Journal of Biomolecular NMR **29:** 101–102, 2004.

Letter to the Editor: NMR assignment of the hypothetical protein HI0004 from *Haemophilus influenzae* – a putative essential gene product

Deok Cheon Yeh, James F. Parsons, Lisa M. Parsons, Fang Liu, Edward Eisenstein & John Orban*

Center for Advanced Research in Biotechnology, University of Maryland, Biotechnology Institute, 9600 Gudelsky Drive, Rockville, MD 20850, U.S.A.

Received 15 September 2003; Accepted 31 October 2003

Key words: Haemophilus influenzae, NMR assignment, structural genomics

Biological context

Haemophilus influenzae is a gram negative bacterium which is known to cause meningitis, sepsis, upper respiratory infections, and inner ear infections, particularly in small children. Its 1.83 Mb genome is relatively small, consisting of approximately 1700 open reading frames, and was the first free-living organism to have its genome sequenced entirely (Fleischmann et al., 1995). At the time the genome sequence was released, approximately 32% of the gene products had unknown or poorly understood functions. A significant fraction of these have sequence homologues in prokaryotes, archaea, and eukaryotes, and many are still classified as hypothetical proteins. In an effort to better understand the biochemical functions of these hypothetical proteins, we have been involved in a concerted program to solve their three dimensional structures using X-ray crystallography or NMR spectroscopy (Eisenstein et al., 2000). A summary of the current project status can be found at http://s2f.umbi.umd.edu

One such protein of interest is the 154 residue HI0004 which has over 150 sequence homologues in other organisms but no known function or three-dimensional structure. Mutagenesis experiments indicate a putative essential role for 259 gene products of unknown function in *Haemophilus influenzae* and HI0004 is in this set, making it an important target for structural and functional annotation (Akerley et al., 2002). Further, the HI0004 family of sequence relatives contains a human homologue, C21orf57, which is a putative transcript located on chromosome 21 near the Down syndrome locus (Reymond et al., 2001). Structure determination of HI0004 will therefore allow homology modeling of a large number of sequences in other organisms and provide the basis for an understanding of their biochemical functions at a molecular level. We report here the ^{1}H, ^{13}C, and ^{15}N chemical shift assignments for HI0004, providing the first step toward detailed structure calculations in solution.

*To whom correspondence should be addressed. E-mail: orban@umbi.umd.edu

Methods and experiments

The native HI0004 gene was cloned into a pET-28a vector using standard procedures and transformed into *Escherichia coli* BL21 (DE3) cells (Novagen). Uniformly $^{13}C/^{15}N$-labeled samples of HI0004 were prepared by growing these cells in minimal media with $^{15}NH_4Cl$ and $^{13}C_6$-glucose as the sole nitrogen and carbon sources, respectively. The media were further supplemented with $ZnCl_2$ (2 mg/L), $MnSO_4$ (1 mg/L), $CoCl_2$ (0.1 mg/L), $CuSO_4$ (0.1 mg/L) and thiamine (1 mg/L). *E. coli* cells were grown to an A_{600} of 0.8 and protein expression was induced with 1 mM IPTG. After an additional 3 h, the cells were harvested, suspended in lysis buffer (0.1 M Tris, 1 mM EDTA, 1 mM DTT, pH 8.0), and lysed by sonication. The lysed cells were centrifuged and ammonium sulfate was added to the supernatant to a 40% saturation level. The resulting precipitate was removed by centrifugation and further ammonium sulfate was added to the supernatant to give a 70% saturation level. The precipitate was collected by centrifugation, dissolved in lysis buffer, and dialysed against a solution of 25 mM HEPES, 1 mM EDTA, and 1 mM DTT at pH 7.2. The desalted solution was loaded on a Poros HQ-50 anion exchange column (Applied Biosystems) and eluted with a 0-800 mM NaCl gradient. The homogeneity of individual fractions was determined by SDS-PAGE

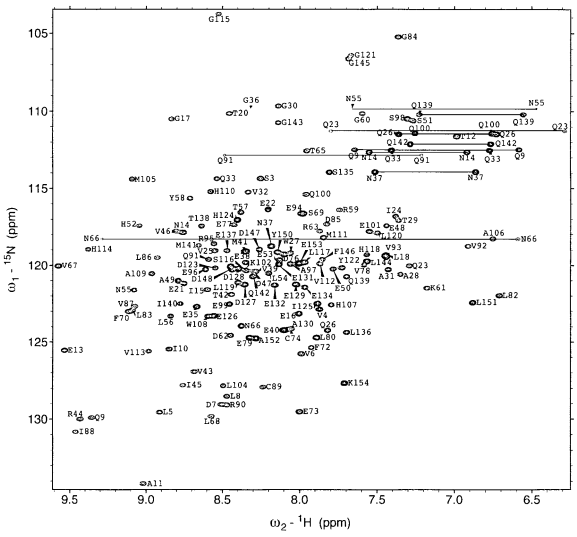

Figure 1. Two-dimensional ^{15}N-HSQC spectrum of HI0004 at 298K and pH 7.0. The downfield peaks due to the indole protons of W27 and W108 are not shown.

and those containing pure HI0004 were combined, dialysed against NMR buffer (50 mM potassium phosphate, 100 mM NaCl, 0.5 mM EDTA, 1 mM DTT, pH 7.0) and concentrated.

Both ^{15}N- and ^{15}N/^{13}C-labeled NMR samples of HI0004 were prepared at concentrations of approximately 1.5 mM in 450 μL of NMR buffer containing 90% H_2O/10% D_2O. The following experiments were acquired: two-dimensional ^{15}N-HSQC,^{13}C-HSQC, and (HB)CB(CGCD)HD/(HB)CB(CGCD)HE, and three-dimensional HNCACB, CBCA(CO)NH, HNCO, HNCA, HBHA(CO)NH, H(CCO)NH-TOCSY, (H)C(CO)NH-TOCSY, HCCH-TOCSY, HCCH-COSY, ^{15}N-NOESY HSQC, and ^{13}C-NOESY HSQC. In addition, a constant time ^{13}C-HSQC spectrum was recorded on a 15% uniformly ^{13}C-labeled sample (1.0 mM) to determine stereospecific assignments of methyl groups in Leu and Val residues (Neri et al., 1989). The methyl groups of 12 out of 14 Val residues and all 17 Leu residues were stereospecifically assigned in this way. With the exception of the ^{15}N-NOESY HSQC, the NMR spectra were collected in a 7 day period at the NMR Facility at Madison (NMRFAM, Wisconsin) on a Bruker DRX-500 equipped with a z-gradient triple resonance cryoprobe. All experiments were recorded at 298K. Spectra were processed using nmrPipe (Delaglio et al., 1995) and analyzed with Sparky (Goddard and Kneller, UCSF).

Extent of assignments and data deposition

Sequence specific backbone assignments were made for 143 of the 145 possible main chain crosspeaks in the ^{15}N-HSQC spectrum of HI0004 (154 residues minus 8 prolines and the N-terminal Met) as illustrated in Figure 1. Residues for which backbone H_N/N assignments were not made are G2 and M133. The chemical shift index (CSI) results suggest that HI0004 contains four beta strands and five helices. Interestingly, the number of these secondary structure elements and their locations in the polypeptide chain from CSI are in very good agreement with sequence-based predictions from Psipred (Jones, 1999). Assignments were also made for about 90% of the possible ^{1}H, ^{13}C, and ^{15}N side chain resonances. The chemical shifts of HI0004 have been deposited in the BioMagResBank database (http://www.bmrb.wisc.edu) under BMRB accession number 5942.

Acknowledgements

We are grateful to NMRFAM at Madison, Wisconsin, for access to their cryoprobe facility and to Dr Frits Abildgaard in particular for assistance with pulse programs. This research was supported by NIH grants GM57890 and 1S10RR15744, and the W. M. Keck Foundation.

References

Akerley, B.J., Rubin, E.J., Novick, V.L., Amaya, K., Judson, N. and Mekalanos, J.J. (2002) *Proc. Natl. Acad. Sci. USA*, **99**, 966–971.

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

Eisenstein, E., Gilliland, G.L., Herzberg, O., Moult, J., Orban, J., Poljak, R.J., Banerjei, L., Richardson, D. and Howard, A.J. (2000) *Curr. Opin. Biotechnol.*, **11**, 25–30.

Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirkness, E.F., Kerlavage, A.R., Bult, C.J., Tomb, J.-F., Dougherty, B.A., Merrick, J.M., McKenney, K., Sutton, G., FitzHugh, W., Fields, C., Gocayne, J.D., Scott, J., Shirley, R., Liu, L.-I., Glodek, A., Kelley, J.M., Weidman, J.F., Phillips, C.A., Spriggs, T., Hedblom, E., Cotton, M.D., Utterback, T.R., Hanna, M.C., Nguyen, D.T., Saudek, D.M., Brandon, R.C., Fine, L.D., Fritchman, J.L., Fuhrmann, J.L., Geoghagen, N.S.M., Gnehm, C.L., McDonald, L.A., Small, K.V., Fraser, C.M., Smith, H.O. and Venter, J.C. (1995) *Science*, **269**, 496–512.

Jones, D.T. (1999) *J. Mol. Biol.*, **292**, 195–202.

Neri, D., Szyperski, T., Otting, G., Senn, H. and Wüthrich, K. (1989) *Biochemistry*, **28**, 7510–7516.

Reymond, A., Friedli, M., Henrichsen, C.N., Chapot, F., Deutsch, S., Ucla, C., Rossier, C., Lyle, R., Guipponi, M. and Antonarakis, S.E. (2001) *Genomics*, **78**, 46–54.