

## A Double-Deletion Method to Quantifying Incremental Binding Energies in Proteins from Experiment: Example of a Destabilizing Hydrogen Bonding Pair

Luis A. Campos,<sup>\*†</sup> Santiago Cuesta-López,<sup>\*‡</sup> Jon López-Llano,<sup>\*†</sup> Fernando Falo,<sup>\*‡</sup> and Javier Sancho<sup>\*†</sup>

<sup>\*</sup>Biocomputation and Complex Systems Physics Institute, <sup>†</sup>Departamento de Bioquímica y Biología Molecular y Celular, and

<sup>‡</sup>Departamento de Física de la Materia Condensada, Facultad de Ciencias, Universidad de Zaragoza, Zaragoza, Spain

**ABSTRACT** The contribution of a specific hydrogen bond in apoflavodoxin to protein stability is investigated by combining theory, experiment and simulation. Although hydrogen bonds are major determinants of protein structure and function, their contribution to protein stability is still unclear and widely debated. The best method so far devised to estimate the contribution of side-chain interactions to protein stability is double mutant cycle analysis, but the interaction energies so derived are not identical to incremental binding energies (the energies quantifying net contributions of two interacting groups to protein stability). Here we introduce double-deletion analysis of 'isolated' residue pairs as a means to precisely quantify incremental binding. The method is exemplified by studying a surface-exposed hydrogen bond in a model protein (Asp96/Asn128 in apoflavodoxin). Combined substitution of these residues by alanines slightly destabilizes the protein due to a decrease in hydrophobic surface burial. Subtraction of this effect, however, clearly indicates that the hydrogen-bonded groups in fact destabilize the native conformation. In addition, molecular dynamics simulations and classic double mutant cycle analysis explain quantitatively that, due to frustration, the hydrogen bond must form in the native structure because when the two groups get approximated upon folding their binding becomes favorable. We would like to remark that 1), this is the first time the contribution of a specific hydrogen bond to protein stability has been measured by experiment; and 2), more hydrogen bonds need to be analyzed to draw general conclusions on protein hydrogen bond energetics. To that end, the double-deletion method should be of help.

### INTRODUCTION

The energetics of biological macromolecules is a central unsolved problem of modern biology that is at the core of highly important phenomena such as protein folding and protein/ligand recognition. Understanding the contribution to protein stability of the various weak interactions between residues that appear in native protein structures will bring insight into the protein folding problem and will help to rationally tailor protein stability (Matsumura et al., 1989; Fersht and Serrano, 1993; Honig and Yang, 1995; Pace et al., 1996; Richards, 1997; Perl et al., 2000; Sanchez-Ruiz and Makhataadze, 2001). Despite advances in recent years, many fundamental questions remain unanswered. As a chief example, it is still unclear whether ubiquitous hydrogen bonds contribute to protein stability. Conflicting views are held on the matter because no available technique can measure the net contribution of any two interacting groups to protein stability (the so-called incremental binding energy (Fersht et al., 1992)). Usually, estimations of the contribution of a given interaction to protein stability are based either on side-chain deletion experiments or on double-mutant cycle analysis. It is clear, however, that simple side-chain deletion experiments aimed at breaking a given interaction and comparing wild-type and mutant stabilities are not informa-

tive because, in most cases, additional interactions are disrupted within the protein (Fersht, 1987; Yang and Honig, 1995). The double-mutant method (Carter et al., 1984; Horovitz et al., 1990) was conceived to alleviate this problem, and although it does not measure incremental binding energies, it allows us to determine an interaction energy between the two side chains, which, for hydrogen bonds, represents a maximum value for the contribution of the interacting groups to protein stability (Fernandez-Recio et al., 1999). In this way, double-mutant cycle analysis provides upper limit values for the incremental binding energy. The problem is that since the various interaction energies so far measured are typically small (0.0 to  $-1.0$  kcal mol<sup>-1</sup>), the actual stabilizing or destabilizing contribution of the bonded groups depends heavily on the value of the solvation energies of polar atoms, which cannot be determined easily.

To solve this problem, we introduce here a different approach, which we term double-deletion analysis. This method focuses on pairs of interacting residues that, beyond their  $\beta$ -carbons, do not establish contacts with other protein residues. We show that when two such residues are simultaneously replaced by alanines, the stability difference between the wild-type and double-mutant protein, properly corrected for small differences in buried hydrophobic area, equals the so-called incremental binding energy. We then apply this double-deletion analysis to quantify, for the first time, the incremental binding energy associated to a pair of surface-exposed, hydrogen-bonded groups in a model protein. Our results, which certainly are not claimed to represent

*Submitted July 23, 2004, and accepted for publication November 4, 2004.*

Address reprint requests to Javier Sancho, Biocomputation and Complex Systems Physics Institute, Facultad de Ciencias, Universidad de Zaragoza, 50009-Zaragoza, Spain. E-mail: jsancho@unizar.es.

© 2005 by the Biophysical Society

0006-3495/05/02/1311/11 \$2.00

doi: 10.1529/biophysj.104.050203

all types of hydrogen bonds in proteins, clearly show that some protein hydrogen bonds destabilize the native conformation. Using classic double-mutant cycle analysis and molecular dynamics simulations, we discuss why they are formed nevertheless.

## MATERIALS AND METHODS

### Theory of double-deletion analysis

Let  $i$  and  $j$  be interacting residues in a protein. In well-chosen cases, where the individual and the simultaneous replacement of these groups by alanines does not alter the local protein structure, a double-mutant cycle can be constructed with the wild-type, single mutants, and double mutants so that an  $ij$  interaction energy is calculated from the conformational stabilities of the four proteins (Fersht et al., 1992).

$$\Delta G_{\text{int}} = \Delta G_{\text{wt}} - \Delta G_{i0} - (\Delta G_{0j} - \Delta G_{00}), \quad (1)$$

where  $\Delta G_{\text{wt}}$ ,  $\Delta G_{i0}$ ,  $\Delta G_{0j}$ , and  $\Delta G_{00}$  are the stabilities of the wild-type, the  $j \rightarrow \text{Ala}$ , the  $i \rightarrow \text{Ala}$ , and the double-mutant protein, respectively. Using energy inventories (Fig. 1), it can be shown that, for nondisruptive mutations, the interaction energy is made of the following terms (all relative to the unfolded state):

$$\Delta G_{\text{int}} = G_{ij} + G_{iw}(ij) + G_{jw}(ij) - G_{iw}(i0) - G_{jw}(0j) + \Delta \Delta G_{\text{Pw}}(ij - i0 - 0j + 00), \quad (2)$$

where  $G_{ij}$  refers to the specific interaction of the two residues,  $G_{iw}(ij)$  and  $G_{jw}(ij)$  are the solvation energies of the two residues in the wild-type protein,  $G_{iw}(i0)$  and  $G_{jw}(0j)$  are the solvation energies of each residue in the single-mutant proteins, and  $\Delta \Delta G_{\text{Pw}}(ij - i0 - 0j + 00)$  summarizes the changes in the solvation of the rest of the protein in the four proteins.

Suppose now that the  $i$  and  $j$  interacting residues do not contact, beyond their  $\beta$ -carbons, any other residue in the protein. If, in addition, long-range electrostatic interactions are masked by working at high ionic strength, the interaction of the  $i$  and  $j$  residues with the rest of the protein (relative to that of alanine) is zero. Thus, if the conformational stability of the double mutant

is subtracted from that of the wild-type protein, a “double-deletion” energy ( $\Delta G_{\text{dd}}$ ) is obtained that, according to the energy inventory (Fig. 1), equals

$$\Delta G_{\text{dd}} = G_{ij} + G_{iw}(ij) + G_{jw}(ij) + \Delta G_{\text{Pw}}(ij - 00). \quad (3)$$

On the other hand, the contribution to protein stability (relative to two alanines) of a pair of residues that interact in the native conformation is given by the incremental binding energy ( $\Delta G_{\text{b}}$ ), defined as (Horovitz et al., 1990; Fersht et al., 1992):

$$\Delta G_{\text{b}} = G_{ij} + G_{iw}(ij) + G_{jw}(ij). \quad (4)$$

Combining Eqs. 3 and 4,

$$\Delta G_{\text{b}} = \Delta G_{\text{dd}} - \Delta G_{\text{Pw}}(ij - 00). \quad (5)$$

Equation 5 is the key to double-deletion analysis because in many cases, as in the example presented in this work, the solvation term ( $\Delta G_{\text{Pw}}(ij - 00)$ ) refers to apolar surface, and its calculation is feasible from known empirical equations (see below). It should be noted that, in  $\Delta G_{\text{Pw}}(ij - 00)$ , the solvation of the mutated residues in the unfolded state does not cancel out, unlike in double-mutant analysis. Since both  $G_{\text{Pw}}(ij)$  and  $G_{\text{Pw}}(00)$  are differential solvation energies (folded minus unfolded),  $\Delta G_{\text{Pw}}(ij - 00)$  can be expressed as

$$\Delta G_{\text{Pw}}(ij - 00) = \Delta G_{\text{Pw}}^{\text{fold}}(ij - 00) - \Delta G_{\text{Pw}}^{\text{unf}}(ij - 00). \quad (6)$$

The  $\Delta G_{\text{Pw}}^{\text{fold}}$  term can be calculated from the surface-exposed areas in the wild-type and double-mutant folded structures. The  $\Delta G_{\text{Pw}}^{\text{unf}}$  term, from the exposure in the unfolded state of the  $\beta$ -carbons of the wild-type  $i$  and  $j$  residues and of the alanine ones in the double mutant. As in classical double-mutant cycle analysis, it is assumed that the mutated residues do not interact in the unfolded state.

### Surface calculations and quantification of solvation energies

The double-deletion method has been applied to determine the contribution to protein stability of a surface-exposed hydrogen bond formed by the Asp96 and Asn128 side chains of the apoflavodoxin from *Anabaena* PCC 7117 (1ftg). To that end, the solvent-accessible surface areas of the wild-type and the D96A/N128A double-mutant proteins have been calculated in two different ways. One way uses the x-ray structure of the wild-type protein and a model of the double mutant that was built by substituting the Asp and Asn residues with Ala. Solvent-accessible surface area is calculated with Naccess 2.1.1 (Hubbard and Thornton, 1993) using a probe sphere of 1.4 Å (Lee and Richards, 1971). The other way uses, as representatives of the proteins, averages of the structures obtained along molecular dynamics simulations (see below). Since the local root mean-square deviations (RMSD) ( $t - t = 0$ ) around the hydrogen bond investigated hardly change along the simulations of the proteins, structures have been averaged that sample the entire trajectories. In this approach, average solvent-accessible surface areas have been calculated using Naccess 2.1.1 (Hubbard and Thornton, 1993), interfaced with CHARMM through a home-made program. The surface-exposed areas of the proteins, calculated by either of the two methods, have been then used to calculate the changes in solvent-exposed area upon mutating D96 and N128 to Ala (excluding the mutated carboxyl and carboxamide groups, which are explicitly excluded in the  $\Delta G_{\text{Pw}}(ij - 00)$  term of Eq. 3 because this term refers to the interactions between the rest of the protein and water).

The surface area of the beta carbons of residues D96, N128, A96, and A128 in the unfolded state have been calculated using data from molecular dynamics simulations of Ala-X-Ala tripeptides (Zielenkiewicz and Saenger,

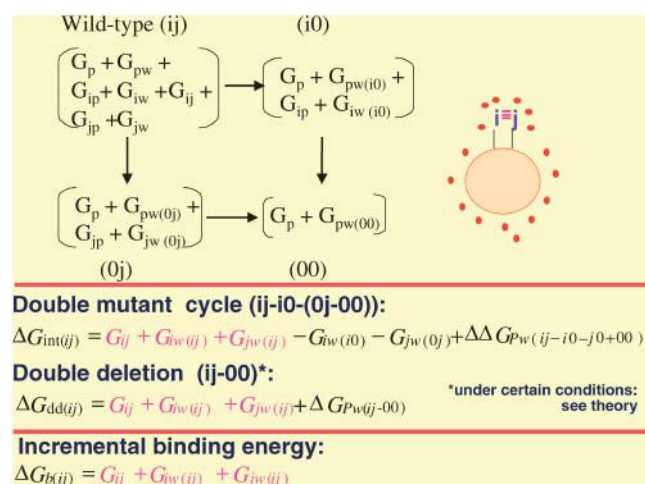


FIGURE 1 Energy inventory in double-mutant cycle and double-deletion analyses. The equations show the relationship between the incremental binding energy from the unfolded state (the contribution of any two groups to protein stability), the double-mutant cycle interaction energy, and the double-deletion energy. See Theory.

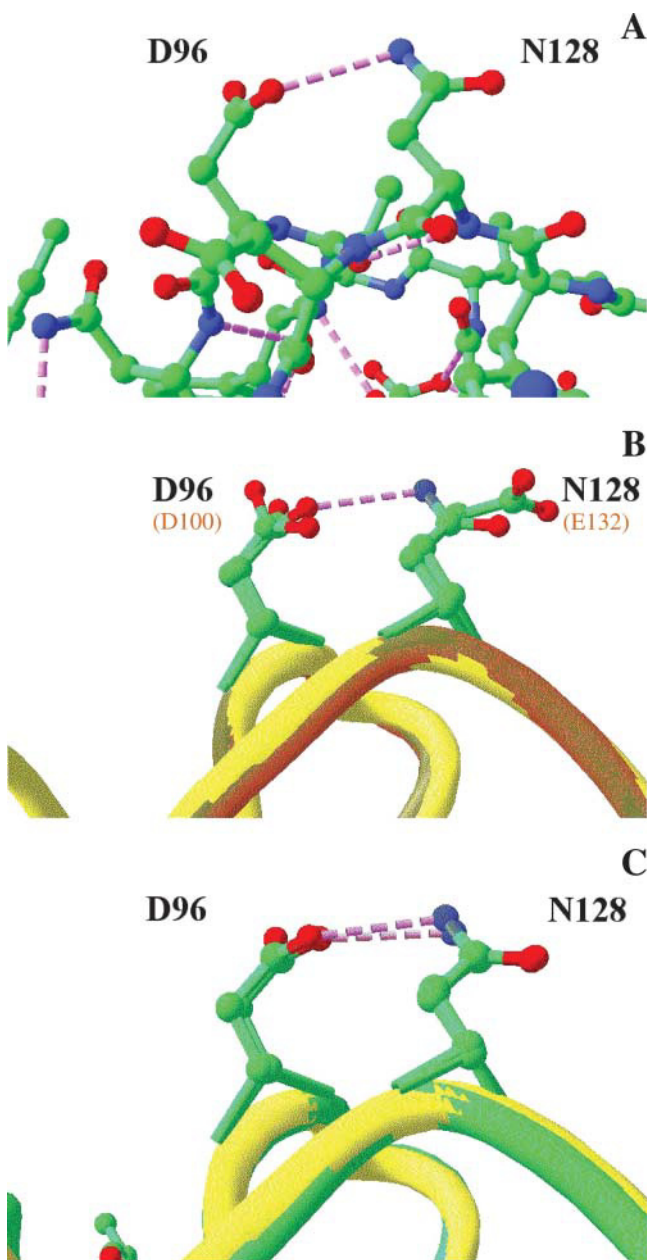


FIGURE 2 (A) Ball and stick representation of the *Anabaena* apoflavodoxin structure (1ftg) showing the hydrogen bonded residues D96 and N128. Hydrogen bonds in magenta. (B) Superposition of the apoflavodoxin from *Anabaena* and holo flavodoxin from *Chondrus crispus* (2fcr) showing the *Anabaena* hydrogen-bonded residues D96 and N128 and their structural equivalents, D100 and E132. The perfect conservation of the structure at the site of mutation in the *Chondrus crispus* protein, where the hydrogen bond is no longer possible, can be appreciated. (C) Superposition of the *Anabaena* apo and holo (1flv) flavodoxin structures showing the conservation of the hydrogen bond upon FMN cofactor binding.

1992). These data agree with those reported for tripeptides by Creamer and co-workers (1995, 1997), who suggest, however, that in longer peptides side-chain exposures are reduced to ~65% of their values in tripeptides.

The quantification of solvation energies (in cal mol<sup>-1</sup>) from changes in solvent area (in Å<sup>2</sup>) has been performed using the following relationship:

$$\Delta G_{Pw}(ij - 00)^{\text{apolar}} = 28.7(\pm 2.8)\Delta ASA(ij - 00)^{\text{apolar}}, \quad (7)$$

where the converting factor is the average ( $\pm$  SE) of 11 different factors proposed since 1991 (Sharp et al., 1991; Eriksson et al., 1992; Schiffer et al., 1992; Blaber et al., 1993; Pinker et al., 1993; Koehl and Delarue, 1994; Vajda et al., 1995; Eisenhaber, 1996; Weng et al., 1997). The very small change in polar area (see Table 2) has not been considered. According to different parameterizations (Vajda et al., 1994; Xie and Freire, 1994), its contribution to  $\Delta G_{Pw}(ij - 00)$  would be between 0.00 and 0.06 kcal mol<sup>-1</sup>.

## Mutagenesis, protein expression and purification, and spectroscopic characterization

PCR-mutagenesis of the *Anabaena* PCC 7119 flavodoxin gene was performed with the QuikChange kit and the mutations identified by sequencing. Expression of the gene (Fillat et al., 1991) was done in *Escherichia coli*. Purification and removal of the FMN prosthetic group was performed as described (Genzor et al., 1996a). Near-ultraviolet (UV) circular dichroism (CD) spectra (260–310 nm) of wild-type and mutant proteins were obtained with a 1-cm cuvette and 30-μM protein solutions in 50 mM MOPS, pH 7. Far-UV CD spectra (200–250 nm) were recorded with a 1-mm cuvette, at the same protein concentration in a 5-mM MOPS, pH 7, buffer containing 15 mM NaCl.

## Stability measurements

The conformational stability of the apoflavodoxin from *Anabaena* has been extensively characterized in our laboratory (Genzor et al., 1996a; Maldonado et al., 1998a,b, 2002; Fernandez-Recio et al., 1999; Irun et al., 2001a,b; Langdon et al., 2001; Lopez-Llano et al., 2004a,b) and its equilibrium urea denaturation has been shown to be two-state (Genzor et al., 1996a; Fernandez-Recio et al., 1999; Irun et al., 2001a,b; Langdon et al., 2001; Maldonado et al., 2002). The stability of wild-type and mutant apoflavodoxins has been measured by urea denaturation as described (Genzor et al., 1996a), but using a ratio of intensities (320/380 nm). Because  $m$  values are typically determined with large errors when urea unfolding curves of proteins are fitted using the linear extrapolation method (Santoro and Bolen, 1988), which is in contrast with the much greater reproducibility of denaturant concentrations of mid-denaturation, protein stability differences are most accurately determined using an average  $m$  value for the different proteins, although this practice is sometimes questioned (Yi et al., 2003). Based on previous work in our laboratory with wild-type and mutant apoflavodoxins, we have estimated (Fernandez-Recio et al., 1999) that the accuracy of stability differences between apoflavodoxin variants calculated using an average  $m$  slope is  $\sim \pm 0.06$  kcal mol<sup>-1</sup> (this applies to  $\Delta G_{dd}$ ; see Eq. 3), and that of stability differences between four variants  $\sim \pm 0.08$  kcal mol<sup>-1</sup> (this applies to  $\Delta G_{im}$ ). If, however, the individual  $m$  values obtained for each protein variant are used, much larger errors are obtained due to the intrinsic poor reproducibility of  $m$  values. In this work we report stability differences calculated using both an average  $m$  value and individual  $m$  values. The two sets of data are in qualitative agreement and point to the same conclusions. We consider the data obtained using an average  $m$  value to be more accurate. Another potential source of inaccuracy in protein stability determinations is batch-related protein stability differences. However, in the particular case of *Anabaena* apoflavodoxin, we have not observed over the years significant differences among different batches of the wild-type protein (not shown).

## Molecular dynamics

Molecular dynamics simulations of the apoflavodoxin wild-type structure (1ftg) and of the modeled double mutant were performed using the CHARMM (c27b2) package (Brooks et al., 1983). An initial step of minimization was applied to both structures, using several cycles of steepest

descent, conjugate gradient, and adopted-basis Newton-Raphson. Solvation of the systems was achieved by placing the protein structures inside a preequilibrated cubic box of TIP3P water molecules (Jorgensen et al., 1983). To reduce edge effects, periodic boundary conditions were applied, and the SHAKE algorithm (Ryckaert et al., 1977) was used to hold rigid the internal geometry of the water molecules, according to the Jorgensen description (Jorgensen et al., 1983). Long-range electrostatic interactions were modeled with the particle-mesh Ewald method (Essmann et al., 1995), using a 12.0-Å cutoff and a grid spacing of  $\sim 1.0$  Å. To achieve an appropriate neutralization of the system,  $\text{Na}^+$  counterions were iteratively placed. Initially, they were randomly positioned, avoiding overlaps with the protein and removing the water molecules located within a 2.5-Å radius of the ions introduced. Then, a short minimization was performed, keeping the protein fixed, to improve the solvation of the ions, and a 10-ps CPT dynamics was run (298 K, 1 atm) (Feller et al., 1995) to allow the solvation cage to expand to avoid internal strains.

Langevin dynamics were used to heat the system and to produce trajectories in the canonical ensemble (Paterlini and Ferguson, 1998; Krivov et al., 2002). The use of Langevin dynamics is cpu time-consuming (as compared to using other traditional algorithms, such as nose-Hoover) but is advantageous in that it guarantees a better representation of the ensemble. Since the aim was the determination of equilibrium properties, the choice of the friction coefficient should not affect the results (provided the fluctuation-dissipation relation is fulfilled), although it can influence the dynamics (see below). A leapfrog Verlet integrator with a time step of 1 fs was used. The friction coefficient  $\gamma$  in the Langevin equations was set to  $64 \text{ ps}^{-1}$  for solvent molecules (Smith et al., 1993) and to  $1.5 \text{ ps}^{-1}$  for protein atoms. This choice allows a fast equilibration of the solvent and speeds up the dynamical processes inside the protein (Zagrovic and Pande, 2003). In addition, it eliminates the spoiling high-frequency modes in the solvent that do not concern our study. The simulations began with a 50-ps, slow, progressive heating to the working temperature (298 K), followed by a production run of 4.5 ns.

## RESULTS AND DISCUSSION

### Identification of suitable hydrogen-bonded “isolated pairs”

Hydrogen bonding is the most important single interaction in proteins, governing protein architecture and function (Jeffrey and Saenger, 1991; Branden and Tooze, 1998; Desiraju and Steiner, 1999; Lesk, 2000). For this reason we have chosen to implement the first application of the double-deletion analysis outlined in the Theory (see above) to quantify the incremental binding energy of a hydrogen bond. To find suitable candidates, we examined several small model proteins used for stability studies: flavodoxin (1ftg), ferredoxin (1fxa), lysozyme (193l), barstar (1a19), ferredoxin-NADP+ reductase (1que), pepsin (4pep), CheY (1ehc), and cytochrome *c* (1crc). For each protein, we have identified all the side chain/side chain hydrogen bonds (18, 7, 12, 7, 28, 28, 2, and 2, respectively) and selected the side chains that only form one hydrogen bond (4, 3, 1, 5, 8, 5, 0, and 0, respectively). Then we calculated their overall solvent exposures. For the pairs with at least one residue with solvent exposure  $>50\%$  (1, 1, 0, 2, 1, 1, 0, and 0, respectively), atom exposures were calculated with MOLMOL (Koradi et al., 1996), and pairs where the exposure of the atoms that would be removed upon mutation to Ala was  $>50\%$  were selected. Only two pairs remained, one in ferredoxin NADP<sup>+</sup> oxidoreductase and one in

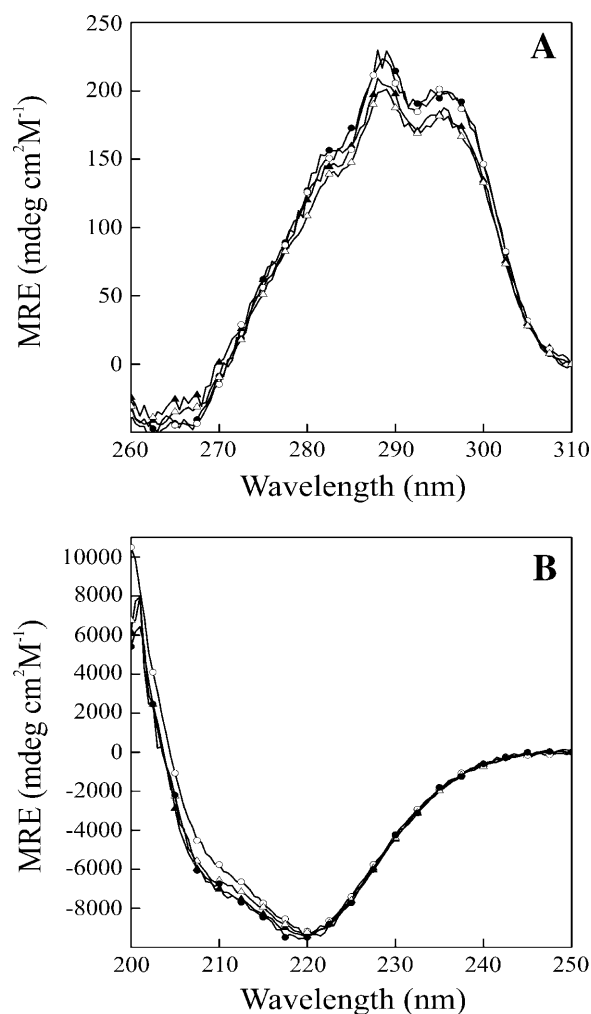


FIGURE 3 Near-UV (A) and far-UV (B) circular dichroism spectra of wild-type (solid circles), D96A (solid triangles), N128A (open circles), and D96A/N128A (open triangles) mutants. Spectra obtained at  $25.0 \pm 0.1^\circ\text{C}$  in MOPS, 50 mM, pH 7.0.

flavodoxin. Of these, only the flavodoxin pair satisfied the “isolation” requisite of double-deletion analysis, that the residues involved in the interaction analyzed do not make contact with any other residue in the protein beyond their  $\text{C}_\beta$ . The pair is formed between the D96 and N128 side chains of the apoflavodoxin from *Anabaena* PCC 7119 (Genzor et al., 1996b) (Fig. 2 A). The pair is also present in the holo form of the protein (Rao et al., 1992) (Fig. 2 C) and in three mutant flavodoxins, one in the apo form and two in the holo form, previously reported in our laboratory (Lostao et al., 2000, 2003). The “isolation” requisite itself makes it unlikely that the combined substitution of the two residues by alanines cause any significant structural rearrangement in the protein because no additional residue loses or gains interactions upon mutation.

As outlined above, the identification of suitable candidates has been carried out by visual inspection. The fact that only 1

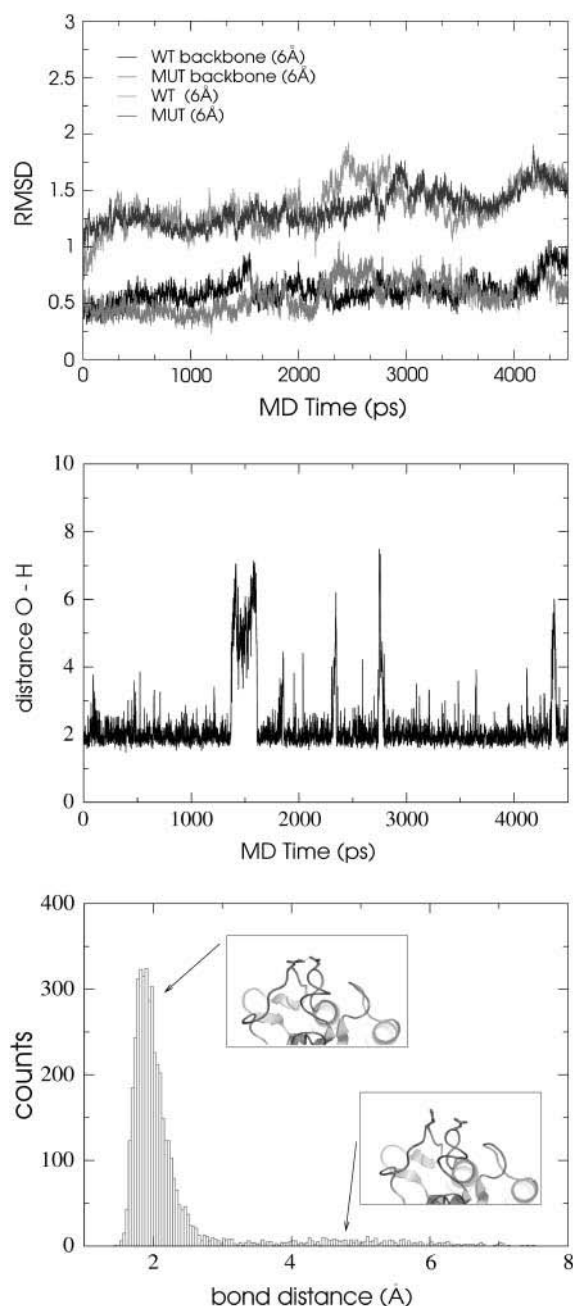


FIGURE 4 Molecular dynamics simulation of wild-type *Anabaena* apoflavodoxin and of the D96A/N128A double mutant. (Top) RMSD of the overall structures (upper traces) and of the atoms within a 6-Å radius of the carboxyl and carboxamide groups removed upon mutation (lower traces). The initial raising of the RMSD traces corresponds to the initial heating to 298 K of the reference minimized structures. (Middle) Evolution of the hydrogen bond H...O distance along the simulation. The shorter of the distances between the side-chain NH hydrogen of N128 and any of the side-chain O atoms of D96 is represented. Hydrogen bond breaking and reforming events are evidenced as peaks from the equilibrium distance baseline. (Bottom) Statistics of hydrogen bond distances during a 4.5-ns simulation of wild-type apoflavodoxin. Counts of distances sampled every picosecond are shown. The main peak represents the conformations that retain the hydrogen bond (see inset) whereas the flatter, wider peak represents conformations with a broken hydrogen bond. A  $2.5 \pm 0.1$ -Å cutoff has been used to calculate the free energy of hydrogen bond formation from the folded state.

in 104 side chain/side chain hydrogen bonds analyzed has turned out to be appropriate suggests that an automated analysis of the Protein Data Bank would be helpful to find candidates to analyze other side-chain interactions and other types of hydrogen bonds. Work is in progress in this direction (F. Pazos, personal communication).

### Integrity of the mutant proteins

The overall integrity of the D96A and N128A single mutants and of the D96A/N128A double mutant has been initially assessed by comparing the fluorescence, far-UV CD, and near-UV CD spectra to those of wild-type. The fluorescence emission (not shown) and the far- and near-UV CD spectra (Fig. 3) of the three mutant proteins are almost identical to those of the wild-type protein. In addition to maintaining the overall fold, double-deletion analysis requires, as double-mutant cycle analysis does, that the local protein structure is not altered by the mutations introduced. Although the x-ray structures of the mutants are not available (they have failed to crystallize) there is firm crystallographic evidence, coming from the structure of a highly related flavodoxin, that the implemented mutations to alanine do not cause local perturbations. As shown in Fig. 2 B, superimposed to the structure of the wild-type *Anabaena* apoflavodoxin (Genzor et al., 1996b), the flavodoxin from *Chondrus crispus* (2fcr) (Fukuyama et al., 1992) contains an aspartic residue (D100) that is structurally equivalent to the D96 in *Anabaena* apoflavodoxin. However, at the position equivalent to N128, the *C. Crispus* flavodoxin displays a glutamate (E132), and therefore hydrogen bonding with its D100 neighbor is not possible. In this respect, and given that E132 and D100 should repel each other due to their charges, the *C. Crispus* flavodoxin exemplifies the structural consequences of a mutation that, potentially, is much more disruptive than the D96A and N128A mutations implemented here. Yet, as Fig. 2 B shows, the *C. Crispus* flavodoxin accommodates the mutation by simply rotating the glutamate side chain so that the carboxyl group points to the solvent. The  $C_{\beta}$  of the *C. Crispus* E132 is at the same position as that of *Anabaena* N128, and remarkably D100 remains unmoved from the position of the structurally equivalent *Anabaena* D96. If the mutation of one of the residues involved in the pair leaves the other unchanged and unpaired (beyond the  $C_{\beta}$ ) it is difficult to envision that the mutations to alanine may cause any local alteration. Based on this fact, we have modeled the structure of the double mutant D96A/N128A by simply mutating in silico the wild-type residues to alanine.

A more direct indication that the double mutation is well tolerated by the protein without significant rearrangements has been obtained from Langevin molecular dynamics simulations of wild-type and double-mutant apoflavodoxin. The simulations have been run for 4.5 ns, which has proved long enough to reach an equilibrium configuration. Fig. 4 (top) shows, for the wild-type and the D96A/N128A double



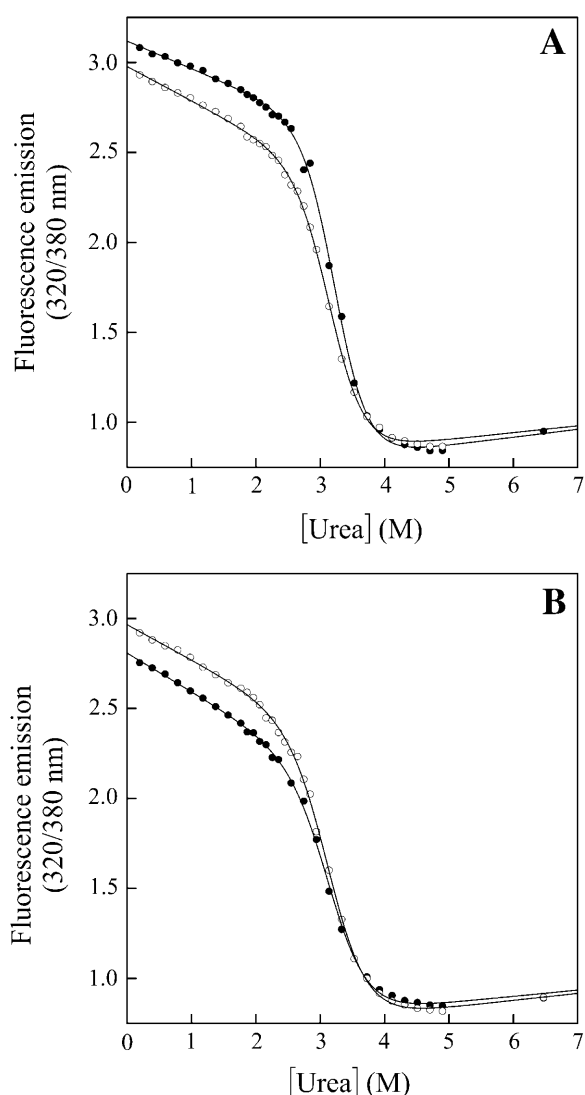


FIGURE 5 Urea denaturation curves of wild-type (solid circles) and D96A/N128A (open circles) apoflavodoxin double mutant (A) and of the D96A (solid circles) and N128A (open circles) single mutants (B). Data were recorded at  $25.0 \pm 0.1^\circ\text{C}$  in MOPS, 50 mM, pH 7.0, with 0.5 M NaCl, and fitted to a two-state equation (Santoro and Bolen, 1988).

mutant, the RMSDs as a function of simulation time (relative to the starting, previously minimized structures) of the residues located within a 6-Å radius of any of the atoms of the D96 carboxyl and N128 carboxamide groups. It is clear that the RMS values hardly change with time and that no noticeable differences can be observed between the wild-type and the double-mutant RMS traces. In fact, when the side chains are included together with the backbone atoms in the RMS calculations, the wild-type trace is somewhat less stable than that of the double mutant, as a consequence of the dynamics of the hydrogen-bonded D96 and N128 side chains. The RMS traces of the backbone atoms are very stable for the two proteins, which can hardly be distinguished. This local stability of the loops bearing the hydrogen-bonded residues

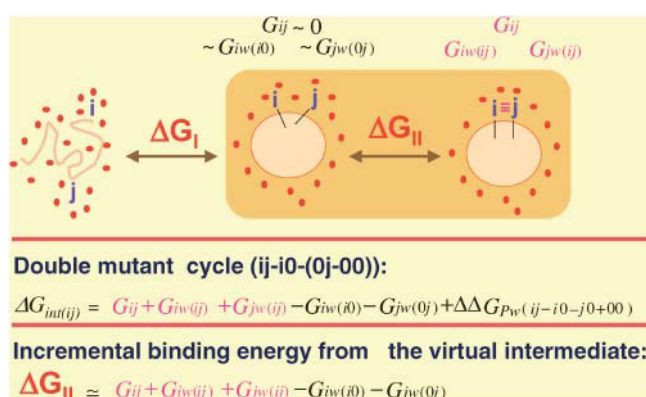


FIGURE 6 Scheme depicting the folding of a protein as divided into two steps. In the first one, with  $\Delta G_I$ , the protein gets folded to a virtual intermediate (essentially folded) where the  $i$  and  $j$  residues do not yet establish an interaction. Here, the solvations of the  $i$  and  $j$  residues are equivalent to those in the folded state of the  $0j$  and  $i0$  single mutants, and the interaction between them is considered close to zero. In the second step, with  $\Delta G_{II}$ , the two residues establish an interaction. The equations show the relationship between  $\Delta G_{II}$ , which represents the incremental binding energy from the virtual, folded, intermediate, and interaction energy, calculated from double-mutant cycle analysis.

correlates with their low B-factor in the x-ray structure (Genzor et al., 1996b). In fact, our simulations reveal larger departures from the x-ray structure in distant regions of the protein that display high B-factors in the crystal (not shown).

### Incremental binding energy of a hydrogen bond

To calculate the incremental binding energy of the mutated bond, the stability of wild-type and D96A/N128A apoflavodoxins has been measured by urea denaturation (Fig. 5 A) as described (Genzor et al., 1996a). To avoid analysis complications arising from long-range electrostatic interactions, we have performed all measurements in the presence of 0.5 M NaCl. Previous work (Maldonado et al., 2002) has shown that this salt concentration effectively masks medium and long-range electrostatic interactions in apoflavodoxin. As indicated by its urea concentrations of mid denaturation (Table 1), the double mutant lacking the hydrogen bond is slightly less stable than the wild-type protein by  $0.21 \pm 0.06 \text{ kcal mol}^{-1}$  (or  $0.29 \pm 0.23 \text{ kcal mol}^{-1}$  if less accurate individual  $m$  values are used instead of an average  $m$  value; see Materials and Methods). This difference in stability (wild-type minus double mutant), that we have termed double-deletion energy, equals the contribution of the two hydrogen bond-forming groups to protein stability (relative to having two alanines) plus a solvation term (see Eqs. 3–5). As indicated in the Theory (see above), the solvation term in Eq. 5 concerns essentially apolar atoms, its sign is known, and its actual value can be calculated from empirical equations with reasonable accuracy. From the differential solvation in the folded state ( $-18.7 \text{ Å}^2$  corresponding to the apolar atoms neighboring the

**TABLE 1** Stability of wild-type and mutant proteins

Protein	$m^*$ (kcal mol <sup>-1</sup> M <sup>-1</sup> )	$U_{1/2}^\dagger$ (M)	$\Delta G^\ddagger$ (kcal mol <sup>-1</sup> )	$\Delta G_{av}^\S$ (kcal mol <sup>-1</sup> )
WT	2.27 ± 0.06	3.270 ± 0.022	7.43 ± 0.16	7.10 ± 0.05
D96A/N128A	2.25 ± 0.05	3.175 ± 0.001	7.14 ± 0.16	6.89 ± 0.03
D96A	2.15 ± 0.21	3.181 ± 0.003	6.85 ± 0.66	6.90 ± 0.01
N128A	2.02 ± 0.06	3.176 ± 0.010	6.41 ± 0.23	6.90 ± 0.02

Urea denaturation was performed at 25.0°C, in 50 mM MOPS, pH 7.0, with 0.5 M NaCl.

\*Slope of a linear plot of  $\Delta G$  versus urea concentration. Mean of two determinations ± SE.

†Urea concentration of mid-denaturation. Mean of two determinations ± SE.

‡Standard free energy of unfolding calculated for each protein as  $m_i$  times  $U_{1/2i}$ . Mean of two determinations ± SE.

§Standard free energy of unfolding calculated for each protein as  $m_{av}$  times  $U_{1/2i}$ , where  $m_{av} = 2.17$ , is the average slope of all determinations. Mean of two determinations ± SE. We consider this data to be more accurate.

96 and 128 side chains in the wild-type and in the modeled double-mutant proteins (Table 2) plus  $-62.5 \text{ \AA}^2$ , corresponding to the D96 and N128  $C_\beta$  atoms (Table 3)), together with the differential solvation in the unfolded state ( $-58.4 \text{ \AA}^2$ , corresponding to the D96 and N128  $C_\beta$  atoms (Table 3)), the overall apolar area change related to the solvation term in Eq. 5 amounts to  $-22.8 \text{ \AA}^2$ . We thus calculate (Eq. 7) this solvation term at  $-0.65 \text{ kcal mol}^{-1}$ . On the other hand, a more accurate calculation of this term can be performed if, instead of considering the solvent exposures in the wild-type crystal structure and in the double-mutant model, averages of the exposure to solvent of the two proteins during the 4.5-ns molecular dynamics trajectories are used (see Table 2). From the averaged exposures of each protein, we calculate that the overall differential area (wild-type minus double mutant) exposed to solvent is  $-14.5 \text{ \AA}^2$ , rather than  $-22.8 \text{ \AA}^2$ , which sets the solvation term in Eq. 5 at  $-0.42 \text{ kcal mol}^{-1}$ . An estimation of the error associated to the solvation term can be obtained if an average ( $\pm$  SE) of the areas calculated by the two methods is considered ( $18.7 \pm 4.2 \text{ \AA}^2$ ) and the uncertainty in the multiplying constant on Eq. 7 is taken into account. For the constant ( $28.7 \text{ cal mol}^{-1} \text{ \AA}^{-2}$ , see Materials

and Methods) we calculate a standard error of  $\pm 2.8 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$  from the values of 11 different factors proposed in recent years (see Materials and Methods). In this way, the solvation term is estimated at  $-0.54 \pm 0.13 \text{ kcal mol}^{-1}$ .

The net contribution of the hydrogen-bonding carboxylate and carboxamide groups of D96 and N128 to protein stability can now be calculated (see Eq. 5) by combining the experimentally determined double-deletion energy and the solvation term, and it turns out to be of  $+0.33 \pm 0.14 \text{ kcal mol}^{-1}$  (using nonaveraged  $m$  slopes a less accurate quantification can be offered at  $+0.25 \pm 0.26 \text{ kcal mol}^{-1}$ ). The contribution of the two hydrogen-bonding groups is thus, in principle, small and destabilizing. However, we would like to point out that the destabilizing contribution of the D96/N128 bonding groups to protein stability could be larger. This is so because we have used in our calculations solvent-exposed areas in the denatured state that are based on the exposures observed in model tripeptides, and therefore could be unrealistically large. As more accurate determinations of solvent exposures in denatured states are being performed (by averaging states populated in molecular dynamics simulations and by considering longer model

**TABLE 2** Incremented surface area per atom type in the folded state

Protein	Amino acid residues with solvent accessible surface changes relative to wild-type	New area exposed to solvent in the folded state by atom type ( $\text{\AA}^2$ )				
		C(C)	C(O)	C(N)	O	N
D96A*	A95, A96, N97, N128, N129	33.5	0.0	0.0	0.0	1.4
N128A*	A95, F127, A128, D129	29.9	3.5	3.3	2.7	0.0
D96A/N128A*	A95, A96, N97, F127, A128, N129	74.4	3.5	3.3	2.6	1.5
		11.9 <sup>‡</sup>	3.5 <sup>‡</sup>	3.3 <sup>‡</sup>	2.6 <sup>‡</sup>	1.5 <sup>‡</sup>
D96A/N128A <sup>†</sup>	Y94, A95, A96, N97, D126, F127, A128, N129	65.7	2.6	1.0	0.8	1.4
		6.8 <sup>‡</sup>	2.6 <sup>‡</sup>	1.0 <sup>‡</sup>	0.8 <sup>‡</sup>	1.4 <sup>‡</sup>

Surface area values refer only to the folded state. They exclude the  $i$  and  $j$  mutated residues and are pertinent to estimate the magnitude of  $\Delta\Delta G_{pw}(ij - i0 - 0j + 00)$  in Eq. 2 because in the double-mutant cycle all terms concerning solvations in the denatured state cancel out.

\*Calculated with Naccess 2.1.1 (Hubbard and Thornton, 1993) from the x-ray structure of the wild-type protein (1FTG) and the models of the single and double mutants generated by replacing in silico the Asp and/or Asn side chains by Ala ones.

†Calculated with Naccess 2.1.1 (Hubbard and Thornton, 1993) from averages of the wild-type and double-mutant structures generated along 4.5-ns Langevin molecular dynamics trajectories.

‡Excluding the  $\beta$ -carbons of residues at 96 and 128. The data concerning carbon atoms (C(C), C(O), and C(N)) are used, together with data in Table 3, for the calculation of the solvation term in Eq. 6.

**TABLE 3** Exposed area of  $C_\beta$  atoms of residues 96 and 128 in the folded and unfolded states ( $\text{\AA}^2$ )

Residue	Folded state	Unfolded state*
D96	18.9	36.3
N128	8.1	38.3
A96	48.4	66.5
A128	41.1	66.5

These areas allow calculation of the contribution of the  $\beta$ -carbons to  $\Delta G_{PW}$  ( $ij - 00$ ) in Eq. 3. To calculate this term, the atoms that are removed by mutation are not pertinent but the solvation of their  $\beta$ -carbons in the folded and unfolded states has to be taken into account because, unlike in double-mutant cycle calculations, it does not cancel out.

\*Unfolded state data from Zielenkiewicz and Saenger (1992).

peptides) the reported values of solvent exposures tend to shrink. We note, in this respect, that if the actual exposures of the  $C_\beta$  in the denatured state were smaller than those used in our calculations, and reported in Table 3, by the amount suggested by Creamer and co-workers (1995) for longer peptides, the solvation term in Eq. 5 would still be negative but significantly larger than the estimated  $-0.54 \pm 0.13 \text{ kcal mol}^{-1}$  (actually, it would amount to  $\sim -1.3 \text{ kcal mol}^{-1}$ ). In this more realistic scenario, the destabilizing contribution of the D96/N128 hydrogen bond would be of  $\sim +1.1 \text{ kcal mol}^{-1}$ .

The reason the wild-type protein is slightly more stable than the double alanine mutant is that a significant stabilization is obtained from an increased hydrophobic effect arising from the shade cast by the carboxylate and carboxamide groups of D96 and N128 on neighboring apolar groups, not directly in contact. This effect does not stabilize the hydrogen bond itself because it would arise to a similar extent in the wild-type protein if the hydrogen bond were not formed.

In agreement with our finding of a destabilizing contribution of hydrogen-bonding groups to protein stability, there is recent work by several laboratories that also points to a destabilizing contribution of hydrogen-bonding groups in proteins (Ma and Nussinov, 2000; Guerois et al., 2002). The same view is represented in detailed calculation (Ben-Tal et al., 1997) and measurement (reviewed in Ben-Tal et al., 1997) of the dimerization energy of model compounds. The contrasting view supporting a stabilizing contribution of hydrogen-bonding groups to protein stability based on the analysis of single-deletion experiments has been reviewed by Myers and Pace (1996). In our view, single-deletion experiments are unlikely to clarify so subtle a matter, among other things because, as is acknowledged by Myers and Pace, “we are left to guess at the hydrogen bonding status of the remaining partner”.

### Why a destabilizing interaction is established

It may seem paradoxical that a destabilizing interaction like this hydrogen bond is present at all in the native structure. The paradox, however, can be easily explained in a quanti-

tative manner by conceptually dividing the folding of the protein into two processes (Fig. 6). First, the protein folds to a virtual intermediate where residues  $i$  and  $j$  are close in space but do not yet interact with each other. In the second step, the  $i$  and  $j$  side chains approach and form a bond. It is the free energy difference of the second step ( $\Delta G_{II}$ ) that governs the stability of the hydrogen bond in the context of the native structure and the fact that the hydrogen bond is observed in the crystal structure merely suggests that  $\Delta G_{II}$  should be negative. To test this interpretation we have quantitated  $\Delta G_{II}$  both from experiment and from simulation.

The experimental approach relies on the similarity of the solvation energies of  $i$  and  $j$  residues in the virtual intermediate depicted in Fig. 6 and those displayed by the same residues in the single mutants ( $i0$  and  $0j$ ) present in the double-mutant cycle scheme (Fig. 1). Although identifying one set of solvation energies with the other is a simplification (because it is likely that the  $i$  and  $j$  residues would be more desolvated in the virtual intermediate than in the single mutants) it provides a useful way to estimate  $\Delta G_{II}$  from classical double-mutant cycle analysis. Assuming that the solvation energies of the  $i$  and  $j$  residues in the single mutants approximate those in the virtual intermediate, the interaction energy measured by double-mutant cycle represents the binding energy of the  $i$  and  $j$  residues interacting from the close-to-native intermediate state ( $\Delta G_{II}$  in Fig. 6) plus a solvation term:  $\Delta \Delta G_{PW} (ij - i0 - j0 - 00)$  that essentially refers to apolar surface and can be estimated independently. We have thus resorted to double-mutant analysis, prepared the two related single apoflavodoxin mutants, and determined their stability by urea denaturation (Fig. 5B). The double-mutant cycle-derived interaction energy is of  $-0.19 \pm 0.06 \text{ kcal mol}^{-1}$  (Table 1; or, less accurately,  $-1.3 \pm 0.7 \text{ kcal mol}^{-1}$ , if individual instead of averaged  $m$  values are used). Since the solvation term amounts in this case to  $+0.32 \pm 0.03 \text{ kcal mol}^{-1}$  ( $11.0 \text{ \AA}^2$ , Table 2),  $\Delta G_{II}$  is calculated at  $-0.51 \pm 0.07 \text{ kcal mol}^{-1}$ ; stabilizing (a larger, but less accurate value of  $-1.6 \pm 0.7 \text{ kcal mol}^{-1}$  would be calculated from individual  $m$  slopes).

In fact, due to the expected greater desolvation of the side chains in the virtual intermediate than in the single mutants, and due to the smaller entropy change of bond formation in the intermediate than in the unfolded state, the calculated value of  $\Delta G_{II} = -0.51 \text{ kcal mol}^{-1}$  underestimates the binding energy of the hydrogen bond within the folded structure. We believe a more accurate determination of  $\Delta G_{II}$  can be achieved by careful analysis of molecular dynamics simulation of the wild-type protein. To that end, we have specifically monitored the dynamics of the D96/N128 bond. The bond can be established by either of the OD1 and OD2 oxygen atoms of the D96 side chain, and, indeed, the alternative involvement in the bond of the two oxygens is observed (not shown). To describe the energetics of a carboxylate/carboxamide hydrogen bond, the two configurations of the bond should not be differentiated.



Monitoring the distances between the D96/N128 residues during the 4.5-ns trajectory reveals that, in addition to the swapping of oxygens, the bond breaks and reforms many times during the sampled trajectory. In some cases, the Asp side chain is observed to bend into the solvent where it establishes new bridges with bulk water molecules. To illustrate the dynamics of the bond, the shortest of the distances between the N128 side chain H atom and any of the D96 OD1 and OD2 atoms is shown in Fig. 4 (*middle*) as a function of time. Some clear breaking events are evident in the trajectory. The fluctuation of O–H distances around the equilibrium position is best observed in the histogram shown in Fig. 4 (*bottom*), where two regions can be distinguished: a narrow peak centered around the equilibrium bond distance (1.8 Å) and a very broad distribution from 2.5 Å to ~8 Å corresponding to the unbound configuration. This is consistent with local two-state behavior and allows quantification of the binding energy of bond formation from the folded state. Using a typical 2.5-Å threshold as the bond breaking O–H distance, we calculate that the hydrogen bond remains formed 85% of the time, which reflects a binding energy of  $-1.0 \pm 0.1 \text{ kcal mol}^{-1}$  (allowing for a 0.1-Å error in the threshold). As was anticipated above, this value of  $\Delta G_{\text{II}}$  is larger than the one calculated from the double-mutant cycle approximation ( $-0.51 \pm 0.07 \text{ kcal mol}^{-1}$ ) and we consider it to be more accurate. Whatever the exact value of  $\Delta G_{\text{II}}$ , both the experimental analysis and the molecular dynamics simulation clearly indicate that forming the hydrogen bond from the compact, partly desolvated, close-to-native state does indeed significantly stabilize the protein. The paradox is thus solved as follows: adding to the apoflavodoxin polypeptide two hydrogen bonding groups (the carboxyl and carboxamide in D96 and N128) that form a hydrogen bond in the native state destabilizes the native protein, and yet the two groups are forced to interact and form the bond because, in the context of the folded protein, bond formation becomes favorable. Why this is so in this particular case is open for interpretation and it is clear from the molecular dynamics simulations that the hydrogen bond can be broken by side-chain rotations. We point out that two potential contributions to the stability of the hydrogen bond in the context of the native structure could be a lower effective concentration of water felt by the interacting residues in the folded state (as compared to the unfolded state) and a reduced entropy change of binding in the native state due to their proximity and to the fact that the side chain of N128 is relatively constrained. Whatever the specific cause, which is difficult to precise, it seems that frustration manifesting in protein folding may similarly drive the formation of other non-stabilizing or even destabilizing interactions that will thus be present in native proteins. Recent work on a salt bridge also points to this direction (Luisi et al., 2003). Thus, statistical potentials derived from contact frequencies in proteins do not necessarily reflect the energetics of pairwise interactions, if the denatured state is taken as the reference.

## CONCLUDING REMARKS

It is tempting to extrapolate the finding that the hydrogen bond analyzed here displays a positive (destabilizing) incremental binding energy to conclude that hydrogen bonds destabilize proteins or at least do not stabilize them. Indeed, our finding agrees well with the fact that, as far as we know, no claims of protein stabilization have been made based on engineering pairs of polar groups to form new hydrogen bonds, which suggests that perhaps proteins cannot be stabilized in this way. It is clear, however, that more hydrogen bonds must be studied to establish whether the picture offered by this Asp/Asn bond can be generalized. This is so because the differential solvation energies ( $G_{\text{iw}}$ ,  $G_{\text{ju}}$ ) will vary with solvent exposure in the native state, and because their values for the various polar groups appearing in proteins are different (Jeffrey and Saenger, 1991), as are the intrinsic strengths of the bonds they establish ( $G_{\text{ij}}$ ). Therefore, a surface Asp/Asn hydrogen bond may be significantly different from a surface bond involving other residue types or from a buried Asp/Asn bond. In terms of overall protein energetics, it would be particularly interesting to see what the trend is for carefully chosen buried hydrogen bonding groups, as they could report on the contribution of the ubiquitous main-chain hydrogen bonds to protein stability. A recent study suggests, from isotope effect measurements, a different contribution to protein stability for main-chain hydrogen bonds located in  $\alpha$ -helices and in  $\beta$ -sheets (Shi et al., 2002), which stresses the subtlety of the balance.

The more important conclusion of this work is that the double-deletion method offers an experimental way to quantify precisely the contribution of side-chain interactions to protein stability. However, it requires a very demanding selection of suitable interacting pairs that makes it unlikely to find, in a particular model protein, more than one useful pair to investigate a given interaction. The method therefore has both advantages and disadvantages compared to double-mutant cycle analysis.

We thank P. J. Martínez for computational support.

This work was supported by the Spanish Ministerio de Ciencia y Tecnología (BMC 2001-252, MCYT-BFM2002-00113) and Diputación General de Aragón (P120/2001). L.A.C. and S.C.-L. were supported by Formación del Personal Investigador fellowships and J.L.-L. by a Basque government fellowship.

## REFERENCES

- Ben-Tal, N., D. Sitkoff, I. A. Topol, A.-S. Yang, S. Buro, and B. Ong. 1997. Free energy of amide hydrogen bond formation in vacuum, in water, and in liquid alkane solution. *J. Phys. Chem. B.* 101:450–457.
- Blaber, M., J. D. Lindstrom, N. Gassner, J. Xu, D. W. Heinz, and B. W. Matthews. 1993. Energetic cost and structural consequences of burying a hydroxyl group within the core of a protein determined from Ala→Ser and Val→Thr substitutions in T4 lysozyme. *Biochemistry.* 32:11363–11373.

- Branden, C., and J. Tooze. 1998. *Introduction to Protein Structure*. Garland Publishing, New York.
- Brooks, B. R., R. E. Bruccoleri, B. D. Olafson, D. J. Status, S. Swaminathan, and M. J. Karplus. 1983. CHARMM: a program for macromolecular energy minimization and dynamics calculations. *J. Comput. Chem.* 4:187–217.
- Carter, P. J., G. Winter, A. J. Wilkinson, and A. R. Fersht. 1984. The use of double mutants to detect structural changes in the active site of the tyrosyl-tRNA synthetase (*Bacillus stearothermophilus*). *Cell*. 38:835–840.
- Creamer, T. P., R. Srinivasan, and G. D. Rose. 1995. Modeling unfolded states of peptides and proteins. *Biochemistry*. 34:16245–16250.
- Creamer, T. P., R. Srinivasan, and G. D. Rose. 1997. Modeling unfolded states of proteins and peptides. II. Backbone solvent accessibility. *Biochemistry*. 36:2832–2835.
- Desiraju, G. R., and T. Steiner. 1999. *The Weak Hydrogen Bond*. Oxford University Press, Oxford, UK.
- Eisenhaber, F. 1996. Hydrophobic regions on protein surfaces. Derivation of the solvation energy from their area distribution in crystallographic protein structures. *Protein Sci.* 5:1676–1686.
- Eriksson, A. E., W. A. Baase, X. J. Zhang, D. W. Heinz, M. Blaber, E. P. Baldwin, and B. W. Matthews. 1992. Response of a protein structure to cavity-creating mutations and its relation to the hydrophobic effect. *Science*. 255:178–183.
- Essmann, U., L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen. 1995. A smooth particle mesh Ewald method. *J. Chem. Phys.* 103:8577–8593.
- Feller, S. E., Y. H. Zhang, R. W. Pastor, and B. R. Brooks. 1995. Constant-pressure molecular-dynamics simulation: the Langevin piston method. *J. Chem. Phys.* 103:4613–4621.
- Fernandez-Recio, J., A. Romero, and J. Sancho. 1999. Energetics of a hydrogen bond (charged and neutral) and of a cation/ $\pi$  interaction in apoflavodoxin. *J. Mol. Biol.* 290:319–331.
- Fersht, A. R. 1987. The hydrogen bond in molecular recognition. *Trends Biochem. Sci.* 12:301–304.
- Fersht, A. R., A. Matouschek, and L. Serrano. 1992. The folding of an enzyme. I. Theory of protein engineering analysis of stability and pathway of protein folding. *J. Mol. Biol.* 224:771–782.
- Fersht, A. R., and L. Serrano. 1993. Principles of protein stability derived from protein engineering experiments. *Curr. Opin. Struct. Biol.* 3:75–83.
- Fillat, M. F., W. E. Borrias, and P. J. Weisbeek. 1991. Isolation and overexpression in *Escherichia coli* of the flavodoxin gene from *Anabaena PCC 7119*. *Biochem. J.* 280:187–191.
- Fukuyama, K., H. Matsubara, and L. J. Rogers. 1992. Crystal structure of oxidized flavodoxin from a red alga *Chondrus crispus* refined at 1.8 Å resolution. Description of the flavin mononucleotide binding site. *J. Mol. Biol.* 225:775–789.
- Genzor, C. G., A. Beldarrain, C. Gomez-Moreno, J. L. Lopez-Lacomba, M. Cortijo, and J. Sancho. 1996a. Conformational stability of apoflavodoxin. *Protein Sci.* 5:1376–1388.
- Genzor, C. G., A. Perales-Alcon, J. Sancho, and A. Romero. 1996b. Closure of a tyrosine/tryptophan aromatic gate leads to a compact fold in apoflavodoxin. *Nat. Struct. Biol.* 3:329–332.
- Gueriois, R., J. E. Nielsen, and L. Serrano. 2002. Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. *J. Mol. Biol.* 320:369–387.
- Honig, B., and A. S. Yang. 1995. Free energy balance in protein folding. *Adv. Protein Chem.* 46:27–58.
- Horovitz, A., L. Serrano, B. Avron, M. Bycroft, and A. R. Fersht. 1990. Strength and cooperativity of contributions of surface salt bridges to protein stability. *J. Mol. Biol.* 216:1031–1044.
- Hubbard, S. J., and J. M. Thornton. 1993. NACCESS, Computer Program, Department of Biochemistry and Molecular Biology, University College, London.
- Irun, M. P., M. M. Garcia-Mira, J. M. Sanchez-Ruiz, and J. Sancho. 2001a. Native hydrogen bonds in a molten globule: the apoflavodoxin thermal intermediate. *J. Mol. Biol.* 306:877–888.
- Irun, M. P., S. Maldonado, and J. Sancho. 2001b. Stabilization of apoflavodoxin by replacing hydrogen-bonded charged Asp or Glu residues by the neutral isosteric Asn or Gln. *Protein Eng.* 14:173–181.
- Jeffrey, G., and W. Saenger. 1991. *Hydrogen Bonding in Biological Structures*. Springer-Verlag, Berlin.
- Jorgensen, W. L., J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. 1983. Comparison of simple potential functions for simulating liquid water. *Chem. Phys.* 79:926–935.
- Koehl, P., and M. Delarue. 1994. Polar and nonpolar atomic environments in the protein core: implications for folding and binding. *Proteins*. 20:264–278.
- Koradi, R., M. Billeter, and K. Wüthrich. 1996. MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.* 14:51–55.
- Krivov, S. V., S. F. Chekmarev, and M. Karplus. 2002. Potential energy surfaces and conformational transitions in biomolecules: a successive confinement approach applied to a solvated tetrapeptide. *Phys. Rev. Lett.* 88:38101–38105.
- Langdon, G. M., M. A. Jimenez, C. G. Genzor, S. Maldonado, J. Sancho, and M. Rico. 2001. *Anabaena* apoflavodoxin hydrogen exchange: on the stable exchange core of the  $\alpha/\beta$ (21345) flavodoxin-like family. *Proteins*. 43:476–488.
- Lee, B., and F. M. Richards. 1971. The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* 55:379–400.
- Lesk, A. M. 2000. *Introduction to Protein Architecture*. Oxford University Press, Oxford, UK.
- Lopez-Llano, J., S. Maldonado, M. Bueno, A. Lostao, M. A. Jimenez, M. P. Lillo, and J. Sancho. 2004a. The long and short flavodoxins. I: Role of the differentiating loop in apoflavodoxin structure and FMN binding. *J. Biol. Chem.* In press.
- Lopez-Llano, J., S. Maldonado, S. Jain, A. Lostao, R. Godoy-Ruiz, J. M. Sanchez-Ruiz, M. Cortijo, J. Fernandez-Recio, and J. Sancho. 2004b. The long and short flavodoxins. II: Role of the differentiating loop in apoflavodoxin stability and folding mechanism. *J. Biol. Chem.* In press.
- Lostao, A., F. Daoudi, M. P. Irun, A. Ramon, C. Fernandez-Cabrera, A. Romero, and J. Sancho. 2003. How FMN binds to *Anabaena* apoflavodoxin: a hydrophobic encounter at an open binding site. *J. Biol. Chem.* 278:24053–24061.
- Lostao, A., M. El-Harrou, F. Daoudi, A. Romero, A. Parody-Morreale, and J. Sancho. 2000. Dissecting the energetics of the apoflavodoxin-FMN complex. *J. Biol. Chem.* 275:9518–9526.
- Luisi, D. L., C. D. Snow, J. J. Lin, Z. S. Hendsch, B. V. Tidor, and D. P. Raleigh. 2003. Surface salt bridges, double-mutant cycles, and protein stability: an experimental and computational analysis of the interaction of the Asp 23 side chain with the N-terminus of the N-terminal domain of the ribosomal protein I9. *Biochemistry*. 42:7050–7060.
- Ma, B., and R. Nussinov. 2000. Molecular dynamics simulations of a beta-hairpin fragment of protein G: balance between side-chain and backbone forces. *J. Mol. Biol.* 296:1091–1104.
- Maldonado, S., M. P. Irun, L. A. Campos, J. A. Rubio, A. Luquita, A. Lostao, R. Wang, E. B. Garcia-Moreno, and J. Sancho. 2002. Salt-induced stabilization of apoflavodoxin at neutral pH is mediated through cation-specific effects. *Protein Sci.* 11:1260–1273.
- Maldonado, S., M. A. Jimenez, G. M. Langdon, and J. Sancho. 1998a. Cooperative stabilization of a molten globule apoflavodoxin fragment. *Biochemistry*. 37:10589–10596.
- Maldonado, S., A. Lostao, M. P. Irun, J. Fernandez-Recio, C. G. Genzor, E. B. Gonzalez, J. A. Rubio, A. Luquita, F. Daoudi, and J. Sancho. 1998b. Apoflavodoxin: structure, stability, and FMN binding. *Biochimie*. 80:813–820.
- Matsumura, M., G. Signor, and B. W. Matthews. 1989. Substantial increase of protein stability by multiple disulphide bonds. *Nature*. 342:291–293.

- Myers, J. K., and C. N. Pace. 1996. Hydrogen bonding stabilizes globular proteins. *Biophys. J.* 71:2033–2039.
- Pace, C. N., B. A. Shirley, M. McNutt, and K. Gajiwala. 1996. Forces contributing to the conformational stability of proteins. *FASEB J.* 10: 75–83.
- Paterlini, M. G., and D. M. Ferguson. 1998. Constant temperature simulations using the Langevin equation with velocity Verlet integration. *Chem. Phys.* 236:243–252.
- Perl, D., U. Mueller, U. Heinemann, and F. X. Schmid. 2000. Two exposed amino acid residues confer thermostability on a cold shock protein. *Nat. Struct. Biol.* 7:380–383.
- Pinker, R. J., L. Lin, G. D. Rose, and N. R. Kallenbach. 1993. Effects of alanine substitutions in alpha-helices of sperm whale myoglobin on protein stability. *Protein Sci.* 2:1099–1105.
- Rao, S. T., F. Shaffie, C. Yu, K. A. Satyshar, B. J. Stockman, J. L. Markley, and M. Sundaralingam. 1992. Structure of the oxidized long chain flavodoxin from *Anabaena 7120* at 2 Å resolution. *Protein Sci.* 1:1413–1427.
- Richards, F. M. 1997. Protein stability: still an unsolved problem. *Cell. Mol. Life Sci.* 53:790–802.
- Rickaert, J. P., G. Ciccotti, and H. J. C. Berendsen. 1977. Numerical integration of cartesian equations of motions of a system with constraints: molecular dynamics of *n*-alkanes. *J. Comput. Physiol.* 23:327–341.
- Sanchez-Ruiz, J. M., and G. I. Makhatadze. 2001. To charge or not to charge? *Trends Biotechnol.* 19:132–135.
- Santoro, M. M., and D. W. Bolen. 1988. Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl alpha-chymotrypsin using different denaturants. *Biochemistry.* 27:8063–8068.
- Schiffer, C. A., J. W. Caldwell, R. M. Stroud, and P. A. Kollman. 1992. Inclusion of solvation free energy with molecular mechanics energy: alanyl dipeptide as a test case. *Protein Sci.* 1:396–400.
- Sharp, K. A., A. Nicholls, R. Friedman, and B. Honig. 1991. Extracting hydrophobic free energies from experimental data: relationship to protein folding and theoretical models. *Biochemistry.* 30:9686–9697.
- Shi, Z., B. A. Krantz, N. Kallenbach, and T. R. Sosnick. 2002. Contribution of hydrogen bonding to protein stability estimated from isotope effects. *Biochemistry.* 41:2120–2129.
- Smith, P. E., B. M. Pettitt, and M. J. Karplus. 1993. Stochastic dynamics simulations of the alanine dipeptide using a solvent modified potential energy surface. *J. Phys. Chem.* 97:6907–6913.
- Vajda, S., Z. Weng, and C. DeLisi. 1995. Extracting hydrophobicity parameters from solute partition and protein mutation/unfolding experiments. *Protein Eng.* 8:1081–1092.
- Vajda, S., Z. Weng, R. Rosenfeld, and C. DeLisi. 1994. Effects of conformational flexibility and solvation on receptor-ligand binding free energies. *Biochemistry.* 33:13977–13988.
- Weng, Z., C. DeLisi, and S. Vajda. 1997. Empirical free energy calculation: comparison to calorimetric data. *Protein Sci.* 6:1976–1984.
- Xie, D., and E. Freire. 1994. Molecular basis of cooperativity in protein folding. V. Thermodynamic and structural conditions for the stabilization of compact denatured states. *Proteins.* 19:291–301.
- Yang, A. S., and B. Honig. 1995. Free energy determinants of secondary structure formation. I. Alpha helices. *J. Mol. Biol.* 252:351–365.
- Yi, F., D. A. Sims, G. J. Pielak, and M. H. Edgell. 2003. Testing hypotheses about determinants of protein structure with high-precision, high-throughput stability measurements and statistical modeling. *Biochemistry.* 42:7594–7603.
- Zagrovic, B., and V. Pande. 2003. Solvent viscosity dependence of the folding rate of a small protein: a distributed computing study. *J. Comput. Chem.* 24:1432–1436.
- Zielenkiewicz, P., and W. Saenger. 1992. Residue solvent accessibilities in the unfolded polypeptide chain. *Biophys. J.* 63:1483–1486.