doi:10.1016/j.jmb.2004.09.047 J. Mol. Biol. (2004) 344, 223–237

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Do Proteins Always Benefit from a Stability Increase? Relevant and Residual Stabilisation in a Three-state Protein by Charge Optimisation

Luis A. Campos[1,2], Maria M. Garcia-Mira[1,3], Raquel Godoy-Ruiz[1,3] Jose M. Sanchez-Ruiz[1,3]* and Javier Sancho[1,2]*

[1]*Biocomputation and Complex Systems Physics Institute University of Zaragoza, 50009 Zaragoza, Spain*

[2]*Departamento de Bioquímica y Biología Molecular y Celular. Fac. Ciencias, University of Zaragoza, 50009 Zaragoza Spain*

[3]*Departamento de Química Física. Fac. Ciencias, University of Granada, 18071 Granada Spain*

**Corresponding authors*

The vast majority of our knowledge on protein stability arises from the study of simple two-state models. However, proteins displaying equilibrium intermediates under certain conditions abound and it is unclear whether the energetics of native/intermediate equilibria is well represented in current knowledge. We consider here that the overall conformational stability of three-state proteins is made of a "relevant" term and a "residual" one, corresponding to the free energy differences of the native to intermediate (N-to-I) and intermediate to denatured (I-to-D) equilibria, respectively. The N-to-I free energy difference is considered to be the relevant stability because protein-unfolding intermediates are likely devoid of biological activity. We use surface charge optimisation to first increase the overall (N-to-D) stability of a model three-state protein (apoflavodoxin) and then investigate whether the stabilisation obtained is realised into relevant or into residual stability. Most of the mutations designed from electrostatic calculations or from simple sequence conservation analysis produce large increases in the overall stability of the protein. However, in most cases, this simply leads to similarly large increases of the residual stability. Two mutations, nevertheless, show a different trend and increase the relevant stability of the protein substantially. When all the mutations are mapped onto the structure of the apoflavodoxin thermal-unfolding intermediate (obtained independently by equilibrium ϕ-analysis and NMR) they cluster perfectly so that the mutations increasing the relevant stability appear in the small unstructured region of the intermediate and the others in the native-like region. This illustrates the need for specific investigation of N-to-I equilibria and the structure of protein intermediates, and indicates that it is possible to rationally stabilise a protein against partial unfolding once the structure of the intermediate conformation is known, even if at low resolution.

Keywords: charge/charge interaction; protein stability; protein stabilisation; protein folding; equilibrium intermediate

Introduction

Over the last 15 years important advances have been made in the understanding of protein stability by combining mutational and thermodynamic studies of the unfolding equilibrium. To simplify the analysis of the role played in stability by specific interactions, the study of small model proteins displaying two-state equilibrium behaviour[1–5] has been largely favoured over that of larger, multi-domain or multi-state proteins.[6–10] The investigation of protein stability principles has concomitantly led to the discovery of methods of protein stabilisation, a most important practical goal with application in fields such as the industrial use of enzymes[11–14] and the use of antibodies for

Abbreviations used: N-to-I, native to intermediate; FMN, flavine mononucleotide; DSC, differential scanning calorimetry; T_m, melting temperature; WT, wild-type.

E-mail addresses of the corresponding authors: jsancho@unizar.es; sanchezr@ugr.es

diagnostics or treatment.[15–17] The issue even shows a promising ramification to the emerging field of molecular remediation of folding diseases.[18] Among the methods proposed and tested in order to rationally stabilise proteins, some exploit entropic differences between folded and unfolded polypeptides to achieve an overall stabilisation through the engineering of disulfide bonds[19–21] or the removal of glycine residues or introduction of proline residues into the sequence.[1,22] Other methods focus on the fairly well understood energetics of α-helices,[6,23–27] or in the engineering of specific interactions, such as the filling of small protein cavities[28–30] or the discharge of hydrogen bonds.[31] Others, yet, are not based on physical insight but rather on sequence comparison of mesophilic and thermophilic homologous proteins.[20,32–37] The story of the application of many of these methods shows that, despite being extremely valuable, they may often fail due to the intrinsic complexity of protein interactions. The search for generally reliable methods may thus be considered a key goal of the field.

This wealth of knowledge on the response of two-state proteins to point mutations aimed at increasing the stability of the native state is in sharp contrast with the lack of guidance on the behaviour of proteins displaying more complex unfolding equilibria, which may not be in minority within the proteome. If we take the simplest of such proteins, one with one equilibrium intermediate, say in the thermal unfolding, two distinct equilibria have to be considered: the one linking the well-folded, native conformation with the less orderly intermediate, and another relating the latter to the unfolded state. If, as it is to be expected, native behaviour is confined to the well-folded state, the free energy difference of the first equilibrium is the one that really counts for the maintenance of function and can thus be termed the "relevant" conformational stability, as opposed to the "residual" stability residing in the second equilibrium.[38] The intriguing question, for which no answer has so far been offered, is: will the strategies devised over the years to stabilise two-state proteins be applicable to increasing the relevant stability of three-state or more complex proteins?

We seek here an answer to this question by using one of the most promisingly general methods of protein stabilisation: the optimisation of surface charge–charge interactions, which has a reputed story of success.[39–43] The method relies in identifying, by theoretical methods of varying complexity, mutations of a protein that are expected to lower the electrostatic energy of the native state.[41,44–46] Our procedure, one of the simplest, uses the Tanford–Kirkwood model[47] to estimate interaction energies between charges, and has been applied successfully to ubiquitin.[39,46] In this work, we design and test charge mutations in the apoflavodoxin from *Anabaena* PCC7119 (Figure 1),[48] a model protein with the interesting characteristic of behaving as a two-state system towards urea-denaturation[49] but as a three-state system towards thermal unfolding.[50] Our analysis of the two-state equilibrium shows that large overall protein stabilisation is achieved easily by the mutations so simply designed: the native to denatured (N-to-D) energy gap is largely increased. The successful mutations are then used to investigate how the stabilisation is partitioned into its relevant and residual components. The analysis of the thermal unfolding clearly indicates that the relevant stability of the protein remains nearly unchanged in most cases. The few mutations that lead to substantial increases of the relevant stability are explained by the structure of the thermal intermediate, as obtained from equilibrium φ-analysis.[90] The principles demonstrated here are applicable to the analysis of other stabilising strategies and illustrate how low-resolution structures of protein intermediates can guide the rational stabilisation of three-state proteins against the occurrence of partly unfolded conformations.

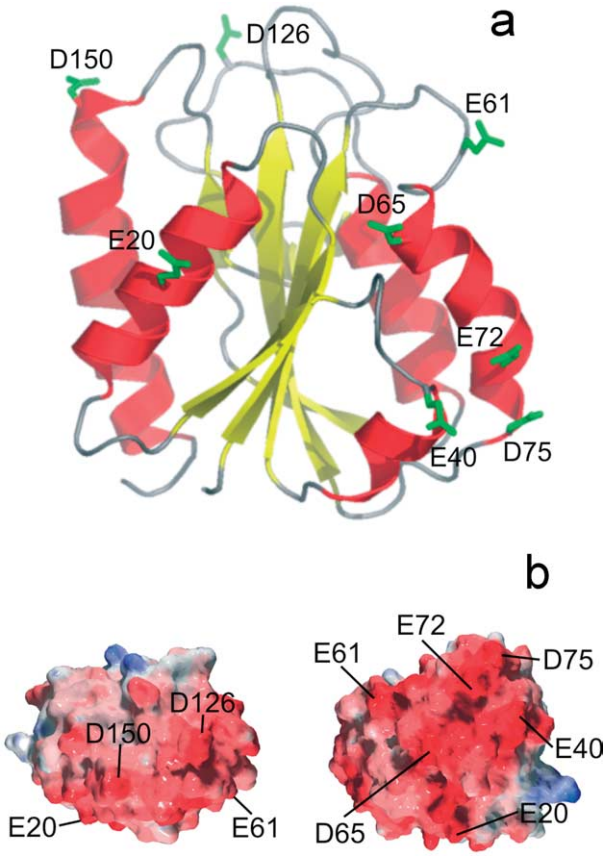

Figure 1. (a) Scheme of the apoflavodoxin structure, 1ftg. The side-chains of the acidic residues mutated to lysine are shown in green. (b) Two views of the electrostatic potential surface (calculated with Weblab) of apoflavodoxin (negatively and positively charged areas are in red and blue, respectively). The location of the mutated residues is indicated.

Results

Structural integrity of mutants and reversibility of the thermal unfolding reactions

All the mutant proteins were purified in their functional holoflavodoxin form, and the flavine mononucleotide (FMN) prosthetic group was subsequently removed as explained. To verify the structural integrity of the mutants we have compared their near-UV CD spectra with that of wild-type (Figure 2(a)). Five of the mutants display near-UV CD spectra identical with that of the wild-type, while the E61K, D65K and D150K mutants show moderate decreases in intensity that are, in principle, compatible with slight modifications in the environment of aromatic residues (we note in this respect that positions 61 and 65 are close to W57 and W66, respectively). The far-UV CD spectra of six of the mutants (Figure 2(b)) are almost identical with that of wild-type and only D65K and D150K show small increments in the signal around 208 nm. This is not related to a partial unfolding because the signal at 222 nm remains unchanged for the two mutants. The spectral differences in D65K and D150K probably reflect a change in the contribution of their aromatic residues to the far-UV CD spectrum.[49]

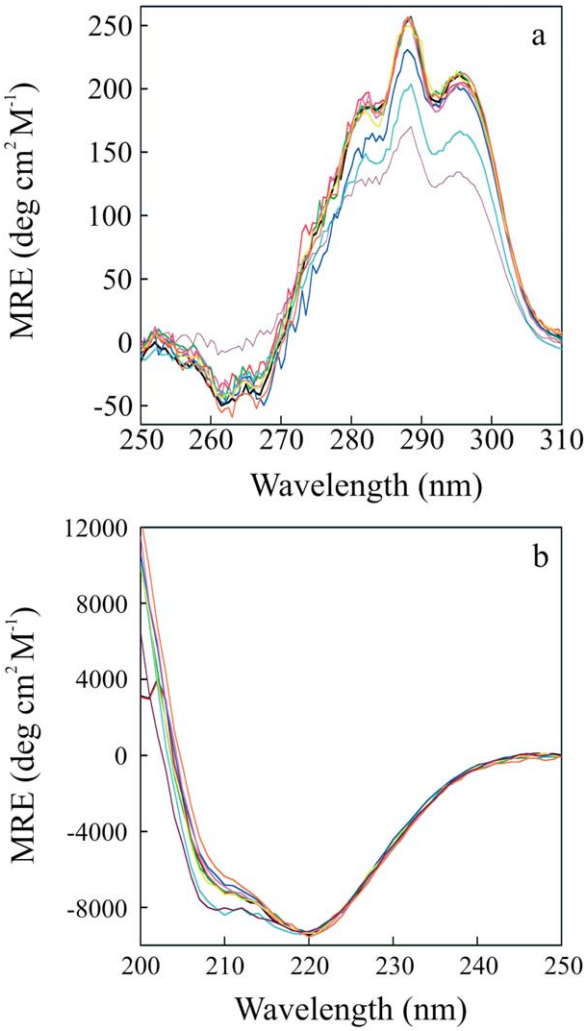

Figure 2. Near-UV (a) and far-UV (b) circular dichroism spectra of wild-type (black), E20K (red), E40K (green), E61K (blue), D65K (cyan), E72K (magenta), D75K (yellow), D126K (orange) and D150K (purple) recorded at 25.0(±0.1) °C, in 50 mM Mops (pH 7.0).

The reversibility of apoflavodoxin thermal unfolding was reported in a previous work for the wild-type protein and eight additional variants.[50] In differential scanning calorimetry (DSC) experiments, 90% of the unfolded wild-type protein was successfully renatured upon cooling. The degree of reversibility observed was even higher when the thermal denaturation was followed spectroscopically at lower concentrations of protein and the unfolding was stopped immediately after the transition. For the mutants described in this work, we have typically recovered after cooling between 85 and 97% of the native near-UV ellipticity (which is lost completely upon heating in all cases). Their thermal unfolding is thus also highly reversible. In addition, we have determined that the transition temperatures are independent of the scan rate (from 0.5 to 1.5 deg.C/minute) with the exception of only the D75K mutant. The data reported for D75K are thus less accurate that that of wild-type and all other mutants.

Global stabilisation of apoflavodoxin by charge reversal mutations

The urea-denaturation curves of the wild-type and the eight apoflavodoxin mutants analysed are shown in Figure 3, and the data derived from their fits to equation (1) are in Table 1. All the mutants display higher $U_{1/2}$ values than the wild-type protein. As it is customary, the stability of each

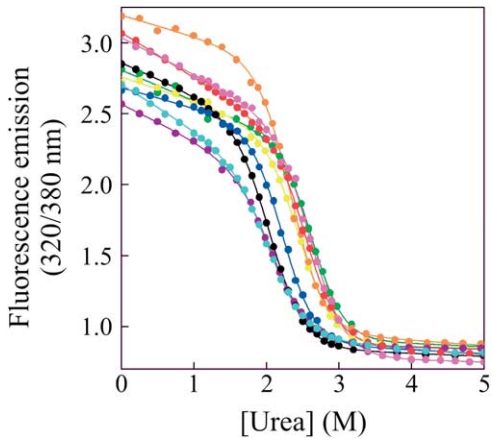

Figure 3. Urea-denaturation curves of wild-type and mutant proteins. The fitting to a two-state equation (continuous lines) and the experimental data of wild-type (black), E20K (red), E40K (green), E61K (blue), D65K (cyan), E72K (magenta), D75K (yellow), D126K (orange) and D150K (purple) are represented. The unfolding at 25.0(±0.1) °C, in 50 mM Mops (pH 7.0) was followed by fluorescence emission.

Table 1. Conformational stabilisation of charge reversal mutant apoflavodoxins relative to wild-type, as determined from two-state (N↔D) urea denaturation: comparison with predicted stabilisations

Protein	Location	U_m[a] (M)	m[a] (kcal mol^{-1} M^{-1})	$\Delta\Delta G_{exp(ND)}$[b] (kcal mol^{-1})	$\Delta\Delta G_{exp(ND)}$[c] (kcal mol^{-1})	$\Delta\Delta G_{pred(ND)}$[d] (kcal mol^{-1})
Wt		2.07±0.01	2.44±0.03	–	–	
E20K	Helix (center)	2.56±0.01	2.58±0.07	−1.60±0.19	−1.27±0.16	−1.39
E40K	Loop	2.66±0.02	2.73±0.17	−2.21±0.45	−1.47±0.17	−1.29
E61K	Loop	2.24±0.01	2.66±0.09	−0.88±0.21	−0.40±0.15	−1.53
D65K	Helix (N1)	2.09±0.01	2.29±0.07	+0.26±0.16	−0.05±0.15	−2.77
E72K	Helix (C-cap)	2.64±0.02	2.40±0.13	−1.28±0.36	−1.40±0.17	−1.72
D75K	Loop	2.47±0.01	2.45±0.09	−1.00±0.23	−1.00±0.16	−1.03
D126K	Loop	2.35±0.01	2.58±0.10	−1.01±0.25	−0.70±0.16	−1.00
D150K	Helix (N′)	2.12±0.01	2.25±0.07	+0.28±0.16	−0.13±0.15	−1.03

[a] The standard deviations have been calculated by interval analysis[86] and propagated to calculate the reported SD of $\Delta\Delta G$.
[b] Experimentally determined $\Delta\Delta G_{ND}$ values calculated using the fitted m values and urea concentration of mid-denaturation of each protein ($\Delta\Delta G = m^{WT}U_m^{WT} - m^{mut}U_m^{mut}$). $\Delta\Delta G$ corresponds to the unfolding free energy difference of the wild-type protein minus that of the mutants. Negative values indicate a greater stability of the mutant.
[c] Experimentally determined $\Delta\Delta G_{ND}$ values calculated using an average value of m=2.49 kcal mol^{-1} M^{-1} ($\Delta\Delta G = m^{av}\Delta U_m^{WT-mut}$). $\Delta\Delta G$ corresponds to the unfolding free energy difference of the wild-type protein minus that of the mutants. Negative values indicate a greater stability of the mutant.
[d] Predicted stabilisation calculated with the Tanford–Kirkwood model.[39–47]

protein and, hence, the change in stability conferred by each mutation can be calculated in two ways: either using the fitted m value for each protein, or using an averaged m value that minimises the extrapolation errors. The $\Delta\Delta G$ values calculated with either approach are in qualitative agreement and are reported in Table 1. Since the m values obtained for each of the proteins (± their corresponding standard deviations (SD)) lay in all cases within the SD interval around the mean m value of all the mutants, we consider the stability differences calculated using the average m value to be more accurate and we will therefore discuss them.

The effectiveness of the method used to select the mutations is illustrated by the fact that six out of the eight mutations predicted to stabilise apoflavodoxin clearly succeed and increase the overall conformational stability of the protein by 0.4–1.5 kcal mol^{-1} (1 cal = 4.184 J). Two mutations (D65K and D150K), however, contradicting prediction, leave the stability barely changed. Perhaps significantly, these two mutations occur near the N terminus of α-helices, as it will be discussed below. In order to further evaluate the reliability of the design method, we compare in Figure 4 the calculated stability changes predicted for the eight flavodoxins mutants analysed here and for nine ubiquitin mutants analysed previously with the observed experimental stability changes.[39] The only clear outlier is the flavodoxin D65K mutant that fails to stabilise the protein. Leaving it aside, the data of the additional 16 mutants can be fit (with R=0.90) to: $\Delta\Delta G_{exp.} = 1.37\Delta\Delta G_{theor.} + 0.75$. Although this is not perfect agreement between prediction and experiment, it shows clearly the usefulness of attempting global protein stabilisation using simple electrostatic calculations.

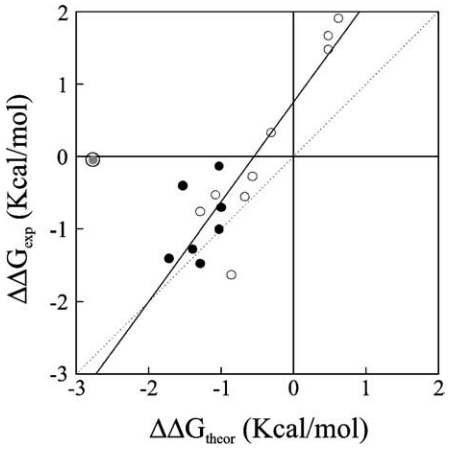

Figure 4. Representation of experimental global unfolding free energy changes (wild-type minus mutant) determined from chemical denaturation *versus* the corresponding theoretical predicted changes for ubiquitin (open circles) and apoflavodoxin (filled circles) mutants. The linear fit of all data points (except that of the apoflavodoxin D65K mutant (surrounded)) is shown as a continuous line. The ideal relationship of $\Delta\Delta G_{exp} = \Delta\Delta G_{theor}$ is shown as a broken line.

Relevant and residual stabilisation of apoflavodoxin by charge reversal mutations

The urea unfolding of apoflavodoxin is a two-state equilibrium where only the native and the fully unfolded states of the protein appear at measurable concentrations.[49] The thermal unfolding, in contrast, is a three-state equilibrium where an intermediate accumulates that, at certain temperatures, constitutes the major species.[31,50] In previous work,[38] the free energy difference of the native to intermediate (N↔I) equilibrium in a three-state protein was termed "relevant stability" (because it is relevant to function) and that of the I↔D equilibrium "residual stability". To investigate the influence of charge reversal mutations on the relevant and residual stabilities of apoflavodoxin, the thermal denaturation of each variant has

been recorded using four different spectroscopic techniques, from which a global analysis of the equilibria energetics can be performed. Qualitatively, the effect of some mutations on the relevant and residual stabilities can be already anticipated from the shifts in the apparent melting temperature (T_m) of individual unfolding curves. This is so because the apparent fluorescence and near-UV CD T_m values are not far from T_{m1} (N↔I), and the apparent absorbance and far-UV CD T_m values are quite close to T_{m2} (I↔D),[50] which reflects the fact that the intermediate displays a close to native secondary structure and UV absorbance, but fluorescence and near-UV CD signals closer to those of the denatured state.[50] Figure 5(b) and (c) show the experimental unfolding curves of the E20K and D126K variants, respectively, which, compared to those of the wild-type protein (Figure 5(a)), illustrate two different trends observed among the mutants. In E20K, the absorbance and far-UVCD curves (those of higher apparent T_m) are clearly shifted towards higher temperatures, relative to wild-type (most stabilising mutants display this trend). In contrast, in D126K, the absorbance and far-UV CD curves are similar to that of the wild-type, while the fluorescence and near-UV CD curves (of lower apparent T_m) are shifted towards higher temperatures.

The global analysis of the thermal unfolding data gathered for each mutant allows us to calculate quite accurately the changes brought about by the charge reversal mutations in the free energy differences of the N↔I and I↔D equilibria (Table 2). The two mutations that hardly modify the global stability (N↔D), as determined by two-state urea unfolding (D65K and D150K, Table 1), cause little change in ΔG_1 or ΔG_2. Of the six mutations that stabilise apoflavodoxin against denaturation by urea, only two (E61K and D126K) stabilise moderately the native state relative to the intermediate of the thermal unfolding (and they do not modify the stability of the intermediate significantly relative to the unfolded state), while four of them (E20K, E40K, E72K, D75K) leave unchanged the stability of the native state, relative to the intermediate, but stabilise markedly the intermediate relative to the unfolded state.

We have checked whether the global stabilisation of the native state measured in the two-state urea unfolding correlates with that calculated by summing the relevant and residual stabilisations determined from the global fitting of the thermal unfolding data. This can occur only if the urea-unfolded and the high temperature-unfolded states are energetically similar with respect to the mutations studied, the stability differences at 44 °C derived from the thermal unfolding global analysis extrapolate to 25 °C (the temperature of the urea-unfolding experiments) and, most importantly, if the simplifications introduced in equations (4) and (5) for the calculation of $\Delta\Delta G_1$ and $\Delta\Delta G_2$ using the data derived from global analysis of the thermal unfolding are reasonable. We show in Figure 6 that

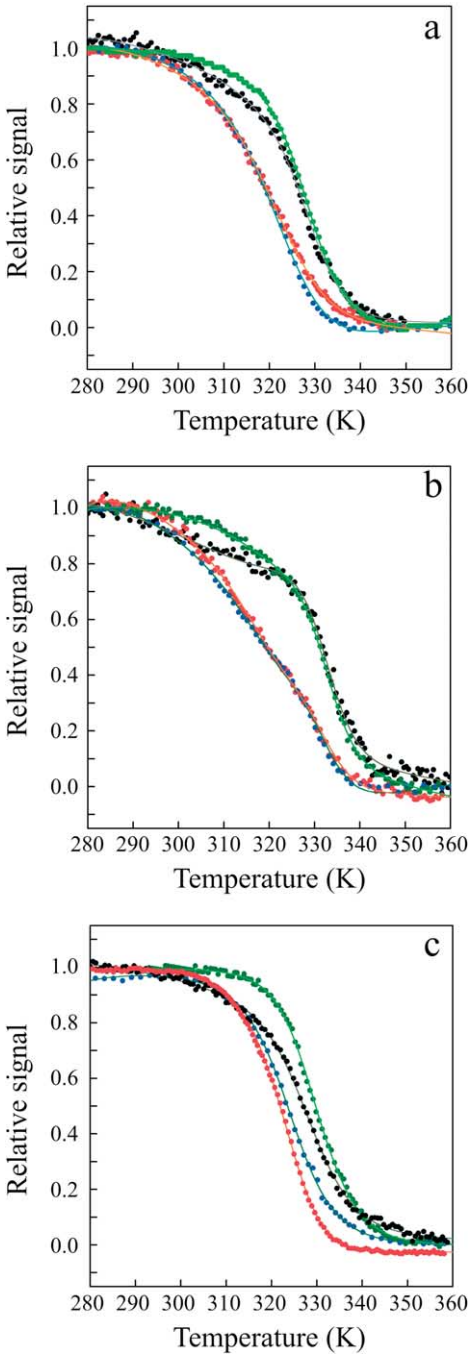

Figure 5. Thermal denaturation curves of wild-type and mutant proteins: wild-type (a), E20K (b) and D126K (c). The global fittings to a three-state equation of the four curves corresponding to each protein are shown as solid lines. The far-UV CD (black dots), near-UV CD (red dots), absorbance (green dots) and fluorescence (blue circles) experimental curves are shown.

the overall stabilisation achieved by the charge reversal mutations and measured by the two-state urea-denaturation curves correlates very well indeed ($R = 0.97$) with that calculated by summing the $\Delta\Delta G_1$ and $\Delta\Delta G_2$ energies derived from the global analysis of the thermal unfolding ($\Delta\Delta G_{thermal} = 0.2 + 1.1\Delta\Delta G_{urea}$), which lends

Table 2. Relevant ($\Delta\Delta G_{N-I}$) and residual ($\Delta\Delta G_{I-D}$) conformational stabilisation in charge reversal mutant apoflavodoxins, relative to wild-type, as determined from three-state thermal denaturation global analysis

Protein	$T_{m1(NI)}$ (K)	$\Delta H_{1(NI)}$ (kcal mol^{-1})	$T_{m2(ID)}$ (K)	$\Delta H_{2(ID)}$ (kcal mol^{-1})	$\Delta\Delta G_{NI}$[a] (kcal mol^{-1})	$\Delta\Delta G_{ID}$[a] (kcal mol^{-1})	$\Delta\Delta G_{ND}$[b] (kcal mol^{-1})
Wt	317.3±0.2	33.9±0.1	329.0±0.1	52.7±0.6	–	–	
E20K	315.8±0.4	30.2±0.3	332.8±0.1	66.1±1.3	+0.14±0.05	−1.20±0.08	−1.06±0.09
E40K	317.9±0.4	33.5±0.1	333.5±0.1	62.5±1.1	−0.06±0.05	−1.16±0.07	−1.22±0.09
E61K	320.9±0.1	41.0±0.9	331.7±0.2	38.0±1.0	−0.46±0.03	+0.22±0.07	−0.24±0.07
D65K	315.2±0.4	30.4±0.1	329.2±0.1	54.9±0.9	+0.20±0.05	−0.11±0.06	+0.09±0.08
E72K	318.5±0.3	33.2±0.2	336.2±0.1	57.8±1.0	−0.13±0.04	−1.38±0.07	−1.50±0.08
D75K	317.4±0.2	34.2±0.1	334.2±0.1	59.6±0.8	−0.01±0.04	−1.14±0.06	−1.15±0.07
D126K	321.9±0.2	40.6±0.2	330.8±0.1	45.1±1.3	−0.58±0.04	+0.03±0.07	−0.55±0.08
D150K	317.2±1.1	30.4±0.7	329.6±0.4	44.7±1.2	+0.01±0.11	+0.21±0.08	+0.22±0.14

The ΔH and T_m values of the two equilibria are derived from a global fit performed, for each protein, on four thermal unfolding curves obtained using fluorescence emission, near-UV CD, near-UV absorbance, and far-UV CD. The standard deviations (SD) have been calculated by interval analysis.[86] Reducing to 1/20th of the data points, the data sets of each protein that are globally fit to the three-state equations leads to SD for the T_m and ΔH values about two and three times larger, respectively, than those reported in the Table. The propagation of those SD to the $\Delta\Delta G$values approximately doubles their SD. Experimental errors for the ΔH and T_m values have been determined for the wild-type protein by independent global analysis of four different sets, each consisting of the four unfolding curves required. They are 2.9 kcal mol^{-1} and 1.1 kcal mol^{-1} for ΔH_1 and ΔH_2, and 0.61 K and 0.19 K for T_{m1} and T_{m2}, respectively. An approximation to the experimental errors associated with the relevant and residual stabilisations reported in this Table can be obtained assuming that the errors are similar for the mutants. Under this assumption, the propagated experimental errors for $\Delta\Delta G_{ID}$ are very similar to those reported in the Table, while those for $\Delta\Delta G_{NI}$ are 1.5 times larger to twice as large.

[a] Free energy differences calculated at 317.3 K (see Materials and Methods). The SD in $\Delta\Delta G_{N-I}$ and $\Delta\Delta G_{I-D}$ have been propagated from those of the enthalpies and mid temperatures as described here.

[b] Calculated as: $\Delta\Delta G_{N-I}+\Delta\Delta G_{I-D}$. SD propagated from those of the individual energy differences.

support to the simplifying assumptions introduced in the calculation of $\Delta\Delta G_1$ and $\Delta\Delta G_2$.

As we discussed previously,[50] the calorimetric fitting of thermograms corresponding to three-state equilibria is particularly difficult when one of the transitions is broad and of low enthalpy, as is the case of the N-to-I transition in apoflavodoxin. As a consequence, the calorimetric fittings performed here require the guessing of the $C_p(I)$ values. In spite of this, the calorimetric fitted data obtained for the wild-type and mutant proteins (Table 3) is in qualitative agreement with the more robust data derived from the spectroscopically followed thermal unfolding and we find reasonable linear correlation between the T_m, ΔH and ΔG values derived from the calorimetric fitting and those deriving from the global spectroscopic fitting (not shown), especially for the second transition of larger enthalpy change. The agreement in the temperatures of the first transition is lower, as expected, because not even a prominent shoulder is observed centred at this temperature in the thermograms of the wild-type and mutants proteins (see Figure 3 of Irun *et al.* [50]).

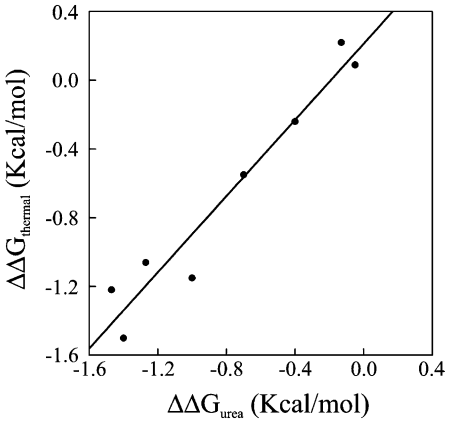

Figure 6. Global stabilisation in apoflavodoxin charge reversal mutants. Correlation of $\Delta\Delta G_{ND}$ energy changes obtained directly from two-state urea denaturation ($\Delta\Delta G_{urea}$) and from global fittings of thermal unfolding curves ($\Delta\Delta G_{thermal}$). The linear fit is represented as a continuous line. The good correlation is indicative of the equivalence of the thermally unfolded and urea-unfolded states of the protein and of the appropriateness of the simplifications introduced in the global analysis of the thermal unfolding curves.

Relationship between stability gain and sequence conservation

Previous work on other proteins has shown that some relationship seems to exist between the level of conservation of a residue at a given position within a family of homologous proteins and the stability increase obtained upon engineering that residue in a member of the family previously displaying a different one. For example, an increase of 18 deg. C in the T_m of GroEL minichaperones was attained by combining six mutations so designed,[51] and an even larger increase of 27 deg. C was obtained by deriving consensus sequences of fungal pythases.[52] The approach has been termed the consensus concept for protein stabilization.[53] An attempt to relate sequence conservation and stability effects quantitatively was done in a study of mutations in the SH3 domain.[54] For charge reversal mutations, such as those studied here, one would expect that some form of the consensus concept applying, the reversal of poorly conserved negative charges would lead to protein stabilisation, while mutating conserved negatively charged

Table 3. Thermodynamic parameters and relevant and residual stabilization free energies for some of the apoflavodoxin variants, as derived from the analysis of DSC thermograms

Protein	$T_{m1(NI)}$ (K)	$\Delta H_{1(NI)}$ (kcal mol^{-1})	$T_{m2(ID)}$ (K)	$\Delta H_{2(ID)}$ (kcal mol^{-1})	$\Delta\Delta G_{NI}$[a] (kcal mol^{-1})	$\Delta\Delta G_{ID}$[a] (kcal mol^{-1})	$\Delta\Delta G_{ND}$[b] (kcal mol^{-1})
Wt	312±2.5	28.0±3.6	326±1.2	60.0±2.4	–	–	–
E20K	309.3±3.1	22.0±5.3	331.2±0.3	66.2±0.7	0.24±0.25	−0.95±0.51	−0.71±0.57
E40K	316.5	32.7	332.5	59.6	−0.4±0.25	−1.12±0.51	−1.52±0.57
E72K	312.8±0.4	27.2±1.0	333.6±0.5	68.4±2.4	−0.07±0.25	−1.40±0.51	−1.47±0.57
D75K	314.8	28.3	330.0	54.3	−0.25±0.25	−0.73±0.51	−0.98±0.57
D126K	320.0	21.8	327.5	39.7	−0.71±0.25	−0.28±0.51	−0.99±0.57
D150K	305.8	25.1	326.9	48.0	0.60±0.25	−0.12±0.51	+0.48±0.57

For the WT form, and the E20K and E72K variants, three DSC experiments were carried out, and average values and the corresponding standard errors are reported for the transition temperatures and transition enthalpy changes. We believe that these reported standard errors are roughly representative of the errors associated to the transition parameters for the other variants. The mutation effects on stabilisation free energies were calculated using the Schellman equation (see the text for details); the errors reported for the $\Delta\Delta G$ values are rough estimates, based on the premise that the main source of error in $\Delta\Delta G$ calculation in this case is the uncertainty in the transition temperatures (estimated to be about 2 K). The error in global stabilisation values: $\Delta\Delta G_{ND}$, have been propagated from the errors of the added free energy differences.

[a] See Table 2.

[b] See Table 2.

residues would produce smaller stabilisations. To test this possibility, we have calculated weighted conservation percentages of negatively charged residues at the mutated positions among the known long-chain flavodoxin sequences. For the purpose of the calculation, Asp and Glu residues have been considered as identical. The overall stabilisation afforded by the charge reversal mutations is represented in Figure 7 *versus* the weighted conservations. A clear linear relationship ($R=0.89$) is found that can be fit to: $\Delta\Delta G=-2.01+0.02\,C$, where C is the percentage of conservation at the position mutated, and the stabilisation is calculated in kcal mol^{-1}. The global stabilisation obtained by the charge reversal mutations can thus be predicted semi-quantitatively from a simple sequence comparison within the flavodoxin family. This and a recently reported study on the effect of carboxylic acid mutations on *Escherichia coli* thioredoxin stability[55] represent, to our knowledge, the best demonstration of the consensus concept at the level of single mutations.

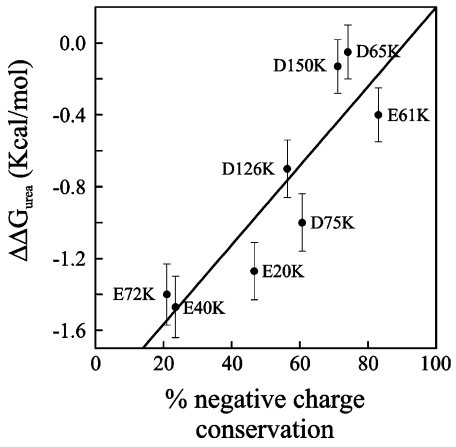

Figure 7. Correlation between the percentage of conservation, within the flavodoxin family, of negatively charged residues in positions equivalent to those mutated in this work and the experimentally determined global stabilisation observed in the mutants analysed. The continuous line is the linear fit. Replacement of poorly conserved acidic residues within the flavodoxin family by positively charged residues leads to large global stabilisations.

Discussion

Rational, easy and global stabilisation of two-state and three-state proteins by charge reversal mutations

Among the different methods devised over the years to increase the conformational stability of proteins, the optimisation of electrostatic interactions stands as one of the more promising. It does not rely on engineering salt-bridges, which usually contribute little to protein stability,[56–59] unless they are carefully engineered.[60] Instead, it is based in optimising medium-range and long-range charge interactions at protein surfaces, and it has been used successfully to stabilise several proteins, such as ubiquitin,[39,46] cold-shock protein,[42,61] the peripheral subunit-binding domain,[41] rubredoxin,[59] and Rnase T_1 and Rnase Sa.[40] We have tested here, using apoflavodoxin (a highly acidic protein), one of the simplest implementations of the electrostatic calculations that constitute the basis of the method.[39,46] Of the eight mutations predicted to stabilise apoflavodoxin that have been tested, six clearly succeed and increase the overall conformational stability by 0.4–1.5 kcal mol^{-1}, while two mutations leave the stability barely changed (Table 1). This is a high rate of success and indicates, as shown for ubiquitin, that even the simplest implementations of electrostatic calculations is an extremely useful guide to achieve protein stabilisation. Furthermore, we find a good qualitative agreement between the calculated stabilisation and that observed experimentally (Figure 4) for both apoflavodoxin and ubiquitin.

Since the method works well for two proteins that are different in several ways (ubiquitin is a very basic protein displaying two-state behaviour towards thermal unfolding, while apoflavodoxin is highly acidic and displays a three-state thermal unfolding), it is clear that it will probably work well for most proteins, provided that their denatured states are truly unfolded. Actually, the simple Tanford–Kirkwood method we use here has been shown to account qualitatively for the effect of charge-deletion and charge-reversal mutations on several proteins,[43] as well as for the salt-induced proton uptake/release behaviour.[62] As for the two mutations that did not stabilise apoflavodoxin (D65K and D150K, Table 2), we notice that they are located near the N terminus of helices 3 and 5, respectively, where the replacement of negatively charged residues by positive ones may exert a substantial destabilising effect through repulsion with the positively charged helix dipoles, which is not taken into account by our electrostatic calculations, as they do not involve partial charges. An estimation of this effect, based on mutational analysis at helical ends,† suggests that an Asp to Lys mutation should destabilise the protein by about 1.6 kcal mol^{-1}, thus significantly opposing the stabilisation afforded by medium and long-range favourable interactions. However, we cannot rule out the possibility that the lack of stabilisation of the D65K and D150K mutations is due to other unforeseen effects on the protein structure (as small changes in the near-UV CD spectra of these mutants could indicate), and it is clear that any current method of rational protein stabilisation will fail when the integrity of the native state is not preserved by the mutation introduced, an eventuality that cannot always be anticipated.

On the other hand, several proteins have been stabilised in recent years, by applying the so-called consensus concept,[53] which consists of introducing at specific positions of a protein the residues that abound in the equivalent position of its homologues. In one instance,[54] a linear relationship was found between the stability of several SH3 single mutants and the logarithm of the percentage of conservation, at the mutated positions, of the residues present in the variants analysed (with $r=0.52$). To investigate if the consensus concept can be applied to predict the stabilisation afforded by charge reversal mutations at the surface of apoflavodoxin, we have applied the rather bold approach of calculating, within the flavodoxin family, percentages of conservation of acidic residues at the mutated sites (without differentiating between Asp and Glu residues). The plot in Figure 7 shows that there is a fair linear relationship between that simple magnitude and the stabilisation afforded by the charge reversal mutations (with $r=0.89$). In practice, this means that we can predict the stability gain of charge reversal mutations (Asp or Glu to Lys) at the surface of apoflavodoxin based solely on this simple conservation criterion.

On a physical basis, the consensus approach for protein stabilisation is another name for the known fact that the statistics of residues in specific locations in proteins often seem to obey Boltzman's law.[55,63,64] In fact, with the appropriate treatment, the relationship between percentage of negative residues and stability gain shown in Figure 7 can be restated in terms of a linear relationship between statistical energies of occurrence of negatively charged residues at the mutated positions and the experimental stability gain, with $r=0.89$ (not shown). This is remarkable because, in the particular case of flavodoxin, the observed high level of conservation of some of the mutated residues (E61, D65, D126 and D150) could be additionally related to a functional role in the binding to its redox partner, the enzyme ferredoxin-$NADP^+$ reductase.[65]

The apoflavodoxin mutations analysed clearly illustrate that performing charge reversal mutations involving the replacement by lysine of negatively charge residues not involved in salt-bridges or hydrogen bonds and located at protein surfaces is a most useful strategy to achieve stabilisation of the native state relative to the unfolded one. As criteria to select the appropriate mutations, both electrostatic calculations and statistical analysis seem to work well. The electrostatic calculations, nevertheless, are superior in that they allow us to estimate with reasonable accuracy the extent of the stabilisation that will be exerted by a given mutation (Figure 4). The feasibility of achieving similarly accurate predictions based on statistics remains to be determined.

The partitioning of global stabilisations into relevant and residual stabilisations

The large global stabilisation of the apoflavodoxin native state, relative to the unfolded one, produced by the charge reversal mutations, as determined from two-state urea-denaturation, has also been observed in the thermal denaturation. Due to the fact that an intermediate accumulates in the thermal unfolding of this protein, the accurate calculation of free energy differences of the N-to-I and of I-to-D equilibria usually requires performing, for each variant, a global fitting of several thermal unfolding curves recorded using different spectroscopic techniques.[66] In this work we have globally fitted fluorescence, near-UV CD, far-UV CD and near-UV absorbance thermal unfolding curves to analyse the energetics of the two equilibria. As is typical of spectroscopically followed thermal unfolding analysis, the fittings for ΔC_p are not very accurate. To increase the accuracy of the calculated changes in free energies associated with the mutations, the ΔC_p values have been constrained to intervals around the means obtained in the analysis of over 40 flavodoxin mutants.[90] Besides, a simplification of equations (4)

† http://wwwbioq.unizar.es/departamento/investigacion/JSS/PFS/alfahelix.html.

and (5) has been used to extrapolate the thermodynamic data obtained for each mutant at its two temperatures of mid-denaturation to a common temperature of comparison (317.3 K). The $\Delta\Delta G_{ND}$ values obtained by summing the $\Delta\Delta G_{NI}$ and $\Delta\Delta G_{ID}$ energies so calculated agree very well with those determined directly from two-state urea unfolding curves (Figure 6; see also Ref. 90), thus supporting the method of analysis, and indicating that the thermally unfolded state is thermodynamically equivalent to the urea-unfolded state.

For proteins with an equilibrium intermediate, the overall free energy difference between the native and denatured states (the global conformational stability) can be partitioned into two terms corresponding to the free energy differences of the two associated equilibria:

$$\Delta G_{ND} = \Delta G_{NI} + \Delta G_{ID}$$

Since the intermediate states of proteins are expected to be deprived of the biologically relevant properties of the native states, the free energy difference between the native and the first intermediate state (ΔG_{NI}) has been termed the relevant stability of the protein, as it is the one that preserves native behaviour, while the free energy difference between the intermediate state and the unfolded one (ΔG_{ID}) was termed residual stability, because it is of little importance for function.[38] For multi-subunit proteins, a similar idea was suggested by Privalov *et al.*,[67] who pointed out that "The stability of the native state of a macromolecule consisting of several independent subunits is not defined by the total Gibbs energy value needed for complete disruption of its structure, but by the stability of the least stable subunit of this system." Somewhat surprisingly, a key issue that has received little attention until now is that of whether the various known stabilising strategies that are successful for increasing the global stability of two-state proteins will be effective for increasing the relevant stability in proteins displaying equilibrium intermediates. This is key for the rational stabilisation of complex proteins against *in vitro* partial unfolding and related closely to the possibility of *in vivo* protein stabilisation through the binding of small ligands.[18] The apoflavodoxin from *Anabaena*, by behaving as two-state with urea-denaturation and as three-state with thermal unfolding, provides an excellent opportunity to investigate the matter. Towards that end we have analysed here one of the most robust strategies of two-state protein stabilisation, i.e. the rational optimisation of electrostatic interactions, and indeed we have found that six of the eight mutations tested have significantly increased the energy gap between the native and the unfolded state. For any of these six successful mutations, an increase of the melting temperature for the first transition will be a reflection of an increase in its relevant stability, while increases of the temperature of the second transition will merely indicate that the thermal intermediate is more resistant to further unfolding. Of the six stabilised mutants, only two (E61K and D126K) display T_{m1} values that are significantly higher than those of the wild-type (Table 2), and can thus be described as having a larger relevant stability than the wild-type protein (by about 0.5 kcal mol^{-1}). In contrast, four mutants (E20K, E40K, E72K and D75K) display almost the same T_{m1} as the wild-type, while their second transition occurs at higher temperatures than it does in the wild-type. Each of these four mutants has almost completely realised a global stabilisation of more than 1 kcal mol^{-1} into increasing the residual stability of the intermediate relative to the denatured state (Figure 8) and, in practice, they are not more stable than the wild-type protein towards thermal inactivation because their N-to-I energy gap has not changed. This result illustrates clearly that even some of the more successful methods of

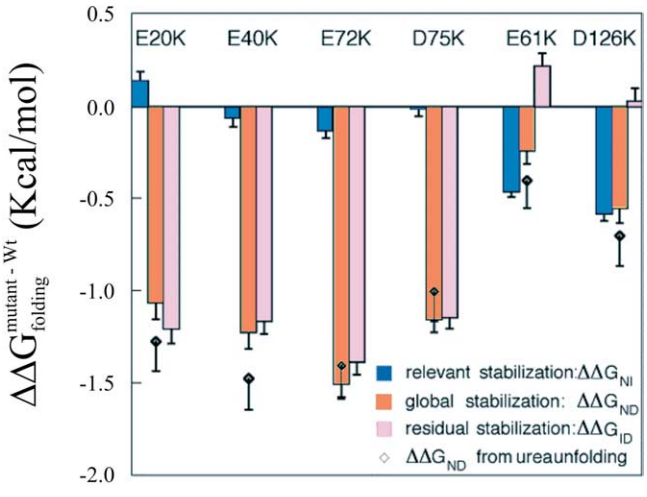

Figure 8. Partitioning of the global stabilisation of the N–D equilibrium in apoflavodoxin mutants into relevant and residual stabilisation of the N–I and I–D equilibria of the thermal unfolding. Orange bars represent the overall stabilisation of the indicated mutants relative to the wild-type protein (as determined from thermal unfolding global analysis of fluorescence, far-UV CD, near-UV CD and near-U absorbance unfolding curves). Dark blue bars represent the associated relevant stabilisations and pink bars the residual stabilisation. The corresponding error bars are shown. The global stabilisation directly calculated from two-state urea unfolding is shown as diamonds (with error bars).

rationally increasing the global stability of proteins may fail in most cases to enhance the relevant stability of complex proteins that display equilibrium intermediates.

The structure of protein intermediates as a guide to increase the relevant stability

Given the fact that achieving a global stabilisation of a protein seems not to guarantee an increase of its relevant stability, it would be helpful to have criteria to help selecting the strategies best suited to stabilise proteins with complex equilibria. It is possible that, through the detailed study of many protein intermediates at the residue level, some of the known strategies may stand out as more effective in order to increase the relevant stability. Meanwhile, guidance should be sought in following the same approach that has been followed to devise rational strategies for overall protein stabilisation; that is, exploiting the structural differences of the two protein conformations in equilibrium; in the case of the relevant stability, those of the native and intermediate states. The problem here is that, unlike the unfolded state that can, in principle, be modelled as essentially deprived from side-chain interactions (which, of course, is not always true), no appropriate general model of a protein intermediate is available. Conjecture based on spectroscopic observations in the early studies leading to the discovery of molten globules favoured a view of their conformation as being rich in secondary structure but lacking native tertiary interactions.[68–70] There is, however, compelling evidence indicating that the conformation of many protein intermediates may be quite close to that of native states.[50,71–74] This vanishing of a clear general picture of the differences between native and intermediate conformations precludes any *a priori* approach to operate on the non-existent obvious differences between the two conformations and, therefore, the rational stabilisation of the native state requires detailed knowledge of the structure of the associated intermediate. This is the main difficulty, because gathering structural information on protein intermediates is usually difficult. In a parallel work,[90] we have used equilibrium ϕ-value analysis[50] to produce a low-resolution structure of the apoflavodoxin thermal intermediate. Basically, the intermediate displays a bipartite structure consisting of a large region with a close to native structure and a smaller region, comprising some 40 residues located in three different loops, which is disordered. This structure satisfactorily explains the contrasting behaviour of the E61K and D126K mutations (which effectively increase the relevant stability) with that of the E20K, E40K, E72K and D75K variants (where the large global stabilisation is merely realised into increasing the residual stability; Figure 8). The latter four mutations appear in regions of the protein that are close to being natively folded in the intermediate (as judged from their ϕ-values) and, hence, strengthening the native conformation there similarly strengthens that of the intermediate without modifying the free energy difference of the two species. The intermediate, nevertheless, is greatly stabilised against further unfolding. In contrast, the two mutations that increase the relevant stability appear located in regions where the intermediate is likely unfolded. For the E61K mutant the evidence is indirect because no information on the structure of the intermediate is available for the loop where the mutated residue is located. Yet, a simple comparison of the X-ray structures of three flavodoxin variants[75,76] singles out this loop as highly flexible and thus very likely to be "unfolded" in the thermal intermediate. For the D126K mutant, the one displaying the clearer pattern of full conversion of the overall stabilisation into relevant stability (Figure 8), the ϕ-structure of the thermal intermediate places residue 126 in a clearly unfolded region, corresponding to the long loop comprising residues 120–140 that is specific to the long-chain flavodoxins.[77] The stabilisation of apoflavodoxin by engineering specific interactions of residues located in regions that becomes disordered in the intermediate can thus stabilise efficiently, in a specific manner, the native state relative to the intermediate. Had the structure of the apoflavodoxin intermediate been available in advance, it could have been anticipated which of the six successful charge reversal mutations were appropriate to increase the relevant stability or rather likely to increase only the residual one. Similarly, any attempt to drive the intermediate conformation of the protein into the native state by small molecule binding should consider the unfolded region of the intermediate as the target. We hope thus that ϕ-structures of equilibrium intermediates obtained by the method developed to investigate the apoflavodoxin thermal intermediate[90] will help define in complex proteins the regions that constitute appropriate targets for successful relevant stabilisation, either by protein engineering or ligand binding.

Materials and Methods

Design of mutants

Calculations of the energies of charge–charge interactions were carried using our implementation of the Tanford–Kirkwood model with the solvent-accessibility correction due to Gurd, as described.[46] This is, admittedly, a very simple model, but has been shown to predict qualitatively the effect of charge-deletion and charge-reversal mutations on the stability of several proteins.[43] We found that, according to these Tanford–Kirkwood calculations, most carboxylic acid groups in the apoflavodoxin molecule are involved in predominantly destabilising interactions with charges of the same sign (as shown by positive values for their charge–charge interaction energy). Among them, we selected for mutation those groups for which the predicted stabilising effect of a charge reversal mutation was about 1 kcal/mol or more and that, in addition, were not involved in hydrogen bonding or salt-bridges. On this basis, we

selected eight mutations (E20K, E40K, E61K, D65K, E72K, D75K, D126K and D150K) for experimental testing. The mutations are spread on several α-helices and loops (Figure 1(a) and Table 1), in regions of negative electrostatic potential (Figure 1(b)). Since the calculation of the stabilising effects does not take into account more subtle electrostatic effects such as those exerted by helical ends,[2,78,79] care has been taken so that the mutations implemented in α-helices represent both internal positions and positions near the helical termini. In this way, the influence of helical position on the predictability of the energetics of lysine-introducing charge reversal mutations can be assessed roughly.

Mutagenesis, protein expression and purification and circular dichroism spectra

The mutations have been introduced in the flavodoxin gene[80] by PCR-based mutagenesis using the Stratagene QuikChange kit. The presence of the mutations was verified by sequencing the entire gene. The protein was purified as described,[49] and the apoprotein prepared by precipitation with trichloroacetic acid.[81] The near-UV and far-UV CD spectra of the mutant apoflavodoxins in 50 mM Mops (pH 7) were measured in a 1 or a 0.1 cm path-length cuvette, respectively, using 30 μM protein solutions.

Urea-denaturation curves

Denaturation curves of wild-type and mutant apoflavodoxins in 50 mM Mops (pH 7) were obtained by measuring the fluorescence of 2 μM protein solutions of different concentrations of urea, using a thermostatically controlled Aminco-Bowman series 2 spectrophotometer from Spectronics Instruments. Fluorescence emission was measured at 25.0(±0.1) °C, with excitation at 280 nm. To minimise protein concentration errors, the ratio of the emission intensities at 320 nm and 380 nm was represented as a function of the concentration of urea.

Thermal denaturation curves

Thermal denaturation of wild-type and mutant apoflavodoxins in 50 mM Mops (pH 7) was followed by far-UV CD at 222 nm, near-UV CD at 288 nm, near-UV absorbance at 288 nm and fluorescence emission (ratio of 320 nm/380 nm, with excitation at 280 nm). We have used the equipment described above for the fluorescence measurements, and a Jasco 710 spectropolarimeter for CD and absorbance measurements. In all the experiments the temperature was raised from 275 °C to 360 °C at a rate of 1.5–2 deg. C/minute. The CD measurements were done with 30 μM protein samples using 1 mm or 1 cm path-length cuvettes for the far-UV and near-UV regions, respectively. The absorbance data were collected during the near-UV CD measurement. The fluorescence data were collected using a protein concentration of 2 μM.

Calorimetric measurements

Differential scanning calorimetry (DSC) experiments were carried out with a VP-DSC calorimeter from Microcal (Northhamton, MA) at scan-rates of 1.5 k/minute or 0.5 k/minute. protein solutions for the calorimetric experiments were prepared by exhaustive dialysis against the buffer. The samples were degassed at room temperature before the calorimetric experiments. calorimetric cells (operating volume ~0.5 ml) were kept under an excess pressure of 30 psi (1 psi≈6.9 kpa) to prevent degassing during the scan. In all measurements, the buffer from the last dialysis step was used in the reference cell of the calorimeter. Several buffer–buffer baselines were obtained before each run with a protein solution in order to ascertain proper equilibration of the instrument. In most experiments, a reheating run was carried out to determine the reversibility of the denaturation process. Finally, an additional buffer–buffer baseline was obtained immediately after the protein runs to check that no significant change in instrumental baseline had occurred when working with aqueous solutions. The level of instrumental baseline reproducibility attained was similar to that described.[50]

Data analyses

Urea-denaturation curves for wild-type and mutant proteins were fit to the following two-state equation:[82]

$$F = \frac{F_N + m_N U + (F_D + m_D U)e^{-\Delta G/RT}}{1 + e^{-\Delta G/RT}} \quad (1)$$

with:

$$\Delta G = m(U_m - U) \quad (2)$$

where F is the fluorescence signal, U is the concentration of urea; U_m is the concentration of urea at mid-denaturation; F_N and F_D are the fluorescence signals of the native and denatured states, respectively, at $[U]=0$, m_N and m_D are the linear dependences of the native and denatured state fluorescence signals, respectively, with concentration of urea; and m is the linear dependence of the free energy with concentration of urea.

The thermal denaturation curves of each protein were fit globally to the following three-state equation that accounts for the presence of an intermediate and is based on the two-state equation from Privalov *et al.*:[83]

$$Y = \frac{Y_N + m_N T + (F_1 + m_1 T)e^{-\Delta G_1/RT} + (F_D + m_D T)e^{-(\Delta G_1 + \Delta G_2)/RT}}{1 + e^{-\Delta G_1/RT} + e^{-(\Delta G_1 + \Delta G_2)/RT}} \quad (3)$$

where ΔG is given by the Gibbs–Hemholtz equation:

$$\Delta G_1 = \Delta H_1(1 - T/T_{m1}) - \Delta C_{p1}((T_{m1} - T) + T\ln(T/T_{m1})) \quad (4)$$

$$\Delta G_2 = \Delta H_2(1 - T/T_{m2}) - \Delta C_{p2}((T_{m2} - T) + T\ln(T/T_{m2})) \quad (5)$$

In equation (3), Y is the spectroscopic signal; Y_N, Y_I and Y_D are spectroscopic signals at $T=0$ K of the native, intermediate and denatured states, respectively; m_N, m_I and m_D are the spectroscopic slopes of the native, intermediate and denatured states, respectively. In equations (4) and (5), ΔH_1, T_{m1} and ΔC_{p1} correspond to the enthalpy change, the melting temperature and the change in heat capacity of the first transition (N↔I); and ΔH_2, T_{m2} and ΔC_{p2} correspond to the enthalpy change, the melting temperature and the change in heat capacity of the second transition (I↔D). For each protein, the four thermal unfolding curves recorded (near-UV and far-UV CD, fluorescence and absorbance) were roughly normalised to represent signal changes between approximately 1 and 0, in order to make each curve contribute similarly to the fitting. Due to the non-negligible slopes of the native and denatured state spectroscopic signals, an exact normalization of a given set of curves, previous to the global fitting, is possible only if individual fits of the different curves are performed. This was avoided because it can bias the subsequent global fitting. The four curves

were then fit globally to equation (3), using the MLAB program (Accelrys Inc.), to obtain a single set of transition temperatures, enthalpy and heat capacity changes describing the two unfolding equilibria, while the spectroscopic values and slopes of the three species were allowed to differ for each curve. It is well known that the fitting of thermal unfolding curves to equations similar to equation (3) does not provide accurate ΔC_p values.[84] Although performing, for each protein, a global fit of several curves increases the reliability of the fitted ΔC_p values, we have still observed too wide a range of fitted ΔC_p values among the different mutants, a clear reflection of their unreliability. Although this hardly affects the accuracy of the fitted $\Delta H(T_m)$ and T_m values, it precludes extrapolation of the data from the transition regions. To improve the fitting of ΔC_p, we have initially calculated the mean and the standard deviations of the ΔC_p values obtained in preliminary fits for 40 different apoflavodoxin mutants analysed in the laboratory. Then, in subsequent fits of the mutants analysed in this work we have imposed the criterion that the fitted ΔC_p values should remain within an interval defined as $|x|-\sigma$ to $|x|+\sigma$, which is 0–2.6 kJ mol^{-1} K^{-1} for ΔC_{p1} and 5.4–7.6 kJ mol^{-1} K^{-1} for ΔC_{p2}.

The calorimetric heat capacity *versus* temperature profiles for the denaturation of apoflavodoxin variants were fit using a three-state model.[50] This fitting procedure assumes that the (temperature-dependent) heat capacity of the intermediate state can be represented as a linear combination of the heat capacities of the native and denatured states, weighted by a parameter Z (for $Z=1$ the heat capacity of the intermediate state would equal that of the denatured state). Fittings were carried out using $Z=0.6$, as suggested by the results reported.[50] The quality of the fittings obtained was similar to that reported by these authors for wild-type (WT) apoflavodoxin.[50] Since the effects of mutations on the overall denaturation free energy were reasonably small, the Schellman equation was used to get an approximation of their values.[85] The Schellman equation can be written as:

$$\Delta\Delta G = \Delta H^0_m \Delta T_m / T^0_m$$

where $\Delta\Delta G$ is the perturbation free energy (i.e. ΔG (variant) $-\Delta G$(WT)), T^0_m and ΔH^0_m are the denaturation temperature and denaturation enthalpy change (at the denaturation temperature) for the WT form, and ΔT_m is the perturbation effect on the denaturation temperature [i.e. T_m(variant) $- T^0_m$]. This procedure was applied to both the N-to-I and the I-to-D transitions using, in each case, the thermodynamic parameters derived from the fittings to the DSC profiles.

Free energy calculations from global analysis of thermal denaturation, and error calculations

The changes produced in the free energies of the N-to-I and I-to-D thermal unfolding equilibria by the mutations analysed can be calculated more accurately from the parameters obtained in the global fitting of the four thermal unfolding curves obtained for each mutant using different spectroscopic techniques. Because of the intrinsically low accuracy of the ΔC_p values, equations (4) and (5) have been simplified by removing their second terms, and an optimal reference temperature has been chosen for comparison of the different mutations (317.3 K, melting temperature for the first transition of the wild-type protein) at which the average contribution of the deleted terms to the changes in free energy differences is at a global minimum.[90] On the other hand, the estimations of the standard deviations provided by the MLAB for each fitted parameter are correct only for linear models, which is not the case of protein unfolding. We have thus calculated standard deviations by a rigorous interval analysis at a 67% confidence level using the method described by Beechem.[86]

Calculation of percentages of conservation of charged residues in the flavodoxin family

A total of 31 sequences of long-chain flavodoxins obtained with a BLAST search[87] against the sequence of the flavodoxin from *Anabaena* PCC7119 were first aligned using ClustalW[88] and the initial alignment was then refined. The 31 sequences of the alignment were weighted following an established procedure,[89] and weighted percentages of conservation of negatively charged residues at each flavodoxin position mutated in this work were calculated.

Acknowledgements

We acknowledge financial support from grants BCM2001-252, P120/2001, BIO2000-1437 and BIO2003-02229. L.A.C. was supported by an FPU fellowship.

References

1. Matthews, B. W., Nicholson, H. & Becktel, W. J. (1987). Enhanced protein thermostability from site-directed mutations that decrease the entropy of unfolding. *Proc. Natl Acad. Sci. USA*, **84**, 6663–6667.
2. Sali, D., Bycroft, M. & Fersht, A. R. (1988). Stabilization of protein structure by interaction of alpha-helix dipole with a charged side-chain. *Nature*, **335**, 740–743.
3. Jackson, S. E. & Fersht, A. R. (1991). Folding of chymotrypsin inhibitor 2.1. Evidence for a two-state transition. *Biochemistry*, **30**, 10428–10435.
4. Khorasanizadeh, S., Peters, I. D., Butt, T. R. & Roder, H. (1993). Folding and stability of a tryptophan-containing mutant of ubiquitin. *Biochemistry*, **32**, 7054–7063.
5. Viguera, A. R., Martinez, J. C., Filimonov, V. V., Mateo, P. L. & Serrano, L. (1993). Thermodynamic and kinetic analysis of the SH3 domain of spectrin shows a two-state folding transition. *Biochemistry*, **33**, 2142–2150.
6. Ganter, C. & Pluckthun, A. (1990). Glycine to alanine substitutions in helices of glyceraldehyde-3-phosphate dehydrogenase: effects on stability. *Biochemistry*, **29**, 9395–9402.
7. Uversky, V. N. & Ptitsyn, O. B. (1994). "Partly folded" state, a new equilibrium state of protein molecules: four-state guanidinium chloride-induced unfolding of beta-lactamase at low temperature. *Biochemistry*, **33**, 2782–2791.
8. Jaenicke, R. (1999). Folding and stability of domain proteins. *Prog. Biophys. Mol. Biol.* **71**, 155–241.
9. Ionescu, R. M., Smith, V. F., O'Neill, J. C. & 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.
10. Pedroso, I., Irun, M. P., Machicado, C. & Sancho, J. (2002). Four-state equilibrium unfolding of an scFv antibody fragment. *Biochemistry*, **41**, 9873–9884.
11. Goddette, D. W., Christianson, T., Ladin, B. F., Lau, M., Mielenz, J. R., Paech, C. *et al*. (1993). Strategy and implementation of a system for protein engineering. *J. Biotechnol.* **28**, 41–54.
12. Masui, A., Fujiwara, N. & Imanaka, T. (1994). Stabilization and rational design of serine protease AprM under highly alkaline and high-temperature conditions. *Appl. Environ. Microbiol.* **60**, 3579–3584.
13. Tange, T., Taguchi, S., Kojima, S., Miura, K. & Momose, H. (1994). Improvement of a useful enzyme (subtilisin BPN′) by an experimental evolution system. *Appl. Microbiol. Biotechnol.* **41**, 239–244.
14. Goetz, G., Iwan, P., Hauer, B., Breuer, M. & Pohl, M. (2001). Continuous production of (*R*)-phenylacetylcarbinol in an enzyme-membrane reactor using a potent mutant of pyruvate decarboxylase from *Zymomonas mobilis*. *Biotechnol. Bioeng.* **74**, 317–325.
15. Gulich, S., Linhult, M., Stahl, S. & Hober, S. (2002). Engineering streptococcal protein G for increased alkaline stability. *Protein Eng.* **15**, 835–842.
16. McHugh, L., Hu, S., Lee, B. K., Santora, K., Kennedy, P. E., Berger, E. A. *et al*. (2002). Increased affinity and stability of an anti-HIV-1 envelope immunotoxin by structure-based mutagenesis. *J. Biol. Chem.* **277**, 34383–34390.
17. McDonagh, C. F., Beam, K. S., Wu, G. J., Chen, J. H., Chace, D. F., Senter, P. D. & Francisco, J. A. (2003). Improved yield and stability of L49-sFv-beta-lactamase, a single-chain antibody fusion protein for anticancer prodrug activation, by protein engineering. *Bioconjug. Chem.* **14**, 860–869.
18. Issaeva, N., Friedler, A., Bozko, P., Wiman, K. G., Fersht, A. R. & Selivanova, G. (2003). Rescue of mutants of the tumor suppressor p53 in cancer cells by a designed peptide. *Proc. Natl Acad. Sci. USA*, **100**, 13303–13307.
19. Matsumura, M., Becktel, W. J., Levitt, M. & Matthews, B. W. (1989). Stabilization of phage T4 lysozyme by engineered disulfide bonds. *Proc. Natl Acad. Sci. USA*, **86**, 6562–6566.
20. Takagi, H., Takahashi, T., Momose, H., Inouye, M., Maeda, Y., Matsuzawa, H. & Ohta, T. (1990). Enhancement of the thermostability of subtilisin E by introduction of a disulfide bond engineered on the basis of structural comparison with a thermophilic serine protease. *J. Biol. Chem.* **265**, 6874–6878.
21. Mansfeld, J., Vriend, G., Dijskstra, B. W., Veltman, R., Van den Burg, B., Venema, G. *et al*. (1997). Extreme stabilization of a thermolysin-like protease by an engineered disulfide bond. *J. Biol. Chem.* **272**, 11152–11156.
22. Hecht, M. H., Sturtevant, J. M. & Sauer, R. T. (1986). Stabilization of lambda repressor against thermal denaturation by site-directed Gly-Ala changes in alpha-helix 3. *Proteins: Struct. Funct. Genet.* **1**, 43–46.
23. Nicholson, H., Tronrud, D. E., Becktel, W. J. & Matthews, B. W. (1992). Analysis of the effectiveness of proline substitutions and glycine replacements in increasing the stability of phage T4 lysozyme. *Biopolymers*, **32**, 1431–1441.
24. Blaber, M., Baase, W. A., Gassner, N. & Matthews, B. W. (1995). Alanine scanning mutagenesis of the alpha-helix 115–123 of phage T4 lysozyme: effects on structure stability and the binding of solvent. *J. Mol. Biol.* **246**, 317–330.
25. Munoz, V., Cronet, P., Lopez-Hernandez, E. & Serrano, L. (1996). Analysis of the effect of local interactions on protein stability. *Fold. Des.* **1**, 167–178.
26. Villegas, V., Viguera, A. R., Aviles, F. X. & Serrano, L. (1996). Stabilization of proteins by rational design of alpha-helix stability using helix/coil transition theory. *Fold. Des.* **1**, 29–34.
27. Lacroix, E., Viguera, A. R. & Serrano, L. (1998). Elucidating the folding problem of alpha-helices: local motifs, long-range electrostatics, ionic-strength dependence and prediction of NMR parameters. *J. Mol. Biol.* **284**, 173–191.
28. Akasako, A., Haruki, M., Oobatake, M. & Kanaya, S. (1997). Conformational stabilities of *Escherichia coli* Rnase HI variants with a series of amino acid substitutions at a cavity within the hydrophobic core. *J. Biol. Chem.* **272**, 18686–18693.
29. Lassalle, M. W., Yamada, H., Morii, H., Ogata, K., Sarai, A. & Akasaka, K. (2001). Filling a cavity dramatically increases pressure stability of the c-Myb R2 subdomain. *Proteins: Struct. Funct. Genet.* **45**, 96–101.
30. Ohmura, T., Ueda, T., Ootsuka, K., Saito, M. & Imoto, T. (2001). Stabilization of hen egg white lysozyme by a cavity-filling mutation. *Protein Sci.* **10**, 313–320.
31. Irun, M. P., Maldonado, S. & Sancho, J. (2001). Stabilization of apoflavodoxin by replacing hydrogen-bonded charged Asp or Glu residues by the neutral isosteric Asn or Gln. *Protein Eng.* **14**, 173–181.
32. Wallon, G., Kryger, G., Lovett, S. T., Oshima, T., Ringe, D. & Petsko, G. A. (1997). Crystal structures of *Escherichia coli* and *Salmonella typhimurium* 3-isopropylmalate dehydrogenase and comparison with their thermophilic counterpart from *Thermus thermophilus*. *J. Mol. Biol.* **266**, 1016–1031.
33. Bogin, O., Peretz, M., Hacham, Y., Korkhin, Y., Frolow, F., Kalb, A. J. & Burstein, Y. (1998). Enhanced thermal stability of *Clostridium beijerinckii* alcohol dehydrogenase after strategic substitution of amino acid residues with pralines from the homologous thermophilic *Thermoanaerobacter brockii* alcohol dehydrogenase. *Protein Sci.* **7**, 1156–1163.
34. Hasegawa, J., Shimahara, H., Mizutani, M., Uchiyama, S., Arai, H., Ishii, M. *et al*. (1999). Stabilization of *Pseudomonas aeruginosa* cytochrome *c*(551) by systematic amino acid substitutions based on the structure of thermophilic *Hydrogenobacter thermophilus* cytochrome *c*(552). *J. Biol. Chem.* **274**, 37533–37537.
35. Perl, D., Mueller, U., Heinemann, U. & Schmid, F. X. (2000). Two exposed amino acid residues confer thermostability on a cold shock protein. *Nature Struct. Biol.* **7**, 380–383.
36. Hasegawa, J., Uchiyama, S., Tanimoto, Y., Mizutani, M., Kobayashi, Y., Sambongi, Y. & Igarashi, Y. (2000). Selected mutations in a mesophilic cytochrome *c* confer the stability of a thermophilic counterpart. *J. Biol. Chem.* **275**, 37824–37828.
37. Nemeth, A., Kamondi, S., Szilagyi, A., Magyar, C., Kovari, Z. & Zavodszky, P. (2002). Increasing the thermal stability of cellulase C using rules learned from thermophilic proteins: a pilot study. *Biophys. Chem.* **96**, 229–241.
38. Sancho, J., Bueno, M., Campos, L. A., Fernandez-Recio, J., Irun, M. P., Lopez, J. *et al*. (2002). The relevant stability of proteins with equilibrium intermediates. *Sci. World J.* **2**, 1209–1215.

39. Loladze, V. V., Ibarra-Molero, B., Sanchez-Ruiz, J. M. & Makhatadze, G. I. (1999). Engineering a thermostable protein *via* optimization of charge-charge interactions on the protein surface. *Biochemistry*, **38**, 16419–16423.
40. Grimsley, G. R., Shaw, K. L., Fee, L. R., Alston, R. W., Huyghues-Despointes, B. M., Thurlkill, R. L. *et al*. (1999). Increasing protein stability by altering long-range coulombic interactions. *Protein Sci.* **8**, 1843–1849.
41. Spector, S., Wang, M., Carp, S. A., Robblee, J., Hendsch, Z. S., Fairman, R. *et al*. (2000). Rational modification of protein stability by the mutation of charged surface residues. *Biochemistry*, **39**, 872–879.
42. Perl, D. & Schmid, F. X. (2001). Electrostatic stabilization of a thermophilic cold shock protein. *J. Mol. Biol.* **313**, 343–357.
43. Sanchez-Ruiz, J. M. & Makhatadze, G. I. (2001). To charge or not to charge? *Trends Biotechnol.* **19**, 132–135.
44. Garcia-Moreno, B., Chen, L. X., March, K. L., Gurd, R. S. & Gurd, F. K. N. (1985). Electrostatic interactions in sperm whale myoglobin. Site specificity, roles in structural elements, and external electrostatic potential distributions. *J. Biol. Chem.* **260**, 14070–14082.
45. Yang, A. S. & Honig, B. (1994). Structural origins of pH and ionic strength effects on protein stability. Acid denaturation of sperm whale apomyoglobin. *J. Mol. Biol.* **237**, 602–614.
46. Ibarra-Molero, B., Loladze, V. V., Makhatadze, G. I. & Sánchez-Ruiz, J. M. (1999). Thermal *versus* guanidine-induced unfolding of ubiquitin. An analysis in terms of the contributions from charge–charge interactions to protein stability. *Biochemistry*, **38**, 8138–8149.
47. Tanford, C. & Kirkwood, J. G. (1957). Theory of protein titration curves. I. General equations for impenetrable spheres. *J. Am. Chem. Soc.* **79**, 5333–5339.
48. Genzor, C. G., Perales-Alcon, A., Sancho, J. & Romero, A. (1996). Closure of a tyrosine/tryptophan aromatic gate leads to a compact fold in apoflavodoxin. *Nature Struct. Biol.* **3**, 329–332.
49. Genzor, C. G., Beldarrain, A., Gomez-Moreno, C., Lopez-Lacomba, J. L., Cortijo, M. & Sancho, J. (1996). Conformational stability of apoflavodoxin. *Protein Sci.* **5**, 1376–1388.
50. Irun, M. P., Garcia-Mira, M. M., Sanchez-Ruiz, J. M. & Sancho, J. (2001). Native hydrogen bonds in a molten globule: the apoflavodoxin thermal intermediate. *J. Mol. Biol.* **306**, 877–888.
51. Wang, Q., Buckle, A. M., Foster, N. W., Johnson, C. M. & Fersht, A. R. (1999). Design of highly stable functional CroEL minichaperones. *Protein Sci.* **8**, 2186–2193.
52. Lehmann, M., Loch, C., Middendorf, A., Studer, D., Lassen, S. F., Pasamontes, L. *et al*. (2002). The consensus concept for thermostability engineering of proteins: further proof of concept. *Protein Eng.* **15**, 403–411.
53. Lehmann, M., Pasamontes, L., Lassen, S. F. & Wyss, M. (2000). The consensus concept for thermostability engineering of proteins. *Biochim. Biophys. Acta*, **1543**, 408–415.
54. Maxwell, K. L. & Davidson, A. R. (1998). Mutagenesis of a buried polar interaction in an SH3 domain: sequence conservation provides the best prediction of stability effects. *Biochemistry*, **37**, 16172–16182.
55. Godoy-Ruiz, R., Perez-Jimenez, R., Ibarra-Molero, B. & Sanchez-Ruiz, J. M. (2004). Relation between protein stability, evolution and structure, as probed by carboxylic acid mutations. *J. Mol. Biol.* **336**, 313–318.
56. Anderson, D. E., Becktel, W. J. & Dahlquist, F. W. (1990). pH-induced denaturation of proteins: a single salt bridge contributes 3–5 kcal/mol to the free energy of folding of T4 lysozyme. *Biochemistry*, **29**, 2403–2408.
57. Sali, D., Bycroft, M. & Fersht, A. R. (1991). Surface electrostatic interactions contribute little on stability of barnase. *J. Mol. Biol.* **220**, 779–788.
58. Ramos, C. H. I., Kay, M. S. & Baldwin, R. L. (1999). Putative interhelix ion pairs involved in the stability of myoglobin. *Biochemistry*, **38**, 9783–9790.
59. Strop, P. & Mayo, S. L. (2000). Contribution of surface salt bridges to protein stability. *Biochemistry*, **39**, 1251–1255.
60. Makhatadze, G. I., Loladze, V. V., Ermolenko, D. N., Chen, X. & Thomas, S. T. (2003). Contribution of surface salt bridges to protein stability: guidelines for protein engineering. *J. Mol. Biol.* **327**, 1135–1148.
61. Makhatadze, G. I., Loladze, V. V., Gribenko, A. V. & Lopez, M. M. (2004). Mechanism of thermostabilization in a designed cold shock protein with optimized surface electrostatic interactions. *J. Mol. Biol.* **336**, 929–942.
62. Garcia-Mira, M. M. & Sanchez-Ruiz, J. M. (2001). pH corrections and protein ionisation in water/guanidinium chloride. *Biophys. J.* **81**, 3489–3502.
63. Fernandez-Recio, J. & Sancho, J. (1998). Intrahelical side-chain interactions in alpha helices: poor correlation between energetics and frequency. *FEBS Letters*, **429**, 99–103.
64. Shortle, D. (2003). Propensities, probabilities, and the Boltzmann hypothesis. *Protein Sci.* **12**, 1298–1302.
65. Nogues, I., Martinez-Julvez, M., Navarro, J. A., Hervas, M., Armenteros, L., de la Rosa, M. A. *et al*. (2003). Role of hydrophobic interactions in the flavodoxin mediated electron transfer from photosystem I to ferredoxin-NADP$^+$ reductase in Anabaena PCC 7119. *Biochemistry*, **42**, 2036–2045.
66. Luo, J., Iwakura, M. & Matthews, C. R. (1995). Detection of a stable intermediate in the thermal unfolding of a cysteine-free form of dihydrofolate reductase from *Escherichia coli*. *Biochemistry*, **34**, 10669–10675.
67. Privalov, P. L., Mateo, P. L., Khechinashvili, N. N., Stepanov, V. M. & Revina, L. D. (1981). Comparative thermodynamic study of pepsinogen and pepsin structure. *J. Mol. Biol.* **152**, 445–464.
68. Herold, M. & Kirschner, K. (1990). Reversible dissociation and unfolding of aspartate aminotransferase from *Escherichia coli*: characterization of a monomeric intermediate. *Biochemistry*, **29**, 1907–1913.
69. Semisotnov, G. V., Rodionova, N. A., Razgulyaev, O. I., Uversky, V. N., Gripas, A. F. & Gilmanshin, R. I. (1991). Study of the "molten globule" intermediate state in protein folding by a hydrophobic fluorescent probe. *Biopolymers*, **31**, 119–128.
70. Kuroda, Y., Kidokoro, S. & Wada, A. (1992). Thermodynamic characterization of cytochrome *c* at low pH. Observation of the molten globule state and of the cold denaturation process. *J. Mol. Biol.* **223**, 1139–1153.
71. Buchner, J., Renner, M., Lilie, H., Hinz, H. J., Jaenicke, R., Kiefhabel, T. & Rudolph, R. (1991). Alternatively folded states of an immunoglobulin. *Biochemistry*, **30**, 6922–6929.
72. Rehaber, V. & Jaenicke, R. (1993). The low-temperature folding intermediate of hyperthermophilic D-glyceraldehyde-3-phosphate dehydrogenase from *Thermotoga maritima* shows a native-like cooperative unfolding transition. *FEBS Letters*, **317**, 163–166.

73. Marmorino, J. L., Lehti, M. & Pielak, G. J. (1998). Native tertiary structure in an A-state. *J. Mol. Biol.* **275**, 379–388.
74. Luo, Y. & Baldwin, R. L. (2001). How Ala↠Gly mutations in different helices affect the stability of the apomyoglobin molten globule. *Biochemistry,* **40**, 5283–5289.
75. Lostao, A., El Harrous, Daoudi, F., Romero, A., Parody-Morreale, A. & Sancho, J. (2000). Dissecting the energetics of the apoflavodoxin–FMN complex. *J. Biol. Chem.* **275**, 9518–9526.
76. Lostao, A., Daoudi, F., Irun, M. P., Ramon, A., Fernandez-Cabrera, C., Romero, A. & Sancho, J. (2003). How FMN binds to anabaena apoflavodoxin: a hydrophobic encounter at an open binding site. *J. Biol. Chem.* **278**, 24053–24061.
77. Mayhew, S. G. & Ludwig, M. L. (1975). *Flavodoxins and electron-transferring flavoproteins*. Academic Press, New York, NY.
78. Nicholson, H., Becktel, W. J. & Matthews, B. W. (1988). Enhanced protein thermostability from designed mutations that interact with alpha-helix dipoles. *Nature*, **336**, 651–656.
79. Sancho, J., Serrano, L. & Fersht, A. R. (1992). Histidine residues at the N- and C-termini of alpha helices: perturbed pK_as and protein stability. *Biochemistry,* **31**, 2253–2258.
80. Fillat, M. F., Borrias, W. E. & Weisbeerk, P. J. (1991). Isolation and overexpression in *Escherichia coli* of the flavodoxin gene from Anabaena PCC7119. *Biochem. J.* **280**, 187–191.
81. Edmonson, D. E. & Tollin, G. (1971). Chemical and physical characterization of the Shethna flavoprotein and apoprotein and kinetics and thermodynamics of flavin analog binding to the apoprotein. *Biochemistry,* **10**, 124–132.
82. Santoro, M. M. & 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.
83. Privalov, P. L. (1979). Stability of proteins: small globular proteins. *Advan. Protein Chem.* **33**, 167–241.
84. Pace, C. N., Shirley, B. A. & Thompson, J. A. (1989). *Measuring the conformational stability of a protein Protein Structure, a Practical Approach*. IRL Press, Oxford pp. 311–330.
85. Schellman, J. A. (1987). The thermodynamic stability of proteins. *Annu. Rev. Biophys. Biophys. Chem.* **16**, 115–137.
86. Beechem, J. M. (1992). Global analysis of biochemical and biophysical data. *Methods Enzymol.* **210**, 37–54.
87. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucl. Acids Res.* **25**, 3389–3402.
88. Higgins, D. G., Thompson, J. D. & Gibson, T. J. (1996). Using CLUSTAL for multiple sequence alignments. *Methods Enzymol.* **266**, 383–402.
89. Henikoff, S. & Henikoff, J. G. (1994). Position-based sequence weights. *J. Mol. Biol.* **243**, 574–578.
90. Campos, L. A., Bueno, M., Lopez-Llano, J., Jiménez, M. A. & Sancho, J. (2004). Structure of stable protein folding intermediates by equilibrium ϕ-analysis: the apoflavodoxin thermal intermediate. *J. Mol. Biol.*, in press.

Edited by F. Schmid

(Received 1 June 2004; received in revised form 31 August 2004; accepted 20 September 2004)