Protein Science (2000), *9*:776–785. Cambridge University Press. Printed in the USA.

Novel disulfide engineering in human carbonic anhydrase II using the PAIRWISE side-chain geometry database

RANDALL E. BURTON,[1] JENNIFER A. HUNT,[3] CAROL A. FIERKE,[2] AND TERRENCE G. OAS[1]

[1]Department of Biochemistry, Duke University Medical Center, Durham, North Carolina 27710
[2]Chemistry Department, University of Michigan, 930 N. University, Ann Arbor, Michigan 48109

(RECEIVED October 1, 1999; FINAL REVISION January 13, 2000; ACCEPTED February 25, 2000)

Abstract

An analysis of the pairwise side-chain packing geometries of cysteine residues observed in high-resolution protein crystal structures indicates that cysteine pairs have pronounced orientational preferences due to the geometric constraints of disulfide bond formation. A potential function was generated from these observations and used to evaluate models for novel disulfide bonds in human carbonic anhydrase II (HCAII). Three double-cysteine variants of HCAII were purified and the effective concentrations of their thiol groups were determined by titrations with glutathione and dithiothreitol. The effects of the cysteine mutations on the native state structure and stability were characterized by circular dichroism, enzymatic activity, sulfonamide binding, and guanidine hydrochloride titration. These analyses indicate that the PAIRWISE potential is a good predictor of the strength of the disulfide bond itself, but the overall structural and thermodynamic effects on the protein are complicated by additional factors. In particular, the effects of cysteine substitutions on the native state and the stabilization of compact nonnative states by the disulfide can override any stabilizing effect of the cross-link.

Keywords: disulfide bond; protein engineering; protein stability

Use of disulfide bonds to stabilize proteins

One important goal of protein engineering is the stabilization of natural protein folds to increase their usefulness as reagents. For example, it has been proposed that human carbonic anhydrase II (HCAII) could be modified into a highly sensitive and specific metal ion biosensor (Thompson & Jones, 1993), but the wild-type protein gradually loses activity in long-term storage and may not remain folded under the wide range of solution conditions that a general-use metal sensor might encounter. Engineering novel disulfide bonds into a protein of known structure is one possible way of increasing the global stability of a protein (Matsumura & Matthews, 1991). Studies of bovine pancreatic trypsin inhibitor (BPTI) (Schwartz et al., 1987), phospholipase A_2 (Zhu et al., 1995), ribonuclease T1 (Pace et al., 1988), and hen egg white lysozyme (Cooper et al., 1992) have shown that removing naturally occurring disulfides by mutagenesis can dramatically destabilize proteins. However, attempts to stabilize proteins by the addition of novel disulfide bonds have been less successful (Betz, 1993), although some increases in stability over wild-type have been observed (Kanaya et al., 1991; Clarke et al., 1995; Akker et al., 1997; Mansfeld et al., 1997).

The considerable variability in the energetic effects of engineered disulfides suggests that several stabilizing and destabilizing components of similar magnitude contribute. The primary stabilizing effect of a disulfide bond is proposed to be a reduction in the configurational entropy of the denatured state (Betz, 1993). In addition, the disulfide group is more hydrophobic than two free thiol groups, providing additional stabilization to disulfide bonds that are buried in the native state but not the denatured state of a protein. Analysis of the thermal denaturation of disulfide-bonded proteins in the reduced and oxidized forms has suggested that differential solvation of the residues near the disulfide bond may also have an effect (Doig & Williams, 1991). One important destabilizing effect of a novel disulfide bond is the introduction of strain in the native state structure due to the stringent geometric requirements of the sulfur–sulfur bond. Structure-based prediction methods are needed to ensure that an engineered disulfide is consistent with the original structure of the protein.

Several methods have been employed to use the known three-dimensional (3D) structure of a protein to choose sites for cysteine mutations. The structural constraints of disulfide cross-links arise

Reprint requests to: Terry Oas, Duke University Medical Center, Department of Biochemistry, DUMC 3711, Durham, North Carolina 27710; e-mail: oas@biochem.duke.edu.

[3]Present address: Novartis Agribusiness Biotechnology Research, Inc., P.O. Box 12257, Research Triangle Park, North Carolina 27709.

from the rigidity of the C-S-S bond angle (~104°), the S-S bond length (2.0 Å), and preference for a C-S-S-C dihedral angle of ±90° (Richardson, 1981). Some investigators have chosen a "manual" method, using interactive graphics to model putative disulfides and to assess their suitability with the native structure (Kanaya et al., 1991; Zhou et al., 1993). Pabo and Suchanek (1986) automated this process using the PROTEUS algorithm to locate acceptable sites for making disulfide-linked dimers of the N-terminal domain of λ repressor. Other algorithms based on geometric criteria have also been proposed (Hazes & Dijkstra, 1988; Watanabe et al., 1991).

We have developed an algorithm to predict the suitability of a given site for a novel disulfide link by analyzing the packing preferences of cysteine pairs that form disulfide bonds found in high-resolution crystal structures. The PAIRWISE algorithm, described below, uses a backbone-based local coordinate system to identify the position of one cysteine residue relative to a closely packed neighbor. The primary advantage of the method is that the calculation is very rapid, allowing a large number of possible sites to be tested. In addition, the method is completely general in that pairwise interactions between any two residue types could be analyzed.

Testing disulfide bond strength

Experimental feedback is essential for refining computational modeling techniques such as the PAIRWISE algorithm. The energetic effect of introducing a novel disulfide bond into a protein can be dissected into the stability effects of introducing two cysteines into the structure ($\Delta\Delta G_{cys}$) and creating a cross-link by oxidizing the thiols ($\Delta\Delta G_{ox}$), as described by the following linked thermodynamic cycles:

$$\begin{array}{ccccc} WT\text{-}N & \xleftrightarrow{\Delta G_{cys,N}} & Mut\text{-}N_{SH}^{SH} & \xleftrightarrow{\Delta G_{ox,N}} & Mut\text{-}N_{S}^{S} \\ \updownarrow & \Delta\Delta G_{cys} & \updownarrow & \Delta\Delta G_{ox} & \updownarrow \\ WT\text{-}D & \xleftrightarrow{\Delta G_{cys,D}} & Mut\text{-}D_{SH}^{SH} & \xleftrightarrow{\Delta G_{ox,D}} & Mut\text{-}D_{S}^{S}. \end{array} \quad (1)$$

The vertical arrows in these cycles are the free energies of stability (*N* vs. *D*) for the wild-type protein (*WT*), as well as the double-cysteine variant in the reduced ($Mut\text{-}N^{SH}$) and oxidized ($Mut\text{-}N^{S}$) states. These experimentally observable energies can be used to infer the energetic effects described by the horizontal transitions. The overall goal of disulfide engineering is to minimize the often destabilizing effects of cysteine substitution ($\Delta\Delta G_{cys} = \Delta G_{cys,N} - \Delta G_{cys,D}$), while maximizing the gain from disulfide formation ($\Delta\Delta G_{ox} = \Delta G_{ox,D} - \Delta G_{ox,N}$). The PAIRWISE algorithm is designed to address the latter problem by optimizing the geometric fit between the native structure and the constraints of the disulfide bond, thereby increasing $\Delta G_{ox,N}$ (the stabilizing effect of disulfide formation on the native state). Therefore, it is desirable to have experimental techniques that are able to assess both the overall success of the engineering attempt and the ability of the algorithm to successfully identify sites in a protein of known structure that minimize the strain on the disulfide bond.

The difference in the free energy of stability between the wild-type protein and the oxidized form of the variant ($\Delta\Delta G_{SS} = \Delta G_{D,Mut_S^S} - \Delta G_{D,WT} = (\Delta G_{cys,N} + \Delta G_{ox,N}) - (\Delta G_{cys,D} + \Delta G_{ox,D}) = \Delta\Delta G_{cys} + \Delta\Delta G_{ox}$) is directly relevant to the goal of engineering a more stable protein. $\Delta\Delta G_{SS}$ can be readily measured by taking the difference in free energy of denaturation between the wild-type protein and the oxidized form of the double-cysteine variant. As mentioned above, $\Delta\Delta G_{SS}$ is not a direct measure of the compatibility of the native structure with the geometric requirements of the disulfide bond as it is affected by slight changes in packing and solvation around the site of the cysteine substitutions, as well as deviations from two-state behavior and residual structure in the denatured state (Betz, 1993). The predictive power of the PAIRWISE algorithm can be assessed explicitly by direct measurement of $\Delta\Delta G_{cys,N}$, given by the effective concentration of the protein thiol groups for intramolecular disulfide formation relative to the bimolecular association of a small thiol compound (Creighton, 1983; Lin & Kim, 1989). The common choice is glutathione, which is a biologically relevant redox compound that is chemically similar to a polypeptide.

Measurement of the effective concentrations of protein disulfides relative to glutathione requires the measurement of the ratio of oxidized (P_S^S) to reduced (P_{SH}^{SH}) protein in the presence of excess reduced (GSH) and oxidized (GSSG) glutathione. This analysis can be complicated by the presence of additional species, such as disulfide-linked protein dimers (PSSP) or mixed disulfides between protein and glutathione (P_{SH}^{SSG} and P_{SSG}^{SSG}). One approach is to separate and quantitate all possible species by high-performance liquid chromatography or native polyacrylamide gel electrophoresis, as has been done for BPTI (Goldenberg et al., 1993) and thioredoxin (Lin & Kim, 1989). However, these techniques require optimization for each protein system studied and may not be applicable for some proteins. To address this difficulty, we have developed a simple protocol using a desalting column followed by thiol quantitation, using conditions that minimize the formation of unwanted disulfide species (PSSP, P_{SH}^{SSG}, and P_{SSG}^{SSG}). These conditions limit the range of effective concentrations that can be measured and require additional experiments to confirm that undesired disulfide species do not accumulate significantly at equilibrium, but the protocol can potentially be used with any protein. This technique was used to measure effective concentrations for novel disulfides introduced in HCAII, which were then used to assess the predictive power of a novel algorithm for structure-based disulfide engineering.

Results

Development of the PAIRWISE database

A database of 377 high-resolution crystal structures has been collected from the Brookhaven Protein Data Bank (PDB) database, starting with the 25% identity cutoff list from Hobohm and Sander (1994), and selecting those structures with ≤2.0 Å resolution and an *R*-factor of 0.2 or better. The PAIRWISE program scans this database for all instances where two specified residues that are more than five residues apart in the primary sequence come in contact. Contact is assumed when the C_β atoms are close enough for the side chains to potentially make van der Waals contact. The longest distance from the C_β atom to any other atom in the same residue is measured for the 20 side chains ($d_{\max}$). Whenever the C_β-C_β distance is less than $d_{\max,1} + d_{\max,2} + 3.0$ Å, the pair is evaluated. The local space around a given residue is uniquely defined by three points, shown in Figure 1. The C_β atom is set at the origin. The C_α atom is placed on the negative *x*-axis, and the carbonyl carbon is placed in the *xy* plane, in positive *y* space. Once the coordinate space around a given residue has been defined, its

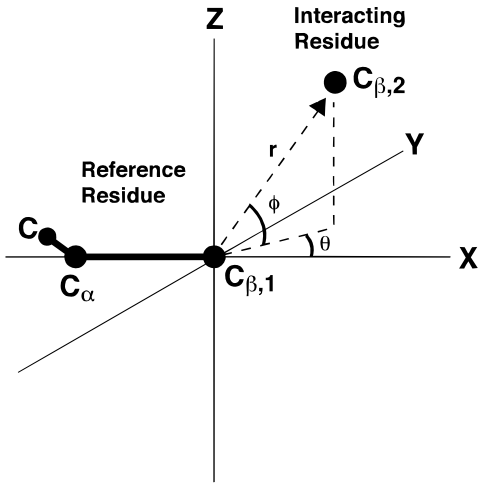

Fig. 1. Description of coordinate system used to analyze side-chain packing orientations. Three backbone atoms (C, C_α, and C_β) are used to define the *XY* plane, and the position of the C_β atom of the other residue is determined using a spherical coordinate system (θ, ϕ, r).

interactions with another residue can be probed by asking where the C_β atom in the other residue lies within that coordinate system. To separate the coordinate information into orientation-based and distance-based components, a spherical coordinate system is used that defines each point by two angles (θ, ϕ) and a distance *r*. In the current implementation, the *r* value is not used, but could potentially be used to generate a distance-based contact potential. The list of (θ, ϕ) points were then used to generate a probability distribution, using a Gaussian kernel estimator (Silverman, 1986), shown below:

$$P(\theta,\phi) = \frac{1}{Nh^2}\sum_{i=1}^{N} e^{-\frac{d^2}{4\pi h^2}}$$

$$d^2 = (x - x_i)^2 + (y - y_i)^2 + (z - z_i)^2$$

$$x = \cos\phi\cos\theta$$

$$y = \cos\phi\sin\theta$$

$$z = \sin\phi \qquad (2)$$

where *N* is the number of points, *h* is a smoothing factor chosen to be 0.25 based on visual inspection and examination of the cross-correlation function of the density matrix at various values of *h* (Silverman, 1986), and *d* is the distance between the grid point (θ, ϕ) and the data point (θ_i, ϕ_i) in Cartesian space. A Gaussian curve centered on each data point is generated, and the sum of these curves is a continuous function that contains information on the types of packing orientations found in the database. Preferred packing arrangements are reflected as peaks in the distribution, and disallowed configurations are indicated by valleys or "plains" with density values near zero. The distributions are sampled on a regularly spaced 40×40 grid in (θ, ϕ) space, and the density value for a given (θ, ϕ) point is calculated by linear interpolation between the nearest grid points.

PAIRWISE analysis of cysteine–cysteine packing orientations

The database of crystal structures was scanned for instances where two cysteines come in contact according to the criterion outlined above. Visual examination of a random sample of these cysteine pairs using INSIGHT II (MSI, Inc., San Diego, California) indicated that nearly all of these residues were joined by disulfide cross-links. The few exceptions involved crystal structures with multiple disulfide bonds close in space, such that two cysteines would be close enough to potentially interact, but favored bond formation with another partner. These exceptions were left in the distribution, because removing them did not significantly affect the final predictions. The PAIRWISE algorithm was used to generate a list of (θ, ϕ) values to define the relative orientation of the cysteine residues. Two sets of (θ, ϕ) points were generated for each pair of residues A and B, one in which residue A was used to define the coordinate system and the (θ, ϕ) coordinates of the C_β atom in residue B was recorded (θ_1, ϕ_1) and another in which residue B defined the coordinate system and the position of residue A was determined (θ_2, ϕ_2) . This list of (θ, ϕ) points was used to generate the probability density function shown in Figure 2. This distribution shows that there are strong preferences for (θ, ϕ) values around $(-\pi/4, \pi/3)$ and weaker preferences for areas around $(\pi/4, 0)$ and $(-\pi/4, -\pi/4)$. Visual examination of selected cysteine pairs with these (θ, ϕ) values indicated that their disulfide bonds adopted bond lengths and angles close to ideal values. The peaks in the density function also roughly correlated with previously observed rotameric preferences for disulfide groups (Richardson, 1981; Thornton, 1981; Harrison & Sternberg, 1996; Petersen et al., 1999). In particular, the large peak at $(-\pi/4, \pi/3)$ consist predominantly of the two most common symmetric rotamers, $g^+g^+\mathbf{g}g^+g^+$ and $g^+g^+\mathbf{d}g^+g^+$, using the terminology of (Harrison & Sternberg, 1996). The peak at $(-\pi/4, -\pi/4)$ seems to dominated by the most common asymmetric rotamer $g^+g^+\mathbf{d}g^-t$, while the $(\pi/4, 0)$ peak contains a wide variety of rotamers, including the immunoglobulin-like conformation. The PAIRWISE algorithm is able to identify the stereochemical preferences of disulfide bonds using a coordinate system that is generalizable to all natural amino acids, suggesting

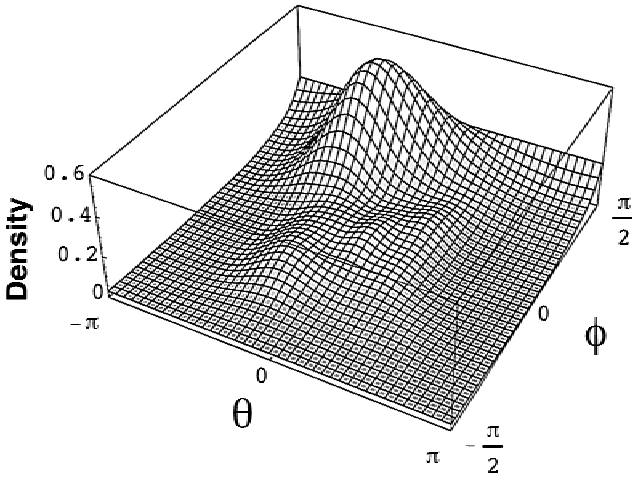

Fig. 2. Cys–Cys probability distribution. The conformational preferences induced by the presence of a disulfide bond are clearly seen as the peaks in the distribution and large, flat areas in (θ, ϕ) space where no points are found.

that this analysis may also be able to identify packing preferences for every side chain.

Predictive pseudo-energy function for disulfide-bonded CysCys pairs

Since the (θ,ϕ) points for a given pair depend only on the backbone atoms, the preferences observed in Figure 2 offer a rapid means for assessing the suitability of a given backbone conformation for disulfide engineering. To test the predictive power of this distribution, 10 randomly-chosen crystal structures of proteins with naturally-occurring disulfide bonds were left out of the database used to calculate the (θ,ϕ) distribution. The resulting distribution was visually identical from a distribution calculated using all structures, and the Kolmogorov–Smirnov test (Press et al., 1992) indicated that these distributions were statistically indistinguishable (data not shown). The disulfides in the ten test structures were evaluated by calculating a psuedo-energy score from their (θ,ϕ) values:

$$score = -RT \ln \frac{p(\theta_1,\phi_1)p(\theta_2,\phi_2)}{p_{\text{random}}^2} \quad (3)$$

where $p(\theta,\phi)$ is the value of the density function at a given coordinate and p_{random} is the probability density derived from a random distribution. Negative scores indicate (θ,ϕ) values that are found more often in the crystal structure database than expected from a random distribution, and positive scores indicate (θ,ϕ) values rarely found in the database. The scores for the 21 naturally-occurring disulfides in this test set were uniformly negative, ranging from $-2{,}900$ to $-11{,}000$ J mol^{-1} (data not shown).

Selection of disulfide bonds in carbonic anhydrase

All possible residue pairs in HCAII were scored in this manner to assess the suitability of putative disulfide bonds throughout the structure. Table 1 shows the PAIRWISE predictions and structural parameters for three putative disulfides in HCAII that appear in different structural contexts and sample a wide range of PAIRWISE scores. These sites were predicted to be either highly favorable (L60C/S173C), neutral (A38C/A258C), or unfavorable (S29C/S197C). In addition, these sites potentially introduce disulfides into diverse structural contexts: joining two β-strands (L60C/S173C), a β-strand to a loop (S29C/S197C), and a unique pair of β-strands involved in a pseudoknot structure (A38C/A258C) (Eriksson et al., 1988). Since the PAIRWISE calculation was performed using the wild-type structure with no allowance for backbone movement in response to the disulfide, relatively rigid sites were chosen as indicated by low β-factors reported in the crystal structure (PDB code: 2CBA (Håkansson et al., 1992)). A naturally-occurring cysteine residue at position 206 was substituted with arginine to prevent complications from a third thiol group and to improve binding of the fluorescent inhibitor dansylamide (Krebs & Fierke, 1993). The C206R sequence was then used as the background for the double-cysteine substitutions and will be referred to as WT* throughout.

Disulfide formation under oxidizing conditions

Table 2 shows that the number of free thiols detected per protein varies from 2.0 ± 0.1 under strongly reducing conditions to 0.1 ± 0.1 in strongly oxidizing conditions for all of the variants studied. The formation of disulfide-linked dimers was ruled out by SDS-PAGE, and no protein–glutathione adducts were detected by electrospray mass spectrometry (data not shown). These results indicate that all cysteine-containing variants studied here are fully reduced in 5 mM dithiothreitol (DTT) and fully oxidized in 1 mM GSSG, allowing both redox states to be characterized under identical conditions for each variant. Disulfide stabilities were determined by measuring the effective concentrations of the protein thiol groups. Effective concentrations have been used as a measure of local strain in variants of thioredoxin (Lin & Kim, 1989, 1991) and staphylococcal nuclease (Hinck et al., 1996). Figure 3 shows the titration of protein thiols in the three-disulfide variants of HCAII with glutathione and DTT. The S29C/S197C variant is readily reduced by glutathione, agreeing with the PAIRWISE prediction that the native backbone structure was not consistent with disulfide bond formation even though the C_β atoms in these residues should be close enough to interact. The A38C/A258C and L60C/S173C variants are more resistant to reduction; no detectable concentration of the thiol forms of these proteins could be found at the [GSH]/[GSSG] ratios studied. Reduction of these variants was possible using DTT, which is a stronger reducing agent. The A38C/A258C variant was fully reduced at all ratios studied, indicating that the redox potential for this disulfide is intermediate between gluathione and DTT. The equilibrium constant for reduction $K_{red} = ([P_S^S][DTT_{SH}^{SH}])/([P_{SH}^{SH}][DTT_S^S])$ of the L60C/S173C variant by DTT was determined to be 0.6, comparable to the K_{red} of 0.1 and 0.5, respectively, observed for the 14–38 disulfide bond in

Table 1. *Results of computational analysis of HCAII variants*[a]

CAII variant	PAIRWISE score (J mol^{-1})	Δ in side-chain volume (Å^3)	Average β-factor	Structural context
C206R (WT*)	—	—	6.3	Active site loop
A38C/A258C	−150	+28	13.9	Bridges pseudoknot structure
L60C/S173C	−8,000	−159	10.9	Edge of central β-sheet
S29C/S197C	+5,000	+14	5.4	Connects end of short β-strand to active site loop

[a]The first column is the predicted change in side-chain volume caused by the cysteine mutations, using standard volumes (Richards, 1977). The PAIRWISE algorithm was used to analyze the suitability of three potential disulfide bonds with the native structure observed in the crystal structure (Håkansson et al., 1992), PDB code: 2CBA, assuming a fixed backbone conformation. The average β-factor for the backbone atoms of the mutated residues and the structural context of the substitution(s) are taken from the wild-type structure.

Table 2. *In vitro assays of disulfide bond formation*[a]

CAII variant	[-SH]:[Protein] (5 mM DTT)	[-SH]:[Protein] (1 mM GSSG)	C_{eff} (M)
C206R (WT*)	0.0 ± 0.1	0.1 ± 0.1	—
A38C/A258C	2.1 ± 0.1	0.1 ± 0.1	36 ± 24[b]
L60C/S173C	1.9 ± 0.1	0.2 ± 0.1	720 ± 120[b]
S29C/S197C	2.1 ± 0.1	0.2 ± 0.1	0.0064 ± 0.0006

[a]The effective concentrations of the thiol groups reflect the strength of the disulfide bonds toward reduction by small molecule disulfides. All measurements were made in 0.1 M $TrisSO_4$, pH 8.0, 0.1 mM $ZnSO_4$ at 25 °C.

[b]Calculated using a C_{eff} of 1,200 M for DTT (Goldenberg et al., 1993).

BPTI (Goldenberg et al., 1993) and a highly stabilizing engineered disulfide in barnase (Clarke et al., 1995). C_{eff} for these proteins can be estimated from the effective concentration of the thiol groups in DTT, 1,200 M (Goldenberg et al., 1993).

Effects of substitutions on enzymatic activity and inhibitor binding

While the effective concentration of the thiol groups is a precise test of the strain imposed on a disulfide bond, stabilization of the overall protein structure requires that the cysteine substitutions have little or no impact on the native-state structure or stability. To test the effects of these substitutions on the gross structural features of the protein, enzymatic activity and inhibitor binding assays were performed on the oxidized and reduced forms of the proteins. Differences between the WT* variant and the reduced form of the double-cysteine variant indicate effects of the amino acid changes. Additional effects upon oxidation are due to disulfide formation. HCAII has two readily measurable activities: *p*-nitrophenyl acetate hydrolysis and CO_2 hydration. In addition, binding of the fluorescent inhibitor dansylamide (DNSA) is sensitive to the presence of the protein-bound zinc atom and the integrity of the active-site cleft (Thompson & Jones, 1993; Nair et al., 1996). As shown in Table 3, DNSA binds the WT* variant almost 1,000-fold more tightly than the wild-type protein, consistent with earlier observations (Krebs & Fierke, 1993). The oxidized forms of two of the double-cysteine variants (L60C/S173C and A38C/A258C) also bind DNSA very tightly. No binding was detected to the S29C/S197C variant in up to 30 μM DNSA, indicating that the critical binding pocket is not present under these conditions.

The activities measured for WT* and the three double cysteine variants, listed in Table 3, also suggest that the C206R, L60C/S173C, and A38C/A258C substitutions have little or no effect on the active site, whereas the S29C/S197C mutation is severely disruptive. The two disulfide variants predicted to be compatible with the wild-type structure (L60C/S173C and A38C/A258C) have similar PNPA hydrolysis and CO_2 hydration activities as the WT* background, indicating that the structure of the active site is only modestly perturbed. However, the S29C/S197C protein is inactive under both reducing and oxidizing conditions, indicating that the cysteine substitutions have disrupted the active site structure, and this perturbation is not relieved by formation of the disulfide bond. Residues Leu60, Ser173, Ala38, and Ala258 are well removed from the active site cleft of HCAII, while Ser29 and

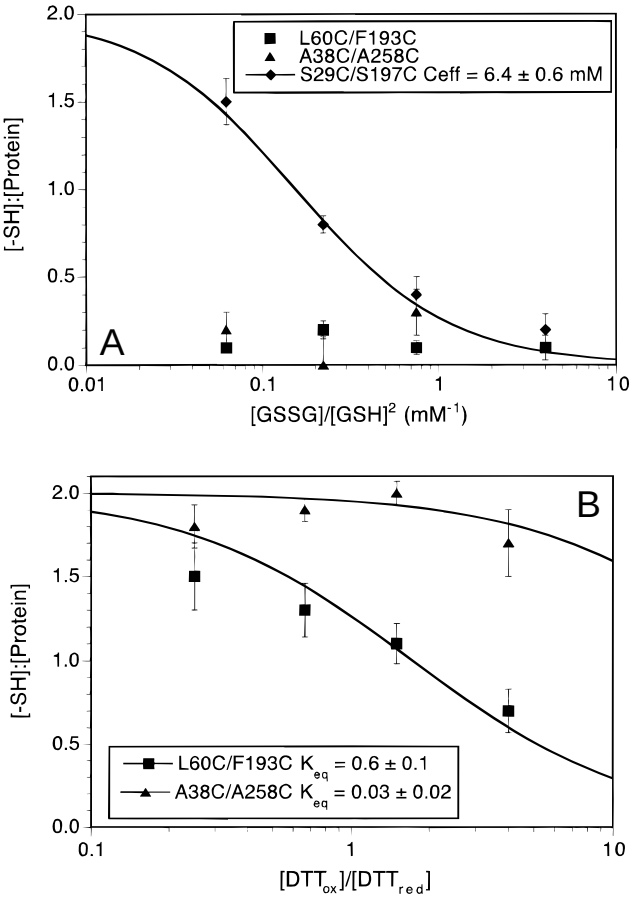

Fig. 3. Effective concentrations of protein thiol groups vs. (**A**) glutathione and (**B**) dithiothreitol. The number of thiol groups per protein molecule was determined by reaction with 2,2′-dithiopyridine under denaturing conditions (Materials and methods). This ratio varies between 2.0 (fully reduced) to 0.0 (fully oxidized) as the redox potential of the solution is changed with the small thiol compounds. These data were fit to Equation 4 in Materials and methods to determine effective concentrations. SDS-PAGE was used to verify that no detectable amounts of intermolecular disulfide bonds were formed, and no protein–glutathione mixed disulfides were detected by mass spectrometry.

Ser197 are immediately adjacent to a critical active site loop (Håkansson et al., 1992; Krebs & Fierke, 1993), which may explain the loss of activity when these sites are altered. The serine residue at position 29 is conserved among all carbonic anhydrases and a single S29C substitution was shown to be extremely destabilizing (Mårtensson et al., 1992). In addition, the residues that contact Ser29 in the wild-type crystal structure are also conserved, including Ser197, although single substitutions of Ser197 have little or no effect on activity (Krebs & Fierke, 1993). The peptide bond between Ser29 and Pro30 is in the *cis* conformation in the native structure (Eriksson et al., 1988). It is possible that even the relatively mild Ser → Cys substitutions at these positions disrupt this strained conformation and inactivate the enzyme.

Circular dichroism spectra of HCAII variants

The circular dichroism (CD) spectra for WT* and the disulfide variants, shown in Figure 4, provide a more global picture of the

Table 3. *Enzymatic activities of WT* and the double cysteine mutants*[a]

	5 mM DTT		1 mM GSSG		
CAII variant	PNPA (M^{-1} s^{-1})	CO_2 ($\times 10^8$ s^{-1})	PNPA (M^{-1} s^{-1})	CO_2 ($\times 10^8$ s^{-1})	DNSA K_D (nM)
C206R (WT*)	920	1.3	1,500	2.0	18
A38C/A258C	940	ND	780	ND	7
L60C/S173C	520	0.5	1,300	1.1	8
S29C/S197C	<50	<0.01	<50	<0.01	NA[b]

[a] Activities for *p*-nitrophenol acetate (PNPA), hydrolysis (k_{cat}/K_M in 0.5 mM PNPA), and CO_2 hydration (k_{cat}) are reported under reducing (5 mM DTT) or oxidizing (1 mM GSSG) conditions. The dissociation constant for the fluorescent inhibitor dansylamide (DNSA) indicates the structural integrity of the active site in the oxidized form of the protein. All measurements were made at 25 °C in 0.1 M $TrisSO_4$, pH 8.0, 0.1 mM $ZnSO_4$.

[b] No detectable binding up to 30 μM DNSA.

structural effects of these amino acid substitutions. The far-ultraviolet (UV) spectra for all four proteins are similar under both reducing and oxidizing conditions and consistent with the predominantly β-sheet structure seen in the wild-type crystal structure (Håkansson et al., 1992). The near-UV spectra, which are sensitive to the tertiary environments of the eight tryptophans (Frekgård et al., 1994), indicate that WT*, L60C/S173C, and A38C/A258C have tightly packed cores, while S29C/S197C does not. Since the near-UV signal is lost in the reduced as well as oxidized form, the disruptive effect is probably due to the insertion of cysteines at

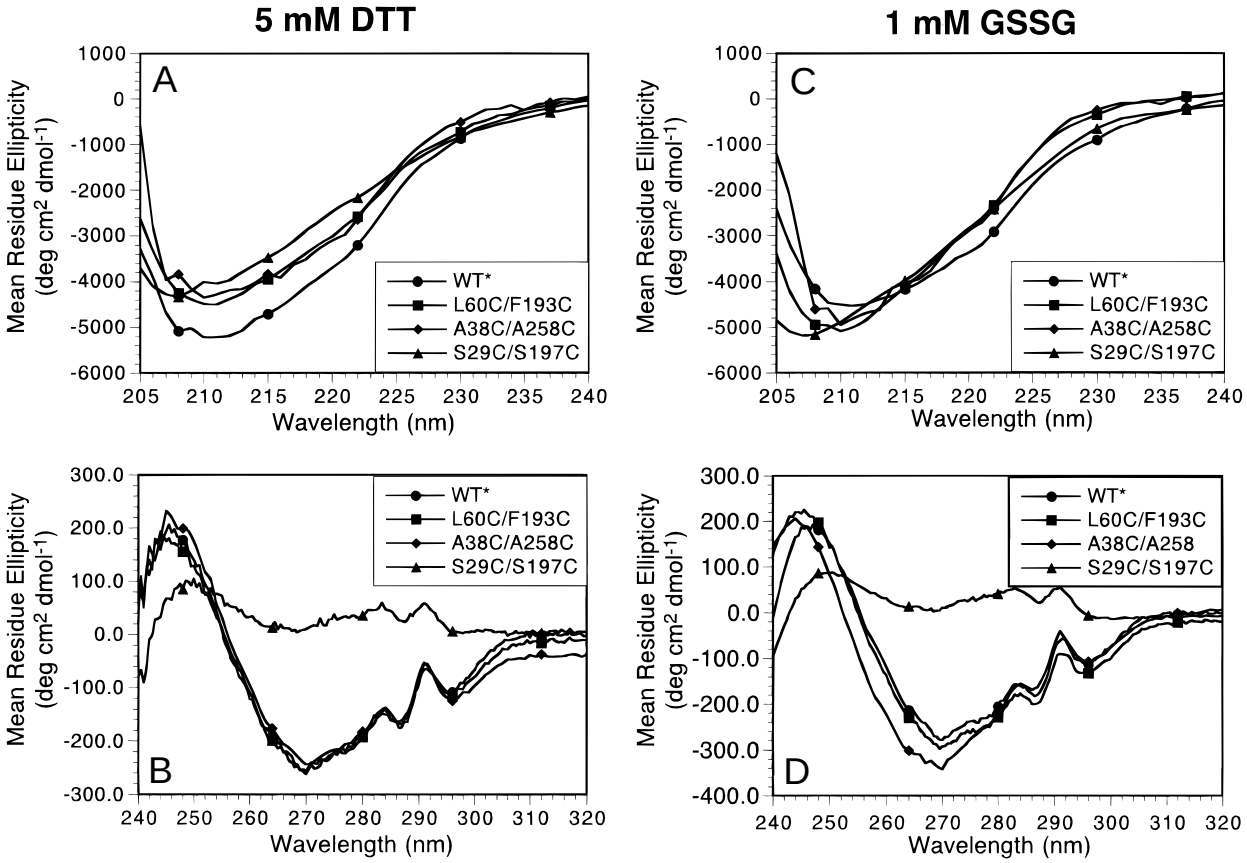

Fig. 4. CD spectra of mutants in (**A**, **B**) 5 mM DTT and (**C**, **D**) 1 mM GSSG. The (**A**, **C**) far-UV spectra indicate that the overall secondary structure for all of the variants studied are nearly identical under both reducing (DTT) and oxidizing (GSSG) conditions. However, the (**B**, **D**) near-UV spectra indicate that the local environment of the tryptophan side chains in the S29C/S197C variant has been perturbed, and that these side chains are probably quite mobile. In contrast, the near-UV spectra for the L60C/S173C and A38C/A258C variants are similar to WT* under all conditions.

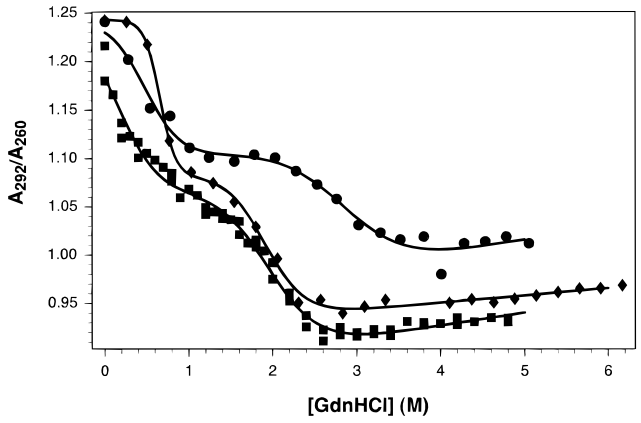

Fig. 5. GdnHCl denaturation of WT* (♦) and the L60C/S173C CAII variant in reduced (■) and oxidized (●) form. The solid and dashed lines represent fits of the data to a three-state ($U \leftrightarrow I \leftrightarrow N$) model. The parameters derived from these fits are shown in Table 4.

these positions, consistent with the observed effects on activity and inhibitor binding.

GdnHCl denaturation of WT and L60C/S173C*

As the L60C/S173C variant forms a strong disulfide bond as measured by C_{eff} and does not appear to have disrupted the global structure of the enzyme, the stability of the protein was determined by titration with guanidine hydrochloride (GdnHCl), followed by the ratio of absorbance at 292 and 260 nm to correct for aggregation of the protein at intermediate GdnHCl concentration, as shown in Figure 5. The titration data were fit to a three-state model (Tweedy et al., 1993), and the results of these fits are shown in Table 4. The $N \rightarrow I$ transition in the reduced form of the L60C/S173C variant occurs at a much lower denaturant concentration than for WT* to the extent that no native baseline is observed. The parameters related to the $N \rightarrow I$ transition are therefore poorly determined, but ΔG_{NI} is clearly much lower that for WT*. The $I \rightarrow U$ transition is quite similar between WT* and reduced L60C/S173C, suggesting that these states are either not perturbed by the substitutions, or they are equally perturbed. Upon oxidation of the disulfide the midpoints of both transitions increase, the absorbances of *I* and *U* increase and the transitions occur over a broader range of [GdnHCl]. These changes suggest that while disulfide formation does have the intended effect of stabilizing the native fold, the cross-link may be influencing the ensemble of conformations that make up the nonnative states, as has been observed previously in other proteins (Pace et al., 1990; Betz et al., 1996).

Discussion

The results for the three double-cysteine variants of HCAII outlined above underscore the difficulties of structure-based disulfide design in a protein that can adopt compact, nonnative states. The destabilizing effects of cysteine substitutions at buried positions in the native state can arise from either changes in side-chain volume or the chemical nature of the cysteine side chain. In particular, the thiolate form of cysteine would be very destabilizing in a hydrophobic core due to the cost of desolvating the charged sulfur atom. The primary force proposed to stabilize disulfide-bonded proteins is a reduction in the conformational entropy of nonnative states (Betz, 1993). This entropic contribution can be estimated by assuming a random coil conformation for the denatured state and calculating the entropy lost by covalently closing the loop between the cysteine residues. Using Equation 5 from Betz (1993), the estimated entropic benefits of forming the three disulfides in this study are between 2.7 and 3.0 kcal mol^{-1} at 298 K. The full amount of this stabilization can only be realized if the disulfide does not induce any strain in the native state and the nonnative states of the protein can be described as random coil. Strain in the native state may account for the lower effective concentrations of the thiol groups in the S29C/S197C and A38C/A258C variants.

However, a disulfide bond will stabilize any compact conformation that allows the disulfide to adopt an appropriate geometry. Oxidation of the 60–173 disulfide bond shifts both the $N \leftrightarrow I$ and $I \leftrightarrow U$ transitions to higher GdnHCl concentrations, which suggests that the *I* state has a structure that is also compatible with the disulfide bond, relative to the *U* state. Comparison of the GdnHCl denaturation profiles of L60C/S173C and WT* suggest that the conformational ensembles that make up the *I* and *U* states in WT* may be different than those populated by the reduced and oxidized forms of L60C/S173C. The fitted values of m_{NI} for WT* and reduced L60C/S173C are different, but in the absence of a good native baseline, the uncertainty in the m_{NI} value for reduced L60C/S173C prevents a detailed interpretation. The fitted m_{IU} value is also lower for the reduced variant, but visual inspection of the titration curves suggest that the $I \rightarrow U$ transitions are quite similar for these proteins (Fig. 5). In contrast, the oxidized L60C/S173C protein has a significantly different GdnHCl denaturation profile. The absorbance of the *I* and *U* states are significantly perturbed by formation of the disulfide cross-link, and the m_{NI} and m_{IU} values

Table 4. *Parameters derived from a three-state fit of the GdnHCl-denaturation data for C206R and L60C/S173C* [a]

Variant	A_U (5 M)	Z	ΔG_{NI} (kcal mol^{-1})	m_{NI} (kcal mol^{-1} M^{-1})	ΔG_{IU} (2M) (kcal mol^{-1})	m_{IU} (kcal mol^{-1} M^{-1})
WT*	0.96	0.51	4.2	6.4	−0.2	2.5
L60C/S173C (reduced)	0.94	0.50	0.4	2.5	−0.1	2.0
L60C/S173C (oxidized)	1.02	0.49	1.4	2.8	1.4	1.6

[a] The absorbance ratio for the fully native protein (A_N) was not well defined for the L60C/S173C data, so both datasets were fit using a constant value of $A_N = 1.243$, the value determined from a complete fit of the WT* data. The absorbances of the native and intermediate states (A_N and Z) were assumed to be independent of [GdnHCl].

are even lower than those observed for the reduced protein. These observations suggest that the *I* and *U* states of oxidized L60C/S173C consist of significantly different conformations from both the reduced and WT* proteins in terms of solvation (*m* values) and the environment of the aromatic side chains (absorbance).

The effects of forming the 60–173 disulfide may give some additional insights into the conformations of the *I* and *U* states of HCAII. Mutagenesis and chemical-labeling studies have shown that the center of the central β-sheet in HCAII is inaccessible to solvent in the intermediate state, but the edges are frayed (Mårtensson & Jonsson, 1993). These experiments also showed that the center of the sheet remains ordered even in the *U* state, although the edges appear to be completely unfolded. Since the 60–173 disulfide would tether strands 9 and 10 at one edge of the β-sheet, the β-sheet structure may persist out to that edge in the *I* and *U* states, attenuating the specific stabilization of the *N* state by the disulfide bond. Therefore, it would appear that the 60–173 disulfide, while a strong bond in terms of the effective concentration of the thiol groups, only modestly stabilizes ($\sim$1 kcal mol^{-1}) the native state relative to the *I* and *U* states, perhaps due to the compact nature of these nonnative states.

The S29C/S197C and A38C/A258C variants were intended as controls for the application of the PAIRWISE potential for disulfide engineering and had the expected characteristics of disulfides that are inconsistent with the original structure. The S29C/S197C variant was completely nonnative by all criteria under both reducing and oxidizing conditions. The stability of this protein was severely compromised by the cysteine substitutions, and disulfide formation could not recover any detectable population of the native state. In addition, the effective concentration of the thiol groups was quite low, consistent with the prediction made by the PAIRWISE algorithm that the native backbone structure was not consistent with a disulfide bond between residues 29 and 197. The low effective concentration is somewhat surprising since the cysteine mutations appear to have disrupted the native conformation, which could allow the disulfide to form more freely than if the structure was intact. The ester hydrolysis activity and CD spectra of the A38C/A258C variant were identical to wild-type CAII under reducing conditions, suggesting that the protein is more tolerant to reduced cysteine substitutions at these sites than at the other two sites. However, the activity decreased significantly in the presence of glutathione, in contrast to WT* and L60C/S173C, suggesting that the structure of the native state is disrupted by the formation of the disulfide bond.

The results for these double-cysteine variants of HCAII suggest a number of ways that the computational screening for potential disulfide engineering sites can be improved. In particular, the destabilizing effects of the cysteine substitutions must be minimized. One way to do this is to avoid substitutions at evolutionarily conserved sites, which would have eliminated the S29C/S197C variant from consideration. Molecular models of the cysteine substitutions in the native crystal structure, made with no allowance for structural rearrangement, indicate that the more disruptive substitutions (A38C/A258C and S29C/S197C) have steric overlaps between the sulfur atoms and other parts of the structure. In particular, the model of S29C/S197C suggests significant overlaps between the cysteine side chains and surrounding backbone atoms. These overlaps are probably relieved by movement of the backbone, which could then disrupt additional packing interactions in that region. A second generation of the disulfide prediction routine would scan the structure for residues that are close enough for a potential disulfide cross-link, use the PAIRWISE score to assess geometric suitability, then explicitly model promising candidates to look for steric overlaps. In this way, the efficient PAIRWISE algorithm could screen a large number of candidates and identify those worth a more detailed examination. Extending the search algorithm to allow movement of backbone atoms may also improve its ability to model disulfides in flexible regions. The speed of the PAIRWISE calculation would make it possible to screen a large number of possible backbone perturbations, and the most promising candidates could then be screened in more detail with an all-atom model.

Conclusions

The work presented here indicates that the simple, rapid PAIRWISE algorithm has been able to identify suitable and unsuitable sites for disulfide bonds in human carbonic anhydrase II. However, it is clear that careful consideration must be made concerning the effects of cysteine substitutions in the cores of proteins, even when the packing effects would be expected to be subtle (e.g., Ser $\rightarrow$ Cys substitutions). The method used for determining the effective concentration of protein thiols should generally applicable to other proteins, unlike previously described techniques (Lin & Kim, 1989; Goldenberg et al., 1993), although additional care must be taken to rule out intermolecular disulfide species. These measurements indicate that at least one of the disulfide variants predicted to be consistent with the native structure (L60C/S173C) is nearly as stable toward reduction as a naturally occurring disulfide bond in BPTI (Goldenberg et al., 1993). The negative control (S29C/S197C), which was predicted to form a strained disulfide bond, indeed showed a low effective concentration of the thiol groups and a nonnative structure. This work indicates that the presumably stabilizing effects of even a very good disulfide bond can be masked by additional factors.

Materials and methods

Mutagenesis, protein expression, and purification

Double-cysteine substitutions were introduced into pACA1 (Nair et al., 1991) containing the gene for human carbonic anhydrase II by Kunkel mutagenesis (Kunkel et al., 1987). The entire coding regions for the altered HCAII genes were sequenced by the dideoxynucleotide termination method (Sanger et al., 1977) to verify that the substitutions had been made and that the remainder of the gene was unaffected.

Cells were grown at 37 °C in rich media (20 g/L tryptone, 10 g/L yeast extract, 5 g/L NaCl, 0.33X M9 salts, 0.37% glucose, 0.6 μM $ZnSO_4$) with 100 mg/L ampicillin to 1.0 OD_{600}, then induced by the addition of 0.25 mM isopropyl-βD-thiogalactoside (IPTG) (final concentration). The temperature was lowered to 30 °C, and the cells were harvested by centrifugation 6 h later. Cells were lysed in a French press at 1,200 psi, and the supernatant was run through a DEAE Sephacel column. The flow-through was dialyzed against 10 mM MES, pH 7.0, and then loaded onto a SP-Sepharose column and eluted with a 0–0.5 M $NaSO_4$ gradient. The pooled fractions were run through a SuperDex G75 gel-filtration column equilibrated in 10 mM TrisSO_4, pH 8.0. Protein stocks were kept at -80 °C until needed.

Enzymatic activity measurements

The activity of WT* and the double-cysteine variants for both *p*-nitrophenyl acetate (PNPA) ester hydrolysis and CO_2 hydration was measured in 0.1 M $TrisSO_4$, pH 8.0, 0.1 mM $ZnSO_4$. The PNPA hydrolysis activity was determined in 0.5 mM PNPA from the initial rate of change in absorbance at 348 nm, using an extinction coefficient of 5,000 M^{-1} for the formation of *p*-nitrophenol (Armstrong et al., 1966). CO_2 hydration was measured by the pH indicator method (Kalifah, 1971) in 175 mM 3-*N*-tris-(hydroxymethyl)methylamino-propanesulfonic acid (TAPS)/m-cresol purple indicator buffer, pH 8.0, 0.35 mM ethylenediaminetetraacetic acid (EDTA), 118 mM Na_2SO_4, using a stopped-flow device from Applied Photophysics, Inc. (Leatherhead, UK).

CD spectroscopy

CD measurements were made on an Aviv 62DS spectrometer with a sample containing 10 μM protein, 10 mM $TrisSO_4$, pH 8.0, 0.1 mM $ZnSO_4$, and 5 mM of either oxidized glutathione or DTT. Since glutathione makes a significant contribution to the CD spectrum in the range studied (200–320 nm), samples were equilibrated in glutathione for 4 h, then dialyzed against 10 mM $TrisSO_4$, pH 8.0, 0.1 mM $ZnSO_4$ for 24 h. A 1 cm pathlength quartz cell was used for the near-UV spectrum (320–240 nm), and a 0.1 cm pathlength cell was used for the far-UV region (240–200 nm).

Thiol content measurements

A 50 μM protein solution was equilibrated in 100 mM $TrisSO_4$, pH 8.0, 0.1 mM $ZnSO_4$, and 5 mM of either GSSG or DTT. The thiol exchange reaction was allowed to proceed for at least 2 h and then quenched by the addition of 1N HCl to pH 2.0. The small molecule thiols were removed on a PD-10 (Pharmacia, Uppsala, Sweden) gel filtration column, and the protein was collected in four fractions. The protein in each fraction was denatured by the addition of GdnHCl to 4 M. The pH was neutralized with 1 M $TrisSO_4$, pH 7.5 to a final pH of 7.0 and the protein concentration was determined from the absorbance at 280 nm, using the extinction coefficient (ϵ_{280} = 54,580 M^{-1}) determined by the number of tyrosine and tryptophan residues (Edelhoch, 1967). The thiol content was assayed by adding 2,2′-dithiodipyridine (Aldrithiol) and measuring the change in absorbance at 343 nm, $\Delta\epsilon_{343\ nm}$ = 7,600 $M^{-1}\,cm^{-1}$ (Grassetti & Murray, 1967). The effective concentrations of the thiol groups were determined by fitting the thiol content data to the following equation, which assumes that protein–glutathione mixed disulfides are not significantly populated under the assay conditions:

$$\frac{[-SH]}{[\text{Protein}]} = \frac{2}{C_{eff}\dfrac{[\text{GSSG}]}{[\text{GSH}]} + 1}. \tag{4}$$

Disulfide stabilities were measured for each variant by determining the effective concentration of the thiol groups. The bimolecular reaction is approximated using glutathione as a reference thiol, and the effective concentration (C_{eff}) is measured as

$$C_{eff} = \frac{[P_S^S][\text{GSH}]^2}{[P_{SH}^{SH}][\text{GSSG}]} \tag{5}$$

where [GSH] and [GSSG] are the concentrations of reduced and oxidized glutathione, and P_{SH}^{SH} and P_S^S are the concentrations of the thiol and disulfide forms of the protein.

DNSA binding assays

Dansylamide binding constants were determined by fluorescence titration (Fierke et al., 1991). A solution of 0.5 μM protein, 0.1 M $TrisSO_4$, pH 8.0, and 0.1 mM $ZnSO_4$ was prepared, and DNSA was progressively added from either a 36 or 390 μM stock. The binding of DNSA to the protein was monitored by fluorescence energy transfer, with the excitation at 280 nm and emission at 470 nm. The data were fit to the following equation:

$$Fl = BP + EP * \frac{A - \sqrt{A^2 - 4 * [\text{DNSA}]_{tot} * [E]_{tot}}}{2 * [E]_{tot}}$$

$$A = K_D + [E]_{tot} + [\text{DNSA}]_{tot} \tag{6}$$

where *Fl* is the observed fluorescence signal, *BP* is the initial signal, *EP* is the final signal at saturating concentrations of DNSA, K_D is the dissociation constant for the enzyme-inhibitor complex, $[E]_{tot}$ is the total concentration of enzyme (free + bound), and $[\text{DNSA}]_{tot}$ is the total concentration of inhibitor. The titration data (*Fl* vs. $[\text{DNSA}]_{tot}$) were fit using nonlinear regression in *Mathematica* (Wolfram, Inc., Champaign, Illinois) with held fixed at 0.5 μM and *BP*, *EP*, and K_D allowed to vary.

GdnHCl denaturation curves

The equilibrium stabilities of the HCAII variants used in this study were determined by titration with GdnHCl, followed by the ratio of absorbance at 292 and 260 nm (Tweedy et al., 1993). Individual samples were prepared for each denaturant concentration in 0.1 M $TrisSO_4$, pH 8.0, 0.1 mM $ZnSO_4$, and incubated at 25 °C for 24 h. GdnHCl concentrations were determined by refractive index measurements (Pace, 1986). The absorbance data were fit to the following model, derived from Equation 2 in (Tweedy et al., 1993):

$$A_{obs} = A_N + \frac{(A_U - A_N)K_{NI}(Z + K_{IU})}{1 + K_{NI} + K_{NI}K_{IU}}$$

$$K_{NI} = e^{-\frac{\Delta G_{NI,0} - m_{NI}}{RT}}$$

$$K_{IU} = e^{-\frac{\Delta G_{IU,0} - m_{IU}}{RT}} \tag{7}$$

where A_N is the A_{292}/A_{260} ratio for the native state, A_U is the ratio for the unfolded state, *Z* is the fractional absorbance of the *I* state ($Z = (A_I - A_N)/(A_U - A_N)$), $\Delta G_{NI,0}$ is the free energy difference between the native and intermediate states extrapolated to 0M GdnHCl, m_{NI} is the [GdnHCl] dependence of this free energy, and $\Delta G_{IU,0}$ and m_{IU} define the free energy difference between the intermediate and the unfolded state and its [GdnHCl] dependence, respectively.

Acknowledgments

This work was supported by NIH grants GM40532 (T.G.O.) and GM40602 (C.A.F.). R.E.B. was partially supported by an ASSERT award from the Office of Naval Research (N00014-95-1-0951) and J.A.H. was partially supported by an NIH postdoctoral fellowship (GM-17467).

References

Akker FVd, Feil IK, Roach C, Platas AA, Merritt EA, Hol WGJ. 1997. Crystal structure of heat-labile enterotoxin from *Escherichia coli* with increased thermostability introduced by an engineered disulfide bond in the A subunit. *Protein Sci 6*:2644–2649.

Armstrong JM, Myers DV, Verpoorte JA, Edsall JT. 1966. Purification and properties of human erythrocyte carbonic anhydrases. *J Biol Chem 241(21)*: 5137–5149.

Betz SF. 1993. Disulfide bonds and the stability of globular proteins. *Protein Sci 2*:1551–1558.

Betz SF, Marmorino JL, Saunders AJ, Doyle DF, Young GB, Pielak GJ. 1996. Unusual effects of an engineered disulfide on global and local protein stability. *Biochemistry 35*:7422–7428.

Clarke J, Henrick K, Fersht AR. 1995. Disulfide mutants of barnase. 1. Changes in stability and structure assessed by biophysical methods and X-ray crystallography. *J Mol Biol 253*:493–504.

Cooper A, Eyles SJ, Radford SE, Dobson CM. 1992. Thermodynamic consequences of the removal of a disulphide from hen lysozyme. *J Mol Biol 225*:939–943.

Creighton TE. 1983. An empirical approach to protein conformation stability and flexibility. *Biopolymers 22*:49–58.

Doig AJ, Williams DH. 1991. Is the hydrophobic effect stabilizing or destabilizing in proteins? The contribution of disulphide bonds to protein stability. *J Mol Biol 226*:819–835.

Edelhoch H. 1967. Spectroscopic determination of tryptophan and tyrosine in proteins. *Biochemistry 6*:1948–1954.

Eriksson AE, Jones TA, Liljas A. 1988. Refined structure of human carbonic anhydrase II at 2.0 Å resolution. *Proteins Struct Funct Genet 4*:274–282.

Fierke CA, Calderone TA, Krebs JF. 1991. Functional consequences of engineering the hydrophobic pocket of carbonic anhydrase II. *Biochemistry 30*:11054–11063.

Frekgård P-O, Mårtensson L-G, Jonasson P, Jonsson B-H, Carlsson U. 1994. Assignment of the contribution of the tryptophan residues to the circular dichroism spectrum of human carbonic anhydrase II. *Biochemistry 33*: 14281–14288.

Goldenberg DP, Bekeart LS, Laheru DA, Zhou JD. 1993. Probing the determinants of disulfide stability in native pancreatic trypsin inhibitor. *Biochemistry 32*:2835–2844.

Grassetti DR, Murray JF Jr. 1967. Determination of sulfhydryl groups with 2,2′-dithiodipyridine. *Arch Biochem Biophys 119*:41–49.

Håkansson K, Carlsson M, Svensson LA, Liljas A. 1992. Structure of native and apo-carbonic anhydrase II and some of its anion-ligand complexes. *J Mol Biol 227*:1192–1204.

Harrison PM, Sternberg MJE. 1996. The disulphide β-cross: From cystine geometry and clustering to classification of small disulphide-rich protein folds. *J Mol Biol 264*:603–623.

Hazes B, Dijkstra BW. 1988. Model building of disulfide bonds in proteins with known three-dimensional structure. *Protein Eng 2*:119–125.

Hinck AP, Truckses DM, Markley JL. 1996. Engineered disulfide bonds in staphylococcal nuclease: Effects on stability and conformation of the folded protein. *Biochemistry 35*:10328–10338.

Hobohm U, Sander C. 1994. Enlarged representative set of protein structures. *Protein Sci 3*:522–524.

Kalifah RG. 1971. The carbon dioxide hydration activity of carbonic anhydrase. I. Stop-flow kinetic studies on the native human isoenzymes B and C. *J Biol Chem 246*:2561–2573.

Kanaya S, Katsuda C, Kimura S, Nakai T, Kitakuni E, Takamura H, Katayanagi K, Morikawa K, Ikehar M. 1991. Stabilization of *Escherichia coli* ribonuclease H by introduction of an artificial disulfide bond. *J Biol Chem 266*:6038–6044.

Krebs JF, Fierke CA. 1993. Determinants of catalytic activity and stability of carbonic anhydrase II as revealed by random mutagensis. *J Biol Chem 268(2)*:948–954.

Kunkel TA, Roberts JD, Zakour RA. 1987. Rapid and efficient site-specific mutagenesis without phenotypic selection. *Methods Enzymol 154*:367–382.

Lin TY, Kim PS. 1989. Urea dependence of thiol-disulfide equilibria in thioredoxin: Confirmation of the linkage relationship and a sensitive assay for structure. *Biochemistry 28*:5282-05287.

Lin TY, Kim PS. 1991. Evaluating the effects of a single amino acid substitution on both the native and denatured states of a protein. *Proc Natl Acad Sci USA 88*:10573–10577.

Mansfeld J, Vriend G, Dijkstra BW, Veltman OR, Burg BV, Venema G, Ulbrich-Hofmann R, Eijsink VGH. 1997. Extreme stabilization of a thermolysin-like protease by an engineered disulfide bond. *J Biol Chem 272(17)*:11152–11156.

Mårtensson LG, Jonsson B-H. 1993. Characterization of folding intermediates of human carbonic anhydrase II: Probing substructure by chemical labeling of SH groups introduced by site-directed mutagenesis. *Biochemistry 32*: 224–231.

Mårtensson LG, Jonsson B-H, Andersson M, Kihlgren A, Bergenhem N, Carlsson U. 1992. Role of an evolutionarily invariant serine for the stability of human carbonic anhydrase II. *Biochem Biophys Acta 1118*:179–186.

Matsumura M, Matthews BW. 1991. Stabilization of functional proteins by introduction of multiple disulfide bonds. *Methods Enzymol 202*:336–356.

Nair SK, Calderone TL, Christianson DW, Fierke CA. 1991. Altering the mouth of a hydrophobic pocket. Structure and kinetics of human carbonic anhydrase II mutants at residue Val-121. *J Biol Chem 266(26)*:17320–17325.

Nair SK, Elbaum D, Christianson DW. 1996. Unexpected binding mode of the sulfonamide fluorophore 5-dimethylamino-1-naphthalene sulfonamide to human carbonic anhydrase II. *J Biol Chem 271*:1003–1007.

Pabo CO, Suchanek EG. 1986. Computer-aided model-building strategies for protein design. *Biochemistry 25*:5987–5991.

Pace CN. 1986. Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods Enzymol 131*:266–280.

Pace CN, Laurents DV, Thomson JA. 1990. pH Dependence of the urea and guanidine hydrochloride denaturation of ribonuclease A and ribonuclease T1. *Biochemistry 29*:2564–2572.

Pace CN, Shirley BA, Thomson JA, Barnett BJ. 1988. Conformational stability of ribonuclease T1 with zero, one and two intact disulfide bonds. *J Biol Chem 263*:11820–11825.

Petersen MTN, Jonson PH, Petersen SB. 1999. Amino acid neighbors and detailed conformational analysis of cysteines in proteins. *Protein Eng 12(7)*:535–548.

Press WH, Teukolsky SA, Vetterling WT, Flannery BP. 1992. *Numerical recipes in C*. 2nd ed. New York: Cambridge University Press.

Richards FM. 1977. Areas, volumes, packing, and protein structure. *Ann Rev Biophys Bioeng 6*:151–176.

Richardson JS. 1981. The anatomy and taxonomy of protein structure. *Adv Prot Chem 34*:167–339.

Sanger F, Nicklen S, Coulson AR. 1977. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA 74*:5463–5467.

Schwartz H, Hinz H, Mehlich A, Tschesche H, Wenzel HR. 1987. Stability studies on derivatives of bovine pancreatic trypsin inhibitor. *Biochemistry 26*:3544–3551.

Silverman BW. 1986. *Density estimation for statistics and data analysis*. London: Chapman and Hall.

Thompson RB, Jones ER. 1993. Enzyme-base fiber optic zinc biosensor. *Anal Chem 65*:730–734.

Thornton JM. 1981. Disulphide bridges in globular proteins. *J Mol Biol 151*:261–287.

Tweedy NB, Nair SK, Paterno SA, Fierke CA, Christianson DW. 1993. Structure and energetics of a non-proline cis-peptidyl linkage in a proline 202 $\rightarrow$ alanine carbonic anhydrase II variant. *Biochemistry 32*:10944–10949.

Watanabe K, Nakamura A, Fukuda Y, Saito N. 1991. Mechanism of protein folding III. Disulfide bonding. *Biophys Chem 40*:293–301.

Zhou NE, Kay CM, Hodges RS. 1993. Disulfide bond contribution to protein stability: Positional effects of substitution in the hydrophobic core of the two-stranded α-helical coiled coil. *Biochemistry 32*:3178–3187.

Zhu H, Dupureur CM, Zhang X, Tsai M-D. 1995. Phospholipase A_2 engineering. The roles of disulfide bonds in structure, conformational stability and catalytic function. *Biochemistry 34*:15307–15314.