REVIEW ARTICLES

Mutant Sequences as Probes of Protein Folding Mechanisms

C. Robert Matthews and Mark R. Hurle

Summary

Mutagenesis makes it possible to examine the effect of amino acid replacements on the folding and stability of proteins. The evaluation of kinetic and equilibrium folding data using reaction coordinate diagrams allows one to determine the roles that single amino acids play in the folding mechanism.

Introduction

The central dogma in the field of protein folding is that the amino acid sequence of a protein determines its three-dimensional structure. Although this hypothesis was first proposed over two decades ago by Anfinsen,[1] there has been rather little progress in understanding the mechanism by which this complex conformational change occurs. Most workers believe that folding is a directed search involving discrete intermediates, and a number of different models have been proposed.[2] However, evidence for structural assignments for the intermediates has proven to be exceedingly difficult to obtain. The principal problem is the high cooperativity of the unfolding reaction. Only the native and unfolded forms are significantly populated at equilibrium during the unfolding of most globular proteins. Intermediate, partially folded forms that would be very useful in identifying the folding pathway generally do not appear at sufficient concentrations to be characterized by high-resolution spectroscopic techniques such as NMR or X-ray crystallography.

The failure to detect intermediates with equilibrium methods has led to the application of kinetic techniques to identify transiently populated species. Extensive studies on the folding of bovine pancreatic ribonuclease A have demonstrated the presence of transient intermediates in folding and some rather general information on their structures.[3] Although these kinetic studies have played a valuable role in demonstrating the presence of intermediates and in providing a broad outline of the mechanism of folding, it seems unlikely that they will be able to provide detailed information on the structures of the intermediates.

Site-directed mutagenesis, which permits the replacement of any amino acid in a protein with any of its naturally occurring counterparts, appears to have the potential to advance dramatically our understanding of protein folding mechanisms. This possibility exists because the concept of discrete intermediates also implies that certain amino acids will play key roles in directing folding. Using methods that are currently being formulated and which are described below, we believe that an analysis of the diverse effects of single amino acid replacements on the kinetic and equilibrium folding properties of a protein will allow one to determine rather precisely the role of a specific amino acid in that protein's folding process.

Hypothetical Effects of Mutations on Folding and Stability

The approach that we have adopted to interpret the effects of mutations on folding and stability involves the application of reaction coordinate diagrams. Consider the simplest case where the folding reaction can be described both kinetically and thermodynamically by a two-state model involving the native confirmation, N, and the unfolded form, U:

$$N \underset{k_R}{\overset{k_U}{\rightleftharpoons}} U$$

The equilibrium constant, K, is the ratio of the two microscopic rate constants, k_U and k_R, i.e. $K = k_U/k_R$, and the relaxation time, τ, for displacements in the equilibrium position is defined by $\tau^{-1} = k_U + k_R$. For unfolding jumps to high denaturant concentrations, $k_U \gg k_R$ and $\tau^{-1} \approx k_U$. Thus, for unfolding jumps, the observed relaxation time is inversely proportional to the unfolding rate constant. For refolding jumps to low denaturant concentrations, $k_R \gg k_U$ and $\tau^{-1} \approx k_R$; the observed relaxation time is inversely proportional to the refolding rate constant. Appropriate equilibrium and kinetic measurements of the unfolding and refolding reactions therefore permit the determination of the equilibrium constant and both microscopic rate constants.

These equilibrium and kinetic constants can be converted to a free energy scale by the relationships

$$\Delta G = -RT \ln K,$$
$$\Delta G^{\ddagger}_U = -RT \ln [hk_U/k_B T],$$
$$\Delta G^{\ddagger}_R = -RT \ln [hk_R/k_B T],$$

where ΔG is the free energy difference between N and U and $\Delta G^{\ddagger}_U$ and $\Delta G^{\ddagger}_R$ are the activation free energies for unfolding and refolding, respectively, as shown in Fig. 1. R, h and k_B are the gas, Planck and Boltzmann constants.

Of interest to the present discussion is the effect of amino acid replacement on the reaction coordinate diagram. Three general cases are possible:

(1) *Equilibrium Mutant*. The effect of the mutation is to alter selectively the free energy of one of the stable states. In

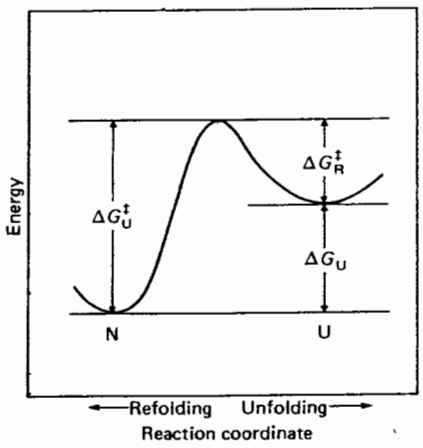

Fig. 1. *Reaction coordinate diagram for an unfolding transition between the native form, N, and the unfolded form, U.*

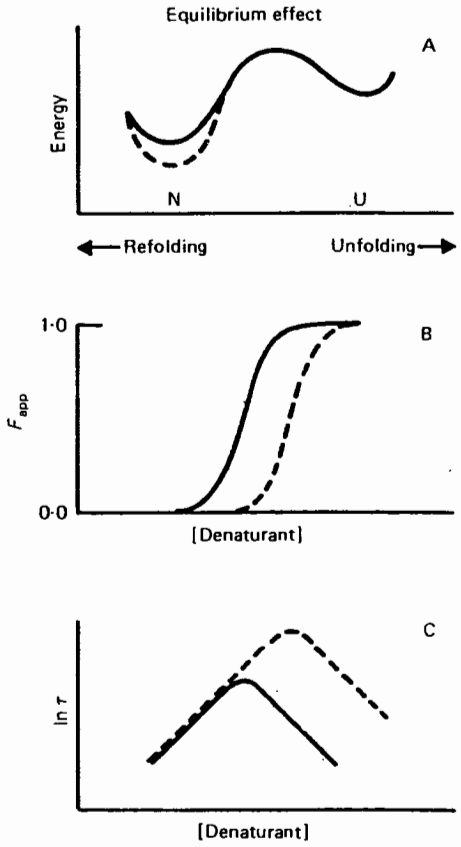

Fig. 2. *Effect on (A) reaction coordinate diagram, (B) equilibrium unfolding curve, and (C) unfolding/refolding relaxation times for a mutation which stabilizes only the native form of an ideal two-state transition.*

the example depicted in Fig. 2A the free energy of the native state has been decreased, perhaps by the introduction of an additional noncovalent interaction, e.g. a salt bridge or hydrogen bond. The energies of the transition state and unfolded forms are unchanged. The increased free-energy difference between N and U implies that the mutant is more stable than wild type and therefore that the equilibrium unfolding transition shifts to higher urea concentration (Fig. 2B). The relaxation time for unfolding is increased; however, that for refolding is unaffected (Fig. 2C). Mutations that have such effects are important in stabilizing the proteins, but do not alter the rate-limiting step in the interconversion of N and U.

Although the sigmoidal shape of the equilibrium unfolding curve is familiar, the inverted V dependence of ln τ versus denaturant concentration is less so. As described above, the relaxation time observed at high denaturant concentrations principally reflects the unfolding microscopic rate constant. This rate constant progressively increases as the equilibrium shifts to favor the unfolded form, resulting in a corresponding decrease in the relaxation time. Under refolding conditions the relaxation time reflects the refolding microscopic rate constant, which progressively increases as the equilibrium shifts to favor the native form. Consequently, the refolding relaxation time decreases as the final denaturant concentration is decreased. In the transition region, where unfolding and refolding rate constants are comparable, ln τ proceeds through a maximum. This type of behavior has been observed in several proteins.[6–8]

(2) *Kinetic Mutant.* The effect of the mutation is to alter selectively the energy of the transition state with respect to the energies of N and U (Fig. 3A). The stability of the protein is unaffected (Fig. 3B); however, the relaxation times for unfolding and refolding are altered in the same way. In the case depicted, both relaxation times are increased (Fig. 3C). Amino acid replacements that fall into this category alter the rate-limiting step between N and U.

(3) *Mixed Equilibrium-Kinetic Mutant.* The mutation alters the energies of the native state, the transition

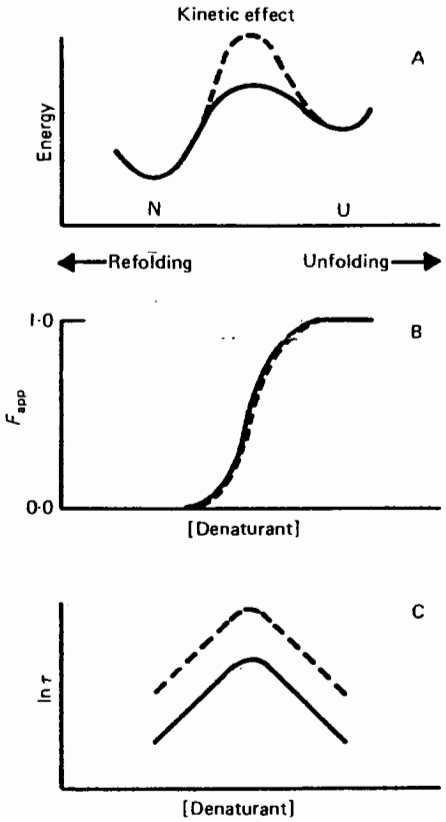

Fig. 3. *Effect on (A) reaction coordinate diagram, (B) equilibrium unfolding curve, and (C) unfolding/refolding relaxation times for a mutation which destabilizes only the transition state.*

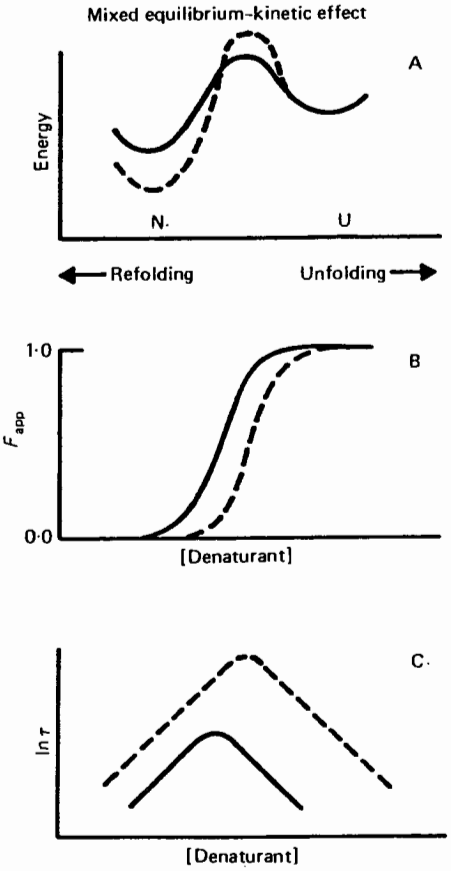

Fig. 4. *Effect on (A) reaction coordinate diagram, (B) equilibrium unfolding curve, and (C) unfolding/refolding relaxation times for a mutation which stabilizes the native form and destabilizes the transition state with respect to the unfolded form.*

state, and the unfolded states relative to each other (Fig. 4A). In the case shown, the increased stability (Fig. 4B) is reflected by non-equivalent increases in the relaxation times for unfolding and refolding (Fig. 4C). Amino acid replacements at such sites affect the stability as well as alter the rate-limiting step in folding.

The reaction coordinate diagrams shown in Fig. 2–4 have been arbitrarily aligned by equating the energies of the unfolded forms. This seems to be a reasonable first approximation since the unfolded forms are devoid of the secondary and tertiary structure whose development these reaction coordinate diagrams monitor. This assumption may not be valid, however, and effects on the energy of the unfolded form may have to be considered.[9] However, realignment of the diagrams for the wild-type and mutant proteins does not change the conclusions with respect to the type of effect or role that a particular amino acid has in the *folding* process. This point will be considered further in the Results section.

REVIEW ARTICLES

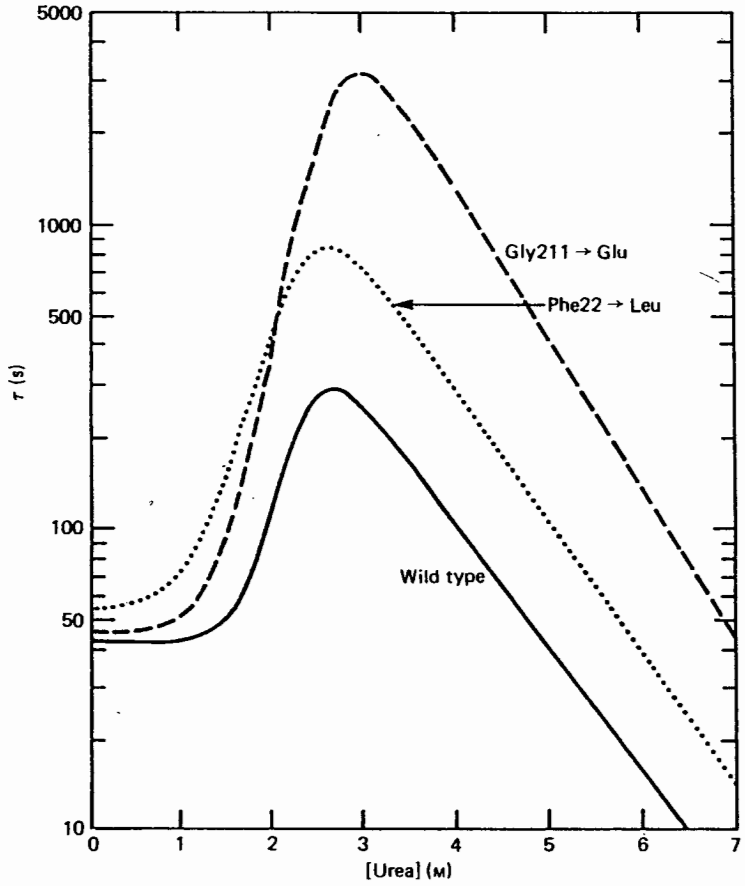

Fig. 5. *Effect on the unfolding/refolding relaxation times of the wild-type α subunit of tryptophan synthase from* E. coli (——) *by the mutations Phe22→Leu* (·····) *and Gly211→Glu* (– – –).[10] *The urea-independence of the curves at low urea concentrations is due to the presence of a rate-limiting proline isomerization step.*[10]

Experimental Results

The above classification scheme can be used to evaluate the effect of amino acid replacements on the folding and stability of the α subunit of tryptophan synthase[10] and dihydrofolate reductase[11,12] from *Escherichia coli* and the effect of a covalent cross-link in bovine pancreatic ribonuclease A.[13] As discussed below, selective effects on the energies of each of the three states have been observed. These effects allow one to discover which roles these amino acids play in the folding pathway.

The effects of the Phe22→Leu mutation in the α subunit of tryptophan synthase on one of the slow phases in folding and on the urea-induced equilibrium transition curve are shown in Fig. 5 and 6A. The kinetic phase shown in Fig. 5 closely corresponds to the transition from the native conformation to a stable intermediate[10], recognizable from the inflection in the equilibrium curve of Fig. 6A. The mutation has very little effect on the stability of the native form relative to the intermediate, but increases the relaxation times for both unfolding and refolding by nearly equal amounts. According to the above classification scheme, Phe22→Leu is very close to being a pure kinetic mutant, with the selective effect on the energy of the transition state. Therefore the amino acid at position 22 plays a key role in folding.

The Gly211→Glu mutation in the α subunit causes rather large increases in the relaxation times for both unfolding and refolding (Fig. 5) and a small increase in the stability of the native form relative to the intermediate (Fig. 6B). The 8-fold increase in the relaxation time for unfolding and the 4-fold increase in the relaxation time for refolding are consistent with the increase in stability and identify this as a mixed equilibrium-kinetic mutant. The observation that positions 22 and 211 both play key roles in the interconversion of the native and intermediate forms was instrumental in identifying the associated structural change as being either domain association or some other type of molecule-wide reaction.[10]

An example of an equilibrium mutant can be found in the Leu28→Arg mutation in dihydrofolate reductase[12] (Fig. 7A,B). The relaxation times for the two observed phases in unfolding are increased by a factor of 3 while those for these same phases in refolding are nearly unchanged (Fig. 7A). These results are consistent with the increase in stability observed in the equilibrium unfolding curve (Fig. 7B). It appears that the amino acid at position 28 does not play a key role in the rate-limiting step of folding.

Another example where a change in protein structure results in an equilibrium effect on folding and stability is provided by the recent results of Scheraga *et al.*[13] on cross-linked ribonuclease A. The introduction of a covalent cross-link between Lys7 and Lys41 results in a 60-fold increase in the relaxation time for unfolding but does not change the rate of refolding; the stability of the protein is increased by an appropriate amount. The increased stability was quantitatively explained by a reduction of the entropy of the unfolded form due to the constraint of the new cross-link, resulting in an increase in its free energy. The absence of an effect on the rate of refolding suggests that the cross-link has a similar influence on the energy of the transition state. The selective effect, therefore, is on the energy of the native state.

A third example of an equilibrium effect is demonstrated by results from our laboratory on the Glu49→Met mutant in the α subunit of tryptophan

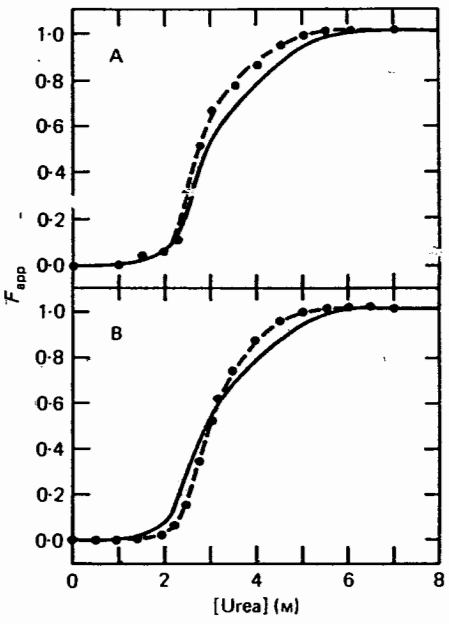

Fig. 6. *Effect on the equilibrium unfolding curve of the wild-type α subunit of tryptophan synthase from* E. coli (——) *by the mutations* (*A*) *Phe22→Leu and* (*B*) *Gly211→Glu*[10] *Data points for the mutants are represented by* (●) *and fits to these points are represented by dashed lines.*

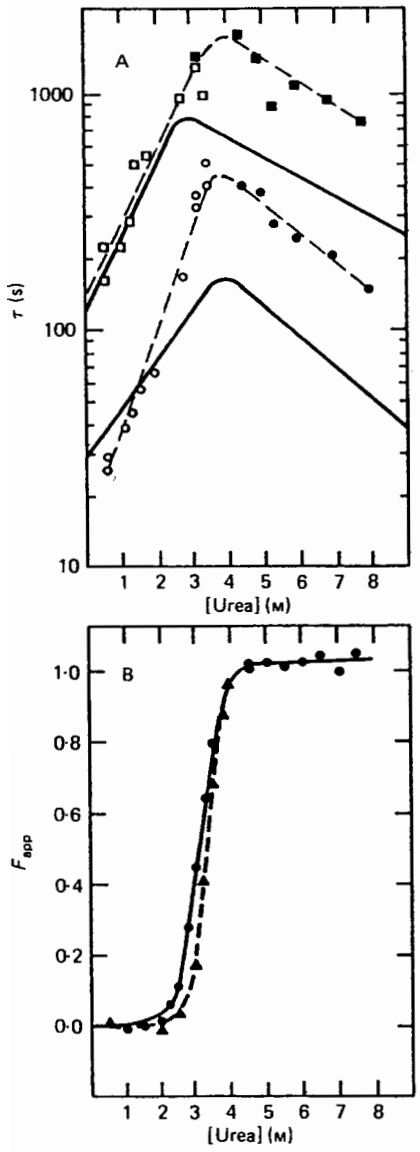

Fig. 7. *(A) Effect on the unfolding/refolding relaxation time of dihydrofolate reductase from* E. coli (——) *due to the mutation Leu28→Arg*[12]. *Unfolding* (●, ■) *and refolding* (○, □) *points for the two relaxation times of the mutant* (----) *are shown. Lines are drawn to aid the eye. (B) Effect on the equilibrium unfolding curve due to the mutation*[12]. *Data points and fitted lines are shown for wild-type* (●, ——) *and Leu28→Arg* (▲, ----) *proteins.*

synthase.[10] In this case, however, it was found that the increase in stability between native and intermediate forms is accounted for by an unchanged relaxation time for unfolding and a decreased relaxation time for refolding. This result implies a selective increase in the energy of the intermediate form with respect to the energies of the native form and the transition state.

Although the kinetic mutants identify the residues that play key roles in the rate-limiting step in folding, the equilibrium mutants can also provide information about mechanisms. Consider again the reaction coordinate diagram in Fig. 2A. Because the additional stabilizing interaction is only realized after the transition state is passed, the amino acid at this position is involved in folding rather late in the process. An alternative method of alignment is to equate the energies of the two native forms. In this case, the destabilizing effect on the unfolded form also exists for the transition state; however, this destabilization is relieved in the native conformation. In both cases the *differential* effect of the mutation is only realized after the transition state is passed in the folding direction.

This rationale was used to explain the results for cross-linked ribonuclease A described above and led to the hypothesis that the amino terminus containing Lys7 associates at a late stage with the remainder of the protein.[13] The effect of the Leu28→Arg mutation in dihydrofolate reductase on the kinetics of unfolding and refolding is similar to that for cross-linked ribonuclease A, suggesting that the residue at position 28 and perhaps the α helix in which it resides is involved in a late stage in folding. The results from the Glu49→Met mutation in the α subunit of tryptophan synthase suggest that position 49 is involved in folding before the transition state is passed.

Future Prospects

The analysis described above leads to the supposition that a systematic study of the effect of mutations at many different sites in a protein on folding and stability will provide a means of mapping the complex conformational change at the level of individual amino acids. Once the roles of individual sites are determined, a series of replacements at each site varying charge, size, hydrophobicity and hydrogen bonding potential will, in principle, permit one to determine the basis for the specific effect. Hopefully, when such results are considered in terms of the X-ray structure, a pattern will emerge that will permit one to develop rules that will apply to the folding of any globular protein.

The mutagenic approach towards determining the mechanism of protein folding is still in its infancy and the basic hypotheses presented above remain to be thoroughly tested. If these hypotheses indeed prove to be valid, our understanding of the factors that direct this essential biological phenomenon should improve dramatically.

This work was funded by the National Institutes of Health, grants GM23303 and 1K04 AG00153, and by the National Science Foundation, grant DMB-8410230.

REFERENCES

1 ANFINSEN, C. B. (1973). Principles that govern the folding of protein chains. *Science* **181**, 223–230.

2 KIM, P. S. & BALDWIN, R. L. (1982). Specific intermediates in the folding reactions of small proteins and the mechanisms of protein folding. *Annu. Rev. Biochem.* **51**, 459–489.

3 KIM, P. S. & BALDWIN, R. L. (1980). Structural intermediates trapped during the folding of ribonuclease A by amide proton exchange. *Biochemistry* **19**, 6124–6129.

4 SCHMID, F. X. & BLASCHEK, H. (1981). A native-like intermediate on the ribonuclease A folding pathway. *Eur. J. Biochem.* **114**, 111–117.

5 SMITH, M. (1985). *In vitro* mutagenesis. *Annu. Rev. Genet.* **19**, 423–462.

6 TANFORD, C., AUNE, K. C. & IKAI, A. (1973). Kinetics of unfolding and refolding of proteins. Results for lysozyme. *J. Mol. Biol.* **73**, 185–197.

7 CRISANTI, M. M. & MATTHEWS, C. R. (1981). Characterization of the slow steps in the folding of the α subunit of tryptophan synthase. *Biochemistry* **20**, 2700–2706.

8 KELLEY, R. F., WILSON, J., BRYANT, C. & STELLWAGEN, E. (1986). Effects of guanidine hydrochloride on the refolding kinetics of denatured thioredoxin. *Biochemistry* **25**, 728–732.

9 HAWKES, R., GRUTTER, M. G. & SCHELLMAN, J. (1984). Thermodynamic stability and point mutations of bacteriophage T4 lysozyme. *J. Mol. Biol.* **175**, 195–212.

10 BEASTY, A. M., HURLE, M. R., MANZ, J. T., STACKHOUSE, T., ONUFFER, J. J. & MATTHEWS, C. R. (1986). The effects of the Phe22→Leu, Glu49→Met, Gly234→Asp and Gly234→Lys mutations on the folding and stability of the α subunit of tryptophan synthase from *Escherichia coli*. *Biochemistry* **25**, 2965–2974.

11 TOUCHETTE, N. A., PERRY, K. M. & MATTHEWS, C. R. (1986). The folding of dihydrofolate reductase from *Escherichia coli*. *Biochemistry* **25**, 5445–5452.

12 MATTHEWS, C. R., PERRY, K. M. & TOUCHETTE, N. A. (1984). The influence of single-site replacements on the folding and stability of dihydrofolate reductase from *Escherichia coli*. *Biochemistry* **23**, 3358.

13 LIN, S. H., KONISHI, Y., NALL, B. T. & SCHERAGA, H. A. (1985). Influence of an extrinsic crosslink on the folding pathway of ribonuclease A. Kinetics of folding and unfolding. *Biochemistry* **24**, 2680–2686.

C. ROBERT MATTHEWS AND MARK R. HURLE *are at the Department of Chemistry, The Pennsylvania State University, University Park, PA 16802, USA.*