Biochemistry

 Volume 37, Number 23 June 9, 1998

Accelerated Publications

Construction of a Family of Cys_2His_2 Zinc Binding Sites in the Hydrophobic Core of Thioredoxin by Structure-Based Design†

Michael S. Wisz, Christopher Z. Garrett, and Homme W. Hellinga*

Department of Biochemistry, Box 3711, Duke University Medical Center, Durham, North Carolina 27710

Received March 30, 1998; Revised Manuscript Received April 22, 1998

ABSTRACT: A semi-automated, rational design strategy has been used to introduce a family of seven single, mononuclear Cys_2His_2 zinc sites at various locations in the hydrophobic core of *Escherichia coli* thioredoxin, a protein that is normally devoid of metal centers. The electronic absorption spectra of the Co^{II} complexes show that five of these designed proteins bind metal with the intended tetrahedral geometry. The designed sites differ in their metal-binding constants and effects on protein stability. Since these designs are constructed within the same host protein framework, comparison of their behavior allows a qualitative evaluation of dominant factors that contribute to metal-binding and metal-mediated protein stabilization. Metal-binding constants are dominated by steric interactions between the buried, designed coordination sphere and the surrounding protein matrix. Metal-mediated stability is the consequence of differential binding to the native and unfolded states. Increased interactions with the unfolded state decrease the stabilizing effect of metal binding. The affinity for the unfolded state is dependent on the placement of the primary coordination sphere residues within the linear protein sequence. These results indicate that a protein fold can have a remarkably broad potential for accommodating metal-mediated cross-links and suggest strategies for engineering protein stability by constructing metal sites that maximize metal binding to the native state and minimize binding to the unfolded state.

Cysteine-rich zinc-binding sites play an important structural role in a variety of proteins ranging from enzymes to superfamilies of transcription factors (*1*−*3*). These centers form stabilizing cross-links in an intracellular, reducing milieu, analogous to the function of disulfide bonds which typically form in the extracellular, oxidizing environment. Structural zinc sites can be formed either by four-coordinate mononuclear centers in which the primary coordination sphere is composed of combinations of cysteines and histidines, ranging from Cys_4, to Cys_3His, Cys_2His_2, and $CysHis_3$, or by binuclear Cys_6 centers (*4*, *5*). The properties of these metal centers are determined by a subtle interplay between the intrinsic character of the primary coordination sphere (choice of ligand, coordination number, and geometry) and the surrounding protein matrix.

Here, we present an automated, structure-based design approach in which single, tetrahedral Cys_2His_2 coordination spheres are introduced de novo at several different locations in the hydrophobic core of *Escherichia coli* thioredoxin, a protein that is normally devoid of metal centers (*6*, *7*), to investigate the major factors that contribute to formation of a stable metal-binding site of the correct geometry and the

† This work was supported by a grant from the National Institutes of Health R29GM49871 to H.W.H. M.S.W. and C.Z.G. acknowledge support from a training grant from the National Institutes of Health 5T32-GM-08487-04.

* To whom correspondence should be addressed. Phone: (919)-681-5885. Fax: (919)-684-8885. E-mail: hellinga@linnaeus.biochem.duke.edu.

S0006-2960(98)00718-1 CCC: $15.00
Published on Web 05/21/1998

stabilization of a protein by metal-mediated cross-links. Previous studies have established that it is possible to introduce single Cys_2His_2 zinc centers into a protein fold by rational, structure-based strategies (*8*−*10*). In this study, we demonstrate that, by comparing the behavior of a family of designs, all constructed within the same host protein framework, that differ in their zinc-binding affinities and stabilities, it is possible to establish some of the interactions that dominate in the proper formation of the site and metal-mediated stabilization of the protein.

MATERIALS AND METHODS

Solutions were prepared with water measuring 18 MΩ resistance. All solutions, unless containing EDTA,[1] were extracted with dithizone:chloroform mixtures (*11*) to further remove any metals. All other chemicals were used without further purification.

Calculations. All calculations were carried out on a NeXT workstation.

Mutagenesis, Protein Expression, and Purification. The mutant thioredoxins (Trx[ZS]) were constructed by oligonucleotide-directed mutagenesis (*12*) and expressed as C-terminal fusions with Maltose Binding Protein mutants (*13*; New England Biolabs) in which the MBP and Trx[ZS] domains are separated by a flexible peptide linker containing a unique thrombin cleavage site. The fully reduced, metal-free apo form of this fusion protein is readily purified by affinity chromatography, as has been described in detail elsewhere (*14*). The resulting protein was judged pure by SDS−polyacrylamide electrophoresis, and protein concentrations were determined spectrophotometrically by absorbance at 280 nm using an extinction coefficient of $\epsilon_{280} = 82.4$ mM^{-1} cm^{-1} as determined by the method of Edelhoch (*15*), and contained two free thiols as determined with Ellman's reagent (*16*).

Metal Complex Formation. All metal complexes were prepared by direct addition of free metal to protein under anaerobic conditions on a Schlenk line with argon-saturated protein, buffer [TS: 20 mM TrisHCl (pH 7.5), 200 mM NaCl], and metal solutions in water. Co^{II}-binding constants were determined by spectrophotometric titration of $CoCl_2$ (Aldrich) into reduced apo-protein solution, fitting the observed absorbance intensities to a binding isotherm that takes into account all species (*17*):

$$A = A_o + (A_{max} - A_o) \times \frac{([P]_t + [M]_t + K_d) - \sqrt{([P]_t + [M]_t + K_d)^2 - 4[P]_t[M]_t}}{2[P]_t} \quad (1)$$

where A_o, A, A_{max} are the absorbances at a given wavelength, in the absence, presence, and at saturation of metal respectively, $[P]_t$ is the total concentration of protein, $[M]_t$ is the total concentration of metal (i.e., free and bound) at each titration point, and K_d is the dissociation constant for the metalloprotein complex. Zn^{II} competition experiments were carried out by titrating in $ZnCl_2$ (Aldrich) into the Co^{II}−protein complex, monitoring bleaching of the absorbance spectrum.

Electronic Absorbance Spectroscopy. All electronic absorbance spectra were recorded at 25 °C on a thermostated Hitachi U-2000. Samples were prepared in TS buffer. Anaerobic spectra were recorded in Teflon-sealed cuvettes, which were deoxygenated by purging with argon on a dual manifold Schlenk line.

Protein Stability. Stability measurements were performed on purified Trx[ZS] protein prepared from the apo-MBP::Trx[ZS] fusion by proteolytic cleavage. Thiols were protected by reversible covalent modification with DTNB throughout the purification procedure (DTNB added in 2-fold excess relative to protein and incubated for 10 min at room temperature). Cleavage of protected apo-MBP::Trx[ZS] (∼300 mg in 50 mL of TS buffer) was carried out by overnight incubation at room temperature with 100 units of plasminogen-free bovine serum thrombin (CalBiochem) in the presence of 2.5 mM $CaCl_2$ and the extent of cleavage determined by SDS−PAGE (typically better than 90% complete). The cleavage reaction was passed over a gel filtration column (P-6, Bio-Rad) to remove excess DTNB and TNB^-, followed by denaturation by addition of solid urea to 8 M. After concentration to a final volume of 10 mL by ultrafiltration (Amicon, YM-10 membrane), Trx[ZS] was separated from MBP using preparative gel electrophoresis (Bio-Rad) under denaturing conditions (9% polyacrylamide, 8 M urea, 2.5 mM EDTA, 25 mM borate, and 50 mM Tris, pH 8.7, 15 °C, 400 V, 12 h; fractions collected by continuous flow of buffer). Protected apo-Trx[ZS] fractions were identified by SDS−PAGE, pooled, concentrated and deprotected by incubation with 50 mM DTT, 50 mM EDTA, and 3.2 M guanidinium chloride (GdnHCl) (1 h, room temperature), and finally purified by gel filtration (P-6, Bio-Rad). The presence of two free thiols was then verified by Ellman's method (*16*).

The stability of the isolated Trx[ZS] was measured at 25 °C by titration with GdnHCl, following the molar ellipticity at 222 nm (10 μM protein in argon purged TS buffer) using an Aviv 62DS circular dichroism spectrophotometer with a nitrogen-flushed chamber, equipped with an automated Microlab 500 titrator (Hamilton). GdnHCl concentrations were measured by their refractive index (*18*). The stability of the apo-protein was measured in the absence of zinc, while the stability of the metalloprotein was measured using the zinc complex (formed by addition of 11 μM $ZnCl_2$). The free energies of unfolding were determined by fitting the data to a two-state model and linearly extrapolating to 0 M denaturant (*18*):

$$\Delta G_U = \Delta G_U(D) - m[D] \quad (2)$$

where ΔG_U is the free energy of unfolding in the presence of denaturant, $\Delta G_U(D)$ the free energy at 0 M denaturant, and m a linear constant.

[1] Abbreviations: DTNB, 5,5′-dithiobis(2-nitrobenzoic acid); DTT, 1,4-dithiothreitol; EDTA, ethylenediaminetetraacetic acid; GdnHCl, guanidinium hydrochloride; ΔG_U, free energy of unfolding; IPTG, isopropyl β-D-thiogalactopyranoside; LMCT, ligand-to-metal charge transfer; MBP, maltose binding protein; Polymin-P, polyethylenimine; SDS−PAGE, sodium dodecyl sulfate/polyacrylamide gel electrophoresis; TrisHCl, tris(hydroxymethyl)aminomethane hydrochloride; Trx, thioredoxin; Trx[ZS], mutant thioredoxin containing a designed Cys_2His_2 zinc binding site.

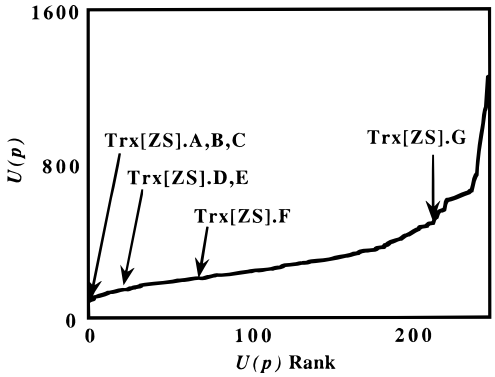


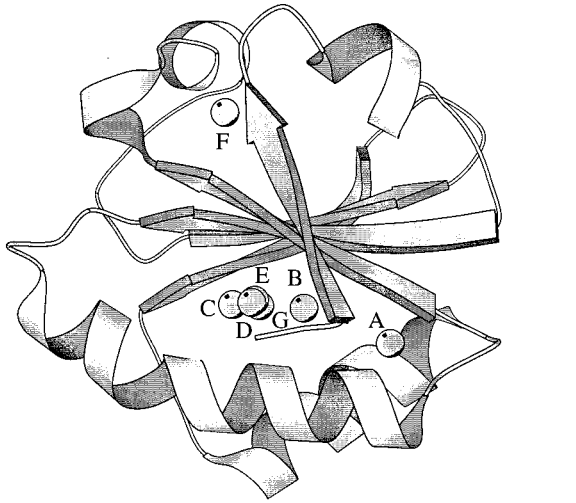
FIGURE 1: Rank ordering of the all the designed sites (*x*-axis) by their *U*(p) scores (weighted least-squares deviation from ideal tetrahedral geometry). The positions within this landscape of the seven designs constructed in this study are indicated by the arrows.

RESULTS

Structure-Based Design of the Cys_2His_2 Zinc Centers. The locations of Cys_2His_2 primary coordination spheres which are intended to form tetrahedral zinc centers were predicted by the Dezymer automated protein design program. Dezymer systematically examines a three-dimensional protein structure to identify sites where the appropriate amino acids can be placed by mutagenesis to introduce a coordination sphere of predetermined geometry while maintaining steric compatibility with the protein fold (*19*). This program has also been used successfully to construct a variety of different metal centers in the hydrophobic core of thioredoxin, including a monunuclear iron−sulfur center (*14*), an Fe_4S_4 iron−sulfur cluster (*20*), an Fe-dependent superoxide dismutase (*21*), and a Blue Copper analogue (*22*).

The Dezymer program predicts the mutations required to construct metal-binding sites using a simple description of the molecular interactions. The tetrahedral zinc primary coordination sphere is specified in geometrical terms (bond lengths, angles, and torsional relationships). The geometry around the metal is described by idealized tetrahedral angles. The geometry of the interaction between the metal and the cysteine or histidine side chains is derived from observations of natural protein structures (*23*). The zinc is allowed to coordinate to either the N_δ or the N_ϵ nitrogens of the histidine ligands and is constrained to be coplanar with the imidazole ring. The C_α-C_β-S_γ-Zn^{II} torsion for the cysteine interaction is unconstrained. The cysteine and histidine side chains are represented as sets of favorable rotamers (*24*). Nonbonded interactions are described by a simple hard-sphere model. In the first instance, sites are selected based on their intrinsic geometry and compatibility with the backbone fold of the protein alone. The outcome of this search is expressed as a *U*(p) score that reflects a weighted least-squares deviation from ideal geometry [$U(p) = 0$ corresponds to a perfect geometry; $U(p) > 0$ is nonideal; an arbitrary cutoff $U(p) = 1300$ was chosen as the limit of acceptability in this particular experiment].

Using this procedure, 250 unique four-residue Cys_2His_2 mutations were identified in the high-resolution X-ray structure of *E. coli* thioredoxin (*7*), which were rank-ordered according to their *U*(p) score (Figure 1). On the basis of visual inspection, seven sites that were buried (judged by their lack of static solvent accessibility, as calculated for each



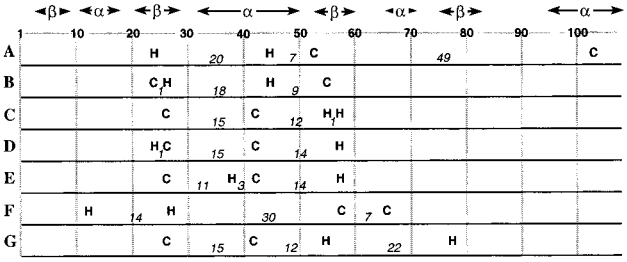
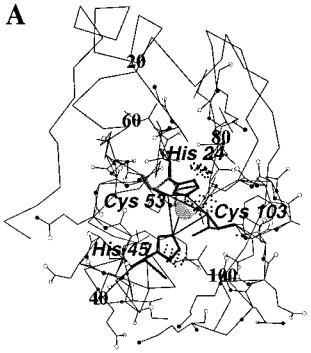


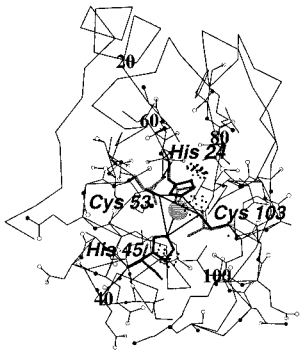
FIGURE 2: The distribution of the designed sites within the hydrophobic core of thioredoxin (letters identify each design). (Top) Distribution of the sites within the protein structure. (Bottom) Distribution along the linear sequence (residue number indicated at top) along with their secondary structural location (H, histidine; C, cysteine; numbers in italics indicate inter-residue spacing within the primary coordination sphere).

model) and that fit in reasonably well with the surrounding residues (assessed subjectively by the apparent absence of severe steric overlap) were selected for experimental evaluation. These designs (Trx[ZS].A−G) were deliberately chosen for their variation in *U*(p) (Figure 1) and are distributed in various locations in the hydrophobic core of thioredoxin (Figure 2 and Table 1). Furthermore, the selected designs vary in the steric compatibility between the introduced primary coordination sphere and the surrounding protein matrix (Figure 3), and no attempt was made at this stage to introduce additional mutations to improve these interactions. The core of thioredoxin is divided into two halves, separated by a central β-sheet. Other than Trx[ZS].F, all the designs are situated in one half, on the same side of the β-sheet that the active-site Cys32−Cys35 disulfide bridge of wild-type thioredoxin is normally located. Trx[ZS].C, D, E, and G all share a common core formed by Cys26 and Cys42. They also share subsets of histidine attachment positions in various combinations.

Construction and Purification of Designed Proteins. Mutants were constructed by oligonucleotide-directed mutagenesis in a thioredoxin background, Trx[ZS]0, that contains four additional mutations relative to wild-type: Asp2Ala mutation to remove an adventitious surface metal-binding site (*25*, *6*), Asp26Leu to increase protein stability (*26*), and Cys32Ser and Cys35Ser to remove the native disulfide bridge. The proteins are expressed as a cleavable fusion protein at the C-terminus of MBP using a derivative of the expression vector pMAL-c2 (*13*; New England Biolabs) in which a flexible peptide linker containing a

A



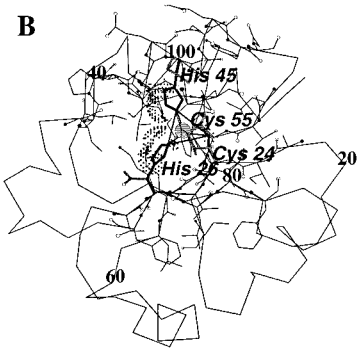


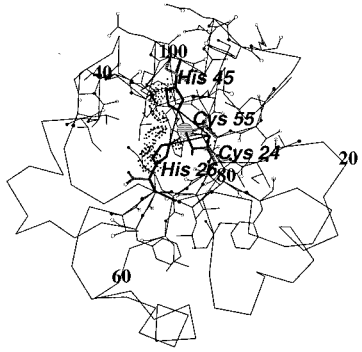
Model Parameters

Side-chain dihedrals		Bond lengths (Å)	
Cys53 χ_1	-50.5	Cys53 S_γ-Zn	2.6
Cys103 χ_1	-174.9	Cys103 S_γ-Zn	2.1
His24 χ_1	-63.3	His24 N_ε-Zn	2.1
His45 χ_1	-162.0	His45 N_ε-Zn	2.6
His24 χ_2	-53.4		
His45 χ_2	84.1		

Zn tetrahedron		Imidazole-Zn coplanarity	
Cys53 S_γ-Zn- Cys103 S_γ	111.0	His24: C_γ- N_δ- N_ε-Zn	-175.4
Cys53 S_γ-Zn- His24 N_ε	106.7	His45: C_γ- N_δ- N_ε-Zn	173.3
Cys53 S_γ-Zn- His45 N_ε	113.1		
Cys103 S_γ-Zn- His24 N_ε	110.8	Thiol-Zn geometry	
Cys103 S_γ-Zn- His 45 N_ε	108.5	Cys53: C_β- S_γ-Zn	106.4
His24 N_ε-Zn- His45 N_ε	106.8	Cys103: C_β- S_γ-Zn	113.1

B

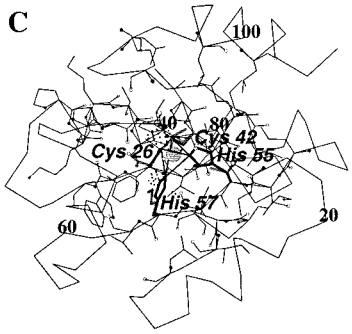


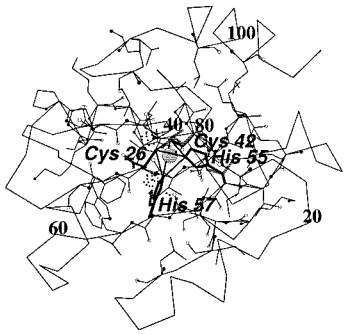


Side-chain dihedrals		Bond lengths (Å)	
Cys24 χ_1	46.6	Cys24 S_γ-Zn	2.7
Cys55 χ_1	-174.2	Cys55 S_γ-Zn	1.4
His26 χ_1	-53.3	His26 N_ε-Zn	1.9
His45 χ_1	69.2	His45 N_ε-Zn	2.0
His26 χ_2	-75.9		
His45 χ_2	-57.8		

Zn tetrahedron		Imidazole-Zn coplanarity	
Cys24 S_γ-Zn- Cys55 S_γ	108.0	His26: C_γ- N_δ- N_ε-Zn	176.0
Cys24 S_γ-Zn- His26 N_ε	112.4	His45: C_γ- N_δ- N_ε-Zn	179.2
Cys24 S_γ-Zn- His45 N_ε	111.3		
Cys55 S_γ-Zn- His26 N_ε	106.2	Thiol-Zn geometry	
Cys55 S_γ-Zn- His 45 N_ε	106.8	Cys24: C_β- S_γ-Zn	114.7
His26 N_ε-Zn- His45 N_ε	111.8	Cys55 : C_β- S_γ-Zn	109.7

C

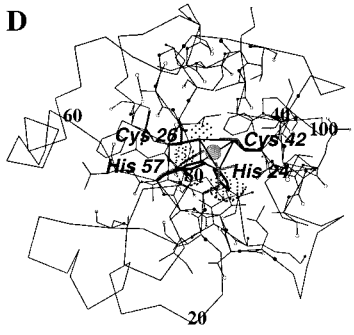


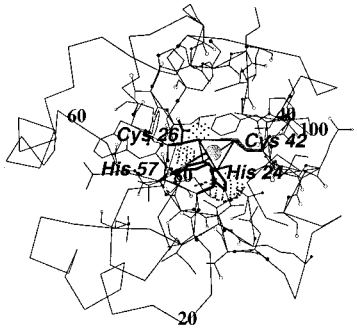


Side-chain dihedrals		Bond lengths (Å)	
Cys26 χ_1	-78.8	Cys26 S_γ-Zn	2.0
Cys42 χ_1	-74.8	Cys42 S_γ-Zn	2.1
His55 χ_1	-159.2	His55 N_ε-Zn	1.5
His57 χ_1	-75.7	His57 N_ε-Zn	2.3
His55 χ_2	65.2		
His57 χ_2	177.9		

Zn tetrahedron		Imidazole-Zn coplanarity	
Cys26 S_γ-Zn- Cys42 S_γ	113.0	His55: C_γ- N_δ- N_ε-Zn	-177.3
Cys26 S_γ-Zn- His55 N_ε	109.1	His57: C_γ- N_δ- N_ε-Zn	173.6
Cys26 S_γ-Zn- His57 N_ε	105.1		
Cys42 S_γ-Zn- His55 N_ε	107.3	Thiol-Zn geometry	
Cys42 S_γ-Zn- His57 N_ε	113.8	Cys26: C_β- S_γ-Zn	108.3
His55 N_ε-Zn- His57 N_ε	108.4	Cys42: C_β- S_γ-Zn	108.3

D





Side-chain dihedrals		Bond lengths (Å)	
Cys26 χ_1	-84.2	Cys26 S_γ-Zn	1.9
Cys42 χ_1	-80.7	Cys42 S_γ-Zn	2.2
His24 χ_1	102.3	His24 N_ε-Zn	2.3
His57 χ_1	-69.8	His57 N_ε-Zn	2.3
His24 χ_2	-104.2		
His57 χ_2	155.9		

Zn tetrahedron		Imidazole-Zn coplanarity	
Cys26 S_γ-Zn- Cys42 S_γ	110.4	His24: C_γ- N_δ- N_ε-Zn	-176.6
Cys26 S_γ-Zn- His24 N_ε	106.7	His57: C_γ- N_δ- N_ε-Zn	175.5
Cys26 S_γ-Zn- His57 N_ε	105.3		
Cys42 S_γ-Zn- His24 N_ε	110.6	Thiol-Zn geometry	
Cys42 S_γ-Zn- His57 N_ε	112.5	Cys26: C_β- S_γ-Zn	109.0
His24 N_ε-Zn- His57 N_ε	111.1	Cys42: C_β- S_γ-Zn	112.3

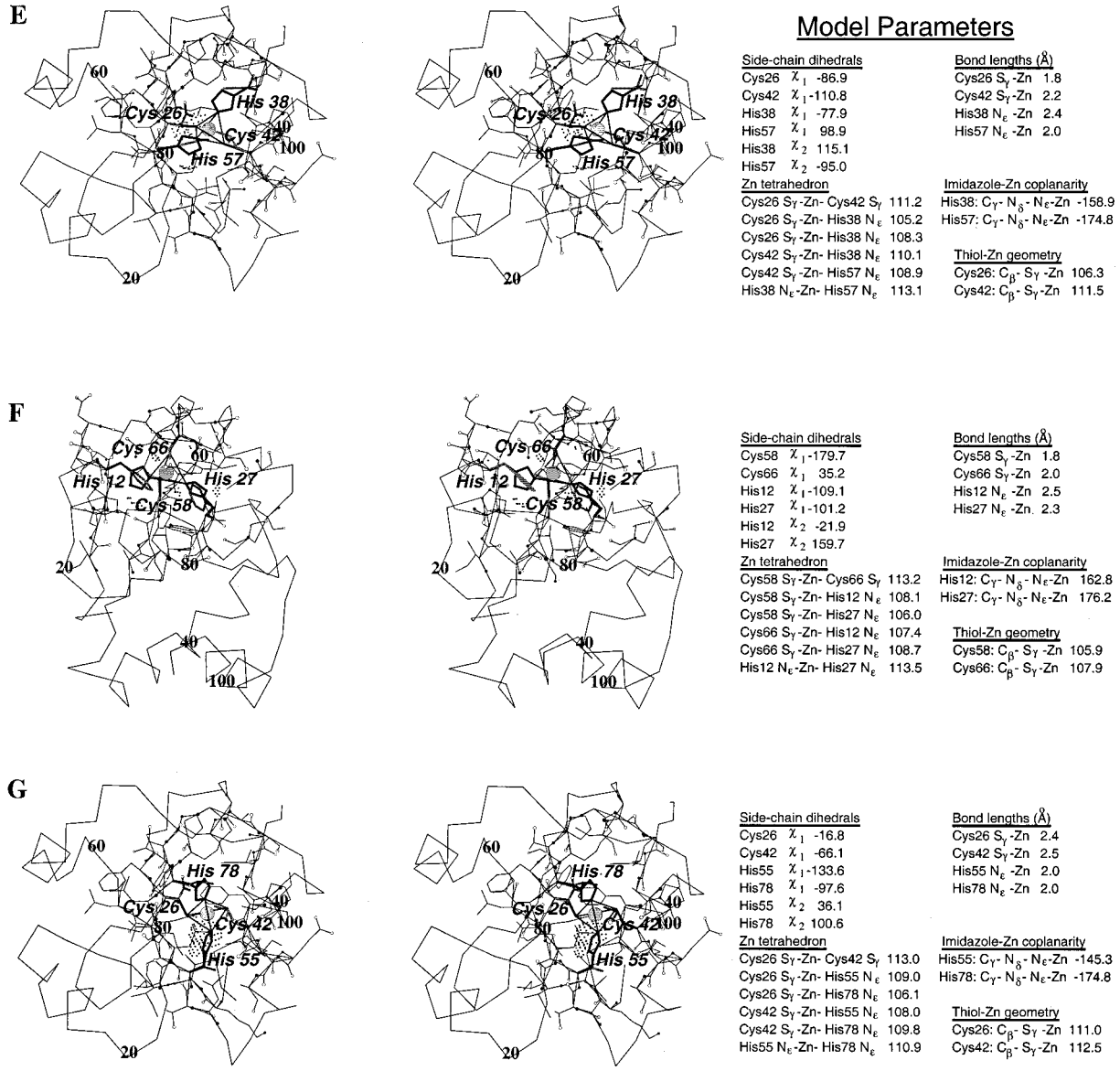


FIGURE 3: Detailed stereoviews of the Trx[ZS] design models (each panel has its own viewing angle). (Thin lines) C_α-trace of the entire protein and residues within the immediate vicinity surrounding the designed site (open circle, oxygen; closed circle, nitrogen). (Thick lines) primary coordination sphere. Gray sphere, zinc. Medium lines, tetrahedron of the primary coordination sphere. Gray lines, wild-type residues replaced by the Cys_2His_2 mutations. Dotted spherical segments, parts of the hard-sphere surface of an atom in the designed center that forms in unfavorable steric interaction with its surrounds. The extent of steric overlap is indicated by the number of dots shown.

unique thrombin cleavage site (Ile-Glu-Gly-Arg-Ile-Ser-Glu-Phe-Gly-Ser-Leu-Val-Pro-Arg-Gly-Ser-Gly-Gly-Ser-Gly-Gly-Ser) has been placed between the MBP and Trx domain. The resulting fusion proteins are readily purified in a single step on an amylose affinity column. All electronic absorption spectra and metal-binding studies were carried out on the intact fusion protein. Protein stability was determined by chemical denaturation of the Trx[ZS] domain prepared from the fusion protein by proteolytic cleavage.

Metal Binding. Five out of the seven proteins were folded in the presence of metal (vide infra). The electronic absorption spectra of the Co^{II} complexes of these five proteins were measured to determine the coordination geometry of the designed primary coordination sphere. A representative spectrum is shown in Figure 4, the others are summarized in Table 1. In all five proteins, the Co^{II} spectrum showed $d-d$ transitions in the 500−700 nm region of energies, intensities, and line-shape characteristic of tetrahedral coordination geometry (*27*). Furthermore, intense absorption bands could be observed in the 320 nm region, consistent with LMCT bands indicative of thiol coordination (*27*). Addition of Co^{II} to the starting construct lacking a designed Cys_2His_2 center (Trx[ZS]0) or to Trx[ZS].D and G (the designs that are unfolded in the presence of metal) showed no evidence of formation of a Co^{II} complex. Thus, if a design is folded in the presence of metal, the site with the correct geometry forms.

Co^{II}-binding constants were determined by direct spectrophotometric titration (inset of Figure 4). Two classes of sites could be distinguished (Table 1): tight binders (K_d = 2−4 μM) and medium binders (K_d = 150−250 μM). Binding of Zn^{II} was determined by displacement of the Co^{II} complex (inset of Figure 4). In all cases, Zn^{II} was found to form a 1:1 stoichiometric complex with the protein. Zn^{II} was found to bind at least 2−3 orders of magnitude more tightly than Co^{II}, as expected (*28*). It was therefore impos-

Table 1: Metal Binding and Stability Summary[a]

site	mutations[b]	$\Delta^{A}G_U$ (kcal/mol)	m_A (kcal/mol/M)	$\Delta^{Zn}G_U$ (kcal/mol)	m_{Zn} (kcal/mol/M)	$\Delta\Delta G_U$ (kcal/mol)	Co^{II} complex, λ_{max} nm (ϵ mM^{-1} cm^{-1})	Co^{II} K_d (μM)	$\Delta\Delta G_{solv}$ (kcal/mol)
0[c]		11.8 ± 0.2	3.8 ± 0.1	12.2 ± 0.2	3.9 ± 0.1	0.3			0
A	L53C, L103C, L24H, I45H	partially unfolded[d]		5.2 ± 0.2	4.0 ± 0.1	5.2[f]	320(1.81), 563(0.25), 622(0.25), 661(0.18sh)	154	4.8
B	L24C, V55C, L26H, I45H	partially unfolded[d]		3.7 ± 0.2	2.8 ± 0.1	3.7[f]	320(1.93), 546(0.30), 630(0.31), 662(0.29sh)	158	2.9
C	L26C, L42C, V55H, K57H	4.4 ± 0.1	3.5 ± 0.1	6.4 ± 0.1	3.1 ± 0.1	2.0	320(3.63), 622(0.40sh), 625(0.41), 675(0.48)	2	1.5
D	L26C, L42C, L24H, K57H	partially unfolded[d]		partially unfolded[d]		0			2.3
E	L26C, L42C, L38H, K57H	6.0 ± 0.1	3.6 ± 0.1	4.5 ± 0.1	2.5 ± 0.1	−1.5	320(1.73), 555(0.18sh), 630(0.50), 675(0.47)	213	1.8
F	L58C, T66C, F12H, F27H	4.1 ± 0.2	4.2 ± 0.2	6.5 ± 0.1	2.5 ± 0.1	2.4	320(1.56), 578(0.28), 622(0.32), 666(0.24sh)	4	4.2
G	L26C, L42C, V55H, L78H	unfolded[e]		unfolded[e]					2.7

[a] $\Delta^{A}G_U$ and $\Delta^{Zn}G_U$ are the free energies of unfolding for the apo-protein and protein−zinc complexes, respectively; m_A and m_{Zn} are the *m*-values (eq 2) of the apo-protein and protein−zinc complexes, respectively; $\Delta\Delta G_U$ is the difference in the unfolding free energies between the protein−zinc complex and the apo-protein (metal-mediated stabilization or destabilization); $\Delta\Delta G_{solv}$ is the calculated difference in free energy of solvation (*34*) of each mutant construct relative to Trx[ZS].0. [b] Mutations relative to Trx[ZS].0 background. [c] The thioredoxin background (Asp2Ala, Asp26Leu, Cys32Ser, and Cys35Ser relative to wild-type thioredoxin) in which the Trx[ZS].A-G mutants were constructed. [d] No pretransition baseline for the folded form is observed, precluding accurate quantitation of $\Delta^{A}G_U$. [e] No data was obtained for the stability of this mutant, due to low levels of expression as well as aggregation in the presence of metal. [f] This value is a minimum estimate of $\Delta\Delta G_U$, using 0 kcal/mol as an upper estimate for $\Delta^{A}G_U$.

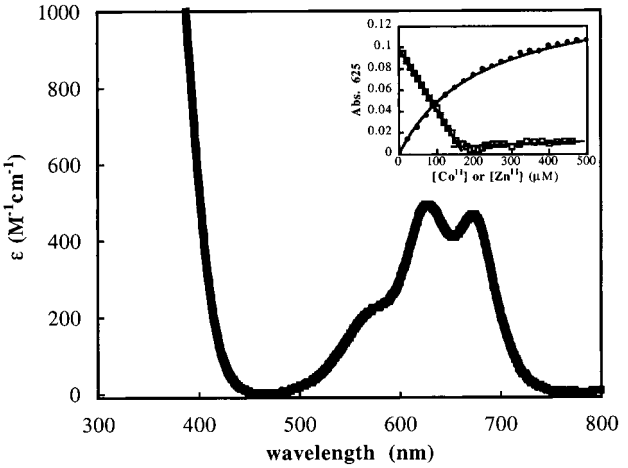


FIGURE 4: Electronic absorbance spectrum of the Co^{II} complex of Trx[ZS].E. (Inset) spectrophotometric titration of Co^{II} (circles) to 200 μM apo-protein and Zn^{II} (squares) into the Co^{II} complex (binding constants determined by fitting to eq 1).

sible to obtain accurate zinc-binding constants by direct spectrophotometric titration, although upper limits could be established by fits to eq 1.

Protein Stability. The stabilities of purified Trx[ZS] domains were determined by titration with GdnHCl in the presence and absence of zinc, using the circular dichroism signal at 222 nm to measure the fraction of native protein. The resulting curves were fit to a two-state model, and the free energies of unfolding were determined by linear extrapolation to 0 M denaturant (eq 2). In all cases, the mutants were less stable than the reference protein, Trx[ZS]0, regardless of metal addition. Four classes of mutants could be distinguished based on their stabilities and the effect of metal addition: 1, apo-protein stable, addition of metal further stabilizes the protein (Trx[Zs].C and F; Figure 5); 2, apo-protein unfolded, addition of metal forms folded protein (Trx[ZS].A and B; Figure 5); 3, apo-protein stable, addition of metal destabilizes (Trx[ZS].E; Figure 5); 4, protein is unfolded, regardless of metal addition (Trx[ZS].D and G; not shown). The stabilities for all mutants are summarized in Table 1.

DISCUSSION

Design Methodology. Rational design is rapidly emerging as a powerful approach to investigate general principles in the relationship between protein structure and function (*29*−*31*). One strategy is to start with the simplest possible molecular model and then to iteratively improve the designs by introducing new layers of complexity guided by experimental results (*22*, *29*). Here, we expand upon this approach and also start with a very simple model, but rather than implementing a single design which is iteratively worked on, we build a small family of related proteins. We demonstrate that, by comparing the experimental results between members of this family, trends can be obtained that give qualitative insights into the contributions dominating the formation and stability of Cys_2His_2 zinc centers. These trends can then be further tested by subsequent incorporation into the next design iteration.

A simple, geometrical representation was used to model the tetrahedral Cys_2His_2 primary coordination sphere. Nonbonded interactions were represented by a hard-sphere model. Sites were modeled to be approximately sterically compatible with the thioredoxin backbone fold, which was kept absolutely fixed during the refinement procedure. No attempt was made in this first design cycle to optimize the geometry of the sites by allowing the backbone to relax. The intent of this design strategy is to exploit deliberately the variation of defects in the designs to obtain a range of binding behaviors to try and uncover dominant effects in the formation of stable sites and metal-mediated protein stabilization.

Inverse Folding Constraints. The automated design algorithm constructs sites that are predicted to be approximately sterically compatible with the surrounding protein matrix into which they have been introduced (*32*). The degree of packing defects introduced by constructing

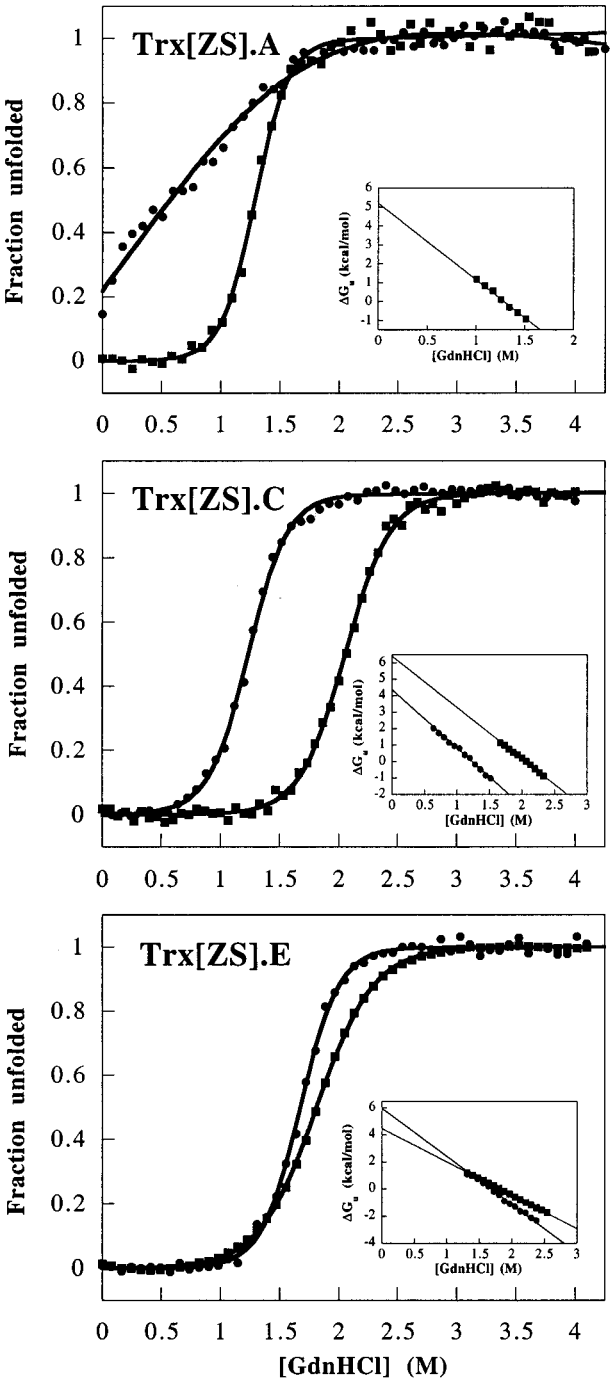


FIGURE 5: GdnHCl unfolding curves for three representative cases in the presence (squares) and absence (circles) of metal. (Insets) linear free energy extrapolations to 0 M GdnHCl. Trx[ZS].A: apo-protein partially folded in the absence of metal; stabilized by metal. Trx[ZS].C: apo-protein stable; stabilized by metal. Trx[ZS].E: apo-protein stable, but destabilized by metal.

the Cys_2His_2 mutations in the hydrophobic core can be assessed experimentally by comparing the free energies of unfolding of the apo-protein, $\Delta^A G_U$ (Table 1), of the various mutants. If consideration of steric compatibility were not an issue, then changes in the $\Delta^A G_U$ values of these mutants in the protein interior should be primarily correlated with changes in the hydrophilicity of the replacements (*33*). This is clearly not the case, since there is no simple correlation between the calculated change in free energy of solvation (*34*), $\Delta\Delta G_{solv}$ (relative to Trx[ZS]0.0, Table 1), and the $\Delta^A G_U$. For instance, Trx[ZS].A and Trx[ZS].F are both predicted to be destabilized by approximately the same amount, based on their $\Delta\Delta G_{solv}$ values, yet Trx[ZS].A is partially unfolded in the absence of metal whereas Trx[ZS].F is still quite stable.

The models indicate that the designs do not fit perfectly into the surrounding protein matrix and are predicted to form steric clashes with neighboring side-chain residues of differing degrees of severity in the various designs (Figure 3). In the actual proteins, these steric imperfections are presumably compensated for by a variety of structural relaxation mechanisms involving both backbone and side-chain movements. It is therefore not straightforward to rationalize quantitatively the $\Delta^A G_U$ values based on the molecular models. One qualitative trend is clear, however: models of designs that show few bad steric interactions have folded apo-proteins (Trx[ZS].C, E, and F), whereas the apo-proteins of models that indicate significant packing defects are unfolded (Trx[ZS].A, B, D, and G). Thus, if steric compatibility is predicted to be maintained in a design, a stable apo-protein is formed reliably. The complex mechanisms by which packing defects are accommodated through structural relaxations make the experimental outcome of designs with known packing imperfections quite uncertain. For instance, Trx[ZS].B and Trx[ZS].D have similar degrees of steric overlaps, yet Trx[ZS].B forms a stable metal complex, whereas Trx[ZS].D remains unfolded in the presence of metal. These results clearly demonstrate that the inverse folding constraints can be reliably satisfied only by avoiding the introduction of unfavorable steric interactions and, therefore, illustrates that detailed packing interactions dominate the stability of a fold, as has also been observed in other experiments on repacking of protein hydrophobic cores (*33*).

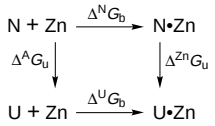
The Geometry and Stability of the Designed Metal Centers. Provided that the protein can form a stable fold in the presence of metal (which is the case in five of the seven designs), the Co^{II} complexes all adopt a tetrahedral geometry. The overall coordination geometry of the metal centers is therefore not affected by packing defects in the designs. However, the stability of a metal center (i.e., its binding constant) is strongly dependent on packing defects. Trx[ZS].C and F have models that indicate few predicted steric overlaps, their apo-proteins are stable, and they form tight complexes with Co^{II} [$K_d(Co^{II}) = 2-4\ \mu M$], very similar in magnitude to natural Cys_2His_2 zinc-binding proteins (*2*). Trx[ZS].A and B are predicted to have significant steric overlap, their apo-proteins are unfolded, and they form Co^{II} complexes that are 100-fold weaker ($\sim$150 μM). Trx[ZS].E is the apparent exception to this simple analysis,[2] since it is not predicted to form many bad steric contacts, its apo-protein is very stable, and yet it has the worst binding constant for Co^{II} (213 μM). Examination of the model shows, however, that there are unfavorable steric interactions *within* the primary coordination sphere itself, between residues Cys26 and His57. This suggests that in this design formation of the tetrahedral geometry imposes "strain" on the coordinating ligands, which might account for the weaker metal binding.

These results indicate that the geometry of the coordination sphere is dictated by the metal center and that the thermodynamic price paid for the necessary structural relaxation

[2] Trx[ZS].D and G are not considered in this analysis, since they do not form stable, folded metalloproteins.

of the protein to accommodate the tetrahedral center is reflected in the binding constant of the metal. This illustrates the balance of forces in the interplay between the intrinsic properties of the metal center and the surrounding protein matrix.

Metal-Mediated Protein Stabilization. Metals can potentially bind both to the native and unfolded state of the protein, resulting in the following thermodynamic linkage relationship (*35*):



where N and U are the apo forms of native and unfolded states, respectively, and N·Zn and U·Zn their metal complexes, $\Delta^N G_b$ and $\Delta^U G_b$ the metal-binding constants of the two states, and $\Delta^A G_U$ and $\Delta^{Zn} G_U$ the free energies of unfolding of the apo- and metallo-proteins. In this linkage relationship, the free-energy change due to metal-mediated protein stability, $\Delta\Delta G_U = \Delta^{Zn} G_U - \Delta^A G_U$, is related to the differential binding of the metal in the native and unfolded states, $\Delta\Delta G_b = \Delta^N G_b - \Delta^U G_b$, by $\Delta\Delta G_U = -\Delta\Delta G_b$. In the limit where there is no binding to the unfolded state ($\Delta^U G_b = 0$), metal-mediated stabilization is entirely due to binding of metal to the native state: $\Delta\Delta G_U = -\Delta^N G_b$. If $\Delta\Delta G_U < -\Delta^N G_b$, we can therefore infer that there must be significant binding of metal to the unfolded state.

Although the determination of the Zn^{II}-binding constants by competitive spectrophotometric titration of the Co^{II} complexes does not allow for accurate quantitation due to their high affinities, we can obtain upper limits for their values which are at least 100-fold stronger than the Co^{II} complex, as expected (*28*). This puts the $K_d(Zn^{II})$ in the 10 nM range for Trx[ZS].C and F and in the 1 μM range for Trx[ZS].A, B, and E (Table 1), corresponding to $\Delta^N G_b \approx -11$ kcal/mol and ~ -8 kcal/mol, respectively. This far exceeds any of the $\Delta\Delta G_U$ obtained for the mutants, and we therefore conclude that there must be significant binding to the unfolded state in all of the mutants.

Further indirect evidence for binding of metal to the unfolded state is provided by more detailed analysis of the unfolding transitions. The factor *m* in the linear free-energy relationship of the unfolding free energy (eq 2) is proportional to the change in solvent-accessible surface area upon denaturation (*36*). Changes in *m* values of a given protein obtained under different conditions are indicative of changes in the nature of the unfolded state. For instance, a decrease in *m* upon disulfide oxidation is consistent with formation of a more compact denatured state due to the introduction of cross-links (*37*). A similar effect is seen in the presence of metal in these designed proteins. Comparison of m_A and m_{Zn} for Trx[ZS].C, E, and F (Table 1) shows that in each case $m_{Zn} < m_A$, indicating the formation of a more compact unfolded state in the presence of zinc, presumably as a result of the introduction of zinc-mediated cross-links in the unfolded state. Interestingly, this is not the case for the wild-type protein (Trx[ZS]0), suggesting that zinc binding to the unfolded state is primarily the consequence of the introduced mutations, rather than to other residues.[3]

Metal binding to the unfolded state appears to be the dominant effect that limits the degree of metal-mediated stabilization, since there is no correlation between the metal affinity to the native state and stabilization. For instance, one of the weaker metal binders, Trx[ZS].A, shows significantly better stabilization than either of the strong binders (Trx[ZS].C and F) and has a significantly larger m_{Zn} value than any of the other designs, consistent with a relatively reduced level of Zn^{II} binding to its unfolded state.

It is unclear at this stage which is the dominant factor that determines binding to the unfolded state. A simple consideration is the spacing between the residues that coordinate to the metal, since increased length of the loops separating the metal-mediated cross-link entropically disfavors binding (*38*). This may account for the apparently low metal affinity of the unfolded state observed in Trx[ZS].A, since this protein has a significantly larger loop separating one of the cysteines from the other residues in the coordination sphere compared to the other designs (Figure 2). However, in none of these other designs is there any obvious correlation between interresidue loop length and variation in $\Delta\Delta G_U$. Furthermore, there is no a priori reason only the designed Cys_2His_2 residues should participate in metal binding in the unfolded state.[3] There may well be a mixture of residues involved, consisting of one or more of the designed, mutant residues as well as other, unchanged, fortuitously nearby residues, even though there is no evidence of metal binding to the unfolded state in wild-type thioredoxin. It is therefore impossible to rationalize the differential binding effects in the different mutants in the absence of more experimental information on the nature of metal binding in the denatured state. These studies are in progress.

The effect of unfavorable steric interactions can also adversely influence metal-mediated protein stabilization. This is clearly observed in the case of Trx[ZS].E, which is *destabilized* upon metal binding ($\Delta\Delta G_U = -1.5$ kcal/mol). As discussed above, this is probably the consequence of unfavorable interactions between the residues within the primary coordination sphere, which manifests itself only when metal binding imposes a relatively rigid geometry upon these residues.

Conclusions. Seven designed Cys_2His_2 tetrahedral zinc centers were constructed at various locations in the hydrophobic of *E. coli* thioredoxin using a structure-based, semi-automated design approach, embodied in the Dezymer program (*19*). Five of the seven designs resulted in the successful construction of tetrahedral metal-binding sites, illustrating the remarkably broad potential for the introduction of such cross-links as well as other metal centers (*14*, *20*−*22*) in the hydrophobic core of a protein that possesses no a priori structural adaptations to accommodate them. A comparative analysis of the designs in this family revealed some of the dominant factors that control the formation of these centers and metal-mediated protein stabilization, illustrating the interplay between the intrinsic properties of the metal coordination sphere and the surrounding protein matrix. Formation of a tetrahedral coordination geometry

[3] Many possible zinc ligands are potentially available for coordination in the unfolded state, including main-chain amides and functional groups in several of the amino acid side chains.

is dominated by the intrinsic properties of the metal coordination sphere. The affinity of the native state for metal is strongly dependent on the steric complementarity between the designed center and the neighboring residues. Metal-binding constants are therefore dominated by interactions with the surrounding protein matrix. Metal-mediated protein stabilization involves a balance between binding to the native (favorable) and unfolded (unfavorable) states. Binding to the unfolded state can be a dominant destabilizing factor.

Two clear strategies can therefore be pursued in the next iteration of the design cycle: stabilization of the binding site by improving packing interactions ("target state optimization") and destabilization of binding to the denatured state ("negative design"). It is becoming increasingly clear that negative design is a crucial, if often cryptic, aspect of protein function (*22*, *29*, *39*, *40*).

ACKNOWLEDGMENT

We thank V. J. Zhang for assistance with mutagenesis and Dr. D. E. Benson for discussion.

REFERENCES

1. Berg, J. M. (1993) *Curr. Opin. Struct. Biol. 3*, 11−16.
2. Berg, J. M., and Godwin, H. A. (1997) *Annu. Rev. Biophys. Biomol. Struct. 26*, 357−371.
3. Berg, J. M., and Shi, Y. (1996) *Science 271*, 1081−1085.
4. Christianson, D. W. (1991) *Adv. Protein Chem. 42*, 281−355.
5. Luisi, B. (1992) *Nature 356*, 379−380.
6. Katti, S. K., LeMaster, D. M., and Eklund, H. (1990) *J. Mol. Biol. 212*, 167−184.
7. Martin, J. L. (1995) *Structure 3*, 245−250.
8. Klemba, M., Gardner, K. H., Marino, S., Clarke, N. D., and Regan, L. (1995) *Nat. Struct. Biol. 2*, 368−373.
9. Klemba, M., and Regan, L. (1995) *Biochemistry 34*, 10094−10100.
10. Regan, L., and Clarke, N. D. (1990) *Biochemistry 29*, 10878−10883.
11. Holmquist, B. (1988) *Methods Enzymol. 158*, 6−13.
12. Kunkel, T. A. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 448−492.
13. Maina, C. V., Riggs, P. D., Grandea, A. G., Slatko, B. E., Moran, L. S., Tagliamonte, J. A., McReynolds, L. A., and Guam, C. (1988) *Gene 40*, 365−373.
14. Benson, D. E., Wisz, M. S., Liu, W., and Hellinga, H. W. (1998) *Biochemistry* (In press).
15. Gill, S. C., and von Hippel, P. H. (1989) *Anal. Biochem. 182*, 319−326.
16. Riddles, P. W., Blakely, R. L., and Zerner, B. (1983) *Methods Enzymol. 91*, 49−60.
17. Segel, I. H. (1975) *Enzyme Kinetics*, pp 72−77, John Wiley & Sons, New York.
18. Pace, C. N., Shirley, B. A., and Thomson, J. A. (1989) in *Protein structure: A practical approach* (Creighton, T. E., Ed.) pp 311−330, IRL Press at Oxford University Press, Oxford.
19. Hellinga, H. W., and Richards, F. M. (1991) *J. Mol. Biol. 222*, 763−785.
20. Coldren, C. D., Hellinga, H. W., and Caradonna, J. P. (1997) *Proc. Natl. Acad. Sci. U.S.A. 94*, 6635−6640.
21. Pinto, A. L., Hellinga, H. W., and Caradonna, J. P. (1997) *Proc. Natl. Acad. Sci. U.S.A. 94*, 5562−5567.
22. Hellinga, H. W. (1998) *J. Am. Chem. Soc.* (Submitted for publication).
23. Pavletich, N. P., and Pabo, C. O. (1991) *Science 252*, 809−817.
24. Ponder, J. W., and Richards, F. M. (1987) *J. Mol. Biol. 193*, 775−791.
25. Hellinga, H. W., Caradonna, J. P., and Richards, F. M. (1991) *J. Mol. Biol. 222*, 787−803.
26. Langsetmo, K., Fuchs, J. A., and Woodward, C. (1991) *Biochemistry 30*, 7603−7609.
27. Bertini, I., and Luchinat, C. (1984) *Adv. Inorg. Biochem. 6*, 71−111.
28. Berg, J. M., and Merkle, D. L. (1989) *J. Am. Chem. Soc. 111*, 3759−3761.
29. Bryson, J. W., Betz, S. F., Lu, H. S., Suich, D. J., Zhou, H. X., O'Neil, K. T., and DeGrado, W. F. (1995) *Science 270*, 935−941.
30. Dahiyat, B. I., and Mayo, S. L. (1997) *Science 278*, 82−87.
31. Hellinga, H. W. (1997) *Proc. Natl. Acad. Sci. U.S.A. 94*, 10015−10017.
32. Pabo, C. (1983) *Nature 301*, 200.
33. Richards, F. M., and Lim, W. A. (1994) *Q. Rev. Biophys. 26*, 423−498.
34. Eisenberg, D., and McClachlan, A. D. (1986) *Nature 319*, 199−203.
35. Kellis, J. T., Jr., Todd, R. J., and Arnold, F. H. (1991) *Bio/Technology 9*, 994−995.
36. Shortle, D. (1992) *Q. Rev. Biophys. 25*, 205−250.
37. Villafranca, J. E., Howell, E. E., Oatley, S. J., Xuong, N.-H., and Kraut, J. (1987) *Biochemistry 26*, 2182−2189.
38. Betz, S. F. (1993) *Protein Sci. 2*, 1551−1558.
39. Hecht, M. H., Richardson, J. S., Richardson, D. C., and Ogden, R. C. (1990) *Science 249*, 884−891.
40. Quinn, T. P., Tweedy, N. B., Williams, R. W., Richardson, J. S., and Richardson, D. C. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 8747−8751.

BI980718F