J. Mol. Model. **1997**, *3*, 359 – 363

Journal of Molecular Modeling

© Springer-Verlag 1997

Homology Modelling of a Newly Discovered Thioredoxin Protein and Analysis of the Force Field and Electrostatic Properties[$]

Günter Raddatz[1,*], Victoria I. Bunik[2], Thomas Scior[1] and Hans Bisswanger[1]

[1]Physiologisch-Chemisches Institut Universität Tübingen, Hoppe-Seyler-Str. 4, 72076 Tübingen, Germany;
Tel: +49-7071-297-4180, Fax: +49-7071-29-3360 (guenter.raddatz@uni-tuebingen.de)

[2]A.N. Belozersky Institute of Physico-Chemical Biology, Moscow State University, Moscow, Russia

Received: 17 June 1997 / Accepted: 4 July 1997 / Published: 14 August 1997

Abstract

Thioredoxin is a small protein (Mr approximately 12,000) found in all living cells from archaebacteria to humans. The active site is highly conserved and has two redox-active cysteine residues in the sequence: -Trp-Cys-Gly-Pro-Cys-. Besides the function of the reduced form as a powerful protein disulfide oxidoreductase, thioredoxin is known to regulate and activate different target enzymes, i.e. ribonucleotide reductase and the mitochondrial 2-oxoacid dehydrogenase multienzyme complexes. Despite the high degree of homology between thioredoxin proteins from different species, there exists a strong variation in the capability of activating target enzymes. This is yet unexplainable, since there still exists no model of a thioredoxin/receptor complex.
On the basis of the recently determined amino acid sequence of the thioredoxin Trx2 from rat mitochondria, which is known to be highly efficient in activating mitochondrial 2-oxoacid dehydrogenase multienzyme complexes, we construct the 3-D structure of this protein by homology modelling methods, using the X-ray structures of thioredoxin from *E. coli* and human as background information. We analyze the differences in the electrostatic properties of the different protein structures and show, that despite the observed homology between the primary sequences, the dipole moment of the protein structures shows significant variations, which might lead to deviations with respect to the binding to the target protein. Using the AMBER 4.0 program package we further investigate and compare the force field energies of the different thioredoxin structures.

Keywords: Thioredoxin, Homology modelling, Electrostatic potential calculations, Protein dipole, Force field energies

Introduction

Thioredoxin is a small protein (Mr approximately 12,000) found in all living cells from archaebacteria to humans. The active site is highly conserved and has two redox-active cysteine residues in the sequence: -Trp-Cys-Gly-Pro-Cys-

[$] *Presented at the 11. Molecular Modeling Workshop, 6 -7 May 1997, Darmstadt, Germany*

* *To whom correspondence should be addressed*

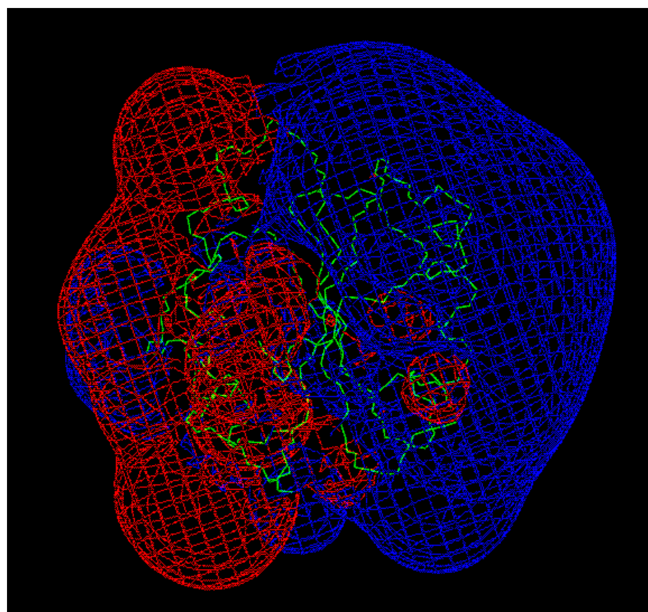

Figure 1. *Electrostatic isopotential surfaces of the constructed model of rat mitochondrial thioredoxin.*

Table 1. *Kollman all atom force field energies of the different thioredoxin structures solvated in a shell of TIP3P-water after 5000 steps conjugate gradient geometry optimization (kcal/mol).*

	Model of of Trx2	*Escherichia coli*	Human cytosol
Total Energy	-9088.3	-9205.2	-9041.6
Bond	165.4	160.2	154.4
Angle	169.4	129.9	115.0
Torsion	143.6	126.5	123.3
VdWaals	-123.3	-167.6	-117.6
Electrostatics	-9269.4	-9279.4	-9146.6
H-Bond	-174.1	-174.9	-170.3

which is localized in a protrusion of the protein. Oxidized thioredoxin Trx-S2 is reduced to Trx-(SH)2 by the flavo-enzyme thioredoxin reductase, thioredoxin and its correspondent thioredoxin reductase forming together the 'thioredoxin system'. Thioredoxins have been implicated in a wide variety of biological functions. Besides the function of the reduced forms as powerful protein disulfide oxidoreductases, thioredoxins can modulate the DNA binding activity of some transcription factors like BZLF1 and NF-kB and act as hydrogen-donor for ribonucleotide reductase (see [1] for an overview). Recently a functional interplay between mitochondrial 2-oxoacid dehydrogenase multienzyme complexes and thioredoxin has been demonstrated experimentally [2], establishing the protective role of thioredoxin during the catalysis induced inactivation of the complexes. It was shown, that despite the strict conservation of the active site environment between thioredoxin from different sources, there exists a significant variation in the efficiency of protection. Thus, mitochondrial thioredoxin shows the protective effect on the 2-oxoacid dehydrogenase multienzyme complexes in concentrations, which are 10-fold lower than those of thioredoxin from *E. coli*. The detailed mechanism of the thioredoxin induced protection is under investigation. Knowledge about the structural and electrostatic differences between the thioredoxins studied would be very useful to develop a rationale about the differences in their action, but no X-ray or NMR structure of mitochondrial thioredoxin has been solved as yet. Moreover, although thioredoxin has been crystallized from a number of organisms, there still does not exist any X-ray structure nor any model of a thioredoxin/target-protein complex. On the basis of the recently determined amino acid sequence of the thioredoxin Trx2 from rat mitochondria [3], we construct the 3-D structure of this protein by homology modelling methods, using the X-ray structures of thioredoxin from *E. coli* and human [4] as background information. The perspective of this work is to gain insights into the structural resp. electrostatical determinants, which could be responsible for the different specificities of the various thioredoxin proteins and to lay the basis for the construction of a complete thioredoxin/target-protein model.

Modelling Methods

The sequences, which show the highest degree of sequence homology with rat mitochondrial thioredoxin Trx2 were obtained with the BLAST search tool, available at the NCBI (http://www.ncbi.nlm.nih.gov/BLAST/). Multiple sequence alignments as prerequisites for the homology based structure construction were performed with the the program CLUSTALW, which is accessible at the WWW-server at the BCM (http://dot.imgen.bcm.tmc.edu:9331). The 3-D structures of *E. coli* resp. human thioredoxin, which showed the highest sequence homology with Trx2 of the proteins with known X-ray structure were used as main templates for the model construction and were obtained from the Protein Data Bank Brookhaven [5] (http://www.pdb.bnl.gov/) under the entries 2TRX resp. 1ERT. The construction of the model was performed within the SYBYL 6.4 modelling program package, using the following strategy: First, the structure of *E. coli* thioredoxin, which shows the highest degree of homology to Trx2 with 35 out of 108 (32.4%) amino acids identical and 24 further conservative mutated was used as the main template for the construction. Due to the high degree of conservation, especially with respect to the amino acids forming the structural core of the protein, the backbone conformation of the *E. coli* protein was adopted for the model. Secondly, sidechain conformations of amino acids changed with respect

Table 2. *Dipole Moments (Debye) of the different thioredoxin structures.*

	Model of Trx2	*Escherichia coli*	Human cytosol
Total Dipole Moment	578.1	297.3	223.5
X component	-556.1	-125.9	-188.4
Y component	-67.1	230.0	-53.2
Z component	-143.0	-140.3	-170.9
Backbone Value	74.3	81.2	74.7
Net charge	-6.0	-5.0	-5.0

to the *E. coli* structure, but conserved with respect to human thioredoxin were adopted from the latter. Thereafter, any amino acid, changed with respect to 2TRX or 1ERT but conserved in any other thioredoxin structure contained in the PDB was modelled according to the conformation in this structure. Third, among the remaining residues those amino acids were searched, which are possible salt bridge partners due to charge and position. If a correlated occurence of these amino acids was detected from the multiple sequence alignment, produced with CLUSTALW, they were modelled as partners of a salt bridge. Finally, the remaining amino acids were modelled to create optimal steric and electrostatic interactions with the environment. The geometry optimization was performed with the program AMBER 4.0 [6] running on a CONVEX-220 using the Kollman all atom force field [7]. The structures were solvated in a shell of TIP3P-water [8] and geometry optimized by 5000 steps conjugate gradient energy minimization. The constructed model of Trx2 was subjected to the programs PROCHECK and SURVOL, available at the EMBL Heidelberg (http://biotech.embl-heidelberg.de:8400), to test the quality of the constructed model. Electrostatics potential surfaces and dipole vectors were computed with the SYBYL 6.4 modelling package using the tools DIPOLE, POTENTIAL and CONTOUR. Potentials are computed by determining a parallelepiped, which completely contains the protein structure and dividing this space into a three-dimensional grid, which spacing is identical in all directions. At each point in the grid, the potential value is computed as

$$V = 322 \cdot \sum_i \frac{q(i)}{\varepsilon \cdot |r(i) - p|}$$

with V being the interaction potential felt by a unit charge at grid point p, with 332 a constant, to convert the resulting energy value to kcal/mol, $q(i)$ the point charge on the i-th atom, ε the dielectric value and $|r(i)–p|$ the distance from the i-th atom to the grid point p. The dipole calculation is based on the point charge distribution of the molecule. All amino acids were considered to have standard protonation states at pH 7.0.

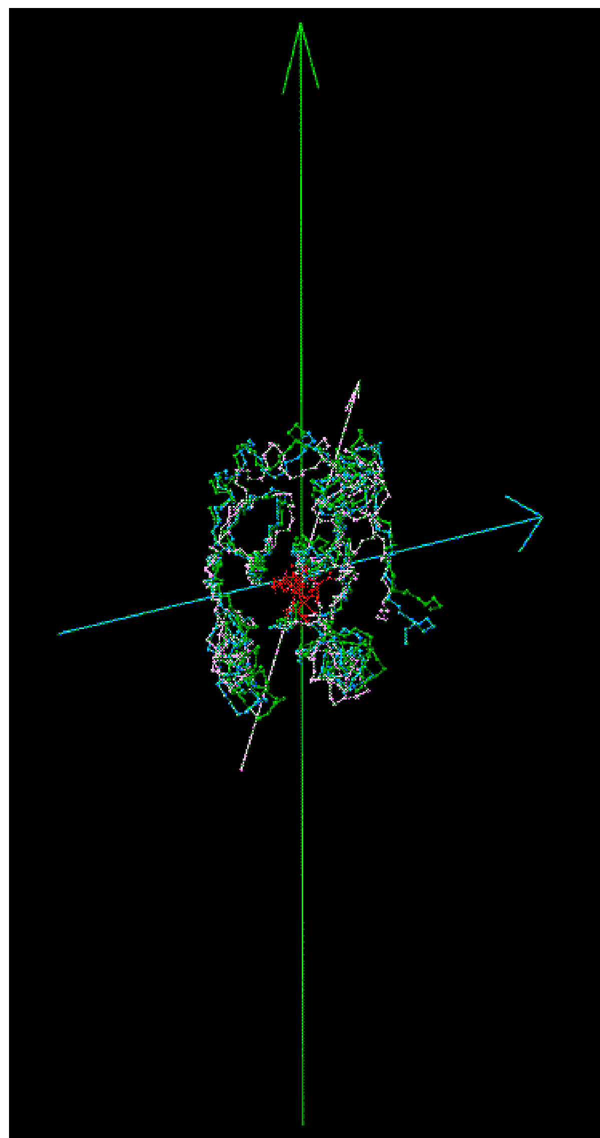

Figure 2. *View of the dipole vectors of the superposi-tioned thioredoxin structures from rat mitochondria (green), E. coli (cyan) and human (white). The active site disulphide bridge is coloured red. All dipole vectors are lying in a plane parallel to the binding site of the protein. The total dipole moments sum up to 578.1 D (rat mitochondria), 297.3 D (E. coli) and 223.5 D (human).*

Results

The check of the quality of the constructed model of Trx2, using the program SURVOL, results in a volume score of 0.8 concerning the backbone and 0.9 concerning the sidechain

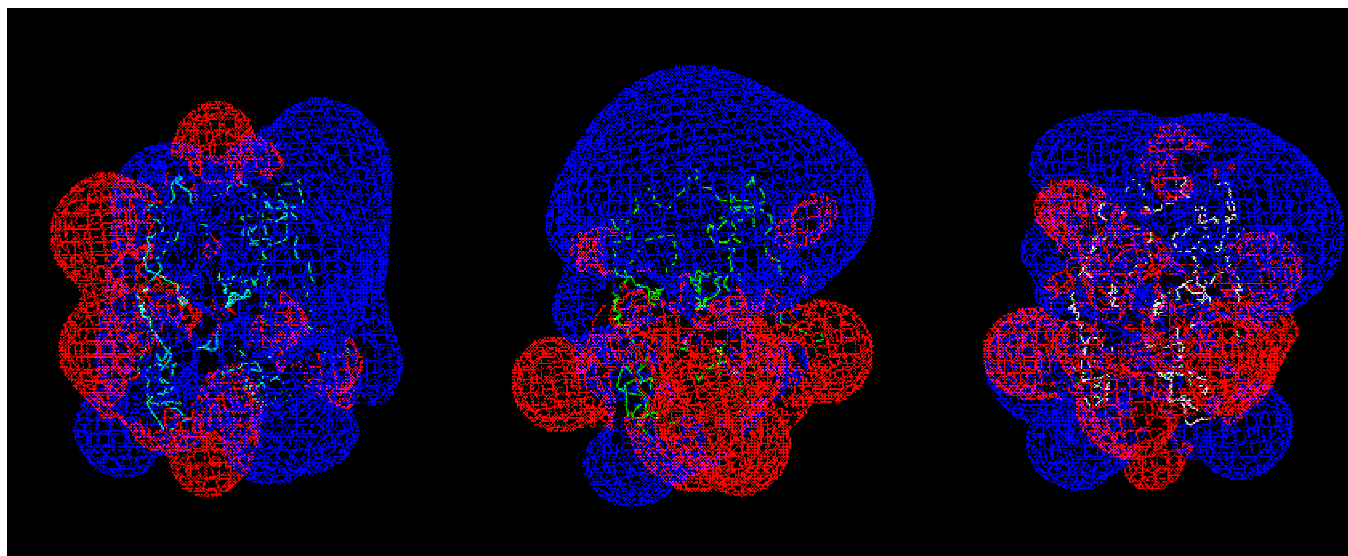

Figure 3. *Electrostatics potential plot produced with the POTENTAL and CONTOUR Tools of SYBYL 6.4 for the thioredoxin from E. coli (left), for the model of mitochondrial Trx2 (middle) and from human (right). Isopotential lines are shown for -10 Kcal/M (blue) and +5 Kcal/M (red).*

atoms of the model. The average value for the whole structure is 0.8, which is near to the optimal value of 0.82. The overall average G factor of the model, computed with PROCHECK, amounts to -0.15. All these values are in the range, which indicates an acceptable model structure. As shown in table 1, there is no essential energetic difference between the created model of Trx2 and the two template structures after geometry optimization, especially if the minimal different chain lengths (107 / 108 / 105 amino acids) are taken into account. The slightly higher van der Waals energy of Trx2 with respect to the *E. coli* protein could perhaps be lowered by further efforts in the model building.

The structural environment of the active site is identical in the mitochondrial thioredoxin and the other two structures, the only difference consisting in the somewhat bigger sidechain of Ile93 in Trx2, which is replaced by alanine in the other proteins. The only two amino acids, for which a possible non-standard ionization at pH 7.0 has been reported, are the buried amino acid Asp26 and the catalytically active Cys32. Jeng and Dyson [9] showed for the thioredoxin from *E. coli*, that Asp26 has an unusal high pKa of about 7.0, whereas according to Follmann [1] the pKa of Cys32 has the unusual low value of about 7.4. Assuming standard amino acid residue protonation states at pH 7, the net charge is -6e on Trx2 and -5e on the proteins from *E. coli* and from human. This agrees with measured pI values for thioredoxins from different organisms, which normally lie in the range between 4.5 - 5.5. In particular, for *E. coli* a pI value of 4.5 was determined in [10]. The most remarkable difference between the three structures consists in the significantly altered charge distribution of the whole protein structure, which leads to the high dipole moment of Trx2, shown in table 2.

Whereas the dipole moments of the *E. coli* resp. human protein resemble the values, which have been reported for a variety of other structures in [11], the dipole moment of Trx2 is about twice as high and nearly reaches the highest value of 637 D, which was determined for carboxypeptidase in [12]. With respect to its orientation the dipole vector of Trx2 resembles the human protein, whereas it is nearly orthogonal to the protein from *E. coli* (Figure 2).

A computation of the dipole vector, which is contributed by the protein backbone alone, shows, that its value is rather low and in orientation different from the dipole vector of the complete structure. The orientation as well as the high value of the dipole vector of Trx2 are mainly caused by a cluster of five uncompensated, negatively charged amino acids, which are contained in the alpha helix a3 (Asp58-His62) and the neighbouring 3/10-helix (Asp64-Glu68). These are also present in the human protein, but completely absent in the one from *E. coli*, which explains the deviation of the orientation of the dipole vector of the *E. coli* structure. The lower magnitude of the dipole vector for the human protein seems to be mainly caused by the presence of the three basic residues Lys73, Lys83 and Lys86, which compensate due to their position somewhat for the dipole moment, induced by the negative charged cluster and which are changed to uncharged residues in Trx2.

The visualization of isopotential surfaces enables further comparison of electrostatic potential characteristics. In Figure 3 the determined isopotential surfaces for the different thioredoxin proteins are shown, where the view-axis is oriented as in Figure 2 and points directly onto the active site. It can be seen, that the high dipole moment of Trx2 is caused by the clear separation of a positive and a negative electrostatics isopotential lobe. The charge distribution around the active site is quite different between Trx2 and the protein

from *E. coli*, where especially the regions of positive potential are more distributed.

Discussion

On the basis of the recently determined amino acid sequence of the thioredoxin Trx2 from rat mitochondria [3], the complete 3-dimensional structure of the protein was modelled. The high degree of homology with respect to the *E. coli* enzyme and the sufficient amount of background information, contributed by thioredoxin structures from other sources, enabled the creation of a model, in which every amino acid could be modelled in a convincing way. The most remarkable difference between Trx2 and the other thioredoxin proteins consists in the enormously high dipole moment of Trx2, which is about twice as high, as with the human resp. the *E. coli* protein and resembles the highest values reported in the literature [11,12]. Concerning the orientation of the dipole moment, Trx2 is similar to the human protein, whereas the dipole vector of the *E. coli* protein is nearly orthogonal to that of the superpositioned Trx2. It was shown in [11] that the computed values of protein dipoles are in good agreement with experimental data, especially for small protein structures. In [13] Janin investigates the possible role of the protein dipole moment in protein/target interactions and proposes a model, in which the contribution of the protein dipole drastically accelerates the formation of a productive binding mode between barnase and barstar. It seems reasonable, that such dipole interactions can support also to the correct binding of thioredoxin to its target enzymes and consequently could allow to tune the specificity of the protein from different sources with respect to the different target proteins. In [14] this idea has been pursued for the interaction of ferredoxin (FN) with its correspondent ferredoxin oxidoreductase (FNR) and it was shown, that by adjusting the protein dipole vectors of FN and FNR, a reasonable binding model for the FD:FNR complex could be developed. Although a detailed analysis of the electrostatic potential distribution with respect to potential binding regions principally could allow a more comprehensive understanding of receptor binding properties, in the case of thioredoxin this is difficult to achieve. At the moment, nearly nothing is known about the contact regions to target proteins, although the immediate environment of the active site is certainly involved in binding. So we consider insight into the dipole properties of the different thioredoxin proteins at the current state of knowledge as a valuable step towards the construction of a binding model of thioredoxin and its target proteins.

Supplementary material: model of Trx2 as PDB-file

Literature

1. Follmann, H.; Haeberlein. I. *BioFactors*. **1996**, *3*, 147-156.
2. Bunik, V. I.; Follmann, H.; Bisswanger, H. *Biochem. Hoppe-Seyler*. in press
3. Spyrou, G.; Enmark, E.; Miranda-Vizuete, A.; Gustafsson, J. A. *J. Biol. Chem.* **1997**, *272*, 2936-2941.
4. Eklund, A.; Gleason, F. K.; Holmgren, A. *Proteins*. **1991**, *11*, 13-28.
5. Bernstein, F. C.; Tasumi, M. *J. Mol. Biol.* **1977**, *112*, 535-540.
6. Kollman, P. A.; Weiner, S. J. *J. Comp. Chem.* **1991**, *2*, 287-299.
7. Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. *J. Am. Chem. Soc.* **1984**, *106*, 765-780.
8. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J.; Impey, R.; Klein, M. *J. Chem. Phys.* **1983**, *79*, 926-931.
9. Jeng, M. F.; Dyson, H. J. *Biochemistry* **1996**, *35*, 1-6.
10. Holmgren, A. *Eur. J. Biochem.* **1968**, *6*, 475-484.
11. Takashima, S.; Asami, K. *Biopolymers* **1993**, *33*, 59-68.
12. Barlow, D. J.; Thornton, J. M. *Biopolymers* **1986**, *25*, 1717-1724.
13. Janin, J. *Protein Science* **1997**, *6 Suppl. 1*. Abstract 105.
14. De-Pascalis, A. R.; Jelesarov, I.; Ackermann, F.; Koppenol, W. H.; Hirasawa, M.; Knaff, D. B.; Bosshard, H. R. *Protein Science* **1993**, *2*, 1126-35.