J. Biochem. 124, 769-777 (1998)

Topological Mutation of *Escherichia coli* Dihydrofolate Reductase[1]

Masahiro Iwakura,[2] Tatsuyuki Takenawa, and Tsutomu Nakamura

National Institute of Bioscience and Human-Technology, 1-1 Higashi, Tsukuba, Ibaraki 305-8566

Received for publication, May 19, 1998

Proteins appear to contain structural elements which determine the folded structure. If such elements are present, the order of structural elements in the primary structure, *i.e.* the chain topology, can be disregarded for building of the folded tertiary structure, when they are properly connected to each other by proper linkers. To experimentally examine this, "topological" mutants (designated as GHF33 and GHF34) of *Escherichia coli* dihydrofolate reductase (DHFR) were designed and constructed by switching two amino acid sequence parts containing the βF strand and βG-βH strands in the primary sequence. In this way, the chain topology of wild-type DHFR, βA→αB→βB→αC→βC→βD→αE→βE→αF→βF→βG→βH, was changed to βA→αB→βB→αC→βC→βD→αE→βE→αF→βG→βH→βF. Such topological mutant proteins were stably expressed and accumulated in *E. coli* cells, and highly purified. Molecular mass measurements of the purified proteins and their proteolytic fragments confirmed that GHF33 and GHF34 had the designed sequence. In terms of k_{cat}, the GHF33 and GHF34 proteins showed about 10 and 20% of the DHFR activity of the wild-type with K_m values of 3.3 μM (GHF33) and 2.1 μM (GHF34), respectively. The topological mutants showed a cooperative two-state transition in urea-induced unfolding experiments with ΔG^{H_2O} values of 4.0 and 4.8 kcal/mol. Whereas, the K_m and ΔG^{H_2O} values for wild-type DHFR were 0.9 μM and 6.1 kcal/mol, respectively. The significance of the topological mutations was discussed with respect to protein folding and protein evolution.

Key words: chain topology, dihydrofolate reductase, protein folding, structural element, topological mutation.

Proteins appear to contain structural elements which determine the folded structure (Fig. 1, [2], [3]) (*1-3*), although this has not been clearly proven yet. Such elements may be distributed throughout the primary sequence of a protein and not localized in a specific part (Fig. 1, [1]). The order of secondary structural elements in the primary sequence, namely, the "chain topology" of the secondary structure, is useful for classification of proteins (*4*). Protein folding includes an intra-molecular "recognition" process at various stages. The intra-molecular association ("recognition") of structural elements may be crucial for the protein folding process (*5, 6*). If substantial information regarding protein structure and folding is encoded by structural elements (Fig. 1, [3]), the chain topology plays only a minor role in determination of the folded structure. The chain connectivity may act to combat entropy during folding by increasing the effective local concentrations of individual structural elements with the fold defined by the interactions among these elements. Thus, the chain topology may not be critical for building of the folded tertiary structure, if they are properly connected to each other (Fig. 1, [4], [5]). Whereas, the chain topology seems to be well conserved within a protein family (*7-9*), and therefore seems to be crucial in determination of the folded structure. We think that the conservation of the chain topology might have been due to a biological mechanism during natural evolution and thus does not reflect the intrinsic properties of a polypeptide. The simplest approach to prove our idea is to experimentally construct an active mutant protein with a different chain topology from the natural one. However, no such successful approaches had been reported before this work was started.

Dihydrofolate reductase (DHFR, EC 1.5.1.3) is a monomeric, two domain protein, and one of the well characterized proteins in terms of structure and function (*10*). DHFR catalyzes the reduction of dihydrofolate (DHF) to tetrahydrofolate (THF), using the reducing cofactor, NADPH. DHFR is a clinically important enzyme not only because it is the target of a number of antifolate drugs, such as trimethoprim and methotrexate, but also because it can be used to produce *l*-leucoverin, an anti-cancer drug (*11*), in a stereospecific manner. The tertiary structure of this enzyme is a doubly-wound, parallel α/β sheet, comprising four α helices and eight β strands (*12*). More than 25 positions have been used for obtaining amino acid substitutions by site-directed mutagenesis (*13-19*), and several circularly permuted variants have been created (*20-22*).

As a primary step to experimentally determine whether the chain topology is crucial in determination of the folded

[1] This research was supported in part by the New Energy and Industrial Technology Development Organization (NEDO) and the Japan Science and Technology Corporation (JST). A part of this work was presented at the Tenth Symposium of the Protein Society held in San Jose, CA, USA, on August 6, 1996 (*40*).

[2] To whom correspondence should be addressed. E-mail: iwakura@nibh.go.jp

Abbreviations: DHF, dihydrofolate; DHFR, dihydrofolate reductase; LC/MS, liquid chromatography/mass spectrometry; THF, tetrahydrofolate.

© 1998 by The Japanese Biochemical Society.

Downloaded from http://jb.oxfordjournals.org/ at Islamic Azad University on October 1, 2012

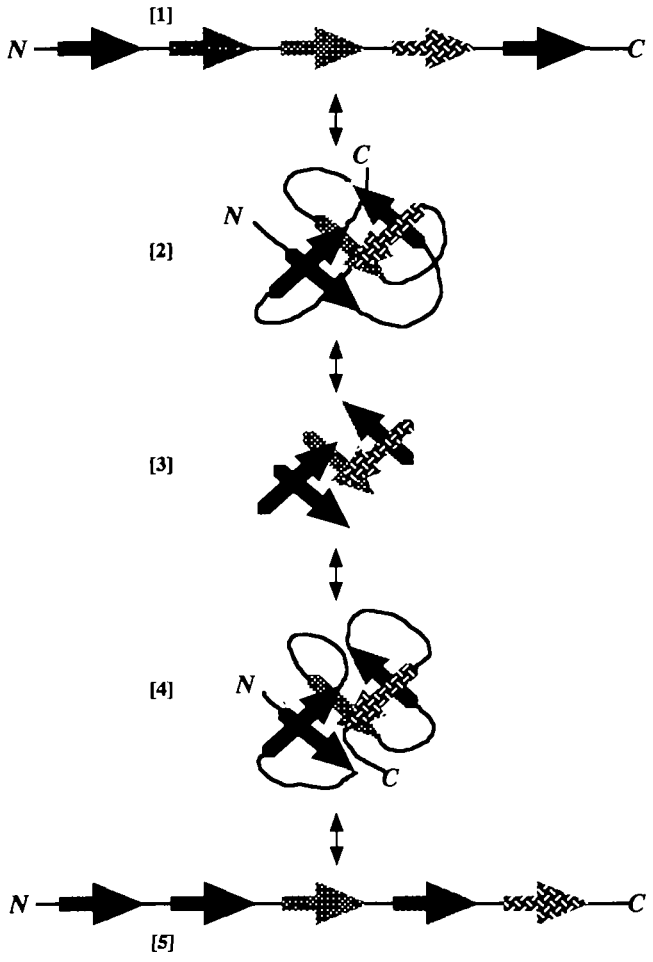

Fig. 1. **Schematic drawing of the role of structural elements and a topological mutation.** For simplification, structural elements are assumed to determine the folded structure [2], and other sequences are assumed only to play a role in connecting each the structural units to avoid undesired inter-molecular associations. Structural elements are separated by connecting sequences in the primary structure, giving a chain topology [1]. If this is true, the fragmented structural elements could form a folded core [3]. Then, the changing connectivity of the structural elements would not destroy the folded core [3], but form the folded structure [4]. The chain topology [5] of the folded structure [4] is completely different from the original chain topology [1]. In this situation, we designated the sequence changes from [1] to [5] as the "topological mutation."

structure or not, a "topological" mutation has been designed using *Escherichia coli* DHFR as a model protein. Our designed mutant protein was stably expressed in *E. coli* cells, and could be highly purified.

MATERIALS AND METHODS

Chemicals—The methotrexate (MTX)-agarose affinity resin was obtained from Sigma. DEAE-Toyopearl 650M was from Tosoh (Tokyo). The restriction enzymes, T4-DNA ligase, Taq polymerase, and *Achromobacter* protease I were obtained from Takara Shuzo (Kyoto). All other chemicals were of reagent grade.

Molecular Cloning—Molecular cloning experiments were carried out according to the manufacturers' protocols or to the method described by Maniatis *et al.* (*23*). Trimethoprim (tmp)-resistant transformants which over-expressed active DHFR were selected on an agar plate containing 8 g/liter tryptone (Difco), 5 g/liter yeast extract (Difco), 5 g/liter NaCl, 50 mg/liter tmp, 200 mg/liter ampicillin, and 15 g/liter Bacto agar (Difco).

Protein Purification and Enzyme Assay—Purification of topological mutant DHFRs was carried out mainly by MTX-affinity chromatography with adequate steps of pre-purification from cell-free extracts (*24*). The purified proteins were homogenous, as judged on SDS-polyacrylamide gel electrophoresis (*25*), and stored in a saturated ammonium sulfate solution as a precipitate at 4°C until use. Protein concentrations were determined as the absorbance at 280 nm using the extinction coefficient ($\varepsilon_{280}=3.11\times10^4\ M^{-1}\cdot cm^{-1}$) (*26*). The activity of DHFR was determined spectrophotometrically at 15°C as the disappearance of NADPH and DHF at 340 nm ($\varepsilon_{340}=11{,}800\ M^{-1}\cdot cm^{-1}$) (*27*). The assay mixture contained 50 μM DHF, 60 μM NADPH, 14 mM 2-mercaptoethanol, MTEN buffer [50 mM 2-morpholinoethanesulfonic acid (MES), 25 mM Tris, 25 mM ethanolamine, 100 mM NaCl], pH 7.0 (*28*), and the enzyme, in a final volume of 2 ml. Michaelis parameters (k_{cat} and K_m) were determined with various concentrations of DHF. Data were analyzed according to the following equation by nonlinear least-square fitting:

$$\text{Initial velocity } (v_{obs}) = k_{cat}\,[\text{DHF}][\text{DHFR}]/([\text{DHF}]+K_m),$$

where [DHF] and [DHFR] are the concentrations of DHF and DHFR, respectively.

Size-Exclusion Chromatography—Size-exclusion chromatography was carried out on a Superdex 75 PC3.2/30 column in a Smart System (Pharmacia, Sweden). Elution was carried out at a constant flow rate of 40 μl/min with 10 mM potassium phosphate buffer, pH 7.8, containing 0.2 mM EDTA at 15°C. Absorbance was monitored at 280 nm. As molecular size (weight) standard proteins, bovine serum albumin, egg albumin, β-lactoglobulin, myoglobin, and ribonuclease were used.

LC-MS Measurements—All mass-spectroscopic (MS) measurements were carried out in the positive-ion mode using a PE Sciex API III triple quadrupole mass spectrometer. For on-line LC/MS measurements the nebulization probe inlet was coupled directly to the splitting device. The split was adjusted to provide a flow of 5 μl/min to the mass spectrometer (total effluent, 0.2 ml/min). Separation was carried out with an acetonitrile gradient, from 10% (0 min) to 60% (45 min), containing 0.1% TFA for an l-column ODS (ϕ2.1×150 mm; Kagaku-hin Kensa Kyoukai, Tokyo). The ionspray tip was held at a potential of 5,000 V. Compressed nitrogen (pressure, 5.7 kg/cm^2; and flow-rate, 0.8 liter/min) was used to assist liquid nebulization. A curtain gas (nitrogen) flow of 1.2 liters/min was used to prevent moisture from reaching the orifice and the quadrupole guidance lens. The MS system was operated with an orifice voltage of 100 V, and the scanning range was m/z from 400 to 2,400, with a step size of 0.5 unit and a dwell time of 1 ms.

Urea-Denaturation—Urea-induced unfolding of proteins was studied by difference ultraviolet spectroscopy (*26*) at 292 nm with a Shimadzu UV-3100PC spectrophotometer. All samples were dialyzed against 10 mM potassium phosphate, pH 7.8, 0.2 mM EDTA, and 1 mM 2-mercapto-

Downloaded from http://jb.oxfordjournals.org/ at Islamic Azad University on October 1, 2012

ethanol. The proteins were diluted to various final urea concentrations in the above buffer and then incubated for at least 30 min at 15°C prior to the measurement of spectra. The signal was found to be invariant after 30 min, demonstrating that equilibrium had been reached. The final protein concentrations ranged from 0.1 to 0.5 mg/ml. The unfolding transition was judged to be fully reversible from the recovery of the absorption spectra at various urea concentrations after incubation in 8 M urea. This reversibility was also demonstrated by the complete recovery of the enzymatic activity after incubation in 8 M urea and dilution with the buffer. Equilibrium unfolding data obtained from the extinction coefficient difference at 292 nm were analyzed based on a two-state model, as described previously (*18, 29*).

CD-Measurements—Circular dichroism spectra were obtained with a 15 s averaging time from 250 to 190 nm using 1 or 2 mm cuvettes with protein concentrations of 0.02-0.2 mg/ml in 10 mM potassium phosphate buffer, pH 7.8, containing 0.2 mM EDTA and 1 mM 2-mercaptoethanol, at 15°C on an Aviv spectrometer, model 62DS. The raw data were converted to mean residue ellipticity (MRE) using the following equation:

$$MRE = (\Theta \times 100)/(C \times D \times N_A)$$

where Θ is the ellipticity in degrees, C the molar protein concentration, D the pathlength in cm, and N_A the number of residues in the protein.

Others—The N-terminal amino acid sequence was determined by Edman degradation with a Beckman System Gold LF3000 protein sequencer equipped with an on-line PTH amino acid analyzer. The cleavage reaction with lysine specific endoprotease was carried out in 20 mM Tris-HCl, pH 9.3, 1 mM EDTA, and 1 M guanidine-HCl at 37°C overnight.

RESULTS

Construction of a Permuted DHFR Gene—DHFR is a two domain protein consisting of a major domain and a nucleotide binding domain (*12*). The major domain is composed of two separate parts; βA-loop-αB and βE-αF-turn-βF-loop-βG-loop-βH, while the nucleotide binding domain is composed of a single part: βB-αC-turn-βC-turn-βD-αE. As already demonstrated by the successful construction of fusion proteins from two or more independent folding units (*30, 31*), a permutation between separate domains is not so interesting for seeking structural elements. Therefore, in

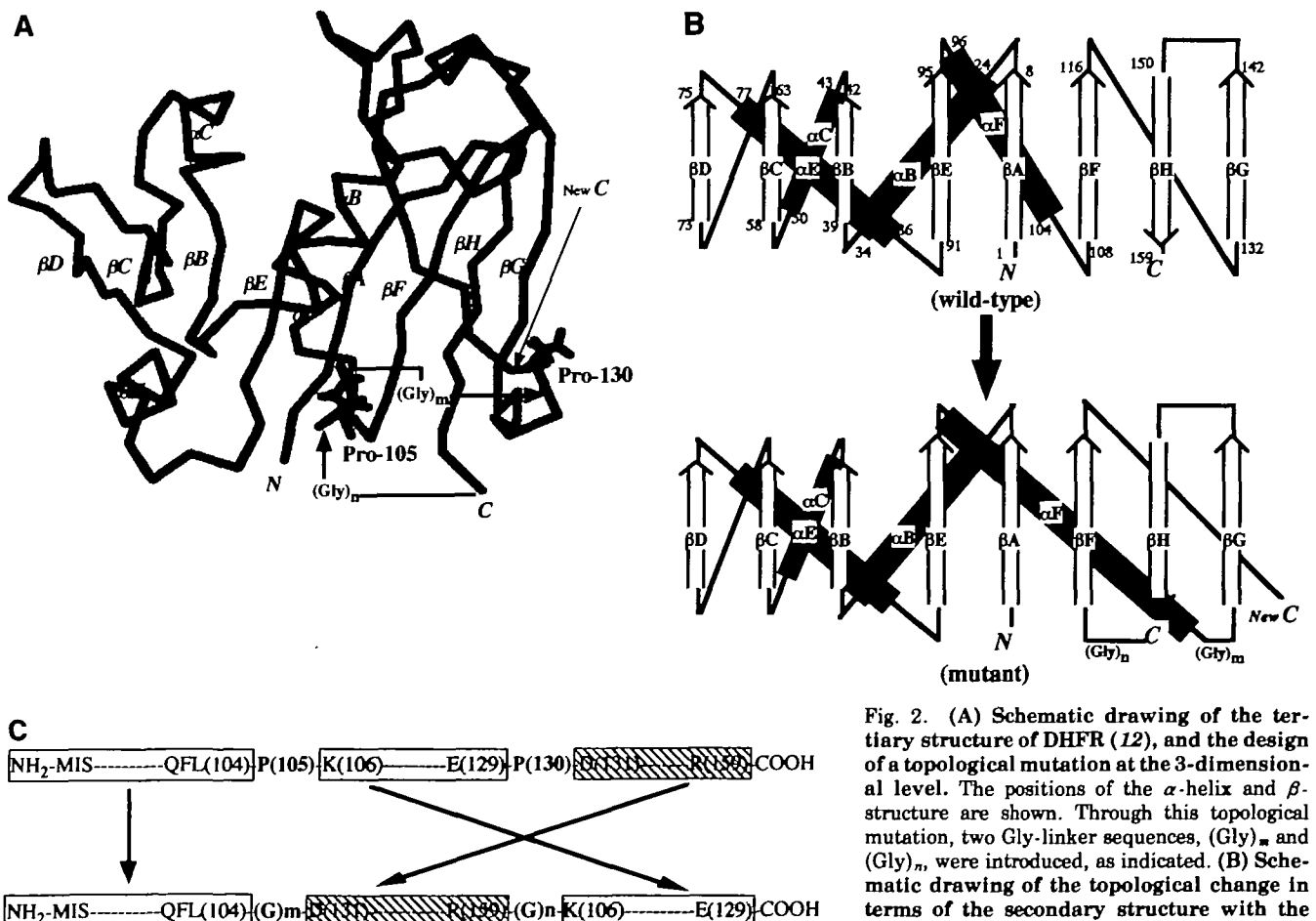

Fig. 2. **(A) Schematic drawing of the tertiary structure of DHFR (*12*), and the design of a topological mutation at the 3-dimensional level.** The positions of the α-helix and β-structure are shown. Through this topological mutation, two Gly-linker sequences, $(Gly)_m$ and $(Gly)_n$, were introduced, as indicated. **(B) Schematic drawing of the topological change in terms of the secondary structure with the topological mutation.** With this mutation, the connectivity of the last three β-strands, (αF)→βF→βG→βH, was changed to (αF)→βG→βH→βF. **(C) Design of a topological mutation at the amino acid sequence level.** The sequences comprising residues 106-129 and 131-159 of wild-type DHFR were switched with each other, and the connecting proline residues were replaced by glycine oligomers.

Downloaded from http://jb.oxfordjournals.org/ at Islamic Azad University on October 1, 2012

this study, we tried to permute the last three β-strands within the second part of the major domain (Fig. 2).

With the permutation, the chain topology of the secondary structures, namely, $\beta A \rightarrow \alpha B \rightarrow \beta B \rightarrow \alpha C \rightarrow \beta C \rightarrow \beta D \rightarrow \alpha E \rightarrow \beta E \rightarrow \alpha F \rightarrow \underline{\beta F} \rightarrow \underline{\beta G} \rightarrow \underline{\beta H}$, is changed to $\beta A \rightarrow \alpha B \rightarrow \beta B \rightarrow \alpha C \rightarrow \beta C \rightarrow \beta D \rightarrow \alpha E \rightarrow \beta E \rightarrow \alpha F \rightarrow \underline{\beta G} \rightarrow \underline{\beta H} \rightarrow \underline{\beta F}$, and the connecting sequences between αF and βF, and βF and βG should be broken. Because the connecting sequences between secondary structural motifs have probably been optimized during natural evolution and because proline residues are present in the turn regions in the connecting sequences between both αF and βF, and βF and βG, the proline residues (P105 and P130) in these sequences were replaced by glycine oligomers to introduce flexibility to the un-optimized connections. If, through this permutation, the ability to form the folded wild-type structure of the $\alpha F \rightarrow \beta F \rightarrow \beta G \rightarrow \beta H$ region was lost, the major domain could not be formed any more, and then the tertiary structure responsible for the DHFR activity would be lost. On the contrary, if the chain topology was not crucial in determination of the folded structure, the folded structure of the $\alpha F \rightarrow \beta G \rightarrow \beta H \rightarrow \beta F$ region should be similar to that of the counterpart in the wild-type protein, and a topological mutant would show DHFR activity.

The sequences of topological mutants (designated as GHF33 and GHF34) of DHFR were designed as shown in Fig. 2C: the two amino acid sequences corresponding to the βF strand and βG-βH strands in the primary sequence were switched, and the two proline residues (Pro105 and Pro130) located between the αF and βF strands, and the βF and βG strands were replaced by glycine oligomers ($m=n=3$ for GHF33; $m=3$, $n=4$ for GHF34). Apparently, amino acid replacements other than Pro105 and Pro130 did not occur with this design.

Permuted DHFR genes were constructed essentially by the overlapping extension PCR method (Fig. 3) with the synthetic primers listed in Table I. The reconstructed DHFR gene on pTZwt1-3 (*18*) was used as a DNA template. The permuted gene thus obtained was inserted to the *Bam*HI site of the pUC118 vector and used to transform *E. coli*. The genes for both GHF33 and GHF34 transformed *E. coli* cells to Tmp^R and Amp^R as a result of overproduction of the active DHFR mutants. Two plasmids (designated as pGHF33 and pGHF34) were isolated and their gene sequences were confirmed by DNA sequencing. From the transformed cells, the mutant proteins (GHF33 and GHF34) were purified to homogeneity mainly by MTX-affinity chromatography. The homogeneity of the proteins was examined by SDS-polyacrylamide gel electrophoresis (data not shown) and LC/MS. Figure 4 shows a typical LC/MS chromatogram of GHF33 and its mass spectrogram. The mass numbers of GHF33 and GHF34 were determined

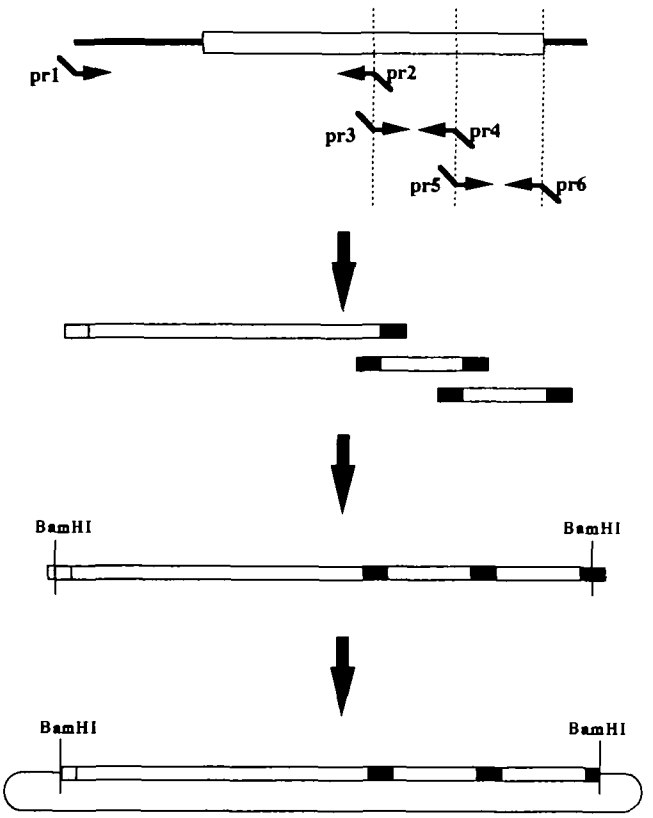

Fig. 3. **Schematic drawing of the construction procedure for a topological mutant DHFR gene.** The reconstructed DHFR gene (including the promoter region) carried on pTZwt (*18*), which overproduces wild-type DHFR, was used as a template for PCR amplification of gene fragments. The sequences of the PCR primers used in this study are presented in Table I. The PCR products were used for the 2nd PCR in which the three gene pieces were combined through homologous recombination. The respective pairs, (pr2 and pr5) and (pr3 and pr6), contained a complementary sequence in the 5′ region for homologous recombination. Primers pr1 and pr4 contain a *Bam*HI site. The *Bam*HI fragment of the 2nd PCR product was used to insert into the *Bam*HI site of a vector pUC118, and the recombinant plasmids were used to transform *E. coli* to Amp^R and Tmp^R.

TABLE I. **List of the primers used in this study.**

Primer	Sequence
Common primer	
pr1	5′-CCCGGATCCG CTCTTGACAA TTAGTTAACT ATTTG-3′
pr4	5′-GGGATCCCTT ACTCGTAATC CGGAAAATGG GTGTCGCC-3′
For construction of GHF33 gene	
pr2	5′-CCAGTCATCG CCGCCGCCCA AGAACTGTTC ATAAACGCGT CCGCCGCCAA-3′
pr3	5′-CGTCGTGGTG GTGGTAAAGC GCAAAAGCTT TATCTGACGC ATA-3′
pr5	5′-TTCTTGGGCG GCGGCGATGA CTGGGAATCG GTATTCAGCG AA-3′
pr6	5′-CGCTTTACCA CCACCACGAC GCTCGAGGAT TTCGAAACAA-3′
For construction of GHF34 gene	
pr2	5′-CCAGTCATCG CCGCCGCCGC CCAAGAACTG TTCATAAACG CGTCCGCCGC CAA-3′
pr3	5′-CGTCGTGGTG GTGGTGGTAA AGCGCAAAAG CTTTATCTGA CGCATA-3′
pr5	5′-TTCTTGGGCG GCGGCGGCGA TGACTGGGAA TCGGTATTCA GCGAA-3′
pr6	5′-CGCTTTACCA CCACCACCAC GACGCTCGAG GATTTCGAAA CAA-3′

Downloaded from http://jb.oxfordjournals.org/ at Islamic Azad University on October 1, 2012

to be 18,145 and 18,202, respectively, which coincided with the molecular masses calculated from the predicted amino acid sequences (Table II). The N-terminal amino acid sequences of GHF33 and GHF34 were exactly the same as that of the wild-type enzyme as far as determined (up to 10 cycles). For further confirmation of the protein sequence, these proteins were digested with a Lys-specific endo-peptidase, *Achromobacter* protease I, and the fragmented products were analyzed by LC/MS. As shown in Fig. 5, five peptide fragments were obtained for both GHF33 and

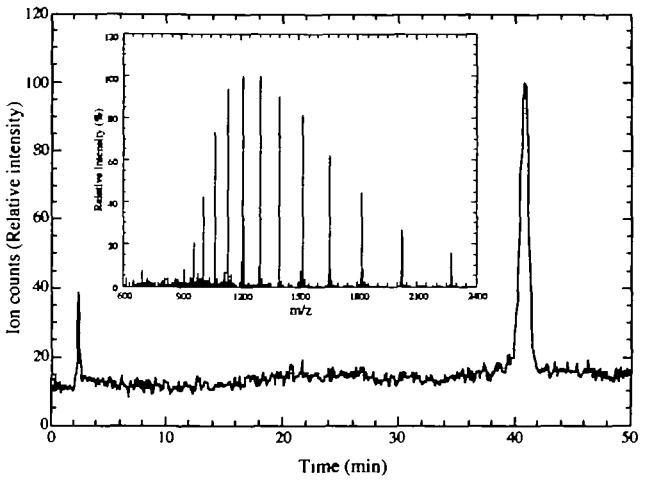

Fig. 4. **On line LC/MS chromatogram of a topological mutant, GHF33.** RP-HPLC was carried out with a linear gradient of 10-60% acetonitrile (in 0.1% TFA) over 45 min at the flow rate of 0.2 ml/min. The inset shows a mass-spectrum of the peak fraction. From the multiple charge mass peaks, the molecular mass of GHF33 was determined to be 18,145 Da.

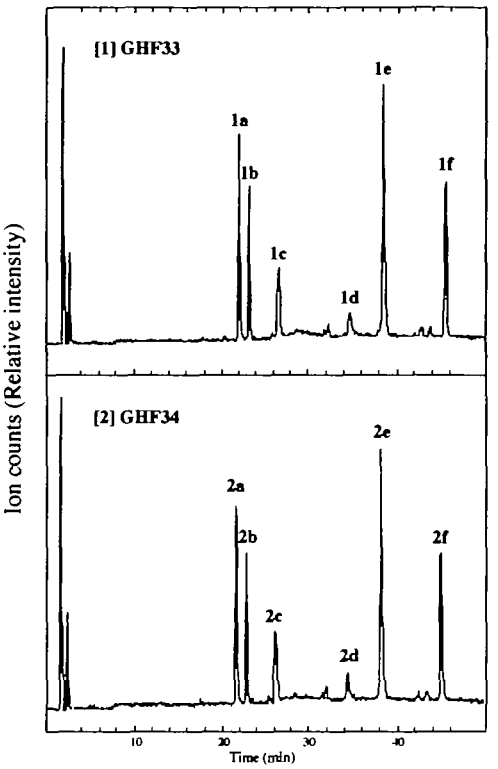

Fig. 5. **On line LC-MS chromatogram of the digested products of GHF33 [1] and GHF34 [2] with lysine specific endoprotease.** Reverse phase HPLC was carried out with a linear gradient of 10-60% acetonitrile (in 0.1% TFA) over 45 min at the flow rate of 0.2 ml/min. The molecular masses of the peak a, b, c, d, and e materials were determined, and are indicated in Table III.

TABLE II. **Molecular masses, Michaelis parameters, and thermodynamic parameters for the urea-induced unfolding reaction of wild-type DHFR and its permuted variants.**

Protein (sequence based on wild-type numbering)	Molecular mass (Da)	Enzyme activity		Urea denaturation	
	measured (calculated)	k_{cat} (s^{-1})	K_m for DHF (μM)	ΔG^{H_2O} (kcal/mol)	A [kcal/mol/(M urea)]
GHF33 ([1-104]-GGG-[131-159]-GGG-[106-129])	18,145 (18,147)	0.53	3.3	4.0	1.6
GHF34 ([1-104]-GGG-[131-159]-GGGG-[106-129])	18,202 (18,204)	1.12	2.1	4.8	2.0
Wild-type ([1-104]-P-[106-129]-P-[131-159])	17,997 (17,999)	5.83	0.9	6.1	2.0

TABLE III. **Summary of molecular mass measurements of proteolytic fragments with Lys-specific endo-peptidase.**

Peak in Fig. 5	Molecular mass (Da)		N-terminal sequence	Assignment
	Measured	Calculated		
[1] GHF33				[M1-L104]+GGG+[D131-R159]+GGG+[K106-E129]
peak 1a	3,014	3,013.7	RNTLN	[R33-K58]
peak 1b	2,028	2,029.1	NIILS	[N59-K76]
peak 1c	2,364	2,364.1	LYLTH	[L110-E129]
peak 1d	2,690	2,691.3	AQKLY	[A107-E129]
peak 1e	6,921	6,922.1	SVDEA	[S77-L104]+GGG+[D131-R159]+GGG+K106
peak 1f	3,555	3,556.9	MISIL	[M1-K32]
[2] GHF34				[M1-L104]+GGG+[D131-R159]+GGGG+[K106-E129]
peak 2a	3,013	3,013.7	RNTLN	[R33-K58]
peak 2b	2,028	2,029.1	NIILS	[N59-K76]
peak 2c	2,364	2,364.1	LYLTH	[L110-E129]
peak 2d	2,690	2,691.3	AQKLY	[A107-E129]
peak 2e	6,977	6,979.2	SVDEA	[S77-L104]+GGG+[D131-R159]+GGGG+K106
peak 2f	3,556	3,556.9	MISIL	[M1-K32]

Downloaded from http://jb.oxfordjournals.org/ at Islamic Azad University on October 1, 2012

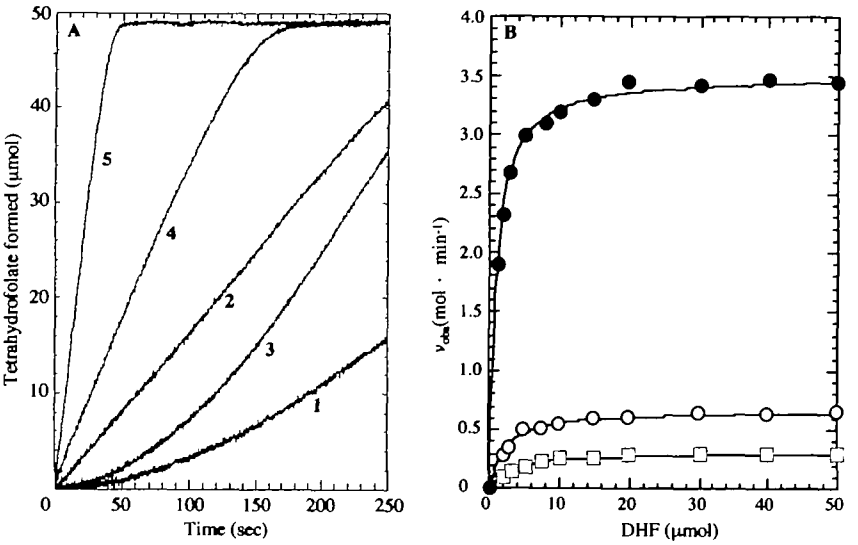

Fig. 6. **The enzymatic activity of topological mutants of DHFR.** (A) Hysteretic behavior of the mutant and wild-type enzymes (at the enzyme concentration of 0.4 μM). Curve 1, GHF33 without pre-incubation with NADPH; curve 2, GHF33 with pre-incubation with NADPH for 10 min; curve 3, GHF34 without pre-incubation with NADPH; curve 4, GHF34 with pre-incubation with NADPH; curve 5, wild-type DHFR without pre-incubation with NADPH. (B) Saturation profile of the permuted and wild-type enzymes. Initial velocity (v_{obs}) was plotted against the initial concentration of DHF. To eliminate the hysteric behavior, the enzymes were pre-incubated with NADPH for more than 5 min prior to initiation of the reaction with dihydrofolate. (□) GHF33; (○) GHF34; (●) wild-type.

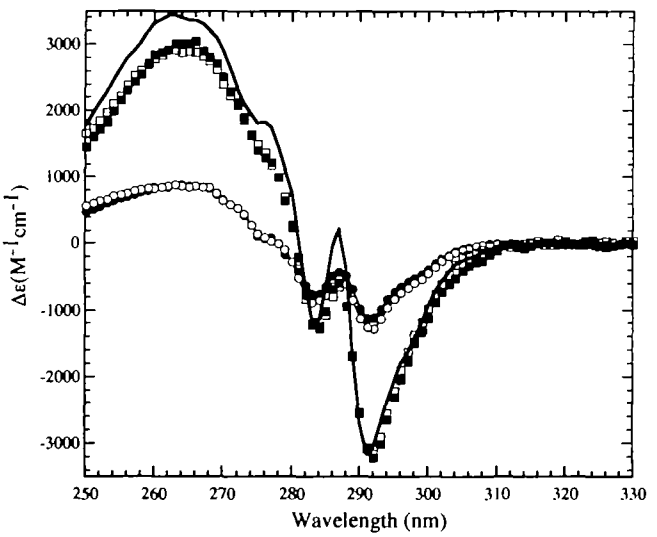

Fig. 7. **Difference absorption spectra in the near UV region of the native and denatured forms of GHF33 and GHF34 derived from wild-type DHFR.** (○) GHF33 (native form)-wild-type DHFR (native form); (●) GHF34 (native form)-wild-type DHFR (native form); (□) GHF33 (denatured in 8 M urea)-wild-type DHFR (native form); (■) GHF34 (denatured in 8 M urea)-wild-type DHFR (native form); (——) wild-type DHFR (denatured in 8 M urea)-wild-type DHFR (native form).

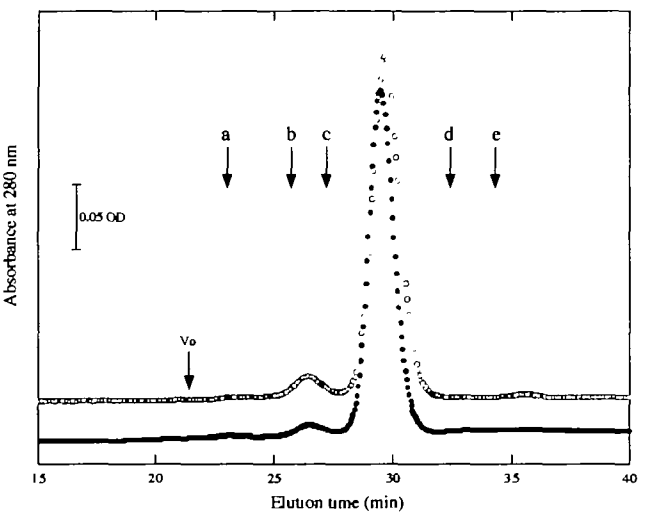

Fig. 8. **Size-exclusion chromatography of GHF34 (○) and wild-type DHFR (●).** Arrows indicate the elution peaks of standard proteins: a, bovine serum albumin (R_s = 33.9 Å); b, egg albumin (R_s = 31.2 Å); c, β-lactoglobulin (R_s = 27.1 Å); d, myoglobin (R_s = 20.2 Å); e, ribonuclease (R_s = 19.3 Å).

GHF34, and their molecular masses and assignments are presented in Table III. Also, their N-terminal amino acid sequences were determined and are summarized in Table III. All the results indicated that GHF33 and GHF34 contained the permuted sequence designed.

Enzymatic Activity—The GHF33 and GHF34 proteins showed about 10 and 20% of the DHFR activity of the wild-type under normal assay conditions, respectively, although the mutated proteins showed an obvious lag phase (Fig. 6A). The lag phase could be eliminated by pre-incubation of the proteins with the coenzyme, NADPH, but not with the substrate, dihydrofolate. The k_{cat} and K_m values for the GHF33 protein were 0.53 (s^{-1}) and 3.3 μM, and those for the GHF34 protein were 1.12 (s^{-1}) and 2.1 μM, at 15°C, respectively (Fig. 6B). While those for the wild-type enzyme were 5.83 (s^{-1}) and 0.9 μM, respectively.

UV Absorbance—The UV absorbance spectra of the GHF33 and GHF34 proteins did not completely coincide with that of the wild-type under the native conditions. The different spectra showed that some tryptophan residues in the mutated proteins were more exposed to the solvent than those in the wild-type enzyme (Fig. 7). On the contrary, under the denatured conditions (8 M urea), the spectra of the mutated and wild-type proteins were similar to each other, indicating that all tryptophan residues were fully exposed in the denatured proteins. It is noteworthy that the differences ($\Delta\varepsilon_{280}$ values) in absorbance at 280 nm, at which the protein concentrations were determined, between GHF34 (and also GHF33) and the wild-type protein under both native and denatured conditions were less than 300 M^{-1}·cm^{-1} (*ca.* 1% of the extinction coefficient

Downloaded from http://jb.oxfordjournals.org/ at Islamic Azad University on October 1, 2012

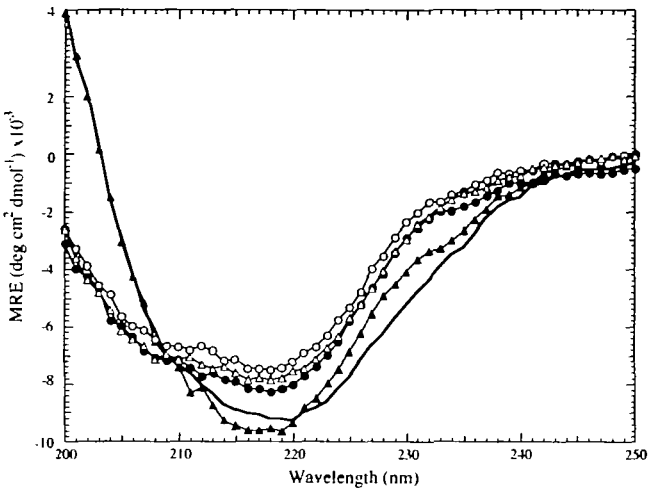

Fig. 9. **Far UV CD spectra of the proteins in the presence and absence of 5 μM NADPH.** (○) GHF33; (●) GHF34; (△) GHF34 with NADPH; (——) wild-type DHFR; (▲) wild-type DHFR with NADPH.

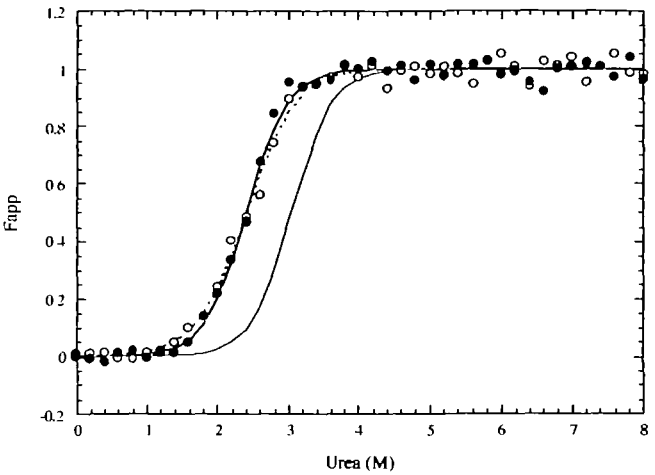

Fig. 10. **Dependence of the apparent fraction of unfolded protein, Fapp, on the concentration of urea for GHF33 (○), GHF34 (●), and wild-type DHFR (——).**

of the wild-type protein, $\varepsilon_{280} = 3.11 \times 10^4\ M^{-1} \cdot cm^{-1}$).

Size-Exclusion Chromatography—Figure 8 shows the elution patterns of GHF34 and wild-type DHFR from a Superdex 75 PC3.2/30 column. Two peaks corresponding to the monomer (major peak) and its associated dimer form (minor peak) were obtained, however, the fractions of the associated dimer form of GHF34 and wild-type DHFR were as low as 5 and 3%, respectively. The elution time for GHF34 was 29.6 min, which was longer by 0.2 min than that of wild-type DHFR (29.4 min), namely, GHF34 has a smaller storks radius than wild-type DHFR, although the difference is quite small. Also, GHF33 showed a similar elution pattern to that of GHF34 (data not shown).

CD Spectra—As shown in Fig. 9, the far UV CD spectra of GHF33 and GHF34 were different from that of wild-type DHFR. Typically, the negative CD signals between 215 and 235 nm were decreased. Roughly speaking, about 20 and 10% of the CD signal at 220 nm was lost in GHF33 and GHF34, respectively. Usually, such a decrease in the negative CD signal can be related to a loss of fractions of α and β structures in mutant proteins. However, in our topological mutation, oligoglycine linkers were introduced to connect the separated polypeptides. In this way, the fraction of random coil may be increased (several percent). In addition, the side chain coupling of Trp47 and Trp74 residues in *E. coli* DHFR led to positive CD in this region (*32*). As shown by the different spectra, some tryptophan residues were more exposed to the solvent in GHF33 and GHF34 (Fig. 7), which might have caused the change in the CD spectra. The real loss of α and β structures may, therefore, be smaller than that expected from the observed CD spectra. The CD spectra of GHF34 in the presence of a coenzyme (NADPH) were essentially the same as those in its absence, while NADPH induced a spectral change for wild-type DHFR. The shape of the wild-type spectra in the presence of NADPH became closer to those of GHF33 and GHF34 (Fig. 9). The substrate, DHF, did not induce a spectral change for GHF33, GHF34, or wild-type DHFR, as far as tested.

Stability—GHF33 and GHF34 showed cooperative 2-state transitions in urea-induced unfolding experiments (Fig. 10). The ΔG^{H_2O} (the free energy difference between the native and unfolded proteins in the absence of urea) and A (the factor representing the urea dependence of the free energy change) values for the GHF33 protein were 4.0 kcal/mol and 1.6 kcal/mol/M (urea), and those for the GHF34 protein were 4.8 kcal/mol and 2.0 kcal/mol/M (urea), respectively. While those for the wild-type enzyme were 6.1 kcal/mol and 2.0 kcal/mol/M (urea), respectively.

DISCUSSION

Although the topological mutants, GHF33 and GHF34, showed destabilized properties, these mutant proteins could fold into a stable structure and showed the same catalytic function as the original wild-type protein. In general, local structural changes at and near the active site of an enzyme cause a drastic decrease in its catalytic function (*2*). Although the far UV CD spectra of the topological mutants were different from that of the wild-type enzyme, the active site conformation (which might be determined by mutual interaction among the postulated structural elements) should be essentially the same. Otherwise, the catalytic activity must be essentially lost. In this context, the changes in the CD spectra may be due to the structural change(s) in connecting region(s) or tryptophan side chain effects. Therefore, we concluded that chain topology is not so critical for building of the active enzyme. To prove this, crystallographic analysis is necessary and we are trying to crystallize the mutant proteins, although we have not been successful yet.

Why the chain topology is not important can be explained simply by assuming the presence of structural elements in the protein sequence, as already shown in Fig. 1. On the contrary, if all the amino acids in the primary sequence equally contribute to determination of the folded structure, our present results are hard to explain. A similar approach to examine the presence of structural elements has been taken involving dissection analysis of proteins (*1, 2, 33-37*). In many cases, complementary fragments can associate and fold to give active molecules. However, such

Downloaded from http://jb.oxfordjournals.org/ at Islamic Azad University on October 1, 2012

dissection analysis only suggests the presence of structural elements, *i.e.* it does not strictly prove the insignificance of the chain topology in determination of the folded structure.

The protein backbone consists of continuous main chain covalent bonds and plays a key role in determination of protein structure and function. The covalent peptide bond has a direction and is not symmetrical. For example, Ala-Gly is different from Gly-Ala. Thus, the order of amino acids is crucial and defines the folded structure. Recently, linguistic analysis of protein folding has been introduced to classify protein folds in terms of chain topology (*38*). A sentence consists of linear strings of letters that build up into words, phrases, clauses, and so on. This is analogous to protein structures. In this sense, to determine what corresponds to a word, phrase and so on in protein structures would be of great help for elucidation of how the primary sequence determines the folded tertiary structure. This study intended in part to apply linguistic analysis to the definition of a protein sub-structure experimentally. The order of words or phrases in a sentence can be permuted (but not always) without losing the original meaning, while the order of letters in a word cannot be permuted, for example, "The mutant was then purified" has the same meaning as "Then the mutant was purified," but not as "The mutant was hten purified." Indeed, as shown in this study, the former is the case in DHFR. The DHFR sequence could be separated into three parts, which might correspond to a phrase. Systematic permutation analysis would provide a "grammar" for protein sequences and, therefore, a useful approach for solving the protein folding problem.

Our present results also suggest that the same enzymatic function, probably resulting from a similar folded structure, can be obtained by intra-molecular swapping of structural units with proper linkers (Fig. 1, [4] and [5]), namely, by means of a topological mutation. It is noteworthy that the topological mutant protein, GHF34, showed as high as 20 and 8% of the original wild-type activity in terms of k_{cat} and k_{cat}/K_m, respectively, even though the mutant structure was not optimized with respect to stability. If an ancestor DHFR protein was formed from mini-peptides (*39*), there should be a number of topological mutant proteins with the same function. The chain topologies of DHFRs from various organisms are essentially the same and it is quite difficult to find a topologically different DHFR among naturally evolved DHFRs. The high conservation of the chain topology in DHFR suggests that the wild-type chain topology was established by chance when the ancestor DHFR arose. Otherwise, DHFRs with different chain topologies such as those presented in this work should exist in natural sources, and this is not the case. Also, the natural mutation process seems to preclude the conservation of the chain topology except for gene shuffling. This may be why a homology search in functionally or structurally related proteins is applicable.

Our present results also suggest that a folded structure can be obtained through combinatorial connecting structural elements with proper linkers. In general, it is quite difficult to find a polypeptide with a stable folded structure derived from a random amino acid sequence. In contrast, if we could define structural elements, it should become possible to construct a certain folded structure through random combination of the structural elements with great efficiency, and such a strategy would be useful for creating novel functional proteins.

The authors are grateful to Ms. Naomi Furuya, National Institute of Bioscience and Human-Technology, for her technical assistance with the protein purification.

REFERENCES

1. Taniuchi, H. and Anfinsen, C.B. (1971) Simultaneous formation of two alternative enzymology active structures by complementation of two overlapping fragments of staphylococcal nuclease. *J. Biol. Chem.* **246**, 2291-2301
2. Bertolaet, B.L. and Knowles, J.R. (1995) Complementation of fragments of triosephosphate isomerase defined by exon boundaries. *Biochemistry* **34**, 5736-5743
3. Fersht, A.R. (1997) Nucleation mechanisms in protein folding. *Curr. Opin. Struct. Biol.* **7**, 3-9
4. Woolfson, D.N., Evans, P.A., Hutchinson, E.G., and Thornton, J.M. (1993) Topological and stereochemical restrictions in beta-sandwich protein structures. *Protein Eng.* **6**, 461-470
5. Ladurner, A.G. and Fersht, A.R. (1997) Glutamine, alanine or glycine repeats inserted into the loop of a protein have minimal effects on stability and folding rates. *J. Mol. Biol.* **273**, 330-337
6. Nolting, B., Golbik, R., Neira, J.L., Soler-Gonzalez, A.S., Schreiber, G., and Fersht, A.R. (1997) The folding pathway of a protein at high resolution from microseconds to seconds. *Proc. Natl. Acad. Sci. USA* **94**, 826-830
7. Kabsch, W. and Sander, C. (1984) On the use of sequence homologies to predict protein structure: identical pentapeptides can have completely different conformations. *Proc. Natl. Acad. Sci. USA* **81**, 1075-1078
8. Sander, C. and Schneider, R. (1991) Database of homology-derived protein structures and the structural meaning of sequence alignment. *Proteins* **9**, 56-68
9. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.* **215**, 403-410
10. 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
11. 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
12. Bystroff, C. and Kraut, J. (1991) Crystal structure of unliganded *Escherichia coli* dihydrofolate reductase. Ligand-induced conformational changes and cooperativity in binding. *Biochemistry* **30**, 2227-2239
13. Villafranca, J.E., Howell, E.E., Voet, D.H., Strobel, M.S., Ogden, R.C., Abelson, J.N., and Kraut, J. (1983) Directed mutagenesis of dihydrofolate reductase. *Science* **222**, 782-788
14. Perry, K.M., Onuffer, J.J., Touchette, N.A., Herndon, C.S., Gittelman, M.S., Matthews, C.R., Chen, J.T., Mayer, R.J., Taira, K., and Benkovic, S.J. (1987) Effect of single amino acid replacements on the folding and stability of dihydrofolate reductase from *Escherichia coli*. *Biochemistry* **26**, 2674-2682
15. Garvey, E.P., Swank, J., and Matthews, C.R. (1989) A hydrophobic cluster forms early in the folding of dihydrofolate reductase. *Proteins* **6**, 259-266
16. Dunn, S.M., Lanigan, T.M., and Howell, E.E. (1990) Dihydrofolate reductase from *Escherichia coli*: probing the role of aspartate-27 and phenylalanine-137 in enzyme conformation and the binding of NADPH. *Biochemistry* **29**, 8569-8576
17. Leontiev, V.V., Uversky, V.N., and Gudkov, A.T. (1993) Comparative stability of dihydrofolate reductase mutants in vitro and in vivo. *Protein Eng.* **6**, 81-84
18. 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
19. Ohmae, E., Iriyama, K., Ichihara, S., and Gekko, K. (1998)

Downloaded from http://jb.oxfordjournals.org/ at Islamic Azad University on October 1, 2012

Nonadditive effects of double mutations at the flexible loops, glycine-67 and glycine-121, of *Escherichia coli* dihydrofolate reductase on its stability and function. *J. Biochem.* **123**, 33-41
20. Buchwalder, A., Szadkowski, H., and Kirschner, K. (1992) A fully active variant of dihydrofolate reductase with a circularly permuted sequence. *Biochemistry* **31**, 1621-1630
21. Protasova, N.Yu, Kireeva, M.L., Murzina, N.V., Murzin, A.G., Uversky, V.N., Gryaznova, O.I., and Gudkov, A.T. (1994) Circularly permuted dihydrofolate reductase of *E. coli* has functional activity and a destabilized tertiary structure. *Protein Eng.* **7**, 1373-1377
22. Iwakura, M. and Nakamura, T. (1998) Effects of the length of a glycine linker connecting the N- and C-termini of a circularly permuted dihydrofolate reductase. *Protein Eng.* **11**, 707-713
23. Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982) *Molecular Cloning. A Laboratory Manual.* Cold Spring Harbor Laboratory, Cold Spring Harbor, New York
24. Iwakura, M., Furusawa, K., Kokubu, T., Ohashi, S., Shimura, Y., and Tsuda, K. (1992) Dihydrofolate reductase as a new "affinity handle." *J. Biochem.* **111**, 37-45
25. Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680-685
26. Touchette, N.A., Perry, K.P., and Matthews, C.R. (1986) Folding of dihydrofolate reductase from *Escherichia coli*. *Biochemistry* **25**, 5445-5452
27. Hillcoat, B., Nixon, P., and Blakely, R.L. (1967) Effect of substrate decomposition on the spectrophotometric assay of dihydrofolate reductase. *Anal. Biochem.* **21**, 178-189
28. Morrison, J.F. and Stone, S.R. (1988) Mechanism of the reaction catalyzed by dihydrofolate reductase from *Escherichia coli*: pH and deuterium isotope effects with NADPH as the variable substrate. *Biochemistry* **27**, 5499-5506
29. Santoro, M.M. and Bolen, D.W. (1988) Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl alpha-chymotrypsin using different denaturants. *Biochemistry* **27**, 8063-8068
30. Iwakura, M. and Matthews, C.R. (1993) Construction and characterization of a single polypeptide chain containing two enzymatically active dihydrofolate reductase domains. *Protein Eng.* **5**, 791-796
31. Iwakura, M. and Kokubu, T. (1995) Production of hetero-dimeric dihydrofolate reductase-thymidylate synthase bifunctional enzyme. *J. Biochem.* **118**, 67-74
32. Kuwajima, K., Garvey, E.P., Finn, B.E., Matthews, C.R., and Sugai, S. (1991) Transient intermediates in the folding of dihydrofolate reductase as detected by far-ultraviolet circular dichroism spectroscopy. *Biochemistry* **30**, 7693-7703
33. Tasayco, M.L. and Carey, J. (1992) Ordered self-assembly of polypeptide fragments to form nativelike dimeric trp repressor. *Science* **255**, 594-597
34. Wu, L.C., Laub, P.B., Elove, G.A., Carey, J., and Roder, H. (1993) A noncovalent peptide complex as a model for an early folding intermediate of cytochrome c. *Biochemistry* **32**, 10271-10276
35. Wu, L.C., Grandori, R., and Carey, J. (1994) Autonomous subdomains in protein folding. *Protein Sci.* **3**, 369-371
36. Peng, Z.Y. and Kim, P.S. (1994) A protein dissection study of a molten globule. *Biochemistry* **33**, 2136-2141
37. Gegg, C.V., Bowers, K.E., and Matthews, C.R. (1997) Probing minimal independent folding units in dihydrofolate reductase by molecular dissection. *Protein Sci.* **6**, 1885-1892
38. Gross, M. (1996) Linguistic analysis of protein folding. *FEBS Lett.* **390**, 249-252
39. Seidel, H.M., Pompliano, D.L., and Knowles, J.R. (1992) Exons as microgenes? *Science* **257**, 1489-1490
40. Iwakura, M. (1996) An intramolecular shuffling of secondary structural motifs in *E. coli* dihydrofolate reductase yields an active enzyme. *Protein Sci.* **5**, Suppl. 1, 68

Downloaded from http://jb.oxfordjournals.org/ at Islamic Azad University on October 1, 2012