The crytal structure of NGOO47 from Neisseria gonorrhoeae reveals a novel protein fold incorporating a helix-turn-helix motif

ARTICLE *in* PROTEINS STRUCTURE FUNCTION AND BIOINFORMATICS · MAY 2010

Impact Factor: 2.63 · DOI: 10.1002/prot.22698 · Source: PubMed

PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

STRUCTURE NOTE

The crystal structure of NGO0477 from *Neisseria gonorrhoeae* reveals a novel protein fold incorporating a helix-turn-helix motif

Jingshan Ren,[1] Sarah Sainsbury,[1] Joanne E. Nettleship,[1] Nigel J. Saunders,[2] and Raymond J. Owens[1]*

[1] The Oxford Protein Production Facility and Division of Structural Biology, University of Oxford, Oxford OX3 7BN, United Kingdom

[2] The Bacterial Pathogenesis and Functional Genomics Group, The Sir William Dunn School of Pathology, University of Oxford, Oxford OX1 3RE, United Kingdom

Key words: Neisseria gonorrhoeae; XRE-family regulator; prophage; novel fold.

INTRODUCTION

Neisseria spp. are Gram-negative Beta-Protobacteria, which include many species found only in humans. Of these, two are important pathogens, *N. meningitidis,* responsible for bacterial meningitis and septicaemia, and *N. gonorrhoeae* the causative agent of gonorrhoea. In common with most bacteria, DNA sequencing has identified prophage sequences integrated into the genomes of both of these species (reviewed[1]). One of these bacteriophage-derived sequences found only in *N. gonorrhoeae* (strain FA1090) is the gene NGO0477 (also designated as XNG0459) a 369 bp opening reading frame on the complementary strand. The sequence is located in a 43 kb region that appears to represent an entire double stranded tailed bacteriophage genome and codes for a putative 16.59 kD transcriptional regulator containing a helix-turn-helix motif.[2] As part of a structural proteomics study of pathogenic *Neisseria*, we have solved the crystal structure of the NGO0477 protein, which reveals an unusual and novel protein fold featuring an intrachain disulphide bridge. Comparison with other proteins indicates that the C-terminus of the protein shares structural homology with DNA-binding proteins. However, examination of the structure reveals that the putative binding helices are orientated towards the inside of the protein and therefore are not available to interact with DNA.

METHODS

Cloning, expression, and purification

The NGO0477 expression construct, consisting of residues 21 to 154 fused to a N-terminal hexahistidine tag, was generated by means of ligation-independent cloning using Gateway technology™ (Invitrogen).[3] The gene was amplified from genomic DNA (*Neisseria gonorrhoeae* strain FA1090) using KOD HiFi™ polymerase (Novagen) using the forward primer 5′-GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGAAGGAGATAGAACCATGGCACATCACCACCACCATCACATCCCCGTTACCCACATAAAATGCC 3′ and the reverse primer 5′GGGGACCACTTTGTACAAGAAAGCTGGGTCTCATCAGCCTGCTCCATATATTCGTACCG 3′ and inserted into the expression vector, pDEST14 (Invitrogen). Selenomethionine-labeled protein was produced in the *E. coli* auxotroph B834(DE3) as previously described.[4] The molecular mass of the purified protein determined by gel filtration column chromatography on a HiLoad

Current address of Sarah Sainsbury is Gene Centre Munich, University of Munich – LMU, Feodor-Lynen-Strasse 25, 81377 Munich, Germany.

Grant sponsors: UK Medical Research Council and Biotechnology and Biological Sciences Research Council (OPPF).

*Correspondence to: Raymond J. Owens, The Oxford Protein Production Facility and Division of Structural Biology, Henry Wellcome Building for Genomic Medicine, University of Oxford, Oxford OX3 7BN, United Kingdom.
E-mail: ray@strubi.ox.ac.uk

Received 7 December 2009; Revised 8 January 2010; Accepted 9 January 2010
Published online 1 February 2010 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.22698

Table I
X-Ray Data Collection and Refinement Statistics

Data Collection Details:		
X-ray source	ESRF BM14	
Data set	Peak	remote
Wavelength (Å)	0.97800	0.90499
Space group	*H3*	
Unit cell (Å)	$a = b = 124.82$, $c = 137.58$	
Resolution range (Å)	30.0 − 2.80 (2.90–2.80)	30.0 − 3.20 (3.31–2.20)
Unique reflections	19515 (1929)	12891 (1287)
Completeness (%)	99.7 (97.3)	98.9 (92.4)
Redundancy	10.7 (8.5)	5.2 (3.2)
Average $I/\sigma I$	10.3 (2.2)	5.8 (1.3)
Rmerge	0.238 (0.923)	0.234 (0.571)
Refinement statistics:		
Resolution range (Å)	30.0 − 2.80 (2.90–2.80)	
No. of reflections (working/test)	18502/1004	
R-factor[a] (R_{work}/R_{free})	0.230/0.279	
No. of atoms (protein/other atoms)	4098/62	
Rms bond length deviation (Å)	0.009	
Rms bond angle deviation (°)	1.2	
Mean B-factor (protein/other atoms [Å^2])	45/39	
Ramachandran plot		
Residues in most favored regions (%)	93.9	
Residues in allowed regions (%)	6.1	
Residues in disallowed regions (%)	0	

[a] R_{work} and R_{free} are defined by $R = \Sigma_{hkl}||F_{obs}| - |F_{calc}||/\Sigma_{hkl}|F_{obs}|$, where *h,k,l* are the indices of the reflections (used in refinement for R_{work}; 5%, not used in refinement, for R_{free}), F_{obs} and F_{calc} are the structure factors, deduced from measured intensities and calculated from the model, respectively.

16/60 Superdex G75 column (GE Healthcare Life Sciences) was 34.6 kD, suggesting that it exists as a dimer in solution. The protein was concentrated to 16.6 mg/mL in 20 m*M* Tris pH 7.5, 200 m*M* NaCl, 1 m*M* Tris (2-carboxyethyl) phosphine (TCEP) and stored flash frozen at −80°C.

Crystallization, data collection, and determination of the structure

Crystals of the purified protein were obtained in a 200 nL sitting drop vapor diffusion crystallization experiment in 0.1 *M* Tris HCl, pH 7.5 containing 22% PEG 400 and 0.3 *M* tri-sodium citrate.[5] Multiple wavelength anomalous dispersion data for two crystals of selenomethionine-labeled NGO0477 protein were collected to 2.8 Å resolution at beamline BM14, the ESRF (Grenoble, France). Following a fluorescence scan, X-ray data of 480° at the peak wavelength (0.97800Å) from one crystal and 180° at a remote wavelength (0.907Å) from a second crystal were collected with oscillation of 1.0° per frame. Crystals, taken directly from the crystallization drop without the addition of a cryoprotectant, were flash frozen and maintained at 100K under a stream of nitrogen gas during data collections. Indexing and integration of data images were carried out using HKL2000. The NGO0477 crystals have a space group of H3 with unit cell dimensions of $a = b = 124.82$ Å and $c = 137.58$ Å. There are four monomers in one crystal asymmetric unit with a solvent content of 62%. The SHELX program suite was used to evaluate the anomalous signal during the course of the MAD data collection and to solve the positions of selenium sites. SOLVE/RESOLVE[6,7] were then used for refinement of selenium positions and phase improvement combined with automated model building. The partially built model by RESOLVE was rebuilt manually using the program COOT[8] and then refined with REFMAC.[9] Tight main z chain NCS restraint for residues 5–19 and 26–130 was applied throughout the refinement. The statistics for the X-ray data and the refined structure are given in Table I. The atomic coordinates of NGO0477 and structure factors have been deposited in the Protein Data Bank under the accession code 3KXA.

Sequence and structure analyzes

The following servers/meta-servers were used to analyze the sequence and structure of NGO0477: BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi), ProFunc (www.ebi.ac.uk/thornton-srv/databases/profunc), Dali (http://ekhidna.biocenter.helsinki.fi/dali_server), OPAL (www.oppf.ox.ac.uk), PISA (http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html) and TOPS++FATCAT (http://fatcat.burnham.org/TOPS/).

RESULTS AND DISCUSSION

Sequence analysis

Blast searches of the Genbank database with the NGO0477 amino acid sequence from *N. gonorrhoeae*

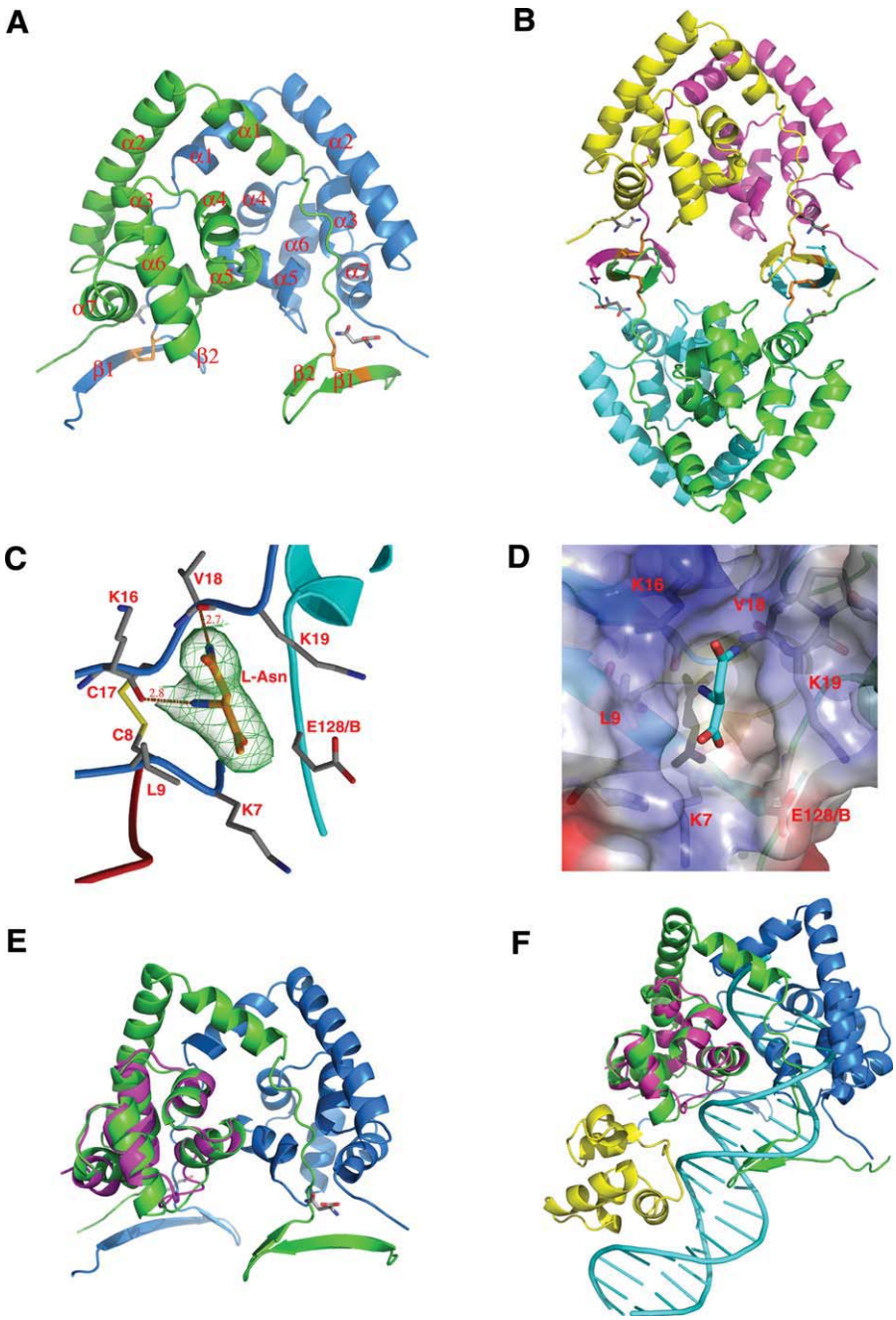

Figure 1

Structure of the NGO0477 protein: (A) Ribbon diagram shows the NGO0477 dimer and overall fold with secondary structure labeled. The disulphide bonds linking the β1 and β2 strands in each monomer are shown as orange sticks. The bound putative L-asparagine residues are shown as sticks and colored by atoms. (B) The NGO0477 tetramer in the crystallographic asymmetric unit. (C) $|2F_o\text{-}F_c|$ map contoured at 1σ showing the well defined electron density for the bound L-asparagine. Residues lining the binding site are shown as sticks and colored by atoms. Hydrogen bonds are draw as broken lines with distances indicated. (D) Semi-transparent protein surface showing the L-asparagine binding pocket (E) Superimposition of NGO0477 (green and blue) with the unliganded phage 434 cro protein (magenta, pdb id: 1r69). (F) Superimposition of NGO0477 with the phage 434 cro protein and DNA complex (magenta, yellow, and cyan; pdb id: 2or1). Both (E) and (F) show that the α5 helix that is buried between the dimer interface in NGO0477 overlaps well with the DNA-binding helices in the phage 434 cro protein.

strain FA1090, returned identical sequences from the genomes of four other strains of *N. gonorrhoeae* (MS11, PID24-1, PID332, and 1291), which have been sequenced. In addition a gene (NEISICOT_01377) encoding a protein with 92% amino acid sequence identity was identified in the recently sequenced genome of the com-

mensal neisseria, *N. sicca* (ATCC 29,256). Bioinformatic analyzes showed that all the proteins had an N-terminal hydrophobic 20 amino acid sequence predicted as a secretion signal by the SignalP algorithm ($p = 0.997$),[10] suggesting that the proteins may be localized to the periplasm. However, they also have a C-terminal domain containing a helix-turn-helix (HtH) motif belonging to the Xenobiotic Response Element (XRE)[11] family of DNA-binding proteins. This family includes bacteriophage transcriptional repressors[12] and the presence of such an HtH motif implies that NGO0477 and related proteins are involved in binding to DNA. The relationship between the sequence and structure of the protein was investigated by X-ray crystallography. For structural studies, the NGO0477 protein was produced in *E. coli* without the putative N-terminal leader sequence.

Overall structure

The X-ray crystal structure of the NGO0477 from *N. gonorrhoeae* (FA1090) was determined using selenomethionine-substituted protein by multiple wavelength anomalous dispersion. The asymmetric unit of the crystal comprises four protein molecules arranged as two back-to-back dimers [Fig. 1(B)]. The dimeric nature of the protein observed in the crystal is consistent with the results of size exclusion chromatography of the purified protein which showed that it behaved as a dimer in solution. The two dimers are related by a 2-fold NCS perpendicular to the 3-fold crystallographic axis, resulting in a pseudo space group of H32 for the crystal. The dimer: dimer interface is formed by the beta hairpins at the N-terminus of each monomer (β1 residues 4–11: β2 14–16) arranged in anti-parallel pairs. The β stranded hairpin is stabilized by an intra-molecular disulphide bridge linking C6 in the β1 strand to C17 at the start of an extended linker that joins the N-terminal region to the α-helical C-terminal domain of the protein. The connecting strand and following two helices run anti-parallel to the equivalent polypeptide of the adjacent monomer [Fig. 1(A)]. The C-terminal domain consists of a five helix bundle (α3–7) and is involved in dimerisation of the protein. Approximately 20% of the total solvent accessible surface of each monomer is buried at the dimer interface (3880 Å^2) with the main point of contact between the two α5 helices. These are arranged in an anti-parallel fashion and interact directly through hydrogen bonds between P88 and N92 on each chain and indirectly through water-mediated hydrogen bonds involving S96 and S91 on each chain [Fig. 1(A)]. The C-terminus of the protein lies close to the N-terminus with a hydrogen bond between E180 and K7. Intriguingly, there is a surface pocket formed by residues K7, L9, K16, K19 from the β hairpin of one chain and E128 from the adjacent chain that contains clear electron density. The best fit for this is an asparagine molecule which makes H-bond interactions with the carbonyl groups of K16 and V18 [Fig. 1(C,D)]. The functional significance of the binding pocket and whether asparagine is a physiological ligand is unclear.

Comparison with other structures

No structural matches to the entire NGO0477 protein were found in SSM,[13] Dali,[14] and TOPS++FATCAT[15] searches of the Protein Data Bank strongly indicating that the protein represents a new fold. The only hits that were obtained were to the C-terminal domain of the protein (residues 63- 125) and were all HtH containing DNA-binding proteins. The closest structural match from the Protein Databank identified by all three searches was the phage 434 cro protein (PDB id 1r69). Superimpositions of the NGO0477 structure with that of the cro repressor (both apo and in complex with DNA, PDB id 1r69 and 2or1) showed root mean square deviations (rmsd) of 1.4 Å and 1.6 Å for 55 and 57 equivalent Cαs, respectively [Fig. 1(E,F)]. Although there is high structural similarity between the C-terminal domain of NGO0477 and the DNA-binding domain of the cro repressor, the overlay shows that the orientation of the domain in NGO0477 towards the interior of the dimer means that the protein would not bind to DNA [Fig. 1(F)]. Reorientation of the domain to enable interaction with DNA would require a major conformational change involving disruption of the extensive dimer interface, which would be energetically unfavorable. Thus the annotation of NGO0477 as a DNA-binding regulator, based on primary sequence analysis is not supported by the tertiary structure of the protein. However, given the structural similarity, it seems likely that the protein is derived in part from a phage encoded repressor though this feature has been elaborated into a new fold of as yet unknown function.

ACKNOWLEDGMENTS

We thank the staff at BM14 for help with data collection.

REFERENCES

1. Snyder LA, Davies JK, Ryan CS, Saunders NJ. Comparative overview of the genomic and genetic differences between the pathogenic *Neisseria* strains and species. Plasmid 2005;54:191–218.
2. Piekarowicz A, Klyz A, Majchrzak M, Adamczyk-Poplawska M, Maugel TK, Stein DC. Characterization of the dsDNA prophage sequences in the genome of *Neisseria gonorrhoeae* and visualization of productive bacteriophage. BMC Microbiol 2007;7:66.
3. Walhout AJ, Temple GF, Brasch MA, Hartley JL, Lorson MA, van den Heuvel S, Vidal M. GATEWAY recombinational cloning: application to the cloning of large numbers of open reading frames or ORFeomes. Methods Enzymol 2000;328:575–592.
4. Ren J, Nettleship JE, Sainsbury S, Saunders NJ, Owens RJ. Structure of the cold-shock domain protein from Neisseria meningitidis

reveals a strand-exchanged dimer. Acta Crystallogr Sect F Struct Biol Cryst Commun 2008;64:247–251.
5. Walter TS, Diprose JM, Mayo CJ, Siebold C, Pickford MG, Carter L, Sutton GC, Berrow NS, Brown J, Berry IM, Stewart-Jones GB, Grimes JM, Stammers DK, Esnouf RM, Jones EY, Owens RJ, Stuart DI, Harlos K. A procedure for setting up high-throughput nanolitre crystallization experiments. Crystallization workflow for initial screening, automated storage, imaging and optimization. Acta Crystallogr D-Biol Crystallogr 2005;61:651–657.
6. Terwilliger TC. Maximum-likelihood density modification. Acta Crystallogr D-Biol Crystallogr 2000;56:965–972.
7. Terwilliger TC, Berendzen J. Automated MAD and MIR structure solution. Acta Crystallogr D-Biol Crystallogr 1999;55:849–861.
8. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallogr D-Biol Crystallogr 2004;60:2126–2132.
9. Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr D-Biol Crystallogr 1997;53:240–255.
10. Bendtsen JD, Nielsen H, von Heijne G, Brunak S. Improved prediction of signal peptides: SignalP 3.0. J Mol Biol 2004;340:783–795.
11. Wood HE, Devine KM, McConnell DJ. Characterisation of a repressor gene (xre) and a temperature-sensitive allele from the Bacillus subtilis prophage, PBSX. Gene 1990;96:83–88.
12. Wintjens R, Rooman M. Structural classification of HTH DNA-binding domains and protein-DNA interaction modes. J Mol Biol 1996;262:294–313.
13. Krissinel E, Henrick K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta Crystallogr 2004;60:2256–2268.
14. Holm L, Kaariainen S, Rosenstrom P, Schenkel A. Searching protein structure databases with DaliLite v.3. Bioinformatics 2008;24:2780–2781.
15. Veeramalai M, Ye Y, Godzik A. TOPS++FATCAT: fast flexible structural alignment using constraints derived from TOPS+ Strings Model. BMC bioinformatics 2008;9:358.