Structural consequences of replacement of an α-helical Pro residue in Escherichia coli thioredoxin

Rudresh1, Rinku Jain1, Vardhan Dani3, Ashima Mitra3, Sarika Srivastava3, Siddhartha P.Sarma3, R.Varadarajan3,4 and S.Ramakumar1,2

1Department of Physics, 2Bioinformatics Center and 3Molecular Biophysics Unit, Indian Institute of Science, Bangalore-560012, India
4To whom correspondence should be addressed.
E-mail: varadar@mbu.iisc.ernet.in

While it is well known that introduction of Pro residues into the interior of protein α-helices is destabilizing, there have been few studies that have examined the structural and thermodynamic effects of the replacement of a Pro residue in the interior of a protein α-helix. We have previously reported an increase in stability in the P40S mutant of Escherichia coli thioredoxin of 1–1.5 kcal/mol in the temperature range 280–330 K. This paper describes the structure of the P40S mutant at a resolution of 1.8 Å. In wild-type thioredoxin, P40 is located in the interior of helix two, a long α-helix that extends from residues 32 to 49 with a kink at residue 40. Structural differences between the wild-type and P40S are largely localized to the above helix. In the P40S mutant, there is an expected additional hydrogen bond formed between the amide of S40 and the carbonyl of residue K36 and also additional hydrogen bonds between the side chain of S40 and the carbonyl of K36. The helix remains kinked. In the wild-type, main chain hydrogen bonds exist between the amide of 44 and carbonyl of 40 and between the amide of 43 and carbonyl of 39. However, these are absent in P40S. Instead, these main chain atoms are hydrogen bonded to water molecules. The increased stability of P40S is likely to be due to the net increase in the number of hydrogen bonds in helix two of E.coli thioredoxin.

Keywords: α-helix/mutant/proline/thermal stability/thioredoxin

Introduction
Proline residues in proteins differ from other amino acid residues in two important respects: Pro residues lack an amide hydrogen atom and also the main chain dihedral angle φ is fixed at approximately −65 ± 11° (Laskowski et al., 1993) because of the presence of the rigid pyrrolidine ring. The lack of an amide hydrogen means that Pro residues cannot participate in main chain hydrogen bonding interactions that require the presence of an amide group. It is for this reason that Pro is seldom found in the hydrogen-bonded, central regions of α-helices or β-strands (MacArthur and Thornton, 1991). In an α-helix the amide proton of residue i forms a hydrogen bond with the carbonyl group of residue i – 4. Pro residues lack an amide proton and hence cannot form such a hydrogen bond. However, it has recently been suggested (Chakrabarti and Chakrabarti, 1998) that Pro residues in helices can form C–H···O hydrogen bonds. Most Pro residues in α-helices are found in the first turn of the helix. Pro residues have been introduced into helical regions of T4 lysozyme (Sauer et al., 1992; Gray et al., 1996) and have been shown to destabilize the protein by about 3–8 kcal/mol. While it is clear that introduction of Pro residues into the interior of protein α-helices is destabilizing (O’Neil and DeGrado, 1990; Strehlow et al., 1991; Sauer et al., 1992; Blaber et al., 1993), there have been few studies that have examined whether substitution of a Pro residue in the interior of an α-helix can result in stabilization of a protein. Substitution of such Pro residues should result in the formation of an additional main chain hydrogen bond between the amide group of the substituted residue and the carbonyl oxygen of the residue four residues N-terminal to the substituted residue.

There have been three previous studies of the effects of replacement of such α-helical Pro residues on protein stability. Replacement of the Pro61 residue in the middle of a long helix in the Escherichia coli Fis protein with Ala and Ser resulted in significant increases in the Tm of the protein (Yuan et al., 1994). However, the crystal structure of the P61A mutant revealed that the helix was still kinked in the mutant protein, showing that factors other than the presence of Pro were responsible for kinking of the helix. In contrast, in T4 lysozyme (Alber et al., 1988), replacement of a helical residue, Pro86, with a variety of other residues did not affect the Tm of the protein, although the kink was retained. A recent study reported (Chakrabarti et al., 1999) an increase in stability in the P40S,M37L mutant of E.coli thioredoxin (hereafter referred to as P40S) of 1–1.5 kcal/mol in the temperature range 280–330 K relative to both wild-type thioredoxin and the M37L single mutant. In this work, we solved the crystal structure of the above thioredoxin mutant, to obtain insight into the structural factors responsible for this stability increase. We also provide solution NMR evidence supporting the use of the wild-type protein for structural comparison with the P40S mutant.

Materials and methods

Crystallization and data collection
The construction, expression and purification of the mutant protein have been described previously (Chakrabarti et al., 1999). The mutant protein, P40S, contains two mutations at positions 37 and 40 where residues Met and Pro have been replaced by Leu and Pro, respectively. The M37L mutation was constructed as a first step to introducing single Met residues at various positions in the protein with the ultimate goal of carrying out fragment complementation studies (Ghoshal et al., 1999). It was previously shown that the M37L mutation did not affect the stability of the protein (Chakrabarti et al., 1999). Attempts were made to crystallize the M37L single mutant, but no suitable crystals could be obtained. This is not surprising as E.coli thioredoxin and its mutants are typically difficult to crystallize. However, NMR studies (see
below) were used to show that the M37L mutation alone has virtually no effect on the structure of the protein. The P40S mutant protein was crystallized by the hanging drop vapor diffusion method, using 25% ethanol as the precipitant in the presence of 10 mM cupric acetate and 100 mM sodium acetate buffer, pH 3.8, at room temperature (Holmgren and Soderberg, 1970). The crystals were obtained in 3 months and diffracted to 1.8 Å. The crystals belong to the triclinic space group P1 with the unit cell parameters \(a = 27.32 \text{ Å}, b = 37.50 \text{ Å}, c = 50.74 \text{ Å}, \alpha = 69.25^\circ, \beta = 79.71^\circ, \gamma = 85.39^\circ\).

X-ray diffraction data were collected using a MAR Research 300 mm image plate detector mounted on a Rigaku rotating anode X-ray generator. The data were collected at room temperature, using Cu Kα radiation. The crystallographic statistics are given in Table I.

Structure solution, model building and refinement

Structure solution was obtained after starting from the coordinates of wild-type thioredoxin (Katti et al., 1990) by molecular replacement using the program AmoRe (Navaza, 1998). The crystals were obtained in 3 months and diffracted to 1.8 Å. The correlation coefficient and R-factor were 58.0% and 37.3%, respectively. The structure solution so obtained was used for model building. Iterative cycles of refinement of the model were done using the program CNS (Brüner et al., 1994). The model was initially subjected to rigid body refinement and positional refinement in the resolution range 2–30 Å. This brought \(R\) and \(R_{\text{free}}\) (Brüner et al., 1992) to 29.08% and 33.43%, respectively. Mutations Met37Leu and Pro40Ser were incorporated in the model using the software O (Jones et al., 1991) by observing the electron density map. This model was improved by simulated annealing refinement using torsion angle dynamics. Slow cooling from an initial temperature of 5000 to 300 K in steps of 25 K was carried out with energy minimization before and after simulated annealing. This decreased \(R\) and \(R_{\text{free}}\) to 26.07% and 29.99%, respectively. Further manual rebuilding of the model, incorporation of cupric ions in the electron density map and simulated annealing refinement after including all the data, decreased \(R\) and \(R_{\text{free}}\) to 24.53% and 28.25% respectively. The addition of 133 water molecules combined with positional and isotropic temperature refinements decreased \(R\) and \(R_{\text{free}}\) to 18.2% and 22.2%, respectively. Throughout, manual rebuilding of the model was carried out after observing the \(2F_\text{o} - F_\text{c}\) and

<table>
<thead>
<tr>
<th>Table I. Data collection statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Radiation used</td>
</tr>
<tr>
<td>Detector</td>
</tr>
<tr>
<td>Temperature</td>
</tr>
<tr>
<td>Crystal system, space group</td>
</tr>
<tr>
<td>Cell parameters</td>
</tr>
<tr>
<td>Resolution</td>
</tr>
<tr>
<td>Unique reflections ([I/\sigma(I)] &gt; 0)</td>
</tr>
<tr>
<td>Multiplicity of data</td>
</tr>
<tr>
<td>Completeness</td>
</tr>
<tr>
<td>(R_{\text{merge}})</td>
</tr>
</tbody>
</table>

![Fig. 1. Overlay of amide proton–α-proton region of 50 ms TOCSY spectra of (A) wt and M37L thioredoxin and (B) wt and P40S thioredoxin. All proteins are in the oxidized state. Correlation peaks for wt thioredoxin are labeled with sequence-specific resonance assignments for \(H^3\) protons only. Also present in this region of the spectrum are resonances from \(H^2\) protons of Cys, Ser and Thr which have not been labeled.](https://academic.oup.com/peds/article/15/8/627/1532104)
Structure of P40S thioredoxin

Fig. 2. Association of two molecules in the asymmetric unit for (a) wild-type and (b) mutant P40S proteins.

$F_o - F_c$ electron density maps. Omit maps were computed for the final model by omitting 20 residues at a time and used to rebuild parts of model, so as to remove model bias. The quality of the final model was assessed by using the programs CNS and PROCHECK (Laskowski et al., 1993). Table II shows the final model statistics (Luzzati, 1952).

**NMR sample preparation**

Samples of wild-type, M37L and M37LP40S proteins for NMR were prepared by dissolving purified, lyophilized proteins in 50 mM sodium phosphate buffer (pH 6.0, 95% H$_2$O–5% D$_2$O) containing 50 mM NaCl. All protein samples were of 2–3 mM concentration.

**NMR data acquisition**

NMR data were recorded on a Bruker DRX 500 spectrometer operating at a proton frequency of 500.13 MHz. All spectra were recorded in the phase-sensitive mode using the TPPI (Marion and Wuthrich, 1983) method of quadrature detection. The sample temperature was maintained at 300 K and a relaxation delay time of 1 s was used in all experiments. Water suppression was achieved using the WATERGATE sequence as the read pulse (Piotto et al., 1992). Two-dimensional 50 ms ‘clean’ TOCSY (Greisinger et al., 1988) using the MLEV-17 spin-lock sequence (Bax and Davis, 1985) and 150 ms NOESY (Kumar et al., 1980) spectra were acquired with proton spectral widths of 8000 Hz in both the $F_1$ and $F_2$ dimensions. All spectra were recorded with 512 $t_1$ increments to give a matrix of 2048 complex $t_2$ points×512 real $t_1$ points.

**NMR data processing and analysis**

All NMR data were processed on an Intel PC workstation running Suse Linux 7.3 using NMRPipe/NMRDraw processing software (Delaglio et al., 1995). A Gaussian filter with a line broadening parameter of 15 Hz was applied in the direct and indirectly acquired dimensions. Data sets were zero-filled in both dimensions to yield final matrices of $2048 \times 2048$ real data points. Digital resolution in the 2D spectrum was 4 Hz per point. NMR data were analysed using ANSIG (Kraulis et al., 1994).

**Results and discussion**

NMR data show that structures of wild-type and M37L are very similar

The structures of both wild-type (wt) thioredoxin (Katti et al., 1990) and P40S (this work) have been solved by X-ray crystallography. P40S also contains an additional M37L mutation. However, this mutation has no effect on the stability of the protein (Chakrabarti et al., 1999). The conservative nature of the substitution coupled with the observations that the side chain of residue 37 is highly exposed (with an accessibility of
about 70% in both wt and P40S) and does not interact with other residues in the protein. It is entirely reasonable to conclude that the M37L mutation does not affect the protein structure. Since it was not possible to crystallize the M37L single mutant, structural comparisons between wt, M37L and P40S were carried out using NMR. The approach to making assignments of oxidized M37L thioredoxin resonances involved transfer of many amide proton resonances from the known assignments of oxidized wt thioredoxin (Dyson et al., 1989) by superposition of $^1$H-$^1$H TOCSY spectra. These assignments were then confirmed through identification of spin systems in the $^1$H-$^1$H TOCSY spectrum of M37L thioredoxin. Sequential connectivities were established through examination of 2D $^1$H NOESY spectra. Figure 1a shows an overlay of the amide proton–$\alpha$-proton region of the $^1$H-$^1$H TOCSY of wild-type and M37L thioredoxin. It is readily apparent that almost all resonances superimpose remarkably well (see Figure 1). Even residues immediately adjacent to the site of mutation on the primary sequence show very small perturbations in chemical shifts. For instance, K36, A39 and P40 show changes of only 0.01, 0 and 0.08 p.p.m., respectively, in $^1$H chemical shift whereas K36 and A39 show changes of only 0.04 and 0.03 p.p.m. in $^1$H chemical shift, respectively. The largest changes in chemical shift are exhibited by L37 and I38, which is not surprising. It should be mentioned that in the case of the M37L mutant, we have not been able to identify I38 unambiguously. It is most likely due to extreme overlap or because of exchange. We are unable to verify this at present. However, given that C35,K36,A39,P40 and 141 exhibit very small changes in chemical shifts, it is entirely reasonable to conclude that the structural perturbations in M37L in the vicinity of the mutation are minimal. Although it is difficult to correlate changes in chemical shift with exact changes in structure, the near identity of chemical shifts indicates structural similarity, especially when analysed in conjunction with NOE data. On the other hand, superposition of similar regions of the NMR spectra of wild-type and P40S thioredoxins (Figure 1B) reveals that there are significant differences in chemical shifts for backbone nuclei and transfer of assignments cannot be as easily accomplished as in the case of the M37L single mutant. Hence it is reasonable to conclude that structural differences between wt and P40S are due to the P40S mutation alone and that the M37L substitution has virtually no effect on the structure of the protein.

**Overall structure of P40S mutant**

There are two molecules A and B in the asymmetric unit. Both molecules were refined independently without the application of non-crystallographic restraints. These molecules were superimposed by the least-squares method using the program ALIGN (Cohen, 1997). The C$\alpha$ trace of superimposed molecules has an r.m.s.d. of 0.28 Å. The molecule is also superimposed on wt thioredoxin. The average r.m.s.d. for the superposition of C$\alpha$ traces of wt and mutant molecules is 0.67 Å. The thioredoxin protein molecule has five $\beta$ strands surrounded by four helices to form two conformational domains $\beta$$\alpha$$\beta$$\alpha$ and $\beta$$\beta$$\alpha$. In P40S, the four helices extend from residue 11 to 17, 34 to 49, 59 to 63 and 95 to 107 in both molecules of the asymmetric unit. The second helix is the longest and in the wt this helix extends from residue 32 to 49, with a kink
in the helix at position 40. The kink was thought to result from the presence of Pro at position 40. The relative orientations of the A and B molecules are different in wt and P40S, although the overall structures of the individual chains are very similar (Figure 2). The only significant differences between wt and P40S are in helix two, and these are discussed in greater detail below. In addition to the P40S mutation, the mutant also contains the M37L mutation. However, residue 37 is highly accessible and does not interact with other regions of the structure in either wt or P40S. Hence differences between the two molecules result exclusively from the P40S mutation.

Comparison of hydrogen bonding in helix two of wt and P40S

The intramolecular hydrogen bonding pattern present in helix two for both wt and P40S is summarized in Table III. In the wt protein Pro40 lacks an amide hydrogen and hence does not form the usual main chain hydrogen bond with the carboxyl of Lys36. However, the CD of Pro40 forms a C–H···O hydrogen bond with the carbonyl group of residue 37 (Chakrabarti and Chakrabarti, 1998). In the mutant structure there are additional bifurcated hydrogen bonds from Ser40 NH to the carbonyls of Leu37 and Lys36 compared with the wt. In addition, the side chain atom OG of Ser40 is located at a hydrogen bonding distance of 2.8 Å from the carbonyl group of residue 36. The superposition of this modified helix on to the corresponding helix in the wt protein is depicted in Figure 3. It can be seen that the kink in the helix is still maintained and the kink angle (Collaborative Computational Project Number 4 (CCP4), 1994) increases by 6° compared with the wt protein. It is clear that in wt, neither residue 40 nor residue 41 amide NHs are involved in hydrogen bonding interactions with other groups in the protein or with water molecules. In P40S, the amide group of residue 41 forms a 3_10-helical hydrogen bond with the carboxyl of residue 38. Unexpectedly in P40S, amide groups of residues Asp43 and Glu44 do not form main chain hydrogen bonds with carbonyl groups of residues of 39 and 40, respectively, although these hydrogen bonds are present in wt. Figure 4 shows a comparison of main chain hydrogen bonding within the helix in the wt and mutant structures. There is no net increase in the number of main chain hydrogen bonds in P40S relative to wt. However, there is an increase of one side chain–main chain hydrogen bond due to the S40 side
Table IV. Comparison of \( \phi \) and \( \psi \) values in helix two of E.coli thioredoxin (values are averaged over the two molecules in the asymmetric unit)

<table>
<thead>
<tr>
<th>Residue</th>
<th>M37L.P40S</th>
<th>Wild-type</th>
<th>( \Delta \phi ) (°)</th>
<th>( \Delta \psi ) (°)</th>
</tr>
</thead>
<tbody>
<tr>
<td>32Cys</td>
<td>–87</td>
<td>–90</td>
<td>–3</td>
<td>0</td>
</tr>
<tr>
<td>33Gly</td>
<td>–60 –58</td>
<td>–61 –54</td>
<td>–1</td>
<td>4</td>
</tr>
<tr>
<td>37Met</td>
<td>–62 –52</td>
<td>–63 –33</td>
<td>–1</td>
<td>19</td>
</tr>
<tr>
<td>38Ile</td>
<td>–63 –38</td>
<td>–92 –9</td>
<td>–29</td>
<td>29</td>
</tr>
<tr>
<td>41Ile</td>
<td>–105 –12</td>
<td>–68 –36</td>
<td>37</td>
<td>–23</td>
</tr>
<tr>
<td>42Leu</td>
<td>–60 –40</td>
<td>–64 –37</td>
<td>–4</td>
<td>3</td>
</tr>
<tr>
<td>44Glu</td>
<td>–68 –42</td>
<td>–62 –42</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>46Ala</td>
<td>–57 –37</td>
<td>–57 –41</td>
<td>0</td>
<td>–4</td>
</tr>
<tr>
<td>47Asp</td>
<td>–75 –46</td>
<td>–70 –49</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>48Glu</td>
<td>–63 –34</td>
<td>–65 –32</td>
<td>–2</td>
<td>2</td>
</tr>
<tr>
<td>49Tyr</td>
<td>–104 14</td>
<td>–110 16</td>
<td>–6</td>
<td>–2</td>
</tr>
</tbody>
</table>

Fig. 5. Root mean square difference between wt and P40S for backbone atoms in helix two of E.coli thioredoxin.

Main chain dihedral angle analysis

The r.m.s.d. plot (Figure 5) for the backbone atoms in this helix region in the mutant shows that the atoms of residues 37–43 have shifts in the mutant relative to the wt. The comparison of Ramachandran angles in the helix region for the wt and mutant structures reveals that the \( \phi, \psi \) values are significantly perturbed for residues 37–41 (Table IV). The \( \phi \) and \( \psi \) values of Ser40 and Ile41 in P40S lie outside of the

chain. The plot shows that replacement of proline by serine has substantially decreased the distance between the carbonyl of residue 36 and the NH of residue 40 from 4.6 to 3.2 Å. A similar kind of decrease is observed between the mutant structure of the P61A and its corresponding wt Fis protein of E.coli (Yuan et al., 1994) and also in mutated structures of P86 by different amino acids (Alber et al., 1988) in T4 lysozyme.
typical values for an α-helix. In P40S, there is a decrease in the ψ angle of Ala39 which results in a loss of the H-bond between residues 39 and 43. There is a similar disruption of the main chain hydrogen bond between CO40 and NH44 in P40S. In order to satisfy their hydrogen bonding potential, both amides 43 and 44 as well as carbonyls 39 and 40 are involved in hydrogen bonding with water molecules (Figure 6). These water molecules are absent in wt. A total of five hydrogen bonds are formed between the above four groups and water molecules. Thus, when protein–water hydrogen bonds are taken into account, there is a net increase of six hydrogen bonds in P40S relative to wt.

Comparisons with other mutant structures

An analysis was also made of structural changes occurring in the two other proteins (Fis protein, mutant P61A and T4 lysozyme, mutant P86A) where replacements of α-helical Pro have been made. In both of the wt proteins, the only main chain group in the relevant helix with an unsatisfied hydrogen bond was the carbonyl group located at position i – 4. In both these cases replacement of the Pro at position i was accompanied by the formation of an additional hydrogen bond between the amide of the substituted residue and the carbonyl of residue i – 4. In addition to this change, in the case of the Fis protein, four additional intrahelical hydrogen bonds are present in the mutant helix. These are as follows: N61–O58, N62–O59, N66–O63, N67–O64 and N69–O66. In contrast, for T4 lysozyme, no other new hydrogen bonds are formed and there is a loss of one hydrogen bond (N88–O85) that was present in the wt protein. Furthermore, the carbonyl group i – 4 to the mutation site is buried in Fis and has no possibility of interacting with solvent. However, in T4 lysozyme, the corresponding carbonyl is exposed, suggesting that it may be hydrogen bonded to aqueous solvent. Although no water molecules were seen within hydrogen bonding distance in the wt T4 lysozyme structure, it is possible that the water was present but disordered. Consistent with this analysis, in the case of Fis protein the mutation had a stabilizing effect whereas in the case of T4 lysozyme no increase in stability was seen (Alber et al., 1988). The experimentally observed increase in protein stability for replacement of an α-helical Pro in the three proteins studied to date is Fis > thioredoxin > T4 lysozyme. This order correlates well with the additional hydrogen bonds formed upon mutation. The additional hydrogen bonds in the case of Fis are all intrahelical ones between main chain atoms, whereas in the case of thioredoxin, they are between protein and solvent. The latter are likely to be less stabilizing because of the unfavorable entropy associated with solvent ordering. This work emphasizes the importance of obtaining high-resolution crystal structures to interpret mutational data as the additional protein–solvent hydrogen bonds would probably not have been detectable at even a slightly lower resolution. The analysis suggests that replacements of α-helical Pro residues can result in protein stabilization. Stabilization is more likely to occur if the carbonyl group at position i – 4 relative to the mutation site is buried and not hydrogen bonded in the wt structure. Although Pro-containing helices are kinked, replacement of the Pro residue does not result in removal of the kink, suggesting that packing constraints with the rest of the protein also play an important role in formation of the kink. In addition to localized changes in hydrogen bonding, there may also be contributions from small changes distributed over the structure of the protein as well as in the unfolded state of the protein. In particular, since Pro is a rigid residue, replacements of Pro with other residues are likely to result in an increase in the conformational entropy of the unfolded state and a consequent decrease in protein stability. However, the results of this work and also other studies described above suggest that in selected cases replacement of Pro residues in a helix interior can lead to protein stabilization. The structural changes in response to such replacements are complex and involve changes in both protein and the surrounding solvent.

Acknowledgements

Computational facilities of the Supercomputer Education and Research Center and DBT supported Interactive Graphics Facility were used for this work. The X-ray data were collected at the Image Plate Facility at the Molecular Biophysics Unit. This work was supported by grants from DST, DAE and CSIR, Government of India, to S.R. and R.V. R.V. is a recipient of a Swarnajayanthi Fellowship from DST and is a Senior Research Fellow of the Wellcome Trust. The crystal structure and structure factors of the P40S structure have been deposited at the RCSB protein databank. S.S. thanks DST and DBT for funding the 500 MHz NMR machine at SIF, Indian Institute of Science.

References


Received September 27, 2001; revised April 9, 2002; accepted April 18, 2002