DSC Studies of a Family of Natively Disordered Fragments from *Escherichia coli* Thioredoxin: Surface Burial in Intrinsic Coils†

Carla Mendoza,‡ Francisco Figueirido,§ and María Luisa Tasayco*,‡

Biochemistry Division, Department of Chemistry, City College of New York, 138th Street and Convent Avenue, New York, New York 10031, and Imagine Software, 233 Broadway, New York, New York 10279

Received November 19, 2002; Revised Manuscript Received January 27, 2003

ABSTRACT: The accumulating data from proteome analysis indicates that numerous proteins have segments and/or domains, involved in regulatory functions of the eukaryotic cell, which are entirely unstructured under physiological conditions, challenging the structure−function paradigm. Although many such natively unfolded proteins have been structurally analyzed by NMR spectroscopy, little is known about solvent inaccessible surfaces in premolten globules and intrinsic coils. Recent DSC studies of two protein fragments have shown a promising way to estimate the predominantly hydrophobic buried surfaces [Georgescu, R. E., García-Mira, M. M., Tasayco, M. L., and Sánchez-Ruiz, J. M. (2001) *Eur. J. Biochem. 268,* 1−10]. Here we report a systematic heat capacity analysis of a family of natively disordered complementary fragments of oxidized *Escherichia coli* thioredoxin (1−31/32−108, 1−37/38−108, 1−50/51−108, and 38−73) which provides insights into the local and nonlocal interactions contributing to the burial of predominantly hydrophobic surface in intrinsic coils.

Numerous proteins have segments and/or domains which are entirely unstructured under physiological conditions but become structured upon binding to the target (*1*). Such segments or domains have been variously called "natively unfolded" (*2*, *3*), "intrinsically unstructured" (*4*), or "intrinsically disordered" (*5*) and frequently correspond to regions involved in regulatory functions of the eukaryotic cell (*1*) such as transcription (*6*−*11*) and signal transduction (*12*−*14*).

During the past decade, the accumulation of results arising from the experimental and computational analysis of proteomes (*15*) has demonstrated the widespread occurrence of natively unfolded regions in proteins from different organisms, especially in higher eukaryots. This has provoked intriguing hypotheses about their role in biological processes (*16*, *17*) since their existence seems to challenge the structure−function paradigm of proteins (*3*, *4*, *15*, *18*). Models such as the protein trinity (*18*) and the protein quartet (*3*) have been proposed in an attempt to provide an alternative view that would encompass these cases.

At least two groups of natively unfolded proteins have been distinguished by Uversky et al. (*3*) on the basis of structural properties: premolten globules and intrinsic coils. These proteins appear to have a combination of low mean hydrophobicity and relatively high net charge (*19*), low content of ordered secondary structure, different dependency of the hydrodynamic dimensions on protein molecular mass (*2*, *19*), and high intramolecular flexibility, intrinsic coils being the ones with the lower degree of intramolecular packing density, residual secondary structure, and compactness (*3*). Despite the increasing number of high-resolution NMR studies of the structure and dynamics of these proteins (see ref *20* and review *21*, and references therein), there is at present no simple way to estimate the buried surface of a natively unfolded protein; in contrast, their determination for natively globular proteins constitutes a standard procedure. There thus seems to be a need for new approaches for further characterizing natively unfolded proteins.

DSC[1] studies have been reported in the literature as being sensitive to the presence of residual structure in natively unfolded proteins (*22*). Recently, a systematic heat capacity analysis of two natively disordered fragments of oxidized *Escherichia coli* thioredoxin (Trx) by Sánchez-Ruiz and Tasayco (*23*) showed that the experimental absolute heat capacity of fragment 1−73 is lower than the one predicted for a fully solvated polypeptide (based on model compounds (*24*, *25*)] by 1.7 kJ K^{-1} mol^{-1} at 25 °C, reflecting the burial of at least 1000 $Å^2$ of predominantly hydrophobic surface. These authors have also shown that the experimental absolute heat capacity of complementary fragment 74−108 is predicted well by these models, indicating that it is a fully solvated polypeptide. Although the presence of residual structure in both fragments was unclear from far-UV CD and ^1H NMR standard chemical shift analysis (*26*−*28*), recent NMR relaxation analysis of these fragments (*28*) has demonstrated the presence of a temperature-dependent transient hydrophobic effect only in fragment 1−73, consistent with the heat capacity analysis.

† This work was supported by a National Science Foundation (NSF) grant to M.L.T., an NSF US-Spain Cooperative Award, and the RCMI grant to City College of New York.

* To whom correspondence should be addressed.

‡ City College of New York.

§ Imagine Software.

[1] Abbreviations: CD, circular dichroism; GnHCl, guanidinium hydrochloride; Trx, oxidized *E. coli* thioredoxin; ASA, accessible surface area; DSC, differential scanning calorimetry.

10.1021/bi0271906 CCC: $25.00
Published on Web 02/27/2003

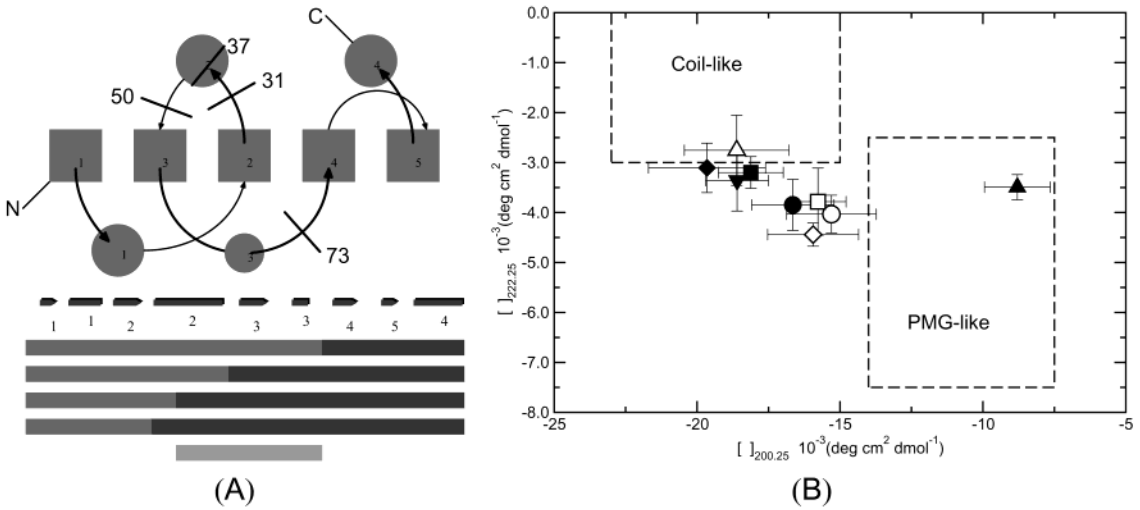

FIGURE 1: Family of natively disordered complementary fragments of Trx. (A) Scheme of the three-dimensional structure of Trx depicting cleavage sites from double variants with a single Arg at different sites. The helices and β-strands are depicted as circles and squares, respectively, as well as along the primary sequence using arrows and rectangles, respectively. The family of N-fragments (1−31, 1−37, 1−50, and 1−73), C-fragments (32−108, 38−108, 51−108, and 74−108), and the middle fragment (38−73) are shown along the primary sequence using horizontal bars in dark gray, black, and light gray, respectively. (B) Plot of molar ellipticity values at 222 nm vs those at 200 nm for the family of natively disordered fragments of Trx: 1−31 (△), 1−37 (○), 1−50 (□), 1−73 (◇), 32−108 (▲), 38−108 (●), 51−108 (■), 74−108 (◆), and 38−73 (▼). The regions most commonly populated by the intrinsic coil-like (Coil-like) and pre-molten globule-like (PMG-like) natively unfolded proteins are shown boxed with dashed lines (*3*).

This proven sensitivity of DSC to the presence of residual structure that cannot be detected by standard low- and high-resolution spectroscopic analysis prompted us to extend the studies to a family of (putatively) natively disordered complementary fragments of Trx with the goal of recognizing intrinsic coils with predominantly hydrophobic buried surface and determining the region, or regions, responsible for that burial. Here we report a systematic heat capacity analysis of a family of complementary fragments of Trx (1−31/32−108, 1−37/38−108, 1−50/51−108, and 38−73) (see Figure 1A) that behave like intrinsic coils under physiological conditions according to far-UV circular dichroism spectroscopy. This analysis not only recognizes the presence of residual structure in the majority of these putative intrinsic coils and provides an estimate for their hydrophobic buried surface but also gives insight into the underlying local and nonlocal contributions to this burial.

MATERIALS AND METHODS

Design and Production of the Complementary Fragments. To produce natively disordered fragments, we used variants of Trx with a single cleavage site, which separates the sequence of β_2 and β_4 and thus prevents the formation of folded fragments (*26*). To accomplish this, we moved the single R73 of Trx to the interconnecting regions between β_2 and α_2 as well as α_2 and β_3 to produce W31R/R73A Trx and Q50R/R73A Trx, respectively (see Figure 1A). Proteolytic cleavage of these double variants gave fragments 1−31, 32−108, 1−50, and 51−108, while chemical and/or proteolytic cleavage of wild-type Trx provided fragments 1−37, 38−108, and 38−73 (see Figure 1A).

Purification of Fragments. Double mutants of Trx were overexpressed in *E. coli* JF521 and purified by gel filtration and reverse phase chromatography (*23*). The wild-type Trx and their double mutants were either proteolytically or chemically cleaved into complementary fragments according to previously reported procedures (*23*), yielding the following fragments: 1−31, 1−37, 1−50, 1−73, 38−73, 32−108, 38−108, 51−108, and 74−108. These fragments were subsequently purified by reverse phase chromatography, identified by electrospray mass spectroscopy, and stored in lyophilized form. The concentrations were determined spectrophotometrically using the average molar mass for all the fragments with the exception of fragment 1−37 (for which the average of the closed and open states of the homoserine was used) and the following molar extinction coefficients: ϵ_{280} = 14 100 for fragment 1−73, ϵ_{215} = 39 700 for fragment 74−108, ϵ_{280} = 2560 for fragment 38−73, ϵ_{280} = 11 500 for fragment 1−37, ϵ_{280} = 2560 for fragment 38−108, ϵ_{280} = 5690 for fragment 1−31, ϵ_{280} = 2560 for fragment 32−108, ϵ_{280} = 1280 for fragment 51−108, and ϵ_{280} = 12 660 for fragment 1−50. The lyophilized samples were resuspended in 7 M guanidine hydrochloride, 10 mM potassium phosphate (KP_i) buffer at pH 7.5 and subsequently dialyzed against KP_i buffer and concentrated. All the solutions were in 10 mM KP_i, with the exception of fragment 32−108 for which 25 mM KP_i was used.

Characterization of Fragments. Serial dilutions of freshly prepared solutions of fragments were made for monitoring the concentration dependence of the isolated fragments using size exclusion chromatography and far-UV circular dichroism spectroscopy. The elution volume of the isolated fragments in solution was determined for a range of concentrations from 15 to 300 μM using a Superdex peptide HR 10/30 column equilibrated with 100 mM KP_i (pH 7.00) and a Biotech model LCC-500 liquid chromatography system. The far-UV CD spectra of the same solutions of fragments from 190 to 250 nm were obtained using an AVIV-60DS instrument with temperature control and cells with a path length of 1 mm at 25 °C. Likewise, the thermograms for the same range of concentrations of these fragments were acquired with a VP-DSC microcalorimeter from MicroCal at a scan rate of 1.5 K/min and a pressure of 207 kPa according to previously reported DSC measurements (*23*).

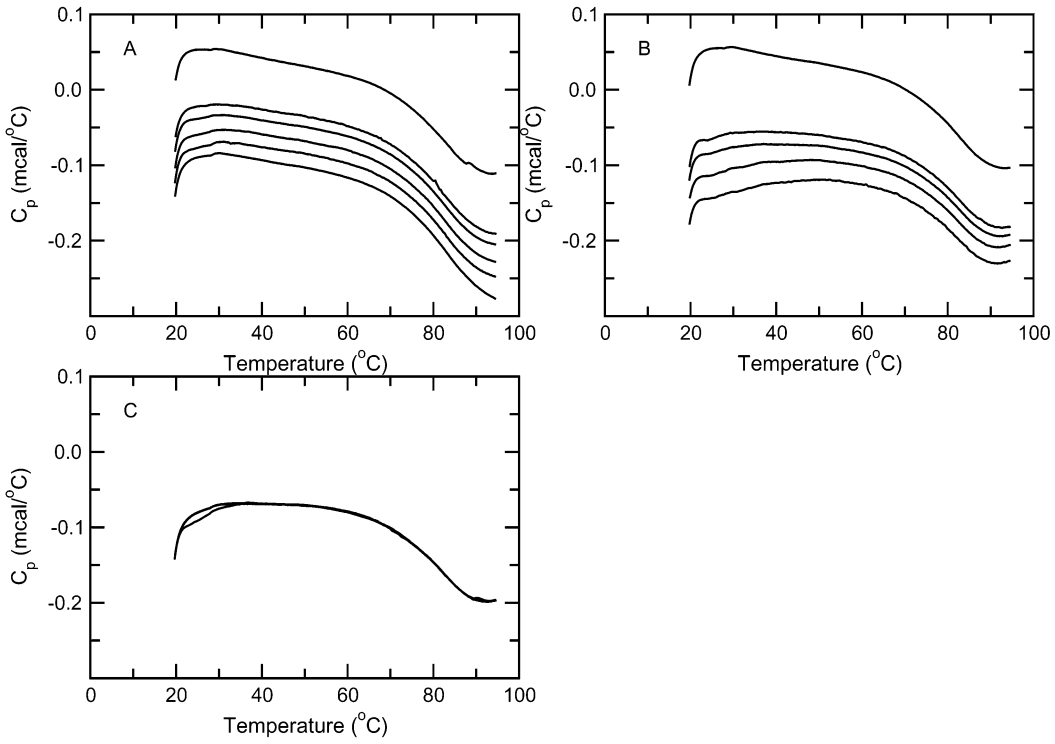

FIGURE 2: Original DSC thermograms of selected fragments in solution with the corresponding buffer−buffer baseline. (A) From top to bottom are shown the last baselines acquired prior to measurements with fragment 1−37 and its DSC thermograms at increasing concentrations of 0.085, 0.102, 0.132, 0.152, and 0.189 mM. (B) From top to bottom are shown the last baselines acquired prior to measurements with fragment 38−108 and its DSC thermograms at increasing concentrations of 0.065, 0.079, 0.096, and 0.120 mM. (C) Overlay of DSC thermograms of heating: first and second reheating for a solution of fragment 38−73 without removal of the sample from the cell of the microcalorimeter.

Fitting the Concentration Dependence of DSC Thermograms of Protein Fragments To Obtain Their Absolute Heat Capacities. Accurate values for the experimental absolute heat capacities ($C_{p,\mathrm{P}}$) of protein fragments were derived from the concentration dependence of the apparent heat capacity function ($\Delta C_{p,\mathrm{app}}$), obtained upon subtracting the buffer−buffer baseline from the heat capacity of a given fragment concentration (*29*). Several DSC thermograms corresponding to different concentrations of fragments were analyzed using eq 1 according to a previously reported procedure (*23*):

$$\Delta C_{p,\mathrm{app}} = CV_0\left(C_{p,\mathrm{P}} - \frac{V_\mathrm{P}}{V_\mathrm{W}}C_{p,\mathrm{W}}\right). \qquad (1)$$

where C is the protein concentration, V_0 is the calorimetric cell volume, and V_P and V_W are the specific volumes of the protein and water, respectively. The V_P values were calculated on the basis of amino acid composition (*30*).

Calculation of the Absolute Heat Capacity of a Completely Solvent-Exposed Fragment. The absolute heat capacity (C_p^U) of a fully hydrated protein fragment was calculated according to the amino acid composition and eq 2 (*24*, *25*). The polynomial representations of the Privalov and Makhatadze data (*24*, *25*) given by Freire (*31*) and the ones corresponding to the Häckel data (*32*) were used to determine the temperature dependence of C_p^U:

$$C_p^\mathrm{U} = \sum_{i=1}^{20} n_i C_{p,i} + (N-1)C_{p,\mathrm{NH-CO}} + C_{p,\mathrm{NH_2}} + C_{p,\mathrm{COOH}} \qquad (2)$$

where $C_{p,i}$ is the contribution from the ith residue (without the end termini) and the other three terms take into account the contribution from the peptide bond and the N- and C-termini, respectively.

Estimate of the Buried Surface Area of a Polypeptide. The buried surface area of a protein, or protein fragment, in the native state is related to the difference between the absolute heat capacity of the unfolded and folded state (ΔC_p) of a given polypeptide through an empirical equation (eq 3) (*33*, *34*), where $\Delta\mathrm{ASA_{np}}$ and $\Delta\mathrm{ASA_p}$ are the changes in solvent-exposed surface area due to the unfolding of nonpolar and polar residues, respectively, which are measured in square angstroms. The coefficients are known to be positive (nonpolar) and negative (polar) with values in kilojoules per mole per square angstrom essentially independent of temperature in the range of 25−40 °C:

$$\Delta C_p = 1.84 \times \Delta\mathrm{ASA_{np}} - 1.05 \times \Delta\mathrm{ASA_p} \qquad (3)$$

While this relation was derived from studies of natively folded soluble proteins (*34*), we will assume that it is also valid for protein fragments with residual structure. The estimates were calculated assuming also that $\Delta\mathrm{ASA_p}$ and $\Delta\mathrm{ASA_{np}}$ regions are buried in the ratio found in native proteins ($\Delta\mathrm{ASA_p}/\Delta\mathrm{ASA_{np}} = 0.59$).

RESULTS

Characterization of the Family of Natively Disordered Complementary Fragments from Oxidized E. coli Thioredoxin. Size-exclusion chromatography and the molar

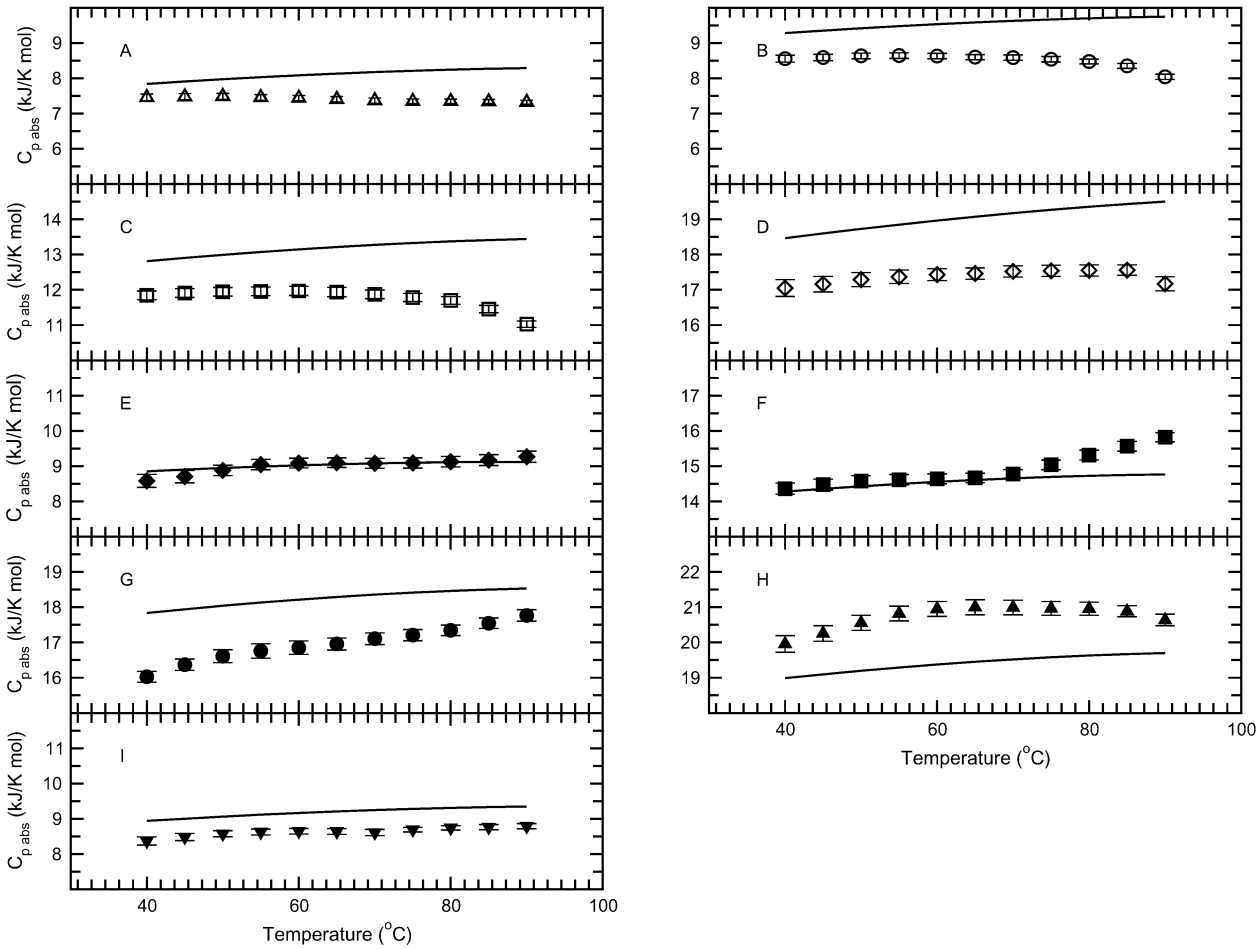

FIGURE 3: Absolute heat capacity vs temperature profiles for the family of fragments. The $C_{p,P}$ and C_p^U values of these fragments are depicted with symbols and solid lines, respectively: (A) 1−31, (B) 1−37, (C) 1−50, (D) 1−73, (E) 74−108, (F) 51−108, (G) 38−108, (H) 32−108, and (I) 38−73. The error bars are shown on the plot, although in most cases they are smaller than the size of the symbols. The $C_{p,P}$ and C_p^U values for fragments 1−73 and 74−108 were taken from ref *23*.

ellipticity from 190 to 250 nm of the far-UV CD spectrum of a freshly prepared solution of fragments throughout a range of concentrations (from 0.3 to 1.3 mg/mL) show no concentration dependence (data not shown) and argue in favor of their monomeric state. In all cases, the far-UV CD spectra of these fragments show a minimum around 200 nm, diagnostic of the so-called "random coil state" (data not shown). Further analysis of their far-UV CD spectra using the double-wavelength plot of the ellipticity values at 222 and 200 nm and comparison with the areas populated by premolten globules and intrinsic coils with known hydrodynamic ratios (*3*) (see Figure 1B) indicate that fragment 32−108 behaves like a somewhat compact natively unfolded protein (premolten globule), while the others behave like coils or intrinsic coils (*3*).

Heat Capacity Analysis of the Family of Fragments. The reproducibility of the thermograms for the solutions of fragments 1−37, 38−73, and 38−108 (see Figure 2) upon heating and reheating them more than once demonstrates the reversibility of the temperature-dependent phenomena from 40 to 90 °C. These traces show no obvious cooperative unfolding transitions. For any given fragment, analysis of these thermograms at increasing concentrations according to eq 1 yields linear plots (see representative plots at 40 °C in Figure 4) with a slope that gives the $C_{p,P}$ value at the selected temperature. In this analysis, the $C_{p,P}$ values, the

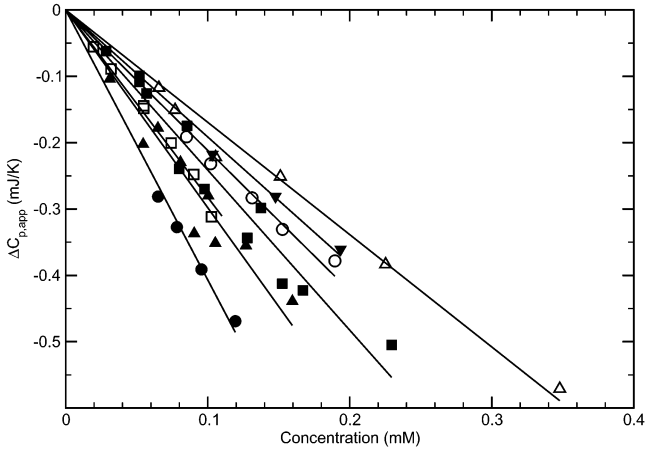

FIGURE 4: Plot of the apparent heat capacity vs the concentration for the family of natively disordered fragments of Trx. The best linear fits with a zero intercept of the data for these fragments are shown at pH 7.5 and 40 °C in a decreasing trend of $\Delta C_{p,app}$ values: 38−108 (●), 32−108 (▲), 1−50 (□), 51−108 (■), 1−37 (○), 38−73 (▼), and 1−31 (△).

C_p^U values obtained with the model of Makhatadze and Privalov (*24*, *25*), and the empirical equation (eq 3) (see Table 1) give ΔC_p values at 40 °C of 0.4 ± 0.1, 0.7 ± 0.1, 1.0 ± 0.1, 0.6 ± 0.1, −1.0 ± 0.2, 1.8 ± 0.1, and −0.1 ± 0.3 kJ/mol for fragments 1−31, 1−37, 1−50, 38−73, 32−

Table 1: Changes in Accessible Surface Area (ASA, in square angstroms) for the Family of Fragments[a]

fragment	ΔC_p[b]	ΔASA_{np}[b]	ΔASA_p[b]	ΔC_p[c]	ΔASA_{np}[c]	ΔASA_p[c]
1−31	0.4 ± 0.1	310 ± 60	180 ± 60	1.4 ± 0.1	1150 ± 60	680 ± 60
1−37	0.7 ± 0.1	600 ± 60	350 ± 60	2.0 ± 0.1	1620 ± 60	960 ± 60
1−50	1.0 ± 0.1	790 ± 60	470 ± 60	2.5 ± 0.1	2030 ± 60	1190 ± 60
1−73	1.4 ± 0.2	1160 ± 110	680 ± 110	3.5 ± 0.2	2890 ± 110	1710 ± 110
32−108	−1.0 ± 0.2	N/A	N/A	1.5 ± 0.2	1200 ± 60	708 ± 60
38−108	1.8 ± 0.1	1480 ± 70	870 ± 70	4.0 ± 0.2	3240 ± 70	1910 ± 70
51−108	−0.1 ± 0.3	N/A	N/A	2.0 ± 0.3	1480 ± 130	870 ± 130
74−108	0.3 ± 0.2	220 ± 70	130 ± 70	1.4 ± 0.2	1120 ± 70	660 ± 70
38−73	0.6 ± 0.1	470 ± 60	270 ± 60	1.6 ± 0.1	1320 ± 60	780 ± 60

[a] Estimates were calculated using eq 3 and assuming that polar (ΔASA_p) and nonpolar (ΔASA_{np}) regions are buried in the ratio found in native proteins ($\Delta ASA_p/\Delta ASA_{np} = 0.59$). The ΔC_p values in kilojoules per kelvin per mole were calculated at 40 °C. ΔC_p values (kilojoules per kelvin per mole) of fragments 1−73 and 74−108 at 40 °C were taken from the literature (*23*). [b] Using Makhatadze et al. (*24*, *25*). [c] Using Häckel et al. (*32*).

108, 38−108, and 51−108, respectively. For most of the fragments, the ΔC_p values are positive (see Figure 3). The exceptions are fragment 32−108, which gives negative values throughout the whole temperature range, and 51−108, which becomes negative beyond 66 °C. It is worth noticing that all these values would be positive if we had instead used the model of Häckel et al. (*22*, *32*) to calculate the C_p^U (see Table 1). All these traces are complex, and it is not clear to us how to interpret the temperature dependence. In any case, the ΔC_p values at 40 °C that are positive regardless of the model have been converted to changes in nonpolar and polar solvent accessible surface area according to eq 3 and are shown in Table 1, which also includes the values previously obtained for fragments 1−73 and 74−108 at the same temperature (*23*). The solvent accessible surface area changes corresponding to the ΔC_p values obtained with the model of Makhatadze and Privalov (*24*, *25*) indicate the essentially fully solvated state of fragments 51−108 and 74−108 and a predominantly hydrophobic surface area that increases with fragment length for fragments 1−31, 38−73, 1−37, 1−50, 1−73, and 38−108. Fragment 32−108 is intriguing because of its anomalous negative ΔC_p value (see Table 1) and its apparently somewhat compact natively unfolded (premolten globule) nature based on the far-UV CD spectra (see Figure 1B). Whether this negative ΔC_p value reflects the fact that premolten globules show a ratio between the solvent accessible surface area changes of polar (ΔASA_p) and nonpolar (ΔASA_{np}) higher than the one for native globular proteins is unclear.

DISCUSSION

ΔC_p Values Are Good Predictors of Predominantly Hydrophobic Buried Surface in Natively Disordered Proteins that Behave like Intrinsic Coils. The ΔC_p value of a given polypeptide depends on the calculation of the C_p^U value of its fully solvated state, which is based on the principle of group additivity and dependent only on its amino acid composition. However, as discussed above, the contributions from the individual amino acids are model-dependent (*25*, *32*), and it is difficult to assess the correct ΔC_p value. For some natively disordered fragments of Trx that behave like intrinsic coils (fragments 1−73 and 74−108), the ΔC_p values calculated using Makhatadze's model (*24*, *25*) are known to correlate well with the presence and absence of the transient hydrophobic effect observed by NMR relaxation measurements (*28*). We take this as proof that the heat capacity analysis recognizes the presence of predominantly hydrophobic buried surface in intrinsic coils, even those that go undetected by far-UV CD and ^{1}H NMR chemical shift analysis (*27*, *28*). An important assumption made in this study is that the ΔC_p values of the family of natively disordered fragments of Trx reflect the behavior of their monomers, despite the general tendency of protein fragments to aggregate (*35*−*41*). In our case, the linear concentration dependence of both the molar ellipticity values (data not shown) and $\Delta C_{p,app}$ values (see Figure 4) demonstrates the predominance of monomers in the freshly prepared solutions of all these fragments at the selected range of concentrations (0.3 and 1.3 mg/mL) and in the reversible region of the DSC measurements.

The heat capacity analysis of the whole family of intrinsic coils derived from Trx (fragments 1−31, 1−37, 1−50, 1−73, 38−108, 51−108, 74−108, and 38−73) shows that the N-fragment and middle fragment have a predominantly hydrophobic buried surface that increases with the length of the fragment, while only one of the C-fragments (38−108) shows this type of burial. No simple explanation is readily available. According to the literature (*19*), one would expect a negative correlation between the buried surface area and the combination of a high net charge and low hydrophobic content, but that is not what we observe. For instance, the N-fragments have a higher ratio of net charge over a number of nonpolar residues than the C-fragments and sometimes show more burial. More unsettled questions involve the reasons for the premolten globule nature and the anomalous negative ΔC_p value of fragment 32−108. A possible source of hydrophobic burial is the formation of a ring between the backbone and the C32−C35 disulfide and its subsequent docking to the nearby W31 and W28 residues, but it cannot be the only reason for hydrophobic burial because fragments 1−31, 38−73, and 38−108 show this kind of burial in its absence. Indeed, the transient hydrophobic effect detected in fragment 1−73 by NMR relaxation analysis remains, although the amount is diminished, upon reduction of the disulfide (*28*). However, the burial in fragment 1−37, obtained by CNBr cleavage of Trx, may be due to both the closed (lactone) form of the C-terminal homoserine and the C32−C35 disulfide ring. The trend of the $C_{p,P}$ values of the N-fragments to decrease with the increase in temperature may be explained by the enhancement of the hydrophobic effect in polypeptides at higher temperatures (*42*). The majority of the remaining fragments, such as fragments 51−

Table 2: Nonadditivity of the ΔC_p Values (in kilojoules per kelvin per mole) for the Family of Natively Disordered Complementary Fragments from Trx[a]

fragments	$\sum\Delta C_p$[b]	$\sum\Delta C_p$[c]
1−50 and 51−108	1.0 ± 0.4	4.3 ± 0.4
1−73 and 74−108	1.7 ± 0.3	4.9 ± 0.3
1−37 and 38−108	2.5 ± 0.2	5.9 ± 0.2
1−37, 38−73, and 74−108	1.7 ± 0.4	5.0 ± 0.4
1−73	1.4 ± 0.2	3.5 ± 0.2
1−37 and 38−73	1.3 ± 0.2	3.6 ± 0.2
38−108	1.8 ± 0.2	4.0 ± 0.2
38−73 and 74−108	0.8 ± 0.3	3.0 ± 0.3

[a] What is shown is the sum of the ΔC_p values in kilojoules per kelvin per mole at 40 °C. [b] Calculated using Makhatadze et al. (*24*, *25*) for the sets of complementary fragments. [c] Calculated using Häckel et al. (*32*) for the sets of complementary fragments.

108 and 38−108, show the inverse trend, and assessing whether it is associated with the burial of polar residues requires further studies.

Additivity of Absolute Heat Capacity Values Provides Insight about Local and Nonlocal Interactions. Since the estimate of the hydrophobic buried surface in fragment 1−73 arising from the ΔC_p values according to the model of Makhatadze and Privalov (*24*, *25*) is consistent with the transient hydrophobic effect found by NMR relaxation analysis (*28*), we will continue our discussion using ΔC_p values based on this model unless stated otherwise.

To evaluate the contribution of nonlocal interactions in the burial of predominantly hydrophobic surface found in this family of natively disordered fragments (see Table 1), we tested the additivity of the ΔC_p values at 40 °C in two ways (see Table 2). First, the ΔC_p values of fragments 1−73 and 38−108 were compared to the sum of the ΔC_p values from their corresponding pairs of complementary fragments. The fact that the ΔC_p values of fragments 1−37 and 38−73 add up to give the ΔC_p of fragment 1−73 demonstrates the additivity of those values and implies that the predominantly buried surface of fragment 1−73 comes from the accumulation of buried surface from both regions of the sequence. In contrast, the nonadditivity for fragment 38−108 reflects that besides the accumulation of buried surface arising from the complementary regions (fragments 38−73 and 74−108), there is a contribution from interactions between those regions. As a second test, we focused on the complementary fragments that comprise the whole sequence of thioredoxin. In this case, we too observe different ΔC_p values, even considering the experimental errors, which demonstrate their nonadditivity and imply the participation of interactions among various regions in the surface burial. In both additivity tests, we have assumed that changes in ΔC_p values at 40 °C are not due to the double mutation of Trx (Q50R/R73A) or the CNBr chemical modification of the C-terminal Met into the open and closed homoserine in fragment 1−37. At least the difference between the theoretical C_p^U sum for fragments 1−73 and 74−108 (obtained from wild-type Trx) and the one for fragments 1−50 and 51−108 (obtained from the W50R/R73A mutant of Trx) is negligible. This does not preclude, however, the possibility that the difference between the sums for fragments 1−37 and 38−108 and the one for fragments 1−73 and 74−108 may be enhanced by the closed form of the C-terminal homoserine in fragment 1−37.

To further the analysis and identify potential local and nonlocal interactions, we have chosen (see the Appendix) to represent these fragments as a combination of five nonoverlapping segments (S1−31, S32−37, S38−50, S51−73, and S74−108) and consider only their binary interactions. The ΔC_p of a given individual fragment, with the exception of fragment 32−108, is represented as the sum of the ΔC_p values of their segments and interactions. As previously indicated, the ΔC_p values have contributions from both polar and apolar groups; the former are expected to be negatively and the latter positively correlated upon solvent exposure (*33*, *34*). In the case of the unfolding of typical soluble folded proteins, the apolar contribution is known to outweigh the polar one (eq 3). Since there is no evidence to the contrary, we will assume that this is the case also for the segments, which implies that the values of ΔC_p for the segment and their interactions are non-negative and the burial is predominantly hydrophobic.

Analysis of the fragments with positive ΔC_p values indicates an increase with increasing fragment length, allowing us to extract useful constraints about the interactions between a fragment, or segment(s), and the segment used for its extension. For example, adding S38−50 to fragment 51−108 produces a larger increase in ΔC_p than adding the same segment to fragment 1−37, revealing that the interactions between the first pair of regions are stronger than the other one. Similar results are obtained when we analyze the extension of fragments 1−50 and 74−108 with the addition of S51−73.

A more systematic analysis (see the Appendix) requires solving the linear equations that relate the segment contributions and their interactions (the unknowns, or dependent variables) to the observed ΔC_p of the fragments (the independent variables). Unfortunately, there are more dependent variables (13) than equations (8). However, these equations can be supplemented with the fact that the contributions cannot be negative if they are to yield constraints on the dependent variables. In the Appendix, a Monte Carlo (*43*) approach is described, in which the statistical distributions of the dependent variables can be obtained, consistent with the experimental uncertainties. In this approach, one generates "random" values for the independent variables (consistent with the experimental uncertainties) and for each such selection solves the equations for the dependent variables. The collection of results thus obtained provides an estimate of their statistical distributions. One complication in this case is, of course, the fact that there are more dependent variables than equations. To get around this and reduce the problem to only eight dependent variables, five of them are arbitrarily chosen and their values treated as random. Without the positivity constraint, this would not help much, since a solution for the eight remaining dependent variables could be found for any random selection. However, the positivity constraint limits the possible solutions, so information about all the dependent variables can be obtained.

This analysis, regardless of the model used to obtain the ΔC_p values, yields constraints on eight of the nine possible binary interactions (see Tables 3 and 4 and the Appendix) and predicts that segment S1−31 contributes more than other segments to the buried surface of the N-fragments (1−37, 1−50, and 1−73), and that the most significant interactions

Table 3: Discrepancy between Experimental and Calculated Absolute Heat Capacity Values (ΔC_p) for Selected Segments of Trx and Their Interactions[a]

T (°C)		1−31	32−37	38−50	51−73	74−108
40		0.3 ± 0.1	0.1 ± 0.1	0.1 ± 0.1	0.1 ± 0.1	0.2 ± 0.1
	1−31		0.2 ± 0.1	0.1 ± 0.1	0.1 ± 0.1	
	32−37			0.1 ± 0.1	0.1 ± 0.1	
	38−50				0.3 ± 0.2	1.0 ± 0.3
	51−73					0.1 ± 0.1
50		0.4 ± 0.1	0.1 ± 0.1	0.1 ± 0.1	0.1 ± 0.1	0.1 ± 0.1
	1−31		0.2 ± 0.1	0.1 ± 0.1	0.1 ± 0.1	
	32−37			0.1 ± 0.1	0.1 ± 0.1	
	38−50				0.3 ± 0.2	0.7 ± 0.3
	51−73					0.1 ± 0.1
60		0.6 ± 0.1	0.1 ± 0.1	0.1 ± 0.1	0.1 ± 0.1	0.1 ± 0.1
	1−31		0.2 ± 0.1	0.1 ± 0.1	0.1 ± 0.1	
	32−37			0.1 ± 0.1	0.1 ± 0.1	
	38−50				0.3 ± 0.2	0.6 ± 0.3
	51−73					0.1 ± 0.1
70		0.8 ± 0.1	0.1 ± 0.1	0.1 ± 0.1	0.1 ± 0.1	0.1 ± 0.1
	1−31		0.2 ± 0.1	0.1 ± 0.1	0.1 ± 0.1	
	32−37			0.1 ± 0.1	0.1 ± 0.1	
	38−50				0.4 ± 0.1	0.4 ± 0.2
	51−73					0.1 ± 0.1

[a] Estimates were calculated for segments 1−31, 32−37, 38−50, 51−73, and 74−108 (see the Appendix) at selected temperatures using ΔC_p values based on Makhatadze et al. (*24*, *25*). ΔC_p values of fragments 1−73 and 74−108 at 40 °C were taken from the literature (*23*).

Table 4: Discrepancy between Experimental and Calculated Absolute Heat Capacity Values (ΔC_p) for Selected Segments of Trx and Their Interactions[a]

T (°C)		1−31	32−37	38−50	51−73	74−108
40		1.4 ± 0.1	0.2 ± 0.2	0.3 ± 0.2	0.3 ± 0.2	1.3 ± 0.2
	1−31		0.3 ± 0.1	0.1 ± 0.1	0.1 ± 0.1	
	32−37			0.1 ± 0.1	0.1 ± 0.1	
	38−50				0.3 ± 0.2	1.0 ± 0.3
	51−73					0.3 ± 0.3
50		1.5 ± 0.1	0.2 ± 0.2	0.4 ± 0.2	0.5 ± 0.3	1.2 ± 0.2
	1−31		0.3 ± 0.1	0.1 ± 0.1	0.1 ± 0.1	
	32−37			0.1 ± 0.1	0.1 ± 0.1	
	38−50				0.6 ± 0.3	0.8 ± 0.3
	51−73					0.2 ± 0.2
60		1.7 ± 0.1	0.2 ± 0.2	0.4 ± 0.2	0.7 ± 0.3	1.2 ± 0.1
	1−31		0.3 ± 0.1	0.1 ± 0.1	0.1 ± 0.1	
	32−37			0.1 ± 0.1	0.1 ± 0.1	
	38−50				0.6 ± 0.3	0.8 ± 0.3
	51−73					0.2 ± 0.2
70		1.8 ± 0.1	0.2 ± 0.1	0.5 ± 0.2	0.7 ± 0.3	1.3 ± 0.2
	1−31		0.3 ± 0.1	0.1 ± 0.1	0.1 ± 0.1	
	32−37			0.1 ± 0.1	0.1 ± 0.1	
	38−50				0.6 ± 0.2	0.5 ± 0.3
	51−73					0.2 ± 0.2

[a] Estimates were calculated for segments 1−31, 32−37, 38−50, 51−73, and 74−108 (see the Appendix) at selected temperatures using ΔC_p values based on Häckel et al. (*32*).

arise between segments S1−31 and S32−37, S38−50 and S51−73, and S38−50 and S74−108. These predictions reflect the contribution of local and nonlocal interactions to the predominantly hydrophobic buried surface of the N-fragments. The predicted predominantly hydrophobic surface of S1−31 and the nonlocal (medium-range) interactions between S1−31 and S32−37 correlate well with the restricted motions of the hydrophobic segment of residues 25−39 found by the NMR relaxation studies of fragment 1−73 (*28*). However, these studies fail to detect the constrained motions of the backbone that one would expect from other local and nonlocal interactions. It appears that far-UV CD spectroscopy and DSC measurements are very sensitive for detection of weak interactions that produce predominantly hydrophobic buried surface within natively disordered polypeptides that behave like intrinsic coils, which can be missed by standard high-resolution spectroscopic experiments, but whose presence can be tested with more elaborate NMR experiments.

Nature of the Predominantly Hydrophobic Buried Surface in the Natively Disordered Fragments from Trx. One of the intriguing questions about natively disordered protein fragments is how well their residual structure is represented in the native state ensemble of the parent protein (*44*). If it is well-represented, one would expect, for example, that fragments containing the rather rigid hydrophobic central β-strands (β_2−β_4) of native Trx (*45*, *46*) will contain a noticeable hydrophobic cluster. Indeed, recent DSC studies of fragments encompassing these central β-strands show a broad unfolding transition (M. L. Tasayco and C. Mendoza, unpublished results). Although the majority of fragments that include the sequence of one or two of these β-strands (1−31, 1−37, 1−50, 1−73, 38−73, and 38−108) show this kind of surface burial, a pair of fragments containing β_3 and/or β_4 (51−108 and 74−108) behave as fully solvent-exposed. It thus appears that the presence of the sequence of β_2 might be sufficient to produce burial, but that is not the case for β_3 or β_4.

The predicted interactions between segments S1−31 and S32−37, S38−50 and S51−73, and S38−50 and S74−108 are, however, consistent with native-like interactions involving helical regions such as the packing of β_2 against the N-terminal portion of α_2 (containing the disulfide ring between C32 and C35 and the interconnecting backbone), the docking of β_3 against α_2, and the packing of α_2 against α_4, respectively (see Figure 1A). Following this line of thought, the presence of the last two interactions may explain the significant burial in fragment 38−108. Likewise, the burial promoted by the disulfide ring preceding α_2 and the packing of α_2 against α_4 may explain the pre-molten globule state of fragment 32−108. These predictions, however, await validation by future NMR studies.

CONCLUSIONS

This work and previous work (*23*) have shown that DSC constitutes a rapid and sensitive tool for determining if a fragment that is an intrinsic coil according to far-UV CD measurements nonetheless has burial of predominantly hydrophobic regions. A positive ΔC_p is a telltale sign that the fragment behaves like a globular soluble protein, which buries its hydrophobic patches to a greater extent than the polar ones. Furthermore, assuming that this behavior extends to the relation between ΔC_p and the changes in accessible surface area (ASA) upon burial, one can obtain estimates of these changes and thus estimate the degree of folding in the intrinsically unstructured fragment.

A systematic analysis of the residual structure in a family of intrinsically unstructured fragments of *E. coli* Trx reveals interactions between segments that reflect existing local and nonlocal interactions between regions of secondary structure in the native parent protein. In our studies, we have found no evidence for non-native-like interactions.

ACKNOWLEDGMENT

We thank Roxana Georgescu for providing the plasmids of the Trx variants, María M. Herrera for expert technical assistance, MLTJ's group for generating Trx and its variants, José Manuel Sánchez-Ruiz from the Chemistry Department of Universidad de Granada (Granada, Spain) for allowing us to carry out the DSC measurements in his lab, and his group for assistance with data analysis. Finally, we thank the mass spectroscopy facility at Hunter College (City University of New York, New York, NY).

APPENDIX

Herein we analyze the equations that relate the difference between the calculated predictions for a fully hydrated polypeptide and the experimentally observed heat capacities (ΔC_p) with the per-segment values and their interactions. For the sake of concreteness, we will perform the analysis at 40 °C.

To simplify the analysis, we will use a notation slightly different than that in the main text. Let's start by numbering the segments as follows:

segment	residues
1	1−31
2	32−37
3	38−50
4	51−73
5	74−108

Using this numbering scheme, we will denote with ΔC_i the ΔC_p of the isolated ith segment and with $\Delta C_{i,j}$ the contribution of the interaction between segments i and j. With this notation we can write, for example

$$\Delta C_p(s_i + s_j) = \Delta C_i + \Delta C_j + \Delta C_{i,j}$$

where we denoted with $s_i + s_j$ the union of segments i and j. Notice that in the absence of interactions the ΔC_p behaves additively, since the nonadditive contribution from the terminal ends is assumed to cancel out when taking the difference between the theoretical and experimental values.

The experimentally observed ΔC_p values (at 40 °C) can thus be expressed in the following way:

$$\Delta C_p(s_1) = 0.38 \pm 0.10$$

$$\Delta C_p(s_1 + s_2) = 0.73 \pm 0.10$$

$$\Delta C_p(s_3 + s_4) = 0.57 \pm 0.11$$

$$\Delta C_p(s_5) = 0.27 \pm 0.18$$

$$\Delta C_p(s_4 + s_5) = 0.08 \pm 0.16$$

$$\Delta C_p(s_3 + s_4 + s_5) = 1.81 \pm 0.15$$

$$\Delta C_p(s_1 + s_2 + s_3 + s_4) = 1.42 \pm 0.24$$

$$\Delta C_p(s_1 + s_2 + s_3) = 0.96 \pm 0.12$$

These translate directly into the following eight equations in the 15 unknowns ΔC_i and $\Delta C_{i,j}$:

$$\Delta C_1 = 0.38 \pm 0.10$$

$$\Delta C_1 + \Delta C_2 + \Delta C_{1,2} = 0.73 \pm 0.10$$

$$\Delta C_3 + \Delta C_4 + \Delta C_{3,4} = 0.57 \pm 0.11$$

$$\Delta C_5 = 0.27 \pm 0.18$$

$$\Delta C_4 + \Delta C_5 + \Delta C_{4,5} = 0.08 \pm 0.16$$

$$\Delta C_3 + \Delta C_4 + \Delta C_5 + \Delta C_{3,4} + \Delta C_{3,5} + \Delta C_{4,5} = 1.81 \pm 0.15$$

$$\Delta C_1 + \Delta C_2 + \Delta C_3 + \Delta C_4 + \Delta C_{1,2} + \Delta C_{1,3} + \Delta C_{1,4} + \Delta C_{2,3} + \Delta C_{2,4} + \Delta C_{3,4} = 1.42 \pm 0.24$$

$$\Delta C_1 + \Delta C_2 + \Delta C_3 + \Delta C_{1,2} + \Delta C_{1,3} + \Delta C_{2,3} = 0.96 \pm 0.12$$

Since there are only eight equations and 15 unknowns, we cannot possibly hope to find a unique solution. However, notice first of all that the two quantities $\Delta C_{1,5}$ and $\Delta C_{2,5}$ appear in none of the equations, the former because it requires the whole protein and the latter because the only fragment that would contribute was not considered in the analysis due to its negative ΔC_p. Therefore, we can exclude them from the analysis, leaving us with only 13 unknowns (of course, this also means that $\Delta C_{1,5}$ and $\Delta C_{2,5}$ are unconstrained by these experiments). Of course, a system of eight equations on 13 unknowns does not have a unique solution. However, the equations do provide constraints and relations among the variables. The easiest way to see this is to consider five (13 − 8) of the unknown quantities as independent variables and to rewrite the eight equations above with those five variables "on the right-hand side". The choice of variables is rather arbitrary; in what follows, we will select $\Delta C_{1,2}$, $\Delta C_{1,3}$, $\Delta C_{1,4}$, $\Delta C_{3,4}$, and $\Delta C_{4,5}$ as independent variables. We can then rewrite the eight equations above in an equivalent manner

$$\Delta C_1 = 0.38 \pm 0.10$$

$$\Delta C_1 + \Delta C_2 = 0.73 \pm 0.10 - \Delta C_{1,2}$$

$$\Delta C_3 + \Delta C_4 = 0.57 \pm 0.11 - \Delta C_{3,4}$$

$$\Delta C_5 = 0.27 \pm 0.18$$

$$\Delta C_4 + \Delta C_5 = 0.08 \pm 0.16 - \Delta C_{4,5}$$

$$\Delta C_3 + \Delta C_4 + \Delta C_5 + \Delta C_{3,5} = 1.81 \pm 0.15 - \Delta C_{3,4} - \Delta C_{4,5}$$

$$\Delta C_1 + \Delta C_2 + \Delta C_3 + \Delta C_4 + \Delta C_{2,3} + \Delta C_{2,4} = 1.42 \pm 0.24 - \Delta C_{1,2} - \Delta C_{1,3} - \Delta C_{1,4} - \Delta C_{3,4}$$

$$\Delta C_1 + \Delta C_2 + \Delta C_3 + \Delta C_{2,3} = 0.96 \pm 0.12 - \Delta C_{1,2} - \Delta C_{1,3}$$

Now we are left with eight equations on eight unknowns (those appearing on the left-hand side), which are dependent, however, on the (also unknown!) five quantities that appear

Table 5: ΔC_p Value of Each Segment

s_1	s_2	s_3	s_4	s_5
0.33 ± 0.11	0.12 ± 0.11	0.11 ± 0.10	0.07 ± 0.07	0.12 ± 0.11

Table 6: Interactions between Segments

	s_2	s_3	s_4	s_5
s_1	0.18 ± 0.15	0.08 ± 0.08	0.12 ± 0.12	
s_2		0.08 ± 0.08	0.12 ± 0.12	
s_3			0.28 ± 0.17	1.07 ± 0.27
s_4				0.08 ± 0.08

on the right-hand side. Of course, we are no closer to finding a solution than we were before. However, if we make the plausible assumption that all the contributions have to be non-negative, we can use these constraints to obtain further information about the interactions. And this time these constraints will give us information about both the dependent and independent variables!

A systematic analysis of the system of equations shown above that takes into consideration the experimental errors as well as the positivity constraint can be done in a variety of ways. An elegant approach that can be extended to arbitrary systems of (not necessarily linear) equations is Monte Carlo simulation (*43*). In one variant of this approach, one generates a large number of samples consistent with the experimental errors and calculates the quantities of interest (in our case, the variables appearing on the left-hand side), obtaining thus an estimation of their statistical distribution. If enough samples have been generated, the solutions that remain after applying these constraints will fill a region of n-dimensional ($n = 13$ in this case) space and the distribution will give us a measure of the uncertainties in each variable. It should be emphasized that the random samples must include both the experimental values and the selected "independent" variables. The former are, as usual, assumed to be normally distributed, with a mean and standard deviation given by the experimental value and uncertainty. In contrast, the statistical distribution of the independent variables is not known (or, rather, much less known) and must be assumed; the weaker the assumptions, the more reliable the constraints that result from the analysis. Fortunately, the positivity constraints do impose limits on their values, and we can therefore make the minimal assumption that each of them is uniformly distributed within these bounds. For each sample thus obtained, the equations are then solved in terms of the independent variables and the positivity constraints are imposed by dropping the solutions that are inconsistent with them (that is, some variable attains a negative value). An analysis of the resulting distributions shows that they are not Gaussian, so the average and standard deviation are not good indicators. Instead, we have chosen to use the median and 75% confidence level to report the values and their errors.

Applying systematically the positivity constraints together with the equations found above, we arrive at the final results shown in Tables 5 and 6. Again notice that no information regarding the interactions between segments s_1 or s_2 and s_5 is available from these experiments. Notice also that some of the experimentally determined values, like the ΔC_p for segment s_1 (1−31), become shifted after the full analysis (although remaining consistent within the experimental uncertainties).

SUPPORTING INFORMATION AVAILABLE

Heteronuclear NOE (NHNOE) of fragment 1−73 from *E. coli* TRX with C32−C35 in the oxidized and reduced forms. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Dyson, H. J., and Wright, P. E. (2002) Coupling of folding and binding for unstructured proteins, *Curr. Opin. Struct. Biol. 12*, 54.
2. Uversky, V. N. (2002) What does it mean to be natively unfolded? *Eur. J. Biochem. 269*, 2.
3. Uversky, V. N. (2002) Natively unfolded proteins: A point where biology waits for physics, *Protein Sci. 11*, 739.
4. Wright, P. E., and Dyson, H. J. (1999) Intrinsically unstructured proteins: Re-assessing the protein structure−function paradigm, *J. Mol. Biol. 293*, 321.
5. Dunker, A. K., Lawson, J. D., Brown, C. J., Williams, R. M., Romero, P., Oh, J. S., Oldfield, C. J., Campen, A. M., Ratliff, C. M., et al. (2001) Intrinsically disordered proteins, *J. Mol. Graphics Modell. 19*, 26.
6. Laity, J. H., Dyson, H. J., and Wright, P. E. (2000) DNA-induced α-helix capping in conserved linker sequences is a determinant of binding affinity in Cys2-His2 zinc fingers, *J. Mol. Biol. 39*, 8747.
7. Laity, J. H., Dyson, H. J., and Wright, P. E. (2000) Molecular basis for modulation of biological function by alternate splicing of the Wilms' tumor suppressor protein, *Proc. Natl. Acad. Sci. U.S.A. 97*, 11932.
8. Agalarov, S. C., Prasad, G. S., Funke, P. M., Stout, C. D., and Williamson, J. R. (2000) Structure of the S15, S6, S18-rRNA complex: Assembly of the 30S ribosome central domain, *Science 288*, 107.
9. Ayed, A., Mulder, F. A., Yi, G. S., Lu, Y., Kay, L. E., and Arrowsmith, C. (2001) Latent and active p53 are identical in conformation, *Nat. Struct. Biol. 8*, 756.
10. Grossman, J. G., Sharff, A. J., O'Hare, P., and Luisi, B. F. (2001) Molecular shapes of transcription factors TFRIIB and VP16 in solution: Implications for recognition, *Biochemistry 40*, 6267.
11. Kartin, D., Longhi, S., Receveur, V., and Canard, B. (2002) The N-terminal domain of the phosphoprotein of Morbilliviruses belongs to the natively unfolded class of proteins, *Virology 296*, 251.
12. He, X., Chow, D., Martick, M. M., and García, K. C. (2001) Allosteric activation of a spring-loaded natriuretic peptide receptor dimer by hormone, *Science 293*, 1657.
13. Kim, A. S., Kakalis, L. T., Abdul-Manan, N., Liu, G. A., and Rosen, M. K. (2000) Autoinhibition and activation mechanisms of the Wiskott-Aldrich syndrome protein, *Nature 404*, 151.
14. Huber, A. H., and Weis, W. I. (2001) The structure of the B-caterin/E-cadherin complex and the molecular basis of diverse ligand recognition by B-catenin, *Cell 105*, 391.
15. Dunker, A. K., Brown, C. J., Lawson, J. D., Iakoucheva, L. M., and Obradovic, Z. (2002) Intrinsic disorder and protein function, *Biochemistry 41*, 6573.
16. Plaxco, K. W., and Gross, M. (1997) The importance of being unfolded, *Nature 386*, 657.
17. Shoemaker, B. A., Portman, J. J., and Wolynes, P. G. (2000) Speeding molecular recognition by using the folding funnel: The fly-casting mechanism, *Proc. Natl. Acad. Sci. U.S.A. 97*, 8868.
18. Dunker, A. K., and Obradovic, Z. (2001) The protein trinity: linking function and disorder, *Nat. Biotechnol. 19*, 805.
19. Uversky, V. N., Gillespie, J. R., and Fink, A. L. (2000) Why are "natively unfolded" proteins unstructured under physiological conditions? *Proteins: Struct., Funct., Genet. 41*, 415.
20. Bracken, C. (2001) NMR spin relaxation methods for characterization of disorder and folding in proteins, *J. Mol. Graphics Modell. 19*, 3.
21. Dyson, H. J., and Wright, P. E. (2001) Nuclear magnetic resonance methods for elucidation of structure and dynamics in disordered states, *Methods Enzymol. 339*, 258.
22. Häckel, M., Konno, T., and Hinz, H.-J. (2000) A new alternative method to quantify residual structure in unfolded proteins, *Biochim. Biophys. Acta 1479*, 155.

23. Georgescu, R. E., García-Mira, M. D. M., Tasayco, M. L., and Sánchez-Ruiz, J. M. (2001) Heat capacity analysis of oxidized *Escherichia coli* thioredoxin fragments (1−73, 74−108) and their noncovalent complex. Evidence for the burial of apolar surface in protein unfolded states, *Eur. J. Biochem. 268*, 1.
24. Makhatadze, G. I., and Privalov, P. L. (1990) Heat capacity of proteins. I. Partial molar heat capacity of individual amino acid residues in aqueous solution: Hydration effect, *J. Mol. Biol. 213*, 375.
25. Privalov, P. L., and Makhatadze, G. I. (1990) Heat capacity of proteins. II. Partial molar heat capacity of the unfolded polypeptide chain of proteins. Protein unfolding effects, *J. Mol. Biol. 213*, 385.
26. Tasayco, M. L., Fuchs, J. A., Yang, X.-M., Dyalram, D., and Georgescu, R. E. (2000) Interaction between two discontiguous chain segments from the β-sheet of *Escherichia coli* thioredoxin suggests initiation site for folding, *Biochemistry 39*, 10613.
27. Yang, X.-M., Georgescu, R., Li, J.-H., Yu, W., Haierhan, and Tasayco, M. L. (1999) Recognition between disordered polypeptide chains from cleavage of an α/β domain: Self- versus non-self-association, in *Proceedings of the 1999 Pacific Symposium on Biocomputing*, Mauna Lani, HI, Jan 4−9, 1999, Vol. 4, pp 590−600, World Scientific, Singapore.
28. Daughdrill, G. W., Vise, P. D., Zhou, H., Yang, X., Yu, W.-F., Tasayco, M. L., and Lowry, D. F. (2003) Monitoring temperature induced changes in the hydrodynamic volume of a partially folded *E. coli* thioredoxin fragment using NMR relaxation, *J. Biomol. NMR*, submitted for publication.
29. Kholodenko, V., and Freire, E. (1999) A simple method to measure the heat capacity of proteins, *Anal. Biochem. 270*, 336.
30. Makhatadze, G. I., Medvedkin, V. N., and Privalov, P. L. (1990) Partial molar volumes of polypeptides and their constituent groups in aqueous solutions over a broad temperature range, *Biopolymers 30*, 1001.
31. Freire, E. (1995) Differential scanning calorimetry, in *Protein Stability and Folding: Theory and Practice* (Shirley, B. A., Ed.) Methods in Molecular Biology, Chapter 9, pp 191−218, Humana Press, Totowa, NJ.
32. Häckel, M., Hinz, H.-J., and Hedwig, G. R. (1999) A new set of peptide-based group heat capacities for use in protein stability calculations, *J. Mol. Biol. 291*, 197.
33. Spolar, R. S., and Record, M. T. (1994) Coupling of local folding to site-specific binding of proteins to DNA, *Science 263*, 777.
34. Gómez, J., Hilser, V. J., Xie, D., and Freire, E. (1995) The heat capacity of proteins, *Proteins 22*, 404.
35. Fernández, A., and de las Mercedes Boland, M. (2002) Solvent environment conducive to protein aggregation, *FEBS Lett. 529*, 298.
36. Liu, Y., and Eisenberg, D. (2002) 3d domain swapping: As domains continue to swap, *Protein Sci. 11*, 1285.
37. Newcomer, M. E. (2002) Protein folding and three-dimensional domain swapping: A strained relationship? *Curr. Opin. Struct. Biol. 12*, 48.
38. Bennett, M. J., Schlunegger, M. P., and Eisenberg, D. (1995) 3d swapping: A mechanism for oligomer assembly, *Protein Sci. 4*, 2455.
39. Sinha, N., Tsai, C.-J., and Nussinov, R. (2001) A proposed structural model for amyloid fibril elongation: Domain swapping forms and interdigitating β-structure polymer, *Protein Eng. 14*, 93.
40. Dima, R. I., and Thirumalai, D. (2002) Exploring protein aggregation and self-propagation using lattice models: Phase diagram and kinetics, *Protein Sci. 11*, 1036.
41. Harrison, P. M., Chan, H.-S., Prusiner, S. B., and Cohen, F. E. (2001) Conformational propagation with prion-like characteristics in a simple model of protein folding, *Protein Sci. 10*, 819.
42. Baldwin, R. L. (1986) Temperature dependence of the hydrophobic interaction in protein folding, *Proc. Natl. Acad. Sci. U.S.A. 83*, 8069.
43. Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., and Teller, A. H. (1953) Equation of state calculations by fast computing machine, *J. Chem. Phys. 21*, 1087.
44. Hilser, V. J., Dowdy, D., Oas, T. G., and Freire, E. (1998) The structural distribution of cooperative interactions in proteins: Analysis of the native state ensemble, *Proc. Natl. Acad. Sci. U.S.A. 18*, 9903.
45. Stone, M. J., Chandrasekhar, K., Holmgren, A., Wright, P. E., and Dyson, H. J. (1993) Comparison of backbone and tryptophan side-chain dynamics of reduced and oxidized *Escherichia coli* thioredoxin using ^{15}N NMR relaxation measurements, *Biochemistry 32*, 426.
46. Jeng, M.-F., and Dyson, H. J. (1995) Comparison of the hydrogen-exchange behavior of reduced and oxidized *E. coli* thioredoxin, *Biochemistry 34*, 611.

BI0271906