Journal of Biomolecular NMR, **20:** 105–110, 2001.
KLUWER/ESCOM

Structure of the N-terminal region of *Haemophilus Influenzae* HI0017: Implications for function

Liping Yu*, Jamey Mack, Phil Hajduk & Stephen W. Fesik
Pharmaceutical Discovery Division, D46Y, AP10/LL, Abbott Laboratories, 100 Abbott Park Road, Abbott Park, IL 60064-6098, U.S.A.

Received 28 December 2000; Accepted 21 March 2001

Key words: Haemophilus influenzae, HI0017 protein, NMR structure, structural genomics

Abstract

Haemophilus influenzae is a gram-negative pathogen that causes infections ranging from asymptomatic colonization of the human upper respiratory tract to serious invasive diseases such as meningitis. Although the genome of *Haemophilus influenzae* has been completely sequenced, the structure and function of many of these proteins are unknown. HI0017 is one of these uncharacterized proteins. Here we describe the three-dimensional solution structure of the N-terminal portion of HI0017 as determined by NMR spectroscopy. The structure consists of a five-stranded antiparallel β-sheet and two short α-helices. It is similar to the C-terminal domain of *Diphtheria* toxin repressor (DtxR). The C-terminal portion of HI0017 has an amino acid sequence that closely resembles pyruvate formate-lyase – an enzyme that converts pyruvate and CoA into acetyl-CoA and formate by a radical mechanism. Based on structural and sequence comparisons, we propose that the C-terminus of HI0017 functions as an enzyme with a glycyl radical mechanism, while the N-terminus participates in protein/protein interactions involving an activase (iron-sulfur protein) and/or the substrate.

Introduction

Haemophilus influenzae is a gram-negative bacterium exclusively found in humans that causes infections such as otitis media, sinusitis, bronchitis, and serious invasive meningitis (Turk, 1982, 1984). Thus, *H. influenzae* represents a clinically important organism for drug discovery. The genome of *H. influenzae* has been completely sequenced (Fleischmann et al., 1995), offering the possibility for finding novel protein targets for the development of new antibacterial agents. One potential protein target from *H. influenzae* is HI0017. This protein has 127 amino acids (Fleischmann et al., 1995) and a molecular weight of about 14 kDa. Although its structure and function are unknown, proteins with similar primary sequences have been found in other microorganisms, including *E. coli* (Blattner et al. 1997), *Serratia liquefaciens* (Givskov and Molin, 1992), and Bacteriophage T4 (Valerie et al., 1986). However, proteins homologous to HI0017 have not been found in higher organisms, suggesting that HI0017 might be a good drug target. The only part of the protein that shows high sequence homology with known proteins is the C-terminal 60 residues. This portion of the protein is homologous to pyruvate formate-lyase, a much larger protein that converts pyruvate and CoA into acetyl-CoA and formate by a radical mechanism (Roedel et al., 1988; Sawers and Watson, 1998).

HI0017 is an attractive broad spectrum antibacterial target, because its sequence is highly homologous to those from other microorganisms (Figure 1). Moreover, it is small and amenable to structural studies by NMR spectroscopy. Here we describe the three-dimensional structure of the N-terminal region of HI0017 as determined by NMR. The structure of this protein was compared to previously determined structures of proteins to provide clues about its function.

*To whom correspondence should be addressed. E-mail: Liping.Yu@abbott.com

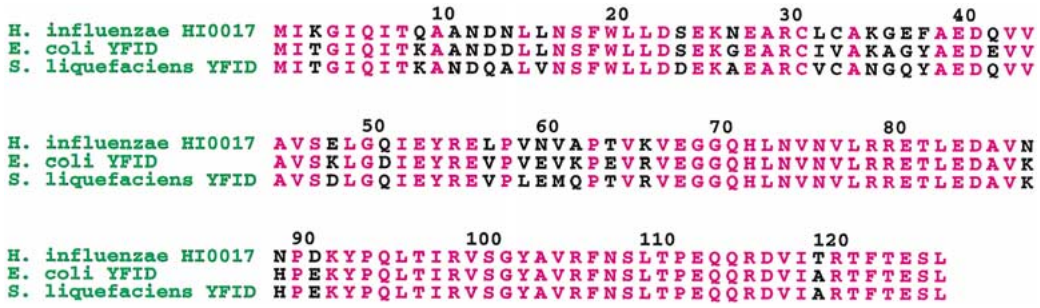

Figure 1. Sequence alignment of *H. influenzae* HI0017, *E. coli* YFID (accession number P33633), and *Serratia liquefaciens* YFID (accession number P18953).

Based on structural and sequence comparisons, we propose a function for HI0017.

Materials and methods

Sample preparation

HI0017 was cloned into pET-21b with a C-terminal His tag and expressed in *E. coli* BL21 (DE3) cells. Uniformly ^{15}N- and ^{15}N,^{13}C-labeled proteins were prepared for the NMR experiments by growing bacteria containing $^{15}NH_4Cl$ with or without [U-^{13}C]-glucose. A uniformly ^{15}N,^{13}C-labeled and fractionally deuterated protein was also prepared by growing cells in 65% $^{2}H_2O$. The labeled recombinant proteins were purified by affinity chromatography on a nickel-IDA column (Invitrogen) followed by ion exchange chromatography on a Q-sepharose column (Pharmacia). Purified proteins were dialyzed and concentrated to 1 mM for NMR studies in a buffer containing 50 mM sodium phosphate (pH 6.7), 50 mM NaCl, and 10 mM DTT.

NMR spectroscopy and structure determination

The NMR spectra were collected at 25 °C on a Bruker DRX600 or DRX800 NMR spectrometer. The ^{1}H, ^{15}N, and ^{13}C resonances of the backbone were assigned with triple resonance experiments (HNCA, HN(CO)CA, HN(CA)CB, HN(COCA)CB, HNCO, and HN(CA)CO) using the uniformly ^{15}N,^{13}C-labeled and 65% fractionally deuterated sample (Yamazaki et al., 1994). The side chain signals were assigned from 3D H(CCO)NH-TOCSY, C(CO)NH-TOCSY, HCCH-TOCSY, and ^{15}N-edited TOCSY experiments (Clore and Gronenborn, 1994).

Structures of HI0017 were generated using a distance geometry/simulated annealing protocol (Nilges et al., 1988; Kuszewski et al., 1992) with the X-PLOR program (Brünger, 1992). Structure calculations for the N-terminal domain employed a total of 1027 proton–proton distances derived from 3D ^{15}N- and ^{13}C-resolved NOESY spectra (Fesik and Zuiderweg, 1988) collected with a mixing time of 80 ms. The NOE-derived distance restraints were given upper bounds of 2.7, 3.3, 4.0, 5.0, and 6.0 Å based upon the measured NOE intensities. From an analysis of the amide exchange rates measured from a series of ^{1}H/^{15}N HSQC spectra recorded after the addition of $^{2}H_2O$, 46 hydrogen bond restraints were added. In addition, 37 ϕ angular restraints derived from an analysis of the C′, N, C^{α}, H^{α}, and C^{β} chemical shifts using the TALOS program (Cornilescu et al., 1999) were included in the structural calculations.

Results and discussion

Assignments

The ^{1}H, ^{13}C, and ^{15}N resonances of HI0017 were assigned from an analysis of several heteronuclear multidimensional NMR spectra. Although the signals from the N-terminal region of HI0017 were nearly completely assigned (98%), the peaks from a portion of the C-terminus of the protein (in particular, residues 95–122) could not be assigned due to line broadening caused by chemical exchange. This exchange broadening may be due to a monomer–dimer equilibrium at the concentration (∼1 mM) used in the NMR experiments. This hypothesis is consistent with sedimentation equilibrium studies, which indicate that HI0017 exists as a monomer at lower protein concentrations (20 μM) and as a dimer at higher protein concentrations with a monomer–dimer equilibrium dissociation constant (K_D) of ∼0.5 mM. It is also consistent with

Table 1. Structure statistics and root-mean-square deviation for 21 structures of the N-terminal portion of *H. influenzae* HI0017

Structural statistics[a]	<SA>	$<\overline{SA}>_r$
Rmsd from experimental distance restraints (Å)[b]		
All (1073)	0.014±0.002	0.010
Intra-residue (247)	0.011±0.004	0.007
Sequential (274)	0.019±0.002	0.015
Medium range (136)	0.011±0.007	0.004
Long range (370)	0.008±0.004	0.006
Hydrogen bond (46)	0.023±0.006	0.019
Rmsd from experimental torsional angle restraints (deg)[c]		
ϕ angles (37)	0.42±0.18	0.41
X-PLOR potential energies (kcal mol^{-1})		
E_{tot}	126±12	122
E_{bond}	11±1	11
E_{ang}	83±7	87
E_{imp}	11±1	11
E_{repel}	9±3	8
E_{noe}	11±3	6
E_{cdih}	0.5±0.4	0.4
Rmsd from idealized geometry		
Bonds (Å)	0.00±0.00	0.00
Angles (deg)	0.37±0.01	0.38
Impropers (deg)	0.38±0.01	0.33
Cartesian coordinate rmsd (Å)	N, C^{α}, and C′	all heavy
<SA> vs. $<\overline{SA}>$[d]	0.74±0.12	1.42±0.18

[a] <SA> is the ensemble of 21 NMR-derived solution structures of the N-terminal domain of *H. influenzae* HI0017; $<\overline{SA}>$ is the mean atomic structure; $<\overline{SA}>_r$ is the energy-minimized average structure. No electrostatic potential was used in the structural calculations, and the X-PLOR F_{repel} function was only used for repulsive van der Waals interactions with a force constant of 4.0 kcal mol^{-1} $Å^{-4}$ with atomic radii set to 0.75 times their CHARMM values (Brooks et al., 1983).

[b] Distance restraints were employed with a square-well potential (F_{noe} = 50 kcal mol^{-1} $Å^{-2}$). Hydrogen bonds were given bounds of 1.8–2.4 Å (H-O) and 2.7–3.3 Å (N-O). No distance restraint was violated by more than 0.3 Å in any of the final structures.

[c] Torsional restraints were applied with values of −120±30° for those angles in β-strands and −60±30° for those angles in α-helices. A force constant of 200 kcal mol^{-1} rad^{-2} was applied for all torsional restraints.

[d] Rmsd for residues 2–57.

the relatively short ^{15}N T_2 relaxation times (∼50 ms) observed for the backbone amides.

Structure determination

The three-dimensional structure of the N-terminal region of HI0017 (residues 1–58) was determined using a distance geometry/simulated annealing protocol (Nilges et al., 1988; Kuszewski et al., 1992) from a total of 1073 NMR-derived restraints. A superposition of 21 low-energy NMR structures is shown in Figure 2, and the structural statistics are given in Table 1. As shown in Figure 2, the structure of the backbone is well defined by the NMR data. The atomic root-mean-square deviation (rmsd) about the mean coordinate positions for residues 2–57 is 0.74±0.12 Å for the backbone atoms and 1.42±0.18 Å for all heavy atoms.

Description of structure

The structure of the N-terminal region of HI0017 consists of a five-stranded antiparallel β-sheet and two short α-helices (Figure 3A). The short α1 helix has only one turn and is located between the β1 and β2 strands, while the slightly longer α2 helix (one-and-a-half turn) connects β4 and β5. The β-sheet is twisted to form a barrel-like shape which is closed at one end by the second α-helix (α2). A β-bulge is present at Thr 8 in β1 which is situated adjacent to α1. The structure of the protein is stabilized by a hydrophobic core consisting of residues I5, I7, L15, W20, L22, F38, V44, V46, L49, I52, and Y54.

Comparison to previously determined structures

Based on a 3D structure search using DALI (Holm and Sander, 1996), the N-terminal region of HI0017 was found to resemble the C-terminal domain of *Diphtheria* toxin repressor protein (DtxR) (Qiu et al., 1996; Pohl et al., 1998) (Figure 3B). The overall folds of both proteins are very similar with an rmsd of 2.3 Å for the superimposition of 45 pairs of backbone atoms (N, C^{α}, C′). Both proteins contain a twisted five-stranded antiparallel β-sheet. However, DtxR contains an additional two-turn α-helix between β1 and β2. Another slight difference is the longer loop in HI0017 that connects β3 and β4. Interestingly, this loop in HI0017 occupies the corresponding space of the additional α-helix in DtxR (Figure 3).

Possible functions of HI0017

The N-terminal protein sequence of HI0017 shares no homology with the sequence of the C-terminal domain of DtxR. Yet, their overall folds are highly similar (Figure 3). This suggests that the domains of these two proteins could have similar functions. DtxR regulates the expression of several iron-sensitive genes in *Corynebacterium diphtheriae* and is a member of a family of related repressor proteins in Gram-positive bacteria, including *Mycobacterium tuberculosis* and *Streptomyces lividans* (Pohl et al., 1998; Wang et al., 1999). The C-terminal domain (residues 148–226)

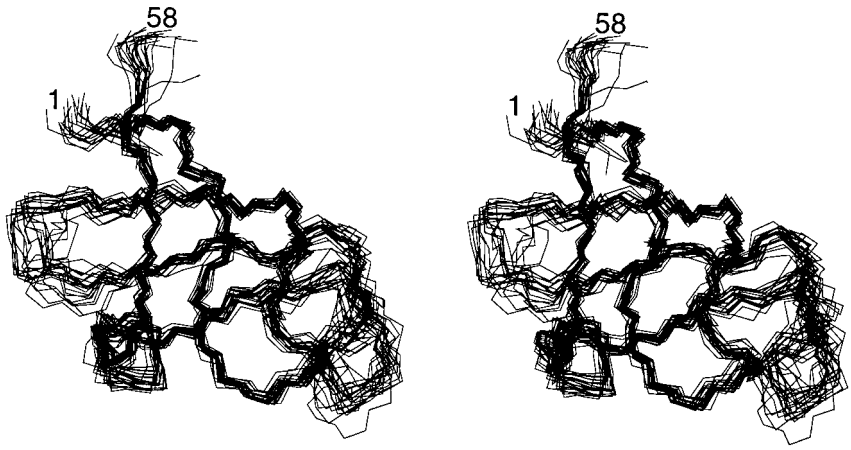

Figure 2. Stereoview of the backbone (N, C^{α}, C') of 21 superimposed NMR-derived structures of the N-terminal portion of *H. influenzae* HI0017 (residues 1–58).

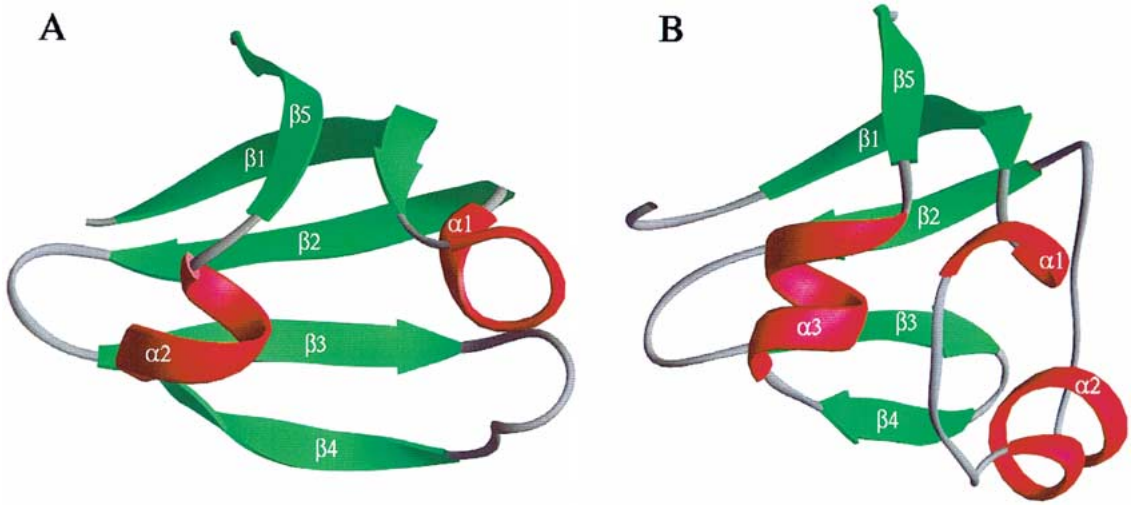

Figure 3. Ribbon plot (Carson, 1987) depicting (A) the averaged minimized NMR structure of the N-terminal portion of *H. influenzae* HI0017 (residues 1–58) and (B) the C-terminal domain of *Diphtheria* toxin repressor protein (DtxR). The structure of the N-terminal portion of HI0017 consists of five β-strands (β1, 2–10; β2, 18–23; β3, 28–32; β4, 41–45; β5, 52–57) and two short α-helices (α1, 13–17; α2, 46–51).

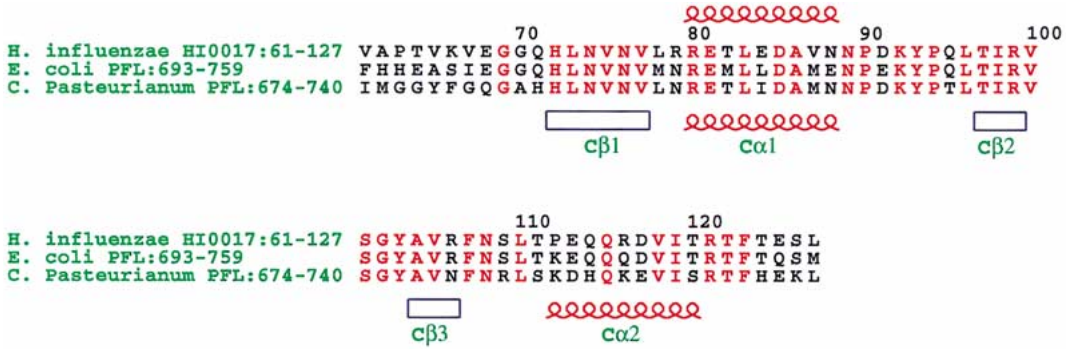

Figure 4. Sequence alignment of the C-terminal segment of *H. influenzae* HI0017 and the C-terminal portion of pyruvate formate-lyases (PFL) from *E. coli* (accession number P09373) and *Clostridium pasteurianum* (accession number Q46266). The secondary structure of PFL is indicated below the sequence. An α-helix observed for HI0017 as well as the HI0017 sequence numbering are indicated above the sequence.

of DtxR has been shown to bind to proline-rich sequences and was postulated to function as a molecular switch by altering the monomer–dimer equilibrium by binding to an internal proline-rich sequence (Wang et al., 1999).

The C-terminal 60 residues of HI0017 have been found to share high sequence homology with the C-terminal sequence of pyruvate formate-lyase (PFL) (Figure 4). PFL, a large protein with ∼760 amino acid residues (Roedel et al., 1988; Sawers and Watson, 1998; Eklund and Fontecave, 1999), catalyzes the reversible conversion of pyruvate and CoA into acetyl-CoA and formate. This enzyme plays a key role in anaerobic glucose fermentation in *E. coli* and other bacteria (Sawers and Watson, 1998). Upon activation by an activase (an iron-sulfur protein), PFL carries a free radical under anaerobic conditions, which is located on a glycyl residue (G734) near the C-terminus (Wagner et al., 1992). The active form of PFL is inactivated by molecular oxygen during aerobic growth to form the radical-free, oxygen-stable form of the protein.

The crystal structure of *E. coli* PFL has recently been determined and consists of a 10-stranded β/α barrel flanked on the outside by α-helices (Becker et al., 1999). The active site residues (C418, C419, G739) are situated on the tips of the two short loops near the center of the barrel. This 10-stranded β/α barrel motif is structurally similar to that previously determined in ribonucleotide reductase (RNR) class I (subunit R1) (Nordlund et al., 1990; Uhlin and Eklund, 1994) and class III (subunit NrdD) (Sun et al., 1995; Logan et al., 1999).

Since HI0017 and *E. coli* PFL share 80% identical residues for the C-terminal 60 amino acids of the protein (Figure 4), they are likely to have similar functions, i.e. to carry a free radical at the highly conserved glycyl residue in the SGYA sequence (Figure 4). Although the solution structure of the C-terminal portion of HI0017 could not be determined due to the broad NMR signals, it is interesting to note that an observed α-helix for residues 80–88 in HI0017 corresponds to an α-helix (residues 712–720) in *E. coli* PFL (Figure 4). This observation further supports the idea that these highly homologous C-terminal sequences perform a similar function as a radical carrier.

Conclusions

HI0017 represents one of the structurally and functionally unknown proteins in a clinically important organism, *H. influenzae*. Here we reported the solution structure of the N-terminal portion of *H. influenzae* HI0017. The structure is composed of a five-stranded β-sheet and two α-helices and resembles the C-terminal domain of *Diphtheria* toxin repressor (DtxR). The C-terminal sequence of HI0017 is found to share high homology with the C-terminal sequence of pyruvate formate-lyase. Therefore, we propose that an unidentified activase (probably an iorn-sulfur protein) binds to HI0017 to form a functional complex. Upon the binding of an activase, HI0017 is activated by the formation of a glycyl radical at the G102 position which corresponds to G734 in *E. coli* PFL. The activated form of HI0017 then participates in protein/substrate catalysis.

Acknowledgements

We thank Dr. John Harlan for performing sedimentation equilibrium studies.

References

Becker, A., Fritz-Wolf, K., Kabsch, W., Knappe, J., Schultz, S. and Wagner, A.F.V. (1999) *Nat. Struct. Biol.*, **6**, 969–975.

Blattner, F.R., Plunkett III, G., Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J., Mau, B. and Shao, Y. (1997) *Science*, **277**, 1453–1474.

Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S. and Karplus, M. (1983) *J. Comput. Chem.*, **4**, 187–217.

Brünger, A.T. (1992) *X-PLOR Manual*, Yale University Press, Cambridge, MA.

Carson, M. (1987) *J. Mol. Graphics*, **5**, 103–106.

Clore, G.M. and Gronenborn, A.M. (1994) *Methods Enzymol.*, **239**, 349–363.

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

Eklund, H. and Fontecave, M. (1999) *Structure*, **7**, R257–R262.

Fesik, S.W. and Zuiderweg, E.R.P. (1988) *J. Magn. Reson.*, **78**, 588–593.

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., Fitzhgh, W., Fields, C.A., Gocayne, J.D., Scott, J.D., Shirley, R., Liu, L.-I., Glodek, A., Kelly, 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.
Holm, L. and Sander, C. (1996) *Science*, **273**, 595–602.
Kuszewski, J., Nilges, M. and Brünger, A.T. (1992) *J. Biomol. NMR*, **2**, 33–56.
Logan, D.T., Andersson, J., Sjöberg, B.-M. and Nordlund, P. (1999) *Science*, **283**, 1499–1504.
Nilges, M., Clore, G.M. and Gronenborn, A.M. (1988) *FEBS Lett.*, **229**, 317–324.
Nordlund, P., Sjöberg, B.-M. and Eklund, E. (1990) *Nature*, **345**, 593–598.
Pohl, E., Holmes, R.K. and Hol, W.G.J. (1998) *J. Biol. Chem.*, **273**, 22420–22427.
Qiu, X., Pohl, E., Holmes, R.K. and Hol, W.G.J. (1996) *Biochemistry*, **35**, 12292–12303.
Roedel, W., Plaga, W., Frank, R. and Knappe, J. (1988) *Eur. J. Biochem.*, **177**, 153–158.
Sawers, G. and Watson, G. (1998) *Mol. Microbiol.*, **29**, 945–954.
Sun, X., Eliasson, R., Pontis, E., Andersson, J., Buist, G., Sjöberg, B.-M. and Reichard, P. (1995) *J. Biol. Chem.*, **270**, 2443–2446.
Turk, D.C. (1982) In *Haemophilus influenzae: Epidemiology, Immunology, and Prevention of Disease* (Eds., Sell, S.H. and Wright, P.F.), Elsevier Biomedical, New York, NY, pp. 3–9.
Turk, D.C. (1984) *J. Med. Microbiol.*, **18**, 1–16.
Uhlin, U. and Eklund, H. (1994) *Nature*, **370**, 533–539.
Valerie, K., Stevens, J., Lynch, M., Henderson, E.E. and de Riel, J.K. (1986) *Nucleic Acids Res.*, **14**, 8637–8654.
Wagner, A.F.V., Frey, M., Neugebauer, F.A., Schafer, W. and Knappe, J. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 996–1000.
Wang, G., Wylie, G.P., Twigg, P.D., Caspar, D.L.D., Murphy, J.R. and Logan, T.M. (1999) *Proc. Natl. Acad. Sci. USA*, **96**, 6119–6124.
Yamazaki, T., Lee, W., Arrowsmith, C.H., Muhandiram, D.R. and Kay, L.E. (1994) *J. Am. Chem. Soc.*, **116**, 11655–11666.