Journal of Biomolecular NMR 30: 225–226, 2004.
© 2004 Kluwer Academic Publishers. Printed in the Netherlands.

Letter to the Editor: Sequence-specific assignment and secondary structure determination of the 195-residue complex formed by the *Mycobacterium tuberculosis* proteins CFP-10 and ESAT-6

Philip S. Renshaw[a], Vaclav Veverka[a], Geoff Kelly[b], Thomas A. Frenkiel[b], Richard A. Williamson[c], Stephen V. Gordon[d], R. Glyn Hewinson[d] & Mark D. Carr[a,*]

[a]*Department of Biochemistry, University of Leicester, University Road, Leicester LE1 7RH, U.K.;* [b]*MRC Biomedical NMR Centre, NIMR, Mill Hill, London NW7 1AA, U.K.;* [c]*Department of Biosciences, University of Kent, Canterbury, Kent CT2 7NJ, U.K.;* [d]*TB Research Group, Veterinary Laboratories Agency, New Haw, Surrey KT15 3NB, U.K.*

Received 26 March 2004; Accepted 2 June 2004

Key words: CFP-10, ESAT-6, NMR resonance assignments, secondary structure, tuberculosis

Biological context

Comparative genomic studies have identified a total of 14 regions of difference between the *Myobacterium bovis* BCG-Pasteur vaccine strain and virulent *M. tuberculosis* bacilli (Behr et al., 1999). One of these regions, designated RD1, is deleted from all BCG substrains but is present in all clinical isolates and virulent strains of *M. bovis* and *M. tuberculosis*. The RD1 region contains the genes for 9 proteins (Rv3871 to Rv3879c), which are clearly implicated in tuberculosis pathogenesis (Pym et al. 2002). This includes the genes for two, sequence related, relatively small secreted proteins, which are known as CFP-10 and ESAT-6.

Neither CFP-10 nor ESAT-6 show any significant sequence similarity with any proteins of known tertiary structure or function. However, they are members of a large family of mycobacterial proteins found in the *M. tuberculosis* complex, which as with CFP-10 and ESAT-6, are encoded by genes arranged in pairs in the genome (Renshaw et al., 2002 and references therein). We have recently reported that CFP-10 and ESAT-6 form a tight, 1:1 complex (20.6 kDa), which induces the folding of both proteins and clearly represents their functional form (Renshaw et al., 2002). In this communication we report the determination of the essentially complete sequence-specific backbone and side chain resonance assignments for the CFP-10 · ESAT-6 complex and identification of the secondary structure of both proteins.

*To whom correspondence should be addressed. E-mail: mdc12@le.ac.uk

Methods and experiments

Uniformly ^{15}N, ^{13}C, and ^{15}N/^{13}C labelled CFP-10 and ESAT-6 were prepared from pET28a- and pET21a-based *E. coli* expression vectors respectively, which were grown in minimal media supplemented with 1 g/l $(^{15}NH_4)_2SO_4$ and 2 g/l U-^{13}C D-Glucose as sole nitrogen and carbon sources. ^{13}C labelled protein samples used in ^{13}C/^{1}H HMQC-NOESY experiments were prepared as described above, but with the addition of non-isotopically labelled aromatic amino acids to the minimal media (His, Phe, Trp and Tyr at 50 mg/l). The recombinant proteins were purified using anion exchange chromatography with complexes of labelled/unlabelled proteins produced as described previously (Renshaw et al., 2002) which were concentrated to give NMR samples of 0.9–1.5 mM.

All NMR data were acquired at 35 °C on either an 800 MHz Varian Inova or a 600 MHz Bruker Avance spectrometer. The 2D and 3D spectra recorded to obtain sequence specific assignments for CFP-10 and ESAT-6 in the complex were: TOCSY with mixing times of 40 and 60 ms; NOESY with an NOE mixing time of 100 ms; ^{15}N/^{1}H HSQC; TOCSY-HSQC with a mixing time of 45 ms; NOESY-HSQC with an NOE mixing time of 85 ms; ^{13}C/^{1}H HCCH-TOCSY with a mixing time of 19.6 ms;

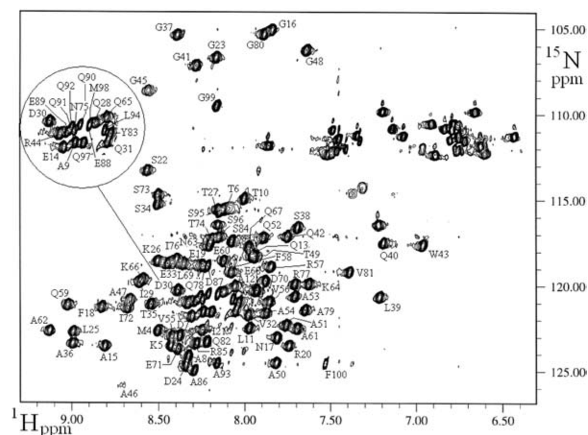

Figure 1. A ^{15}N/^{1}H HSQC spectrum recorded from a 1.3 mM sample of uniformly ^{15}N labelled CFP-10 in a 1:1 complex with unlabelled ESAT-6 at 35 °C. The assignments of signals from the backbone amide groups are indicated by residue type and number.

HMQC-NOESY with an NOE mixing time of 100 ms; and ^{15}N/^{13}C/^{1}H HNCACB, CBCA(CO)NH and HBHA(CBCACO)NH (Bax. 1994 and references therein). Typical acquisition times in F_1 and F_2 for the 3D experiments were 10–17 ms for ^{15}N, 6.3–9.5 ms for ^{13}C and 15–20 ms for ^{1}H, with F_3 (^{1}H) corresponding to about 60 ms for ^{15}N/^{13}C labelled samples and 150 ms for ^{15}N labelled. The majority of the 3D spectra were acquired over about 88 h, 2D ^{1}H experiments over about 20 h and ^{15}N/^{1}H HSQC spectra over about 30 min. Typical acquisition times in 2D experiments were either 41.7 ms (^{15}N) or 35 ms (^{1}H) in F_1 and 150 ms in F_2 (^{1}H). The 3D NMR data were processed using NMRPipe (Delaglio et al., 1995) and analysed using the XEASY package (Bartels et al., 1995).

CFP-10 and ESAT-6 in the complex give rise to well-resolved spectra as illustrated by the ^{15}N/^{1}H HSQC spectrum (Figure 1). The good sensitivity and dispersion observed for signals in the 3D spectra allowed essentially complete backbone resonance assignments to be made for the complex, which, together with patterns of sequential and medium range NOEs involving backbone amides, was used to map the secondary structure. The data clearly show that both CFP-10 and ESAT-6 exist as two long helices (residues Ala 8 to Gln 40 and Ala 46 to Gly 80 for CFP-10, and residues Phe 8 to Trp 43 and Glu 49 to Ala 84 for ESAT-6) joined by a short loop or turn (residues Gly 41 to Gly 45 and Gly 44 to Ser 48 respectively).

Extent of assignments and data deposition

Backbone amide signals (^{15}N and ^{1}H) were assigned for all residues in CFP-10 except Ala 2 and Glu 3 (98%) and for all residues in ESAT-6 except Met 1 and Thr 2 (98%). In addition to backbone amide resonances, ^{13}Cα, ^{1}Hα, ^{13}Cβ and ^{1}Hβ signals were assigned for all residues in both proteins (100%). Essentially complete assignments (^{13}C and ^{1}H) were obtained for all the non-exchangeable aliphatic side chain signals of residues apart from Ile 21 (Cγ1) and Leu 69 (Cγ, Hγ) in CFP-10 and Ile 25 (Cγ) in ESAT-6 (98.5% of the complex). Complete proton assignments were obtained for the aromatic ring signals from His, Trp and Tyr, but for four of the five Phe residues the Hζ signals could not be identified.

The comprehensive ^{15}N, ^{13}C and ^{1}H resonance assignments obtained for CFP-10 and ESAT-6 together in the complex have been deposited at the BioMagResBank database (accession number 5680).

Acknowledgements

This work was initially supported by the award of a CASE studentship to Philip Renshaw from the BBSRC/VLA. Recent support has been provided by a project grant from the Wellcome Trust (066047).

References

Bartels, C., Xia, T.-H., Billeter, M., Güntert, P. and Wüthrich, K. (1995) *J. Biomol. NMR*, **5**, 1–10.

Bax, A. (1994) *Curr. Opin. Struct. Biol.*, **4**, 738–744.

Behr, M.A., Wilson, M.A., Gill, W.P., Salamon, H., Schoolnik, G.K., Rane, S. and Small, P.M. (1999) *Science*, **284**, 1520–1523.

Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S.V. et al. (1998) *Nature*, **393**, 537–544.

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

Hsu, T., Hinigley-Wilson, S.M., Chen, B., Chen, M., Dai, A.Z., Morin, P.M., Marks, C.B., Padiyar, J., Goulding, C., Gingery, M., Eisenberg, D., Russell, R.G., Derrick, S.C., Collins, F.M., Morris, S.L., King, C.H. and Jacobs, Jr., W.R. (2003) *Proc. Natl. Acad. Sci. USA*, **100**, 12420–12425.

Pym, A.S., Brodin, P., Brosch, R., Huerre, M. and Cole, S.T. (2002) *Mol. Microbiol.*, **46**, 709–717.

Renshaw, P.S., Panagiotidou, P., Whelan, A., Gordon, S.V., Hewinson, R.G., Williamson, R.A. and Carr, M.D. (2002) *J. Biol. Chem.*, **277**, 21598–21603.

Wards, B.J., de Lisle, G.W. and Collins, D.M. (2000) *Tubercle Lung Dis.*, **80**, 185–189.