PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

STRUCTURE NOTE

Crystal structure of VC1805, a conserved hypothetical protein from a *Vibrio cholerae* pathogenicity island, reveals homology to human p32

Md. Arif Sheikh,[1] Jane A. Potter,[1] Kenneth A. Johnson,[1] Robert B. Sim,[2] E. Fidelma Boyd,[3] and Garry L. Taylor[1]*

[1] Centre for Biomolecular Sciences, University of St Andrews, St Andrews, Fife, KY16 9ST, United Kingdom

[2] Department of Biochemistry, MRC Immunochemistry Unit, University of Oxford, Oxford OX1 3QU, United Kingdom

[3] Department of Biological Sciences, Wolf Hall, University of Delaware, Newark, DE 19711

Key words:* *Vibrio cholerae; VC1805; human p32; C1q.

INTRODUCTION

The severe diarrhoeal disease cholera is caused by the gram-negative bacterium *Vibrio cholerae* and continues to be a major cause of morbidity and mortality. Like many *Vibrio* species *V. cholerae* inhabits an aquatic ecosystem, and most *V. cholerae* isolates do not possess the ability to cause cholera. Of more than 200 known serotypes, only O1 and O139 serogroups are highly pathogenic and acknowledged to cause epidemic disease.[1] The O1 serogroup can be divided into two groups: classical and El Tor, with the first cholera pandemic, beginning in Asia in 1817, and the subsequent five pandemics probably caused by the classical biotype. The seventh and present pandemic began in 1961 caused by the El Tor biotype.[2] In 1992 a novel O-serogroup, O-139, emerged to cause epidemic cholera.[3] All *V. cholerae* O1 and O139 serogroups encode the major virulence factors cholera toxin and toxin coregulated pilus, the latter being encoded on a pathogenicity island, named Vibrio Pathogenicity Island-I (VPI-1).[4] Several other genomic regions have been identified that occur mainly among epidemic O1 and O139 serogroup isolates, including the so-called *Vibrio* seventh pandemic island-I (VSP-I) encoding ORFs VC0175 to VC0185, VSP-II encoding ORFs VC0490 to VC0516 and VPI-2 encoding ORFs VC1758 to VC1809.[5–7] A microarray analysis of *V. cholerae* El Tor isolates was used to identify VSP-I and VSP-II encompassing genes that are possibly responsible for the unique characteristics of the seventh pandemic (El Tor) strains.[5] An evolutionary genetic analysis of clinical and environmental isolates suggested that pandemic strains arose from a common O1 serogroup progenitor through the successive acquisition of new virulence regions.[8]

VPI-2 is a 57.3 kb region which consists of 52 ORFs present in all toxigenic O1 and O139 serogroup isolates, but lacking in non-O1 and non-O139 nontoxigenic isolates.[6] VPI-2 encodes a type-I restriction modification system, a *nan-nag* region of genes involved in sialic acid metabolism that may play a nutritional role,[9] a sialidase/neuraminidase known to convert intestinal higher-order gangliosides to GM1,[10] and a region with homology to Mu phage. In addition, VPI-2 contains several genes that code for hypothetical proteins. One of these proteins, VC1805, is a 148 amino acid protein with no function revealed so far through sequence analysis. VC1805 is also encoded in the reduced 20 kb VPI-2 region present

Grant sponsors: BBSRC; Scottish Funding Council; University of St Andrews; European Synchrotron Radiation Facility; European Union.
*Correspondence to: Garry Taylor, Centre for Biomolecular Sciences, University of St Andrews, St Andrews, Fife, KY16 9ST, UK.
E-mail: glt2@st-andrews.ac.uk
Received 1 December 2007; Revised 30 December 2007; Accepted 31 December 2007
Published online 25 February 2008 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.21993

among most O139 serogroup isolates that are missing ORFs VC1761 to VC1788.[6] A paralogue of VC1805 exists within VSP-II, VC0508 a 147 residue protein that shares 59% amino acid sequence identity with VC1805. A search of the sequence database using PSI-BLAST reveals homologues of VC1805 in several *Vibrio* species: *V. vulnificus, V. splendidus, V. alginolyticus* and *V. fischeri*, and orthologues in *Altermonas macleodii, Aeromonas hydrophilia*, and some *Shewanella* species. In addition, PSI-BLAST reveals that the adjacent hypothetical protein VC1804, with 104 residues, is a homologue of VC1805 sharing 26% sequence identity. Similarly the hypothetical protein VC0509 is a homologue of the adjacent protein VC0508. The four proteins VC0508, VC0509, VC1804, and VC1805 are therefore likely to share the same protein fold and be functionally related.

As part of a structural genomics approach to understanding the function of hypothetical proteins within *V. cholerae* genomic islands we have determined the crystal structure of VC1805 to a resolution of 2.1 Å using heavy atom isomorphous replacement. The structure reveals a similarity to the human mitochondrial protein p32 that is known to have several binding partners, including the human complement system protein C1q. This study shows that VC1805 does bind C1q, suggesting potential biological roles for the protein and its homologues.

MATERIALS AND METHODS

Cloning

The VC1805 gene was amplified by the polymerase chain reaction (PCR) from genomic DNA of *Vibrio cholerae* O1 El Tor strain N16961 using standard protocols. Forward and reverse primers were 5′-GGAGGCATAAC CATGGCTAACATTTTG and 5′-CTTGATGGAATTCAC TAAGGAGCCGAG with specific recognition sites *Nco*I and *EcoR*I. After amplification, the PCR product was digested with *Nco*I and *EcoR*I enzymes and cloned into the similarly digested pEHISTEV vector (an engineered variant of pET30 with an N-terminal 6× His tag that is cleavable using Tobacco Etch Virus (TEV) protease, Dr H. Liu, University of St Andrews, personal communication).

Protein expression and purification

The plasmid DNA containing the VC1805 gene was transformed into *Escherichia coli* strain BL21 (DE3). A transformed expression host colony was used to inoculate 100 mL LB containing antibiotic kanamycin (50 μg/mL) followed by shaking incubation at 37°C overnight. After incubation, a 50-mL culture was used to inoculate 5 L of LB containing kanamycin followed by shaking incubation at 37°C for 2–3 h until an OD_{600} of 0.5–0.6 was reached. At this point, expression of the cloned gene was induced by the addition of 0.5 m*M* IPTG (isopropyl-β-D-thiogalactopyranoside) followed by further incubation overnight at 25°C.

The cell pellet was resuspended in PBS buffer containing DNase and EDTA-free protease inhibitor cocktail (Roche Applied Science) and sonicated. Cell lysate was syringe filtered through a 0.22-μm filter unit (Millipore) and VC1805 was purified using a HisTrap™ HP column (GE Healthcare), according to the manufacturer's instruction manual. Protein samples were dialyzed against 3 L of PBS buffer containing 50 m*M* Tris-HCl, pH 7.5, 0.5 m*M* EDTA, 1 m*M* DTT (dithiothreitol), 20 m*M* imidazole, and 0.3*M* NaCl for 4 h at 4°C, with two buffer changes. Next, TEV protease was added into the dialysis tube containing His-tagged protein (1 μg protease for 1 mg protein) at 25°C overnight. The digested sample was dialyzed against the same dialysis buffer without EDTA and DTT for 2 h at 4°C and the dialysis buffer was changed twice. Protein samples were syringe filtered through a 0.22-μm filter unit (Millipore) and loaded onto the Nickel column as before. The flow-through containing VC1805 was dialyzed against 3 L of PBS buffer containing 50 m*M* Tris-HCl pH 7.5, 20 m*M* imidazole, and 20 m*M* NaCl overnight at 4°C. After dialysis, the samples were concentrated to 5 mL using a 3000-Da MWCO centrifugal concentrator (Sigma-Aldrich) for further purification by size exclusion on a Hiload™ 26/60 Superdex™ 200 gel filtration column. Protein purity was analyzed by SDS PAGE (sodium dodecyl sulfate polyacrylamide gel electrophoresis) and the correct identity confirmed by mass spectrometry.

Crystallization

A precrystallization test (Hampton Research) indicated the appropriate protein concentration (8 mg/mL) for crystallization screening. Three screens from Hampton research (Index, Crystal Screen 1 and Crystal Screen 2) were used in initial crystallization trials of native VC1805. This experiment was carried out using the sitting drop diffusion method following standard procedure supplied by the manufacturers. Initial hits were optimized through variation of the concentration of salt, precipitant and protein. The best quality crystals were obtained in an optimization of condition 41 of Crystal Screen 1 (0.1*M* HEPES-Na pH 7.5, 10% v/v 2-propanol, 24% w/v polyethylene glycol 4000).

Structure solution and refinement

Crystals were cryoprotected in crystallization buffer with 20% glycerol added before being flash-frozen in a nitrogen gas stream at 100 K. Data, which extended to a maximum resolution of 2.13 Å, were collected on station ID14-1 of the European Synchrotron Radiation Facility (ESRF), Grenoble, France. The structure was solved by

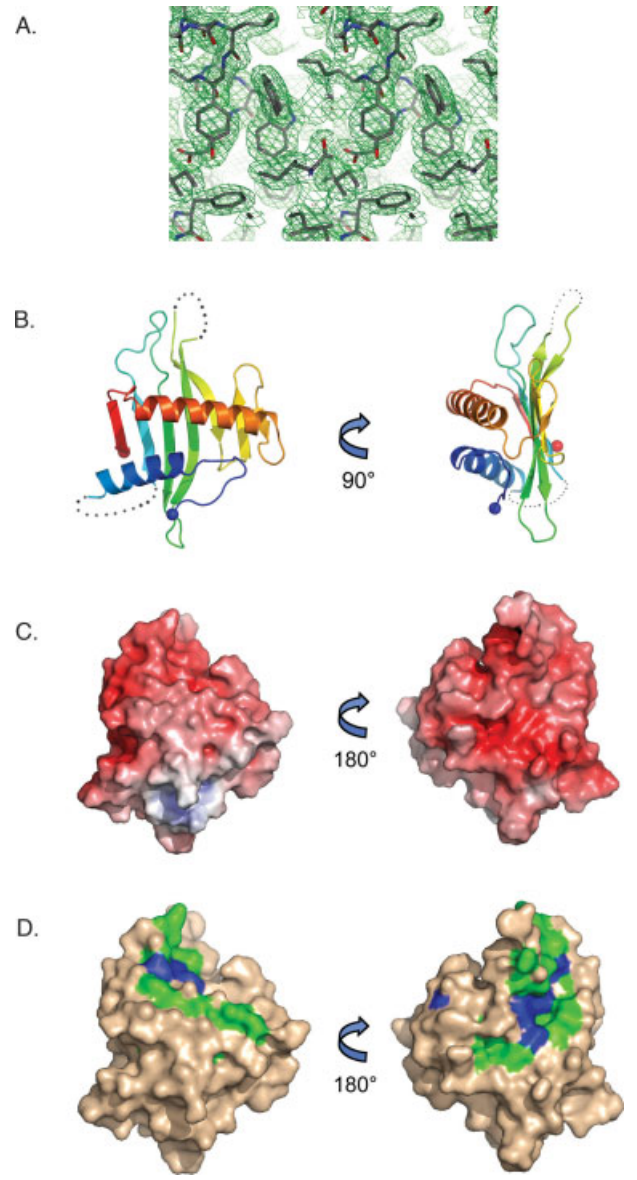

Figure 1

*The structure of VC1805. (**A**) Stereo image of part of the core hydrophobic region showing the final $2F_o$–F_c electron density map contoured at 1.0 σ. (**B**) Orthogonal views of a schematic drawing of the VC1805 fold in rainbow colours from blue at the N-terminus to red at the C-terminus. The disordered regions are shown as black dotted lines. (**C**) Two views of the electrostatic potential mapped onto the surface from −10KT (deepest red) to +10KT (deepest blue). (**D**) Conserved residues mapped onto the surface. Blue residues are totally conserved across all family members shown in Figure 2, and green are conserved across all members except VC1804 and VC0509. The left hand images are in the same orientation.*

multiple heavy atom isomorphous replacement using data from a crystal soaked for 30 min in 5 m*M* mercury nitrate or 5 m*M* PCMBS (*p*-chloromercuri benzene sulphonate). Data were indexed and integrated using the program MOSFLM.[11] The integrated data were merged and scaled using the programs of the CCP4[12] suite. The initial phases were obtained using the program SOLVE[13] with density modification using the program RESOLVE,[14–16] confirming the space group as $P3_121$. Manual model rebuilding and refinements were conducted using the programs O,[17] COOT,[18] and REFMAC5.[19] Part of the final $2F_o$–F_c electron density map is shown in Figure 1(A). The data collection and refinement statistics for native and heavy atom derivatives are summarized in Table I.

C1q binding assay

Using a 96-well microtiter plate (Maxisorb, NUNC, Denmark), wells were coated with 50 µL of a twofold serial dilution of VC1805 in PBS (from 4 to 0.125 µg/mL). For negative controls, wells were also coated with 50 µL

Table I
Data Collection and Refinement Statistics

	Crystal		
	Native	Mercury nitrate	PCMBS
Unit cell parameters (Å)	$a = b = 78.3$, $c = 42.3$	$a = b = 78.4$, $c = 41.3$	$a = b = 78.7$, $c = 42.3$
Space group	$P3_121$		
Maximum resolution (Å)	2.13	2.9	2.3
R_{merge}	0.057 (0.375)	0.079 (0.359)	0.038 (0.278)
Observed reflections	75,489	17,207	35,617
Unique reflections	8549	3223	6496
Completeness (%)	99.4 (99.4)	94.8 (94.8)	94.6 (96.7)
$\langle I/\sigma I \rangle$	23.2 (6.4)	21.8 (5.3)	29.2 (4.6)
Multiplicity	8.8 (9.0)	5.3 (5.3)	5.5 (4.7)
No. of heavy atom sites		2	1
⟨figure of merit⟩ after SOLVE	0.43		
Refinement			
Resolution limits	39.2–2.13		
Number of reflections	8131		
Overall *R* factor	0.235		
Free *R* factor	0.316		
RMSD ideal bond length (Å)	0.013		
RMSD ideal bond angles (°)	1.39		
No. of protein atoms	1054		
No. of water atoms	35		
Average B-factor for protein (Å^2)	39.2		
Average B-factor for water (Å^2)	40.4		

$R_{merge} = \sum_h \sum_j |I_{hj} - \langle I_h \rangle| / \sum_h \sum_j$, where I_{hj} is the intensity of the *j*th observation of unique reflection *h*. Overall *R* factor $\sum h||F_{oh}| - |F_{ch}|| / \sum h|F_{oh}|$, where F_{oh} and F_{ch} are the observed and calculated structure factor amplitudes for reflection *h*. Free *R* factor is equivalent to overall *R* factor, but is calculated using 5% of reflections excluded from the maximum-likelihood refinement stages.

of a twofold and fivefold dilution of milk in PBS buffer and with 50 μL of the archaeal DNA-binding protein hssb1 at 4 and 2 μg/mL, diluted in PBS buffer. Coating was done overnight at 4°C. The wells were then washed four times with PBS +0.1% Tween 20 (Buffer A). The remaining protein binding sites on the wells were saturated with 360 μL of blocking buffer (10% milk in PBS) and incubated for 1 h at room temperature, then wells were washed four times with buffer A. About 100 μL biotinylated anti-human C1q IgG in blocking buffer which contain 5 μg of antibody were added to the wells and incubated at 37°C for 1 h. After incubation, washed the wells four times with Buffer A. Streptavidin HRP (horseradish peroxidase, KPL Inc.) was added to the wells and incubated 37°C for 1 h, followed by four washes times with buffer A. About 100 μL of TMB peroxide substrate (prepared as per manufacturer's instructions, KPL) was added to the wells and kept at room temperature. As a positive control for C1q binding, wells were coated with hen ovalbumin (as below), then reacted with rabbit-anti-hen ovalbumin antiserum, to form immune complexes in the wells. C1q was purified from human plasma.[20] Rabbit anti-C1q IgG was biotinylated with biotin *N*-hydroxysuccinimide.

Complement consumption assay

NUNC Maxisorp 96-well microtitre plate wells were coated with 100 μL (5 μg) of VC1805 in PBS. As a positive control, wells on the same plate were coated with 100 μL per well of 100 μg/mL hen ovalbumin (Sigma) in 0.1*M* sodium carbonate buffer, pH 9.6. The plate was incubated overnight at 4°C, then washed four times with buffer A. The remaining sites for protein binding on the plate were saturated by blocking buffer and incubated for 1 h at room temperature before again washing four times with buffer A. Rabbit anti-ovalbumin antiserum (MRC Immunochemistry Unit) (heat-treated for 30 min at 56°C to inactivate rabbit C1q) was diluted 10-fold in PBS and 100 μL added to the ovalbumin-coated wells and incubated for 1 h at room temperature. Wells were washed four times with buffer A. The plates were now ready for the complement consumption assay. Reagent and buffer details were as previously described.[20] NHS (normal human serum) was diluted with 1 vol of DGBV++ buffer. The diluted serum (300 μL) was added in triplicate to wells coated with VC1805, ovalbumin anti-ovalbumin and noncoated wells followed by incubation at 37°C for 30 min. Serum was removed from wells and kept on ice until ready to assay for total complement haemolytic activity.[20] Briefly, serial twofold dilutions of serum were incubated with antibody-sensitised sheep erythrocytes and the degree of lysis determined after 1-h incubation at 37°C. The dilution of serum required to give 50% lysis was calculated.

RESULTS AND DISCUSSION

Overall structure

The crystal structure of VC1805 reveals a monomeric protein composed of a flat seven-stranded anti-parallel β-sheet with two helices sitting on the same face of the sheet [Fig. 1(B)]. The first helix (α1) is at the N-terminus, followed by six β-strands (β1 to β6) of varying length that form a simple consecutive up-down topology. The second helix (α2) then stretches back across these six β-strands, leading to a seventh strand (β7) at the C-terminus that completes the sheet. There are four disordered regions where no electron density was observed: the first four residues (1–4), the connection between α1 and β1 (31–39), the loop connecting β3 and β4 (84–86) and the final three residues (146–148). VC1805 has a pI of 4.7 and this is reflected in the negatively-charged surface potential exhibited by the protein [Fig. 1(C)].

Conserved amino acids

Figure 2 shows a CLUSTAL-W alignment of representative VC1805 homologues and paralogues. All of the proteins listed in Figure 2 carry a negative charge having isoelectric points between 4.3 and 5.4. *The V. vulnificus, V. spendidus, V. alginolyticus*, and *V. fischeri* homologues share sequence identities of 65, 60, 47, and 43%, respectively, with VC1805. The *Alteromonas macleodii, Aeromonas hydrophilia*, and *Shewanella* orthologues share identities of 38, 36, and 36%, respectively with VC1805. The *V. cholerae* paralogue VC0508 shares 59% identity with VC1805. Finally the *V. cholerae* homologue VC1804 and its paralogue VC0509 share 24 and 21% identity respectively with VC1805, while they share 52% identity with each other. There are 10 amino acids that are invariant among this family and these are mapped onto the surface of VC1805 and coloured blue in Figure 1(D). In addition, if VC1804 and VC0509 are excluded, then several other amino acids are conserved which are also mapped onto the surface in green. There appear to be two conserved patches, both carrying a negative potential, on opposite sides of the protein. One patch forms a groove formed between α2 and the loop connecting β1 and β2 with the invariant Tyr50 and His58 coming close in space at one end of the groove [Fig. 1(D), left hand image]. The other patch is centred around another groove on the opposite face that includes a deep hole lined with invariant residues Pro59, Glu61, Phe80 [Fig. 1(D), right hand image]. It is interesting to note that VC1804 would be predicted to be missing the first helix, a region that does not contain any conserved residues.

Structural homology to human p32

Sequence comparisons and threading had failed to suggest a fold or function for VC1805. The crystal structure

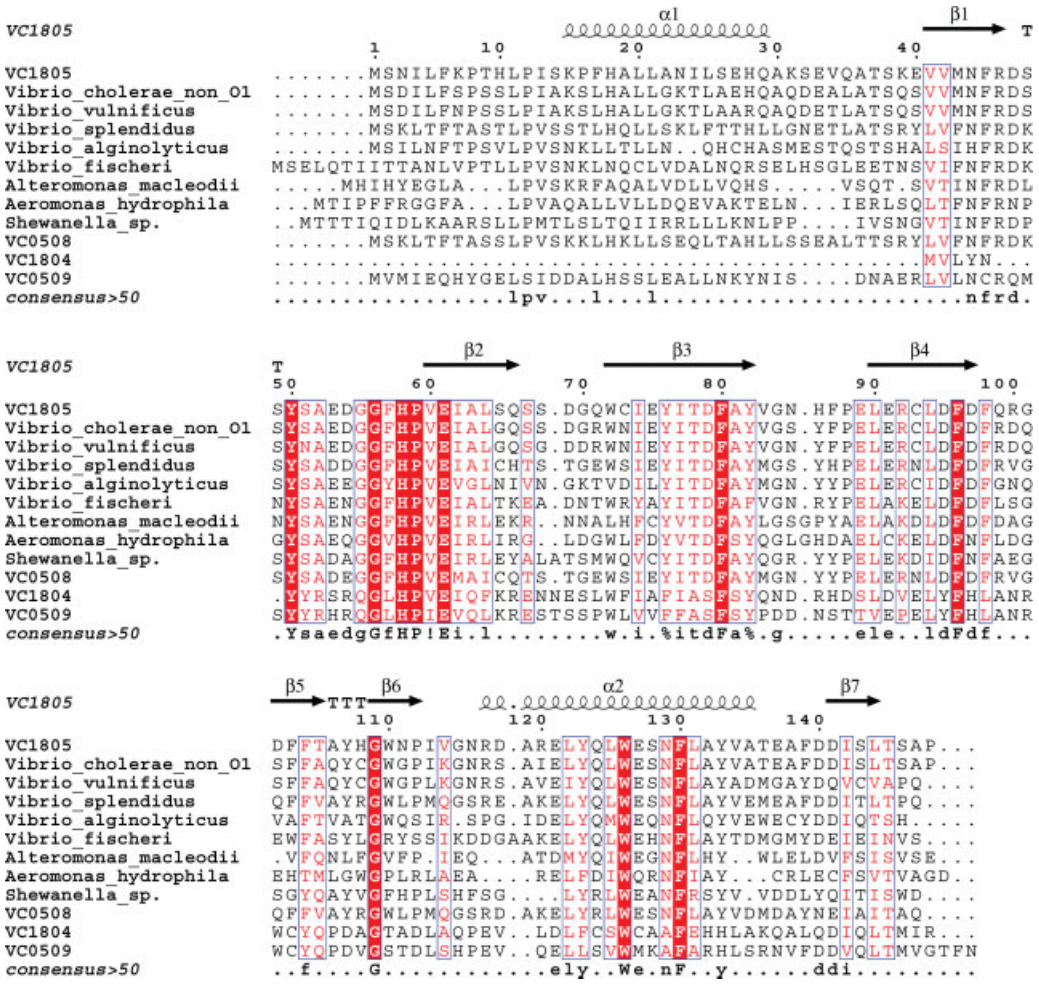

Figure 2

A CLUSTAL-W alignment of VC1805 homologues, orthologues and paralogues. The top line shows the secondary structure of VC1805. NCBI entry names and genome annotation name for the sequences are V. cholerae VC1805 (NP_231440, VC1805), V. cholerae VC0508 (NP_230159, VC0508), V. vulnificus (NP_935046, VV2253), V. splendidus (ZP_00989426, V12B01_02775), V. alginolyticus (ZP_01262456, V12G01_09492), V. fischeri (YP_206478, VFA0520), Alteromonas macleodii (ZO_01108644, MADE_11760), Aeromonas hydrophilia (YP_855622, AHA_1080), Shewanella sp. (YP_963582, Sputw31181_2201), V. cholerae VC1804 (NP_231439, VC1804), V. cholerae VC0509 (NP_230160, VC0509).

of VC1805 was submitted to the ProFunc server that revealed a structural homology with human p32 (PDB code 1P32),[21] an acidic mitochondrial matrix protein. This protein is an evolutionarily conserved eukaryotic protein originally discovered in association with the nuclear pre-mRNA splicing factor SF2/ASF, although no evidence for a splicing function has been reported.[22] Human p32 is a promiscuous protein, reported to bind to many viral proteins,[23–28] transcription factor IIB,[24] the lamnin B receptor,[29] high molecular weight kininogen and factor XII,[30] vitronectin,[31] hyaluronic acid[32] and the globular head domain of the plasma complement component C1q.[33] Human p32 is a homotrimer with each chain containing three helices and a seven-stranded anti-parallel β-sheet, and has a pI of 4.[34] VC1805 and p32 share 16% sequence identity. The conserved fold between the two proteins involves the seven-stranded β-sheet and the helix (α2) that straddles the sheet, although the connectivity is slightly different (see Fig. 3). The rmsd fit between VC1805 and p32 is 2.51 Å over 70 Cα atoms using SSM.[35] Human p32 has two long helices at its N- and C-termini that are involved in the association with other monomers in the p32 trimer. The first β-strand after the N-terminal helix of p32 is in the same location as β7 in VC1805, but from then on the topology is the same for both proteins until the structurally conserved helix (α2) which in p32 is slightly shorter and turns 90° to form the C-terminal helix that associates with the long N-terminal helix. As noted above, in VC1805 the final strand of the sheet (β7) follows helix α2. Although p32 has a low pI, the charge distribution is asymmetric with one face of the p32 trimer carrying a

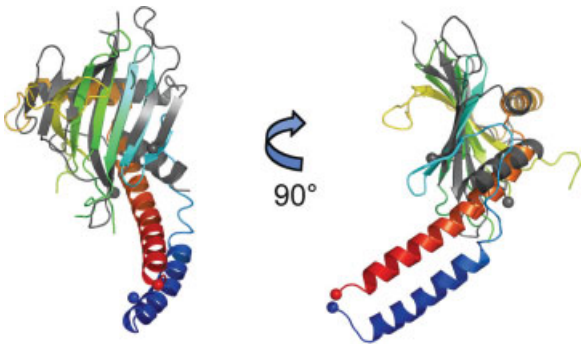

Figure 3

Superposition of VC1805 with a monomer of human p32. VC1805 is shown in grey. The A chain of human p32 is coloured from blue at the N-terminus to red at the C-terminus.

predominantly negative charge, and the other face far less negative. There is little sequence similarity between VC1805 and p32, and the residues forming the two patches conserved within the VC1805 family are not conserved in p32. Conversely the two regions conserved among the eukaryotic homologues of p32 are not conserved in VC1805.[34]

The only homologous structure to p32 in the protein databank is that of a hypothetical protein LMAJ011689 from *Leishmania major* (PDB code 1YQF) that although it shares only 16% sequence identity with human p32, it also has a low pI of 4.5, and possesses the same trimeric structure as p32 with monomers having an RMSD fit of 1.64 Å over 155 Cα atoms using SSM.[35] The p32 fold is therefore an unusual fold in the current protein databank, so the fact that VC1805 shares the same core fold of a helix straddling a seven-stranded anti-parallel β-sheet, and also has a dominant negative charge potential, suggests that the proteins may have a related function.

VC1805 and C1q

The presence of VC1805 in only toxigenic strains of *V. cholerae*, and the presence of a paralogue in the VSP-II pathogenicity island acquired in the seventh and current

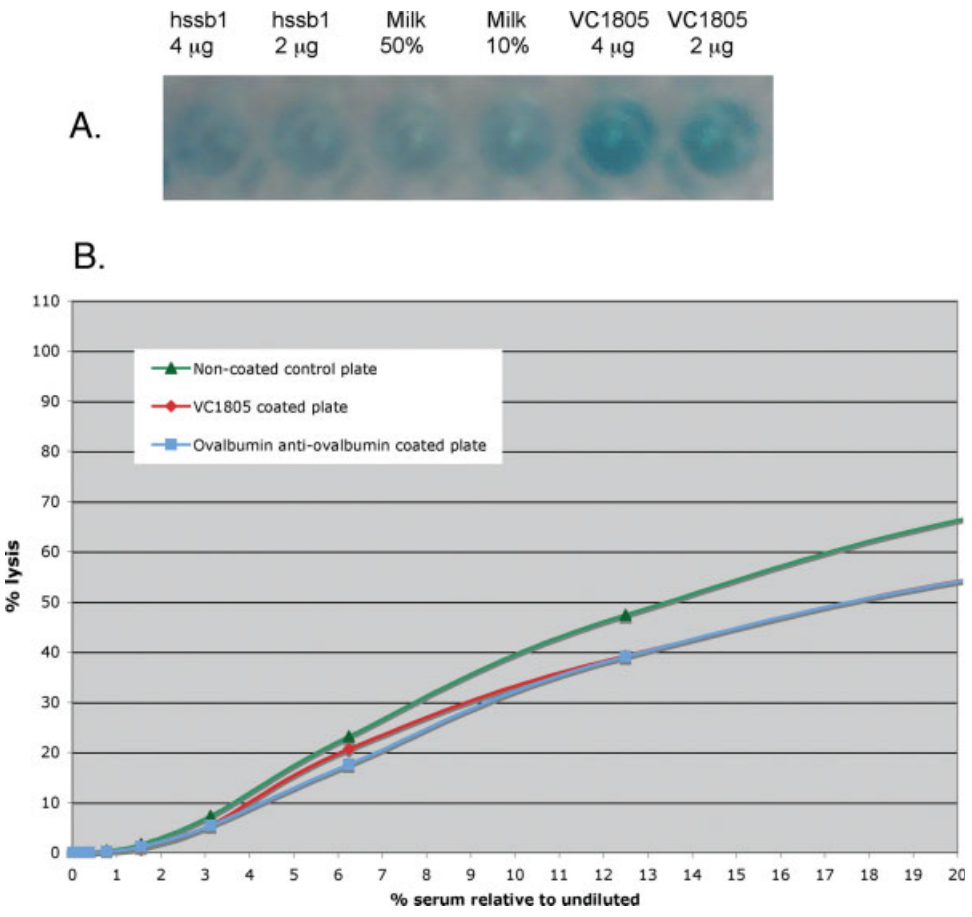

Figure 4

*Binding of VC1805 to C1q. (**A**) ELISA assay showing C1q binding VC1805 with negative controls. (**B**) Complement consumption assay.*

pandemic isolates, suggests that these proteins must have a role that gives an advantage to those strains of *V. cholerae* possessing them. What about the other *Vibrio sp.* and γ-*Proteobacteria* that have homologues of VC1805? *V. vulnificus* is a marine pathogen and etiologic agent of human mortality from seafood-borne infections. *V. spendidus* and *V. alginolyticus* are pathogens of certain shellfish, including oysters and clams. *V. fischeri* is a nonpathogenic beneficial marine species that enters into a mutualistic symbiosis in the light organ of the bobtail squid, *Euprymna scolopes*. *Alteromonas macleodii* and *Shewanella sp.* W3-18-1 are deep sea marine bacteria. *Aeromonas hydrophilia* is found in all freshwater environments and brackish water, and is a pathogen of fish and amphibians that can also cause disease in humans. Of the many reported binding partners of p32, the binding to the head domain of the complement component C1q appeared worthy of exploration. *V. cholerae* is generally accepted to be a noninvasive gut pathogen, and thought to be relatively non-inflammatory[36,37] although a more recent report described an inflammatory response in both O1 and O139 cholera infections.[38] Could VC1805 therefore be engaging C1q as part of a strategy to modify complement attack, or to mediate cell adhesion?

C1q bound to VC1805 immobilised on microtitre plate wells [Fig. 4(A)]. Immobilised VC1805 also caused complement consumption when incubated with human serum [Fig. 4(B)], indicating that VC1805 binds C1 and activates the complement classical pathway. Complement consumption by VC1805 was at a similar level (25–30%) to consumption by immune complexes. C1, a complex of C1q with the serine protease proenzymes C1r and C1s, is the form in which C1q circulates in blood. There was no consistent dose-dependent complement inhibitory effect when soluble VC1805 was added to serum (data not shown). This is consistent with the requirement for multivalent binding of C1q to a target array to initiate complement activation. Thus soluble VC1805, which is a monomer in solution, may bind weakly to sites on the C1q globular heads, but engagement of two or more heads of C1q, and summation of multiple weak interactions, as can occur with an array of immobilized VC1805, is required for high avidity C1q binding and complement activation.

It would not be advantageous for *V. cholerae* to have a protein that would activate complement, but this would be consistent with *V. cholerae* being non-invasive. It is not known if VC1805 is secreted, and if so whether it remains attached to the bacterial surface. What therefore could be the relevance of binding to C1q? It is known that some other complement-related components (DAF-CD55 and MCP-CD46) serve as anchors for gut microbes, but DAF and MCP are cell-surface proteins, not soluble like C1q. C1q is synthesized in gut epithelial cells[39,40] and may be present in the gut lumen, although this has been shown only in inflammatory bowel disease.[41] C1q can associate with membranes via C1q receptors which bind the collagenous region of C1q. The identification of C1q receptors has been controversial, but it is now widely accepted that Calreticulin, bound to cell surfaces via CD91, HLA class 1 heavy chain, or CD59, acts as a receptor for C1q and the related "defence collagens" SPA, SPD and MBL.[42–44] C1q receptor activity is present on lung epithelial cells,[45] but it does not appear to have been investigated in gut epithelium. Perhaps a role for VC1805 is to aid in cell adhesion and to act as a bridge between membrane immobilized C1q and some ligand on the bacterial surface - this may be consistent with the two conserved surface patches on opposite sides of VC1805.

CONCLUSIONS

The structure of VC1805 has an unusual topology with only two structural homologues in the current protein databank: human p32 and a hypothetical protein from *Leishmania major*. The four homologues/paralogues VC0508, VC0509, VC1804, and VC1805 in the *V. cholerae* genome of toxigenic strains are predicted to share the same fold, and all carry a predominantly negative charge, something they also share with human p32. The mitochondrial p32 protein binds to a variety of ligands, and we have chosen to focus on just one of the reported binding partners of human p32, showing that VC1805 does bind to the complement protein C1q. If this is relevant *in vivo*, then one possible role of VC1805 might be to aid adherence of the bacteria to the intestinal wall via C1q bound to C1q receptors. C1q can bind a remarkable variety of immune and non-immune ligands through charge pattern recognition mediated mainly through its globular head domain.[46] The crystal structure of the globular head domain of human C1q revealed the details of the heterotrimeric organization of the three chains each with similar jelly-roll topology.[47] The three modules within the C1q head domain show clear differences in their electrostatic surface potentials, with one module carrying a predominantly positive charge that has been implicated in the C1q-IgG interaction.[47] However, the details of the interactions between C1q and its various ligands remain to be elucidated.[45] It may be fortuitous that VC1805, being negatively charged, happens to be able to bind to C1q *in vitro*, and this may be completely unrelated to its biological function *in vivo*. Human p32 also binds to hyaluronic acid, a non-sulfated glycosaminoglycan found in the extracellular matrix of epithelial tissues. Perhaps hyaluronic acid is the ligand recognized by VC1805. This structural study on VC1805 has stimulated further investigations that will focus on quantifying the interaction with C1q, carrying out pull-down experiments to identify other binding partner(s) of VC1805,

and performing gene knockout experiments to explore the consequences of deleting VC1805 and its paralogues.

ACKNOWLEDGMENTS

The structure of VC1805 was produced using the resources of the Scottish Structural Proteomics Facility at St Andrews. Conor O'Kane (Oxford) kindly provided the biotinylated anti-C1q. We thank Malcolm White (St Andrews) for the gift of hssb1 used as a negative control in the C1q binding assay.

REFERENCES

1. Reidl J, Klose KE. Vibrio cholerae and cholera: out of the water and into the host. FEMS Microbiol Rev 2002;26:125–139.
2. Kaper JB, Morris JG Jr, Levine MM. Cholera. Clin Microbiol Rev 1995;8:48–86.
3. Albert MJ, Siddique AK, Islam MS, Faruque AS, Ansaruzzaman M, Faruque SM, Sack RB. Large outbreak of clinical cholera due to *Vibrio cholerae* non-O1 in Bangladesh. Lancet 1993;341:704.
4. Karaolis DK, Johnson JA, Bailey CC, Boedeker EC, Kaper JB, Reeves PR. A *Vibrio cholerae* pathogenicity island associated with epidemic and pandemic strains. Proc Natl Acad Sci USA 1998;95:3134–3139.
5. Dziejman M, Balon E, Boyd D, Fraser CM, Heidelberg JF, Mekalanos JJ. Comparative genomic analysis of *Vibrio cholerae*: genes that correlate with cholera endemic and pandemic disease. Proc Natl Acad Sci USA 2002;99:1556–1561.
6. Jermyn WS, Boyd EF. Characterization of a novel *Vibrio* pathogenicity island (VPI-2) encoding neuraminidase (nanH) among toxigenic *Vibrio cholerae* isolates. Microbiology 2002;148(Part 11):3681–3693.
7. O'shea YA, Finnan S, Reen FJ, Morrissey JP, O'Gara F, Boyd EF. The *Vibrio* seventh pandemic island-II is a 26.9 kb genomic island present in *Vibrio cholerae* El Tor and O139 serogroup isolates that shows homology to a 43.4 kb genomic island in *V. vulnificus*. Microbiology 2004;150 (Part 12):4053–4063.
8. O'shea YA, Reen FJ, Quirke AM, Boyd EF. Evolutionary genetic analysis of the emergence of epidemic *Vibrio cholerae* isolates on the basis of comparative nucleotide sequence analysis and multilocus virulence gene profiles. J Clin Microbiol 2004;42:4657–4671.
9. Vimr ER, Kalivoda KA, Deszo EL, Steenbergen SM. Diversity of microbial sialic acid metabolism. Microbiol Mol Biol Rev 2004;68: 132–153.
10. Galen JE, Ketley JM, Fasano A, Richardson SH, Wasserman SS, Kaper JB. Role of *Vibrio cholerae* neuraminidase in the function of cholera toxin. Infect Immun 1992;60:406–415.
11. Leslie AGW. Recent changes to the MOSFLM package for processing film and image plate data. Joint CCP4 and ESF-EAMCB newsletter on protein crystallography, 26. Daresbury, Warrington WA, England: SERC Laboratory; 1992.
12. CCP4. The CCP4 Suite: programs for protein crystallography. Acta Crystallogr 1994;D50:760–763.
13. Terwilliger TC, Berendzen J. Automated MAD and MIR structure solution. Acta Crystallogr D Biol Crystallogr 1999;55(Part 4):849–861.
14. Terwilliger TC. Maximum-likelihood density modification. Acta Crystallogr D Biol Crystallogr 2000;56(Part 8):965–972.
15. Terwilliger TC. Automated main-chain model building by template matching and iterative fragment extension. Acta Crystallogr D Biol Crystallogr 2003;59(Part 1):38–44.
16. Terwilliger TC. Automated side-chain model building and sequence assignment by template matching. Acta Crystallogr D Biol Crystallogr 2003;59(Part 1):45–49.
17. Jones TAZJ, Cowan SW, Kjeldgaard M. Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallogr A 1991;47(Part 2):110–119.
18. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr 2004;60(Part 12, Part 1):2126–2132.
19. The CCP4 suite: programs for protein crystallography. Acta Crystallogr D Biol Crystallogr 1994;50(Part 5):760–763.
20. Salvador-Morales C, Flahaut E, Sim E, Sloan J, Green ML, Sim RB. Complement activation and protein adsorption by carbon nanotubes. Mol Immunol 2006;43:193–201.
21. Laskowski RA, Watson JD, Thornton JM. ProFunc: a server for predicting protein function from 3D structure. Nucleic Acids Res 2005;33(Web Server issue):W89–W93.
22. Krainer AR, Mayeda A, Kozak D, Binns G. Functional expression of cloned human splicing factor SF2: homology to RNA-binding proteins. U1 70K, and Drosophila splicing regulators Cell 1991;66:383–394.
23. Desai K, Loewenstein PM, Green M. Isolation of a cellular protein that binds to the human immunodeficiency virus Tat protein and can potentiate transactivation of the viral promoter. Proc Natl Acad Sci USA 1991;88:8875–8879.
24. Yu L, Loewenstein PM, Zhang Z, Green M. *In vitro* interaction of the human immunodeficiency virus type 1 Tat transactivator and the general transcription factor TFIIB with the cellular protein TAP. J Virol 1995;69:3017–3023.
25. Tange TO, Jensen TH, Kjems J. *In vitro* interaction between human immunodeficiency virus type 1 Rev protein and splicing factor ASF/SF2-associated protein, p32. J Biol Chem 1996;271:10066–10072.
26. Wang Y, Finan JE, Middeldorp JM, Hayward SD. P32/TAP, a cellular protein that interacts with EBNA-1 of Epstein-Barr virus. Virology 1997;236:18–29.
27. Bruni R, Roizman B. Open reading frame P—a herpes simplex virus gene repressed during productive infection encodes a protein that binds a splicing factor and reduces synthesis of viral proteins made from spliced mRNA. Proc Natl Acad Sci USA 1996;93:10423–10427.
28. Matthews DA, Russell WC Adenovirus core protein V interacts with p32—a protein which is associated with both the mitochondria and the nucleus. J Gen Virol 1998;79(Part 7):1677–1685.
29. Simos G, Georgatos SD. The lamin B receptor-associated protein p34 shares sequence homology and antigenic determinants with the splicing factor 2-associated protein p32. FEBS Lett 1994;346:225–228.
30. Herwald H, Dedio J, Kellner R, Loos M, Muller-Esterl W. Isolation and characterization of the kininogen-binding protein p33 from endothelial cells. Identity with the gC1q receptor. J Biol Chem 1996;271:13040–13047.
31. Lim BL, Reid KB, Ghebrehiwet B, Peerschke EI, Leigh LA, Preissner KT. The binding protein for globular heads of complement C1q, gC1qR. Functional expression and characterization as a novel vitronectin binding factor. J Biol Chem 1996;271:26739–26744.
32. Deb TB, Datta K. Molecular cloning of human fibroblast hyaluronic acid-binding protein confirms its identity with P-32, a protein co-purified with splicing factor SF2. Hyaluronic acid-binding protein as P-32 protein, co-purified with splicing factor SF2. J Biol Chem 1996;271:2206–2212.
33. Ghebrehiwet B, Lim BL, Peerschke EI, Willis AC, Reid KB. Isolation, cDNA cloning, and overexpression of a 33-kD cell surface glycoprotein that binds to the globular "heads" of C1q. J Exp Med 1994;179:1809–1821.
34. Jiang J, Zhang Y, Krainer AR, Xu RM. Crystal structure of human p32, a doughnut-shaped acidic mitochondrial matrix protein. Proc Natl Acad Sci USA 1999;96:3572–3577.
35. Krissinel E, Henrick K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta Crystallogr D Biol Crystallogr 2004;60(Part 12, Part 1):2256–2268.

36. Satchell KJ. Activation and suppression of the proinflammatory immune response by *Vibrio cholerae* toxins. Microbes Infect 2003;5: 1241–1247.
37. Silva TM, Schleupner MA, Tacket CO, Steiner TS, Kaper JB, Edelman R, Guerrant R. New evidence for an inflammatory component in diarrhea caused by selected new, live attenuated cholera vaccines and by El Tor and Q139 *Vibrio cholerae*. Infect Immun 1996;64: 2362–2364.
38. Qadri F, Bhuiyan TR, Dutta KK, Raqib R, Alam MS, Alam NH, Svennerholm AM, Mathan MM. Acute dehydrating disease caused by *Vibrio cholerae* serogroups O1 and O139 induce increases in innate cells and inflammatory mediators at the mucosal surface of the gut. Gut 2004;53:62–69.
39. Loos M, Martin H, Petry F. The biosynthesis of C1q, the collagen-like and Fc-recognizing molecule of the complement system. Behring Inst Mitt 1989;84:32–41.
40. Morris KM, Colten HR, Bing DH. The first component of complement. A quantitative comparison of its biosynthesis in culture by human epithelial and mesenchymal cells. J Exp Med 1978;148: 1007–1019.
41. Halstensen TS, Mollnes TE, Garred P, Fausa O, Brandtzaeg P. Surface epithelium related activation of complement differs in Crohn's disease and ulcerative colitis. Gut 1992;33:902–908.
42. Sim RB, Moestrup SK, Stuart GR, Lynch NJ, Lu J, Schwaeble WJ, Malhotra R. Interaction of C1q and the collectins with the potential receptors calreticulin (cC1qR/collectin receptor) and megalin. Immunobiology 1998;199:208–224.
43. Ogden CA, deCathelineau A, Hoffmann PR, Bratton D, Ghebrehiwet B, Fadok VA, Henson PM. C1q and mannose binding lectin engagement of cell surface calreticulin and CD91 initiates macropinocytosis and uptake of apoptotic cells. J Exp Med 2001;194:781–795.
44. Ghiran I, Klickstein LB, Nicholson-Weller A. Calreticulin is at the surface of circulating neutrophils and uses CD59 as an adaptor molecule. J Biol Chem 2003;278:21024–21031.
45. Malhotra R, Haurum J, Thiel S, Jensenius JC, Sim RB. Pollen grains bind to lung alveolar type II cells (A549) via lung surfactant protein A (SP-A). Biosci Rep 1993;13:79–90.
46. Ghai R, Waters P, Roumenina LT, Gadjeva M, Kojouharova MS, Reid KB, Sim RB, Kishore U. C1q and its growing family. Immunobiology 2007;212:253–266.
47. Gaboriaud C, Juanhuix J, Gruez A, Lacroix M, Darnault C, Pignol D, Verger D, Fontecilla-Camps JC, Arlaud GJ. The crystal structure of the globular head of complement protein C1q provides a basis for its versatile recognition properties. J Biol Chem 2003;278: 46974–46982.