0.302	0.549
	V40	10	2.64	3.44	1.86
	N59	17	3.33	11.1	3.34
	I155	13	1.58	0.299	0.547
	WT (reference)[b]	8	1.10	0.0860	0.693
k_{cat}/K_m	M16	18	6.78	15.1	3.89
	M20	15	3.17	11.5	3.38
	W30	10	6.45	1.08	1.04
	V40	10	3.21	3.19	1.79
	N59	17	2.60	3.70	0.978
	I155	13	4.19	1.83	1.35
	WT (reference)[b]	8	5.62	1.09	2.47
ΔG	W30	10	4.05	0.687	0.829
	V40	10	4.52	0.739	0.860
	N59	17	5.32	0.956	0.978
	I155	13	3.29	1.673	1.29
	WT (reference)[c]	3	6.25	0.017	0.565[d]
C_m	W30	10	2.26	0.118	0.343
	V40	10	2.56	0.195	0.441
	N59	17	3.10	0.0905	0.301
	I155	13	2.15	0.157	0.397
	WT (reference)[c]	3	3.11	0.00291	0.232[d]
m	W30	10	1.78	0.0148	0.122
	V40	10	1.78	0.0410	0.202
	N59	17	1.71	0.0464	0.215
	I155	13	1.48	0.0883	0.297
	WT (reference)[c]	3	2.01	0.00407	0.274[d]

[a]Enzymatic activity measurements were performed at 15°C. The urea-induced equilibrium unfolding transitions of the wild-type and mutant DHFRs (for parameters of ΔG, C_m and m-value) were also monitored as fluorescence spectra at 15°C. Note that ΔG is the Gibbs free energy change between the native and unfolded states, C_m is the urea concentration of the transition midpoint, and m-value is the sensitivity of the free energy to the denaturant concentration (the co-operativity index). [b]The reference data sets of enzymatic activity (k_{cat}, K_m and k_{cat}/K_m) for the wild type, including standard deviation corrected in terms of the number of data, are the same as those used in Table 1 and the EXPERIMENTAL PROCEDURES section in the main text. [c]The reference data for the urea-induced equilibrium unfolding transition (ΔG, C_m and m) of the wild type are cited from three previous studies, Arai and Iwakura (*5*), O'Neill *et al.* (*40*), and Gekko *et al.* (*41*). [d]As a correction for the effect of the number of data in the reference data set, a coefficient of 4.3027 is used for the correction as the 95% confidential coefficient estimated from the t-distribution table, where the degree of freedom is 2 (three data samples are used) for the estimation as described above.

factor related to the description of the well-fitted AA index is effective for the mutational sensitivity, and then we can guess that a certain interaction related to the AA index participates mainly in the observed property at the specified site. In this way, to elucidate the role or participation of the interaction at each amino residue of the entire protein will be of great help for understanding the protein folding problem mechanism more precisely. In this respect, this approach was successful for the C-terminal end, as described below.

Role of the C-terminal Amino Acid Residues of DHFR—It is known that usage of amino acid residues at the C-terminal ends of naturally occurring proteins is relatively biased (*6*, *7*, *8*). Exhaustive statistical analyses of the C-terminal ends of proteins of *E. coli*, yeast, and *Homo sapiens* showed that charged amino acid residues are present at high frequency (*6*, *7*), and this is the case for DHFR from various sources.

As the translation process terminates at the C-terminus, the last few amino acids forming the C-terminus motif might interact with various release factors related to translational termination (*26*), and after translation, they can be processed through the post-translational modification involved in a variety of cell-biological processes through interactions with other proteins and/or non-protein biomolecules such as lipids (*27*). Using GST-fusions with *Saccharomyces cerevisiae* proteins, it has been demonstrated that positively charged amino acid residues at C-termini were suitable for attachment of a protein to membrane components such as phospholipids, thereby stabilizing the spatial structure of the protein or contributing to its subcellular localization after separation from a ribosome at the end of protein synthesis (*8*). However, the *E. coli* DHFR is not a membrane or secreted protein, and thus it is hard to say that the C-terminal arginine residue participates in the attachment of the DHFR to membrane components.

The N- and C-terminal regions of some proteins, such as α-lactalbumin, lysozyme, apomyoglobin and cytochrome *c* have been found to be in close proximity in the native state (*28–31*). It has been proposed that the termini of polypeptides may be held in close proximity during folding *in vivo*, and that the C-terminus regions may contain information critical for the folding and stability of the native protein structures (*32–34*). Early compaction of a DHFR molecule to yield a specific folding intermediate with the N and C-terminal regions in close proximity is suggested to be a crucial event in folding (*35*). In this study, we showed that T_{m1} data sets are mutational sensitive for the C-terminal region of the DHFR, but not other data sets. In particular, it is noteworthy that the T_{m1} data sets for both R158 and R159 are linearly and highly correlated with AA indices related charge effects, such as the isoelectric point and net charge. The T_{m1} parameter may be related to a stability measure of the loop subdomain, which includes the C-terminal region, of the two domain DHFR protein. So, we can suggest the role of the C-terminal amino acid residues in the DHFR are as follows:

(i) R158 and R159 participate in stabilization of the loop subdomain of the DHFR through some sort of

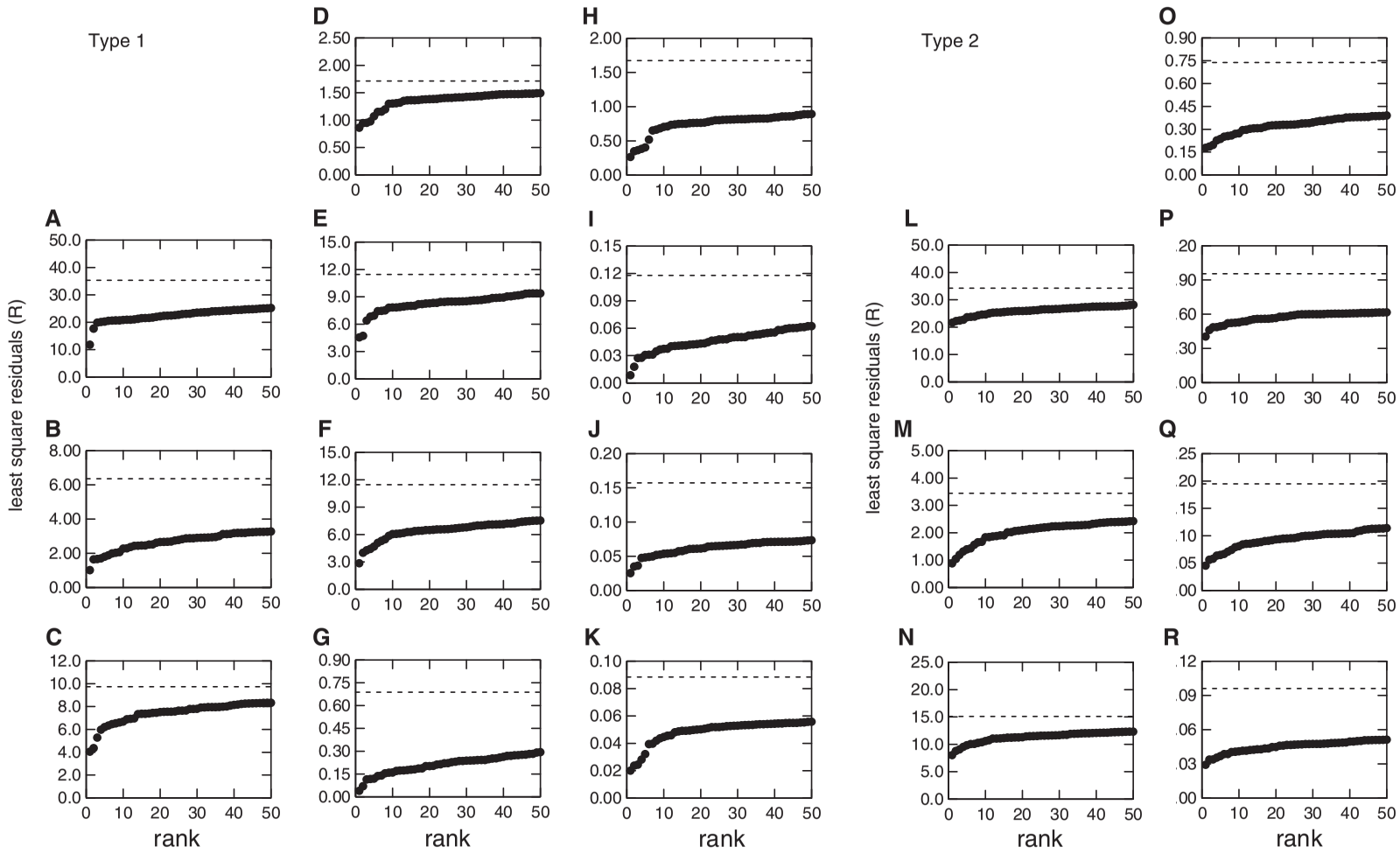

Fig. 6. **Ranking plots for the sensitive data sets of M16X, M20X, W30X, V40X, N59X and I155X.** Ranking plots with sorted values of least square residuals ranking in the top 50, for the observed properties considered as being 'mutational sensitive' data sets for M16, M20, W30, V40, N59 and I155, are presented. These plots are classified into type 1, the k_{cat} data sets of M20 (A), W30 (B) and V40 (C), K_m data sets of M20 (D) and N59 (E), k_{cat}/K_m data set of M20 (F), ÿ ΔG data sets of W30 (G) and I155 (H), C_m value data sets of W30 (I) and I155 (J), and m-value data set of I155 (K); and type 2, the k_{cat} data set of M16 (L), K_m data sets of V40 (M), k_{cat}/K_m data set of M16 (N), ΔG data set of V40 (O) and N59 (P), and C_m value data sets of V40 (Q) and N59 (R).

non-covalent interaction related to the electrostatic properties of amino acid residues.

(ii) In terms of the T_{m1} response, the C-terminal residues of the wild-type DHFR are the best among the naturally occurring amino acids. This means the C-terminal sequence of the present *E. coli* DHFR became optimized during the natural evolution process, and the selection pressure probably comprises the stability of the major domain.

(iii) Other physicochemical indices could not be related to the observed responses significantly. Therefore, steric changes at the C-terminal, for example, peptide extension at the C-terminal end and formation of a fusion protein through the C-terminal, are permissive with relatively small effects on the stability of the DHFR itself (see below).

The C-termini of proteins are employed as polypeptide fusion sites for many tags (*36*); conventionally, maltose-binding protein (MBP), His-tag, Strep-tag (II), glutathione S-transferase (GST) and also DHFR are widely utilized for increases in expression and solubility, simple and quick purification, improvement in fineness (*15*), and facilitation of detection and immobilization (*17*, *37*) of a target protein. The biased sequences in naturally occurring proteins with charged amino acids may serve for stabilization of the proteins, probably through an electrostatic interaction (relatively long range interaction) (*38*) but not through a steric factor (short range interaction) as in the case of DHFR.

Construction of a New AA Index Based on Properties Observed in This Study—In this study, the T_{m1} data sets for both R158 and R159 were found to be linearly and highly correlated with ZIMJ680104; the AAindex for the isoelectric point of the amino acid property. Also, the T_{m1} data sets for R158 and R159 were only mutational sensitive as far as tested, and thus it seems that very small numbers of factors of amino acid residues participate in the change. If it can be assumed that the observed changes in T_{m1} with single mutations would reflect simply changes in a certain factor of each amino acid residue, it would be possible to construct a new AA index independent of any other factors by best fitting to the two T_{m1} data sets as a first approximation. It should be noted that it is hard to say that the AA indices used in this study are independent of each other. We constructed a new AA index, as described in Supplementary Table 2. The resulting new AA index, named the 'charge AA index', showed the highest correlation to the data sets for both R158 and R159 among all AA indices used in this study with R-value = 2.41, $R = 0.974$, $P < 1E\text{-}9$ for T_{m1} of R158 and R-value = 2.56, $R = 0.946$, $P < 1E\text{-}9$ for T_{m1} of R159, respectively.

In conclusion, our novel approach proposed in this study involving the sequence perturbation concept is useful and powerful to find a main (or unique) type of factor (AA index) which manages a specified characteristic at a specified site of protein sequence when the effect of the factor is dominant. In addition, our method is useful for attaining rough information of the number of factors which contribute to a specified characteristic at a specified site, for example, type 1 plot may suggest that a few factors are involved, and type 2 plot may suggest that multiple factors are involved. However, the methodology for the assignment of the factors in terms of AA indices in the case of type 2 plot remains to be developed. The feasibility of our method to the overall sequence of the protein is underway.

SUPPLEMENTARY DATA

Supplementary data are available at JB online.

ACKNOWLEDGEMENTS

We thank N. Furuya, A. Fujisawa, J. Suzuki, M. Oose and C. Yamane (National Institute of Advanced Industrial Science and Technology) for the protein purification, DNA sequencing, and enzyme assays. We are also grateful to Professor C.R. Matthews (University of Massachusetts) for critical reviewing of this article.

CONFLICT OF INTEREST

None declared.

REFERENCES

1. Tramontano, A. (2007) Worth the effort. An account of the Seventh Meeting of the Worldwide Critical Assessment of Techniques for Protein Structure Prediction. *FEBS J.* **274**, 1651–1654
2. Haber, E. and Anfinsen, C.B. (1961) Regeneration of enzyme activity by air oxidation of reduced subtilisin-modified ribonuclease. *J. Biol. Chem.* **236**, 422–424
3. Tomii, K. and Kanehisa, M. (1996) Analysis of amino acid indices and mutation matrices for sequence comparison and structure prediction of proteins. *Protein Eng.* **9**, 27–36
4. Kawashima, S., Pokarowski, P., Pokarowska, M., Kolinski, A., Katayama, T., and Kanehisa, M. (2008) AAindex: amino acid index database, progress report 2008. *Nucleic Acids Res.* **36**, D202–D205
5. Arai, M. and Iwakura, M. (2005) Probing the interactions between the folding elements early in the folding of *Escherichia coli* dihydrofolate reductase by systematic sequence perturbation analysis. *J. Mol. Biol.* **347**, 337–353
6. Berezovsky, I.N., Kilosanidze, G.T., Tumanyan, V.G., and Kisselev, L. (1997) COOH-terminal decamers in proteins are non-random. *FEBS Lett.* **404**, 140–142
7. Berezovsky, I.N., Kilosanidze, G.T., Tumanyan, V.G., and Kisselev, L.L. (1999) Amino acid composition of protein termini are biased in different manners. *Protein Eng.* **12**, 23–30
8. Scheglmann, D., Werner, K., Eiselt, G., and Klinger, R. (2002) Role of paired basic residues of protein C-termini in phospholipid binding. *Protein Eng.* **15**, 521–528
9. Jennings, P.A., Finn, B.E., Jones, B.E., and Matthews, C.R. (1993) A reexamination of the folding mechanism of dihydrofolate reductase from *Escherichia coli*: verification and refinement of a four-channel model. *Biochemistry* **32**, 3783–3789
10. Sawaya, M.R. and Kraut, J. (1997) Loop and subdomain movements in the mechanism of *Escherichia coli* dihydrofolate reductase: crystallographic evidence. *Biochemistry* **36**, 586–603
11. Wagner, C.R., Thillet, J., and Benkovic, S.J. (1992) Complementary perturbation of the kinetic mechanism and

catalytic effectiveness of dihydrofolate reductase by side-chain interchange. *Biochemistry* **31**, 7834–7840
12. Bolin, J.T., Filman, D.J., Matthews, D.A., Hamlin, R.C., and Kraut, J. (1982) Crystal structures of *Escherichia coli* and *Lactobacillus casei* dihydrofolate reductase refined at 1.7 Å resolution. I. General features and binding of methotrexate. *J. Biol. Chem.* **257**, 13650–13662
13. Blakley, R.L. (1984) *Dihydrofolate Reductase in* Folates and Pteridines (Blakley, R.L. and Benkovic, S.J., eds.) Vol. 1, pp. 191–253, Wiley, New York
14. Uwajima, T., Oshiro, T., Eguchi, T., Kuge, Y., Horiguchi, A., Igarashi, A., Mochida, K., and Iwakura, M. (1990) Chemo-enzymatic synthesis of optically pure *l*-leucovorin, an augmentor of 5-fluorouracil cytotoxicity against cancer. *Biochem. Biophys. Res. Commun.* **171**, 684–689
15. Iwakura, M., Furusawa, K., Kokubu, T., Ohashi, S., Tanaka, Y., Shimura, Y., and Tsuda, K. (1992) Dihydrofolate reductase as a new "affinity handle". *J. Biochem.* **111**, 37–45
16. Iwakura, M., Jones, B.E., Luo, J., and Matthews, C.R. (1995) A strategy for testing the suitability of cysteine replacements in dihydrofolate reductase from *Escherichia coli*. *J. Biochem.* **117**, 480–488
17. Takenawa, T., Oda, Y., Ishihama, Y., and Iwakura, M. (1998) Cyanocysteine-mediated molecular dissection of dihydrofolate reductase: occurrence of intra- and inter-molecular reactions forming a peptide bond. *J. Biochem.* **123**, 1137–1144
18. Hillcoat, B.L., Nixon, P.F., and Blakley, R.L. (1967) Effect of substrate decomposition on the spectrophotometric assay of dihydrofolate reductase. *Anal. Biochem.* **21**, 178–189
19. Touchette, N.A., Perry, K.M., and Matthews, C.R. (1986) Folding of dihydrofolate reductase from *Escherichia coli*. *Biochemistry* **25**, 5445–5452
20. Pace, C.N., Vajdos, F., Fee, L., Grimsley, G., and Gray, T. (1995) How to measure and predict the molar absorption coefficient of a protein. *Protein Sci.* **4**, 2411–2423
21. Ionescu, R.M., Smith, V.F., O'Neill, J.C. Jr, and Matthews, C.R. (2000) Multistate equilibrium unfolding of *Escherichia coli* dihydrofolate reductase: thermodynamic and spectroscopic description of the native, intermediate, and unfolded ensembles. *Biochemistry* **39**, 9540–9550
22. Gill, S.C. and von Hippel, P. H. (1989) Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem.* **182**, 319–326
23. Zimmerman, J.M., Eliezer, N., and Simha, R. (1968) The characterization of amino acid sequences in proteins by statistical methods. *J. Theor. Biol.* **21**, 170–201
24. Fauchere, J.L., Charton, M., Kier, L.B., Verloop, A., and Pliska, V. (1988) Amino acid side chain parameters for correlation studies in biology and pharmacology. *Int. J. Pept. Protein Res.* **32**, 269–278
25. Klein, P., Kanehisa, M., and DeLisi, C. (1984) Prediction of protein function from sequence properties. Discriminant analysis of a data base. *Biochim. Biophys. Acta* **787**, 221–226
26. Bjornsson, A., Mottagui-Tabar, S., and Isaksson, L.A. (1996) Structure of the C-terminal end of the nascent peptide influences translation termination. *EMBO J.* **15**, 1696–1704
27. Chung, J.J., Shikano, S., Hanyu, Y., and Li, M. (2002) Functional diversity of protein C-termini: more than zipcoding? *Trends Cell Biol.* **12**, 146–150
28. Forge, V., Wijesinha, R.T., Balbach, J., Brew, K., Robinson, C.V., Redfield, C.V., and Dobson, C.M. (1999) Rapid collapse and slow structural reorganisation during the refolding of bovine α-lactalbumin. *J. Mol. Biol.* **288**, 673–688
29. Segel, D.J., Bachmann, A., Hofrichter, J., Hodgson, K.O., Doniach, S., and Kiefhaber, T. (1999) Characterization of transient intermediates in lysozyme folding with time-resolved small-angle X-ray scattering. *J. Mol. Biol.* **288**, 489–499
30. Eliezer, D., Jennings, P.A., Wright, P.E., Doniach, S., Hodgson, K.O., and Tsuruta, H. (1995) The radius of gyration of an apomyoglobin folding intermediate. *Science* **270**, 487–488
31. Roder, H., Elove, G.A., and Englander, S.W. (1988) Structural characterization of folding intermediates in cytochrome *c* by H-exchange labelling and proton NMR. *Nature* **335**, 700–704
32. Christopher, J.A. and Baldwin, T.O. (1996) Implications of N and C-terminal proximity for protein folding. *J. Mol. Biol.* **257**, 175–187
33. Thornton, J.M. and Sibanda, B.L. (1983) Amino and carboxy-terminal regions in globular proteins. *J. Mol. Biol.* **167**, 443–460
34. Robben, J., Van der Schueren, J., and Volckaert, G. (1993) Carboxyl terminus is essential for intracellular folding of chloramphenicol acetyltransferase. *J. Biol. Chem.* **268**, 24555–24558
35. Arai, M., Kataoka, M., Kuwajima, K., Matthews, C.R., and Iwakura, M. (2003) Effects of the difference in the unfolded-state ensemble on the folding of *Escherichia coli* dihydrofolate reductase. *J. Mol. Biol.* **329**, 779–791
36. Terpe, K. (2003) Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. *Appl. Microbiol. Biotechnol.* **60**, 523–533
37. Iwakura, M., Nakamura, D., Takenawa, T., and Mitsuishi, Y. (2001) An approach for protein to be completely reversible to thermal denaturation even at autoclave temperatures. *Protein Eng.* **14**, 583–589
38. Perry, K.M., Onuffer, J.J., Gittelman, M.S., Barmat, L., and Matthews, C.R. (1989) Long-range electrostatic interactions can influence the folding, stability, and cooperativity of dihydrofolate reductase. *Biochemistry* **28**, 7961–7968
39. Koradi, R., Billeter, M., and Wuthrich, K. (1996) MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.* **14**, 51–55
40. O'Neill, J.C. Jr, and Matthews, C.R. (2000) Localized, stereochemically sensitive hydrophobic packing in an early folding intermediate of dihydrofolate reductase from *Escherichia coli*. *J. Mol. Biol.* **295**, 737–744
41. Gekko, K., Kunori, Y., Takeuchi, H., Ichihara, S., and Kodama, M. (1994) Point mutations at glycine-121 of *Escherichia coli* dihydrofolate reductase: important roles of a flexible loop in the stability and function. *J. Biochem.* **116**, 34–41