The adaptability of *Escherichia coli* thioredoxin to non-conservative amino acid substitutions

R. O'BRIEN,¹ R. WYNN,² P. C. DRISCOLL,¹ B. DAVIS,¹ K. W. PLAXCO,³
J. M. STURTEVANT,⁴ AND J. E. LADBURY¹

¹Department of Biochemistry and Molecular Biology, University College London, Gower Street, London, WC1E 6BT, United Kingdom

²Dupont Merck Pharmaceutical Company, Route 141 and Henry Clay Road, Experimental Station, E336/241B, Wilmington, Delaware 19880

³Oxford Centre for Molecular Sciences, University of Oxford, New Chemistry Laboratory, South Parks Road, Oxford, OX1 3QT, United Kingdom

⁴Department of Chemistry and Department of Molecular Biophysics and Biochemistry, Yale University, P.O. Box 208107, 225 Prospect Street, New Haven, Connecticut 06520

(RECEIVED December 26, 1996; ACCEPTED March 6, 1997)

Abstract

The adaptability of *Escherichia coli* thioredoxin to the substitution of a series of non-natural amino acids has been investigated. Different thiosulfonated alkyl groups were inserted into the hydrophobic core of the protein in position 78 via disulfide bonding with a buried cysteine residue as previously described (Wynn R, Richards FM. 1993. Unnatural amino acid packing mutants of *Escherichia coli* thioredoxin produced by combined mutagenesis/chemical modification techniques. *Protein Sci* 2:395–403). The side chains added to the cysteine included methyl, ethyl, n-propyl, n-butyl, n-pentyl, and cyclo-pentyl derivatives. The side chains appear to exploit the presence of the large cavities to incorporate these variant forms, enabling the protein to fold and have some activity. Solution structural and kinetic data suggested that these substitutions had little effect on the overall fold of the protein. Thermodynamic data revealed that the entropic effect of restricting the side chains in the folded protein has an effect on the stability. The variant forms of thioredoxin have different propensities to form dimers despite the limited structural perturbations. Molecular modeling studies allow the conformation of the side chains to be assessed.

Keywords: calorimetry; cysteine modification; differential scanning; kinetics; protein folding; thermodynamics; thioredoxin; unnatural amino acids

To address the protein folding problem it is not sufficient to just determine which amino acids tolerate each other in the spatial juxtaposition of the folded protein; it is also necessary to establish where it is appropriate to allow cavities to exist. The close-packing of residues in the hydrophobic core of the protein is a requisite for protein stability (Richards, 1977; Matthews, 1991; Richards & Lim, 1994); however, cavities are frequently observed in proteins. Site-directed mutagenesis-based studies have revealed much about their role in protein structure by observing the thermodynamic effects of filling or creating cavities (Eriksson et al., 1992; Hubbard et al., 1994).

In this work we explore the remarkable tolerance to structural variation in a protein and show how this is afforded by the presence

of cavities. *Escherichia coli* thioredoxin is a small protein of approximately 11.7 kDa which performs a range of biological functions including disulfide-dithiol oxidoreductase activity based on two cysteine residues in positions 32 and 35 (Holmgren, 1985, 1989), and its participation in T7 DNA polymerase activity (Mark & Richardson, 1976). The X-ray crystallographic (Holmgren & Soderberg, 1970; Katti et al., 1990) and solution (Dyson et al., 1989, 1990) structures have been determined for *E. coli* thioredoxin. The structures of the oxidized and reduced forms of the protein are very similar with no significant differences observed in solvent-excluded regions of the protein. The hydrophobic core of thioredoxin has been shown to tolerate a wide range of non-conservative amino acid substitutions (Langsetmo et al., 1991; Hellinga et al., 1992; Wynn & Richards, 1993). Wynn and Richards (1993) reported the substitutions of large unnatural amino acid groups using a combination of site specific mutagenesis and chemical modification techniques. This method involved replacing the active site cysteine residues by serine and then substituting a cysteine residue into the hydrophobic core of thioredoxin. This cysteine could then be chemically modified by attaching alkyl thiosulfonated side chains by oxidation, linking them

Reprint requests to: John E. Ladbury, Department of Biochemistry and Molecular Biology, University College London, Gower Street, London, WC1E 6BT, United Kingdom; e-mail: j.ladbury@biochem.ucl.ac.uk.

Abbreviations: NMR, nuclear magnetic resonance; DSC, differential scanning calorimetry; DTT, dithiol threitol; MW, molecular weight; ΔH , change of molar enthalpy; ΔC_p , change in heat capacity.

through a disulfide bond. In this way, non-natural amino acids with linear side chains incorporating up to eight covalent bonds from the α -carbon were inserted into the protein. These variants were shown to fold into native-like structures and to exhibit DNA polymerase enzymatic activity.

In this work, the remarkable adaptability of thioredoxin structure is investigated for the type of variant described above with the L78C substitution and added thiosulfonated alkyl groups ranging from methyl to n-pentyl and cyclopentyl. Residue 78 in thioredoxin forms part of a β -sheet, which is buried in the hydrophobic core of the protein. The thermodynamic, kinetic, and structural determinants of stability have been studied. Our data suggest that the high tolerance of the protein to non-conservative substitutions at position 78 is due to an extensive network of cavities within the core of the protein that can accommodate a range of structures without significantly altering the overall three-dimensional fold.

Results and discussion

The thermodynamic, kinetic, and structural properties of chemically altered mutant forms of thioredoxin incorporating various chain-length thiosulfonated alkyl groups introduced into the hy-

drophobic core of the protein were investigated. Modifications were made at position 78, which is occupied by a leucine in the wild-type protein and is located on the central β -sheet, buried by two proximal α -helices. Both structural and kinetic studies confirm previous observations (Wynn & Richards, 1993) that the non-conservative substitutions do not significantly perturb the protein fold. However, thermodynamic data on the stability of the variants and the different propensities of the variants to self-associate suggest some minor adjustments to the structure do occur as a result of the modifications.

Analysis of DSC data suggests that the variants have different propensities to dimerize

The DSC data for the L78C_{ss}Me and L78C_{ss}Et forms of the protein are best fit by a simple two-state unfolding in which monomeric native protein unfolds to give monomeric denatured protein (Table 1). The ratio β /MW is unity for a strictly two-state unfolding process, whereas values greater than this are indicative of intermolecular cooperativity or slow unfolding (Sturtevant, 1987). For both the L78C_{ss}Me and L78C_{ss}Et forms the β /MW ratio is very close to unity, indicating that no significant concentrations of inter-

Table 1. Data from individual DSC scan on the variant forms of thioredoxin^a

Concentration		$t_{1/2}$ (°C)	% dimer ^b		ΔH_{cal} (kcal/mol)	β /MW	ΔC_p (kcal/(mol K))	SD (%)
μ M	mg/mL		N	D				
L78C*								
330	3.86	65.24	—	—	62.6	1.49	—	—
162	1.90	65.20	—	—	84.6	0.98	—	—
120	1.40	66.91	—	—	86.7	1.01	—	—
120	1.40	66.30	—	—	74.1	1.20	—	—
L78C_{ss}Me								
346	4.05	69.33	—	—	100.5	0.92	1.87	0.52
310	3.63	69.36	—	—	106.9	0.87	1.86	0.37
198	2.31	69.74	—	—	94.5	0.96	0.96	0.58
L78C_{ss}Et								
467	5.46	69.34	—	—	94.0	1.00	1.45	0.37
333	3.89	69.65	—	—	97.6	0.96	1.24	0.41
258	3.01	69.69	—	—	97.1	0.97	1.48	0.30
134	1.57	69.80	—	—	96.8	0.96	1.55	0.43
L78C_{ss}Pr								
226	2.65	67.30	—	—	67.0	1.35	3.06	1.28
83	0.97	69.25	—	—	76.4	1.15	1.86	1.05
37	0.43	70.34	—	—	102.0	0.91	0.41	1.03
L78C_{ss}Bu								
327	3.83	63.07	29	53	53.5	1.34	1.87	0.52
232	2.72	63.72	24	47	52.9	1.41	2.19	0.55
110	1.29	65.49	15	34	56.5	1.37	2.18	0.53
L78C_{ss}Pn								
97	1.14	62.63	42	79	71.8	0.77	2.80	2.04
73	0.85	63.43	37	76	51.0	1.31	2.45	1.44
L78C_{ssc}Pn								
73	0.85	69.59	—	—	84.6	1.11	0.85	0.76
31	0.36	69.56	—	—	81.1	1.21	1.54	1.15

^aData were fit as described in Results and discussion.

^bThe percentage dimer is given in the native, N, and denatured, D, states. L78C*, L78C_{ss}Me, L78C_{ss}Et, and L78C_{ssc}Pn were fit using a model for a simple monomer denaturation. L78C_{ss}Pr, L78C_{ss}Bu, and L78C_{ss}Pn were fit using a model incorporating incomplete dissociation/association to a dimer in both the native and denatured states.

mediate states are populated throughout the unfolding process. The DSC scans for the cyclo-pentyl derivative of thioredoxin were best fit by the same model used for the L78C_{ss}Me and L78C_{ss}Et forms.

By way of contrast, there is a noticeable dependence of $t_{1/2}$ on protein concentration for the L78C_{ss}Pr, L78C_{ss}Bu (Fig. 1), and L78C_{ss}Pn derivatives. This is indicative of the protein being in a higher oligomeric form in its denatured compared to that of the native state. To fit the DSC scans a model which incorporates incomplete dissociation/association to a dimer in both the native and denatured states of the protein described by

was used. This model had previously been invoked to describe the unfolding equilibria of wild-type thioredoxin (Ladbury et al., 1993). Although other more complex models were considered, this is the simplest that best describes the data obtained. Using this model for the L78C_{ss}Pr form made no significant difference to the overall precision of the fit. This would suggest that the dimerization constants for the n-propyl form are negligible, however, to explain the $t_{1/2}$ -concentration dependence K_D has to be greater than K_N . Small dimerization constants were also observed for the wild-type protein at pH above 7.25 (Ladbury et al., 1993). The butyl and pentyl derivatives are significantly better fit by the dimerization model. The fits indicate that a significant amount of dimer forms in the native and denatured states. As expected from the dependence of $t_{1/2}$ on concentration, the denatured form has a greater propensity to dimerize. The potential of the variants of thioredoxin to dimerize increases with increasing chain length of the substituted alkyl group.

This tendency of the protein to dimerize was observed in the case of the wild-type protein (Ladbury et al., 1993, 1994) and in the L78K and L78R mutants (Ladbury et al., 1995) previously studied. Analytical ultracentrifugation studies suggest dimerization of the wild-type protein (Laurent et al., 1964; J.E. Ladbury & J.R. Livingstone, unpubl. data). A recent NMR study also invokes a

monomer-dimer equilibrium in wild-type and L78K mutants to explain observed spectral density parameters (de Lorimier et al., 1996). Some additional evidence for the potential to form oligomeric states in the native state of L78C_{ss}Bu and L78C_{ss}Pn can be derived from the circular dichroism, CD, spectroscopic study of Wynn and Richards (1993). A dramatic reduction in CD signal in the region below 210 nm is observed for these two forms. The CD of proteins in the far-UV has been shown not only to be dependent on secondary structural conformation but also on the relative positions of such motifs to one another. Since there is no evidence of a net change in secondary structural content in the region above 210 nm (including the region associated with signals corresponding to α -helix and β -sheet), the lower CD signal may be due to a minor local disruption of structure of residues surrounding position 78 (Wynn & Richards, 1993) or association of the native form (Manning & Woody, 1987; O'Brien, 1994).

These data indicate that relatively minor changes in primary structure can result in alterations to the oligomeric state of both the native and denatured states. Since dimerization is observed in the wild-type thioredoxin under these conditions (Ladbury et al., 1993), the lack of dimerization in the methyl and ethyl variants suggests that these alkyl side chains interfere with this process. Since there is no change in the charge of the variant forms of the protein over that of the wild-type, the loss of dimerization observed in the methyl and ethyl forms is presumably explained by the disruption of the dimerization interface. Thus, reducing the content of the cavity in which the substituted residues are accommodated appears to change the way the molecules interact. This has to be some long range effect and possibly results from subtle movement of residues around the cavity proximal to the protein surface. However, NMR structural investigation of the variant proteins compared to the wild type does not reveal any direct evidence for this (see below).

The thermodynamics of unfolding of the variant forms

The change in the thermodynamic parameters of unfolding for the variants with respect to L78C* are shown in Table 2. The data for the wild type and the wild type with the active site cysteine residues (C32 and C35) substituted for serine have been included for comparison (data derived from Ladbury et al., 1993, 1994). The reason for the large destabilization of the L78C* mutant compared to the wild-type thioredoxin is not obvious, although the most likely reason is that it is the effect of insertion of a polar group into a position normally occupied by a hydrophobic side chain. Similar sized reductions in stability were observed for L78R and L78K mutations (Ladbury et al., 1995). At 120 μ M protein concentration the L78C_{ss}Me, L78C_{ss}Et, and L78C_{ss}Pr variants are all more stable than the L78C* mutant. There is a dramatic drop in $t_{1/2}$ and ΔG° on extending the mutant side chain to the butyl form. The L78C_{ss}Pn and L78C_{ss}Cp data have been omitted from this table due to the limited data set; however, it should be noted that the $t_{1/2}$ values for the unfolding of the L78C_{ss}Pn form are the lowest of all the mutants even at the low concentrations at which measurements were made. Interestingly, the L78C_{ss}Cp form has a $t_{1/2}$ similar to that of the L78C_{ss}Me and L78C_{ss}Et variants.

There are several potential sources of the destabilizing effects derived from increasing the side chain size at position 78. One of these that is likely to have a significant effect is that the burial of a side chain can lead to reduction in the rotational and vibrational degrees of freedom, which become restricted by steric hindrance. On unfolding, the restricted conformation is removed; thus the

Fig. 1. The dependence of $t_{1/2}$ on concentration shown in DSC data of the L78C_{ss}Bu form of thioredoxin. The figure shows DSC data concentration normalized. From left to right, peaks correspond to 327, 232, and 110 mM, respectively.

Table 2. Thermodynamic data for the unfolding of thioredoxin and variants of thioredoxin^a

Protein	$\Delta T_{1/2}$ (°C)	$\Delta\Delta G^\circ$ (kcal/mol)	$\Delta\Delta H_{cal}$ (kcal/mol)	$T\Delta\Delta S^\circ$ (kcal/mol K)	$\Delta\Delta G_f^\ddagger$ (kcal/mol)	$\Delta\Delta G_{int}^\ddagger$ (kcal/mol)
Wild-type	19.4	4.1	-22.9	-27.0	-0.20 ± 0.06	0.30 ± 0.12
C32S/C33S	9.4	2.3	12.8	10.5		
L78C*	0	0	0	0	0	0
Methyl	2.9	0.8	14.8	14.0		
Ethyl	3.0	0.8	10.2	9.4		
Propyl	1.9	0.4	-2.7	-3.1	0.11 ± 0.15	-0.06 ± 0.09
Butyl	-1.3	-0.2	-22.0	-21.8	-0.05 ± 0.05	0.07 ± 0.14

^aComparison of the thermodynamics of unfolding for each species must be performed at the same temperature. The dependence of enthalpy and $t_{1/2}$ on concentration for the butyl form requires that the data be normalized to an arbitrary concentration. We have selected 120 μ M as a standard concentration and the apparent $t_{1/2}$ of the L78C* mutant as the temperature for the comparison. This value was obtained from the linear relationship between $1/T_{1/2}$ ($T = t_{1/2} + 273.15$) and the natural logarithm of concentration as previously described (Ladbury, 1993; Tanaka, 1993) and ΔH_m (ΔH at $T_{1/2}$) from the plot of ΔH_{cal} versus concentration. The free energy of unfolding was calculated as described elsewhere (Ladbury et al., 1993). The ΔCp employed was 1.88 kcal mol⁻¹ K⁻¹ previously determined for the wild type. The thermodynamic and kinetic data are expressed in terms of $\Delta\Delta J^\circ = \Delta J_{mutant}^\circ - \Delta J_{L78C}^\circ$, where J is the thermodynamic parameter: free energy (ΔG°), enthalpy (ΔH), and entropy (ΔS°). $\Delta\Delta G_f^\ddagger$ and $\Delta\Delta G_{int}^\ddagger$ are the activation energy for the first step in the sequential folding pathway and the activation energy of folding of the sequential intermediate to the native protein, respectively.

unfolded form is entropically favored. As the size of the side chain increases so do the number of degrees of freedom that are affected in this way; thus the mutant proteins become progressively destabilized with side chain size. This explanation for the reduction in stability is particularly appealing, since the inclusion of the cyclopentyl derivative—which, although having the same number of side-chain atoms, has a reduced number of rotamers—leads to an increase in $t_{1/2}$ compared to the n-pentyl group (Table 1). It should be noted that opposing this will be the thermodynamic consequences derived from the hydrophobic effect. The inclusion of larger hydrophobic groups should increase the stability since there is a greater favorable entropy arising from the liberation of ordered water into the bulk phase on folding of the thioredoxin. Comparison of the $t_{1/2}$ obtained from the inclusion of the n-pentyl and the cyclopentyl side chains, which have similar hydrophobic surface area, suggests that the hydrophobic effect cannot be the major determinant since the latter has a higher $t_{1/2}$.

Another possible destabilizing effect could result from the varying propensity to associate into putative dimers in both the native and denatured states as described above. The greater propensity to form these associated forms is observed in the denatured state. This would have the effect of destabilizing the protein since the denatured dimer conformation would be favored. This explanation is enhanced by the observation of the reduction of the positive change in ΔH between the values for n-butyl, n-pentyl, and—to a lesser extent—the n-propyl forms, which could reflect the fact that the protein is not fully unfolded or some intermolecular interactions persist in the unfolded form with respect to the ethyl and methyl forms. However, CD data obtained at temperatures in the post-transition region of the DSC scans show no evidence of residual secondary structure (data not shown).

In addition, since the residual structure that occurs in the denatured states of some proteins appears to manifest itself in the form of hydrophobic interactions (Dill & Shortle, 1991; Evans et al., 1991), increasing the hydrophobic surface area by increasing the mutant side-chain length in position 78 could have a greater capacity to induce this residual structure. This would also have the effect of stabilizing the denatured form of the protein and thus

reducing the relative stability of the folded form. This effect might be expected to be limited in the case of the substitutions described in this work since the changes in hydrophobic surface areas from one form to another are small.

The ΔCp for the unfolding of the mutant forms of thioredoxin was determined from the individual DSC scans. The average ΔCp for the 21 scans measured in this work is 1.91 ± 0.80 kcal mol⁻¹ K⁻¹, which is very close to that determined for the oxidized wild-type thioredoxin obtained by Ladbury et al. (1993) of 1.88 ± 0.62 kcal mol⁻¹ K⁻¹ and 1.66 ± 0.05 kcal mol⁻¹ K⁻¹ by Santoro and Bolen (1992). Although there are large uncertainties associated with estimation of ΔCp from individual scans it appears that there is no significant difference between the ΔCp of the wild type and the mutants under study here. The substitution of charged residues (Lys and Arg) at the 78 position significantly reduced the ΔCp for the unfolding of thioredoxin (Ladbury et al., 1995). This effect was hypothesized to be the result of limited disruption of the native structure such that water could penetrate into the hydrophobic core and thus more hydrophobic surface area was exposed to solvent prior to unfolding. The hydrophobic substitutions in the 78 position described in this work appear not to cause this effect.

Wild-type and variants have similar folding kinetics

Changes in the thermodynamic properties of proteins are often accompanied by changes in their folding kinetics (Fersht, 1995; Mines et al., 1996). In order to investigate the effects of the modifications to thioredoxin on the relative free energy of their folding transition states we have measured the folding kinetics of the wild type and three of the variants (L78C*, L78C_{ss}Pr, and L78C_{ss}Bu). The folding of thioredoxin is multi-phasic, presumably representing both parallel and sequential folding pathways (Kelley et al., 1984, 1986, 1987). The majority of the molecules appear to fold to their native conformation via a sequential pathway featuring a single significantly populated intermediate. The remainder of the molecules refold via a slow, parallel pathway limited by the isomerization of a single proline peptide bond at position 76 (Kelley & Richards, 1987). Consistent with this model the activation energy

for the fast folding to a transition state for all the variants studied is increased by approximately $0.2 \text{ kcal mol}^{-1}$, presumably reflecting a relatively non-specific collapse that would not be significantly affected by the alkyl side chain. The second reaction rate in the sequential pathway, however, is thought to reflect the reorganization of a collapsed form to give the native protein. The transition state for this reaction would presumably be more tightly packed than the first transition state and therefore more significantly perturbed by the presence of the different alkyl-cysteine side chains. This appears to be the case for all the variants, which exhibit an approximately $0.3 \text{ kcal mol}^{-1}$ increase in this activation energy to go from this intermediate state to the final folded form and refold at around 60% of the rate of the wild-type protein. However, it is important to note that these increases in transition state energy are very small compared to the overall change in free energy for the folding process (see below and Table 2), and the effectively identical folding rates of the L78C*, L78C_{ss}Pr, and the L78C_{ss}Bu variants suggest that these substitutions are even more easily accommodated in folding transition states than in the native protein. The very slow refolding phase, which is thought to reflect the isomerization of the polypeptide backbone at residue Pro76, is not changed in the variants, as would be expected of a proline isomerization-limited folding phase.

Wild-type and variants exhibit structural integrity

Figure 2 shows a comparison of the aliphatic regions of the $^2\text{H}_2\text{O}$ 2D ^1H TOCSY NMR spectra (Bax & Davis, 1985) of three representative forms of thioredoxin: wild type, L78C*, and L78C_{ss}Bu. The spectra are essentially superimposable and show that the structural integrity of the three proteins is maintained as evidenced by the similar patterns of dispersion of resonances and cross-peaks (including the slowly exchanging amide NH and aromatic protons (not shown)). Thus, the secondary structure and side chain packing in the mutants is not significantly different from that of oxidized or

reduced wild-type thioredoxin (Dyson, 1990). This limited structural change resulting from quite dramatic substitutions in thioredoxin is also seen for the L78K mutant (de Lorimier et al., 1996). X-ray crystallographic studies on cavity creating mutants of T4 lysozyme (Eriksson et al., 1992) and the insertion of alkyl side chains into staphylococcal nuclease (in a similar way as described here; Wynn et al., 1996) indicated that there is a tendency for the surrounding residues to move toward the space left in a cavity. It is not clear whether the subtle localized structural changes observed for T4 lysozyme would be observed using NMR spectroscopy. The only set of cross peaks observed to be very sensitive to the mutations are those corresponding to the $\text{CaH}-\text{C}\beta\text{H}_2$ moiety of residue 78 (Leu in the wild type and Cys in the mutants, which are boxed in Fig. 2). The changes in chemical shifts of these signals is more likely to result from the change in the chemical nature of the side chain rather than from any gross conformational differences. The CaH proton chemical shift is downfield of the water protons in each case which is indicative of retention of the residue in a β -sheet secondary structural element.

Conformation of side chains in cavity explored by molecular modeling

To enable us to speculate on how the range of the unnatural amino acid side chains inserted can be accommodated in the structure, molecular modeling studies of the mutated forms of thioredoxin were performed as described above. The thioredoxin crystal structure solved at 1.6 Å resolution (Katti et al., 1990) indicates that position 78 is in close proximity to several spacious cavities in the core of the protein. By modeling the structure of wild-type thioredoxin with different chemically modified side chains and using simulated annealing, a range of energy minimized conformations of the side chain in the cavities could be obtained (see Materials and methods). Each panel of Figure 3 shows the five lowest energy conformations for each mutant protein (with the single lowest

Fig. 2. Two-dimensional TOCSY NMR spectra for (A) wild type, (B) L78C*, and (C) L78C_{ss}Bu, showing the high level of structural integrity in each form. The major change in the cross-peak positions among the various forms (boxed) corresponds to the $\text{CaH}-\text{C}\beta\text{H}_2$ of residue 78.

Fig. 3. Side chain of residue 78 for (A) wild-type, (B) L78C_{ss}Me, (C) L78C_{ss}Et, (D) L78C_{ss}Pr, (E) L78C_{ss}Bu, (F) L78C_{ss}Pn, and (G) L78C_{ss}Cp. The large cavity from the L78C* mutant is shown in grid lines, and the cavity present in the modeled structure is shown shaded gray. In the figures of the variant forms, the five lowest energy conformations are shown in lines with the single lowest energy structure shown in ball-and-stick representation. Individual cavities are arbitrarily labeled A, B, C, and D.

energy conformation given in the ball and stick representation). The cavity for the wild-type thioredoxin is shown in grid form and the cavities remaining after insertion of the modified side chains in the lowest energy conformation are depicted in grey shading. The figures show the effect of the progressive increase in the size of the modified side chain. From these data it is clear that thioredoxin has sufficient space in its hydrophobic core to accommodate large groups although, based on energy minimization data, they will

adopt different conformations. Four cavities are present in the wild-type structure in the vicinity of residue 78 (A–D; see Fig. 3) with a calculated total volume of $133.6 \text{ \AA}^3$. In the L78C mutant these cavities combine to form a single large cavity of $192.5 \text{ \AA}^3$ (not shown). As the size of the side chain is increased the cavities become occupied. After adding the n-pentyl group there is still a significant proportion of unoccupied cavity volume suggesting that thioredoxin could potentially accommodate even larger side chains.

Formation of cavities in T4 lysozyme by Leu to Ala mutations led to a destabilizing effect based on the void volume created, which amounted to 24–33 cal mol⁻¹ Å⁻³ (Eriksson et al., 1992). Our data do not correspond to a correlation of this type. This may be the result of the way in which the mutant side chains explore different space and do not probe into one cavity in an ordered way.

Cavity D, which is the smallest of the four in the wild-type structure, appears to be preferred by the aliphatic side chains incorporating more than two carbons except the cyclo-pentyl group. As described above, the DSC data are best fit with models based on dimerization from the n-butyl and n-pentyl derivative with some evidence of this for the propyl form. This coincides with side chain occupancy of cavity D as shown in Figure 3 by the reduction in the size of the available space in that cavity. It could be speculated therefore that this cavity has some effect on the dimerization interface.

Summary

The hydrophobic core around position 78 in thioredoxin is able to accommodate the insertion of a diverse range of side chains. In the absence of high resolution three-dimensional structural data for the variant forms of the protein, we suggest that this can be achieved by the occupation of space provided by a network of cavities observed in the hydrophobic core of the protein. Increasing the size of the linear thiosulfonated side chains up to a pentyl group leads to increasing destabilization of the protein. This can be due to a number of effects although the limited effects on structure and kinetics of folding of the variants indicates that the reduction in stability is not the result of gross structural perturbation.

We have observed a potential for the protein to dimerize with inserted modified side chains larger than propyl (although the propyl form may exhibit limited effects). This effect could be due to side chains probing the D cavity of which a certain volume is required to produce the dimerization interface. This work is in contrast to the recent study on insertion of unnatural amino acids into staphylococcal nuclease, SNase (Wynn et al., 1996), in which the crystal structures of the protein with up to the pentyl side chain are reported. In the case of SNase, the cavity becomes filled after the addition of the ethyl group. This closer packing in SNase indicates that the protein structure has to exhibit a high degree of plasticity to allow larger groups to be inserted. The structure of thioredoxin would appear not to need such plasticity of structure, since even on adding a pentyl group there is still a large cavity volume remaining (107 Å⁻³).

Materials and methods

Protein preparation

The mutants of *E. coli* thioredoxin were prepared as described earlier (Wynn & Richards, 1993). The only cysteine residues in the wild-type thioredoxin are the two found in the disulfide-dithiol oxidoreductase active site at positions 32 and 35. To preclude any possible side reactions with thiosulfonated alkyl groups these residues were substituted for serines. In addition leucine 78 was mutated to a cysteine residue. Thus, a substituted thioredoxin with the three changes C32S, C35S, and L78C (hereafter known as L78C*) was produced. Six modified forms of L78C* thioredoxin were produced by linkage through a disulfide bond onto Cys78. These were with the thiosulfonated alkyl group as a methyl (L78C_{SS}Me), an ethyl (L78C_{SS}Et), an n-propyl (L78C_{SS}Pr), an n-butyl (L78C_{SS}Bu), an

n-pentyl (L78C_{SS}Pn), or a cyclo-pentyl (L78C_{SS}Cp). Samples were stored as lyophilized powders prior to use.

Differential scanning calorimetry (DSC)

Protein samples were dissolved in 50 mM sodium phosphate at pH 7.0 and dialyzed exhaustively against this buffer at 4 °C. Protein samples were in the concentration range of 0.36 to 5.46 mg mL⁻¹ (see Table 1) as determined by optical density based on a molar extinction coefficient at 280 nm of 13.7 mM⁻¹ cm⁻¹ (Reutimann et al., 1981). All DSC scans were performed using the MCS or the MC2 calorimeter (MicroCal, Inc., Northampton, Massachusetts) at a scan rate of 1 K min⁻¹. The reference cell was filled with the buffer solution used to dialyze the protein samples and the instrumental baselines were determined with buffer in both cells. DTT (2 mM) was added to the buffer for the L78C* mutant. The various forms of thioredoxin were found to be 94–100% reversible if the sample was not heated to more than 5 °C above *t*_{1/2}.

DSC data analysis

The calorimetric data, with baselines subtracted, were analyzed by non-linear least-squares curve fitting procedures described elsewhere (Sturtevant, 1987; Ladbury et al., 1993). As in previous studies using DSC we have adopted the simplest model that best fits the data. More complex models in some cases can improve the fit; however, the inherent increase in the number of adjustable parameters leads to a number of possible interpretations. For all the substitutions in thioredoxin we assumed a curve-fitting model based on two-state unfolding. In the fitting program there are three adjustable parameters; *t*_{1/2} (the temperature in °C at which the unfolding is half completed), Δh_{cal} (the calorimetric enthalpy expressed in calories per gram), and the ratio β ($= \Delta H_{vH} / \Delta h_{cal}$ where ΔH_{vH} is the van't Hoff enthalpy).

Kinetics of refolding

The refolding kinetics of thioredoxin were monitored by changes in the fluorescence of its two core tryptophan residues. Rapid kinetic fluorescence data were obtained using a SX17-MV stopped flow fluorimeter (Applied Photophysics, Leatherhead, United Kingdom) thermostated at 20.0 ± 0.1 °C. The excitation wavelength adopted was 280 ± 3 nm and emission was detected above 310 nm by use of the appropriate filters. Data on the slowest kinetic phase were obtained with a Perkin Elmer LS50B fluorimeter using manual mixing techniques. Conditions were as above except that detection was at 350 ± 10 nm. The thioredoxin variants were dissolved in 6.0 M GuHCl and established as being completely unfolded (Wynn & Richards, 1993). They were then diluted 11-fold with 50 mM sodium phosphate at pH 7.0. Reported values and confidence levels represent the average and standard deviation respectively of three traces each comprised of the sum of five experiments. The data were fitted with the software provided with the instrument or with Kaleidograph (Albelbeck software).

Nuclear magnetic resonance (NMR) spectroscopy

Protein samples were dissolved in water to which 5% (v/v) ²H₂O was added. NMR spectra were recorded on a Varian UNITYplus spectrometer operating at a ¹H frequency of 500 MHz. The TOCSY spectra were recorded using 2*oversampling in F2 (sweepwidth = 12,000 Hz) and with F1 (sweepwidth = 6,000 Hz). The spectra (2,048 *t*₂ by 256 *t*₁ complex points) were recorded at 20 °C. Data

were processed using FELIX.2.3 software. Assignments of the proton NMR spectrum were made by comparison with the reported assignments for oxidized wild-type thioredoxin (Dyson et al., 1989).

Molecular modeling

Initial structures of the L78C* mutant and the chemically modified variants L78C_{ss}Me, L78C_{ss}Et, L78C_{ss}Pr, L78C_{ss}Bu, L78C_{ss}Pn, and L78C_{ss}CpN were generated using the Discover module of InsightII (Biosym). Variant structures were based on the wild-type crystal structure of *E. coli* thioredoxin by Katti and coworkers (1990). The Leu side chain in the wild-type structure was replaced by a Cys in the same orientation and aliphatic groups added onto this in their default (extended) conformation. An ensemble of ten model structures were generated from each of these initial structures using a simulated annealing/molecular dynamics protocol in X-PLOR 3.1 (Brünger, 1992), based on the standard protocol used to calculate solution structures from NMR data. The positions of all atoms other than those in residue 78 were fixed, and calculated energetic interactions were restricted to those residues within 5 Å of residue 78. Since no charged groups were introduced into the structure, electrostatic interactions were not considered. A subset of five of the model 10 structures was analyzed in detail. Each subset of structures was selected by consideration of overall and van der Waals energy terms. The lowest energy structure of each ensemble was used as a representative structure for further calculations, and the locations and volumes of cavities close to residue 78 were determined. Cavities within each model structure were identified using the program SURFNET (Laskowski, 1995). This program measures the volumes of cavities in proteins by attempting to fill the cavity with a series of spheres of gradually decreasing radius down to the minimum probe size specified. The probe size for the measurement of cavity volumes was allowed to range from 4.0 Å to 1.0 Å. Clearly the minimum probe size will dictate the size of the cavity. In some cases the use of 1.4 Å (equivalent to the radius of a water molecule) has been adopted; however, since the cavity will not be filled with water we see no reason to adopt this value. Cavities close to residue 78 were selected using the program MASK within SURFNET. The geometry of these cavities was then analyzed with InsightII.

Acknowledgments

J.E.L. is a Wellcome Trust Research Career Development Fellow; P.C.D. is a Royal Society University Research Fellow.

References

- Bax A, Davis DG. 1985. Simplification of ¹H NMR spectra by selective excitation of experimental subspectra. *J Mag Reson* 65:355–360.
- Brünger AT. 1992. *X-PLOR Manual, Version 3.1*. New Haven, Connecticut: Yale University.
- de Lorimier R, Hellinga HW, Spicer LD. 1996. NMR studies of structure, hydrogen exchange, and main-chain dynamics in a disrupted-core mutant of thioredoxin. *Protein Sci* 5:2552–2565.
- Dill KA, Shortle D. 1991. Denatured states of proteins. *Annu Rev Biochem* 60:795–825.
- Dyson HJ, Gippert GP, Case DA, Holmgren A, Wright PE. 1990. Three-dimensional solution structure of the reduced form of *Escherichia coli* thioredoxin determined by nuclear magnetic resonance spectroscopy. *Biochemistry* 29:4129–4136.
- Dyson HJ, Holmgren A, Wright PE. 1989. Assignment of the proton NMR spectrum of reduced and oxidized thioredoxin—Sequence specific assignments, secondary structure, and global fold. *Biochemistry* 28:7074–7087.
- Eriksson AE, Baase WA, Zhang XJ, Heinz DW, Blaber M, Baldwin EP, Matthews BW. 1992. Response of a protein structure to cavity-creating mutations. *Science* 255:178–183.
- Evans PA, Topping KD, Woolfson DN, Dobson CM. 1991. Hydrophobic clustering in nonnative states of a protein: Interpretation of chemical shifts in NMR spectra of denatured states of lysozyme. *Proteins Struct Funct Genet* 9:248–266.
- Fersht AR. 1995. Mapping the structures of transition-states and intermediates in folding: Delineation pathways at high resolution. *Phil Trans R Soc* 348:11–15.
- Hellinga HW, Wynn R, Richards FM. 1992. The hydrophobic core of *E. coli* thioredoxin shows a high tolerance to non-conservative single amino-acid substitutions. *Biochemistry* 31:11203–11209.
- Holmgren A. 1985. Thioredoxin. *Ann Rev Biochem* 54:237–271.
- Holmgren A. 1989. Thioredoxin and glutaredoxin systems. *J Biol Chem* 264:13963–13966.
- Holmgren A, Soderberg B. 1970. Crystallization and preliminary crystallographic data for thioredoxin from *Escherichia coli*. *J Mol Biol* 54:387–390.
- Hubbard SJ, Gross KH, Argos P. 1994. Intermolecular cavities in globular proteins. *Protein Eng* 7:613–626.
- Katti SK, LeMaster DM, Eklund H. 1990. Crystal structure of thioredoxin from *Escherichia coli* at 1.68 Å resolution. *J Mol Biol* 212:167–181.
- Kelley RF, Richards FM. 1987. Replacement of proline-76 with alanine eliminates the slowest kinetic phase in thioredoxin folding. *Biochemistry* 26:6765–6774.
- Kelley RF, Shalongo W, Jagannadham MV, Stellwagen E. 1987. Equilibrium and kinetic measurements of conformational transition of reduced thioredoxin. *Biochemistry* 26:1406–1411.
- Kelley RF, Stellwagen E. 1984. Conformational transitions of thioredoxin in guanidine hydrochloride. *Biochemistry* 23:5095–5202.
- Kelley RF, Wilson JB, Bryant C, Stellwagen E. 1986. Effects of guanidine-hydrochloride on the refolding kinetics of denatured thioredoxin. *Biochemistry* 25:728–732.
- Ladbury JE, Kishore N, Hellinga HW, Wynn R, Sturtevant JM. 1994. Thermodynamic effects of reduction of the active-site disulfide of *Escherichia coli* thioredoxin explored by differential scanning calorimetry. *Biochemistry* 33:3688–3692.
- Ladbury JE, Wynn R, Hellinga HW, Sturtevant JM. 1993. Stability of oxidized *Escherichia coli* thioredoxin and its dependence on protonation of the aspartic acid residue in the 26 position. *Biochemistry* 32:7526–7530.
- Ladbury JE, Wynn R, Thomson JA, Sturtevant JM. 1995. Substitution of charged residues into the hydrophobic core of *Escherichia coli* thioredoxin results in a change in heat capacity of the native protein. *Biochemistry* 34:2148–2152.
- Langsetmo K, Fuchs JA, Woodward C. 1991. The conserved buried aspartic acid residue in oxidized *Escherichia coli* thioredoxin has a pK_a of 7.5. Its titration produces a related shift in global stability. *Biochemistry* 30:7603–7609.
- Laskowski J. 1995. SURFNET: A program for visualizing molecular surfaces, cavities, and intermolecular interactions. *J Mol Graph* 3:323–330.
- Laurent TC, Moore EC, Reichard P. 1964. Enzymatic syntheses of deoxyribonucleotides. IV. Isolation and characterization of thioredoxin, the hydrogen donor from *E. coli* B. *J Biol Chem* 239:3436–3444.
- Manning MC, Woody RW. 1987. Theoretical determination of the CD of proteins containing closely packed antiparallel beta sheets. *Biopolymers* 26:1731–1752.
- Mark DF, Richardson CC. 1976. *Escherichia coli* thioredoxin: A subunit of bacteriophage T7 DNA polymerase. *Proc Natl Acad Sci USA* 73:780–784.
- Matthews BW. 1991. Mutational analysis of protein stability. *Curr Opin Struct Biol* 1:17–21.
- Mines GA, Torbjörn P, Lee SC, Winkler JA, Gray HB. 1996. Cytochrome C folding triggered by electron transfer. *Chem & Biol* 3:491–497.
- O'Brien RRJ. 1994. The biochemical and biophysical properties of xylanases I and II from *Trichoderma viride*. [PhD thesis]. London: University of Greenwich.
- Reutimann H, Straub B, Luisi P-L, Holmgren A. 1981. A conformational study of thioredoxin and its tryptic fragments. *J Biol Chem* 256:6796–6803.
- Richards FM. 1977. Areas, volumes packing, and protein structures. *Annu Rev Biophys Bioeng* 6:151–176.
- Richards FM, Lim WA. 1994. An analysis of packing in the protein folding problem. *Annu Rev Biophys Bioeng* 26:423–498.
- Santoro MM, Bolen DW. 1992. A test of the linear extrapolation of unfolding free energy change over an extended denaturant concentration range. *Biochemistry* 31:4901–4907.
- Sturtevant JM. 1987. Biochemical applications of differential scanning calorimetry. *Annu Rev Phys Chem* 38:463–488.
- Tanaka A, Flanagan J, Sturtevant JM. 1993. Thermal unfolding of staphylococcal nuclease and several mutant forms thereof studies by differential scanning calorimetry. *Protein Sci* 2:567–576.
- Wynn R, Harkins PC, Richards FM, Fox RO. 1996. Mobile unnatural amino acid side chains in the core of staphylococcal nuclease. *Protein Sci* 5:1026–1031.
- Wynn R, Richards FM. 1993. Unnatural amino acid packing mutants of *Escherichia coli* thioredoxin produced by combined mutagenesis/chemical modification techniques. *Protein Sci* 2:395–403.