NMR structure of HI0004, a putative essential gene product from *Haemophilus influenzae*, and comparison with the X-ray structure of an *Aquifex aeolicus* homolog

DEOK CHEON YEH, LISA M. PARSONS, JAMES F. PARSONS, FANG LIU,
EDWARD EISENSTEIN, AND JOHN ORBAN

Center for Advanced Research in Biotechnology, University of Maryland Biotechnology Institute,
Rockville, Maryland 20850, USA

(RECEIVED September 2, 2004; FINAL REVISION November 3, 2004; ACCEPTED November 4, 2004)

Abstract

The solution structure of the 154-residue conserved hypothetical protein HI0004 has been determined using multidimensional heteronuclear NMR spectroscopy. HI0004 has sequence homologs in many organisms ranging from bacteria to humans and is believed to be essential in *Haemophilus influenzae*, although an exact function has yet to be defined. It has a α - β - α sandwich architecture consisting of a central four-stranded β -sheet with the α 2-helix packed against one side of the β -sheet and four α -helices (α 1, α 3, α 4, α 5) on the other side. There is structural homology with the eukaryotic matrix metalloproteases (MMPs), but little sequence similarity except for a conserved region containing three histidines that appears in both the MMPs and throughout the HI0004 family of proteins. The solution structure of HI0004 is compared with the X-ray structure of an *Aquifex aeolicus* homolog, AQ_1354, which has 36% sequence identity over 148 residues. Despite this level of sequence homology, significant differences exist between the two structures. These differences are described along with possible functional implications of the structures.

Keywords: *Haemophilus influenzae*; NMR; structural genomics; MMP; hydrolase

Structural genomics involves the determination of three-dimensional structures of proteins on a genome-wide scale. The targets for such structure determination efforts can vary from proteins with well-defined functions to “hypothetical” proteins for which functions have yet to be found. In the case of hypothetical proteins, the structure can often provide insights into the possible function since protein folds are more highly conserved than sequences (e.g., Lim et al.

2001; Parsons et al. 2004). Most of the targets chosen for structural genomics typically have sequence homologs in other organisms, and therefore the structures obtained can be used to model the folds of relatively large sequence families for which there had previously been no information. The current structural genomics project at our institute has focused on hypothetical proteins from the human pathogen *Haemophilus influenzae* (Eisenstein et al. 2000; Parsons and Orban 2004). This microbial organism was the first to have its genome completely sequenced (Fleischmann et al. 1995) and is responsible for a number of ailments including meningitis, ear infections, and upper respiratory problems.

As part of our structural genomics effort, we describe here the solution NMR structure of the 154-residue protein HI0004 from *Haemophilus influenzae*. HI0004 belongs to a large family of more than 150 protein sequences with homologs spanning organisms from bacteria to humans. Re-

Reprint requests to: John Orban, Center for Advanced Research in Biotechnology, University of Maryland Biotechnology Institute, 9600 Gudelsky Drive, Rockville, MD 20850, USA; e-mail: orban@umbi.umd.edu; fax: (301) 738-6255.

Abbreviations: NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; NOESY, NOE spectroscopy; EDTA, ethylenediaminetetraacetic acid; HSQC, heteronuclear single quantum coherence; RMSD, root mean square deviation.

Article published online ahead of print. Article and publication date are at <http://www.proteinscience.org/cgi/doi/10.1110/ps.041096705>.

cent mutagenesis experiments in *Haemophilus influenzae* indicate that HI0004 is an essential gene (Akerley et al. 2002), but no function has yet been assigned. The structural analysis presented here provides insights into the function of this protein. In addition, the NMR structure is compared with the recent X-ray structure of a homolog from *Aquifex aeolicus* with 36% sequence identity (Oganesyan et al. 2003). A significant difference in the position of one of the helical elements is observed between the two structures despite the relatively high sequence similarity. The possible functional relevance of these structural differences is discussed.

Results and Discussion

Structure determination and statistics

Chemical shift assignments for HI0004 have been described in detail previously and are deposited in the BioMagRes-Bank (accession no. 5942) (Cheon Yeh et al. 2004). A total of 1845 interresidue NOE restraints were used for the structure calculation. The structure statistics are summarized in Table 1 and structural data are shown in Figure 1. On average, approximately 13 interresidue NOE restraints were obtained for each amino acid from residues 4–146. In addition, 252 dihedral angle restraints (ϕ and ψ) were obtained from an analysis of chemical shift data using TALOS (Cornilescu et al. 1999). Hydrogen bond restraints were also used, but these were only employed at the later stages of refinement where the identity of hydrogen bond donor–acceptor pairs was readily apparent. Residual dipolar couplings (RDCs) provided an independent verification of the

Figure 1. Summary of structure and dynamics data for HI0004. (A) Distribution of long range (filled columns), medium range (shaded columns), and sequential (unfilled columns) restraints per residue. (B) Plot of ^{15}N - R_2 relaxation parameters for each residue. (C) The steady-state ^1H - ^{15}N nuclear Overhauser effect (NOE). (D) Backbone RMSDs (N, C_α , CO) per residue with the secondary structure shown beneath the graph.

Table 1. Statistics for the ensemble of 20 structures of HI0004

A. Experimental restraints		
NOE restraints		
All NOE		1845
Sequential ($ i - j = 1$)		806
Medium range ($1 < i - j \leq 5$)		454
Long range ($ i - j > 5$)		585
Hydrogen bond		98
Dihedral angle restraints		252
Total restraints		2195
B. RMSDs to mean structure (Å)		
Backbone atoms (N, C_α , CO) ^a	0.96 ± 0.22	
All non-hydrogen atoms ^a	1.58 ± 0.24	
Secondary structures	0.67 ± 0.23	
C. Measures of structure quality ^b		
Ramachandran distribution		
Most favored (%)	78.6 ± 1.0	
Additionally allowed (%)	21.4 ± 1.0	
Bad contacts/100 residues	2.0 ± 1.4	
Overall dihedral G factor	-0.04 ± 0.02	

^a For residues 4–146.

^b Analyzed using Procheck (Laskowski et al. 1966).

solution structure (Bax 2003). The overall backbone RMSD (residues 4–146) in the final ensemble of 20 structures is 0.96 Å, and the backbone RMSD in the secondary structure regions is 0.67 Å.

Solution structure and backbone dynamics of HI0004

HI0004 is a monomer in solution based on its mobility during size exclusion chromatography and its NMR line-width characteristics. The solution structure has a α - β - α sandwich architecture consisting of a central four-stranded β -sheet with the $\alpha 2$ -helix packed against one side of the β -sheet and four α -helices ($\alpha 1$, $\alpha 3$, $\alpha 4$, $\alpha 5$) on the other side (Fig. 2A,B). The $\alpha 1$ -, $\alpha 4$ -, and $\alpha 5$ -helices make extensive contacts with the β -sheet and each other, whereas the long axis of $\alpha 3$ is oriented almost perpendicularly to the β -sheet. The arrangement of the secondary structure elements is β - α - β - α - β - α - α - α with a β -strand order of 1243. The $\beta 1$ -, $\beta 2$ -, and $\beta 4$ -strands are parallel and $\beta 3$ is antiparallel.

Figure 2. Comparison of the HI0004 NMR structure with the AQ_1354 X-ray structure. (A) Backbone superposition of the 20 lowest energy structures of HI0004 determined by NMR spectroscopy. (B) Ribbon diagram of HI0004 with annotated secondary structure elements. The $\alpha 5$ -helix is highlighted in yellow, and the conserved histidine side chains, H114 and H118 in $\alpha 4$ and H124 in the $\alpha 4$ - $\alpha 5$ loop, are shown in blue. (C) X-ray structure of AQ_1354 (Oganesyan et al. 2003) showing the position of the three conserved histidine residues, H115, H119, and H125, in blue. The $\alpha 5$ -helix is highlighted in yellow.

The most defined regions correspond with secondary structure elements while the loop regions tend to be more conformationally flexible. Loops $\alpha 1$ - $\beta 2$, $\beta 3$ - $\beta 4$, and $\alpha 4$ - $\alpha 5$ are particularly mobile based on ^{15}N -transverse relaxation rates (Fig. 1B) and $\{^1\text{H}\}$ - ^{15}N steady-state NOE measurements (Fig. 1C). Short segments of secondary structure at the N terminus of $\beta 2$ and the first turn of $\alpha 5$ also exhibit motions on the nanosecond-to-picosecond timescale. In addition, the C-terminal residues 147–154 and the N-terminal residues 1–3 are disordered and have few detectable inter-residue NOE contacts (Fig. 1A).

Structural homologs of HI0004

An initial search for structural homologs of HI0004 using CE (Shindyalov and Bourne 1998) revealed that the closest structures were matrix metalloprotease (1mmr, Z-score 4.2) and human fibroblast collagenase (1ayk, Z-score 4.2). These proteins are both zinc-dependent matrix metalloproteases (MMPs), share a similar β -sheet topology with HI0004, and have comparable locations for two of their helices. The C_α superposition of these structures with HI0004 gave RMSDs of 3.9 Å. The structural homology with MMPs suggests a hydrolase function for HI0004. Although there are structural similarities, very little sequence homology (8%–14%) exists between the eukaryotic MMPs and HI0004. The exception is a conserved region containing three His residues that appears in both the MMPs and throughout the HI0004 family of proteins (Fig. 3). In the MMP structures, the three histidines are close to each other and bind Zn^{2+} with tetrahedral coordination, the fourth ligand being a water molecule. The catalytic zinc in the active site is essential for full protease activity. In the structure of HI0004, two of the histidines, H114 and H118, are proximal but the third, H124, is over 10 Å away (Fig. 2B). However,

it must be emphasized that the NMR structure of HI0004 was determined in the absence of zinc ions due to the use of EDTA in the purification buffers (Cheon Yeh et al. 2004).

Recently, an X-ray structure was determined for the *Aquifex aeolicus* homolog of HI0004, AQ_1354 (36% sequence identity over 148 residues), also in the absence of zinc ions (Oganesyan et al. 2003). To our surprise, a significant difference exists between the structures of HI0004 and AQ_1354 despite the level of sequence identity. These differences are discussed below.

Comparison with the X-ray structure of AQ_1354 from *Aquifex aeolicus*

The HI0004 and AQ_1354 structures both have the same number and order of secondary structure elements (Fig. 2B,C). The arrangement of the four β -strands is the same and four of the five α -helices are positioned similarly. Superposition of the C_α atoms of HI0004 and AQ_1354 gives a RMSD of 3.4 Å (Z-score 5.3). However, there is an obvious difference in the position of the $\alpha 5$ -helix. In the HI0004 NMR structure, $\alpha 5$ is oriented at an angle of about 30° with respect to the β -strands, is between the $\alpha 1$ - and $\alpha 4$ -helices, and has some contacts to the β -sheet (Fig. 2B). By contrast, the $\alpha 5$ -helix in the AQ_1354 X-ray structure is approximately perpendicular to the β -strands, packing on top of $\alpha 1$ and $\alpha 4$, and making no contacts with the β -sheet (Fig. 2C). The position of the $\alpha 5$ -helix in HI0004 is well-defined and is constrained by more than 40 long-range reciprocal NOE contacts. These NOEs are inconsistent with the conformation of the $\alpha 5$ -helix seen in the X-ray structure.

Other differences are also noteworthy. The $\alpha 1$ -helix is shorter in the NMR structure of HI0004 (10 residues vs. 16 in AQ_1354) and is also oriented at about 80° relative to $\alpha 4$, allowing $\alpha 5$ to pack between these two helices. In

Figure 3. Alignment of representatives in the HI0004 sequence family. Invariant residues are shown in dark columns and conserved residues are shown in gray columns. The secondary structure elements found in HI0004 are depicted at the top of the figure. Sequence alignment was carried out with CLUSTALW (Thompson et al. 1994) and the figure was generated using ESPript (<http://prodes.toulouse.inra.fr/>).

AQ_1354, the $\alpha 1$ - and $\alpha 4$ -helices are more closely packed and almost parallel. Further, the $\beta 1$ - $\alpha 1$ loop is longer in HI0004 (10 residues vs. 6 in AQ_1354) as is the $\alpha 1$ - $\beta 2$ loop (6 residues vs. 2 in AQ_1354). The orientation of $\alpha 2$, the lone helix on the other side of the β -sheet, is more parallel with the β -strands in the NMR structure of HI0004. The $\beta 3$ - and $\beta 4$ -strands are shorter in HI0004, leaving a longer, flexible loop (10 residues vs. 4 in AQ_1354) connecting these strands.

Since both the X-ray and NMR structures were determined in the absence of zinc, it is not yet clear why there is such a difference in their structures. This may be due to the different techniques used for structure determination or there may be sufficient differences in the sequences of HI0004 and AQ_1354 to cause the structural variance. For example, A109 is located in the $\alpha 4$ -helix of HI0004, and its buried methyl group has extensive hydrophobic contacts with A28, I140, and F146. These interactions help to constrain the position of the $\alpha 5$ -helix between $\alpha 4$ and $\alpha 1$ in the NMR structure. In AQ_1354, the corresponding residue to A109 is K110, and this would be expected to disrupt an equivalent set of hydrophobic interactions for steric and electronic reasons. Other residues may also possibly contribute to the structural differences.

Binding of Zn^{2+} to HI0004

Since the structurally related MMPs are zinc-dependent, we examined the effect of Zn^{2+} on HI0004 using chemical shift

perturbation. Binding of zinc to the three conserved histidine residues in the HI0004 NMR structure (H114, H118, and H124) would require a conformational change, since H124, located in the $\alpha 4$ - $\alpha 5$ loop, is well removed from the other two conserved histidines in the $\alpha 4$ -helix, H114, and H118 (Fig. 2B). In contrast, the corresponding three conserved histidines in the X-ray structure of AQ_1354 (H115, H119, and H125) are in close proximity so that little change from this conformation would be needed for zinc binding (Fig. 2C). Figure 4A shows that significant shift changes were detected for some peaks in the ^{15}N HSQC spectrum of HI0004 upon zinc addition. Binding of zinc resulted in two distinct sets of peaks in the HSQC spectrum corresponding with the free and bound forms of HI0004 in slow exchange on the chemical shift time scale. The HSQC spectrum did not change beyond the addition of 1.4 equivalents of Zn^{2+} . Since peak positions could not be followed as a function of zinc concentration due to slow exchange, backbone resonances were assigned for the zinc-bound form of HI0004 using standard methods (Grzesiek and Bax 1992a,b). The chemical shift changes are shown as a function of sequence in Figure 4B and are mapped onto the HI0004 structure in Figure 4C. Changes in chemical shift occur for the conserved histidines H114, H118, and H124 and adjacent residues as might be expected. However, shift perturbations are also observed in the $\alpha 4$ - $\alpha 5$ loop, $\alpha 5$ -helix, $\alpha 1$ -helix, and β -sheet. Indeed, many of the residues with shift changes either make contacts to or are in the $\alpha 5$ -helix or $\alpha 4$ - $\alpha 5$ loop. In addition, a large change is observed in the main chain C_{α}

Figure 4. Effect of zinc addition on HI0004. (A) Overlaid ${}^{15}\text{N}$ HSQC spectra of free (red) and zinc-bound (green) HI0004. (B) Histogram plot of total chemical shift change (from equation 1) as a function of residue number. (C) Ribbon diagram showing the residues with the largest chemical shift changes ($\Delta\delta_{\text{tot}} \geq 0.2$ ppm) mapped onto the NMR structure of HI0004 in yellow.

shift of P149, which is in the disordered C-terminal tail of HI0004. The shift perturbation results suggest that a structural change involving the $\alpha 4$ – $\alpha 5$ loop and $\alpha 5$ -helix may occur upon zinc binding.

We investigated the possibility that the zinc-induced changes in HI0004 result in a conformation similar to that seen in the X-ray structure of AQ_1354. Main chain amide RDCs (${}^1D_{\text{NH}}$) were measured for zinc-free HI0004, and these experimental values were compared with the expected values calculated for both the NMR structure of HI0004 (Fig. 5A) and also for a model of HI0004 based on the X-ray structure of AQ_1354 (Fig. 5B). From an inspection of these figures it can be seen that the experimental ${}^1D_{\text{NH}}$ values of zinc-free HI0004 are more consistent with the NOE-derived structure than with a model based on the X-ray data. Next, ${}^1D_{\text{NH}}$ RDCs were measured for zinc-bound HI0004, and these were compared with calculated ${}^1D_{\text{NH}}$ values for the NMR structure of HI0004 (Fig. 5C) and for the HI0004 model based on the X-ray structure (Fig. 5D). In this case, the zinc-bound conformation of HI0004 appears to be more in agreement with a model based on the X-ray structure. The RDC results are therefore consistent with the chemical shift perturbation data and with the idea that binding of zinc to HI0004 causes a reorientation of the $\alpha 5$ -helix from its position in the NMR structure to a conformation similar to that observed in the X-ray structure of AQ_1354.

Functional implications

In further comparing the structures of HI0004 and AQ_1354, we noticed in the X-ray structure of AQ_1354 an

Figure 5. Main chain amide RDCs determined experimentally (D^{meas}) vs. model-dependent calculated values (D^{calc}). The diagonal line in each panel shows where the 1:1 correspondence between D^{meas} and D^{calc} would be. RDC data from $\alpha 5$ -helical residues are shown as open circles. (A) D^{meas} for HI0004 (no zinc) vs. the corresponding D^{calc} values obtained from the NOE-derived NMR structure of HI0004 (Q-factor 31%). (B) D^{meas} for HI0004 (no zinc) vs. D^{calc} values from a model of HI0004 based on the X-ray structure, AQ_1354 (Q-factor 54%). (C) D^{meas} for HI0004 with zinc vs. the D^{calc} values obtained from the NMR structure of HI0004 (Q-factor 97%). (D) D^{meas} for HI0004 with zinc vs. D^{calc} obtained using a model of HI0004 based on the X-ray structure, AQ_1354 (Q-factor 47%).

extensive solvent-accessible surface of hydrophobic residues immediately adjacent to the putative active site (Fig. 6A). Many of the residues in this hydrophobic surface (V41, F72, P73, L84, L120, L121, G122, Y123, and L138) are conserved in HI0004 as well as in numerous other members of the sequence family (Fig. 3; see also <http://s2f.umbi.umd.edu/>). Most residues in the hydrophobic patch of AQ_1354 are buried in the NMR structure of HI0004 (Fig. 6B), forming part of the surface that packs against the amphipathic α 5-helix. Consequently, there is only a small region of corresponding exposed hydrophobic surface for HI0004. Based on the RDC results, the conformation of the zinc-bound form of HI0004 is likely similar to the X-ray structure of AQ_1354, and this is probably the active conformation. Therefore, it appears that zinc-binding to HI0004 could serve at least two functions: It may provide the catalytic site metal ion required for enzymatic activity and may also promote the conformational change needed to expose the hydrophobic patch. This solvent-accessible hydrophobic surface is likely to be biologically relevant for molecular recognition since it is both well conserved and next to the putative active site. Although the specific enzymatic function of these proteins is still unknown (Oganessian et al. 2003), the structural analysis presented here suggests that the substrate(s) for AQ_1354 and HI0004 may have a significant hydrophobic component. Moreover, it has been noted that homologs of HI0004 are found on a short operon together with diacylglycerol kinase in other organisms (Chen et al. 1998), raising the possibility that proteins in this family may act on intermediates in membrane lipid biosynthetic pathways. Further work will be needed to test these ideas.

Materials and methods

NMR spectroscopy

Sample preparation and experiments used for NMR assignment have been described previously (Cheon Yeh et al. 2004). Spectra

were recorded at 300 K on a Bruker DRX 500-MHz spectrometer with a cryoprobe or a Bruker DRX-600 equipped with a three-axis gradient probe. Pulsed-field gradients were used for coherence selection and solvent suppression. Data were processed on a Linux workstation using NMRPipe (Delaglio et al. 1995) and analyzed using Sparky (T.D. Goddard and D.G. Kneller, UCSF). Proton chemical shifts were directly referenced to the methyl resonance of 2,2-dimethyl-2-silapentane-5-sulfonate (DSS), while ^{13}C and ^{15}N chemical shifts were indirectly referenced.

^{15}N -transverse relaxation rates (R_2) were obtained from 10 CPMG experiments, with relaxation delays ranging from 16 msec to 144 msec. Experiments were acquired in two single interleaved matrices to ensure uniformity of experimental conditions. Steady-state $\{^1\text{H}\}$ - ^{15}N NOEs were determined from the peak intensity ratio of two experiments, with and without saturation. The relaxation delay was 2 sec, with 3 sec of saturation for NOE experiments. A reference spectrum was acquired with a 5-sec relaxation delay. For titration experiments, a zinc chloride solution was added to HI0004 in 10 mM Tris buffer (pH 7.0), in steps of 0.2 equivalents up to 2 mol equivalents and a ^{15}N HSQC spectrum was recorded after each addition. The total chemical shift change upon zinc addition was calculated using equation 1

$$\Delta\delta_{\text{tot}} = ((\Delta\delta_{\text{HN}}W_{\text{HN}})^2 + (\Delta\delta_{\text{N}}W_{\text{N}})^2 + (\Delta\delta_{\text{C}\alpha}W_{\text{C}\alpha})^2)^{1/2} \quad (1)$$

with the following weighting factors $W_{\text{HN}} = 1$, $W_{\text{N}} = 0.154$, and $W_{\text{C}\alpha} = 0.276$ (Ayed et al. 2001).

Main chain amide RDCs ($^1D_{\text{NH}}$) were measured using a ^1H - ^{15}N IPAP-HSQC experiment (Ottiger et al. 1998) with Pf1 phage as the alignment medium (Hansen et al. 1998). Resolved RDCs of all NH bond vectors in well-defined areas of secondary structure were used in combination with the high resolution NOE-derived structure to calculate the axial and rhombic components of the alignment tensor using PALES (Zweckstetter and Bax 2000). Theoretical RDCs were calculated using the singular value decomposition method (Losonczi et al. 1999) with either the NMR structure of HI0004 or a homology model of HI0004 based on the X-ray structure of AQ_1354 as inputs. Q-factors were determined as described previously (Bax 2003).

Structure calculations

Interproton distance restraints were obtained from 3D ^{15}N -NOESY (τ_m 100 msec) and 3D ^{13}C -NOESY (τ_m 100 msec) spectra run on a sample in water and a homonuclear 2D NOESY (τ_m 120 msec) spectrum collected on a sample in D_2O . Distance restraints were divided into five classes based on peak intensities: very strong (1.8–2.7 Å), strong (1.8–3.1 Å), medium (1.8–3.5 Å), weak (2.3–4.2 Å), and very weak (2.8–5.0 Å). For equivalent protons or nonstereospecifically assigned protons, pseudo-atoms were introduced. Prochiral groups were given floating stereospecific assignments until they could be unambiguously assigned from the structure. Backbone dihedral angle restraints were identified using TALOS (Cornilescu et al. 1999) and chemical shift data. Amide hydrogen exchange was investigated by recording ^{15}N HSQC spectra of lyophilized protein freshly dissolved in D_2O . Hydrogen bond restraints were derived using standard criteria on the basis of the amide deuterium exchange experiment, chemical shifts, and NOE data (Parsons et al. 2001). Structures were calculated initially from an extended polypeptide chain using CNS 1.1 (Brünger et al. 1998) and standard simulated annealing and torsion

Figure 6. Comparison of the AQ_1354 and HI0004 surfaces highlighting the hydrophobic patch (purple) and invariant histidine residues (blue). The view in both panels is from the left of the structures shown in Figure 2, B and C ($\sim 90^\circ$ rotation along the Y-axis). (A) Molecular surface of AQ_1354. Conserved residues in the hydrophobic patch are labeled. The conserved histidine, H125, is obscured in this view. (B) Molecular surface of HI0004 showing the diminished hydrophobic patch due to the altered packing of the α 5-helix.

angle dynamics protocols. Final force constants were 1000 kcal mol⁻¹ Å⁻² for bond angles, 500 kcal mol⁻¹ rad⁻² for angles and improper torsions, 40 kcal mol⁻¹ Å⁻² for experimental distance restraints, 200 kcal mol⁻¹ rad⁻² for dihedral angle restraints, and 4 kcal mol⁻¹ Å⁻⁴ for the van der Waals repulsion term. The 20 best structures were chosen based on a low total energy, no NOE distance violations greater than 0.35 Å, fewer than 10 bad contacts per 100 residues, and a dihedral angle G factor greater than -0.5. Structures were analyzed with PROCHECK (Laskowski et al. 1996) and QUANTA (Molecular Simulations Inc.) and displayed with PyMOL (DeLano Scientific).

Protein Data Bank accession code

Atomic coordinates for the ensemble of 20 structures have been deposited in the RCSB Protein Data Bank (accession code 1xax).

Acknowledgments

This research was supported by NIH grants GM57890 and 1S10RR15744, and the W.M. Keck Foundation. We thank Eugene Melamud for generating the model of HI0004 based on the X-ray structure of AQ₁₃₅₄.

References

- Akerley, B.J., Rubin, E.J., Novick, V.L., Amaya, K., Judson, N. and Mekalanos, J.J. 2002. A genome-scale analysis for identification of genes required for growth or survival of *Haemophilus influenzae*. *Proc. Natl. Acad. Sci./ITL* **99**: 966–971.
- Ayed, A., Mulder, F.A., Yi, G.S., Lu, Y., Kay, L.E., and Arrowsmith, C.H. 2001. Latent and active p53 are identical in conformation. *Nat. Struct. Biol.* **8**: 756–760.
- Bax, A. 2003. Weak alignment offers new NMR opportunities to study protein structure and dynamics. *Protein Sci.* **12**: 1–16.
- Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse, K.R., 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.
- Chen, P., Novak, J., Qi, F., and Caufield, P.W. 1998. Diacylglycerol kinase is involved in regulation of expression of the antibiotic mutacin II of *Streptococcus mutans*. *J. Bacteriol.* **180**: 167–170.
- Cheon Yeh, D., Parsons, J.F., Parsons, L.M., Liu, F., Eisenstein, E., and Orban, J. 2004. NMR assignment of the hypothetical protein HI0004 from *Haemophilus influenzae*—A putative essential gene product. *J. Biomol. NMR* **29**: 101–102.
- 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.
- Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., and Bax, A. 1995. NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**: 277–293.
- Eisenstein, E., Gilliland, G.L., Herzberg, O., Moul, J., Orban, J., Poljak, R.J., Banerjee, L., Richardson, D., and Howard, A.J. 2000. Biological function made crystal clear—Annotation of hypothetical proteins via structural genomics. *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., et al. 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269**: 496–512.
- Grzesiek, S. and Bax, A. 1992a. An efficient experiment for sequential backbone assignment of medium-sized isotopically enriched proteins. *J. Magn. Reson.* **99**: 201–207.
- . 1992b. Improved 3D triple-resonance NMR techniques applied to a 31 kDa protein. *J. Magn. Reson.* **96**: 432–440.
- Hansen, M.R., Mueller, L., and Pardi, A. 1998. Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. *Nat. Struct. Biol.* **5**: 1065–1074.
- 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.
- Lim, K., Zhang, H., Tempczyk, A., Bonander, N., Toedt, J., Howard, A., Eisenstein, E., and Herzberg, O. 2001. Crystal structure of YecO from *Haemophilus influenzae* (HI0319) reveals a methyltransferase fold and a bound S-adenosylhomocysteine. *Proteins* **45**: 397–407.
- Losonczi, J.A., Andrec, M., Fischer, M.W.F., and Prestegard, J.H. 1999. Order matrix analysis of residual dipolar couplings using singular value decomposition. *J. Magn. Reson.* **138**: 334–342.
- Oganesyan, V., Busso, D., Brandsen, J., Chen, S., Jancarik, J., Kim, R. and Kim, S.H. 2003. Structure of the hypothetical protein AQ₁₃₅₄ from Aquifex aeolicus. *Acta Crystallogr. D Biol. Crystallogr.* **59**: 1219–1223.
- Ottiger, M., Delaglio, F., and Bax, A. 1998. Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra. *J. Magn. Reson.* **131**: 373–378.
- Parsons, L. and Orban, J. 2004. Structural genomics and the metabolome: Combining computational and NMR methods to identify target ligands. *Curr. Opin. Drug Discov. Dev.* **7**: 62–68.
- Parsons, L., Eisenstein, E., and Orban, J. 2001. Solution structure of HI0257, a bacterial ribosome binding protein. *Biochemistry* **40**: 10979–10986.
- Parsons, L.M., Yeh, D.C., and Orban, J. 2004. Solution structure of the highly acidic protein HI1450 from *Haemophilus influenzae*, a putative double-stranded DNA mimic. *Proteins* **54**: 375–383.
- Shindyalov, I.N. and Bourne, P.E. 1998. Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. *Protein Eng.* **11**: 739–747.
- Thompson, J.D., Higgins, D.G., and Gibson, T.J. 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**: 4673–4680.
- Zweckstetter, M. and Bax, A. 2000. Prediction of sterically induced alignment in a dilute liquid crystalline phase: Aid to protein structure determination by NMR. *J. Am. Chem. Soc.* **122**: 3791–3792.