Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structural studies of human glioma pathogenesis-related protein 1

Oluwatoyin A. Asojo,[a]* Raymond A. Koski[b] and Nathalie Bonafé[b]

[a]Department of Pathology and Microbiology, College of Medicine, Nebraska Medical Center, Omaha, NE 68198-6495, USA, and [b]L2 Diagnostics LLC, 300 George Street, New Haven, CT 06511, USA

Correspondence e-mail: oasojo@unmc.edu

Received 21 June 2011
Accepted 14 July 2011

PDB References: sGLIPR1, 3q2u; Zn^{2+} complex, 3q2r.

Human glioma pathogenesis-related protein 1 (GLIPR1) is a membrane protein that is highly upregulated in brain cancers but is barely detectable in normal brain tissue. GLIPR1 is composed of a signal peptide that directs its secretion, a conserved cysteine-rich CAP (cysteine-rich secretory proteins, antigen 5 and pathogenesis-related 1 proteins) domain and a transmembrane domain. GLIPR1 is currently being investigated as a candidate for prostate cancer gene therapy and for glioblastoma targeted therapy. Crystal structures of a truncated soluble domain of the human GLIPR1 protein (sGLIPR1) solved by molecular replacement using a truncated polyalanine search model of the CAP domain of stecrisp, a snake-venom cysteine-rich secretory protein (CRISP), are presented. The correct molecular-replacement solution could only be obtained by removing all loops from the search model. The native structure was refined to 1.85 Å resolution and that of a Zn^{2+} complex was refined to 2.2 Å resolution. The latter structure revealed that the putative binding cavity coordinates Zn^{2+} similarly to snake-venom CRISPs, which are involved in Zn^{2+}-dependent mechanisms of inflammatory modulation. Both sGLIPR1 structures have extensive flexible loop/turn regions and unique charge distributions that were not observed in any of the previously reported CAP protein structures. A model is also proposed for the structure of full-length membrane-bound GLIPR1.

1. Introduction

Glioma pathogenesis-related protein 1 (GLIPR1), also known as related to testis-specific, vespid and pathogenesis protein 1 (RTVP-1), is a member of the CAP (**c**ysteine-rich secretory proteins, **a**ntigen 5 and **p**athogenesis-related 1 proteins) superfamily of proteins (Gibbs *et al.*, 2008). As is the case in other mammals, the 31 human CAP proteins belong to nine families: cysteine-rich secretory protein (CRISP), GLIPR1, glioma pathogenesis-related protein 2 (GLIPR2), otherwise referred to as Golgi-associated pathogenesis-related 1 (GAPR-1), peptidase inhibitor 15 (PI15), peptidase inhibitor 16 (PI16), cysteine-rich secretory protein lectin domain 1 (CRISPLD1), cysteine-rich secretory protein lectin domain 2 (CRISPLD2), mannose receptor-like/CAP domain and CTL domain containing-like (CAPCL), and R3H domain containing-like peptidase inhibitor (R3HDML). All of these proteins have a CAP domain, but vary in the other domains that they possess. Overall, the human CAP-protein subfamilies only share limited sequence identity with each other; however, members of subfamilies share higher sequence identity in their CAP domains. A comprehensive review has described mammalian CAP proteins and their roles in reproduction, cancer and immune defense (Gibbs *et al.*, 2008).

Table 1
Statistics of data collection and model refinement.

Values in parentheses are for the highest resolution shell.

	sGLIPR1 (PDB entry 3q2u)	Zn^{2+} complex (PDB entry 3q2r)
Space group	$P2_12_12$	$P2_12_12$
Unit-cell parameters (Å)	a = 85.1, b = 79.5, c = 38.8	a = 85.9, b = 79.7, c = 38.8
Resolution limits (Å)	29.0–1.85 (1.95–1.85)	27.8–2.20 (2.30–2.20)
$\langle I/\sigma(I)\rangle$	20.1 (3.2)	10.4 (5.0)
No. of reflections	279138 (16237)	86445 (12294)
No. of unique reflections	23229 (3305)	18698 (2680)
Multiplicity	12.0 (4.9)	6.1 (6.1)
$R_{\rm merge}$† (%)	9.3 (49.6)	14.5 (40.2)
Completeness (%)	99.9 (99.4)	99.9 (100)
Refinement		
$R_{\rm cryst}$‡/$R_{\rm free}$§	13.4/18.4 (10.0/22.2)	15.8/19.2 (14.6/20.4)
Correlation coefficient		
$F_{\rm o} - F_{\rm c}$	0.965	0.933
$F_{\rm o} - F_{\rm c}$ (free)	0.942	0.910
R.m.s. deviations		
Bond lengths (Å)	0.026	0.007
Bond angles (°)	1.888	1.884
Model composition		
Monomers	1	1
Residues	193	192
Waters	286	258
Zn^{2+}	0	1
Ramachandran (%)		
Preferred regions	96.34	94.2
Outliers	1.0	0.53

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ and $\langle I(hkl)\rangle$ are the intensity of the ith measurement and the mean intensity of the reflection with indices hkl, respectively. ‡ $R_{\rm cryst} = \sum_{hkl} \big||F_{\rm obs}| - |F_{\rm calc}|\big| / \sum_{hkl} |F_{\rm obs}|$, where $F_{\rm obs}$ are observed and $F_{\rm calc}$ are calculated structure-factor amplitudes. § The $R_{\rm free}$ set uses a randomly chosen 5% of reflections.

In humans there are three core members of the GLIPR1 family: GLIPR1, GLIPR1-like 1 (GLIPR1L1) and GLIPR1-like 2 (GLIPR1L2). These have a conserved N-terminal CAP domain and isoforms with variations in the C-terminal extension. Their genes lie in a p53-regulated cluster (Gibbs *et al.*, 2010). While GLIPR1 and GLIPR1L2 are expressed in various tissues, GLIPR1L1 is mainly expressed in testis (Ren *et al.*, 2006). The GLIPR1-family members GLIPR1 and GLIPR1L2 have a signal peptide followed by a CAP domain and a transmembrane domain. These two predicted carboxyl-terminal transmembrane domains are unique to mammalian CAP proteins and are not present in all GLIPR1-family isoforms. RVTP-1b, a splice variant of GLIPR1, has no sequence conservation beyond the N-terminal 141 amino-acid residues (Gibbs & O'Bryan, 2007; Gibbs *et al.*, 2008, 2010).

GLIPR1 is expressed in low levels in diverse human tissues but is highly expressed in astrocytic brain malignancies (Rosenzweig *et al.*, 2006; Xiang *et al.*, 2007). Its expression levels correlate with tumor-cell invasiveness, *i.e.* glioblastoma > anaplastic astrocytomas > low-grade astrocytomas > normal brain (Rosenzweig *et al.*, 2006). GLIPR1 has been proposed to act as a tumor suppressor that undergoes epigenetic inactivation in prostate cancer (Gibbs *et al.*, 2008; Ren *et al.*, 2002, 2004; Satoh *et al.*, 2003; Thompson, 2010), in which the *Glipr1* gene has been shown to be methylated and down-regulated compared with normal human prostate. *Glipr1* gene-transfer approaches, which were initially used to study tumor-suppressor functions, are now being evaluated in clinical trials for prostate cancer gene therapy (Thompson, 2010). Characterizing the structure of GLIPR1 is a first step to clarifying the functions of this CAP protein.

The 15 kDa CAP domain is a structurally conserved cysteine-rich domain historically referred to as the SCP (sperm coating glycoprotein), NCBI cd00168 or Pfam PF00188 domain. It is found in proteins secreted under conditions ranging from plant responses to pathogens, human brain tumor growth, host entry by hookworms and other parasites, to sperm–egg interactions, as well as in venom proteins from insects and reptiles (Gibbs *et al.*, 2006; Ding *et al.*, 2000; Hawdon *et al.*, 1999; Zhan *et al.*, 2003; Gao *et al.*, 2001; Hotez *et al.*, 2003). CAP proteins share limited sequence identity with each other and they have two signature PROSITE-recognized motifs, which are referred to as CRISP motifs (http://www.expasy.ch/prosite). The consensus sequence for the CRISP1 motif is (GDER)(HR)(FYWH)(TVS)(QA)(LIVM)-(LIVMA)W*xx*(STN) and that for the CRISP2 motif is (LIVMFYH)(LIVMFY)*x*C(NQRHS)Y*x*(PARH)*x*(GL)N(LIVNFYWDN). The CRISP2 motif plays key roles in sperm–egg interactions (Cohen *et al.*, 2008). Mammalian CAP proteins have two additional signature motifs, HN*xx*R and G(EQ)-N(ILV), or CRISP3 and CRISP4 motifs, respectively. The functions of these motifs are unknown; however, these residues are in proximity to or lie in the conserved central CAP cavity observed in all reported structures of CAP proteins (Asojo, 2011).

There are several reported structures of CAP proteins in the Protein Data Bank. The vast majority of reported CAP structures are of snake-venom CRISPs, including PDB entries 2dda, 2ddr, 1xx5, 1xta, 3mz8, 1wvr and 1rc9 (Suzuki *et al.*, 2008; Wang *et al.*, 2005, 2010; Shikamoto *et al.*, 2005; Guo *et al.*, 2005). Two reported CAP structures, PDB entries 1u53 and 1nt8, are of *Ancylosoma*-secreted proteins from hookworms (Asojo *et al.*, 2005; Asojo, 2011). There are also two structures of venoms secreted by yellow jackets and fire ants (PDB entries 1qnx and 2vzn; Padavattan *et al.*, 2008; Henriksen *et al.*, 2001) and one structure of a pathogenesis-related protein from tomato (PDB entry 1cfe; Fernández *et al.*, 1997). There is currently only one reported three-dimensional structure of a mammalian CAP protein (PDB entry 1smb), that of the GLIPR2 (GAPR-1) protein, which is also the smallest of all the mammalian CAP-protein members (Serrano *et al.*, 2004). These structures reveal homology in a conserved α–β–α sandwich core, while each structure has variations in the length of the helices and strands and the location and orientation of the helices (Asojo, 2011). These structural differences are likely to affect the ability of CAP proteins to bind to different binding partners (Asojo *et al.*, 2005). Despite the low primary sequence similarity, all reported CAP-protein structures are characterized by a conserved central CAP cavity. This CAP cavity has been shown to be the major Zn^{2+}-binding site in the Zn^{2+} and heparin-sulfate dependent mechanisms of inflammatory modulation by the cobra CRISP natrin (Wang *et al.*, 2010). It is unknown whether similar mechanisms are

conserved in mammalian CAP proteins. While there are extensive structural studies of snake-venom CRISPs, there is an absence of structural data on the various mammalian CAPs. Thus, we have solved the structure of recombinant soluble human GLIPR1 (sGLIPR1).

2. Materials and methods

2.1. Deglycosylation

Recombinant sGLIPR1 protein was assessed for glycosylation by digestion with peptide *N*-glycosidase F (PNGase F) and endoglycosidase H (Endo H). Aliquots of sGLIPR1 were treated with 2000 units of each enzyme in its respective buffer according to the manufacturer's instructions (New England BioLabs) or without enzyme as a control. Digestions were carried out for 3 h at 310 K. Samples were then analyzed by SDS–PAGE (12% gels) and Coomassie Blue staining.

2.2. Structure solution and refinement

Recombinant sGLIPR1 expression, purification, crystallization and data collection have been described elsewhere (Bonafé *et al.*, 2010). In addition to the previously reported conditions, crystals with the same morphology and similar unit-cell parameters were obtained using precipitant consisting of 0.17 *M* acetate, 0.085 *M* Tris–HCl pH 8.5, 25.5% PEG 4000 and 15% glycerol as well as by replacing the acetate with a comparable concentration of ammonium sulfate. Crystallographic data are shown in Table 1. Initial phases were obtained by molecular replacement (MR) with the program *Phaser* (McCoy *et al.*, 2007) using a truncated polyalanine search model containing only the α-helices and strand regions of the CAP domain of stecrisp (PDB entry 1rc9; Guo *et al.*, 2005). The initial MR solution resulted in one monomer per asymmetric unit, which, based on the volume of the unit cell being 26 992 Å^3, corresponds to a Matthews coefficient of 3.11 Å^3 Da^{-1} (60% solvent). MR was followed by automatic model building using the *CCP*4 statistical protein chain-tracing program *Buccaneer* (Cowtan, 2006). The structure was improved through iterative manual model-building cycles using *Coot* (Emsley & Cowtan, 2004) followed by refinement using both *PHENIX* (Afonine *et al.*, 2005; Adams *et al.*, 2010) and *REFMAC*5 (Murshudov *et al.*, 1999) with free *R* (Brünger, 1992) to yield a final model with the statistics listed in Table 1.

The complex with Zn^{2+} was generated by soaking preformed crystals for 10 min with 0.2 m*M* zinc chloride in crystallization buffer comprised of 0.085 *M* sodium cacodylate pH 6.5, 25.5%(*w*/*v*) PEG 8000, 0.17 *M* ammonium sulfate, 15%(*w*/*v*) glycerol. Data from the Zn^{2+}-soaked crystals were collected using a four-circle κ platform Xcalibur PX Ultra with a 165 mm diagonal Onyx CCD detector and a high-brilliance sealed-tube Enhance Ultra (Cu) X-ray source (Oxford Diffraction, Oxford, England) operating at 50 kV and 50 mA at a crystal-to-detector distance of 65 mm and with an exposure time of 150 s per 0.5° oscillation. The data sets were processed using the program *CrysAlis*^{Pro} (Oxford Diffraction). The structure was solved by MR using the structure of sGLIPR1 as the search model. The Zn^{2+} ion was located in the initial MR $F_o - F_c$ electron-density maps at 7σ contour levels. This was followed by iterative cycles of model building with *Coot* as well as structure refinement with *REFMAC*5 and *PHENIX* to yield a model with the statistics shown in Table 1.

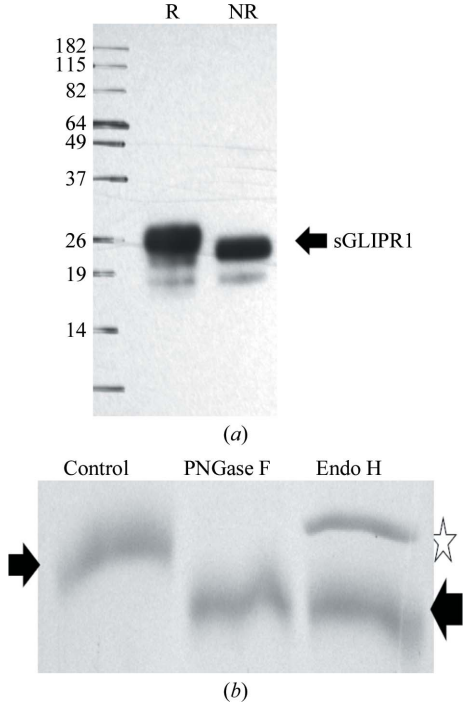

Figure 1
GLIPR1 was expressed in *P. pastoris* as a glycosylated monomeric protein. (*a*) SDS–PAGE Western blot analysis of purified sGLIPR1 using anti-GLIPR1$_{75–95}$ peptide antibody under reducing (R) or nonreducing (NR) conditions. The positions of molecular-weight markers are indicated in kDa. (*b*) Coomassie-stained SDS–PAGE gel depicting sGLIPR1 deglycosylation by peptide *N*-glycosidase F (PNGase F) and endoglycosidase H (Endo H), which resulted in faster migration of the protein (right arrow) compared with nondigested control (left arrow). The band indicated with a star is Endo H.

3. Results and discussion

3.1. Recombinant sGLIPR1 protein

Human GLIPR1 (NCBI accession No. NP_006842) is naturally synthesized as a 255-amino-acid precursor with a signal peptide and a transmembrane domain that localizes the mature protein to the cell membrane in glioblastoma cells (N. Bonafé, unpublished work). The first 21 residues make up the signal peptide, while the C-terminal residues form the predicted membrane-spanning domain. The recombinant sGLIPR1 protein consists of amino acids 22–220 of human GLIPR1 with a *Pichia pastoris* signal peptide and the linker amino-acid sequence EAEAEF added to the N-terminus by the cloning procedures (Bonafé *et al.*, 2010). The recombinant protein was truncated before the C-terminal GLIPR1 transmembrane domain and no purification tag was added. Our choice of numbering for our structural model is that the first residue of the mature peptide is residue number 22. Purified recombinant sGLIPR1 is a monomer with an observed

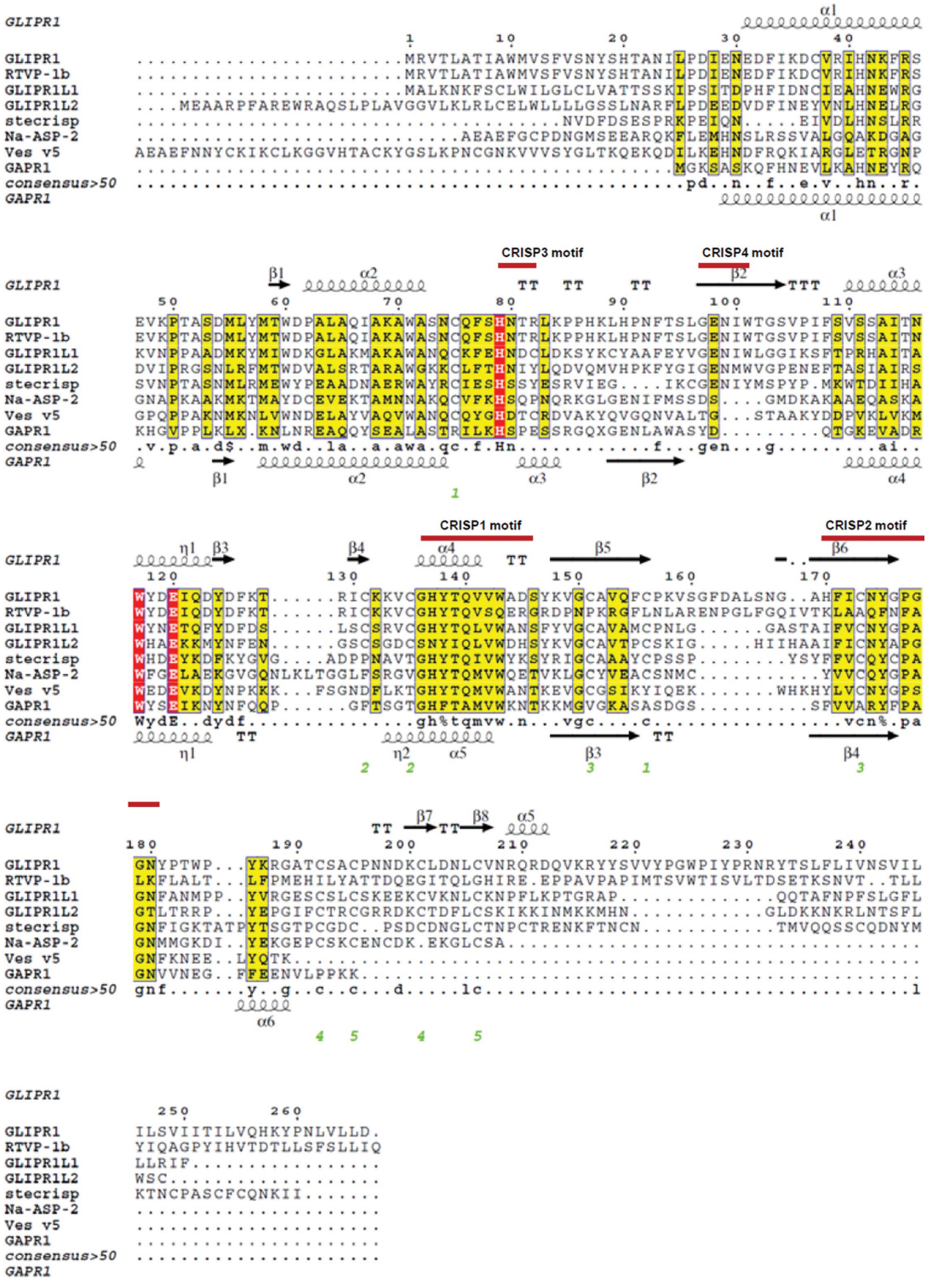

Figure 2
Structural features of sGLIPR1 and primary sequence alignment of sGLIPR1 with other members of the GLIPR1 family and representative CAP-protein structures. This figure was generated with *ESPript* (Gouet *et al.*, 2003). Also shown are the structural elements of GAPR-1. Secondary-structure elements are shown as follows: α-helices are shown as large squiggles labeled α, 3_{10}-helices are shown as small squiggles labeled η, β-strands are shown as arrows labeled β and β-turns are labeled TT. Identical residues are highlighted in red and conserved residues are highlighted in yellow. The locations of the cysteine residues involved in sGLIPR1 disulfide bonds are numbered in green and the signature CRISP motifs are identified with red bars. The members of the GLIPR1 subfamily used are GLIPR1 (NP_006842), RTVP-1b (ABV21587.1), GLIPR1L1 (NP689992.1) and GLIPR1L2 (NP689649). RTVP-1b is a GLIPR1 splice variant with C-terminal divergence. The highest variability for this group is in the signal peptide and loop regions. The representative CAP structures are Na-ASP-2 (PDB entry 1u53), Ves v 5 (PDB entry 1qnx), GAPR-1 (PDB entry 1smb) and stecrisp (PDB entry 1rc9). The AEAEF leader sequences in Na-ASP-2 and Ves v 5 are from the expression vector. In these representative structures, the helices and strand regions of the core CAP domain are well conserved, with the highest variability in the loop regions of sGLIPR1.

molecular mass greater than the theoretical molecular weight of 23 295 Da (Fig. 1*a*). Electrospray mass-spectrometric analyses (data not shown) and deglycosylation experiments suggested the presence of glycosylation (Fig. 1*b*). There is one predicted N-linked glycosylation site at residue Asn92. There is insufficient electron density in proximity to Asn92 to model the glycans in any of the electron-density maps.

3.2. Structure of sGLIPR1

All of our MR attempts using multiple CAP search models containing loop regions failed. Structure determination was accomplished by MR using a truncated polyalanine model of the core helices and strands of snake-venom stecrisp (PDB entry 1rc9) as the search model. This allowed us to phase with a search model that had only 31% sequence conservation. After refinement and model building, a model was obtained in which 193 amino acids had unambiguous main-chain density in *REFMAC*5-weighted $2F_o - F_c$ maps contoured at 1σ. The first GLIPR1 residue Ala22 only has visible electron density in 0.8σ-contoured maps. The C-terminus (KRYYS) and plasmid-incorporated linker sequence (EAEAEF) are disordered and cannot be modeled. The only additional disordered regions in 0.8σ-contoured maps are the side chains of residues that are solvent exposed and lack substantial contacts with other amino acids, notably Lys84, Lys133, Asn197, Asn198 and Asn204. The structure has a high solvent content (60%), but there are also regions of density that we did not model that are clearly not water molecules. Some could be modeled as glycerol molecules, but others had irregular shape and were not modeled. The coordinates and structure factors for the native and zinc chloride-soaked structures have been deposited in the RCSB Protein Data Bank with accession codes 3q2u and 3q2r, respectively.

The topology of sGLIPR1 consists of an N-terminal loop followed directly by the CAP domain and a cysteine-rich C-terminal region (Figs. 2 and 3). The tertiary structure of the conserved core CAP motif of sGLIPR1 is a three-stranded antiparallel β-sheet sandwiched between two layers of α-helices. As is typical of all CAP domains, one α-helical layer is composed of two parallel α-helices ($\alpha1$ and $\alpha3$), while the other has a solitary α-helix ($\alpha2$) (Fig. 3). While $\alpha1$ and $\alpha2$ are

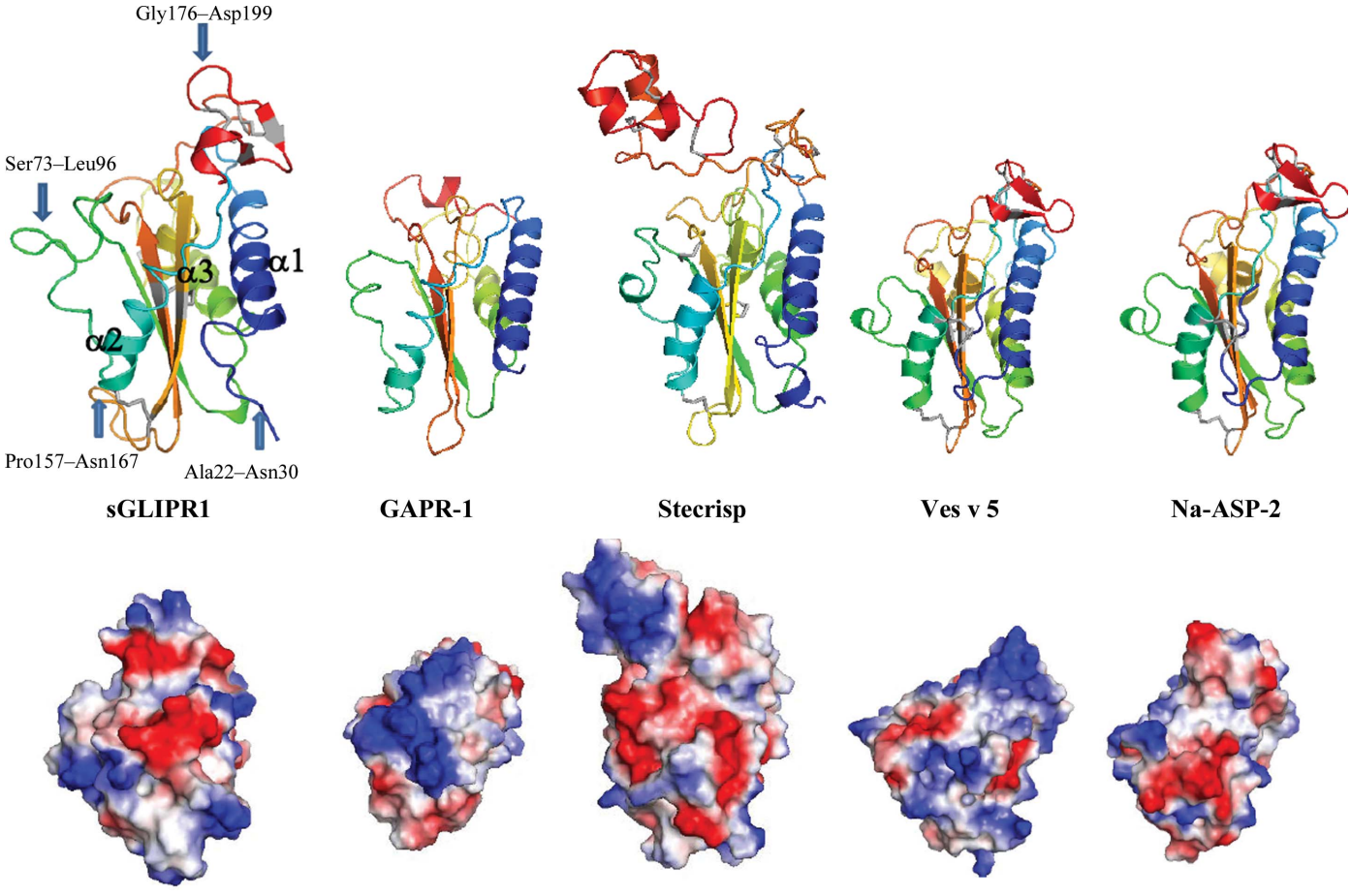

Figure 3
Comparison of the sGLIPR1 structure with representative CAP structures. The top row shows a ribbon diagram of sGLIPR1 (PDB entry 3q2u), revealing a conserved core CAP domain similar to those observed in the representative CAP structures GAPR-1, stecrisp, Ves v 5 and Na-ASP-2. This core α–β–α sandwich is formed by the three core β-strands flanked by the labeled helices. Arrows indicate the locations of loop/turn regions in sGLIPR1 that are longer than in the other structures. The locations of disulfide bonds are shown in gray. The bottom row reveals that the surface-charge distributions differ for these representative CAP structures. These are colored from red for negatively charged regions to blue for positively charged regions. The same view is shown for all CAP structures.

Table 2
Comparison of sGLIPR1 with other CAP proteins.

(*a*) R.m.s.d. between sGLIPR1 and CAP structures (Å).

Chains	3q2r *A*, sGLIPR1–Zn^{2+}	3q2u *A*, sGLIPR1	1qnx *A*, Ves v 5	1rc9 *A*, stecrisp	1smb *A*, GAPR-1	1u53 *A*, Na-ASP-2
3q2r *A*		0.49	1.71	1.48	2.96	1.92
3q2u *A*	0.49		1.70	1.44	2.94	1.83
1qnx *A*	1.71	1.70		1.43	2.54	1.36
1rc9 *A*	1.48	1.44	1.43		2.72	1.48
1smb *A*	2.96	2.94	2.54	2.72		2.62
1u53 *A*	1.92	1.83	1.36	1.48	2.62	

(*b*) Sequence-identity comparison between GLIPR1 and CAP structures (%).

Protein	GLIPR1	sGLIPR1	Stecrisp	Ves v 5	GAPR-1	Na-ASP-2
GLIPR1	100	100	28	21	16	26
sGLIPR1	100	100	31	22	16	26
Stecrisp	28	31	100	21	20	30
Ves v 5	21	22	21	100	26	25
GAPR-1	16	16	20	26	100	18
Na-ASP-2	26	26	30	25	18	100

(*c*) Sequence-identity comparison of human GLIPR1-family proteins (%).

Protein	GLIPR1	RTVP-1b	GLIPR1L1	GLIPR1L2
GLIPR1	100	53	39	32
RTVP-1b	53	100	24	19
GLIPR1L1	39	39	100	31
GLIPR1L1	32	19	31	100

regular α-helices, α3 contains an α-helix that turns into a 3_{10}-helix, reminiscent of the Na-ASP-2 structure (Fig. 3). The CAP-domain antiparallel β-sheet (Fig. 3) is formed by the three longest β-strands: β2 (Gly97–Ser104), β5 (Lys148–Cys156) and β6 (Gly168–Gly176). Directly following the CAP domain are residues that are likely to serve as an anchor between the extracellular CAP domain and the transmembrane domain. This C-terminal region is structurally similar to those of other CAPs, notably Na-ASP-2 and all reported snake-venom CRISP structures. It consists of a two-stranded β-sheet linked by disulfide bonds 4 (Cys192–Cys201) and 5 (Cys195–Cys206) and terminates in a short two-turn helix. Connecting these secondary-structure motifs are unique loop regions. All cysteines in sGLIPR1 are involved in disulfide bonds except for Cys37. There are a total of five disulfide bonds in sGLIPR1 (Fig. 2).

Of the 193 residues in the sGLIPR1 model, 54.4% are in loop or turn regions, including Ala22–Asn30, Glu47–Met58, Ser73–Leu96, Val105–Ser109, Phe126–Arg129, Pro157–Asn167, Gly176–Asp199 and Leu202–Asn204 (Fig. 2). These are mostly on the surface of the structure (Fig. 3). The unique lengths and orientations of these loop/turn regions are likely to explain why phasing by MR failed when the loops were not removed from the initial search model. These structurally nonhomologous regions made significant contributions to the initial MR search model, as we had observed in our structural studies of hookworm CAP proteins (Asojo *et al.*, 2005; Asojo, 2011). These regions are quite different from those predicted in the homology model of sGLIPR1 (Szyperski *et al.*, 1998). It is important to point out that the lengths of the helices and strands, and the orientations and locations of the loops, in sGLIPR1 could not be predicted based on any previous CAP structures.

3.3. Comparison of sGLIPR1 with representative CAP structures

The structures of various CAP proteins have been solved and representative structures have varied amino-acid sequence identities between a high of 31% and a low of 16% (Table 2*b*). Comparison of the sGLIPR1 structure with those of representative members of CAP-protein families, Na-ASP-2 (PDB entry 1u53; Asojo *et al.*, 2005), GAPR-1 (PDB entry 1smb; Serrano *et al.*, 2004), Ves v 5 (PDB entry 1qnx; Henriksen *et al.*, 2001) and stecrisp (PDB entry 1rc9; Guo *et al.*, 2005), reveals that the core secondary-structure elements of the α–β–α sandwich in the CAP domain are conserved, albeit with different lengths (Fig. 3). Additionally, three of the five disulfide bonds in sGLIPR1 are conserved in both Na-ASP-2 and stecrisp. The importance of the disulfides is tempered by the observation that the CAP core of crystallized recombinant cysteine-free GAPR-1 is stable without any disulfide bonds (Serrano *et al.*, 2004). Despite the tertiary-structure similarity of the CAP domain of these representative structures, there does not appear to be any consensus in the overall charge distribution among representative CAP structures. The charge distribution of sGLIPR1 is more clustered than those of the other structures. Furthermore, the loop regions have charge variations across these representative structures (Fig. 3).

Uniquely, sGLIPR1 has longer, differently positioned and more protrusive loop/turn regions linking the secondary-structure regions than any of the previously reported CAP structures (Figs. 2 and 3). These regions in sGLIPR1 reveal the greatest differences when compared with other CAP sequences, whereas the two signature *PROSITE* (http://www.expasy.org/cgi-bin/prosite/PSScan.cgi) CRISP motifs are largely conserved. Apart from the loop/turn regions, the C- and N-terminal residues of these representative CAP structures have the greatest variation (Fig. 2). The differences in these representative CAP structures were quantified as r.m.s.d. calculated by structural alignment using *MOE* (http://www.chemcomp.com). Although GLIPR1 and GAPR-1 are both human proteins, they are the most structurally dissimilar, with main-chain atom r.m.s.d.s of greater than 2.9 Å and only 16% identity (Tables 2*a* and 2*b*). Owing to the extensive variation of the loops/turns and terminal regions, none of the other CAP-protein structures have an r.m.s.d. of lower than 1.3 Å when compared with sGLIPR. Overall, sGLIPR1 is most structurally similar to snake-venom stecrisp, followed by Ves v 5 and Na-ASP-2 (Tables 2*a* and 2*b*, Figs. 2 and 3).

3.4. The CAP cavity

Despite the differences in overall charge distribution and loop regions, CAP proteins are characterized by a large central cavity containing key conserved charged residues. Among these residues in GLIPR1 is His137, which is part of the *PROSITE*-recognized CRISP signature motif 1 (Fig. 2).

The conserved charged residues that lie in the putative binding cavity of sGLIPR1, His137, Glu120, His79 and Glu98, align well with those in representative CAP structures (Fig. 4). These conserved residues have previously been hypothesized to be part of the active-site residues that could form the catalytic triad of a putative serine protease (Asojo *et al.*, 2005; Gibbs *et al.*, 2008). In some of the reported CAP structures there is a serine in the cavity that is oriented such that it is incapable of forming a typical serine protease catalytic triad with the proximal histidine. GLIPR1 lacks this postulated catalytic serine; in its place is Asn80, which is likewise oriented away from the conserved His79 (Fig. 4). The crystallographic dimers of sGLIPR1 are incapable of forming the catalytic triad of a conventional serine protease as is the case for all other CAP structures currently deposited in the PDB. Unsurprisingly, no proteolytic activity has been detected for sGLIPR1 and it does not have any significant sequence similarity to any known peptidase in the MEROPS database (http://merops.sanger.ac.uk/).

Another role of this conserved cavity was suggested by studies which revealed it to be the major Zn^{2+}-binding site in the Zn^{2+} and heparin-sulfate dependent mechanisms of inflammatory modulation by the cobra CRISP natrin (PDB entry 3mz8; Wang *et al.*, 2010). The two conserved histidines that directly coordinate the Zn^{2+} have similar positions and orientations in the sGLIPR1 structure (Fig. 4). Cobra CRISP natrin has an additional serine in close proximity to the cavity. This serine is not conserved in GLIPR1 or other mammalian CAP proteins. Instead, GLIPR1 has Asn99, which aligns well with the serine (Fig. 4). The structure of Zn^{2+}-bound pseudecin (PDB entry 2epf; Suzuki *et al.*, 2008) reveals that a snake-venom CRISP with an asparagine instead of a serine has the ability to form the same network of bonds (Fig. 4); thus, it appeared quite plausible that Zn^{2+} will bind in a very similar fashion in sGLIPR1. To investigate whether Zn^{2+} binds to sGLIPR1, we soaked preformed crystals of sGLIPR1 with 0.15 m*M* $ZnCl_2$ and determined the structure of the complex. The putative binding cavity of sGLIPR1 forms a similar network of bonds with Zn^{2+} as observed in natrin and pseudecin (Fig. 4). The structures of uncomplexed and Zn^{2+}-complexed sGLIPR1 superpose well, with an r.m.s.d. of 0.15 Å for all main-chain atoms and 0.49 Å for all atoms. Zn^{2+} does not result in changes in the orientation of the residues in the putative CAP binding cavity. Since CAP proteins are often produced under conditions involving host immune responses, it is plausible that the conserved cavities in different CAP proteins have similar roles in Zn^{2+}-dependent modulation of inflammation or other chronic diseases such as cancer. This central cavity is exposed in all reported CAP proteins and is accessible to inflammatory agents and other molecules that may bind to these proteins. Additional studies are required to elucidate which molecules bind and how they facilitate inflammatory modulation.

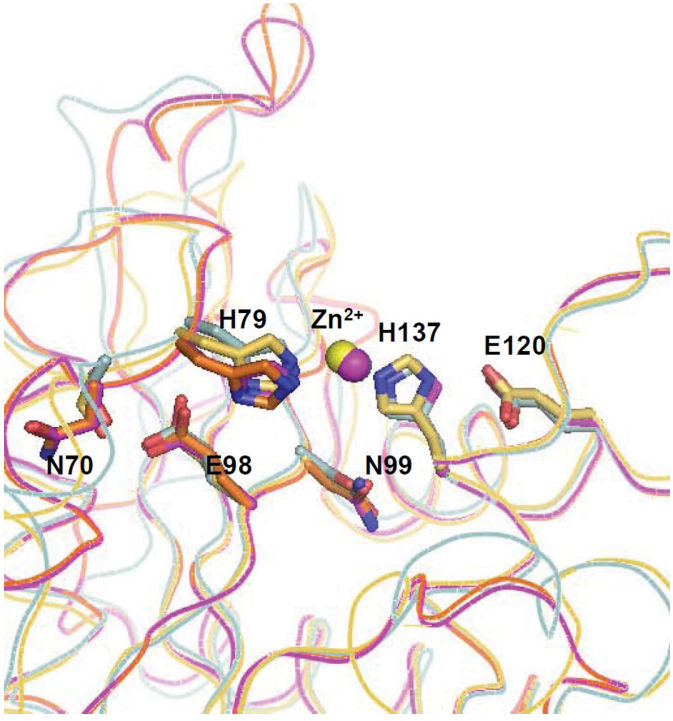

Figure 4
Comparison of CAP-protein central cavities. The cavity of sGLIPR1 complexed with Zn^{2+} (magenta) forms similar interactions with Zn^{2+} as both snake-venom CRISP structures: natrin with Zn^{2+} (yellow) and pseudecin with Zn^{2+} (cyan). The uncomplexed structure of sGLIPR1 (orange) superposes well with that of the Zn^{2+} complex.

3.5. Putative model of functional full-length GLIPR1

Based on our structural analysis, we can propose a model for functional membrane-bound full-length GLIPR1. This model is after the removal of the signal peptide and translocation to the cell membrane. The carboxyl-terminus of GLIPR1 is predicted to have a single transmembrane helix using several membrane-prediction programs within ExPASy including *TopPred*, *TMHMM*, *TMpred* (http://expasy.org/tools/) and *TMRPres2d* (http://biophysics.biol.uoa.gr/TMRPres2D/). The predicted transmembrane helix, 233-YTSLFLIVNSVILILSVIITIL-254, contains the typical hydrophobic residues expected in a membrane-spanning helix. The structure of the intracellular C-terminal domain is unknown and cannot be modeled owing to a lack of homology to any known structure. The 'cytoplasmic tail' of GLIPR1, 255-VQHKYPNLVLLD-265, may serve as a link for signaling between the extracellular and intracellular domains of cells expressing GLIPR1. The flexible hinge motif includes the residues KRYYS, which are disordered in our structure. The hinge motif and preceding linker loop region allow GLIPR1 to be positioned so as to bind various ligands that may be on the cell surface or in the extracellular environment. The flexibility of the hinge and linker regions may allow binding to ligands that also interact with secreted CAP proteins (Fig. 5).

3.6. The GLIPR1 family

Our structure represents the first structure of a member of the GLIPR1 family of mammalian CAP proteins and reveals conserved structural features. Sequence alignment using secondary-structure elements of sGLIPR1 using *ESPript*

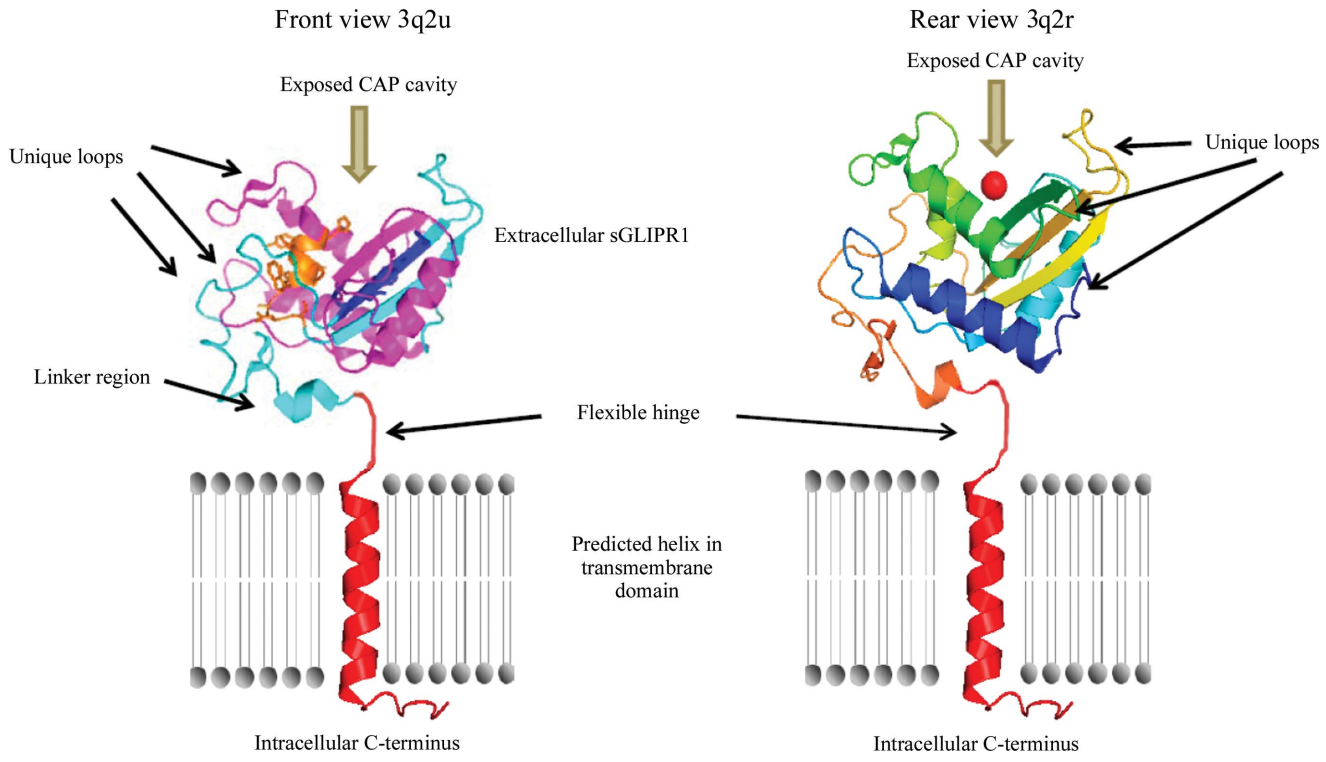

Figure 5
Putative model of functional full-length GLIPR1. This model is after removal of the signal peptide and translocation to the cell membrane. Left panel, front view of model of full-length GLIPR1 with conserved regions from all GLIPR1 protein-family members shown in magenta; the CRISP1 motif is shown in orange. The side chains of the putative CAP cavity residues are shown. The CRISP2 motif that is implicated in sperm–oocyte binding in other CAP proteins is shown in blue. The transmembrane helix (red) links the intracellular and extracellular portions of GLIPR1. Right panel, a rear view of the model rainbow-colored from blue (carboxy-terminus) to red (amino-terminus) reveals an alternative view of full-length GLIPR1 with Zn^{2+} in the CAP cavity.

(Gouet *et al.*, 2003) suggests that sGLIPR1 is a suitable structural model for the CAP domains of other members of the GLIPR1 family. Within the GLIPR1 family, the first 141 amino-acid residues are well conserved, including the loop regions (Fig. 2). The largest primary structural divergence is in the loop regions linking β5 and β6 as well as β6 and β7 (Fig. 2). The overall sequence identity among representative members of this family of CAP proteins varies between 19 and 53% (Table 2*c*). The conserved residues from the GLIPR1 family include approximately 50% of all residues in the extracellular region represented in sGLIPR1. There are structural differences in the N-terminal extensions of GLIPR1L2 (Fig. 2). RTVP-1b, GLIPR1L1 and GLIPR1L2 have two inserted amino-acid residues that GLIPR1 lacks in the sixth β-strand region. Additional N- and C-terminal variations are present in GLIPRL1 and GLIPRL2 isoforms not shown in Fig. 2. The carboxy-terminal linker and transmembrane regions are poorly conserved or absent in some members of the GLIPR1 family. Despite these sequence variations, the conserved residues are present in surface-exposed regions and as such are exposed for possible ligand binding. There is sufficient structural variation among the members of the GLIPR1 family to require further structural investigation of other members of the GLIPR1 family, especially in light of the interest in the roles of these proteins in cancer and reproduction.

4. Conclusions

We have solved the first structures of a member of the GLIPR1 family of mammalian CAP proteins. The structure of sGLIPR1 could only be solved by MR after the removal of flexible regions. This allowed us to solve the phase problem with a search model with only 31% sequence similarity. Compared with other CAP-protein structures, sGLIPR1 has a divergent charge distribution as well as structural variation of the loop regions that are potential starting points for the design of novel anti-GLIPR1 therapeutics. GLIPR1, like other CAP proteins, has a conserved α–β–α sandwich core structure. This core structure has a conserved central cavity *via* which CAP proteins are likely to bind to shared ligands.

This work was supported in part by National Cancer Institute SBIR grant R43CA123991 (NB). OAA was supported by a National Cancer Institute Mentored Career Development Grant for Underrepresented Minorities (K01CA113486).

References

Adams, P. D. *et al.* (2010). *Acta Cryst.* D**66**, 213–221.
Afonine, P. V., Grosse-Kunstleve, R. W. & Adams, P. (2005). *CCP4 Newsl.* **42**, contribution 8.
Asojo, O. A. (2011). *Acta Cryst.* D**67**, 455–462.

Asojo, O. A., Goud, G., Dhar, K., Loukas, A., Zhan, B., Deumic, V., Liu, S., Borgstahl, G. E. & Hotez, P. J. (2005). *J. Mol. Biol.* **346**, 801–814.
Bonafé, N., Zhan, B., Bottazzi, M. E., Perez, O. A., Koski, R. A. & Asojo, O. A. (2010). *Acta Cryst.* F**66**, 1487–1489.
Brünger, A. T. (1992). *Nature (London)*, **355**, 472–475.
Cohen, D. J., Busso, D., Da Ros, V., Ellerman, D. A., Maldera, J. A., Goldweic, N. & Cuasnicu, P. S. (2008). *Int. J. Dev. Biol.* **52**, 737–742.
Cowtan, K. (2006). *Acta Cryst.* D**62**, 1002–1011.
Ding, X., Shields, J., Allen, R. & Hussey, R. S. (2000). *Int. J. Parasitol.* **30**, 77–81.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Fernández, C., Szyperski, T., Bruyère, T., Ramage, P., Mösinger, E. & Wüthrich, K. (1997). *J. Mol. Biol.* **266**, 576–593.
Gao, B., Allen, R., Maier, T., Davis, E. L., Baum, T. J. & Hussey, R. S. (2001). *Int. J. Parasitol.* **31**, 1617–1625.
Gibbs, G. M., Lo, J. C., Nixon, B., Jamsai, D., O'Connor, A. E., Rijal, S., Sanchez-Partida, L. G., Hearn, M. T., Bianco, D. M. & O'Bryan, M. K. (2010). *Endocrinology*, **151**, 2331–2342.
Gibbs, G. M. & O'Bryan, M. K. (2007). *Soc. Reprod. Fertil. Suppl.* **65**, 261–267.
Gibbs, G. M., Roelants, K. & O'Bryan, M. K. (2008). *Endocr. Rev.* **29**, 865–897.
Gibbs, G. M., Scanlon, M. J., Swarbrick, J., Curtis, S., Gallant, E., Dulhunty, A. F. & O'Bryan, M. K. (2006). *J. Biol. Chem.* **281**, 4156–4163.
Gouet, P., Robert, X. & Courcelle, E. (2003). *Nucleic Acids Res.* **31**, 3320–3323.
Guo, M., Teng, M., Niu, L., Liu, Q., Huang, Q. & Hao, Q. (2005). *J. Biol. Chem.* **280**, 12405–12412.
Hawdon, J. M., Narasimhan, S. & Hotez, P. J. (1999). *Mol. Biochem. Parasitol.* **99**, 149–165.
Henriksen, A., King, T. P., Mirza, O., Monsalve, R. I., Meno, K., Ipsen, H., Larsen, J. N., Gajhede, M. & Spangfort, M. D. (2001). *Proteins*, **45**, 438–448.
Hotez, P. J. *et al.* (2003). *Int. J. Parasitol.* **33**, 1245–1258.
McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
Murshudov, G. N., Vagin, A. A., Lebedev, A., Wilson, K. S. & Dodson, E. J. (1999). *Acta Cryst.* D**55**, 247–255.
Padavattan, S., Schmidt, M., Hoffman, D. R. & Marković-Housley, Z. (2008). *J. Mol. Biol.* **383**, 178–185.
Ren, C., Li, L., Goltsov, A. A., Timme, T. L., Tahir, S. A., Wang, J., Garza, L., Chinault, A. C. & Thompson, T. C. (2002). *Mol. Cell. Biol.* **22**, 3345–3357.
Ren, C., Li, L., Yang, G., Timme, T. L., Goltsov, A., Ren, C., Ji, X., Addai, J., Luo, H., Ittmann, M. M. & Thompson, T. C. (2004). *Cancer Res.* **64**, 969–976.
Ren, C., Ren, C. H., Li, L., Goltsov, A. A. & Thompson, T. C. (2006). *Genomics*, **88**, 163–172.
Rosenzweig, T., Ziv-Av, A., Xiang, C., Lu, W., Cazacu, S., Taler, D., Miller, C. G., Reich, R., Shoshan, Y., Anikster, Y., Kazimirsky, G., Sarid, R. & Brodie, C. (2006). *Cancer Res.* **66**, 4139–4148.
Satoh, T., Timme, T. L., Saika, T., Ebara, S., Yang, G., Wang, J., Ren, C., Kusaka, N., Mouraviev, V. & Thompson, T. C. (2003). *Hum. Gene Ther.* **14**, 91–101.
Serrano, R. L., Kuhn, A., Hendricks, A., Helms, J. B., Sinning, I. & Groves, M. R. (2004). *J. Mol. Biol.* **339**, 173–183.
Shikamoto, Y., Suto, K., Yamazaki, Y., Morita, T. & Mizuno, H. (2005). *J. Mol. Biol.* **350**, 735–743.
Suzuki, N., Yamazaki, Y., Brown, R. L., Fujimoto, Z., Morita, T. & Mizuno, H. (2008). *Acta Cryst.* D**64**, 1034–1042.
Szyperski, T., Fernández, C., Mumenthaler, C. & Wüthrich, K. (1998). *Proc. Natl Acad. Sci. USA*, **95**, 2262–2266.
Thompson, T. C. (2010). *Yonsei Med. J.* **51**, 479–483.
Wang, Y.-L., Kuo, J.-H., Lee, S.-C., Liu, J.-S., Hsieh, Y.-C., Shih, Y.-T., Chen, C.-J., Chiu, J.-J. & Wu, W. (2010). *J. Biol. Chem.* **285**, 37872–37883.
Wang, J., Shen, B., Guo, M., Lou, X., Duan, Y., Cheng, X. P., Teng, M., Niu, L., Liu, Q., Huang, Q. & Hao, Q. (2005). *Biochemistry*, **44**, 10145–10152.
Xiang, C., Sarid, R., Cazacu, S., Finniss, S., Lee, H.-K., Ziv-Av, A., Mikkelsen, T. & Brodie, C. (2007). *Biochem. Biophys. Res. Commun.* **362**, 612–618.
Zhan, B., Liu, Y., Badamchian, M., Williamson, A., Feng, J., Loukas, A., Hawdon, J. M. & Hotez, P. J. (2003). *Int. J. Parasitol.* **33**, 897–907.