Probing the Structure and Stability of a Hybrid Protein: The Human−*E. coli* Thioredoxin Chimera[†]

John M. Louis,[‡] Roxana E. Georgescu,[§,||] Maria Luisa Tasayco,[§] Olga Tcherkasskaya,[⊥] and Angela M. Gronenborn*,[‡]

Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892, Department of Chemistry, City College of the City of New York, New York, New York 10031, and Laboratory of Experimental and Computational Biology, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892

Received April 12, 2001

ABSTRACT: The structure and stability of a hybrid protein composed of N-terminal human and C-terminal *E. coli* thioredoxin domains were investigated by NMR, fluorescence, and circular dichroism spectroscopy. We demonstrate that the chimeric protein is correctly folded and exhibits the common thioredoxin architecture. However, the stability of the hybrid protein toward thermal and chemical denaturation is clearly reduced when compared with both parent proteins. Altogether, our data indicate that the interface between the two folding units of thioredoxin is tolerant toward changes in exact interdigitation of side chains, allowing for the formation of the unique overall thioredoxin fold. Further, the gene encoding the human−*E. coli* chimera was tested in vivo whether it supports the assembly of filamentous phages. No complementation of a thioredoxin-deficient *E. coli* mutant for the replication of the phages M13 or fd was observed, suggesting that parts of the overall protein structure in the N-terminal domain are crucial for this activity.

Recognition between segments of polypeptide chains is of fundamental importance for the understanding of signaling via protein−protein interactions, protein folding, and also protein design. To dissect and understand the complexity of specific interactions within proteins, numerous mutagenesis studies have been carried out, resulting in a large body of data relating the specific packing of amino acid residues in the hydrophobic cores of natural proteins to protein stability (*1*−*4*). Likewise, protein design efforts, that initially were marred by their inability to yield proteins exhibiting native-like, well-ordered tertiary structures (*5*−*7*), are now encompassing design strategies that pay serious attention to side chain packing (*8, 9*). This has led to a burst of successful generations of novel, well-folded, and stable proteins (*10*−*13*).

Intrachain domain−domain interfaces constitute an ideal template for studying the physical and chemical determinants linking structure and stability. In this regard, results obtained with complementary protein fragments have contributed substantially to our understanding of protein folding and design (*14*−*20*). In addition, analysis of protein structures in the PDB (*21*) provides crucial information with respect to structural determinants for interchain (*22*−*24*) as well as intrachain (*25*) domain interfaces. It was shown that domain interfaces within a single protein or between different proteins are remarkably alike, exhibiting amino acid compositions more similar to those on protein surfaces than in protein cores. Despite this wealth of data, several aspects of molecular recognition are still obscure, and studies on well-defined model systems are invaluable in determining the factors that govern the structure, stability, and kinetics of association for each individual protein. For a variety of reasons, thioredoxin has long been employed as such a model system. Three-dimensional structures of thioredoxins from different organisms have been determined, either by X-ray crystallography or by NMR[1] (*26*−*35*), and equilibrium as well as kinetic folding studies have been carried out (*19, 36*−*40*). All thioredoxins share very similar three-dimensional structures, comprising a central core of five β-strands surrounded by four α-helices. For example, the rms difference for the backbone atoms between oxidized human and *E. coli* proteins is ∼1.6 Å (*30*), despite an only moderate degree of sequence identity of 25% (*41*). For the *E. coli* thioredoxin, it was shown that fragments comprising two subdomains can reassemble into a native-like structure (*17, 42, 43*), providing evidence that these subdomains might exist as independent folding units. Moreover, interactions across the β-sheet were implied to be crucial for the correct folding of the protein (*44*). It therefore seemed of interest to investigate whether it was possible to create a hybrid protein composed of amino acid sequences from both human and

[†] This work was supported in part by the AIDS Targeted Antiviral Program of the Office of the Director of the National Institutes of Health to A.M.G.

* To whom correspondence should be addressed. Telephone: (301) 496-5414. Fax: (301) 496-1690. E-mail: gronenborn@nih.gov.

[‡] Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases.

[§] Department of Chemistry, City College of the City of New York.

[||] Present address: Department of Physics, City College of the City of New York, New York, NY 10031.

[⊥] Laboratory of Experimental and Computational Biology, National Cancer Institute.

[1] Abbreviations: CD, circular dichroism; DTT, dithiothreitol; EDTA, ethylenediaminetetraacetic acid; GdnHCl, guanidine hydrochloride; NMR, nuclear magnetic resonance; TDOR, thiol−disulfide oxidoreductases; Trx, thioredoxin; PBS, phosphate-buffered saline; IPTG, isopropyl-β-thiogalactopyranoside.

10.1021/bi010745x CCC: $20.00
Published on Web 08/18/2001

E. coli thioredoxins that correspond to these *E. coli* folding subdomains and characterize its structure, stability, and function.

We cloned and expressed a chimeric human−*E. coli* protein and characterized its global architecture and thermodynamic stability. Structural characterization of the chimeric protein reveals that no gross conformational changes within the thioredoxin fold occurred. Specifically, those parts of the structure that contain the human amino acid sequence are very similar to the human structure, whereas those containing the *E. coli* sequence resemble the *E. coli* structure. Indeed, this hybrid can be viewed as consisting of two folding units: the N-terminal $\beta\alpha\beta\alpha\beta\alpha$ motif derived from the human sequence and the C-terminal $\beta\beta\alpha$ motif derived from the *E. coli* sequence, which reassemble into the thioredoxin fold when located on the same polypeptide chain. The stability toward denaturants as well as the thermal stability of the hybrid protein is, however, reduced compared to the parent thioredoxins. It appears that the interface between the two folding units of thioredoxin is tolerant toward changes in exact interdigitation of side chains, allowing for the formation of the unique overall fold. Elucidation of the structural details at the hybrid interface will have to await the determination of the high-resolution three-dimensional structure.

MATERIALS AND METHODS

Construction of the Hybrid Thioredoxin Gene. The human thioredoxin gene was derived from an expression system comprising a cDNA clone (*45*) and the *E. coli* gene from the plasmid pBHK8 (*46*). Using the internal *Hha*I restriction site (corresponding to amino acids A67, P68) in the *E. coli* gene and an engineered *Hha*I site at the corresponding position in the human gene (A66, S67P) from the expression plasmid (*41*), we assembled the gene sequence of the hybrid gene by ligation of the N-terminal *Nco*I−*Hha*I human portion to the C-terminal *Hha*I−*Xba*I *E. coli* portion. The resulting gene was inserted first into a plasmid with a λP_l promoter and for high expression purposes into pET11a (Novagen, Madison, WI).

Expression and Purification of Thioredoxin Variants. Cells were grown at 37 °C either in Luria−Bertani medium or in a modified minimal medium for uniform (>99%) ^{15}N labeling with $^{15}NH_4Cl$ as the sole nitrogen source, and induced for expression in mid-log phase. Human and *E. coli* thioredoxins were purified according to previously published procedures (*28*). Cells derived from 1 L of culture expressing the hybrid protein were suspended in 80 mL of B-PER (bacterial protein extraction reagent, Pierce, Rockford, IL) containing 5 mM benzamidine (Sigma Chemical Co., St. Louis, MO) and sonicated at 4 °C. The insoluble recombinant protein was washed by resuspension in 70 mL of buffer containing 50 mM Tris-HCl (pH 8.2), 10 mM EDTA, and 2 mM DTT. In both cases, the insoluble fraction was pelleted by centrifugation at 20000*g* for 30 min at 4 °C. The final pellet was solubilized in 5 mL of 50 mM Tris-HCl (pH 8), 7.5 M GdnHCl, 5 mM EDTA, and 10 mM DTT to yield a protein concentration of about 20 mg/mL. Thirty milligrams of protein was applied to a Superdex-75 column (HiLoad 2.6 cm × 60 cm, Amersham Pharmacia Biotech, Piscataway, NJ) equilibrated in 50 mM Tris-HCl (pH 8), 4 M GdnHCl, 5 mM EDTA, and 1 mM DTT and at a flow rate of 3 mL/min at ambient temperature. Peak fractions were pooled, and 5−6 mg of protein was folded at room temperature against 2 L of buffer at each step by the following dialysis scheme: 2 M GdnHCl, 0.1 M sodium phosphate buffer (pH 7) for 2 h; 1 M GdnHCl, 0.1 M sodium phosphate buffer (pH 7) for 1 h; 0.5 M GdnHCl, 0.05 M sodium phosphate buffer (pH 7) for 1 h; 0.05 M sodium phosphate buffer (pH 7) for 2 h. The last dialysis step was repeated overnight. Protein was concentrated and applied to a Superdex-75 column in 50 mM sodium phosphate buffer (pH 7). The folded hybrid protein elutes as a single major peak corresponding to a retention volume of a monomer. A minor contaminant (<5%) elutes just after the void volume, well away from the folded monomeric form of the hybrid. Mass spectrometry of the peak fraction yielded masses of 11 680.2 and 11 549.2 at a ratio of 55:45. These experimental masses agree well with the expected values of 11 680.2 and 11 548.8 for the Met+ and Met− forms of the protein. The presence of the two different species in approximately equal proportion is also apparent from the HSQC spectrum (Figure 2), in which doubling of a substantial number of resonances is observed. In particular, note the labeled peaks for E47. Similar observations were made previously for the human thioredoxin (*28*).

NMR Measurements. ^1H-^{15}N HSQC spectra (*47*, *48*) were recorded on a Bruker DMX500 spectrometer equipped with an (*x*, *y*, *z*)-shielded gradient triple resonance probe at 25 °C. A total of 128 × 512 complex points in the indirect (^{15}N) and acquisition (^1H) dimensions, respectively, with total acquisition times of 64 ms in both dimensions were collected. Samples contained ∼1 mM hybrid protein in 100 mM sodium phosphate buffer and 0.02% NaN_3.

Circular Dichroism Measurements. Circular dichroism measurements were carried out on an AVIV-60DS spectropolarimeter (AVIV Instruments Inc.) equipped with a thermostatic cell holder. The spectropolarimeter was calibrated with (+)-10-camphorsulfonic acid, and measurements were performed with a bandwidth of 1 nm, response time of 0.5 s, and scan speed of 20 nm/min. Far-UV spectra were recorded over 180−260 nm in 0.1 and 0.5 cm cells, and near-UV spectra were obtained over 240−340 nm in a 0.5 cm cell. Each spectrum (the average of 4 scans) was corrected for the appropriate solvent-blank. Raw data were expressed as molar ellipticity in units of deg·cm^2·dmol^{-1}. Experiments were carried out at 20 °C in 10 mM potassium phosphate buffer (pH 7) for protein concentrations covering the range of 5−130 μM. No changes in CD data were observed, demonstrating that no significant protein aggregation occurs.

Steady-State Fluorescence Measurements. Fluorescence spectra were obtained on a PTI spectrofluorometer equipped with a temperature-controlled water bath. Measurements were performed with a scan speed of 120 nm/min and data intervals of 1 nm. Emission was measured in the ratio mode and corrected for the appropriate solvent-blanks, as well as for wavelength-dependent bias of the optics and detection system. An excitation wavelength of 295 nm was used in all experiments to avoid tyrosine absorbance. The emission spectrum averaged over four scans was characterized by the position of the intensity maximum and the fluorescence intensity integrated over the range of 305−405 nm. All the experiments were performed at 25 °C in 10 mM potassium phosphate buffer, pH 7.

Chemical and Thermal Denaturation Measurements. Equilibrium (un)folding induced by GdnHCl was monitored by steady-state fluorescence and far-UV circular dichroism for three proteins, i.e., human, *E. coli*, and hybrid thioredoxins. Each sample was dissolved in potassium phosphate buffer containing the desired amount of GdnHCl. Protein concentration was estimated by absorbance using a Hewlett-Packard 8452A diode array spectrophotometer. The molar extinction coefficient at 280 nm was calculated from the tryptophan and tyrosine composition and the standard values of 5690 and 1280 $M^{-1}\cdot cm^{-1}$ reported for those amino acids (*49, 50*). These estimates were supported by measurements of the protein concentration by means of amino acid analysis. Ultrapure GdnHCl was purchased from US Biochemicals Inc. and used without further purification. The concentration of GdnHCl in the sample was determined by refractive index measurements using an AO Abbe refractometer. Experiments were carried out for protein concentrations of 4 and 10 μM. No changes in recovered transition curves were observed, demonstrating that no significant protein aggregation is present.

Thermal denaturation was monitored by measuring the far-UV molar ellipticity at 222 nm. Data were collected as a function of temperature with a scan rate of 1 °C/min over the range of 20−100 °C in 10 mM potassium phosphate buffer, pH 7. The concentrations of hybrid protein were 10, 20, and 40 μM, and for the human protein, concentrations of 10 and 20 μM were used. The concentration of *E. coli* thioredoxin was 20 μM, and its (un)folding profile was used as a reference to compare the present results with those reported previously (*40*). In addition, temperature-induced unfolding curves of human protein were acquired for a variety of GdnHCl concentrations below 4 M (see below). The drift in ellipticity of buffer−GdnHCl mixtures was monitored over the entire temperature range for GdnHCl concentrations of 0.5 and 2 M.

The analysis of (un)folding curves was performed in accord with the procedure reported previously in considerable detail (*40, 51*). The apparent fraction of unfolded molecules, f_U, was extracted from raw data, assuming the two state model:

$$f_U = \frac{X_N - X}{X_N - X_U} \quad (1)$$

In eq 1, the experimental parameter (e.g., fluorescence intensity, ellipticity, etc) measured under given denaturing conditions (e.g., GdnHCl, temperature, etc) is denoted as X, while X_N and X_U are values of the X characteristic in the native and unfolded conformations, respectively, under identical experimental conditions. The equilibrium constant, K_U, and the free energy change, ΔG_U, for the (un)folding reaction were calculated using the basic equations

$$K_U = \frac{f_U}{1 - f_U} \quad (2)$$

and

$$\Delta G_U = -RT \ln K_U = -RT \ln\left(\frac{X_N - X}{X - X_U}\right) \quad (3)$$

where R is the gas constant (1.987 $cal\cdot deg^{-1}\cdot mol^{-1}$) and T is the absolute temperature (K). Employing eqs 1−3 allows one to recover the thermodynamic parameters of the (un)folding reaction by fitting routines applied to either the raw data, X, or (un)folding profiles, f_U (*39*). The effect of denaturing conditions on the characteristics of the native, X_N, and unfolded, X_U, states needs to be addressed. During chemical denaturation, monitored by both fluorescence and far-UV CD, the pre- and post-transition baselines were essentially flat. Hence, the X_N and X_U parameters were assumed to be independent of GdnHCl concentration. For fluorescence data, this assumption is supported by experimental results, which reveal that the emission of the indole fluorophore in solution shows no significant change with GdnHCl concentration (*52*). Thermodynamic parameters were extracted from the Gibbs energy function:

$$\Delta G_u = \Delta G^\circ_{H_2O} - m[\text{GdnHCl}] \quad (4)$$

where $\Delta G^\circ_{H_2O}$ is the free energy in the absence of denaturant and m is the measure of the dependence of free energy on denaturant concentration, [GdnHCl]. For the thermal denaturation monitored by far-UV molar ellipticity, Θ, the pre- and post-transition baselines exhibited a distinct linear change with temperature, which necessitated attention. Specifically, the data were calculated assuming that there is a linear dependence of the ellipticity on temperature for the native and unfolded conformations:

$$\Theta_T = (\Theta_N + m_N T) - [(\Theta_N + m_N T) - (\Theta_U + m_U T)]f_U \quad (5)$$

and the slopes of those dependences were denoted as m_N and m_U, respectively. Equation 5 (including a combination of eqs 1−3) was fitted to the raw data, using the Levenbern−Marguardt algorithm implemented in the commercial program Origin (Microcal). Thermodynamic parameters were extracted from the Gibbs energy function:

$$\Delta G_u(T) = \Delta H_m\left(1 - \frac{T}{T_m}\right) + \Delta C_p\left[(T - T_m) - T\ln\left(\frac{T}{T_m}\right)\right] \quad (6)$$

with ΔC_p, ΔH_m, and T_m being the temperature-independent heat capacity, enthalpy change, and midpoint temperature, respectively.

Bacterial Strains and Phages. The thioredoxin-deficient strain of *E. coli* A307 (*53*) was generously supplied by Marjorie Russel and transformed with the expression plasmids for *E. coli*, human, and hybrid proteins. The ability of the resulting strains to support filamentous phage growth was tested for M13 and fd. One hundred microliters of serial dilutions of phage supernatants in PBS (pH 7.4) was mixed with 200 μL of fresh *E. coli* cells bearing the different thioredoxin plasmids ($OD_{600} \sim 0.5$) and 4 mL of top agar. IPTG (final concentration of 1 mM) and ampicillin (50 μg/mL) were added, and the mixture was plated and grown overnight. Plaque formation was assessed for four different dilutions.

RESULTS AND DISCUSSION

Design of the Thioredoxin Hybrid. The amino acid sequences of human, *E. coli*, and chimeric proteins and the

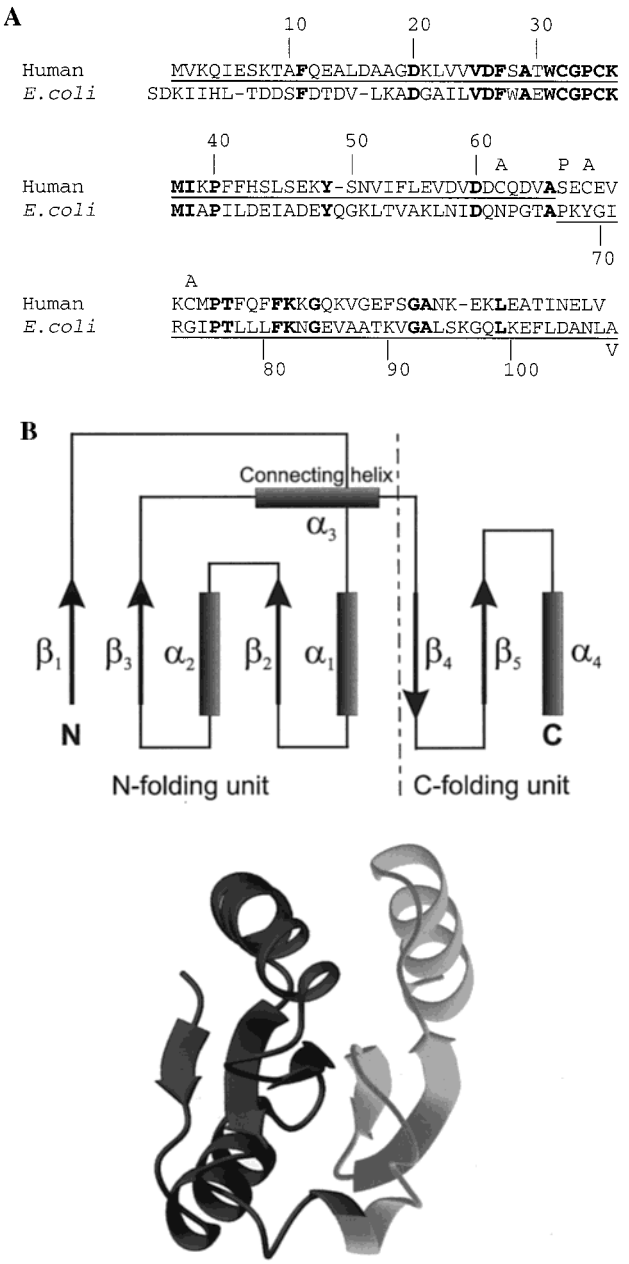

FIGURE 1: Design of a hybrid protein composed of the N- and C-terminal domains of human and *E. coli* thioredoxins, respectively. (A) Alignment of the primary sequences of human (top) and *E. coli* (bottom) thioredoxins. The primary sequence of the hybrid protein is underlined. Conserved residues between human and *E. coli* thioredoxins are shown in boldface letters. Amino acids Ala66 and Pro67 of the human thioredoxin and Ala67 and Pro68 of *E. coli* thioredoxin correspond to the restriction endonuclease site *Hha*I in the DNA sequence. The human protein contains the mutations Cys62Ala, Cys69Ala, and Cys73Ala. Residue 108 in the hybrid protein is Val instead of Ala. (B) Schematic representation of the secondary structure elements in the thioredoxin fold (top) and structure of the hybrid protein based on molecular modeling (bottom).

subdomains of the thioredoxin fold are displayed in Figure 1. The crossover point between the two different amino acid sequences is located at position Ala66 (human numbering) within helix α3 (human structure), that runs approximately perpendicular to the direction of the five-stranded *β*-sheet. This helix can be viewed as constituting the bridge between the N- and C-motifs of the thioredoxin fold (*54*). This position is slightly N-terminal to the boundary of the two subdomains of *E. coli* thioredoxin (amino acids 73/74), previously identified by proteolytic cleavage (*17, 40*). In the *E. coli* structure, residues 59−63 are α-helical, whereas residues 66−70 form a 3_{10}-helix (*27*). In the human structure, residues 58−62 form a helical turn, followed by an α-helix for residues 63−69 (*30*). Inspection of both structures suggested that most residues at the newly created domain interface in a chimera might be accommodated without severe steric clash. In particular, the seven amino acids forming *β*-strand 2, Leu22 to Ser28 of the human sequence, share three identities and two conservative changes with the equivalent positions in *E. coli*. In addition, we reasoned that the replacement of the two larger residues Ile23 and Leu24 by the smaller Val residues would perhaps weaken the interface to some degree, but not disrupt it. Residues on the two α-helices, α2 and α4, involved in helix−helix packing, also exhibit similar characteristics in the two parent molecules and the hybrid. In the *E. coli* structure, Ile41 interacts with Leu99, Ile45 with Leu103, and Tyr49 with Leu107. A pivotal hydrophobic contact in the human structure is formed between Leu45 and Ile101. In the hybrid interface, these contacts would be replaced by Phe40−Leu99, Leu45−Leu103, and Tyr49−Leu107.

Structure Assessment of the Thioredoxin Hybrid by NMR. We expressed and purified the hybrid thioredoxin to homogeneity, defined conditions for its folding in vitro, and characterized its structural properties. Figure 2 shows the ^{1}H-^{15}N HSQC spectrum of uniformly ^{15}N-labeled protein. As can be appreciated, the spectrum exhibits all the hallmarks of a well-folded polypeptide, displaying significant dispersion of cross-peaks and narrow line widths. Comparison with the equivalent spectra of human and *E. coli* thioredoxins (*40, 55*) reveals that those resonances arising from the *E. coli* portion of the protein exhibit similar chemical shifts to their counterparts in the *E. coli* spectrum while those from the human portion (in brackets) are similar to the equivalent ones in the human spectrum. Naturally, some small differences exist, as well as more pronounced changes for residues at the interface. Since chemical shifts are extremely sensitive parameters reporting on a particular structural environment around the associated atoms, a qualitative assessment of structural identity and similarity can be carried out. For all three proteins, the observed similarities are enough to conclude that the overall three-dimensional structure of the hybrid protein is similar to the human and *E. coli* structures with respect to secondary structure elements and fold. A structure of the hybrid protein based on molecular modeling is shown in ribbon representation in Figure 1B. The modeling is based on the respective structures of the parent proteins and energy minimization of the resulting chimera. Initial assessment of the model based on NMR parameters indicates that indeed the fold agrees with the experimental observations. A detailed evaluation of the interface in structural terms will, however, have to await the determination of a refined structure of the hybrid protein.

Circular Dichroism and Steady-State Fluorescence. Further assessment of structural features of thioredoxins was carried out by circular dichroism and fluorescence spectroscopy. Figure 3 displays near- and far-UV CD spectra of all three proteins under native conditions. As can be appreciated, an unusually large ellipticity in the near-UV range is

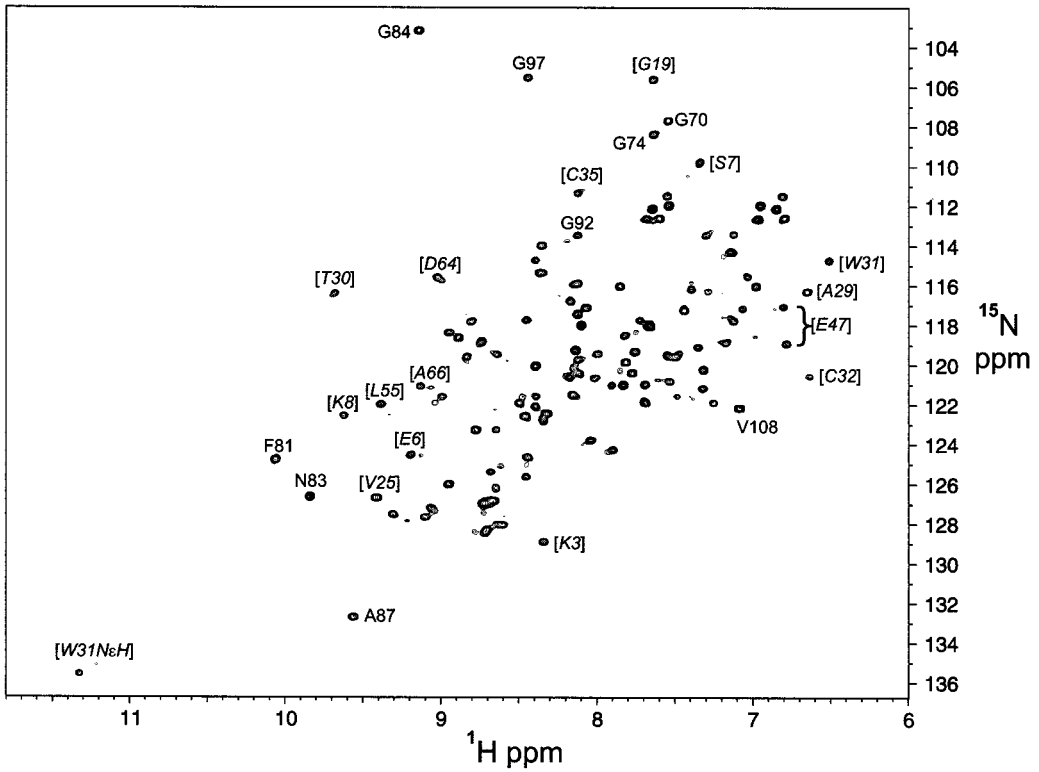

FIGURE 2: ^{1}H-^{15}N HSQC spectrum of the hybrid thioredoxin composed of human and *E. coli* domains. The protein concentration is ~1 mM in 100 mM sodium phosphate buffer (pH 6) at 25 °C. Selected cross-peaks are labeled with the amino acid type and position in the human (in brackets) and *E. coli* portions of the hybrid protein sequence (see Figure 1). The two cross-peaks observed for Glu47 arise from the N-terminal heterogeneity of the protein, i.e., the Met+ and Met− forms, respectively.

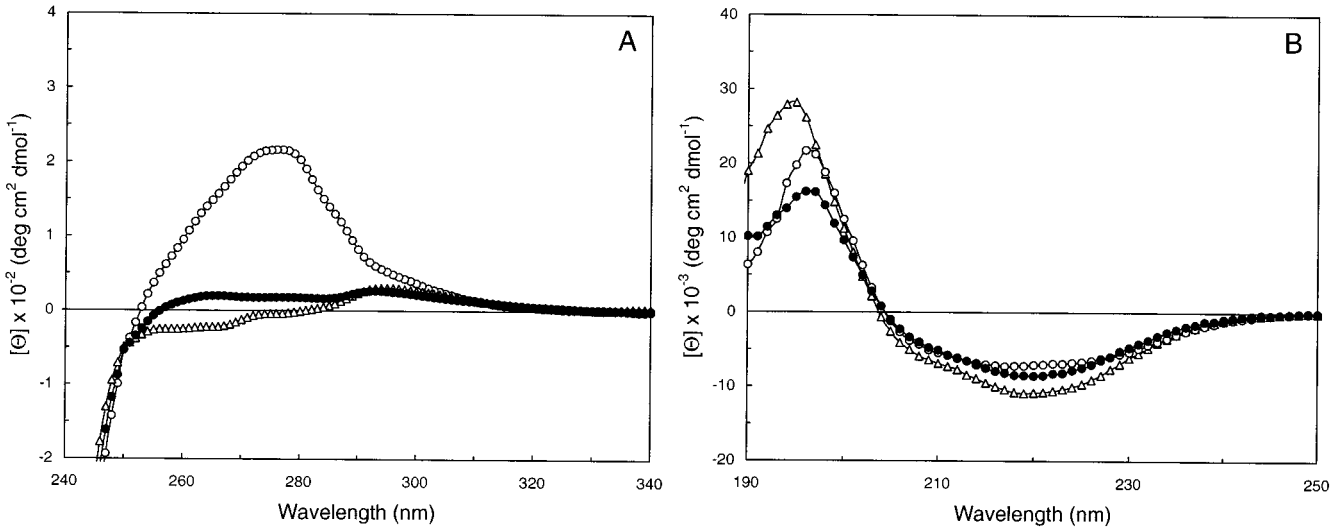

FIGURE 3: Near-UV (A) and far-UV (B) circular dichroism spectra of thioredoxin proteins: human (△), *E. coli* (○), and hybrid (●). Measurements were carried out at 20 °C in 10 mM potassium phosphate buffer at pH 7. Other experimental conditions were as described under Materials and Methods.

observed for the *E. coli* protein, which is absent in the human and hybrid proteins (Figure 3A). The most straightforward explanation for this finding relates to the number of aromatic residues in the above proteins and their locations within the three-dimensional structures. For instance, *E. coli* thioredoxin contains two tryptophan residues, i.e., Trp28 and Trp31, which are close in sequence and space. Furthermore, Trp28 is immobilized in the structure and therefore might be responsible for the large ellipticity. The two other thioredoxins, i.e., human and hybrid, contain only a single tryptophan, Trp31, which is located on the reactive site loop, exposed to solvent. This may explain why the near-UV CD spectra of the human and hybrid proteins exhibit features associated with a symmetric, fluctuating environment. The far-UV CD spectra of the thioredoxins are shown in Figure 3B. Clearly, the CD spectrum of the hybrid protein is very similar to the *E. coli* one, with the human spectrum exhibiting somewhat larger ellipticity in the 220 nm region (Figure 3B). Despite those small differences, the far-UV CD data are consistent with the known content of overall secondary structure in the human and *E. coli* proteins, and the hybrid protein will contain similar amounts. The intrinsic fluores-

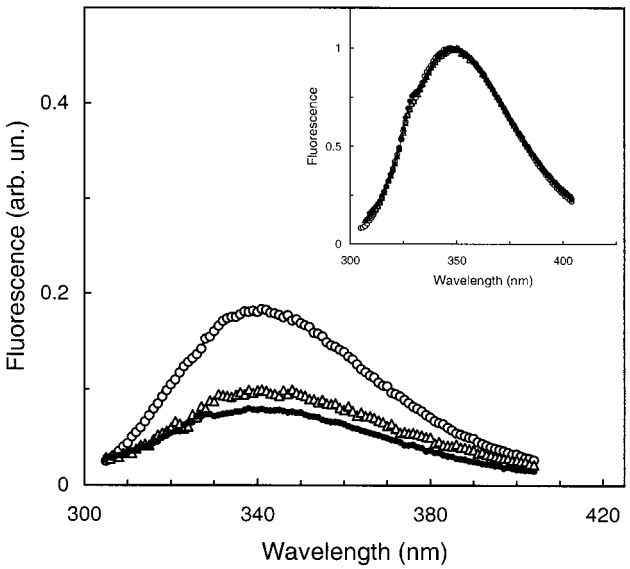

FIGURE 4: Intrinsic fluorescence of thioredoxin proteins: human (△), *E. coli* (○), and hybrid (●). Excitation was at 295 nm, and the spectra were recorded at 25 °C for samples containing 20 μM protein in 10 mM potassium phosphate buffer, pH 7. The inset shows normalized emission spectra of unfolded proteins in 6 M GdnHCl.

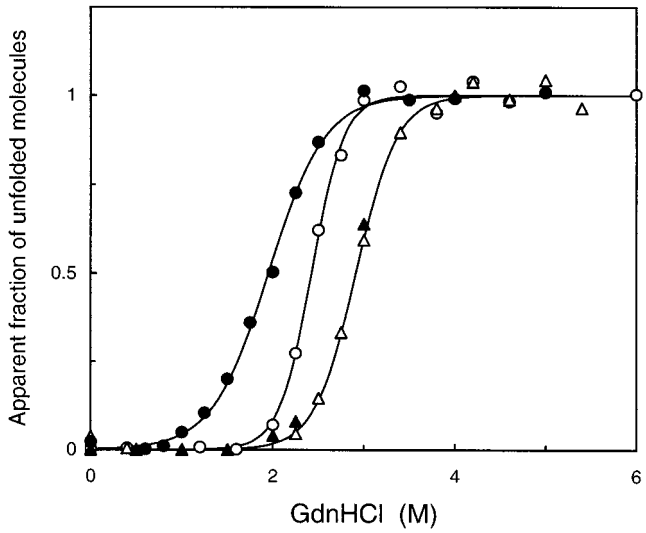

FIGURE 5: Equilibrium GdnHCl denaturation curves for thioredoxin proteins: human (△), *E. coli* (○), and hybrid (●). The (un)folding transition was followed by the intrinsic fluorescence excited at 295 nm. Raw data were converted to the fraction of unfolded molecules and plotted against GdnHCl concentration. Superpositions of the (un)folding profiles monitored by fluorescence (△) and far-UV molar ellipticity at 222 nm (▲) are shown for human thioredoxin. The solid lines represent nonlinear least-squares fits of the data to the two-state model. The recovered thermodynamic parameters are summarized in Table 1. All measurements were carried out at 25 °C on samples containing 4 and 10 μM protein in 10 mM potassium phosphate buffer, pH 7.

cence of the three thioredoxins also closely resemble each other (Figure 4). Specifically, the emission spectra of the folded proteins are centered around 340 nm, and unfolding by GdnHCl results in a large increase in the fluorescence quantum yield and a 10 nm red shift of the emission spectra. Overall, evaluation of all the data from NMR, fluorescence, and CD experiments allows us to conclude that no gross conformational changes within the thioredoxin fold have occurred in the hybrid protein.

Denaturation by Chaotropic Agents. The stability of the hybrid protein toward chemical denaturants was probed by tryptophan fluorescence and far-UV circular dichroism. Figure 5 displays the (un)folding profiles generated for human, *E. coli*, and chimeric thioredoxins in the presence of increasing amounts of GdnHCl. As shown for the human protein, the CD profiles superimpose perfectly on those derived from fluorescence, indicating that the formation of the coillike environment of the Trp residues (fluorescence) and the disruption of native secondary structure (far-UV CD) occur under identical conditions. Simple two-state fits of the individual curves yield denaturant concentrations at the midpoint of the transitions, C_m, of 1.9, 2.5, and 2.9 M GdnHCl for the hybrid, *E. coli*, and human proteins, respectively. The extrapolated Gibbs free energy in the absence of denaturant, $\Delta G^\circ_{H_2O}$, has values of 3.9, 7.9, and 7.2 kcal mol^{-1} (see Table 1). The steepness of the unfolding curves is clearly different for all three proteins, and recovered *m*-values, characterizing the dependence of the free energy on denaturant concentration, are also not identical. The most cooperative transition is observed for the *E. coli* protein with an *m*-value of 3.2 kcal (mol·M)$^{-1}$, whereas both the human and the hybrid proteins exhibit smaller *m*-values of about 2.2 kcal (mol·M)$^{-1}$. Comparison the (un)folding profiles generated for the parent proteins illustrates the well-known fact that a higher C_m is not necessarily an indication of higher thermodynamic stability: For example, the *E. coli* protein has a lower C_m than the human protein, although the extrapolated Gibbs energy is higher than that of the human protein (Table 1). In fact, the *m*-value is assumed to be proportional to the solvent-exposed surface area upon unfolding of a protein (*56*), and a small *m*-value is thought to be correlated with a more compact denatured state (*57*) or an increased population of folding intermediates (*58*, *59*). Since the three-dimensional structures of the human and *E. coli* proteins closely resemble each other, the accessible surface area upon unfolding is expected to be also very similar. In this regard, calculation of the accessible surface area for the unfolded chains using the methodology of Creamer et al. (*62*) yields quite similar upper values of 14 909, 14 780, and 15 059 Å^2 for human, *E. coli*, and hybrid thioredoxin, respectively. Given the noticeable difference in the *m*-values of the two parent proteins, it seems likely that the present equilibrium experiments hint at different putative intermediates on the folding pathway of these thioredoxins. Further support for the notion that folding intermediates may

Table 1: Thermodynamic Parameters of the GdnHCl-Induced (Un)Folding of Thioredoxin Proteins at 25 °C in 10 mM Potassium Phosphate Buffer, pH 7[a]

thioredoxin	C_m[b] (M)	ΔG°[c] (kcal·mol^{-1})	m [kcal (mol·M)$^{-1}$]	$\Delta\Delta G_u$[d] (kcal·mol^{-1})
human	2.9 ± 0.2	7.2 ± 0.5	2.5 ± 0.2	
E. coli	2.5 ± 0.2	7.9 ± 0.4	3.2 ± 0.2	−
hybrid	1.9 ± 0.2	3.9 ± 0.4	2.0 ± 0.2	−3.3 ± 0.5
				−4.0 ± 0.5

[a] The error in the values of recovered parameters is less than 10% as judged by the quality of the fit and repeated experiments. [b] Concentration of GdnHCl at the midpoint of transition. [c] Standard Gibbs free energy in the absence of denaturant. [d] Difference in standard free energy between hybrid and human (first row) or *E. coli* (second row) thioredoxin proteins.

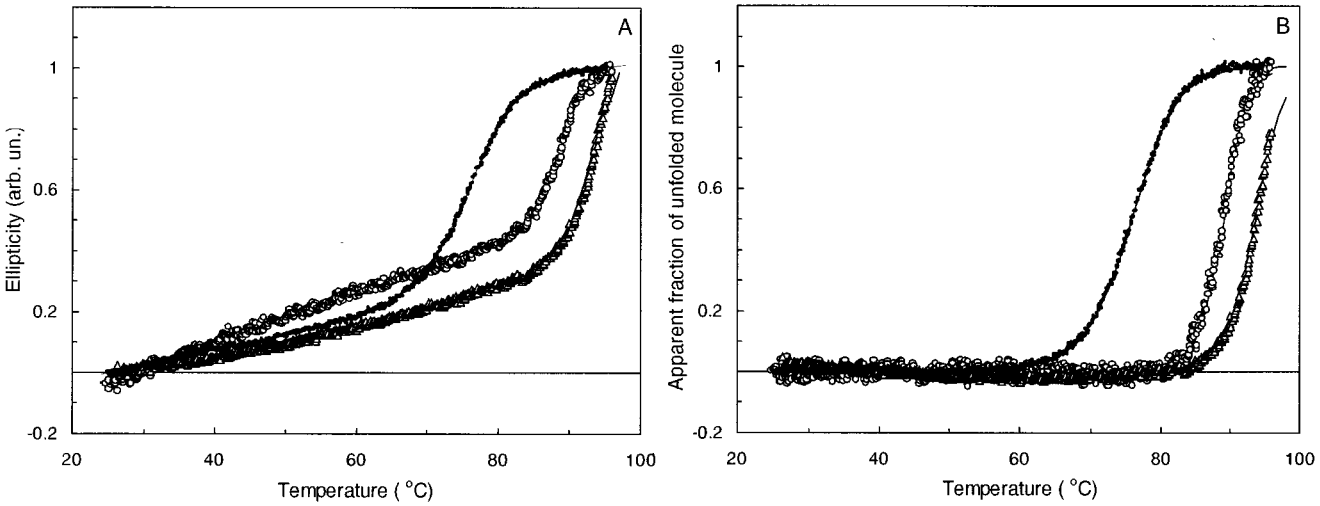

FIGURE 6: Thermal denaturation curves for thioredoxin proteins: human (△), *E. coli* (○), and hybrid protein (●). Both raw data (A) and the apparent fraction of unfolded molecules (B) are displayed. The recovered thermodynamic parameters are summarized in Table 2. Other experimental conditions were as described under Materials and Methods.

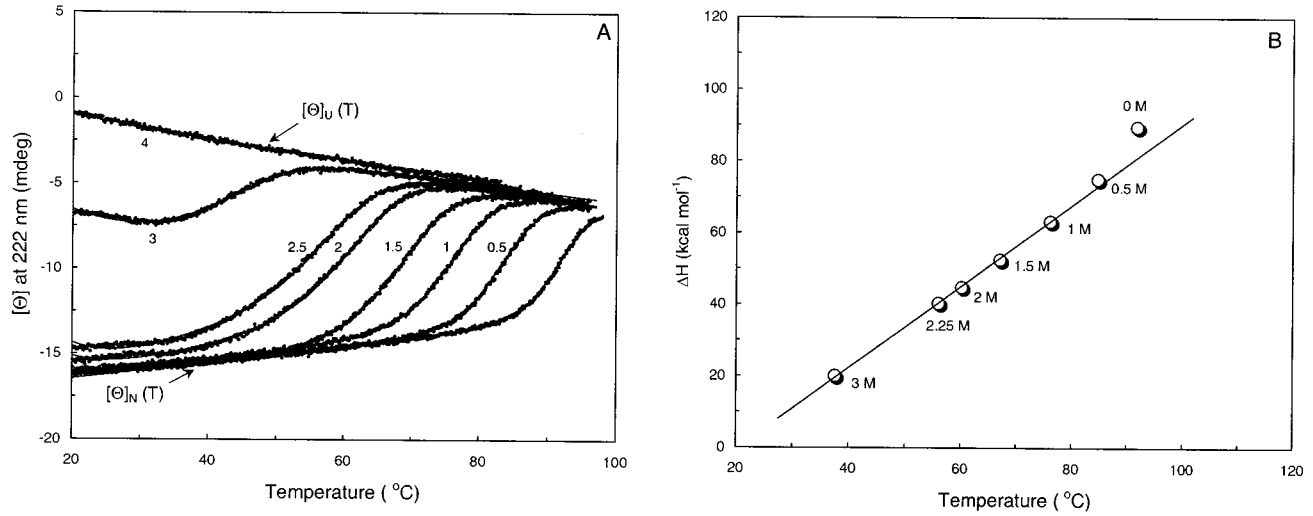

FIGURE 7: Effect of GdnHCl on the thermal denaturation of human thioredoxin. Both raw CD data collected for GdnHCl concentrations below 4 M (A) and effect of the unfolding enthalpy (ΔH_m) on the transition temperature (T_m) for each of the above transitions (B) are shown. The data refer to the GdnHCl concentration in molar units. Linear extrapolation of the data in panel B yields an apparent heat capacity (ΔC_p) of 1.18 ± 0.04 kcal (mol·K)$^{-1}$.

be present is provided by results obtained for a mutant chimeric protein, in which Gly70 was substituted by valine. Unfolding profiles of this variant exhibit a lower apparent midpoint denaturation concentration of about 1 M GdnHCl, accompanied by a distorted sigmoidal transition (data are not shown). This finding strongly suggests the presence of an equilibrium unfolding intermediate during the chemical denaturation of the hybrid protein. Overall, GdnHCl denaturation clearly demonstrates the reduced thermodynamic stability of the hybrid protein compared with both parent proteins. This is readily understood from the predicted structural imperfections of the chimeric interface. Interestingly, the (un)folding curve of the hybrid protein resembles more the one observed for the human thioredoxin, than that of the *E.coli* protein. In particular, the steepness is very similar. This may reflect the dominance of the N-terminal domain in the formation of any putative folding intermediate.

Thermal Denaturation of Thioredoxin Proteins. Unfolding of the thioredoxins induced by temperature was monitored by far-UV molar ellipticity at 222 nm. Melting curves obtained for all three proteins are displayed in Figure 6. As can be appreciated, the hybrid thioredoxin exhibits the lowest midpoint temperature (T_m = 76 °C) followed by the *E. coli* (T_m = 89 °C) and the human (T_m = 91 °C) proteins. The value obtained for the *E. coli* protein agrees well with previous results (*40*). The unfolding curve for human thioredoxin precluded the extraction of an ellipticity value, Θ_U, for the unfolded protein (Figure 6) due to the high T_m value. Additional melting experiments in the presence of varying amounts of GdnHCl were carried out to provide estimates for the unfolded ellipticity (Figure 7A). These experiments also provided independent estimates for the temperature-independent heat capacity, ΔC_p, in the absence of denaturant for the human protein. Specifically, a ΔC_p value of 1.18 kcal (mol·K)$^{-1}$ was calculated from the linear dependence of the unfolding enthalpy, ΔH, on the transition temperature, T_m, obtained for different GdnHCl concentrations (Figure 7B). Unfolding curves were analyzed assuming a two-state transition, and the recovered thermodynamic parameters are listed in Table 2.

Table 2: Thermodynamic Parameters of the Thermal (Un)Folding of Thioredoxin Proteins in 10 mM Potassium Phosphate Buffer at pH 7[a]

thioredoxin	T_m (°C)	ΔH_m (kcal mol^{-1})	ΔC_p [kcal (K·mol)$^{-1}$]
human	91.2 ± 0.3	99 ± 4	1.59[c]
E. coli[b]	88.7 ± 0.1	116 ± 4	1.71
hybrid	75.7 ± 0.1	73 ± 1	1.58[d]

[a] Thermodynamic parameters are calculated by fitting the thermal denaturation curve followed by far-UV circular dichroism to a two-state model. The error in the values of the recovered parameters is less than 5% as judged by the quality of the fit and repeated experiments. [b] The values of $\Delta H_m = 107$ and $\Delta C_p = 1.46$ were obtained from calorimetry measurements. [c] The value of 1.18 was extracted from the slope of the ΔH vs T_m (see Figure 7B). [d] Theoretical value of $\Delta C_p = 1.34$.

Thermal denaturation, like GdnHCl denaturation, again reveals a reduced thermodynamic stability of the hybrid protein compared to both parent thioredoxins. It should be noted, that the temperature unfolding curve for the *E. coli* protein exhibits a pronounced slope in the pre-transition region, similar to that observed previously (*39*). The effect of temperature on the baselines of both the human and hybrid proteins is clearly less pronounced, and appears to be similar (Figure 6A). Given the uncertainty in interpretation of the linear change in ellipticity before and after an unfolding transition, we can only speculate as to the origin of this phenomenon in thioredoxins. Possible explanations may involve partial unfolding of the proteins, changes from well-folded to molten globule-like states, changes in the contribution of aromatic side chains, or other conformational effects. Generally, thermal denaturation is assumed to be associated with the disruption of intrachain hydrogen bonds, whereas hydrophobic interactions become stronger with increasing temperature (*59*). In fact, unfolding by temperature will depend on the balance between different interactions contributing to the protein stability. It thus may be possible that the pronounced slope on the folded side of the thermal unfolding transition in thioredoxins arises from partial unfolding of the protein, counteracted by an increase in hydrophobic interactions. A possible area in which such local unfolding or increased flexibility prior to complete unfolding may occur, may comprise the connecting helix between the N- and C-terminal domains in the protein, although other regions could also contribute. In this regard, the strong similarities between the features observed for the hybrid protein and the human parent suggest that they might originate from rearrangements involving the common, N-terminal domain. At the present time, there is no straightforward explanation for the steeper baseline observed for the *E. coli* protein. However, since no accompanying changes in the near-UV CD melting curve were observed (*39*), they probably do not involve the cluster around Trp28, which is buried in the protein structure.

Biological Activity of Thioredoxin Proteins. Although human thioredoxin is active as a general protein disulfide reductase (*28, 45*), it is incapable of supporting filamentous phage assembly in *E. coli* (*28*). Based on the proposed interaction area between thioredoxin and other proteins (*60*), as well as mutagenesis data (*53*), the crucial structural determinants for phage assembly of *E. coli* thioredoxin were mapped to a type VIb reverse turn, containing the conserved cis-proline (residues 74−77) and the loop at the end of β-strand 5 (residue 92). These two loops are positioned across a cleft, opposing the active site. In the present hybrid, these parts comprise *E. coli* sequences, and if only specific primary sequences were important, one would have expected the hybrid to exhibit *E. coli* like activity. This was observed for an *Anabaena*−*E.coli* thioredoxin chimera, in which the first one-third of the sequence was *Anabaena*-derived, and the second two-thirds *E. coli*-derived, but not for the reverse orientation (*61*). We tested the present hybrid thioredoxin in a *trx* deletion strain, transformed with plasmids containing *E. coli trx*, human *trx*, and the hybrid *trx* genes. No plaque formation was observed for strains containing the human or the hybrid *trx* gene after transfection with M13 or fd. All three proteins were made in essentially equal amounts in the different transformants as judged by SDS−PAGE of cell extracts. Since the human−*E. coli* thioredoxin hybrid contains a considerably smaller portion of *E. coli* sequences than the *Anabaena*−*E. coli* chimera, it is most likely that the inability to support filamentous phage assembly is due to subtle differences across the hybrid interface, despite retaining the crucial cis-proline and the overall thioredoxin fold. Since the precise nature of the involvement of *E. coli* thioredoxin in filamentous phage assembly is unknown, other causes for the loss of activity of the hybrid protein cannot be ruled out. The elucidation of the high-resolution three-dimensional structure of the hybrid human−*E. coli* thioredoxin will be necessary to substantiate our hypothesis.

CONCLUDING REMARKS

Using a variety of spectroscopic techniques, we demonstrate that the present human−*E. coli* thioredoxin chimera adopts the common thioredoxin fold. This illustrates, at least for the case presented here, the structural tolerance of an internal protein interface. Elucidation of the subtle conformational changes which most likely are present in the packing of residues at the hybrid interface will require further high-resolution structural studies. The overall stability of the hybrid protein toward chaotropic agents and temperature is reduced, compared to both parent molecules, albeit not drastically. This is most likely due to imperfections in side-chain packing at the interface. Interestingly, some of the qualitative features observed in the unfolding curves suggest that the hybrid protein resembles more closely the human thioredoxin than the *E. coli* protein. Whether this simply reflects the fact that a larger portion of the amino acid sequence is derived from the human one or whether indeed the N-terminal domain is dominant in imparting certain features on the folding pathway is unclear at the present time. Future studies using the human−*E. coli* hybrid protein as a starting molecule for developing strategies to evolve a more stable protein interface may provide answers to this question.

ACKNOWLEDGMENT

We thank Elizabeth Murphy for initial cloning efforts, Stewart Durell for the initial hybrid protein model, and Bindi Dangi and Laura Barrientos for recording NMR spectra. We also are very grateful to Norma J. Greenfield for enlightening discussions about CD data analysis.

REFERENCES

1. Richards, F. M., and Richmond, T. (1977) *Ciba Found. Symp.*, 23−45.

2. Eriksson, A. E., Baase, W. A., Zhang, X. J., Heinz, D. W., Blaber, M., Baldwin, E. P., and Matthews, B. W. (1992) *Science 255*, 178−183.
3. Lim, W. A., and Sauer, R. T. (1991) *J. Mol. Biol. 219*, 359−376.
4. Fersht, A. R., and Serrano, L. (1993) *Curr. Opin. Sruct. Biol. 3*, 75−83.
5. Regan, L., and DeGrado, W. F. (1988) *Science 241*, 976−978.
6. Hecht, M. H., Richardson, J. S., Richardson, D. C., and Ogden, R. C. (1990) *Science 249*, 884−891.
7. Betz, S. F. (1993) *Protein Sci. 2*, 1551−1558.
8. Desjarlais, J. R., and Handel, T. M. (1995) *Protein Sci. 4*, 2006−2018.
9. Dahiyat, B. I., and Mayo, S. L. (1996) *Protein Sci. 5*, 895−903.
10. Malakauskas, S. M., and Mayo, S. L. (1998) *Nat. Struct. Biol. 5*, 470−475.
11. Lazar, G. A., Desjarlais, J. R., and Handel, T. M. (1997) *Protein Sci. 6*, 1167−1178.
12. DeGrado, W. F., Summa, C. M., Pavone, V., Nastri, F., and Lombardi, A. (1999) *Annu. Rev. Biochem. 68*, 779−819.
13. Johnson, E. C., Lazar, G. A., Desjarlais, J. R., and Handel, T. M. (1999) *Struct. Folding Des 7*, 967−976.
14. Taniuchi, H., and Anfinsen, C. B. (1971) *J. Biol. Chem. 246*, 2291−2301.
15. Dyson, H. J., Sayre, J. R., Merutka, G., Shin, H. C., Lerner, R. A., and Wright, P. E. (1992) *J. Mol. Biol. 226*, 819−835.
16. Gay, G. D., and Fersht, A. R. (1994) *Biochemistry 33*, 7957−7963.
17. Tasayco, M. L., and Chao, K. (1995) *Proteins: Struct., Funct., Genet. 22*, 41−44.
18. Kobayashi, N., Honda, S., Yoshii, H., Uedaira, H., and Munekata, E. (1995) *FEBS Lett. 366*, 99−103.
19. Chaffotte, A. F., Li, J. H., Georgescu, R. E., Goldberg, M. E., and Tasayco, M. L. (1997) *Biochemistry 36*, 16040−16048.
20. Honda, S., Kobayashi, N., Munekata, E., and Uedaira, H. (1999) *Biochemistry 38*, 1203−1213.
21. Bernstein, F. C., Koetzle, T. F., Williams, G. J., Meyer, E. F., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., and Tasumi, M. (1977) *Eur. J. Biochem. 80*, 319−324.
22. Jones, S., and Thornton, J. M. (1995) *Prog. Biophys. Mol. Biol. 63*, 31−65.
23. Tsai, C. J., Lin, S. L., Wolfson, H. J., and Nussinov, R. (1996) *Crit. Rev. Biochem. Mol. Biol. 31*, 127−152.
24. Tsai, C. J., Lin, S. L., Wolfson, H. J., and Nussinov, R. (1997) *Protein Sci. 6*, 53−64.
25. Jones, S., Marin, A., and Thornton, J. M. (2000) *Protein Eng. 13*, 77−82.
26. Holmgren, A., Soderberg, B. O., Eklund, H., and Branden, C. I. (1975) *Proc. Natl. Acad. Sci. U.S.A. 72*, 2305−2309.
27. Katti, S. K., LeMaster, D. M., and Eklund, H. (1990) *J. Mol. Biol. 212*, 167−184.
28. Forman-Kay, J. D., Clore, G. M., Wingfield, P. T., and Gronenborn, A. M. (1991) *Biochemistry 30*, 2685−2698.
29. Stone, M. J., Chandrasekhar, K., Holmgren, A., Wright, P. E., and Dyson, H. J. (1993) *Biochemistry 32*, 426−435.
30. Qin, J., Clore, G. M., and Gronenborn, A. M. (1994) *Structure 2*, 503−522.
31. Jeng, M.-F., Campbell, A. P., Begely, T., Holmgren, A., Case, D. A., Wright, P. E., and Dyson, H. J. (1994) *Structure 2*, 853−868.
32. Saarinen, M., Gleason, F. K., and Eklund, H. (1995) *Structure 3*, 1097−1108.
33. Weichsel, A., Gasdaska, J. R., Powis, G., and Montfort, W. R. (1996) *Structure 4*, 735−751.
34. Mittard, V., Blackledge, M. J., Stein, M., Jacquot, J. P., Marion, D., and Lancelin, J. M. (1997) *Eur. J. Biochem. 243*, 374−383.
35. Nicastro, G., De Chiara, C., Pedone, E., Tato, M., Rossi, M., and Bartolucci, S. (2000) *Eur. J. Biochem. 267*, 403−413.
36. Reutiman, H., and Luisi, P. L. (1983) *Biopolymers 22*, 107−111.
37. Kelly, R. F., and Richards, F. M. (1987) *Biochemistry 26*, 6765−6774.
38. Borden, K. L., and Richards, F. M. (1990) *Biochemistry 29*, 3071−3077.
39. Georgescu, R. E., Li, J. H., Goldberg, M. E., Tasayco, M. L., and Chaffotte, A. F. (1998) *Biochemistry 37*, 10286−10297.
40. Georgescu, R. E., Braswell, E. H., Zhu, D., and Tasayco, M. L. (1999) *Biochemistry 38*, 13355−13366.
41. Forman-Kay, J. D., Clore, G. M., Stahl, S. J., and Gronenborn, A. M. (1992) *J. Biomol. NMR 2*, 431−445.
42. Slaby, I., and Holmgren, A. (1975) *J. Biol. Chem. 250*, 1340−1347.
43. Reutimann, H., Straub, B., Luisi, P. L., and Holmgren, A. (1981) *J. Biol. Chem. 256*, 6796−6803.
44. Tasayco, M. L., Fuchs, J., Yang, X. M., Dyalram, D., and Georgescu, R. E. (2000) *Biochemistry 39*, 10613−10618.
45. Wollman, E. E., d'Auriol, L., Rimsky, L., Shaw, A., Jacquot, J. P., Wingfield, P., Graber, P., Dessarps, F., Robin, P., Galibert, F., et al. (1988) *J. Biol. Chem. 263*, 15506−15512.
46. Lunn, C. A., Kathju, S., Wallace, B. J., Kushner, S. R., and Pigiet, V. (1984) *J. Biol. Chem. 259*, 10469−10474.
47. Bax, A., Ikura, M., Kay, L. E., Torchia, D. A., and Tschudin, R. (1990) *J. Magn. Reson. 86*, 304−318.
48. Piotto, M., Saudek, V., and Sklenar, V. (1992) *J. Biomol. NMR 2*, 661−665.
49. Edelhoch, H. (1967) *Biochemistry 6*, 1948−1954.
50. Gill, S. C., and von Hippel, P. H. (1989) *Anal. Biochem. 182*, 319−326.
51. Pace, C. N. (1986) *Methods Enzymol. 131*, 266−280.
52. Eftink, M. R. (1994) *Biophys. J. 66*, 482−503.
53. Russel, M., and Model, P. (1986) *J. Biol. Chem. 261*, 14997−15005.
54. Martin, J. L. (1995) *Structure 3*, 245−250.
55. Chandrasekhar, K., Krause, G., Holmgren, A., and Dyson, J. H. (1991) *FEBS Lett. 284*, 178−183.
56. Myers, J. K., Pace, C. N., and Scholtz, J. M. (1995) *Protein Sci. 4*, 2138−2148.
57. Shortle, D. (1995) *Nat. Struct. Biol. 2*, 91−93.
58. Pace, C. N. (1975) *CRC Crit. Rev. Biochem. 3*, 1−43.
59. Privalov, P. L. (1996) *J. Mol. Biol. 258*, 707−725.
60. Eklund, H., Cambillau, C., Sjöberg, B. M., Holmgren, A., Jörnvall, H., Höög, J. O., and Brändén, C. I. (1984) *EMBO J. 3*, 1443−1449.
61. Lim, C. J., Gleason, F. K., Jacobson, B. A., and Fuchs, J. A. (1988) *Biochemistry 27*, 1401−1408.
62. Craemer, T. P., Srinivasan, R., and Rose, G. D. (1997) *Biochemistry 36*, 2832−2835.

BI010745X