Journal of Biomolecular NMR (2005) 31: 167–172

Letter to the Editor: NMR structure note – Solution structure of a bacterial BolA-like protein XC975 from a plant pathogen *Xanthomonas campestris* pv. campestris

Ko-Hsin Chin[a], Fu-Yang Lin[b], Yu-Chen Hu[a], Kong Hung Sze[c], Ping-Chiang Lyu[d] & Shan-Ho Chou[a*]

[a]*Institute of Biochemistry, National Chung-Hsing University, Taichung, 40227, Taiwan;* [b]*Department of Life Science, National Central University, Jung-Li, Taoyuan, Taiwan;* [c]*Department of Chemistry, The University of Hong Kong, Pokfulam;* [d]*Department of Life Science, National Tsing Hua University, Hsin-Chu, Taiwan*

Received 28 October 2004; Accepted 6 December 2004

Key words: biofilm, BolA-like, NMR structure, structural genomics, *Xanthomonas campestris* pv. campestris

Biological context

Xanthomonas campestris (XC) is a Gram-negative bacterium that is phytopathogenic to cruciferous plants and causes worldwide agricultural loss (da Silva et al., 2002). However, it also produces exopolysaccharide (xanthan gum) that is of enormous industrial importance. To identify and characterize novel protein structures and functions in *Xanthomonas campestris,* we have initiated a structural genomics program of a local strain *Xanthomonas campestris* pv. campestris 17 (Tseng et al., 1999). Its genome has been successfully sequenced using the shotgun method (P. Tsai, unpublished result, 2004) and the encoded ORFs were predicted using a bioinformatics approach (http://xcc.life.nthu.edu.tw).

XC975 is a conserved hypothetical protein that contains 80 amino acids and has a molecular weight of 10 kDa. It is classified either as a member of BolA-like protein (PFO1722) by the Pfam database (Bateman et al., 2000) or as a member of stress-induced morphogen (COG0271) by the COG database (Tatusov et al., 2001), based on a bioinformatics approach. The two classifications are in fact quite comparable. The BolA-like proteins are widely spread in most bacteria and eukaryotes, and the *E. coli* BolA gene was first identified as a stationary-phase morphogene induced under stress (Santos et al., 1999), which causes *E. coli* cells to adopt round morphology when over-expressed. It was later found that its over-expression could induce biofilm development (Vieira et al., 2004), which increases bacterial resistance to many toxic substances, such as antibiotics, detergents, and host immune defense response products (Watnick and Kolter, 2000). Given the important medical and economic consequences of this phenomenon, understanding the biofilm development process could help in the design of methods to prevent its formation. In this regard, it is important to note that after biofilm development, the biofilm dispersal in *Xanthomonas campestris* was required for full virulence to plants (Dow et al., 2003). Therefore the expression level of BolA protein may be correlated with the pathogenicity degree of *Xanthomonas campestris*. We now report the NMR resonance assignments and solution structure of XC975. These data can serve as a basis for further structural and functional characterization of the BolA-like protein family.

*To whom correspondence should be addressed. E-mail: shchou@nchu.edu.tw

Methods and results

The Xcc975 gene fragment was PCR amplified directly from the Xcc genome. It was cut with the *Eco*RI and *Xho*I restriction enzymes and cloned into a pGEX-4T-1 vector (Amersham Biosciences) cut with the same enzymes. The final construct codes for a fusion protein consisting of a N-terminal GST (glutathione S-transferase) and a C-terminal Xcc gene product under the control of a tac promoter (Shih et al., 2002). The transformed *E. coli* BL21 (DE3) host cell was then grown in a M9 minimal medium containing either 0.5 g/l of $^{15}NH_4Cl$ and 4 g/l of D-glucose or 0.5 g/l of $^{15}NH_4Cl$ and 2 g/l of [$^{13}C_6$]- d-glucose as the sole nitrogen and carbon sources at 37 °C until an OD value of 0.9 was reached. Over-expression of the fusion protein was induced by the addition of 1 mM IPTG at 20 °C for 20 h to prevent inclusion body formation. The cells were harvested, resuspended in equilibration buffer (10 mM Na_2HPO_4 140 mM, NaCl, 2.7 ml KCl, 1.8 mM KH_2PO_4, pH 8.5), and lysed by microfluidizer (Microfluidics). After centrifugation, the fusion protein was purified by affinity chromatography using a Glutathione Sepharose 4B column (Amersham Biosciences). Cleavage of the target protein from GST was achieved by employing the thrombin protease (Amersham Biosciences). The final purified target protein contains five extra amino acids (GSPEF) at the N-terminal end. Fractions containing the target protein were concentrated by ultrafiltration (Amicon) and dissolved in a 500 μl buffer solution containing 100 mM NaCl, 10 mM NaH_2PO_4 (pH 6.5), 0.025% NaN_3, and 10% D_2O.

NMR spectra were acquired at 298 °K on a Varian UNITY Inova 800 MHz spectrometer at a protein concentration of ~1 mM. The ^{1}H, ^{13}C, and ^{15}N chemical shifts were indirectly referenced to DSS (Wishart and Sykes, 1994). Sequential backbone assignments were achieved by means of 3D HNCACB, CBCA(CO)NH, HNCO, and HCACO spectra. Side-chain assignments were accomplished by using 3D HBHA(CBCACO)NH, H(CCCO)NH, CC(CO)NH (Montelione et al., 1992; Grzesiek et al., 1993), and NOESY-^{13}C-HSQC experiments. The aromatic Cδ/Hδ and Cε/Hε chemical shifts were assigned by correlating them with the corresponding Cβ/Hβ chemical shifts (Yamazaki et al., 1993). Spectra were processed and ana1yzed using the NMRPipe/NMRDraw (Delaglio et al., 1995) and SPARKY (Goddard and Kneller) software packages. Backbone atoms for 70 out of 75 non-proline residues have been assigned (Figure 1). Only the N-terminal MRKR residues and residue H77 were non-detectable in the ^{1}H-^{15}N-HSQC spectrum. Intriguingly, two cross peaks were observed for residues L15 and D16 (marked in Figure 1 with arrows), respectively, possibly due to the cis- and trans-P14-L15 peptide bond swapping in the N-termina1 end. Side chain ^{1}H and ^{13}C resonance assignments are nearly complete for those residues with assignable backbone resonances. The ^{1}H, ^{15}N, and ^{13}C chemical shifts have been deposited in the BioMagResBank (http://www.bmrb.wisc.edu) under accession number 6358.

The solution structure of XC975 was determined using the program CYANA incorporating combined automated NOE assignment and torsion angle dynamics functions (Hermann et al., 2002). ^{15}N-edited and ^{13}C-edited (aliphatic- and aromatic-selective) NOESY spectra ($\tau_m = 100$ ms) were acquired to obtain the NOE distance constraints. The NOE cross peaks were first partially assigned, then further refined in an iterative manner exploiting the automated NOE assignment capability of the CYANA program for the ambiguous cases. A manual check of all NOE assignments was carried out prior to the final

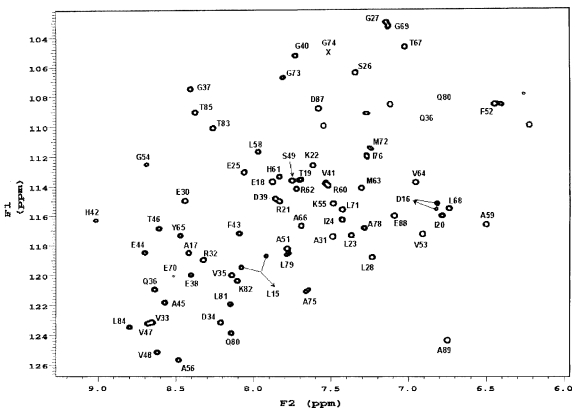

Figure 1. ^{1}H-^{15}N HSQC spectrum of XC975 collected at 298 °K. The cross peaks of residues E70 and G74 are detectable at a lower contour level. Two cross peaks were observed for each residue of L15 and D16, and were marked with arrows.

calculations. Backbone dihedral angle constraints were obtained based on the chemical shift index (CSI) of the assigned backbone $^{1}H\alpha$, $^{13}C\alpha$, $^{13}C\beta$, and $^{13}C'$ chemical shifts (Wishart and Sykes, 1994) or from the average values predicted by the program TALOS (Cornilescu et al., 1999). Some ϕ torsional angles were also obtained from the quantitative HNHA experiment (Bax et al., 1994). For determination of hydrogen bonds, amide protons exchanging slowly with deuterons were identified by time course ^{1}H-^{15}N HSQC experiments on a freeze-dried ^{15}N-labeled sample of XC957 dissolved in D_2O buffer. A total of 924 distance constraints from 1369 assigned NOE cross peaks, including 551 short-range (intra-residue and sequential), 161 medium-range, and 212 long-range NOE-based constraints, as well as 107 TALOS-based ϕ/Ψ and 39 HNHA-based ϕ torsion angle constraints, along with 22 hydrogen bond constraints, were used to refine the structure. The rmsd values of the ensemble structures based on residues 18–89 were 0.48 Å for the backbone atoms and 0.98 Å for all heavy atoms, respectively, for the 20 best final structures from the 100 initial random embeds (Figures 2a, b). As usual, the secondary structural elements converge better than the loop regions (top of Figure 2a, b). Ramachandran plot of the determined dihedral torsion angles gives 88% residues in the allowed region, with another 12% in the additionally allowed region, and no residues in the generously allowed and disallowed regions.

The overall topology of XC975 is $\alpha_1\beta_1\beta_2\alpha_2\alpha_3\beta_3$, with the second helix being more akin to a 3_{10}-turn. A β sheet is formed with three β strands, with the β_2 strand running anti-parallel to the β_1 strand but parallel to the β_3 strand (Figure 3b). All helices and 3_{10}-like turn are anchored on one side of the β-sheet. Many highly conserved hydrophobic residues such as I20, I24, V35, V47, F43, F52, V64 are located in the core region (Figures 2a, b, and 3a), accounting for the major stabilization force between the α helix and β-sheet interaction in XC975. In particular, residues I20 and I24 of helix 1 interact nicely with residues V35 and V33 of β-strand 1, respectively, while residue F52 in the 3_{10}-like turn interacts with residue V47 of β-strand 2, and residues V64 of helix 3 interact nicely both with residues V47 of β-strand 2 and residue L81 of β-strand 3 (Figure 2b). In addition, residue Y65 of helix 3 also

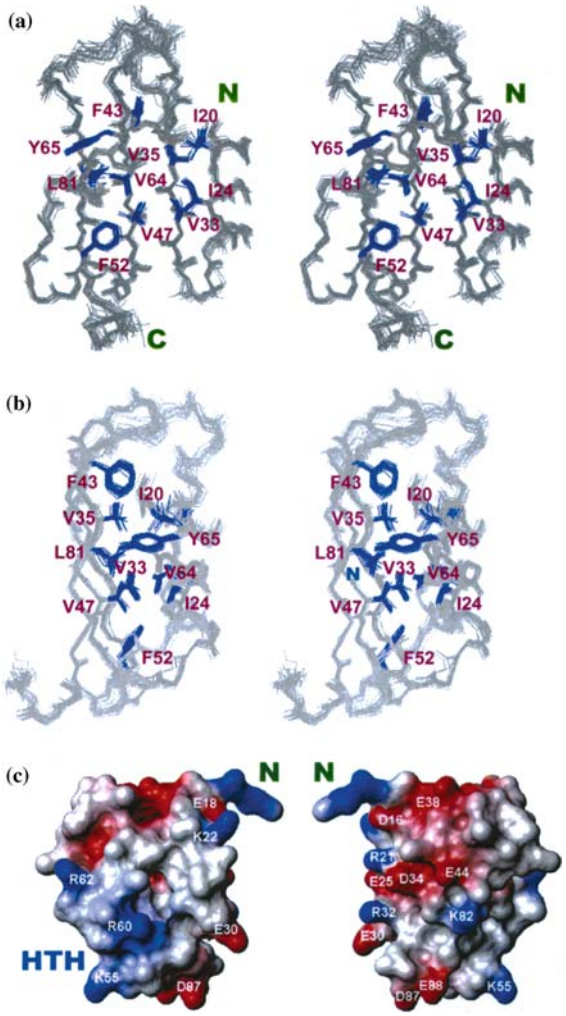

Figure 2. (a) Superposition of the twenty final NMR structures of XC97 5 in line drawing. The backbone atoms are colored in gray, while the side-chain atoms of selected hydrophobic residues in blue to illustrate the hydrophobic interaction in XC975. (b) 90° counterclockwise rotation of figure a. The β-sheet plane is on the left hand side while α-helix plane on the right hand side. Interactions between α-helices and β-sheet are obvious in this perspective. (c) The electrostatic surface views of XC975 in 'helix side' (left) and 'β-sheet side' (right). Residues are colored in red for negatively charged surfaces and blue for positively charged surfaces. The HTH region is also marked. Figures were produced using the MOLMOL program (Koradi et al., 1996).

interacts nicely with residue L81 of β-strand 3. Therefore the helix plane packs nicely against the β-sheet plane of XC975 through the abundant conserved hydrophobic residues to stabilize this protein molecule. As usual, most charged and hydrophilic residues are located in the surface to

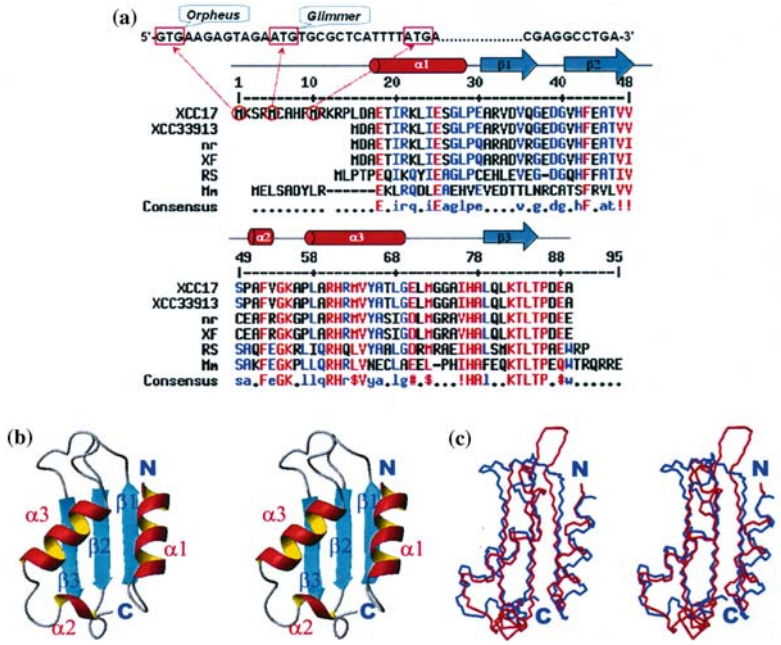

Figure 3. (a) Multiple sequence alignment of the plant pathogen and mouse BolA-like families. Identical residues are colored in red and similar residues in blue, respectively. XCC17: BolA from *Xanthomonas campestris* pv. campestris 17. Xcc33913: *Xanthomonas campestris* pv. campestris str ATCC33913. nr: *Xanthomonas campestris* pv. campestris 9a5c. XF: *Xylella fastidiosa*. RS: *Ralstonia solanacearum*. Mm: mouse *Mus musculus*. The secondary structures of XC975 are indicated at the top of the alignment. The XC975 gene sequence was shown at the top of the figure with the three possible starting codons boxed. (b) A ribbon representation of one of the lowest energy final NMR structure of XC975. (c) Overlaying of XC975 (in blue) with mouse Mm BolA-like protein (in red) (Kasai et al., 2003). These protein sequences bear a 33.8% sequence identify, hence the secondary structure elements are similar, but the loop regions exhibit considerable differences

interact with the solvent molecules (Figure 2c). The C-terminal end is well structured even when the number of these residues is too small to form a α-helix. The atom coordinates have been deposited in the Protein Data Bank (http://www.rcsb.org; PDB ID code IXS3).

Discussion and conclusion

The starting site of the XC975 gene segment 5′-**GTG**AAGAGTAGA**ATG**TGCGCTCATTTTAT**GA**...3′ (Figure 3a) was predicted simultaneously by both the Orpheus and Glimmer programs but with different results. While Orpheus predicted an ORF starting from the first GTG codon, Glimmer predicted a shorter ORF starting from the first ATG codon (Figure 3a). However, when both of these gene segments were cloned into the *E. coli,* host and expressed, no cross peak for the putative amino acid residues preceding the second ATG codon was detectable in the corresponding ^{1}H-^{15}N HSQC spectra, indicating that the extra N-terminal amino acid residues 1–9 are probably not existing in the actual protein. Hence, only the short protein starting from the second ATG codon was employed for later NMR data collection and structural studies.

Recently the solution structure of an eukaryotic BolA-like protein (Mm) from mouse was reported (Kasai et al., 2003). Since XC975 assumes a 33.8% sequence identify with this mouse protein, the two structures are therefore quite similar; both adopt the same topology of $\alpha_1\beta_1\beta_2\alpha_2\alpha_3\beta_3$ with all helices anchoring to the same side of the only β-sheet (Figure 3c). However, the loop regions exhibit considerable differences. For example, the 3_{10} helix present in Mm between the $\alpha3$ and $\beta3$ secondary structures is not observed in XC975, which may be due to the difference in amino acid sequences between the Mm (LPHI) and XC975 (MGGAI) proteins (Figure 3a). In fact, the second helix of XC975 adopts a 3_{10}-like turn in every generated structure

Table 1. Statistics for the ensemble of final NMR structures of XC975[a]

NOE distance constraints	924
Short range (intraresidue and sequential)	551
Medium range ($2 \leqq \|i-j\| \leqq 4$)	161
Long range ($\|i-j\| > 4$)	212
Hydrogen bond constraints	22
Dihedral angle constraints	107
Φ	54
Ψ	53
Number of violation	0
Distance violations (> 0.2 Å)	0
Dihedral angle violations (> 2°)	0
Structure coordinates rmsd(Å)	
Backbone atoms from the first CYANA cycle[b]	2.50
Rmsd drift between the first and the last CYANA cycles[c]	2.28
Backbone atoms for the final ensemble	0.48
All heavy atoms for the final ensemble	0.98
Ramachandran plot	
Most favored regions	63 (88%)
Additional allowed regions	9 (12%)
Generously allowed regions	0
Disallowed regions	0

[a]Statistics are based on residue 18 to residue 89 of the 20 final conformers with the lowest overall energy selected from 100 initial embeds.
[b]For the CYANA calculation to be meaningful, the rmsd value for the backbone atoms of cycle 1 ensemble is suggested to be below 3.0 Å (Hermann et al., 2002).
[c]For the CYANA calculation to be meaningful, the rmsd drift for the backbone atoms between cycle 1 ensemble and cycle 7 ensemble is suggested to be below 3.0 Å (Hermann et al., 2002).

(Figure 3b), possibly due to the different PAFV sequence in this region.

As with the mouse BolA-like protein, the 3_{10}-like turn and α3 helix of the bacterial XC975 protein form an HTH motif, and are arranged at an angle of approximately 60° to each other (Figure 3b). This arrangement is quite similar to the class II KH fold that is commonly found among the nucleic acid binding proteins (Grishin, 2001). However no GXXG loop is detected between the two adjacent helices. Instead, a possible nucleic acid binding HTH motif is formed in this region. This is consistent with the uneven charge distribution of XC975, as shown in the electrostatic surfaces in Figure 2c; the helix side (left), especially that at the HTH region, is highly basic, while the β-sheet side (right) is, on the contrary, highly acidic. In fact, three basic residues K55, R60, and R62 in this region are highly conserved in the BolA-like protein families (Figure 3a). However, since there are also a bundle of highly conserved amino acid residues in the C-terminal β-strand region, the BolA-like protein may serve functions other than plain nucleic acid binding. This is also consistent with the fact that BolA-like proteins are a special type of proteins possessing multiple functions (Santos et al., 1999; Vieira et al., 2000).

Acknowledgements

This works is supported by the Academic Excellence Pursuit grant from the Ministry of Education, and by the National Science Council, Taiwan, ROC to SH Chou and PC Lyu. KH Sze acknowledges support by grants from the Hong Kong University Research Grant and the Research Grants Council of Hong Kong (HKU 7350/04M). We also thank the National Protein Expression facility in the Academic Sinica, Taiwan for providing us the original vectors used in this study.

References

Bateman, A., Birney, E., Durbin, R., Eddy, S.R., Howe, K.L. and Sonnhammer, E.L.L. (2002) *Nucl. Acids Res*., **28**, 263–266.

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

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

da Silva, A.C.R., Ferro, J.A., Relnach, F.C., Farah, C.S., Furlan, L.R., Quaggio, R.B., Monteiro-Vitorello, C.B. et al., (2002) *Nature*, **411**, 459–463.

Delaglio, F, Grzesiek S., Vuister, G.W., Zhu, G., Pfeifer, J. and Bax, A. (1995) *J. Biomol. NMR,* **6**, 277–293.

Dow, J.M., Crossman, L., Findlay, K., He, Y.-Q., Feng, J.-X. and Tang, J.-L. (2003) *Proc. Natl. Acad. Sci. USA*, **100**, 10995–11000.

Goddard, T.D. and Kneller, D.G. University of California, San Francisco, CA.

Grishin, N.V. (2001) *Nucl. Acids. Res.,* **29**, 638–643.

Grzesiek, S., Anglister; J. and Bax, A. (1993) *J. Magn. Reson*., **B 101**, 114–119.

Hermann, T., Güntert, P. and Wüthrich, K. (2002) *J. Mol. Biol*., **319**, 209–227.

Kasai, T., Inoue, M., Koshiba, S., Yabuki, T., Aoki, M., Nunokawa, E. et al., (2003) *Protein Sci*., **13**, 545–548.

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

Montelione, G.T., Lyons, B.A., Emerson, S.D. and Tashiro, M. (1992) *J. Am. Chem. Soc*., **114**, 10974–10975.

Santos, J.M., Freire, P., Vicente, M. and Arraiano, C.M. (1999) *Mol. Microbiol*. **32**, 789–798.

Shih, Y.-P., Kung, W.-M., Chen, J.-C., Yeh, C.-H., Wang, A.H.-J. and Wang, T.-F (2002) *Protein Sci*., **11**, 1714–1719.

Tatusov, R.L., Natale, D.A., Garkavtsev, I.V., Tatusova, T.A., Shankavaram, U.T., Rao, B.S., Kiryutin, B. et al., (2001) *Nucl. Acids. Res*., **29**, 22–28.

Tseng, Y.-H., Choy, K.-T., Hung, C.-H., Lin, N.-T., Liu, J.-Y, Lou, C.-H., Yang, B.-Y. et al., (1999) *J. Bacteriol*. **181**, 117–125.

Vieira, H.L.A., Frire, P. and Arraiano, C.M. (2004) *Appl. Environ. Microbiol*. **70**, 5682–5684.

Watnick, P. and Kolter, R. (2000) *J. Bacteriol*. **182**, 2675–2679.

Wishart, D.S. and Sykes, B.D. (1994) *Meth. Enzymol*. **239**, 363–392.

Yamazaki, T., Forman-Kay, J.D. and Kay, L.E. (1993) *J. Am. Chem. Soc*., **115**, 11054–11055.