PROTEIN STRUCTURE REPORT

A novel member of the YchN-like fold: Solution structure of the hypothetical protein Tm0979 from *Thermotoga maritima*

JOE A. GASPAR,¹ CHENGSONG LIU,^{1,2} KENRICK A. VASSALL,¹ GABRIELA MEGLEI,¹
RICARDO STEPHEN,¹ PETER B. STATHOPOULOS,¹ ANTONIO PINEDA-LUCENA,²
BIN WU,² ADELINDA YEE,² CHERYL H. ARROWSMITH,² AND
ELIZABETH M. MEIERING¹

¹Guelph-Waterloo Centre for Graduate Studies in Chemistry and Biochemistry, University of Waterloo,
Waterloo, Ontario N2L 3G1, Canada

²Department of Medical Biophysics and Ontario Cancer Institute, University of Toronto,
Toronto, Ontario M5G 2M9, Canada

(RECEIVED August 25, 2004; FINAL REVISION August 25, 2004; ACCEPTED August 27, 2004)

Abstract

We report herein the NMR structure of Tm0979, a structural proteomics target from *Thermotoga maritima*. The Tm0979 fold consists of four β/α units, which form a central parallel β -sheet with strand order 1234. The first three helices pack toward one face of the sheet and the fourth helix packs against the other face. The protein forms a dimer by adjacent parallel packing of the fourth helices sandwiched between the two β -sheets. This fold is very interesting from several points of view. First, it represents the first structure determination for the DsrH family of conserved hypothetical proteins, which are involved in oxidation of intracellular sulfur but have no defined molecular function. Based on structure and sequence analysis, possible functions are discussed. Second, the fold of Tm0979 most closely resembles YchN-like folds; however the proteins that adopt these folds differ in secondary structural elements and quaternary structure. Comparison of these proteins provides insight into possible mechanisms of evolution of quaternary structure through a simple mechanism of hydrophobicity-changing mutations of one or two residues. Third, the Tm0979 fold is found to be similar to flavodoxin-like folds and β/α barrel proteins, and may provide a link between these very abundant folds and putative ancestral half-barrel proteins.

Keywords: Northeast Structural Genomics Consortium; *Thermotoga maritima*; DsrH; YchN; protein–protein interaction; protein fold evolution; half-barrel; flavodoxin-like fold

The structure of the *Thermotoga maritima* protein Tm0979 (GI, 4981518; PMID, 10360571; GenBank, AAD36058; SP-EMBL, Q9X074) was determined by NMR spectroscopy as part of an ongoing structural proteomics initiative for this ancient bacterium (Yee et al. 2002). Tm0979 be-

longs to the DsrH family of conserved hypothetical proteins found in bacteria and archaea (Pfam 04077, Bateman et al. 2004; COG 2168) (Fig. 1). The *dsr* locus encodes various proteins involved in sulfur metabolism (Pott and Dahl 1998). It has been shown that mutations in the *dsrH* gene in the phototrophic bacterium *Chromatium vinosum* completely abolish the ability of cells to oxidize intracellular sulfur; however, the molecular function of the protein is not known (Pott and Dahl 1998). Based on structural and sequence analysis, we discuss possible roles of this protein in intracellular sulfur oxidation and the evolutionary significance of the Tm0979 structure.

Reprint requests to: Elizabeth M. Meiering, Guelph-Waterloo Centre for Graduate Studies in Chemistry and Biochemistry, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada; e-mail: meiering@uwaterloo.ca; fax: (519) 746-0435.

Article and publication are at <http://www.proteinscience.org/cgi/doi/10.1110/ps.041068605>.

Figure 1. Alignment of primary sequences for DsrH proteins (Pfam 0477) including Tm0979, as well as for Mth1491 and YchN. PSI-BLAST analysis reveals ~44 sequence relatives for Tm0979; the 24 sequences included in Pfam are shown. The sequence alignments for Tm0979, Mth1491, and YchN are based on Dali structural alignments. Alignments within the DsrH family members (using ClustalX) and coloring for conserved residues in all proteins (blue, small and hydrophobic amino acids [A, V, L, I, M, F, W]; green, hydroxyl and amine amino acids [S, T, N, Q]; magenta, charged amino acids [D, E, R, K] and cysteine [C]; cyan, H and Y; orange, G; yellow, P) are from the Pfam database (<http://www.sanger.ac.uk/Software/Pfam>) (Bateman et al. 2004). Residues in white typeface for Mth1491 and YchN form a patch of highly conserved residues in a putative functional site (see text). β -Strands and helices are indicated by β and h, and residues in secondary structural elements are bold and underlined. For Tm0979, Mth1491, and YchN every 10th residue is indicated by • above or below the sequence.

Results and Discussion

Tm0979 forms a stable homodimer in solution

NMR experiments showed broader linewidth than expected for a protein of this size, suggesting that Tm0979 is not a monomer. The oligomerization state of the protein was investigated further using dynamic light scattering, gel filtration chromatography, and isothermal titration calorimetry; detailed results of these experiments will be reported elsewhere. A Debye plot analysis of light scattering data gave an apparent molecular weight of 22.2 kDa, i.e., consistent with a dimer structure since the molecular weight of the monomer is 12.7 kDa. Gel filtration chromatography showed that the elution time for the protein varied strongly with protein concentration due to formation of a dimer with a K_d of $\sim 5 \mu\text{M}$. Isothermal titration calorimetry experiments were consistent with the gel filtration results, and showed that dimerization changed little with temperature between 25°C and 75°C. The optimum growth temperature for *T. maritima* is $\sim 80^\circ\text{C}$ (Huber et al. 1986), and so the dimer form of Tm0979 is very likely to be biologically relevant.

Tm0979 is a novel member of the YchN-like fold

Based on the biochemical data and observation of many intermolecular NOEs in half-filter experiments, the structure for Tm0979 was refined and calculated as a dimer (see Materials and Methods). The resulting structure (Fig. 2A,B) was in good agreement with the experimental data and had good structural statistics (Table 1). The monomer fold consists of four β/α units, arranged to form a four-stranded parallel β -sheet with strand order 1234. The first three helices pack toward one side of the sheet, while the fourth packs on the opposite side. The dimer interface is formed by the fourth helices packing parallel against one another, sandwiched between the two β -sheets (Fig. 2A,B).

Comparative analysis of the Tm0979 structure was performed using the structural classification databases Dali (Holm and Sander 1998) and SCOP (Murzin et al. 1995). A Dali search identified 133 structurally similar proteins with Z scores higher than 2.0. The structures with the most significant similarity to Tm0979 were Mth1491 ($Z = 7.6$), a conserved hypothetical protein from *Methanobacterium thermoautotrophicum* (PfamB 4177, COG 1416) (Christendat et al. 2002), and the hypothetical protein YchN from *Escherichia coli* (DsrE, Pfam 02635, COG 1553) (Shin et al. 2002). In SCOP, Mth1491 and YchN have been classified as the only members of the YchN-like fold. Both Mth1491 and YchN have folds consisting of four β/α units, as in Tm0979, followed by an additional fifth β -strand. The data for Tm0979 were carefully analyzed for evidence of formation of a fifth β -strand in this protein. The C-terminal residues (82–87) exhibit relatively few NOEs, are not pro-

tected against exchange with solvent, and have chemical shifts indicative of random coil; thus they appear to be unstructured.

Although the structural comparisons identified no proteins with exactly the same number and arrangement of secondary structural elements as Tm0979, based on considerations of topology, packing, and sequence similarities (Murzin 1998) it is likely that Tm0979 represents a novel member of the YchN-like fold. Although Tm0979, Mth1491, and YchN do not have detectable sequence homology in PSI-BLAST searches and have been classified into different families in various databases, structure-based sequence alignment of Tm0979, Mth1491, and YchN confirms that the sequences are related (Fig. 1). Sequence analysis has also suggested evidence for similarities among the DsrEFH proteins in *Chromatium vinosum* (Pott and Dahl 1998).

Quaternary structure analysis of YchN-like proteins

The quaternary structures of the YchN-like proteins are strikingly different: Tm0979 is a dimer, whereas Mth1491 is a trimer, and YchN is a dimer of Mth1491-like trimers (Fig. 2B,C,D). The dimer interface in Tm0979, formed by the fourth helices packing against one another and in between the β -sheets, is fundamentally different from the subunit interfaces observed in Mth1491 and YchN. In the trimer structures of Mth1491 and YchN, the fourth helices are far apart, and pack against the first helices; the center of the trimer structure is formed by the packing together of the fifth β -strands. The two trimers associate in YchN through interactions between structural elements at the C-terminal ends of the β -strands: the loop following the first β -strand, the second helix and following loop, and the loop following helix 3 (Fig. 2D). It should be noted that NMR as well as light scattering and gel filtration data for Tm0979 are not compatible with a trimer-based structure. In particular, half-filter experiments revealed extensive NOEs between the fourth helices, which would not be observed in the trimer. Furthermore, the dimer interface is well defined by a large number of NOEs (Table 1).

The size and nature of the subunit interfaces were analyzed using the program GetArea 1.1 (<http://www.scsb.utmb.edu/getarea>) (Fraczkiewicz and Braun 1998). The dimer interface in Tm0979 has an area of $1104 \text{ \AA}^2$, contains 22% polar atoms, and accounts for 19% of the total surface area of the monomer. This interface is relatively large and nonpolar, comparable to interfaces observed in other homodimeric proteins that tend to be obligate dimers (Nooren and Thornton 2003). The K_d in the μM range is comparable, though somewhat higher, than values measured for other dimers with similar interface characteristics (Nooren and Thornton 2003). For Mth1491, the total interface surface per monomer is $1318 \text{ \AA}^2$, consists of 35% polar atoms, and

Figure 2. Structures of YchN-like proteins. (A) Stereo diagram of backbone atoms for an ensemble of 10 refined Tm0979 dimer structures. The two monomers are colored green and blue, and numeric labels denote residue numbers. (B) Ribbon representation of a single structure of Tm0979, with conserved side chains at a putative functional site shown in CPK representation. The view is similar to that shown in A. (C) Ribbon representation of the crystal structure of the Mth1491 trimer (PDB code 1lls). Each chain is shown in a different color (green, blue, or red). Conserved residues in a putative active site are located at the subunit interfaces, and are shown for the interface between the green and the blue monomers. The residues shown for the green monomer are the ones corresponding to those shown in B. (D) Ribbon representation of the crystal structure of the YchN hexamer (PDB code 1jx7). Chains in both trimers are colored similar to Mth1491 (green, blue and red), with chains in the second trimer being lighter in color. The trimer represented by the darker colors (*lower* half of shown structure) corresponds to a rotation about the horizontal *X*-axis of $\sim 90^\circ$ relative to the view of Mth1491 in C. The second helix (h2 in Fig. 1), as well as the three loops following the first strand, the second helix, and the third helix, are involved in the packing between the trimers and are shown in black for the light green and the dark blue monomers. Conserved residues at the interface between the green and the blue subunits are shown in CPK representation and labeled. Structural representations were generated using MolMol (Koradi et al. 1996).

corresponds to 22% of the total surface area of the monomer. The higher polarity of this interface is similar to those typically observed in heterooligomer or transient dimer systems (Nooren and Thornton 2003). For YchN, considering monomer to hexamer, 2292 Å² are buried, with 33% polar residues in the interfaces and corresponding to 37% of total surface area of the monomer. Considering monomer to trimer for YchN, 1387 Å² are buried, with 33% polarity corresponding to 23% of the total area. These values are very similar to those for Mth1491. Unfortunately no experimental measurements of the strength of interactions between subunits have been reported for Mth1491 and YchN.

Insights into the origin of the different quaternary structures can be obtained from consideration of the primary

sequences of the proteins, the monomer structures, the pattern of conserved residues, and charge distribution on the surface of the proteins (Figs. 1, 3). Despite very low sequence identity, the arrangement of secondary structural elements and the distribution of surface charge are remarkably similar in Mth1491 and YchN (Fig. 3B,C); this logically underlies the similarity in trimer structure for these proteins (Fig. 2C,D). Noteworthy differences in structures for these two proteins occur in the loops following the first β -strand and second helix; the loops are longer in YchN and form a hydrophobic surface (Figs. 2C,D, 3C). In contrast, in Mth1491 the loops contain many acidic residues (D11, E12, D13, D14, E15 in the loop following β 1, and D52 and E54 in the loop following helix 2) that form an acidic patch on

Table 1. Structural statistics for the ensemble of 10 refined conformers of Tm0979 homodimer

Distance restraints (per monomer)	
All	2213
Intraresidue	483
Sequential ($ i - j = 1$)	533
Medium range ($2 \leq i - j \leq 4$)	359
Long range ($ i - j > 4$)	551
Intermolecular	235
Hydrogen bonds ^a	26 × 2
Dihedral angle restraints (per monomer)	
All	98
ϕ	49
ψ	49
DYANA target function, ($\text{\AA}^2$) ^b	4.10 ± 0.41
Residual violations per structure	
Distance constraints ($>0.5 \text{ \AA}$)	0
Dihedral angle restraints ($>5 \text{ \AA}$) ^c	0.3
R.m.s.d. from experimental restraints	
Distance ($\text{\AA}$)	0.0142 ± 0.0007
Dihedral angle ($^\circ$)	0.80 ± 0.10
R.m.s.d. from idealized covalent geometry	
bond ($\text{\AA}$)	0.00361 ± 0.00007
bond angle ($^\circ$)	0.55 ± 0.02
CNS energy (kcal/mol)	
Total	−6982 ± 130
van der Waals ^c	−660 ± 19
Electrostatic	−7646 ± 124
Pairwise R.m.s.d. for residues 2–80	
Backbone atoms	0.90 ± 0.14
All heavy atoms	1.43 ± 0.21
Ramachandran plot (%)	
Residues in most favored regions	86.3
Residues in additional allowed regions	9.4
Residues in generously allowed regions	2.9
Residues in disallowed regions	1.3

^a Two restraints were used for each hydrogen bond ($d_{\text{O-HN}} = 2 \text{ \AA}$, $d_{\text{O-N}} = 3 \text{ \AA}$).

^b Values for ensemble of the 20 lowest energy structures out of 200 calculated.

^c Angle constraint ψ LYS 7 violated in three of 10 structures.

the surface of the protein (Fig. 3B). The hydrophobic patch in YchN likely facilitates packing of the loops to form the interfaces between the two trimers, whereas the acidic patch in Mth1491 would prevent this. Tm0979 on the other hand is considerably smaller (87 residues per monomer) than the other proteins (113 and 117 residues per monomer for Mth1491 and YchN, respectively), and its secondary structural elements and loops tend to be shorter or absent (e.g., 2' helix and fifth β -strand are absent) (Fig. 2A,B). Also, the relative positions of the structural elements are different, particularly for the first and fourth helices, and the surface of the monomer is more highly charged. It is noteworthy that residues in the dimer interface for Tm0979, particularly in helix 4, are quite conserved (Fig. 3A), as tends to be observed for interface residues (Elcock and McCammon 2001; Valdar and Thornton 2001).

The strong similarities between the structures of Mth1491 and YchN suggest that the YchN-like proteins may have evolved from a common ancestor having a trimer-based structure. Comparing primary sequences in the region closest to the C terminus (Fig. 1), one can see that the sequences of the Tm0979-related proteins tend to be more hydrophilic than the sequences for the Mth1491 and YchN families. This suggests an attractive mechanism for changing the oligomer structure from trimer to dimer: Mutation of one or two residues in the fifth strand of the ancestral protein to a hydrophilic residue would effectively disrupt packing together of the fifth strands that are sequestered from solvent in the center of the trimer structure. The monomer form of the protein would then have exposed hydrophobic surfaces on the first and fourth helices. If the fourth helices pack against each other instead, this would lead to the new subunit interface observed in Tm0979, with accompanying destabilization of the first helix in the absence of a binding partner. Consistent with this hypothesis, the first helix in Tm0979 is less well formed than in the other proteins (Fig. 2B). Furthermore, none of the amides in this helix are protected against exchange with solvent, indicating that it is unstable. Mutations involving changes in residue hydrophobicity can also be envisioned for evolution of dimer to trimer and for evolution of these proteins from an ancient dimeric half-barrel structure (see below).

Insights into the evolution of protein–protein interactions may be gained by considering subunit interfaces observed in different protein families. Many families contain proteins with different quaternary structures and subunit interfaces (Nooren and Thornton 2003). Naturally occurring and designed substitutions that change the hydrophobicity of one or two amino acids have been shown to change the oligomer state of various structurally diverse proteins (Davison et al. 2001; O'Neill et al. 2001; LeFevre and Cordes 2003). These studies and the characteristics observed for the YchN-like proteins are consistent with a general mechanism for evolution of interfaces by simple mutational changes that alter residue hydrophobicity of one or two residues. These may be mutations of hydrophobic residues in buried parts of structure to hydrophilic residues, as described above, or mutations of polar surface residues to hydrophobic residues to create a new hydrophobic interface.

Possible functions of YchN-like proteins

Possible functions for Mth1491 and YchN have been proposed based on sequence and structural analyses (Christendat et al. 2002; Shin et al. 2002). Although structurally similar to dehydrogenases, amidohydrolases, and oxidoreductases, Mth1491 was shown not to belong to the former two families (Christendat et al. 2002). Based on high conservation of C72, N74, and N27 (Fig. 1), which are clustered together at a subunit interface (Fig. 2C), and

Figure 3. Monomer structures for (A) Tm0979, (B) Mth1491, and (C) YchN. The structural representations for each protein are shown in similar orientations. (*Right*) Ribbon diagram with selected side chains shown in stick representation. The overall folds for all three proteins are very similar, in particular with respect to the lengths of the first to fourth β -strands, and the orientation of the second and the third helices with respect to the β -sheet. Highly conserved residues located at the putative functional site are shown in black, and selected residues that tend to be conserved at subunit interfaces are shown in blue. For Tm0979 (A), Y72, F75, I76, and E80 in the fourth helix make intermolecular interactions with residues in the fourth helix and first β -strand, such as L3 and K7, of the other monomer. For Mth1491 (B), residues in the fifth strands, A109, Y110, I111 and R112, pack together in the middle of the trimer structure, while L28 in the first helix and V97 and I100 in the fourth helix pack into the interface formed between these helices. For YchN (C), residues in the fifth strand (V114, L115, F117) pack together in the center of the trimer, while L24, L28 in the first helix and L101 in the fourth helix pack into the interfaces formed between these helices. (*Middle*) van der Waals' surfaces for residues colored according to residue conservation calculated by ConSurf (Glaser et al. 2003). The colors range from dark blue for the most conserved residues (score of 9) to very light blue for variable residues (score of 1). Highly conserved putative functional residues are colored black. Sequences for calculating conservation were taken from Pfam full listings. (*Left*) van der Waals' surfaces for atoms colored according to charge. Positively charged atoms are in blue, neutral atoms are in white, and negatively charged atoms are in red. All structural representations were generated using MolMol (Koradi et al. 1996).

observation of an alteration in crystal form in the presence of ammonium sulfate, suggestive of an interaction with sulfate, Mth1491 was proposed to play a role in sulfur metabolism. A cluster of conserved residues was also found in YchN (Figs. 1, 2D; Shin et al. 2002) at a location similar to that in Mth1491, near the protein surface at the interface between subunits. Although the residues conserved in the putative functional site region are generally different between Mth1491 and YchN, the cluster for YchN also includes conserved cysteine (C78 and C81). Proposed possible functions for YchN included redox or hydrolase activity.

Tm0979 also has highly conserved residues, D54, A57, R58, G59 (Figs. 1, 2B), in the region of the conserved cysteines in Mth1491 and YchN; however, the subunit interface has been altered. Furthermore, there is no conserved cysteine in the Tm0979 family (Fig. 1). The conserved D54 is in a similar location as the conserved cysteines, and may have a similar role in protein function. It has been observed in other protein families that equivalent functional roles can be performed by different residues (Todd et al. 2001). Titrations of Tm0979 with various possible ligands monitored by HSQC showed, however, that SO_4^{2-} has no effect on Tm0979 (unlike Mth1491) (Christendat et al. 2002). Other

possible ligands tested (SO_3^{2-} , Cl^- , acetate, Na^+ , Mg^{2+} , Ca^{2+} , NH_4^+ , histidine, ribose, glucose) also showed no significant binding effects. Tm0979 showed no significant activity when screened for catalytic function using a broad range of general enzymatic assays designed to enable the identification of enzyme subclasses (A. Yakunin, pers. comm.).

Based on analyses of genes occurring in various organisms, and the in vivo consequences of altering *dsr* genes, it has been suggested that DsrEFH proteins may have a function in the assembly, folding, or stabilization of sirohaem proteins (Pott and Dahl 1998). The structure of Tm0979 would not be inconsistent with such a function. It is possible that Tm0979 may have a different function from its structural homologs; such a phenomenon has been observed in a substantial proportion of other fold families (Todd et al. 2001).

Common evolutionary origin of YchN-like, TIM barrel, and flavodoxin folds?

Among the many β/α structures containing parallel β -sheets identified by Dali as being similar to the structure of Tm0979, flavodoxin-like proteins (strand order 21345) were particularly common, and had the highest Z scores (up to 4.1 for PDB code 1kgs) after Mth1491 and YchN. Another particularly interesting related structure is the HisF protein from *T. maritima*, which has a TIM barrel fold (strand order 1234578, PDB code 1thf, Z = 3.0, alignment of 72 residues with RMSD 3.5 Å) comprised of two homologous "half-barrel" structures (Lang et al. 2000; Hocker et al. 2001). Based on studies of this and related proteins, it has been proposed that the extremely high sequence divergence of the remarkably abundant TIM barrel fold may arise from the combination of half-barrel proteins; however, to date a natural half-barrel protein has not been identified. Tm0979 may be structurally related to such a fold. A half-barrel protein would be expected to dimerize through a hydrophobic interface between the four-stranded β -sheets. The face of the β -sheet in Tm0979 is hydrophobic as required. Furthermore, the fourth helix of Tm0979 is not so well packed, nor is the dimer very stable. It is possible that Tm0979 may have evolved from a half-barrel ancestor in which one or two mutations, for example, of a hydrophobic residue to a hydrophilic residue in the interface between the fourth helix and the β -sheet, resulted in repositioning of this helix to the other side of the sheet. Furthermore, Sterner and coworkers (Hocker et al. 2002) have proposed that the half-barrels of HisF and HisA are evolutionarily related to the flavodoxin fold, which is also structurally similar to Tm0979. Thus, the structure of Tm0979 may provide important clues regarding relationships between ancient protein folds and modes of fold evolution.

Materials and methods

Protein expression and purification

The Tm0979 target was selected based on published NMR screening criteria (Yee et al. 2002). The gene sequence for Tm0979 was PCR-amplified from *T. maritima* genomic DNA and subcloned immediately after the N-terminal (His)₆ tag and the thrombin cleavage site (with sequence MGSS(H)₆SSGLVPRGSH) of the pET15b vector; then protein was expressed in *Escherichia coli* strain BL21(GoldλDE3) and purified using Ni^{2+} affinity chromatography (Qiagen), essentially as described (Yee et al. 2002), except that ZnCl_2 was not added to the M9 growth medium and no protease cleavage step was performed. NMR samples typically contained ~1 mM ^{15}N - and ^{13}C -labeled protein in 450 mM NaCl, 25 mM $\text{NaH}_2\text{PO}_4/\text{Na}_2\text{HPO}_4$ (pH 6.5), with 0.01% sodium azide, and 5% D_2O .

NMR spectroscopy

All NMR data were acquired at 27°C on a Bruker Avance DMX 600 MHz spectrometer equipped with a triple resonance xyz-gradient probe. Spectra were processed using Felix97.0 (MSI, Inc.). The backbone assignments were obtained using HNCO, CBCA(CO)NH, CBCANH, HNHA (Bax et al. 1994; Kay 1997) and 3D ^{15}N and ^{13}C TS H-CN-H NOESY-HSQC (Xia et al. 2003). Aliphatic side-chain assignments were derived from CC(CO)NH, HCCH-TOCSY, HCC-TOCSY (Bax et al. 1994; Kay 1997). Aromatic ring resonances were assigned using 2D ^1H - ^1H TOCSY and ^1H - ^1H NOESY on samples in D_2O . Intermolecular NOEs were identified using an isotope-filtered 3D ^{13}C HMQC-NOESY experiment (Lee et al. 1994). Slowly exchanging amide protons were monitored by dissolving the protein in D_2O and acquiring a series of ^{15}N -HSQC spectra.

Structure determination

Evaluation of spectra and manual assignments were performed using XEASY (Bartels et al. 1995). Assignments were obtained for 99% of the backbone and C β s, and 80% of the side-chain chemical shifts. Dihedral angle restraints were calculated from chemical shifts using the program TALOS (Cornilescu et al. 1999). Distance restraints for structure calculations were derived from cross-peaks in a 3D ^{15}N and ^{13}C TS H-CN-H NOESY-HSQC and 3D ^{13}C HMQC-NOESY experiments (Lee et al. 1994) (τ_m = 100 msec).

The structure calculation was performed as described (Wu et al. 2003) with the following modifications. An initial fold for the monomer of Tm0979 was obtained as described (Wu et al. 2003) using the program CYANA (Herrmann et al. 2002). The NOE peak list and angle constraint file were duplicated and intermolecular NOEs from isotope-filtered experiments were added for subsequent structure calculations of the homodimer structure. Several iterative structure calculations were performed using NOAH/DYANA and additional assignments that were consistent with the generated structures were added. In the final structure calculation using DYANA, NOEs, dihedral angles, and hydrogen bond restraints per monomer, along with intermolecular NOEs for the dimer were used (numbers summarized in Table 1). Hydrogen bond restraints were based on observation of regular secondary structural elements with characteristic chemical shifts and NOE patterns, and slowed hydrogen exchange. The dimer structure was refined using the default simulated annealing protocol in DYANA. Two hundred structures were calculated, from which the 20 struc-

tures with the lowest target functions were selected and subjected to molecular dynamics simulation in explicit water with the program CNS (Brunger et al. 1998). The 10 structures with the lowest NOE energies were retained and validated by the program PROCHECK-NMR (Laskowski et al. 1996) and NESG validation software (<http://www.nesg.org>) (A. Bhattacharya and G.T. Montelione, unpubl.). Structures were visualized using the program MOLMOL (Koradi et al. 1996).

Accession numbers

The coordinates for the 10 water-minimized CNS structures together with the NMR constraints have been deposited in the RCSB PDB with accession code 1x9a. The NMR chemical shifts were deposited in BioMagResBank (BMRB), accession code BMRB-6324.

Acknowledgments

This study was supported by NSERC (E.M.M.), Ontario Research and Development Challenge Fund, and NIH Protein Structure Initiative (grant P50-GM62413-02) (C.H.A.) and NESG.

References

- Bartels, C., Xia, T.-H., Billeter, M., Güntert, P., and Wüthrich, K. 1995. The program XEASY for computer-supported NMR spectral analysis of biological macromolecules. *J. Biomol. NMR* **5**: 1–10.
- Bateman, A., Coin, L., Durbin, R., Finn, R.D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E.L., et al. 2004. The Pfam protein families database. *Nucleic Acids Res.* **32**: D138–D141.
- Bax, A., Vuister, G.W., Grzesiek, S., Delaglio, F., Wang, A.C., Tschudin, R., and Zhu, G. 1994. Measurement of homo- and heteronuclear J couplings from quantitative J correlation. *Methods Enzymol.* **239**: 79–105.
- Brünger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., et al. 1998. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.* **54**: 905–921.
- Christendat, D., Saridakis, V., Kim, Y., Kumar, P.A., Xu, X., Semesi, A., Joachimiak, A., Arrowsmith, C.H., and Edwards, A.M. 2002. The crystal structure of hypothetical protein Mth1491 from *Methanobacterium thermoautotrophicum*. *Protein Sci.* **11**: 1409–1414.
- Cornilescu, G., Delaglio, F., and Bax, A. 1999. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR* **13**: 289–302.
- Davison, T.S., Nie, X., Ma, W., Lin, Y., Kay, C., Benchimol, S., and Arrowsmith, C.H. 2001. Structure and functionality of a designed p53 dimer. *J. Mol. Biol.* **307**: 605–617.
- Elcock, A.H. and McCammon, J.A. 2001. Identification of protein oligomerization states by analysis of interface conservation. *Proc. Natl. Acad. Sci.* **98**: 2990–2994.
- Fraczkiewicz, R. and Braun, W. 1998. Exact and efficient analytical calculation of the accessible surface areas and their gradients for macromolecules. *J. Comp. Chem.* **19**: 319–333.
- Glaser, F., Pupko, T., Paz, I., Bell, R.E., Bechor-Shental, D., Martz, E., and Ben-Tal, N. 2003. ConSurf: Identification of functional regions in proteins by surface-mapping of phylogenetic information. *Bioinformatics* **19**: 163–164.
- Herrmann, T., Güntert, P., and Wüthrich, K. 2002. Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J. Mol. Biol.* **319**: 209–227.
- Hocker, B., Beismann-Driemeyer, S., Hettwer, S., Lustig, A., and Sterner, R. 2001. Dissection of a ($\beta\alpha$)8-barrel enzyme into two folded halves. *Nat. Struct. Biol.* **8**: 32–36.
- Hocker, B., Schmidt, S., and Sterner, R. 2002. A common evolutionary origin of two elementary enzyme folds. *FEBS Lett.* **510**: 133–135.
- Holm, L. and Sander, C. 1998. Touring protein fold space with Dali/FSSP. *Nucleic Acids Res.* **26**: 316–319.
- Huber, R., Langworthy, T.A., Koenig, H., Thomm, M., Woese, C.R., Sleytr, W.B., and Stetter, K.O. 1986. *Thermotoga maritima* sp. nov. represents a new genus of unique extremely thermophilic eubacteria growing up to 90°C. *Arch. Microbiol.* **144**: 324–333.
- Kay, L.E. 1997. NMR methods for the study of protein structure and dynamics. *Biochem. Cell Biol.* **75**: 1–15.
- Koradi, R., Billeter, M., and Wüthrich, K. 1996. MOLMOL: A program for display and analysis of macromolecular structures. *J. Mol. Graph.* **14**: 51–55.
- Lang, D., Thoma, R., Henn-Sax, M., Sterner, R., and Wilmanns, M. 2000. Structural evidence for evolution of the β/α barrel scaffold by gene duplication and fusion. *Science* **289**: 1546–1550.
- Laskowski, R.A., Rullmann, J.A., MacArthur, M.W., Kaptein, R., and Thornton, J.M. 1996. AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* **8**: 477–486.
- Lee, W., Revington, M.J., Arrowsmith, C., and Kay, L.E. 1994. A pulsed field gradient isotope-filtered 3D ^{13}C HMQC-NOESY experiment for extracting intermolecular NOE contacts in molecular complexes. *FEBS Lett.* **350**: 87–90.
- LeFevre, K.R. and Cordes, M.H. 2003. Retroevolution of λ Cro toward a stable monomer. *Proc. Natl. Acad. Sci.* **100**: 2345–2350.
- Murzin, A.G. 1998. How far divergent evolution goes in proteins. *Curr. Opin. Struct. Biol.* **8**: 380–387.
- Murzin, A.G., Brenner, S.E., Hubbard, T., and Chothia, C. 1995. SCOP: A structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* **247**: 536–540.
- Nooren, I.M. and Thornton, J.M. 2003. Diversity of protein–protein interactions. *EMBO J.* **22**: 3486–3492.
- O'Neill, J.W., Kim, D.E., Johnsen, K., Baker, D., and Zhang, K.Y. 2001. Single-site mutations induce 3D domain swapping in the B1 domain of protein L from *Peptostreptococcus magnus*. *Structure (Camb.)* **9**: 1017–1027.
- Pott, A.S. and Dahl, C. 1998. Sirohaem sulfite reductase and other proteins encoded by genes at the *dsr* locus of *Chromatium vinosum* are involved in the oxidation of intracellular sulfur. *Microbiology* **144**: 1881–1894.
- Shin, D.H., Yokota, H., Kim, R., and Kim, S.H. 2002. Crystal structure of a conserved hypothetical protein from *Escherichia coli*. *J. Struct. Funct. Genomics* **2**: 53–66.
- Todd, A.E., Orengo, C.A., and Thornton, J.M. 2001. Evolution of function in protein superfamilies, from a structural perspective. *J. Mol. Biol.* **307**: 1113–1143.
- Valdar, W.S. and Thornton, J.M. 2001. Protein–protein interfaces: Analysis of amino acid conservation in homodimers. *Proteins* **42**: 108–124.
- Wu, B., Yee, A., Pineda-Lucena, A., Semesi, A., Ramelot, T.A., Cort, J.R., Jung, J.W., Edwards, A., Lee, W., Kennedy, M., et al. 2003. Solution structure of ribosomal protein S28E from *Methanobacterium thermoautotrophicum*. *Protein Sci.* **12**: 2831–2837.
- Xia, Y., Yee, A., Arrowsmith, C.H., and Gao, X. 2003. $^1\text{H}(\text{C})$ and $^1\text{H}(\text{N})$ total NOE correlations in a single 3D NMR experiment. ^{15}N and ^{13}C time-sharing in t_1 and t_2 dimensions for simultaneous data acquisition. *J. Biomol. NMR* **27**: 193–203.
- Yee, A., Chang, X., Pineda-Lucena, A., Wu, B., Semesi, A., Le, B., Ramelot, T., Lee, G.M., Bhattacharya, S., Gutierrez, P., et al. 2002. An NMR approach to structural proteomics. *Proc. Natl. Acad. Sci.* **99**: 1825–1830.