Identification of a Novel Archaebacterial Thioredoxin: Determination of Function through Structure†

Sudeepa Bhattacharyya,‡ Bahram Habibi-Nazhad,‡ Godwin Amegbey,‡ Carolyn M. Slupsky,§ Adelinda Yee,|| Cheryl Arrowsmith,|| and David S. Wishart*,‡

Faculty of Pharmacy and Pharmaceutical Sciences, University of Alberta, Edmonton, Alberta, Canada T6G 2N8, Department of Biochemistry, University of Alberta, Edmonton, Alberta, Canada T6G 2H7, and Division of Molecular and Structural Biology, Ontario Cancer Institute, 610 University Avenue, Toronto, Ontario, Canada M5G 2M9

Received July 20, 2001; Revised Manuscript Received November 28, 2001

ABSTRACT: As part of a high-throughput, structural proteomic project we have used NMR spectroscopy to determine the solution structure and ascertain the function of a previously unknown, conserved protein (*MtH*895) from the thermophilic archeon *Methanobacterium thermoautotrophicum.* Our findings indicate that *MtH*895 contains a central four-stranded β-sheet core surrounded by two helices on one side and a third on the other. It has an overall fold superficially similar to that of a glutaredoxin. However, detailed analysis of its three-dimensional structure along with molecular docking simulations of its interaction with T7 DNA polymerase (a thioredoxin-specific substrate) and comparisons with other known members of the thioredoxin/glutaredoxin family of proteins strongly suggest that *MtH*895 is more akin to a thioredoxin. Furthermore, measurement of the pK_a values of its active site thiols along with direct measurements of the thioredoxin/glutaredoxin activity has confirmed that *MtH*895 is, indeed, a thioredoxin and exhibits no glutaredoxin activity. We have also identified a group of previously unknown proteins from several other archaebacteria that have significant (34−44%) sequence identity with *MtH*895. These proteins have unusual active site -CXXC- motifs not found in any known thioredoxin or glutaredoxin. On the basis of the results presented here, we predict that these small proteins are all members of a new class of truncated thioredoxins.

The exponential growth in genome sequence data has placed increasing pressure on protein chemists to rapidly identify the function of many unknown or unclassified proteins. In cases where sequence comparisons fail to identify potential homologues or functional analogues, structural studies may go a long way toward revealing the function of the protein of interest (*1*−*4*). Because of the potential applications in functional classification, structural biologists are beginning to develop high-throughput X-ray and NMR methods for rapid functional and structural characterization of proteins. Indeed, a number of international structural genomic initiatives are now underway aimed at solving the structure (and identifying the function) of a large number of proteins from a variety of model organisms (*4*). One such model organism is the archaebacterium *Methanobacterium thermoautotrophicum* (ΔH) (*MtH*),[1] a small thermophilic bacterium first sequenced in 1996 (*5*). This particular archeon was chosen for this pilot project not only for its phylogenetic uniqueness but also because it offered an opportunity to better understand the structural basis of the differential thermostability between thermophilic and mesophilic proteins (*1*). To date, more than a dozen protein structures have been solved and characterized for this particular organism (*1*).

The genome of *M. thermoautotrophicum* (*MtH*) contains about 1870 proteins, of which fewer than 50% have been assigned functions based on BLAST sequence analysis (*6*). *MtH*895 is a small 77-residue protein identified in 1999 as a conserved hypothetical protein with unassigned function. Because of its small size and good solution behavior, this protein was chosen for detailed structural analysis by NMR spectroscopy. Here we wish to report on the high-resolution structure of *MtH*895 and the subsequent identification of this protein (through sequential, structural, and biochemical comparisons) as what appears to be the smallest known member of the thioredoxin family.

† Supported by the Natural Sciences and Engineering Research Council (GP01957270), the Protein Engineering Network of Centres of Excellence (PENCE), and the Ontario Cancer Institute (OCI).

* To whom correspondence should be addressed. Tel: 780-492-0383. Fax: 780-492-5305. E-mail: david.wishart@ualberta.ca.

‡ Faculty of Pharmacy and Pharmaceutical Sciences, University of Alberta.

§ Department of Biochemistry, University of Alberta.

|| Division of Molecular and Structural Biology, Ontario Cancer Institute.

[1] Abbreviations: BLAST, basic local alignment search tool; CE, combinatorial extension; DSS, 2,2-dimethyl-2-silapentane-5-sulfonic acid; DTNB, 5,5′-dithiobis(2-nitrobenzoic acid); DTT, dithiothreitol; EDTA, ethylenediaminetetraacetic acid; Grx, glutaredoxin; HSQC, heteronuclear single-quantum coherence spectroscopy; MD, molecular dynamics; *MtH*, *Methanobacterium thermoautotrophicum* (strain ΔH); NADPH, nicotinamide adenine dinucleotide phosphate (reduced); NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; NOESY, nuclear Overhauser effect spectroscopy; PSI-BLAST, position-specific iterative BLAST; RMSD, root mean square deviation; SCOP, structural classification of proteins; TOCSY, total correlation spectroscopy; Trx, thioredoxin.

10.1021/bi0115176 CCC: $22.00 © 2002 American Chemical Society
Published on Web 03/22/2002

MATERIALS AND METHODS

NMR Sample Preparation. Unlabeled, uniformly ^{15}N-labeled, and uniformly ^{13}C/^{15}N doubly labeled protein samples were generous gifts from the laboratory of Dr. Cheryl Arrowsmith (University of Toronto). The 234 bp *MtH*895 gene was expressed using the pET15b (Novagen) expression system in *Escherichia coli* cells [BL21(DE3)] and purified by affinity chromatography using a Ni-NTA column as described elsewhere (*7*). The NMR samples were prepared by dissolving ~5 mg of protein in 0.5 mL of 25 mM potassium phosphate buffer (pH 7) containing 200 mM NaCl and 10% D_2O. About 0.1 mM 2,2-dimethyl-2-silapentane-5-sulfonic acid (DSS) was added to the samples as a chemical shift reference (*8*). The presence of two cysteines in the protein sequence suggested that thiol oxidation could lead to solubility problems. To keep the protein in a reduced state, excess (20 mM) dithiothreitol (DTT) was added to the NMR tube, which was then sealed with parafilm after degassing and refilling the tube with argon gas.

NMR Spectroscopy. NMR experiments were recorded at 25 °C on a Varian Unity 500 MHz spectrometer equipped with a 5 mm triple resonance probe and pulsed field gradient accessories. Two-dimensional ^{1}H NOESY (with mixing times of 75 and 150 ms) (*9*), ^{1}H−^{15}N HSQC, 3D ^{1}H−^{15}N NOESY−HSQC (mixing time of 75 ms), ^{1}H−^{15}N TOCSY−HSQC (*10*, *11*), and HNHA (*12*) spectra were acquired with the unlabeled or ^{15}N-labeled sample. In addition, ^{13}C/^{15}N-edited NOESY−HSQC (mixing time of 75 ms) (*13*), HNCACB (*14*), CBCA(CO)NH (*15*), and HCCH−TOCSY (*16*) as well as a 2D ^{1}H−^{13}C constant time HSQC experiment were collected on a ^{13}C/^{15}N doubly labeled sample. Data were processed and analyzed using NMRPipe (*17*) and PIPP (obtained from Dr. Garret, NIH), respectively.

Assignments and Experimental Restraints. Beginning with the identification of ^{15}N and ^{1}HN chemical shifts from the ^{1}H−^{15}N HSQC spectra (Figure 1), spin systems were initially identified using ^{1}H−^{15}N TOCSY−HSQC data. Subsequently, backbone sequential assignments were completed using HNCACB and CBCA(CO)NH spectra using standard methods (*18*) and confirmed using ^{15}N/^{13}C NOESY−HSQC data. Side chain assignments were completed using data from HCCH−TOCSY as well as 3D TOCSY and NOESY−HSQC spectra. Stereospecific assignments of ^{1}Hβ protons were based on the intensity of ^{1}HN−^{1}Hβ cross-peaks in ^{15}N-edited TOCSY and NOESY−HSQC spectra (*19*). The methyl groups of Val and Leu were assigned stereospecifically on the basis of the intensity of ^{1}HN−^{1}Hγ, ^{1}Hα−^{1}Hγ cross-peaks and the NOE intensity of the stereospecifically assigned ^{1}Hβ protons to the δ-methyl protons of Leu. NOE distance restraints were obtained using 3D ^{15}N NOESY−HSQC and 3D simultaneous ^{13}C/^{15}N NOESY−HSQC as well as two-dimensional ^{1}H NOESY spectra, all recorded with a τ_{mix} = 75 ms. The assigned NOE restraints were classified into four distance ranges: 1.8−2.8, 1.8−3.5, 1.8−5.0, and 1.8−6.0 Å corresponding to strong, medium, weak, and very weak NOE intensities, respectively. NOE peak intensities were measured by volume integration. Pseudoatom corrections were added to the upper distance limits where appropriate (*20*). A 0.5 Å correction was applied to the upper bounds for NOEs involving methyl protons and non-stereospecifically assigned methylene protons. Torsion angle restraints were predicted using an in-house program called SHIFTOR (*21*). SHIFTOR calculates both ϕ and ψ torsion angles from observed ^{13}Cα, ^{13}Cβ, ^{1}Hα, and ^{1}HN chemical shifts and is available at www.redpoll.pharmacy.ualberta.ca. Additional ϕ torsion angle information was obtained from $^3J_{HNH\alpha}$ coupling constants measured from a 3D HNHA experiment. A total of 41 ϕ backbone torsion angle restraints were used. These were assigned an uncertainty of ±10° for residues in well-defined helical or β-sheet regions. Backbone ψ dihedral angle restraints were derived using SHIFTOR as well as from an analysis of $d_{N\alpha}/d_{\alpha N}$ ratios (*22*) and assigned an uncertainty of ±70°. Hydrogen bond restraints (d_{O-HN} = 1.6−2.4 Å, d_{O-N} = 2.6−3.4 Å) for slowly exchanging amide protons were identified from the pattern of sequential and interstrand NOEs involving ^{1}HN and C_αH protons and from the chemical shift index (*23*).

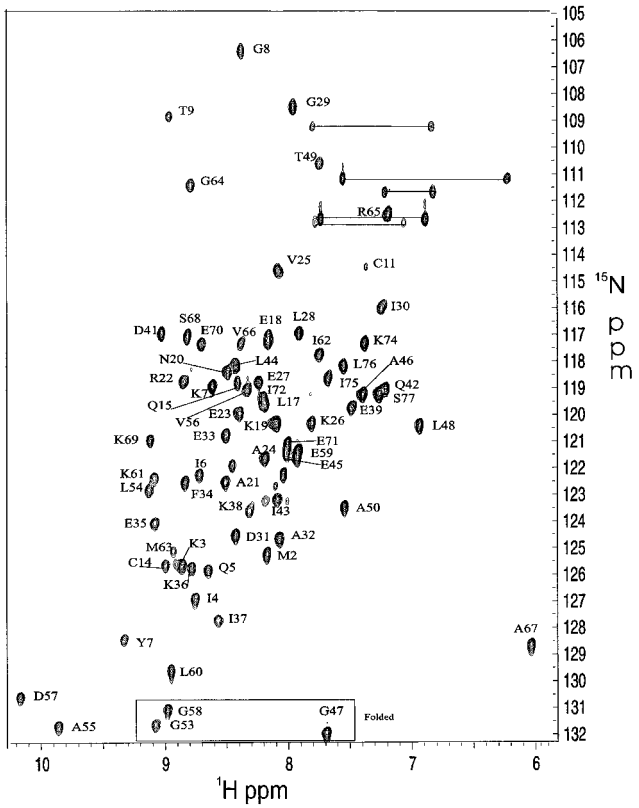

FIGURE 1: ^{1}H−^{15}N HSQC spectrum from 1 mM *MtH*895 in 25 mM KH_2PO_4, 200 mM NaCl, 0.5 mM EDTA, and 20 mM DTT at pH 7 collected at 25 °C on a Varian Unity 500 MHz spectrometer.

Structure Calculations. Structures for *MtH*895 were calculated using X-PLOR 3.851 (*24*). Only NOE-derived internuclear distance constraints were used in the first stage of structure generation. Initially, a set of 50 structures was generated using the ab initio simulated annealing protocol applied to a random coordinate set. These structures were then regularized using the simulated annealing protocol of Nilges et al. (*25*). After this initial step, several ambiguous long-range NOE assignments were clarified by analyzing the resulting 50 structures. Subsequently, the same simulated annealing protocol was repeated (three times) with additional and/or corrected NOE data. For the second stage of refinement, torsion angle restraints were introduced. The final set of 20 structures was selected on the basis that no interproton distance restraint violation could be greater than 0.5 Å and no torsion angle restraint violation could be greater than 5°.

A total of 862 NOE-derived distance restraints [237 long range, 323 medium and short range ($i - i + 1$, 2, 3, or 4), and 302 intraresidue], 102 dihedral angle restraints, and 46 hydrogen bond restraints were used to generate the final structural ensemble. The final set of 20 structures was analyzed with PROCHECK-NMR (*26*) and VADAR (*27*). MOLMOL (*28*) was used to visualize all the structures and to calculate RMSD values.

Modeling of the Interaction between MtH895 and T7 DNA Polymerase. Molecular dynamics simulations of the binding interaction between *MtH*895 and T7 DNA polymerase were carried out using several approaches. Coordinates from the lowest energy conformer of *MtH*895 along with coordinates from the X-ray crystal structure of T7 DNA polymerase bound to *E. coli* thioredoxin (PDB entry: 1T7P) were used in these docking simulations. The probable binding sites and key residues involved in the ligand−enzyme interaction were identified from a detailed analysis of the 1T7P structure. Three proteins, *MtH*895, *E. coli* thioredoxin (PDB entry: 1THO), and *E. coli* glutaredoxin (1GRX), were separately docked onto T7 DNA polymerase, both manually and automatically, using Swiss-PDB Viewer (*29*) and DOCKVISION (*30*), respectively. Different initial orientations of the ligands on T7 DNA polymerase were explored in order to obtain the best possible interactions of the complexes, which were then further optimized to achieve the best geometry. The docked structures were solvated with an average of 2000 simple point charge (*31*) water molecules and subjected to steepest descent and conjugate gradient energy minimization using the GROMACS molecular simulation package (*32*). Initially, a 10 ps MD run was performed with positional restraints followed by a 1000 ps MD run without restraints. Weak coupling of the protein to a solvent bath of constant temperature (300 K) and constant pressure (1 bar) was maintained with a coupling time of 1.0 ps. For all energy minimization and MD simulations GROMOS-43B1 (*33*) and/or GROMACS (ffgmx) (*32*) force fields were used. The 20 lowest energy conformers were selected and energy minimized further using a conjugate gradient algorithm. All models were analyzed using WHAT IF (*34*) and PROCHECK (*35*). Quantitative comparisons between the binding interactions of the three proteins to T7 DNA polymerase were carried out using the Structural Thermodynamics Calculator (STC) (*36*).

Measurement of Thiol Ionization by Ultraviolet Absorbance. The thiol ionization constants for *MtH*895 were measured using the protocol of Dyson et al. (*37*, *38*). The thiolate ion exhibits a stronger absorption at 240 nm (with an ϵ_{240} of about 4000 M^{-1} cm^{-1}) than the un-ionized thiol group. Therefore, the pK_a values of the thiol groups can be monitored by UV absorption during pH titration. Specifically, 0.5 mL of a 1 mM *MtH*895 sample was reduced using a 10× molar excess of DTT. Excess DTT was removed by dialysis against the same buffer under an argon atmosphere. After dialysis, the *MtH*895 sample was diluted to a concentration of ~30 μM in a 0.1 mM EDTA, pH 6.0, and 100 mM potassium phosphate buffer. Thiol ionization was monitored by measuring the protein absorbance at 240 nm on a Pharmacia Biotech Ultraspec 3000 UV/vis spectrophotometer. Small aliquots of 1 M NaOH or 2 M HCl were added to adjust the pH up or down. After each addition the change in protein concentration was noted at 280 nm. The concentration of *MtH*895 was calculated at 280 nm using an extinction coefficient of 1615 M^{-1} cm^{-1}.

Table 1: Structural Statistics of the 20 Lowest Energy Structures of *MtH*895

restraint type	no. of restraints
total unambiguous NOE distances	862
intraresidue	302
sequential	171
medium range	152
long range	237
hydrogen bond constraints	46
dihedral angles Φ_1	41
backbone stereochemistry	percent residues in
most favorable regions	81.1%
allowed regions	17.1%
generously allowed regions	1.4%
nonallowed regions	0.4%
structure ensemble	pairwise RMSD (Å)
backbone (all)	0.56
heavy atoms (all)	1.42

Measurement of Thioredoxin/Glutaredoxin Activity by Ultraviolet Absorbance. Thioredoxin/glutaredoxin activity measurements were performed using the methods described by Holmgren (*39*). To assay for thioredoxin activity, a solution containing 750 μL of 100 mM Tris-HCl (pH 8.0) with 1 mM EDTA, 0.35 mM NADPH, and either 0.02 mM *MtH*895 or *E. coli* thioredoxin (Sigma, Oakville, ON) was prepared. After the solution was allowed to equilibrate for 3 min, 20 μL of a 0.01 mM solution of *E. coli* thioredoxin reductase (Sigma) was added and the oxidation of NADPH monitored at 340 nm using a Pharmacia Biotech Ultraspec 3000 UV/vis spectrophotometer. After 15 min, DTNB (40 μL of a 10 mM solution) was added, and the change in absorbance at 412 nm was further monitored. To assay for glutaredoxin activity, a solution containing 750 μL of 100 mM Tris-HCl (pH 8.0) with 1 mM EDTA, 0.35 mM NADPH, 1 mM glutathione, and 0.75 nM yeast glutathione reductase (Sigma) was prepared. After 5 min of equilibration, a quantity (20 μL of 0.2 mM) of either *MtH*895, *E. coli* thioredoxin, or T4-glutaredoxin *(40)* was added to the reaction mixture and the change in absorbance monitored at 340 nm.

RESULTS

Solution Structure. Statistical parameters for the ensemble of 20 calculated structures are presented in Table 1. All structures (Figure 2A) exhibit good covalent geometry as indicated by low RMS deviations from idealized values and by low NOE, dihedral angle, and van der Waals energies. For all 20 structures, ~99.6% of the main chain (ϕ, ψ) angles fall in the core or allowed regions of the Ramachandran plot (Table 1) as determined using PROCHECK-NMR and VADAR. With the exception of the active site loop (residues 9−13) and the C terminus (residues 75−77), all portions of the *MtH*895 structure are well-defined by NMR standards. The complete set of 20 *MtH*895 structures has been deposited with the Protein Data Bank (PDB accession: 1ILO). The complete set of ^{1}H, ^{13}C, and ^{15}N shifts has been deposited with the BioMagResBank (BMRB accession number 4991).

*MtH*895 is composed of a four-stranded β-sheet sandwiched between two helices on one side and a third on the other (Figure 2B). The topological arrangement of the secondary structural elements (β-α-β-α-β-β-α), the small size

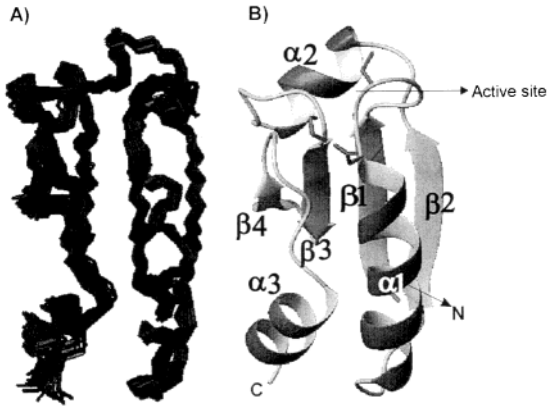

FIGURE 2: (A) Ensemble of 20 minimum energy structures of *MtH*895. (B) Ribbon diagram of the representative structure of *MtH*895. The active site residues have their side chains illustrated.

(<90 residues), and the overall 3D structure clearly indicate that *MtH*895 has a glutaredoxin-like fold *(41)*. Typically a glutaredoxin fold is similar to a thioredoxin-like fold, except the N-terminal residues (forming a β-α secondary structure) are truncated. The four β-strands in *MtH*895 include residues 2−7 (strand β1) and 32−37 (strand β2) which run parallel to each other and antiparallel to the β-hairpin formed by residues 53−56 (strand β3) and residues 59−62 (strand β4). The first helix (helix α1) is the longest helix and runs from residues 14 to 27. Helix α2, which is on the same side of the central β-sheet as helix α1, is composed of residues 40−46 while the third helix (helix α3), connecting strands β2 and β3, comprises residues 69−76. The axes of helices α1 and α3 run parallel to the β-sheet but in a direction opposite to the two parallel β-strands. The redox active site (Cys11-Ala12-Asn13-Cys14) is located just in front of the N-terminal side of the first helix. Like other thioredoxins and glutaredoxins (*41*, *42*) this redox active site is solvent accessible on one side of the molecule but solvent inaccessible on the other. Close inspection of the active site structure shows that the side chains of Ile6, Tyr7, Gln15, Met16, Leu17, and Val66 effectively block any approach to Cys14 (Figure 3). *MtH*895 also has an exposed hydrophobic surface on the opposite side of its redox active site containing Cys11, Ala12, Asn13, Met40, Thr49, Ala50, Leu51, Pro52, Val66, and Ala67. This surface is similar to that seen for other thioredoxins and glutaredoxins and is thought to have an important role in facilitating substrate binding and redox interactions (*43*, *44*). The relative positioning and general topology of this hydrophobic surface is presented in Figure 4A. Opposite to this hydrophobic face is a highly acidic surface (Figure 4B).

Structural Comparison with Other Thioredoxins and Glutaredoxins. A structural comparison was carried out between *MtH*895 and six other representatives of the thioredoxin/glutaredoxin superfamily: (1) T4-glutaredoxin, (2) *E. coli* thioredoxin (Trx), (3) *E. coli* glutaredoxin (Grx), (4) human glutaredoxin (5) *Pyrococcus furiosus* protein disulfide oxidoreductase subunits N and C, and (6) thioredoxin-2 from *Anabaena* sp. This comparison revealed obvious similarities among all seven structures. Figure 5 shows the structure-based sequence alignment, in which the secondary structural elements belonging to the common glutaredoxin/thioredoxin fold are numbered. For all seven structures, the four β-strands forming the core β-sheet are well conserved. Overall, the secondary structures superimpose well with most insertions occurring at the N-termini or loop regions.

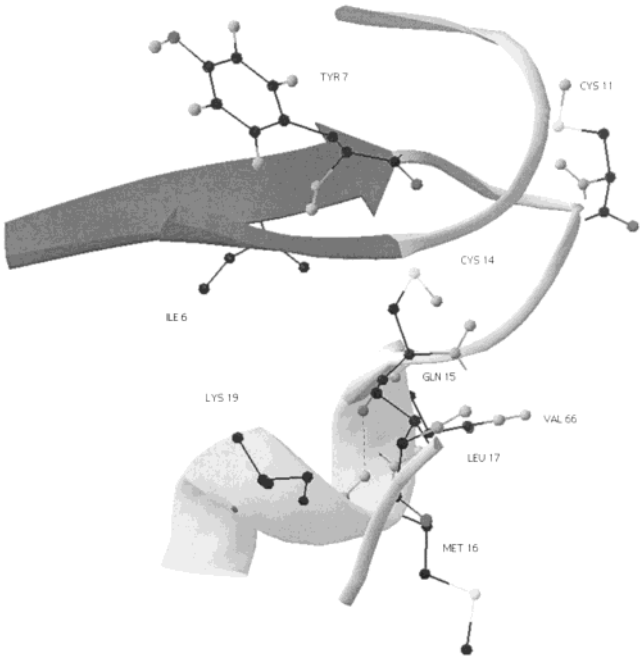

FIGURE 3: Detailed view of the active site region of MtH895. The accessible surface area of Cys14 is 2.9 Å^2, and the accessible surface area of Cys11 is 116 Å^2.

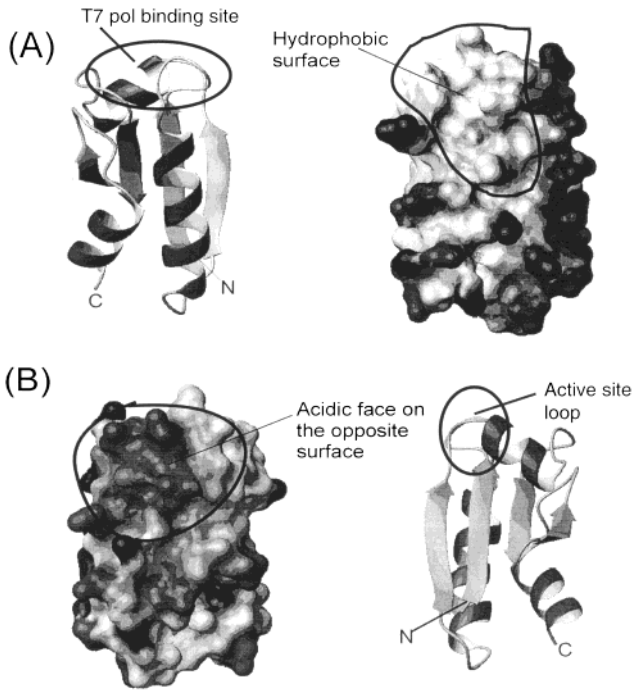

FIGURE 4: Electrostatic potential surface of *MtH*895 showing the exposed hydrophobic surface on one side (A) and an acidic surface on the opposite face (B).

Structures similar to *MtH*895 were also identified by submitting a representative *MtH*895 structure to be searched against the Dali (*45*, *46*), SCOP (*47*), and CE (*48*) databases. Table 2 lists the top hits from the Dali database search. These same hits were also among the top hits in the CE database search. The structures were assessed by their *Z* scores, number of matching residues, and RMSD values. Interest-

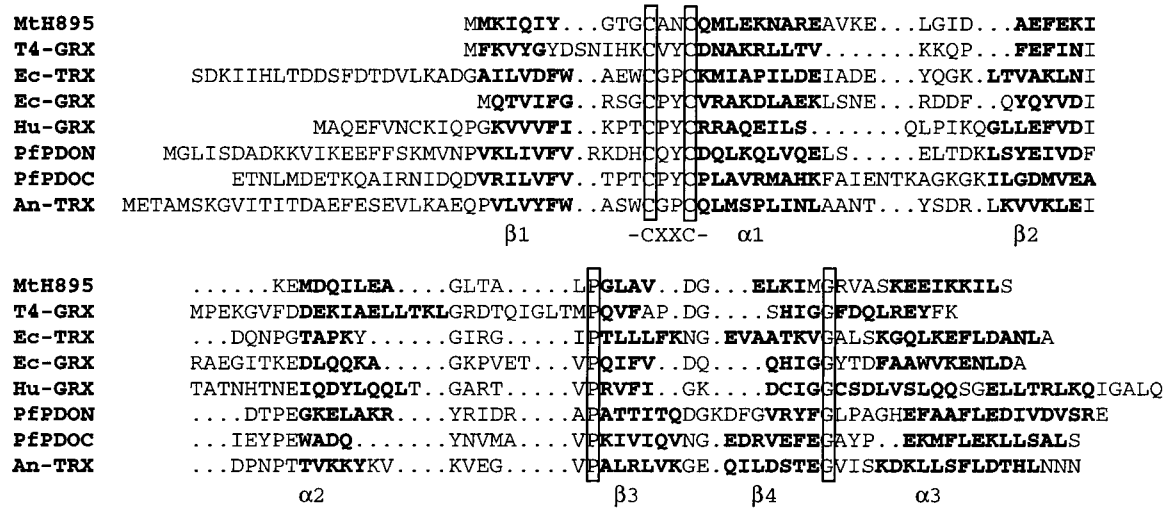

FIGURE 5: Structure-based sequence alignment of *MtH*895, T4-glutaredoxin (T4-Grx), *E. coli* thioredoxin (Ec-Trx), *E. coli* glutaredoxin (Ec-Grx), human glutaredoxin (Hu-Grx), *P. furiosus* protein disulfide oxidoreductase subunits N (PfPDON) and C (PfPDOC), and *Anabaena* thioredoxin II (An-Trx). The α-helices and *β*-strands are marked in bold. The 100% conserved residues are enclosed in boxes. The thioredoxins contain additional secondary elements in the N-terminal region, which are not shown here.

Table 2: Protein Structures Similar to *MtH*895 Derived from a Dali Database Search

proteins	PDB code	*Z* score	RMSD (Å)	alignment length	sequence length	% identity
thioredoxin 2 (*Anabaena* sp.)	1THX	5.3	2.6	72	108	13
thioredoxin *Mj*307 (*M. jannaschii*)	1A8L	5.3	2.6	68	105	19
alkyl peroxide reductase, sub F (*S. typhimurium*)	1HYU	5.2	2.6	74	521	15
calsequestrin[a] (rabbit)	1A8Y	5.1	2.9	69	338	17
glutaredoxin-1 (*E. coli*)	1QFN	4.8	3.3	69	85	14
glutaredoxin-like NrdH (*E. coli*)	1H75	4.8	3.2	64	76	20
glutaredoxin-2 (*E. coli*)	1G7O	4.7	2.9	61	215	13

[a] Has three thioredoxin-like domains.

ingly, the top hits from all three databases belong to known thioredoxins or to proteins with more than one thioredoxin domain. Notably glutaredoxins were largely absent or at the bottom these lists. A comparison between *MtH*895, a structurally related thioredoxin (1THX), and a structurally related glutaredoxin (1QFN) is shown in Figure 6 to illustrate the similarities and differences.

Primary Sequence Analysis. When *MtH*895 was first cloned in early 1999, a detailed assessment using BLAST (*6*), PROSITE (*49*), and other database searches indicated no detectable similarity to any known structure or to any protein with a known function. These negative results led us to undertake the structural studies described here. After structural analysis by NMR (completed in late 2000) revealed the existence of a glutaredoxin-like fold, we conducted a more detailed sequence analysis in March 2001 using PSI-BLAST (*6*). With the addition of three new *MtH*895 homologues since 1999 and the doubling in database size, this later database search clearly revealed that *MtH*895 belongs to the thioredoxin superfamily. After seven PSI-BLAST iterations a list of approximately 150 proteins could be identified as *MtH*895 orthologues, with the vast majority being bacterial or archaebacterial thioredoxins/glutaredoxins. However, the level of sequence identity of *MtH*895 to any known member of the thioredoxin superfamily rarely exceeded 20%.

Interestingly, *MtH*895 exhibits a significant level of sequence identity (34−44%) to a group of previously unidentified proteins from several archaea including *Methanococcus jannaschii*, *Thermotoga maritima*, and the cyanobacterium *Anabaena* sp. (Figure 7). It is worth noting that these primitive organisms already have previously identified thioredoxins/glutaredoxins, but the level of sequence identity of their known thioredoxins/glutaredoxins to any of the members of this new group of proteins does not exceed 28%. For example, *MtH*895 has 28% sequence identity to the putative *Methanobacter* thioredoxin, *MtH*807. In *M. jannaschii*, the conserved hypothetical protein, *Mj*0581, has 42% sequence identity to *MtH*895, as opposed to 21% sequence identity to its recently characterized thioredoxin-like paralogue *Mj*0307 (*50*).

Interaction between MtH895 and T7 DNA Polymerase. Glutaredoxins can be distinguished from thioredoxins on the basis of their interaction (or lack thereof) with T7 DNA polymerase. To assess if *MtH*895 could favorably interact with T7 DNA polymerase (pol), we carried out molecular docking studies between T7 DNA polymerase and *MtH*895, *E. coli* thioredoxin (*51*), and *E. coli* glutaredoxin. The binding interactions of each complex were quantitatively evaluated by calculating accessible surface area and thermodynamic parameters, as well as association and dissociation constants (K_a and K_d, respectively) using the STC program (*36*). The results are presented in Table 3. The calculated free energy change on binding between *MtH*895 and T7 DNA pol is comparable to that between *E. coli* Trx and T7 DNA pol. On the other hand, the predicted binding energy between *E. coli* Grx and T7 DNA pol is approximately five times weaker. Similarly, the STC calculated K_a is only on the order of 10^3 for the T7 DNA pol−*E. coli* Grx interaction whereas it is on the order of 10^{17} and 10^{20} for the T7 DNA pol−*E. coli* Trx and T7 DNA pol−*MtH*895 interactions, respectively.

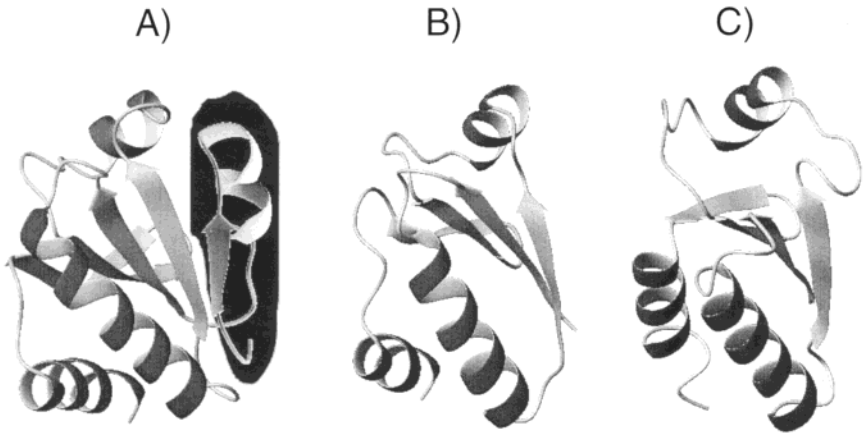

FIGURE 6: (A) Ribbon diagram of thioredoxin-2 (*Anabaena sp.*). (B) Ribbon diagram of *MtH*895 (*M. thermoautotrophicum*). (C) Ribbon diagram of glutaredoxin-1 (*E. coli*). All three molecules have been oriented to show maximal similarity in their overall folds. The N terminus (~30 residues) of thioredoxin-2 has been shaded.

```
MtH895        MMKIQIYGTGCANCQMLEKNAREAVKELGIDAEFEKIKE
Nostoc      MNAIKIEILGTGCKKCQQLEANAKEAVTNLNLIAEVSHITD
Mj0581    MVRVMVVIRIFGTGCPKCNQTYENVKKAVEELGIDAEIVKVTD
Tm_CHP      MAKKVEILGKGCPRCKQTEKIVRMAIEELGIDAVVEKVQD
Af_CHP        MKIKVVGPGCARCKATFDVVKKVVEKEGLDVELEYVTD
Pa0941      MTKAIYYHAGCAICVEAERSLLPLLDRKQVDIEVVHLAEQSAR
Conserved      K     G GC   C           V   G DAE     D

MtH895     MDQILEAGLTALPGLAVDGELKIMGRVASKEEIKKILS
Nostoc     PIEIAKRGVMSTPAMAINGKVVSKGQVISTEQIQPLLQR
Mj0581     VNEIAE WVFVTPGVAFDDVIVFEGKIPSVEEIKEELKSYLEGK
Tm_CHP     INEIVSRGVVATPAVAVDGKVVISGKIPSLDEVKKVLQQA
Af_CHP     MNEAIELGVVATPAVWVDGKVVIQGKIPKESEILEIIKK
Pa0941     IAEAEKAGVKSVPALVVDGQVLHLNFGAALSDLK
Conserved  E     GV   TP    VDG VV  G    S   E
```

FIGURE 7: Primary sequence alignment of the group of previously unidentified, highly conserved proteins from *M. thermoautotrophicum* (*MtH*895), *Nostoc* sp., *M. jannaschii* (*Mj*0581), *T. maritima* (Tm_CHP), *Archeoglobus fulgidus* (Af_CHP), and *Pseudomonas aeruginosa* (Pa0941). CHP stands for conserved hypothetical proteins. The active site -CXXC- is enclosed in a box.

Table 3: Comparative Results of Binding Interactions of T7 DNA Polymerase (pol) with *E. coli* Thioredoxin (Trx), *MtH*895, and *E. coli* Glutaredoxin (Grx)

structure	ΔH (kcal/mol)	$T\Delta S$ (kcal/mol)[e]	ΔG (kcal/mol)	K_a (M)	K_d (M)
T7pol + EcTrx (X-ray)[a]	3.74	27.53	−23.79	3×10^{17}	3×10^{-18}
T7pol + EcTrx[b]	3.99	28.25	−24.26	7×10^{17}	1×10^{-18}
T7pol + *MtH*895[c]	8.20	35.79	−27.62	2×10^{20}	5×10^{-21}
T7pol + EcGrx[d]	−1.48	3.81	−5.29	7×10^{3}	1×10^{-4}

[a] X-ray crystal structure of T7 DNA pol bound to *E. coli* Trx (PDB 1T7P). [b] Energy-minimized structure of T7 DNA pol bound to *E. coli* Trx. [c] Energy-minimized structure of T7 DNA pol bound to *MtH*895. [d] Energy-minimized structure of T7 DNA pol bound to *E. coli* Grx.[e] T = 300 K.

This result shows that MtH895, on the basis of its predicted ability to bind with T7 DNA pol, resembles a thoiredoxin much more than a glutaredoxin. Figure 8 shows the interactions that have been identified between *MtH*895 and T7 DNA pol from the majority of the conformers analyzed via our molecular dynamics and molecular docking simulations.

Thiol Ionization by UV Spectroscopy. The stability of the active site disulfide bond is known to vary widely among the different members of the thioredoxin and glutaredoxin family (*52*). This variability has been shown to bear direct correlation with the pK_a value of the most accessible of the two cysteine thiols. The N-terminal cysteine of the active site -CXXC- motif is known to be solvent exposed and to have higher reactivity and a lower pK_a than the C-terminal cysteine. Normally, in glutaredoxins, the thiol pK_a value of the more exposed cysteine is less than 5 whereas in thioredoxins it is mostly above 6.5 (*52*). On the other hand, the pK_a value of the more buried cysteine is usually very high (above 9.0) in both thioredoxins and glutaredoxins (*53*−*57*). In an effort to determine whether *MtH*895 was more similar to a glutaredoxin or a thioredoxin, we determined the pK_a values of its two thiol groups using UV spectroscopy. On titration, the ϵ_{240} increased by a little over 8000 M^{-1} cm^{-1}, indicating that both thiol groups were fully ionized (Figure 9). Interestingly, both thiol groups titrated at the same pH, giving a pK_a of ~6.7 (as determined by the Henderson−Hasselbach equation). These pK_as are more typical of a thioredoxin than a glutaredoxin.

Thioredoxin/Glutaredoxin Activity. The strong structural evidence suggesting that *MtH*895 was a thioredoxin and not

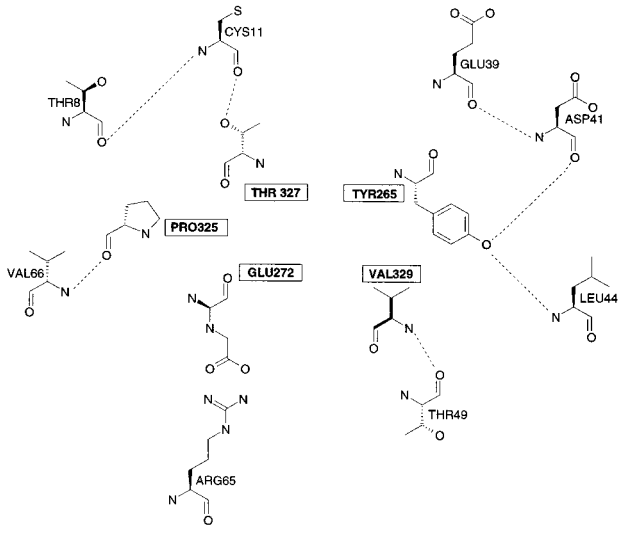

FIGURE 8: Schematic representation of the interaction between T7 DNA polymerase and *MtH*895. Only a few interactions are shown here. Dashed lines represent H-bonds. Residues Cys11, Thr8, Val66, Arg65, Thr49, Leu44, Asp41, and Glu39 belong to *MtH*895 while Thr327, Tyr265, Val329, Glu272, and Pro325 belong to T7 DNA polymerase.

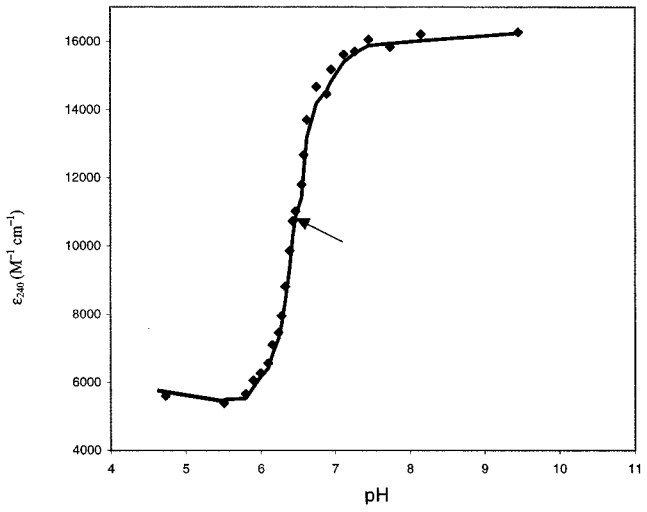

FIGURE 9: Measurement of thiol ionization in *MtH*895 by ultraviolet absorbance.

a glutaredoxin was further confirmed with the results of two standard spectrophotometric assays for glutaredoxin and thioredoxin activity. As seen in Figure 10A, *MtH*895 shows a growing, time-dependent DTNB absorbance at 412 nm which exactly parallels that of *E. coli* thioredoxin. While not quite as active as *E. coli* thioredoxin, *MtH*895 clearly exhibits thioredoxin activity. In contrast, Figure 10B shows how both *E. coli* thioredoxin and *MtH*895 exhibit no detectable glutaredoxin activity whereas T4-glutaredoxin, which was used as a positive control, shows the expected strong NADP absorbance at 340 nm, shortly after it is added to the glutathione/glutathione reductase reaction mixture. Both sets of experiments were repeated multiple times, with little variability in the results.

DISCUSSION

Thioredoxin or Glutaredoxin? The structure of *MtH*895 was solved as part of a pilot structural proteomic project

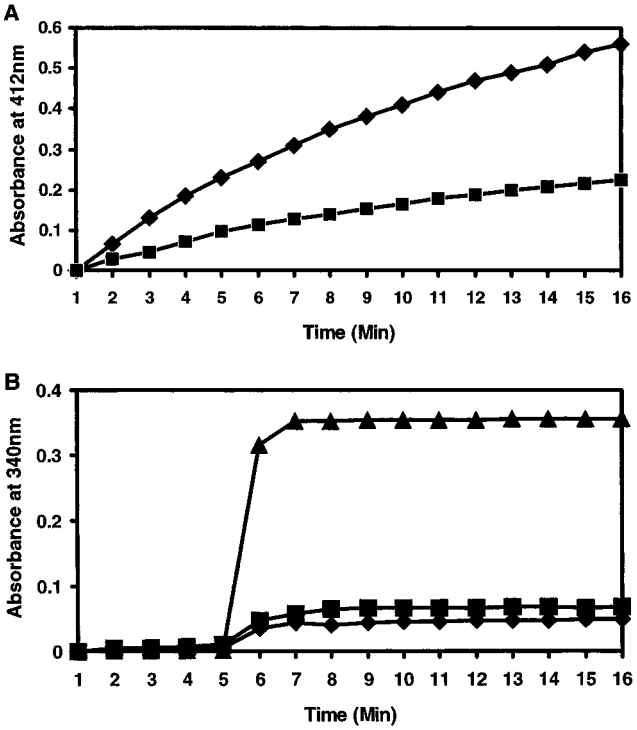

FIGURE 10: Measurement of thioredoxin and glutaredoxin activity via ultraviolet absorbance. (A) Thioredoxin activity kinetics (measured at 412 nm) plotted for *MtH*895 (■) and *E. coli* thioredoxin (◆). (B) Glutaredoxin activity kinetics (measured at 340 nm) for T4-glutaredoxin (▲), *E. coli* thioredoxin (◆), and *MtH*895 (■).

initiated by the Ontario Cancer Institute (*1*). It was aimed at determining the feasibility of using NMR to solve the structures of a large number of proteins and to assess the extent to which these structures could provide insights into their function. Before the structure of *MtH*895 was solved, this protein had no known sequence homologue with either an assigned function or a known three-dimensional structure. After the solution structure was determined, it became obvious that *MtH*895 exhibited a glutaredoxin-like fold. However, the question quickly arose, is it really a glutaredoxin? Glutaredoxins are small, ubiquitous proteins that are important for redox regulation of protein function and signaling. They accept protons from glutathione and transfer those protons to various protein substrates. Our suspicions were raised about the true function of this protein when a literature search revealed that *M. thermoautorophicum* does not contain glutathione or a glutathione-like cystolic thiol (*44*). Glutathione is essential for glutaredoxin function. Furthermore, all previous studies of putative archebacterial glutaredoxins concluded that these putative glutaredoxins functioned essentially as thioredoxins rather than glutaredoxins. That is, they appeared to exhibit no activity as glutathione mixed disulfide reductants (*44*, *50*). To confirm our suspicions, we undertook a detailed investigation to determine the true nature of *MtH*895 through (1) sequence comparison, (2) structural comparison, (3) structural analysis, and (4) biochemical analysis.

Sequence Comparison. As indicated previously, there is little similarity in the primary sequences between *MtH*895 and the members of the thioredoxin/glutaredoxin superfamily. Among those exhibiting the highest level of sequence identity was a group of previously unidentified or putative arche-

bacterial thioredoxins. Sequence comparisons between *MtH*895 and "standard" thioredoxins indicate that only a few residues are fully conserved. These include the two active site cysteines, a cis-proline at the loop preceding β3, a glycine residue between β3 and β4, and a second glycine immediately following β4. As a rule, the active site sequence -Trp-Cys-Gly-Pro-Cys- is highly conserved in thioredoxins whereas glutaredoxins have the conserved -Cys-Pro-Tyr-Cys in all but two members (T4-glutaredoxin has a Val in place of Pro, and pig glutaredoxin has a Phe in place of Tyr). *MtH*895, however, has a unique -Cys-Ala-Asn-Cys- motif which is not found in any known member of the thioredoxin or glutaredoxin superfamily. We believe the absence of a Pro residue in the *MtH*895 active site may confer some conformational flexibility and therefore may have some effect on the stability of the redox reaction process in this enzyme. This is supported by some recent work by Ren et al. (*58*), who have looked at the crystal structure and thermal *B* factors of an archael disulfide oxidoreductase with an unusual active site motif. Interestingly, all members of this putative group of archaebacterial thioredoxins have unique -CXXC- motifs distinctly different from either thioredoxin or glutaredoxin active site motifs (Figure 7). Evidently, archaebacteria appear to have relaxed their requirement for an absolutely conserved sequence -Trp-Cys-Gly-Pro-Cys- active site for redox function.

Structure Comparison. While direct sequence comparisons were not generally able to distinguish between the two possible functions for *MtH*895, structural comparisons were much clearer. As a general rule, significant differences exist between thioredoxins and glutaredoxins at their N-termini where the former typically has additional 15−20 residues. These extra residues form well-defined secondary structures that lay on either side of the central β-sheet core. However, these extra residues have not yet been implicated as having any major role in the redox functionality of these proteins. *MtH*895 does not contain these additional N-terminal secondary structural elements, suggesting that it is superficially more like a glutaredoxin. However, when the structure of *MtH*895 was submitted to the Dali, SCOP, and CE fold identification servers, the top hits for all three database comparisons were members of the thioredoxin family, not the glutaredoxin family. These results strongly suggested that even though *MtH*895 has a glutaredoxin fold, it is more similar to the thioredoxins in structure.

Structure Analysis. A more detailed structural analysis of the active site region of *MtH*895 revealed some important functional information as well. Most glutaredoxins have a highly conserved positively charged residue (Lys or Arg at position 8) in their protein−protein interaction site followed by a positively charged residue on the C-terminal side of their -CXXC- active site (*43*). An aspartic acid at the N-terminus of the third helix is also found in most glutaredoxins. These charged residues, which all lay near the active site, have been implicated in the participation of specific ionic interactions with glutathione (*59*). In contrast to glutaredoxins, most thioredoxins have neutral or hydrophobic residues in these sites. This difference arises because the interacting substrate for thioredoxins is a protein (thioredoxin reductase) rather than a small molecule like glutathione (for glutaredoxins). Interestingly, *MtH*895 does not have any of the characteristic charged residues of glutaredoxins, but rather it has mostly neutral or hydrophobic residues in these sites. This strongly suggests that *MtH*895 cannot possibly interact with glutathione, and therefore, by elimination, it must function as a thioredoxin.

Interaction with T7 DNA Polymerase. A key distinguishing feature between thioredoxins and glutaredoxins is their differential ability to bind T7 DNA polymerase. Most prokaryotic thioredoxins are known to interact with T7 DNA polymerase including *E. coli* Trx (*51*) and thioredoxins from *M. jannaschii* (*50*), *Thiobacillus ferroxidans* (*60*), *Corynebacterium nephridii* (*61*), and *Rhodobacter sphaeroides* (*62*). On the other hand, neither *E. coli* glutaredoxin nor T4-glutaredoxin (*63*) supports T7 phage growth as glutaredoxins appear to be incapable of binding T7 DNA polymerase.

Figure 8 shows the interactions that have been predicted to occur between *MtH*895 and T7 DNA pol. Both the location and the predicted interactions are very similar to those seen in T7 DNA pol bound to *E. coli* Trx (*51*). In both *MtH*895 and *E. coli* Trx, the site of interaction is located on the surface near the vicinity of the active site. Several amino acid residues in the loops containing the -CXXC- motif and those between β2 and α2, α2 and β3, and β4 and α3 (using the numbering scheme presented in Figure 5) are involved in binding. The CO group of Pro325 in T7 DNA pol is predicted to form a hydrogen bond with the NH of Val66 in *MtH*895, similar to the hydrogen bond it forms with the Ala93 NH group located in an equivalent position in *E. coli* Trx. The Val329 NH group of T7 DNA pol forms a hydrogen bond with the Arg73 CO group in *E. coli* Trx. A similar hydrogen bond is also predicted to form between the Val329 NH group and Thr49 CO of *MtH*895, whose sequential position is identical to that of Arg73 in *E. coli* Trx. The second helix of *MtH*895 is predicted to participate in several important interactions with T7 DNA pol, similar to those seen in the *E. coli* protein. The backbone amide and carbonyl groups of the active site cysteines in both proteins are predicted to form several inter- and intramolecular hydrogen bonds. Interestingly, N-terminal active site cysteines in both complexes face outside and are more accessible compared to the C-terminal cysteines which seem to be rigidly trapped in a cavity. Nearly 65% of the residues in T7 DNA pol that are predicted to interact with *MtH*895 can be seen to interact with *E. coli* Trx in the X-ray crystal structure (*51*). The small difference seen in the interactions is probably due to the flexibility of the extended loop between helices H and H1 of T7 DNA polymerase. Important charge−charge interactions predicted in a majority of the conformers involve Arg65 (of *MtH*895) interacting with Glu272, Tyr326, Lys290, and Tyr320 of T7 polymerase. Similar interactions are also predicted between Lys61 and Asn13 of *MtH*895 with Glu259, Trp264, Glu330, Tyr286, and Thr327, respectively, of the polymerase. The corresponding residues, Lys90 and Pro34 in *E. coli* Trx, participate in similar interactions with the T7 DNA polymerase. This high degree of structural equivalence along with the very favorable binding energies calculated via STC strongly suggests that *MtH*895 functions as a thioredoxin rather than a glutaredoxin.

Thiol Ionization. Our measurements indicate that *MtH*895's active site thiols have a pK_a around 6.7, which is more similar to the pK_a value of the exposed cysteine in thioredoxins than glutaredoxins (*57*, *64*). However, the pK_a value of the buried

cysteine is much higher in both of these classes of proteins than in *MtH*895. Interestingly, two other thioredoxin-like proteins appear to exhibit thiol pK_a values similar to those observed for *MtH*895. A thioredoxin-like protein (*Mj*0307) from a related archaeon, *M. jannaschii*, has a single pK_a value of about 6.28 for both of its thiols (*50*). In addition, an *E. coli* thioredoxin D26A mutant shows simultaneous titration of both of its thiols at pH 7.5 (*65*). In wild-type *E. coli* thioredoxin, the N- and C-terminal active site cysteines have pK_a values of 7.1 and 7.9 respectively (*66*). The partial positive charge induced by a helix dipole from helix II has been implicated in lowering the pK_a value of the N-terminal active site cysteine in *E. coli* Trx. In the case of *MtH*895, the active site cysteines are in the same secondary structural elements as in *E. coli* Trx and therefore would be expected to experience a similar helix dipole effect. However, in *E.coli* Trx the presence of Asp26 has been suggested to play a significant role in lowering the pK_a value of the N-terminal active site cysteine compared to the other cysteine (*65*). *MtH*895 differs from *E. coli* Trx because it lacks a negatively charged residue at the equivalent position that could influence the N-terminal cysteine. The lack of a nearby negatively charged residue has also been suggested to be the reason for the simultaneous titration of the active site thiols in *M. jannaschii* thioredoxin (*50*). Other factors, such as the two nonstandard amino acids between the *MtH*895 active site cysteines, may also influence their pK_a values. For example, Grauschopf et al. (*67*) have shown that mutation of the active site dipeptide Pro-His to Pro-Pro in DsbA (which is also a member of the thioredoxin superfamily) significantly changed the pK_a value of Cys30. Wang (*40*) suggested that the scarcity of positive electrostatic potential surrounding the active site thiols may be the reason for the unusually high pK_a of Cys14 in T4-glutaredoxin as compared to the other glutaredoxins. Indeed, *MtH*895 also has very few positively charged residues in and around the active site (Figures 4, 5, and 7), which may be the reason for its relatively high pK_a value.

Glutaredoxin/Thioredoxin Activity. While all other indications (structural, sequential, etc.) provided strong evidence that *MtH*895 was most likely a thioredoxin, the results from the thioredoxin activity assays clearly show that *MtH*895 is a thioredoxin. As seen in Figure 10A, *MtH*895 appears to be about one-half as active as *E. coli* thioredoxin. This reduced activity might be expected, given its short length (77 vs 108 residues), abbreviated structure (missing N-terminal β-α segment), and nonstandard active site sequence (Cys-Ala-Asn-Cys vs Cys-Gly-Pro-Cys). Alternately, the lower activity may have arisen simply because the enzymatic reducing agent (thioredoxin reductase) was obtained from *E. coli* and not from *M. thermoautotrophicum*. Nevertheless, the fact that such a small glutaredoxin-like protein clearly exhibits thioredoxin activity suggests that *MtH*895 and its homologues may represent a group of ancient proteins that were ancestral to both thioredoxins and glutaredoxins. One branch appeared to lengthen to give rise to eubacterial and eukaryotic thioredoxins while the other branch changed its active site, but maintained its size, to give rise to eubacterial and eukaryotic glutaredoxins. While this may be hard to prove, the considerable variablility or plasticity seen in the active site sequences of these homologues (Figure 7) suggests a kind of "founder effect" typical of very ancient sequences.

Thermal Stability of MtH895. One of the reasons for working with proteins isolated from *M. thermoautotrophicum* was to attempt to determine the underlying causes of thermostability in proteins. A number of structural and sequence-specific features have previously been suggested to confer high thermal stability to proteins in thermophilic organisms. For instance, the presence of an increased number of surface ion pairs, an increased number of hydrogen bonds, a higher proportion of aliphatic amino acids, and a larger polar surface area (for increased hydrogen-bonding density with water) have all been suggested as underlying reasons for increased thermal stability in proteins (*68*). In *MtH*895, we have identified four surface ion pairs [according to the definition of Vogt (*69*)]: Lys3 to Glu33, Lys3 to Asp57, Lys26 to Asp31, and Lys74 to Glu70. However, it is unclear whether these are in any higher proportion than what might be expected for a mesophilic protein of similar size. Assessments of the amino acid composition of *MtH*895 suggest that it has a higher than average proportion of aliphatic residues (Ala, Met, Ile, Leu) relative to most mesophilic proteins. The number of hydrogen bonds and proportion of polar surface area for *MtH*895 also appear to be higher than normal. Recently, Thompson and co-workers (*70*) noted that thermophilic proteins are generally smaller in size and more compact in structure. They are also more likely than their mesophilic counterparts to have deletions in exposed loop regions. Given that most thioredoxins and thioredoxin domains are all more than 100 residues in length, *MtH*895, at just 77 residues, is by far the smallest known thioredoxin. Most of the deletions in *MtH*895 appear to occur at the N-terminus and within its loop regions (Figure 5). Overall, it appears that many factors incrementally contribute to the thermostability of *MtH*895. However, given its unusually short length and compact size, we suspect structural compactness may be the most important contributor to the high thermal stability of this particular protein.

CONCLUSION

This project was undertaken to address a key question in structural proteomics: that is, can we determine the function of a protein through its structure? In attempting to answer this question, we initially restricted ourselves to using only the *MtH*895 3D structure and readily available docking or 3D database tools to ascertain its function. These structural analyses clearly indicated that *MtH*895 was a member of the glutaredoxin/thioredoxin family and that it was most likely a thioredoxin. Only after these initial hypotheses were generated did we attempt to confirm them using sequence comparisons (PSI-BLAST), literature surveys, and biochemical methods (thiol titrations, thioredoxin/glutaredoxin activity assays). These latter experiments confirmed and helped to rationalize what we had predicted through our earlier structural analyses. While we were fortunate to have worked with a protein that had at least some resemblance to a known family of proteins, it is still important to remember that this resemblance was only ascertained after the structure was determined, not before. Furthermore, even though a functional assignment was made, we still do not know the exact metabolic role of *MtH*895 in *M. thermoautotrophicum*. Despite this caveat, we believe this work reaffirms the potential for structural proteomics to be used in protein

functional assignment. However, it is not clear at this time whether an NMR-based route or an X-ray-based route may be the most efficient approach for functional assignment. Bottlenecks in both rapid assignment and rapid structure generation, as well as in high-volume protein preparation, still persist, and these problems certainly need to be addressed before structural proteomics can become a reality.

ACKNOWLEDGMENT

We thank B. D. Sykes for advice and assistance with the early phases of the NMR work and D. Waldman for assistance in proofing the manuscript.

REFERENCES

1. Christendat, D., Yee, A., Dharamsi, A., Kluger, Y., Savchenko, A., Cort, J. R., Booth, V., Mackereth, C. D., Saridakis, V., Ekiel, I., Kozlov, G., Maxwell, K. L., Wu, N., McIntosh, L. P., Gehring, K., Kennedy, M. A., Davidson, A. R., Pai1, E. F., Gerstein, M., Edwards, A. M., and Arrowsmith, C. H. (2000) *Nat. Struct. Biol. 7*, 903−909.
2. Skolnick, J., Feltrow, S. J., and Kolinski, A. (2000) *Nat. Biotechnol. 18*, 283−287.
3. Brenner, S. E., and Levitt, M. (2000) *Protein Sci. 9*, 197−200.
4. Sali, A. (1998) *Nat. Struct. Biol. 5*, 1029−1032.
5. Smith, D. R., Doucette-Stamm, L. A., Deloughery, C., Lee, H., Dubois, J., Aldredge, T., Bashirzadeh, R., Blakely, D., Cook, R., Gilbert, K., Harrison, D., Hoang, L., Keagle, P., Lumm, W., Pothier, B., Qiu, D., Spadafora, R., Vicaire, R., Wang, Y., Wierzbowski, J., Gibson, R., Jiwani, N., Caruso, A., Bush, D., and Reeve, J. N. (1997) *J. Bacteriol. 179*, 7135−7155.
6. Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) *Nucleic Acids Res. 25*, 3389−3402.
7. Yee, A., Booth, V., Dharamsi, A., Engel, A., Edwards, A. M., and Arrowsmith, C. H. (2000) *Proc. Natl. Acad. Sci. U.S.A. 97*, 6311−6315.
8. Wishart, D. S., Bigam, C. G., Yao, J., Abilgaard, F., Dyson, H. J., Oldfield, E., Markley, J. L., and Sykes, B. D. (1995) *J. Biomol. NMR 6*, 135−140.
9. Jeener, J., Meier, B. H., Bachmann, P., and Ernst, R. R. (1979) *J. Chem. Phys. 71*, 4546−4553.
10. Kay, L. E., Keifer, P., and Saarinen, T. (1992) *J. Am. Chem. Soc. 114*, 10663−10665.
11. Zhang, O., Kay, L. E., Olivier, J. P., and Forman-Kay, J. D. (1994) *J. Biomol. NMR 4*, 845−858.
12. Kuboniwa, H., Grzesiek, S., Delaglio, F., and Bax, A. (1994) *J. Biomol. NMR 4*, 871−878.
13. Pascal, S., Muhandiram, T., Yamazaki, T., Forman-Kay, J. D., and Kay, L. E. (1994) *J. Magn. Reson. 101*, 197−201.
14. Kay, L. E., Xu, G. Y., and Yamazaki, T. (1994) *J. Magn. Reson., Ser. A 109*, 129−133.
15. Grzesiek, S., and Bax, A. (1992) *J. Am. Chem. Soc. 114*, 6291−6293.
16. Kay, L. E., Xu, G. Y., Singer, A. U., Muhandiram, D. R., and Forman-Kay, J. (1993) *J. Magn. Reson. Ser. B 101*, 133−136.
17. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) *J. Biomol. NMR 66*, 277−293.
18. Fairbrother, W. J., Champe, M. A., Christinger, H. W., Keyt, B. A., and Starovasnik, M. A. (1997) *Protein Sci. 6*, 2250−2260.
19. Clore, G. M., Bax, A., and Gronenborn, A. M. (1991) *J. Biomol. NMR 1*, 13−22.
20. Wüthrich, K., Billeter, M., and Braun, W. (1983) *J. Mol. Biol. 169*, 949−961.
21. Wishart, D. S., and Case, D. A. (2001) *Methods Enzymol. 338*, 3−34.
22. Gagne, S. M., Tsuda, S., Li, M. X., Chandra, M., Smillie, L. B., and Sykes, B. D. (1994) *Protein Sci. 3*, 1961−1974.
23. Wishart, D. S, and Sykes, B. D. (1994) *Methods Enzymol. 239*, 363−392.
24. Brünger, A. T. (1993) *X-PLOR Manual*, version 3.1, Yale University, New Haven, CT.
25. Nilges, M., Clore, G. M., and Gronenborn, A. M. (1988) *FEBS Lett. 229*, 317−324.
26. Laskowski, R. A., Rullmann, J. A. C., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996) *J. Biomol. NMR 8*, 477−486.
27. Wishart, D. S., Willard, L., and Sykes, B. D. (1995) University of Alberta (http://redpoll.pharmacy.ualberta.ca).
28. Koradi, R., Billeter, M., and Wüthrich, K. (1996) *J. Mol. Graphics 14*, 51−55.
29. Guex, N., and Peitsch, M. C. (1997) *Electrophoresis 18*, 2714−2723.
30. Hart, T. N., Ness, S. R., and Reid, R. J. (1997) *Proteins, Suppl. 1*, 205−209.
31. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., and Hermans, J. (1981) Interaction models for water in relation to protein hydration, in *Intermolecular forces* (Pullman, B., Ed.) pp 331−342, D. Reidel Publishing Co., Dordrecht, The Netherlands.
32. van Gunsteren, W. F., Daura, X., and Mark, A. E. (1998) *Encycl. Comput. Chem. 2*, 1211−1216.
33. Berendsen, H. J. C., van der Spoel, D., and van Drunen, R. (1995) *Comput. Phys. Commun. 91*, 43−56.
34. Vriend, G. (1990) *J. Mol. Graphics 8*, 52−56.
35. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Crystallogr. 26*, 283−291.
36. Lavigne, P., Bagu, J. R., Boyko, R., Willard, L., Holmes, C. F., and Sykes, B. D. (2000) *Protein Sci. 9*, 252−264.
37. Dyson, H. J. (1995) *Methods Enzymol. 252*, 293−306.
38. Dyson, H. J., Feng, M. F., Tennant, L. L., Slay, I., Lindell, M., Cui, D. S., Kuprin, S., and Holmgren, A. (1997) *Biochemistry 36*, 2622−2236.
39. Holmgren, A. (1979) *J. Biol. Chem. 254*, 3664−3671.
40. Wang, Y. (1999) Ph.D. Thesis, University of Alberta, Edmonton, AB, Canada.
41. Holmgren, A. (1989) *J. Biol. Chem. 264*, 13963−13966.
42. Eklund, H., Gleason, F. K., and Holmgren, A. (1991) *Proteins 11*, 13−28.
43. Eklund, H., Cambillau, C., Sjoberg, B. M., Holmgren, A., Jornvall, H., Hoog, J. O., and Branden, C. I. (1984) *EMBO J. 3*, 1443−1449.
44. McFarlan, S. C., Terrell, C. A., and Hogenkamp, P. C. (1992) *J. Biol. Chem. 267*, 10561−10569.
45. Holm, L., and Sander, C. (1993) *J. Mol. Biol. 233*, 123−138.
46. Holm, L., and Sander, C. (1996) *Science 273*, 595−602.
47. Murzin, A. G., Brenner, S. E., Hubbard, T., and Chothia, C. (1995). *J. Mol. Biol. 247*, 536−540.
48. Shindyalov, I. N., and Bourne, P. E. (1998) *Protein Eng. 11*, 739−747.
49. Hofmann, K., Bucher, P., Falquet, L., and Bairoch, A. (1999) *Nucleic Acids Res. 27*, 215−219.
50. Lee, D. Y., Ahn, B.-Y., and Kim, K.-S. (2000) *Biochemistry 39*, 6652−6659.
51. Doublie, S., Tabor, S., Long, A. M., Richardson, C. C., and Ellenberger, T. (1998) *Nature 391*, 251−258.
52. Takahashi, N., and Creighton, T. E. (1996) *Biochemistry 35*, 8342−8353.
53. Gan, Z. R., Polokoff, M. A., Jacobs, J. W., and Sardana, M. K. (1990) *Biochem. Biophys. Res. Commun. 168*, 944−951.
54. Mieyal, J. J., Starke, D. W., Gravina, S. A., and Hocevar, B. A. (1991) *Biochemistry 30*, 8883−8891.
55. Jeng, M. F., and Dyson, H. J. (1996) *Biochemistry 35*, 611−619.
56. Dyson, H. J., Feng, M. F., Tennant, L. L., Slay, I., Lindell, M., Cui, D. S., Kuprin, S., and Holmgren, A. (1997) *Biochemistry 29*, 4129−4136.
57. Sun, C. H., Berardi, M. J., and Bushweller, J. H. (1998) *J. Mol. Biol. 280*, 687−701.
58. Ren, B., Tibbelin, G., de Pascale, D., Rossi, M., Bartolucci, S., and Ladenstein, R. (1998) *Nat. Struct. Biol. 5*, 602−611.

59. Nordstrand, K., Aslund, F., Holmgren, A., Otting, G., and Berndt, K. D. (1999) *J. Mol. Biol. 286*, 541−552.
60. Powles, R. E., Deane, S. M., and Rawlings, D. E. (1995) *Microbiology 141*, 2175−2181.
61. Lim, C. J., Sa, J. H., and Fuchs, J. A. (1996) *Biochim. Biophys. Acta 1307*, 13−16.
62. Pille, S., Assemat, K., Breton, A. M., and Clement-Metral, J. D. (1996) *Eur. J. Biochem. 235*, 713−720.
63. Holmgren, A. (1985) *Annu. Rev. Biochem. 54*, 237−271.
64. Gan, Z. R., and Wells, W. W. (1987) *J. Biol. Chem. 262*, 6704−6707.
65. Vohnik, S., Hanson, C., Tuma, R., Fuchs, J. A., Woodward, C., and Thomas, G. J., Jr. (1998) *Protein Sci. 7*, 193−200.
66. Li, H., Hanson, C., Fuchs, J. A., Woodward, C., and Thomas, G. J., Jr. (1993) *Biochemistry 32*, 5800−5808.
67. Grauschopf, U., Winther, J. R., Korber, P., Zander, T., Dallinger, P., and Bardwell, J. C. (1995) *Cell 83*, 947−955.
68. Yee, A., Booth, V., Dharamsi, A., Engel, A., Edwards, A. M., and Arrowsmith, C. H. (2000) *Proc. Natl. Acad. Sci. U.S.A. 97*, 6311−6315.
69. Vogt, G., Woell, S., and Argos, P. (1997) *J. Mol. Biol. 269*, 631−643.
70. Thompson, M. J., and Eisenberg, D. (1999) *J. Mol. Biol. 290*, 595−604.

BI0115176