Journal of Biomolecular NMR **28:** 81–84, 2004.
KLUWER/ESCOM

Letter to the Editor: Solution structure of hypothetical protein TA1414 from *Thermoplasma acidophilum*

Daniel Monleón, Adelinda Yee[b], Chen Song Liu[b], Cheryl Arrowsmith[b] & Bernardo Celda[a,*]
[a]*Department of Physical Chemistry, University of Valencia, C/ Dr. Moliner 50, Burjassot 46100 Valencia, Spain;*
[b]*Division of Molecular and Structural Biology, Ontario Cancer Institute and Department of Medical Biophysics, University of Toronto, 610 University Avenue, Toronto, ON, M5G2M9, Canada*

Received 13 June 2003; Accepted 18 July 2003

Key words: heteronuclear NMR, structural genomics, structural proteomics, *Thermoplasma acidophilum*

Biological context

At the present time, the genomes of more than 100 prokaryotic and eukaryotic organisms have been sequenced. One common approach for elucidation of the molecular function of a protein is to determine its three-dimensional conformation by either X-ray crystallography or NMR, and then to compare the solved structure against known structures. Therefore, completion of sequencing efforts of many genome projects has shifted attention towards rapid structure and function determination for all the proteins encoded by newly discovered genes. Structural proteomics is an emerging scientific field aiming to obtain one or more representative 3D structures for every structural domain family in nature by application of high throughput structure determination techniques. These structures, and the corresponding protein production vectors and resonance assignments, will provide a valuable resource for structural and functional studies of the hundreds of proteins and their homologues that are targeted by the international structural proteomics efforts (Yee et al., 2003).

Here we describe the solution structure of TA1414 as part of a structural proteomics project on the feasibility of the high-throughput generation of samples from *Thermoplasma acidophilum* for structural studies. *T. acidophilum* is a thermoacidophilic archaeon that inhabits a hot and highly acidic environment in which few organisms are viable. The genome of *T. acidophilum* is one of the smallest among free-living organisms (Ruepp et al., 2000). TA1414 is an archaeal hypothetical protein of unknown function that shows some sequence identity to YhhP from *E. coli* (Katoh et al., 2000), another archaeal protein with known structure but also unknown function. The structure of TA1414 reveals a new variation of a common fold. Structural similarities have been detected between TA1414 and the C-terminal of translation initiation factor IF3. Despite fold resemblance suggests a common function for all three proteins TA1414, YhhP and IF3-C, differences in electrostatic charges distribution at the protein surface could indicate otherwise.

*To whom correspondence should be addressed. E-mail: bernardo.celda@uv.es

Methods and results

A recombinant protein consisting of the full sequence of TA1414 (78 amino acids) was expressed in *E. coli* BL21-Gold (DE3) cells containing the pET-15b expression vector (Novagen). Cells were grown in a U-^{15}N and U-^{13}C,^{15}N-labeled 2X-M9 media at 37 °C to an OD_{600} of 1.2 and induced with 1 mM IPTG for overnight at 15 °C. The protein was purified to homogeneity using metal affinity chromatography. The purified protein contained the complete sequence of TA1414 plus the full N-terminal histidine affinity tag. ^{15}N-labeled or ^{13}C/^{15}N-labeled protein solution was prepared in 25 mM sodium phosphate (pH = 6.5), 450 mM NaCl, 1 mM DTT, 95% H_20/5% D_2O. The concentration of the purified protein ranged between 1.5–2.0 mM.

All NMR spectra were recorded at 25 °C on a Bruker AVANCE DRX 500 MHz spectrometer equipped with pulse-field gradient triple-resonance probes. Linear prediction to double number of points was used in the ^{1}H, ^{13}C and ^{15}N indirect dimensions to improve digital resolution. Spectra were processed

Table 1. Structural statistics for the ensemble of 10 lowest energy structures out of 100 calculated for TA1414

Distance restraints	
All	1352
Intra residue	164
Sequential ($\mid i-j \mid = 1$)	343
Medium range ($2 \leq \mid i-j \mid \leq 5$)	302
Long range ($5 \leq \mid i-j \mid$)	543
Dihedral angle restraints	
All	141
ϕ, ψ	70,70
ω	1
Pair wise r.m.s.d	
All residues	
Backbone atoms	0.41 ± 0.16
Heavy atoms	1.17 ± 0.31
Regular secondary structure residues	
Backbone atoms	0.27 ± 0.08
All heavy atoms	0.69 ± 0.11

using the XwinNMR 3.1 software package. Sparky 3.91 (Goddard and Kneller, 1999) and home made shell and perl scripts were used for semi automated peak picking and peak lists filtering (Monleon et al., 2002). Backbone assignments of TA1414 were obtained mainly by combined analysis of CBCA(CO)NH and HNCACB and verified using a HNCA experiment (Bax et al., 1994; Kay, 1997). Peaks lists obtained were initially analyzed with AutoAssign (Moseley et al., 2001) and manually confirmed and extended. In the ^{1}H-^{15}N HSQC, 96% backbone amide resonances were assigned. The nearly complete side-chain proton resonances were identified manually by analysis of H(CC)(CO)NH and HCCH-TOCSY. The ^{1}H, ^{13}C and ^{15}N chemical shifts were referenced to DSS according to the IUPAC recommendation (Markley et al., 1998), and have been deposited in the BioMagResBank (accession code BMRB-5797).

For structure calculation purposes, ^{13}C-NOESY-HSQC (τ_m of 100 and 200 ms) and ^{15}N-NOESY-HSQC (τ_m of 100 ms) spectra were recorded. NOE cross-peak assignment was obtained using a combination of manual and automatic methods. A preliminary fold was calculated on the basis of manually unambiguously assigned NOEs. NOE assignments were extended and structure was further refined by spectra/structure iterative semi-automated analysis with the NOAH module in the program DYANA (Güntert et al., 1997). Pro17 was detected to be in a *cis* conformation by direct observation of sequential H_α-H_α NOE cross-peaks. Peak lists of the NOESY spectra were obtained by interactive peak picking using the 'restricted peak picking' option of the program SPARKY. Backbone dihedral restraints were derived from the $^{1}H_\alpha$ and $^{13}C_\alpha$ secondary chemical shifts using TALOS (Cornilescu et al., 1999). A summary of the final set of structural restraints used for torsion angle dynamics calculations together with other statistics for the ensemble of the 10 lowest energy conformers is reported in Table 1. The program MOLMOL (Koradi et al., 1996) was used to analyze the 10 energy-minimized conformers with lowest NOE violations and to calculate solvent accessible surface and electrostatic charges distributions. MOLMOL and MOLSCRIPT (Kraulis et al., 1991) were used to prepare drawings of the structures. Superposition of 10 lowest violations calculated conformers is shown in Figure 1.

Discussion and conclusions

TA1414 folds into a compact two-layered α/β-sandwich structure with a βαβαββ fold. One layer comprises two α-helices (I:Cys16 to Gln29 and II:Ala43 to Gln54) and the other contains a mixed β-sheet (4 β-strands). First two adjacent β-strands are connected by helix I and second two adjacent β-strands are connected by helix II. A small β-strand (Lys4 to Asp6) is located at the N-terminal tail of the protein and maintains an orientation perpendicular and very close to the β-sheet. The four β-strands of the sheet comprise Arg7 to Gly13, Val32 to Asp42, Gly57 to Val61 and Gly68 to Val77. Both α-helices present some slight bending at approximately 1/3 of their length in coincidence with respective proline residues. However, these proline residues, Pro17 and Pro50 for helix I and II, respectively, have been found in different *cis/trans* conformations. TA1414 fold is almost identical to that reported for closest homologue with known structure, YhhP from *E. coli* (PDB accession code 1DCJ, Figure 2B) with 23% of sequential identity, as it can be seen in Figure 2. RMSD for C_α atoms in secondary structure segments between mean structures of both proteins is 2.5 Å. In YhhP, bending of α-helices is only observed in helix II forced by Pro53, equivalent to Pro50 in TA1414. On the other hand, N-

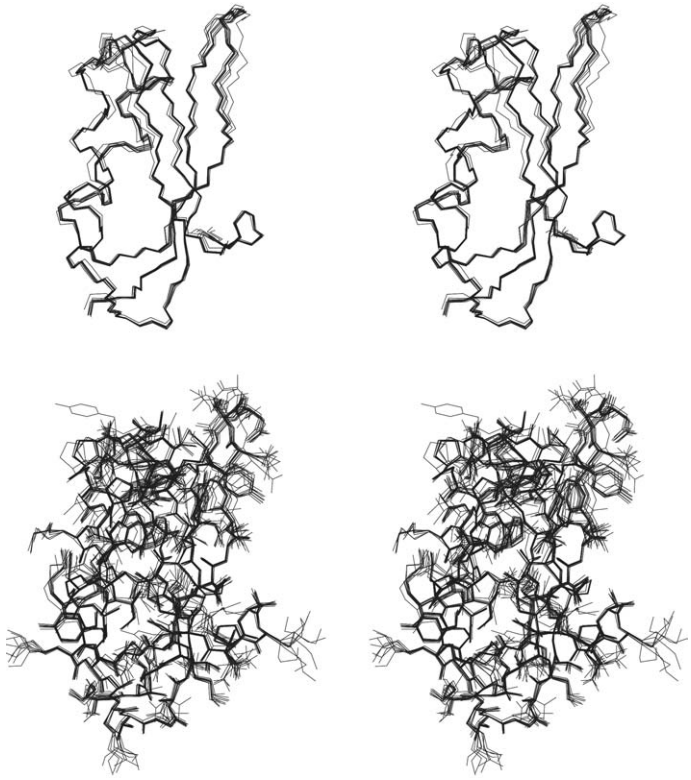

Figure 1. Stereo view of the backbone (top) and heavy atoms (bottom) of 10 lowest NOE violations NMR-derived structures of TA1414 of *Thermoplasma acidophillum* (residues 1–78).

terminal tail is much closer to the body of the protein in the case of TA1414 than in YhhP and the β-strand extended conformation is lost in the latest. Overall, TA1414 folding seems slightly more compact than that reported for YhhP although structural similarity can be considered very high.

A three-dimensional structure search using DALI (Holm and Sanders, 1995) showed that, besides YhhP *E. coli* protein, TA1414 shares some structural homology (DALI Z-score of 3.5) with the C-terminal segment of the translation initiation factor IF3 of *Bacillus stearothermophilus* (PDB accession code 1TIG, Figure 2C). Only 9% amino acid sequence identity has been found between TA1414 and IF3-C. IF3-C contains a βαβαββ fold with 2 α-helices and 4 β-strands arranged in a parallel α/β sandwich fashion like TA1414. As shown in Figure 2, helices and strands run in the same sequence direction in all three homologues. Main structural differences are found in the relative orientation of the α-helices. Otherwise, structures of TA1414 and IF3-C are very similar with a RMSD value for C_α atoms of secondary structure segments between mean structures of both proteins of 3.5 Å. However, a closer examination reveals a significantly different pattern in the electrostatic charges distribution at the protein surface (Figure 3). As it is shown, in the location of a strong positive cluster at the helical side of TA1414 and YhhP, a negative patch is found for IF3-C. Positive residues at this cluster include Lys28 and Arg31 in TA1414 meanwhile in IF3-C, Glu111 and Asp114 are found at these locations. On the other hand, the north part of this side shows a similar effect with negative residues Asp44, Asp45 and Asp51 in TA1414 and positive residues Arg121, Lys123 and Arg125 in similar positions. The β-sheet side of the sandwich, with a general negative character in all three proteins, does not show such a dramatic change in the electrostatic charges at the surface.

In summary, we present the solution structure of TA1414, a functionally unknown protein in *Thermoplasma acidophillum*. TA1414 has a common βαβαββ fold with a α/β sandwich three-dimensional arrangement very similar to that reported for the C-terminal segment of IF3 protein. However, this homology does

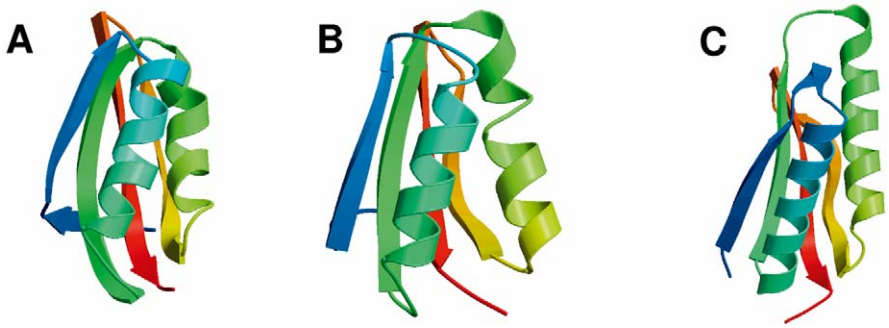

Figure 2. Ribbon diagram in rainbow-coloring criterion (blue, N-terminal; red, C-terminal) depicting (A) lowest energy NMR structure of TA1414 of *Thermoplasma acidophillum* (PDB accession code 1PAV), (B) averaged minimized NMR structure of YhhP of *Escherichia coli* (PDB accession code 1DCJ), and (C) X-ray crystal structure of IF3-C of *Bacillus stearothermophilus* (PDB accession code 1TIG).

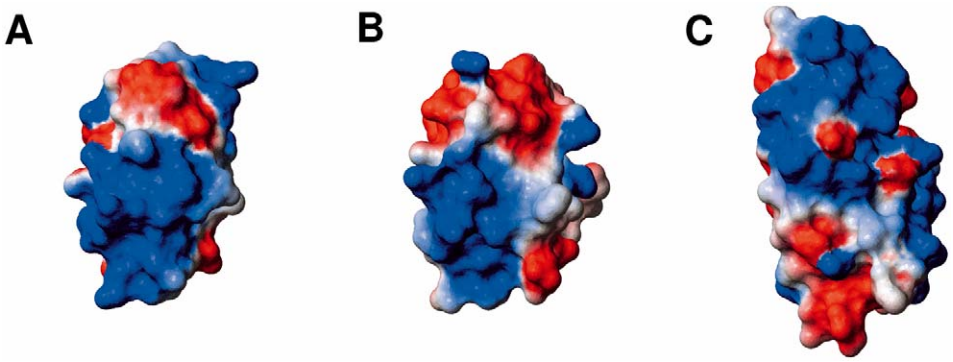

Figure 3. Charge distribution at the solvent accessible protein surface as calculated by the GRASP module of the program MOLMOL (Koradi et al., 1996) for (A) lowest energy NMR structure of TA1414 of *Thermoplasma acidophillum*, (B) averaged minimized NMR structure of YhhP of *Escherichia coli*, and (C) X-ray crystal structure of IF3-C of *Bacillus stearothermophilus*.

not extend to the dramatically different electrostatic charges distribution at the surface. Therefore, it is not possible to infer directly any function for TA1414 on the only basis of the NMR structure reported here. Structures ensemble has been submitted to the Protein Data Bank (PDB accession code 1PAV).

Acknowledgements

The authors wish to thank the SCSIE of the University of Valencia for providing access to the NMR facility and high performance computing facilities. We also thank Bruker España S.A. for technical and economic support as part of the agreement with the University of Valencia for the development of new techniques in biomolecular NMR. This work was also partially funded by the Ontario Research and Development Challenge Fund.

References

Bax, A., Vuister, G.W., Grzesiek, S., Delaglio, F., Wang, A.C., Tschudin, R. and Zhu, G. (1994) *Meth. Enzymol.*, **239**, 79–105.

Cornilescu, F., Delaglio, F. and Bax, A. (1999) *J. Biomol. NMR*, **13**, 289–302.

Frishman, D., Stocker, S., Lupas, A.N. and Baumeister, W. (2000) *Nature*, **407**, 508–513.

Goddard, T.D. and Kneller, D.G. (1999) *SPARKY 3*, University of California, San Francisco.

Güntert, P., Mumenthaler, C. and Wüthrich, K. (1997) *J. Mol. Biol.*, **273**, 283–298.

Holm, L. and Sander, C. (1995) *Trends Biochem. Sci.*, **20**, 478–480.

Katoh, E., Hatta, T., Shindo, H., Ishii, Y., Yamada, H., Mizuno, T. and Yamazaki, T. (2000) *J. Mol. Biol.*, **304**, 219–229.

Kay, L.E. (1997) *Biochem. Cell Biol.*, **75**, 1–15.

Koradi, R., Billeter, M. and Wüthrich, K. (1996) *J. Mol. Graphics*, **14**, 51–55.

Kraulis, P.J. (1991) *J. Appl. Crystall.*, **24**, 946–950.

Markley, J.L., Bax, A., Arata, Y., Hilbers, C.W., Kaptein, R., Sykes, B.D., Wright, P.E. and Wüthrich, K. (1998) *J. Mol. Biol.*, **280**, 933–952.

Monleon, D., Colson, K., Moseley, H.N.B., Anklin, C., Oswald, R., Szyperski, T. and Montelione, G.T. (2002) *J. Funct. Struct. Genom.*, **2**, 93–101.

Moseley, H.N.B., Monleon, D. and Montelione, G.T. (2001) *Meth. Enzymol.*, **339**, 91–108.

Ruepp, A., Graml, W., Santos-Martínez, M-L., Koretke, K.K., Volker, C., Mewes, H.W., Yee, A., Pardee, K., Christendat, D., Savchenko, A., Edwards, A.M. and Arrowsmith, C.H. (2003) *Acc. Chem. Res.*, **36**, 183–189.