SHORT COMMUNICATION

Generation and analysis of proline mutants in protein G

Eun Jung Choi[1] and Stephen L.Mayo[1,2,3]

[1]Division of Biology, California Institute of Technology, 1200 E. California Boulevard, Pasadena, CA and [2]Howard Hughes Medical Institute and Division of Chemistry and Chemical Engineering, California Institute of Technology, 1200 E. California Boulevard, Pasadena, CA 91125, USA

[3]To whom correspondence should be addressed.
E-mail: steve@mayo.caltech.edu

The pyrrolidine ring of the amino acid proline reduces the conformational freedom of the protein backbone in its unfolded form and thus enhances protein stability. The strategy of inserting proline into regions of the protein where it does not perturb the structure has been utilized to stabilize many different proteins including enzymes. However, most of these efforts have been based on trial and error, rather than rational design. Here, we try to understand proline's effect on protein stability by introducing proline mutations into various regions of the B1 domain of Streptococcal protein G. We also applied the Optimization of Rotamers By Iterative Techniques computational protein design program, using two different solvation models, to determine the extent to which it could predict the stabilizing and destabilizing effects of prolines. Use of a surface area dependent solvation model resulted in a modest correlation between the experimental free energy of folding and computed energies; on the other hand, use of a Gaussian solvent exclusion model led to significant positive correlation. Including a backbone conformational entropy term to the computational energies increases the statistical significance of the correlation between the experimental stabilities and both solvation models.
Keywords: proline/protein/design/protein G/protein stability

Introduction

Proline is the only naturally occurring amino acid in which the side chain is bonded to the backbone nitrogen, forming a five-membered pyrrolidine ring. This pyrrolidine ring restricts the rotation of the N–C_α bond, decreasing the backbone conformational entropy of the unfolded form of the protein relative to other naturally occurring amino acids. This allows proline substitution to increase the stability of a protein by decreasing the entropic difference between the unfolded and the folded form, thereby increasing the free energy difference (Nemethy *et al.*, 1966; Matthews *et al.*, 1987). Based on this concept, different residues in various proteins have been mutated to prolines, resulting in increased stability (Matthews *et al.*, 1987; Watanabe *et al.*, 1994).

On the other hand, prolines are also notorious for destabilizing proteins. It is well known that prolines, located internally in α-helices or β-sheets, break the secondary structures, thus destabilizing the protein. Two main factors cause prolines to break secondary structures. One is the absence of hydrogen on the amide nitrogen, which prohibits prolines from acting as a donor in a hydrogen bond. Another is the steric constraint placed on proline and the neighboring residues by the pyrrolidine ring, hindering secondary structure formation. The phi and psi angles preferred by prolines are far from the typical range of those for β-sheets and thus distort the strand significantly. In addition, steric restriction drives the residue preceding proline to prefer the beta conformation, thus limiting the occurrence of prolines in α-helices.

Because of the contradictory effects described above, stabilization of proteins by proline incorporation has typically been achieved by trial and error. The general understanding is that proline mutation is stabilizing because of entropic factors but this behavior can be masked and even reversed by destabilizing enthalpic changes. Thus, protein stabilization by prolines has been achieved by placing prolines in relatively solvent-exposed locations, where they would not disturb the stabilizing interactions of the protein, for example in loops and turns or the first turn of an α-helix (Watanabe and Suzuki, 1998). In this article we examine the effects of proline mutations in protein G and attempt to 'predict' these effects using computational tools with the aim of reasonably incorporating proline in computational protein design efforts.

Materials and methods

Mutagenesis and protein purification

Mutants of the B1 domain of protein G (Gβ1) were constructed using inverse PCR in plasmid pET11A, expressed using BL21 (DE3) cells and purified as described previously (Malakauskas and Mayo, 1998). Two forms, 56- and 57-residue species, of Gβ1 resulted due to incomplete processing of the N-terminal methionine. In this study the 56-residue species of Gβ1 mutants and the wild type were used. Molecular weights were verified using mass spectrometry.

CD analysis

CD data were collected on an Aviv 62DS spectrometer equipped with a thermoelectric unit. Thermal denaturation experiments were monitored at 218 nm from 1 to 99°C by 1°C increments with an equilibration time of 1.5 min using 50 μM of protein in 50 mM sodium phosphate at pH 5.5. The midpoint of the thermal unfolding transition (T_m) was determined from a two-state analysis of each denaturation curve (Minor and Kim, 1994). Guanidinium chloride denaturations were performed at 25°C using 5 μM of protein

in 50 mM sodium phosphate at pH 5.5. Data were collected for 5 min and averaged. Free energies of folding (ΔG_f) and error estimates were obtained by fitting the denaturation data to a two-state transition model (Santoro and Bolen, 1988) using Kaleidagraph (Synergy Software). Chemical and thermal melting curves for protein G and its variants are presented in the Supplemental data available at *PEDS* online.

Computational analysis

The crystal structure of wild-type Gβ1 (PDB ID: 1pga) was used as the starting template for energy calculations. Explicit hydrogens were added using MolProbity (Lovell *et al.*, 2003) and the structure was energy minimized for 50 steps to remove any steric clashes (Mayo *et al.*, 1990). For each mutant, proline was substituted at the selected position, and the protein design program, ORBIT (Dahiyat and Mayo, 1997a, b; Dahiyat *et al.*, 1997; Street and Mayo, 1998; Pierce *et al.*, 2000) was used to optimize the structure (selecting the optimal rotamer for proline as well as for all the other residues in the protein) and to calculate energies. Solvation energies were calculated using the method of either Street and Mayo (1998) or Lazaridis and Karplus (1999).

Results and discussion

Proline mutants of Gβ1

Prolines tend to have a phi angle of $\sim-63°$, while the psi angle clusters around two regions in the Ramachandran map, $-35°$ (α region) and $150°$ (β region) (MacArthur and Thornton, 1991). We selected 10 Gβ1 residues having phi and psi angles compatible with proline for mutation: Thr2, Gly9, Lys10, Val21, Ala23, Ala24, Thr25, Val29, Asp36 and Ala48. In order to explore the effect of prolines in different structural environments, these residues were selected from various regions of the protein (Figure 1). Thr2 and Gly9 are located in a β-strand. Ala23, Ala24 and Thr25 are the first three N-terminal residues on the α-helix, Val29 is in the middle of the helix and Asp36 is the C-terminal residue of the helix. The remaining residues are located in the loops and turns connecting the secondary structural elements. The phi and psi angles of the preceding residue of Thr2, Gly9, Lys10, Val21 and Ala23 are in the β region of the Ramachandran

map. Residues preceding proline prefer the β region because their C_β and amide nitrogen sterically conflict with the C_δ of proline (Schimmel and Flory, 1968; Matthews *et al.*, 1987; Hurley *et al.*, 1992).

Stability studies and analysis

The stability of each of the mutants was determined by performing thermal and chemical denaturation experiments (Table I and Supplementary data available at *PEDS* online). The far UV CD spectra before and after thermal denaturation indicate that all mutants except for K10P, T25P and V29P fold reversibly (data not shown). The post-transition region of the melting curves for V21P and V21P/A23P extends beyond the experimental range of 99°C, which leads to large estimated errors.

As expected from the fact that proline is a well-known secondary structure breaker, most of the proline mutants were less stable than the wild-type protein. An exception to this was V21P, which exhibited a ΔG_f enhancement of 0.5 kcal/mol and a T_m increase of 6°C compared with the wild-type protein. This value agrees well with the expected energy of stabilization generated by the entropic difference between Val and Pro in the unfolded state. According to the method of Nemethy *et al.* (1966) a Val to Pro mutation should increase stability by 0.5 kcal/mol, while the method of Stites and Pranata (1995) suggests an increase of 0.3 kcal/mol at 25°C. Comparisons between mesophilic and thermophilic proteins and mutational studies indicate that proline residues located in loops help to increase the rigidity of the loop, thus increasing the stability of the protein (Vieille and Zeikus, 1996). Val21 is the first residue in a two-residue loop, which connects one of the edge strands of the β-sheet and the N-terminus of the α-helix. It is solvent-exposed and does not interact with other residues; thus, mutation to proline does not disturb any energetically favorable interaction. Given these observations, it is not surprising that the increase in stability for V21P is close to the expected value.

Two residues, Lys10 and Ala48 are located in the $i + 1$ position of a β-turn. The fact that proline is the most favored residue for the $i + 1$ position of β-turns has been rationalized by the analysis of protein structures. These studies reveal that prolines have phi angles that are favored in that position
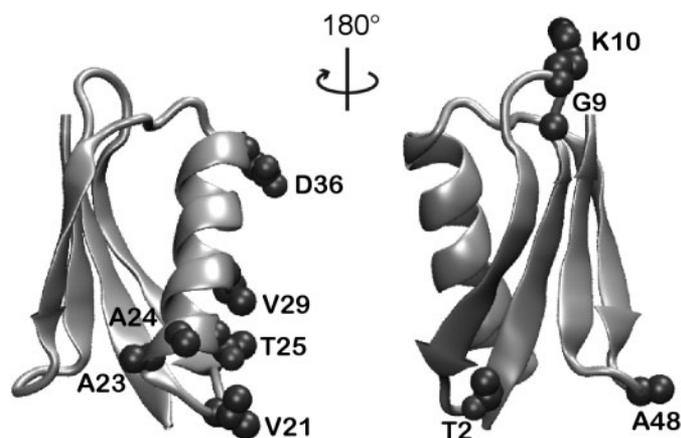

Fig. 1. Views of the 10 positions in Gβ1 that were mutated to proline. These structural figures were generated using VMD (Humphrey *et al.*, 1996).

Table I. Midpoint of thermal unfolding transition (T_m), free energy of folding (ΔG_f) at 25°C and computed energy for Gβ1 variants (errors determined from non-linear fits)

Name	T_m (°C)	ΔG_f (kcal/mol)	$\Delta\Delta G_f$ (kcal/mol)	E_{calc} (SM)[a] (kcal/mol)	E_{calc} (LK)[b] (kcal/mol)
Wild type	89.6 ± 2.6	−5.9 ± 0.4	–	−90.1	−72.2
T2P	83.0 ± 1.2	−3.2 ± 0.3	2.7	−90.5	−72.9
G9P	72.6 ± 0.7	−3.5 ± 0.3	2.4	120.2	146.9
K10P	81.2 ± 1.1	−5.7 ± 0.4	0.2	−89.0	−71.9
V21P	95.8 ± 14.3	−6.4 ± 0.4	−0.5	−90.5	−73.2
A23P	88.2 ± 1.9	−5.6 ± 0.5	0.3	−93.5	−72.7
A24P	85.0 ± 0.8	−5.4 ± 0.3	0.5	−90.1	−71.2
T25P	68.7 ± 0.4	−3.1 ± 0.2	2.8	−77.5	−60.4
V29P	67.0 ± 0.5	−2.4 ± 0.2	3.5	1464.7	1487.1
D36P	68.6 ± 0.8	−2.8 ± 0.3	3.1	3310.3	3330.0
A48P	82.8 ± 0.7	−5.2 ± 0.3	0.7	−86.4	−70.8
V21P/A23P	96.5 ± 28.4	−5.8 ± 0.4	0.1	−93.9	−73.8

[a]ORBIT energy using the Street and Mayo (1998) solvation model.
[b]ORBIT energy using the Lazaridis and Karplus (1999) solvation model.

(Hutchinson and Thornton, 1994). Also, mutational studies report an increase in thermostability with proline substitution in this position (Watanabe and Suzuki, 1998). In our case, these two proline mutants were slightly destabilizing compared with the wild-type protein. This may be partly due to the loss of a hydrogen bond between the amide nitrogen of Ala48 and the carboxylate of Asp46, and for Lys10 the loss of electrostatic interactions with negatively charged residues in the vicinity (Asp40, Glu56 and the C-terminal carboxyl group).

Prolines occur in the first turn of α-helices with high frequency. It has been suggested that proline residues are not destabilizing in this position because the amide hydrogens in the first turn do not make backbone/backbone hydrogen bonds within the helix (von Heijne, 1991). Within the first turn, prolines exist predominantly at the N1 position because steric clashes would result if the preceding residue is in a helical conformation (Yun et al., 1991; Cochran et al., 2001). In our experiments, the proline mutation at the N2 position (A24P) was more destabilizing than the one at the N1 position (A23P), consistent with previous observations. However, both proline mutations were slightly destabilizing. Contrary to some mutational studies, but consistent with our results for A23P and A24P as well as for K10P and A48P, peptide helicity measurements demonstrated that prolines do not stabilize N terminal residues of α-helices or the $i + 1$ residue of β-turns. Instead, it was suggested that prolines occur frequently in certain locations because they are tolerated and not because they are stabilizing (Cochran et al., 2001). Proline substitution at Thr25, which is located in the N3 position, was highly destabilizing with $\Delta\Delta G_f$ of 2.8 kcal/mol and a T_m decrease of 21°C compared with the wild-type protein. The amide nitrogen of the residues in the first turn of an α-helix frequently forms hydrogen bonds with a nearby side chain, such as the N-cap residue, in order to satisfy its hydrogen bond donor potential. This is particularly true for the amide nitrogen of an N3 residue, which has frequently been observed to form a hydrogen bond with the N-cap side chain (Penel et al., 1999). In the wild-type Gβ1 structure, the backbone and side chain of Thr25 make extensive hydrogen bonds with the carboxylate group of the N-cap residue, Asp22. This interaction stabilizes and caps the first turn of the helix (Gronenborn et al., 1991). Substituting proline at this position eliminates this stabilizing interaction, thus destabilizing the protein.

It has been reported that the cost of introducing a proline into an α-helix is ~3.4 kcal/mol (Oneil and Degrado, 1990; Yun et al., 1991). Consistent with this, our Val29 to proline mutation destabilized the protein by 3.5 kcal/mol and decreased the T_m by 23°C.

Proline substitution in the β-sheet at Thr2 and Gly9 destabilizes the protein by distorting the secondary structure of the β-sheet. The far UV CD spectra of these two mutants deviate significantly from the wild-type spectrum, suggesting a change in secondary structure content (data not shown).

Introduction of increasing numbers of prolines, up to nine, additively increased the stability of oligo-1,6-glucosidase (Watanabe et al., 1994). To explore whether the stabilizing/destabilizing effects of our proline mutants were additive, we constructed a double mutant, V21P/A23P. To a first approximation, the effect was additive, resulting in a near zero effect on stability, as can be seen in Table I.

Comparison of computational and experimental energies

The energy of each of the mutants and the wild-type protein was calculated by substituting proline at the respective positions and optimizing all side chains using Optimization of Rotamers By Iterative Techniques (ORBIT). We used two different methods of calculating solvation energy, a surface area dependent solvation model from Street and Mayo (1998) and a Gaussian solvent exclusion model from Lazaridis and Karplus (1999). The energies are reported in Table I. Mutants T2P, G9P, V29P and D36P showed significant backbone movement in the mutated region after minimization of the optimized structure determined by ORBIT and/or exhibited large deviations of their far UV CD spectra compared with the wild-type protein. Since ORBIT utilizes the static backbone of the wild-type crystal structure for its calculations, the ORBIT energy of these mutants is not likely to reflect the energy of the true structure. Thus, they were not considered in the correlation analysis between the calculated ORBIT energy (E_{calc}) and ΔG_f obtained using experiment. However, we would like to point out that ORBIT does predict the destabilizing effects of three (G9P, V29P, D36P) of the four mutants excluded. They all have very high computed energies, due to large van der Waals clashes between side chain and the backbone. We propose that this is the reason for the destabilization of these mutants and deviation of their CD spectra.

Excluding the mutants mentioned above, the agreement between the ORBIT energy difference between the mutant (P) and the wild type (WT) [$\Delta E_{calc}(P - WT)$] and the experimentally determined free energy difference [$\Delta\Delta G_f(P - WT)$] resulted in an R^2 value of 0.79 using the Street and Mayo (SM) solvation model (Figure 2A) and of 0.94 using the Lazaridis and Karplus (LK) solvation model (Figure 3A). Correlation between T_m and ORBIT energy resulted in R^2 values of 0.78 and 0.79 for the SM and LK models, respectively (Figures 2B and 3B). In order to estimate the prediction error of ORBIT energies more accurately and to consider whether the correlation for the dataset is dominated by the result for T25P, we used the 'leave-one-out' cross-validation method on the free energy dataset. The cross-validation estimate of prediction error was 14.5 for the SM method while the LK method gave a significantly lower value of 0.12. Thus the Lazaridis and Karplus excluded volume solvation model based ORBIT energies show greater correlation to experimental ΔG_u and a lower estimate of prediction error than the surface area dependent solvation model based energies, suggesting that the LK model performs better in describing the proline mutants. Overall, although some false positives occur, ORBIT is reasonably predictive in ranking the stabilities of the various mutants as indicated by Spearman's rank correlation coefficients of 0.75 ($P < 0.01$) and 0.93 ($P < 0.01$) for the SM and LK free energy correlations, respectively (data not shown).

In our study, prolines have typically not been included in the set of amino acids that ORBIT considers in protein design calculations because the potential energy function used in ORBIT does not include a conformational entropy term. We tested whether including a backbone conformational entropy term to the computational energy increases the rank correlation between the experimental stabilities and computational energies. Using the backbone entropy scale from Stites and Pranata (1995), the weighting factor for the
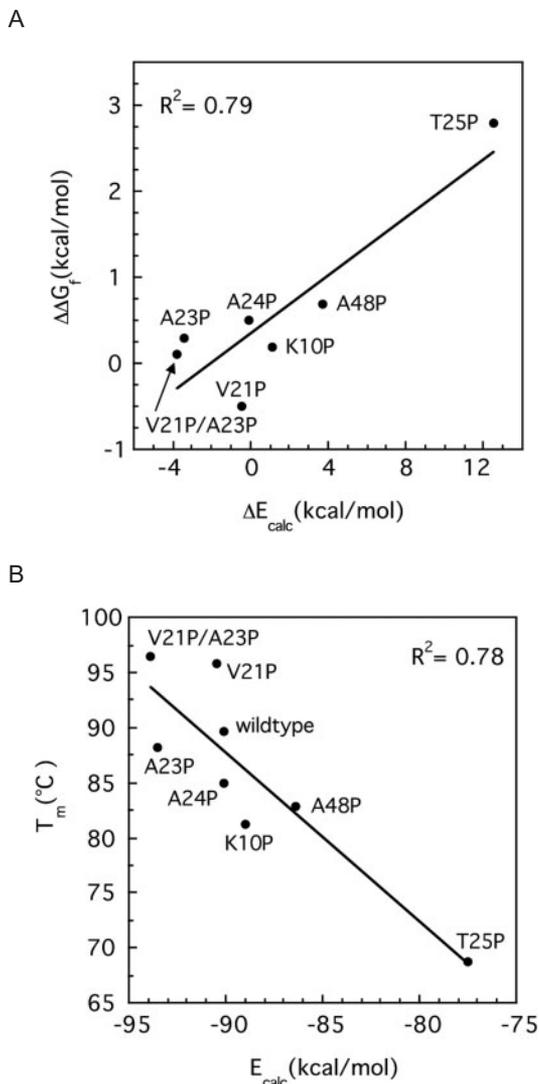
E.J.Choi and S.L.Mayo

A

B

Fig. 2. Correlation between calculated energies, using the Street and Mayo (1998) solvation model, and experimental results. (**A**) Comparison of experimental ($\Delta\Delta G_f$) and calculated (E_{calc}) stability change between the mutant and the wild type. (**B**) Comparison of T_m and calculated energy.

A

B

Fig. 3. Correlation between calculated energies, using the Lazaridis and Karplus (1999) solvation model, and experimental results. (**A**) Comparison of experimental ($\Delta\Delta G_f$) and calculated (E_{calc}) stability change between the mutant and the wild type. (**B**) Comparison of T_m and calculated energy.

entropy term was determined by optimizing the rank correlation between the experimental and computational results. For computational energies calculated with the LK model, a weighting factor in the range of 2.9–12.9 gave a rank correlation of 0.96 ($P < 0.01$). For those calculated with the surface area based method, a weighting factor in the range of 15.01–15.2 also gave a rank correlation of 0.96 ($P < 0.01$).

As described above, proline stabilizes a protein by decreasing the backbone conformational entropy of the unfolded state. ORBIT predicts the stability of proline mutations reasonably well without an entropic term. This is likely due to the dominance of enthalpic contributions over entropic contributions in protein stability modulation by prolines, which overshadows the missing entropic term in the energy function. Nevertheless, addition of an entropic term with appropriate weighting factor increases the correlation between computational energy and experimental energy, especially for the surface area based solvation method.
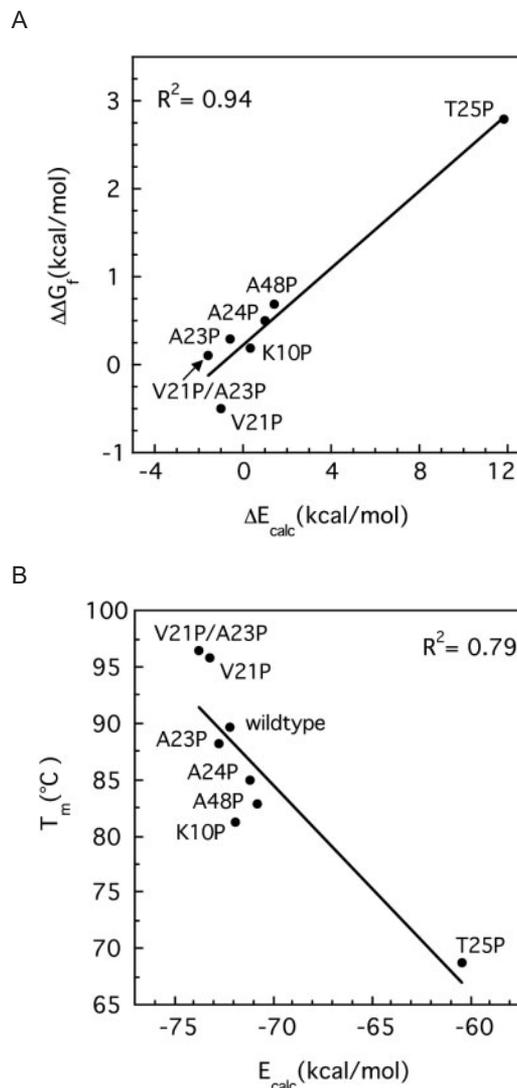
Acknowledgements

References

Cochran,D.A., Penel,S. and Doig,A.J. (2001) *Protein Sci.*, **10**, 463–470.
Dahiyat,B.I. and Mayo,S.L. (1996) *Protein Sci.*, **5**, 895–903.
Dahiyat,B.I. and Mayo,S.L. (1997a) *Science*, **278**, 82–87.
Dahiyat,B.I. and Mayo,S.L. (1997b) *Proc. Natl Acad. Sci. USA*, **94**, 10172–10177.
Dahiyat,B.I., Gordon,D.B. and Mayo,S.L. (1997) *Protein Sci.*, **6**, 1333–1337.
Gronenborn,A.M., Filpula,D.R., Essig,N.Z., Achari,A., Whitlow,M., Wingfield,P.T. and Clore,G.M. (1991) *Science*, **253**, 657–661.
Humphrey,W., Dalke,A. and Schulten,K. (1996) *J. Mol. Graph.*, **14**, 33–38.
Hurley,J.H., Mason,D.A. and Matthews,B.W. (1992) *Biopolymers*, **32**, 1443–1446.
Hutchinson,E.G. and Thornton,J.M. (1994) *Protein Sci.*, **3**, 2207–2216.
Lazaridis,T. and Karplus,M. (1999) *Proteins*, **35**, 133–152.

Lovell,S.C., Davis,I.W., Arendall,W.B., 3rd, de Bakker,P.I., Word,J.M., Prisant,M.G., Richardson,J.S. and Richardson,D.C. (2003) *Proteins*, **50**, 437–450.

MacArthur,M.W. and Thornton,J.M. (1991) *J. Mol. Biol.*, **218**, 397–412.

Malakauskas,S.M. and Mayo,S.L. (1998) *Nat. Struct. Biol.*, **5**, 470–475.

Matthews,B.W., Nicholson,H. and Becktel,W.J. (1987) *Proc. Natl Acad. Sci. USA*, **84**, 6663–6667.

Mayo,S., Olafson,B. and Goddard,W. (1990a) *J. Phys. Chem.*, **94**, 8897–8909.

Minor,D.L., Jr and Kim,P.S. (1994) *Nature*, **367**, 660–663.

Nemethy,G., Leach,S. and Scheraga,H. (1966) *J. Phys. Chem.*, **70**, 998–1004.

Oneil,K.T. and Degrado,W.F. (1990) *Science*, **250**, 646–651.

Penel,S., Hughes,E. and Doig,A.J. (1999) *J. Mol. Biol.*, **287**, 127–143.

Pierce,N.A., Spriet,J.A., Desmet,J. and Mayo,S.L. (2000) *J. Comput. Chem.*, **21**, 999–1009.

Santoro,M. and Bolen,D. (1988) *Biochemistry*, **27**, 8063–8068.

Schimmel,P.R. and Flory,P.J. (1968) *J. Mol. Biol.*, **34**, 105–120.

Stites,W.E. and Pranata,J. (1995) *Proteins*, **22**, 132–140.

Street,A.G. and Mayo,S.L. (1998) *Fold Des.*, **3**, 253–258.

Vieille,C. and Zeikus,J.G. (1996) *Trends Biotechnol.*, **14**, 183–190.

von Heijne,G. (1991) *J. Mol. Biol.*, **218**, 499–503.

Watanabe,K. and Suzuki,Y. (1998) *J. Mol. Catal. B Enzym.*, **4**, 167–180.

Watanabe,K., Masuda,T., Ohashi,H., Mihara,H. and Suzuki,Y. (1994) *Eur. J. Biochem.*, **226**, 277–283.

Yun,R.H., Anderson,A. and Hermans,J. (1991) *Proteins*, **10**, 219–228.