LIBERTAS ACADEMICA
FREEDOM TO RESEARCH

ORIGINAL RESEARCH

Drug Target Exploitable Structural Features of Adenylyl Cyclase Activity in *Schistosoma mansoni*

Andreas N. Mbah[1,2], Henri L. Kamga[3], Omotayo R. Awofolu[2] and Raphael D. Isokpehi[1]

[1]Center for Bioinformatics & Computational Biology, Department of Biology, Jackson State University, Jackson, MS, USA.
[2]Department of Environmental Sciences, College of Agriculture and Environmental Sciences, University of South Africa, South Africa. [3]Faculty of Health Sciences, University of Buea, Buea, South West Region, Cameroon.
Corresponding author email: raphael.isokpehi@jsums.edu

Abstract: The draft genome sequence of the parasitic flatworm *Schistosoma mansoni* (*S. mansoni*), a cause of schistosomiasis, encodes a predicted guanosine triphosphate (GTP) binding protein tagged Smp_059340.1. Smp_059340.1 is predicted to be a member of the G protein alpha-s subunit responsible for regulating adenylyl cyclase activity in *S. mansoni* and a possible drug target against the parasite. Our structural bioinformatics analyses identified key amino acid residues (Ser53, Thr188, Asp207 and Gly210) in the two molecular switches responsible for cycling the protein between active (GTP bound) and inactive (GDP bound) states. Residue Thr188 is located on Switch I region while Gly210 is located on Switch II region with Switch II longer than Switch I. The Asp207 is located on the G3 box motif and Ser53 is the binding residue for magnesium ion. These findings offer new insights into the dynamic and functional determinants of the Smp_059340.1 protein in regulating the *S. mansoni* life cycle. The binding interfaces and their residues could be used as starting points for selective modulations of interactions within the pathway using small molecules, peptides or mutagenesis.

Keywords: adenylyl cyclase pathway, biomarkers, GTP binding, novel drug targets, protein domain interactions, pathway regulation, *Schistosoma*, switch I, switch II

This article is available from http://www.la-press.com.

Introduction

Schistosoma mansoni (*S. mansoni*) is a dioecious trematode and one of the etiologic agents of schistosomiasis, the second most significant tropical disease after malaria of public health significance.[1] The endemicity of this neglected tropical disease of poverty still remains high with about 200 million people affected worldwide (http://www.who.int/ctd/schisto/epidemio.htm). The urinary manifestation of schistosomiasis can become bladder cancer.[2] Praziquantel (PZQ) is the most effective and acceptable drug against schistosomiasis, but its effectiveness is limited due to its inability to kill parasite stages within 2 to 4 weeks post-infection. Resistance to praziquantel has been observed in some *S. mansoni* isolates from endemic regions in Africa.[3,4] The availability of the draft genome sequence of *S. mansoni*[5] and the wealth of basic molecular biology research presents new opportunities to identify schistosome protein drug targets, and to subsequently determine drug exploitable distinctive protein structural features. The purpose of the reported research was to determine drug exploitable distinctive structural features of the Smp_058340.1 protein, which has been identified as an attractive drug target in the *S. mansoni* genome.[6]

The life cycle of *S. mansoni* starts with the eggs in urine or feces found in water, which then hatch to produce miracidia.[7,8] The miracidia will eventually penetrate the snail tissue and transform into sporocysts. The sporocysts in the snail will develop into cercariae, which are then released into the water. The cercaria penetrates human skin, loses its tail during this process, and transforms into schistosomulum. Schistosomulum enters the blood stream and migrates to the liver, bladder and intestines and matures into an adult worm. The adults mate and produce eggs, which are released from the body through urine or feces to restart the life cycle.[8]

During this complex developmental cycle in the mammalian host, the juvenile parasites pursue a systematic and complicated transition that goes through the heart, lungs and finally reaches the hepatic portal system, which is the typical residence of the adult worm.[8,9] This programmed movement is essential for the continuous survival of the schistosome.[9] Furthermore, the movement is highly coordinated by both the organism's neuromuscular system and the participation of a host-derived signal (serotonin (5-hydroxytryptamine: 5HT) that modulates schistosome motility.[9,10] *Schistosoma* species have been documented to take up 5HT from their immediate host circulatory system through a transporter-mediated mechanism.[11,12] This carrier mechanism is thought to contribute massively to the control of the schistosome motility in the host blood system and is very essential for the continuous survival of the parasite[9,13] and thus its pathogenicity. The effects of serotonin have been extensively studied in some members of the platyhelminthes group such as *Hymenolepis diminuta*[9,14] and *S. mansoni*.[9,15,16]

Serotonin alters the parasite muscular tone,[10] strength[12] and carbohydrate metabolism.[14,17,18] In the intramolluscan larval stages, serotonin elevates the muscular activity of the mother sporocysts of *S. mansoni*,[12] leading to stimulation and release of the daughter sporocysts[19] and eliciting turning behavior in miracidia.[20] The 5HT binds directly to the exposed surface 5HT receptor (Smp_126730) on the parasite muscles and then activates adenylyl cyclase activity, whose main function is the regulation of cyclic adenosine monophosphate (cAMP) of the anaerobic glycolysis pathway. This metabolic pathway is the main source of energy in the parasite.[7,8,14] The binding of guanosine triphosphate (GTP) to the *S. mansoni* 5HT receptor Smp_126730 is necessary for serotonin activation.[21] The adult stage of the parasite has the highest adenylyl cyclase activity, while the cercariae stage has a very low adenylyl cyclase activity,[22] and the schistosomula activation is comparable to that of the adult parasite.[23] The serotonin receptor undergoes a series of activated pathways while carrying out its function. The receptor-activated G proteins are bounded to the inner surface of the cell membrane.

Generally, G proteins are known to serve as mediators between signal receptors and effector enzymes. As such, they are regarded as regulator of metabolic pathways. The G protein complex is made up of the Gα and the Gβγ subunits and subunit has a functional role in controlling the activity of adenylyl cyclase. The Gα is further divided into 4 subclasses: Gαs, Gαi, Gαq/11, and Gα12/13. The Gαs subclass is the G protein component responsible for regulating the adenylyl cyclase activity pathway in *S. mansoni*. The binding of serotonin to the extracellular receptor triggers the release of GDP from the Gαs subunit and

initiates the binding of GTP.[24] This serotonin binding causes the Gβγ complex to dissociate from the Gαs-GTP complex. The activated Gαs can stimulate the adenylyl cyclase activity, while the Gβγ complexes can either stimulate or inhibit it. The Gαs-GTP complex will stimulate the adenylyl cyclase activity through the enzyme binding site of the Gαs protein. The adenylyl cyclase activity will finally lead to the production of cAMP from ATP. The process involves removing the pyrophosphatase group from ATP. This metabolic pathway is the main source of energy in the *S. mansoni* throughout its developmental stages.[7,8,14]

The GTP-binding protein Smp_059340.1 has been identified as an attractive drug target in the *S. mansoni* genome.[6] The availability of diverse structural bioinformatics tools provides opportunities to determine the structure-function relationship of the Smp_059340.1. The G (GTP binding) proteins have fingerprints called consensus sequences (G motifs) that have served as valuable tools in their identification. However, any subtle change in these consensus sequences will be reflected in functional differences among these proteins. It is therefore very important to determine the distinctive structural features of the Smp_059340.1 protein that map to its biological function in *S. mansoni*.

The purpose of the reported research investigation was to determine the distinctive structural features regulating the functioning of Smp_059340.1 protein at the molecular level. We hypothesized that determining the physicochemical features and conserved protein domain organization of the *S. mansoni* Smp_059340.1 protein would help identify putative new functional biomarkers and regulatory points of the parasite adenylyl cyclase activity. The impact of conserved domains and their interaction was considered to be crucial for understanding the biological network of the pathway control by Smp_059340.1.

The structural bioinformatics research on the possible characteristic features, structural model and function of the Smp_059340.1 protein revealed that it is stable and hydrophilic, carrying a net positive charge on its surface which might have antigenic properties. The quality of the modeled Smp_059340.1 structure measured in terms of reliability using the QMEAN score4 range was 0.68, and the Ramachandran plot analysis indicated that 95.72% (320 residues) of the model residues were located in the core or favored

region. As such, the predicted model could be assumed to be of good quality and biologically informative. According to the reliability of the modeled protein structure, the key residues regulating Smp_059340.1 function were predicted to be Ser53, Thr188, Gly210 and Asp207 respectively.

Methods

Sequence retrieval, amino acid and physicochemical parameters analysis

The *S. mansoni* smp_059340.1 (UniProt identification (ID): C4QDC7|C4QDC7_SCHMA; Entrez Gene ID: Smp_059340.1) reviewed sequence was retrieved from Swiss-Prot database (http://expasy.org/sprot/). The amino acid composition of the sequence was computed using the ProtParam tool (http://www.expasy.ch/cgi-bin/protparam). The ProtParam tool was also used to compute the physicochemical parameters such as theoretical isoelectric point (Ip), molecular weight, total number of positive and negative residues, extinction coefficient, half-life, instability index, aliphatic index and grand average hydropathy (GRAVY). The percentages of hydrophobic and hydrophilic residues were calculated from the primary structure analysis and the hydrophobicity plot was created using both Hopp-Woods and Kyte-Doolottle scales for possible antigenicity.

Prediction of secondary structure elements and conserved domains

The SOPMA tool (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sopma.html) was used for the secondary structure prediction. The TMPRED server (http://www.ch.embnet.org/software/TMPRED_form.html) performed the identification of possible transmembrane helix regions. The predicted transmembrane helices region was visualized and analyzed using helical wheel plots generated by the program Pepwheel (http://emboss.bioinformatics.nl/cgi-bin/emboss/pepwheel), included in the EMBOSS 2.7 suite.

The possible conserved domain search was carried out using the public server at http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi for the Smp_059340.1, and the individual domains present were analyzed and binding residues were compared within the domains.

Homology modeling and visualization of 3D structure

The three-dimensional (3D) structure of the Smp_059340.1 encoded protein was modeled using the PDB template 1TL7C (chain C). The 3D structure of the Smp_059340.1 protein was generated using the SwissModel server (http:// swissmodel.expasy. org/) and the Phyre/Phyre2 server (http://www.sbg. bio.ic.ac.uk/phyre2/html/page.cgi?id=news).

The quality of the model was evaluated with Ramachandran plot data, based on the phi–psi torsion angles of all the residues in the model using Ramachandran plot2 assessment server (http://dic-soft1.physics.iisc.ernet.in/rp/) and RAMPAGE server (http://mordred.bioc.cam.ac.uk/~rapper/rampage. php). The Rasmol tool (http://www.openrasmol.org/) was used to visualize the modeled 3D structures and to identify any possible "SS" bonds including the distribution of the secondary structures.

The three-dimensional (3D) LigandSite residue(s) and the predicted ligand(s) of the Smp_059340.1 protein were determined using 3DLigandSite server at (http://www.sbg.bio.ic.ac.uk/3dligandsite). The Rasmol tool was further used in locating the positions of the conserved domains and the keys residues involved in regulatory mechanism on the 3D and folded structure. Particularly the key residues involved in its regulatory mechanism. With the Rasmol tool, the actual locations of all domains were mapped on the protein 3D structure and folded structure.

Results

Amino acid content and physicochemical parameters

The analysis of amino acid content and physicochemical parameters suggests that Smp_059340.1 is moderately hydrophilic, due to the presence of high polar amino acid residues (50.7%) against non-polar (hydrophobic) amino acids residues (34.95%) in the sequence (Table 1). The protein is made up of 379 amino acid residues, with 20 amino acids present in the primary structure, constituting an average molecular weight of 44045.5 Da. The present analysis indicates that there are a large proportion of leucine (Leu), glutamate (Glu), lysine (Lys), arganine (Arg), isoleucine (Ile) and serine (Ser) in that order (Table 1). The atomic composition (6191) consists of 1961

Table 1. Amino acid composition of Smp_059340.1 computed using ProtParam server.

Amino acid	Composition (%)	Hydrophilic (%)	Hydrophobic (%)
Ala	6.1		6.1
Arg	7.4	7.4	
Asn	5.8	5.8	
Asp	5.8	5.8	
Cys	2.9		
Gln	3.2	3.2	
Glu	8.2	8.2	
Gly	4.2		4.2
His	2.1	2.1	
Ile	7.2		7.2
Leu	9.2		9.2
Lys	7.9	7.9	
Met	1.6		
Phe	5.3		
Pro	2.9		2.9
Ser	6.1	6.1	
Thr	4.2	4.2	
Trp	1.1		
Tyr	3.7		
Val	5.3		5.3
Total	100.0	50.7	34.9

Notes: The composition of each amino acid residue is indicated as a percentage. The composition of hydrophilic amino acids is 50.7% while hydrophobic amino acids constitute 34.9%. The protein can be described as moderately hydrophilic.

carbons (C), 3093 hydrogen (H), 547 nitrogen (N), 573 oxygen (O) and 17 sulfur atoms, with a molecular formula of $C_{1961}H_{3093}N_{547}O_{573}S_{17}$ (Table 2). The 17 sulfur atoms were made up of 11 cysteine (Cys) and 6 methionine (Met) residues present in the primary structure.

The computed Isoelectric point (pI) was 8.51 (pI > 7), indicating that this protein is basic in nature. The ProtParam extinction coefficient, at a wavelength of 280 nm measured in water, is favorable because proteins have stronger absorption capabilities at this wavelength than other substances commonly found in the solution. The extinction coefficient for this protein was computed with respect to Cys, tryptophan (Trp) and tyrosine (Tyr) as they were present in the primary structure. The estimated half-life of this protein with Met as the N-terminal of the sequence was 30 (>20) hours. The predicted instability index was 48.15 and 82.17 for aliphatic index, with a very low GRAVY index of −0.458. The Hopp-Woods scale identified three regions on this polypeptide predicted to be highly hydrophilic. The hydrophilic regions are

Table 2. Physicochemical properties of Smp_059340.1 computed using ProtParam server.

ProtParam parameters	Values
No. of amino acids	379
Molecular weight	44045.5Da
Theoretical pI	8.51
Number of negative charge residues	53
Number of positive charge residues	58
Formula	$C_{1961}H_{3093}N_{547}O_{573}S_{17}$
Extinction coefficient	43485 m^{-1} cm^{-1}
Estimated half life	30 hours
Instability index	48.15
Aliphatic index	82.17
Grand average of hydropathicity (GRAVY)	−0.458
Total number of atoms	6191

Notes: The physicochemical parameters define the protein chemical and physical properties in its native state. Smp_059340.1 has a net positive charge and is basic in nature (pI > 7).

shown to have peak values greater than 0 (Fig. 1), which is an indication that Smp_059340.1 could be drug target for schistosomiasis.
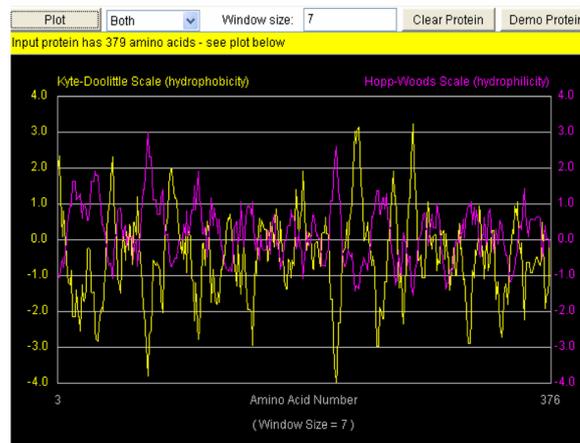
Secondary structure characterization

The SOPMA tool at Expasy (http://expasy.org/tools/) was used to predict secondary structures (Table 3). The data presented here indicates that this protein has mixed secondary structures made up of alpha helix (Hh) (54.62%), extended strand (Ee) (11.35%), beta turn (Tt) (4.75%) and random coil (Cc) (29.29%). The high alpha helix content may be due to the rich alanine and other hydrophobic alpha helix residues, like the phenylalanine (Phe), Leu and Ile content of the protein (Table 1). The TMPRED server at Expasy (http://expasy.org/tools/) predicted possible transmembrane helices, with two possible orientation models (Table 4). The first model, which is strongly preferred based on a total score of 1345, suggests that the N-terminus of the protein is inside and that the region starts from the amino acid residue 218–237, spanning a total length of 20 amino acid residues in an inside → outside (i → o) orientation. The alternative model suggests an outside → inside (o → i) orientation spanning 23 amino acid residues from position 219–241, with a lower score of 854. According to the prediction accuracy of the server, the protein has transmembrane helices in an outside → inside orientation.


```
  1 MVIGCCTLNNISEDAKTRSDAMKQIEKLIEKEKKNFKSTHRLLLLGAGESGKSTIVKQMR
 61 ILHIDGFSEREKKEKIDAIRKNLRDAICSIAGAMGSLKPPVKLELSENRKLRDYILETAS
121 KPDFDYPPEFFTYCAKLWKDGGIQETFERSNEYQLIDCAKYFLDKALEVGAPNYIPSEQD
181 ILRCRVLTSGIFETKFSVDKVNFHMFDVGGQREERRKWIQCFNDVTAIIFVAACSSYNMV
241 LREDPSQNRVKESLELLASIWNNRWLRNISVILFLNKQDLLTEKVLAGKSKIEVYFPHYA
301 TYQAPADTLAEYRHENSEVVRARFFFRDEFLKVTSNNNGGRHYCYPHLTCAVDTENIRRV
361 FNDCRDIIQRMHLRQYELL
```

Figure 1. The hydropathy plot for *Schistosoma mansoni* protein Smp_059340.1.
Notes: The yellow plot is the Kyte-Doolittle hydrophobicity plot. Sections of the plot with high values > 0.0 are highly hydrophobic or membrane spanning segments. The magenta plot is the Hopp-Wood hydrophilicity plot. Higher values above > 0.0 predict rich charge exposed regions with potential antigenic site. Smp_059340.1 gene shows potential antigenic sites with values ≥ 2. Above the plot is the Smp_059340.1 amino acids sequence.

The transmembrane regions are rich in hydrophobic amino acids, while the hydrophilic regions are predicted as potential antigenic segments.

The transmembrane regions are rich in hydrophobic amino acids. This region is well indicated within the Kyte and Dolittle average hydrophobicity plot of the protein residues (Fig. 1), in which all points were shown to have values above the 0 line. On the same plot (Fig. 1) the Hopp-Woods scale predicts the potential antigenic regions

Table 3. Secondary structure composition for *Schistosoma mansoni* protein Smp_059340.1.

Secondary structures	Representation	Composition (%)
Alpha helix	Hh	54.62
Extended stand	Ee	11.35
Beta turn	Tt	4.75
Random coil	Cc	29.29

Notes: Secondary structure composition in percentages for Smp_059340.1 were computed using the SOPMA server at Expasy (http://expasy.org/tools/). The secondary structure elements are composed of four units (alpha helix, extended strand, beta turns and random coil). The structure is made up of more alpha helices and fewer beta turns.

Table 4. Possible transmembrane region and orientation for *Schistosoma mansoni* protein Smp_059340.1.

Parameter	inside → outside (i → o)	outside → inside (o → i)
Length in amino acid position	218–837 (20 residues)	219–241 (23 residues)
Score	1345	854
Orientation preference	Strong (++)	Weak (+)

Notes: Possible transmembrane region and orientation identified by TMPRED server at Expasy (http://expasy.org/tools/). Two transmembrane orientations were predicted. The protein strongly preferred inside → outside (i → o) orientation base on a total score of 1345 with 20 amino acid residues spanning residues 218–237. The alternative model suggest an outside → inside (o → i) orientation spanning 23 amino acid residues from position 219–241 with a lower score of 854.

of the protein. The hydrophilic regions have values greater than 0, and they are predicted as potential antigenic regions. On the Kyte-Dolittle plot, regions with values less than 0 are hydrophilic in nature. Hence the antigenic regions of proteins are hydrophilic. The transmembrane helices predicted using the prediction of transmembrane region and orientation (TMPRED) server were visualized with the Pepwheel tool (Fig. 2) using the 20 amino acids in an inside → outer orientation. The residues in the blue square (I, L, V) and the violet residues (C, A, W, F, Y) represent the hydrophobic amino
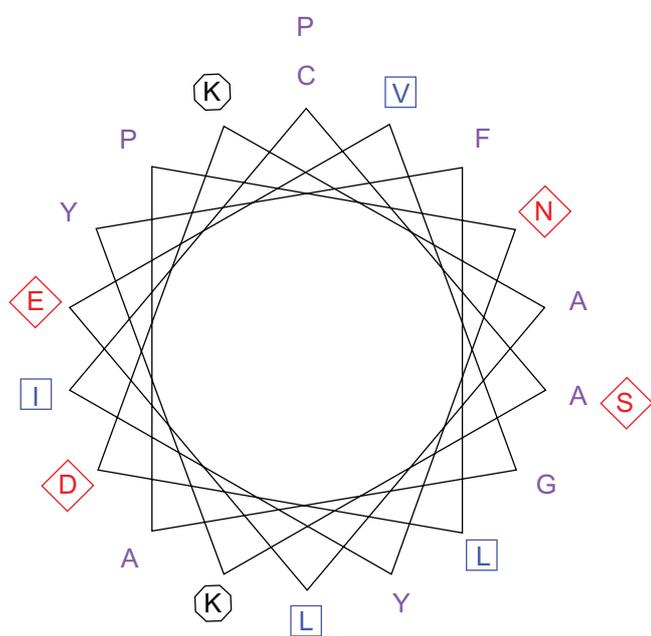

Figure 2. Transmembrane helices plot for *Schistosoma mansoni* Smp_059340.1.
Notes: The transmembrane helices were predicted with TMPRED server at Expasy (http://expasy.org/tools/). The popwheel displays the 20 amino acids transmembrane sequence in a helical representation as if looking down the axis of the helix. This representation is useful for highlighting amphipathicity and other properties of residues around a helix. The aliphatic residues in blue square (I, L, V) and the violet residues (C, A, W, F, Y) are hydrophobic, while hydrophilic residues are marked with diamond (red). The positive charge residues (both K) are with octagonal shape.

acids while those in diamond (red) are the acidic residues (polar residues; N, T, S, D, Q).

Conserved domain and functional analysis of Smp_059340.1

The Conserved Domain search at the National Center for Biotechnology Information (NCBI) website revealed 12 functional units and their full residues (Table 5). The conserved domain search for Smp_059340.1 revealed 12 functional domains scattered within its amino acids sequence length (Fig. 3, Table 5). Close observation of Table 5 shows that these functional units share many residues in common, suggesting that the Smp_059340.1 protein functions as a highly interconnected network of functional units. The next sections provide descriptions of the structural and functional domain interactions of the overall regulatory mechanism of the adenylyl cyclase pathway that controls the general developmental processes of the schistosome.

The Smp_059340.1 modeled 3D structural analysis and verification

The crystal structure of *S. mansoni* C4QDC7_SCHMA (Smp_059340.1) has not yet been determined; here we present the 3D homology model structure of the protein using chain C (1TL7C) as a template with 93% of the residues modeled at a 90% confidence level, as analyzed using the Phyre/Phyre2 server (Fig. 4A). The statistics of Smp_059340.1 (target) and the template showing their comparison was derived from the SwissModel server (Table 6). Residue 39 to 373 of Smp_059340.1 was aligned with the template, covering 88% of its entire sequence length. Irrespective of the length difference between the template and the target, they both share 70% sequence identity. The quality of the model was measured

Table 5. Function of the 12 domains and their constituted residues for *Schistosoma mansoni* protein Smp_059340.1.

| Conserved domain	Component residue(s)	Function
Mg2+/GTP binding site	G46, G51, K52, S53, T188, D207, G210, N276, K277, D279, C350, A351	The Mg2+ interacts with alpha subunits of the G proteins in the present of GTP to form a complex from which nucleotide dissociate slowly
Adenylyl cyclase interaction site	I191, R216, I219, Q220, N223	Stimulates the adenylyl cyclase by binding to it
Beta—gamma complex interaction site	S189, I191, E193, H204, F206, G210, Q211, R212, E214, K217, W218, Q220, C221, F222, N223	The beta/gamma complex can also stimulate and inhibit the adenylyl cyclase activity, but no model of its function had been established
GoLoco binding site	A47, G48, E49, D77, K81, D85, C88, A91, G92, R112, I115, I156, R185, G210, W218, F222	It binds G$\alpha_{i/o}$/GDP complex and prevents the spontaneous release of GDP by Gα, thus acting as a guanine nucleotide dissociation inhibitor (GDI)
Putative receptor binding site	H342, Y343, C344, Y345, P346, H347, L348, T349, C350, A351, V352, D353, E355, N356, I357, R358	Binds the protein to the serotonin receptor (Smp_126730)
Switch I region	R183, C184, R185, V186, L187, T188, S189, G190, I191	Molecular switches (called the effector loop)
Switch II region	V208, G209, G210, Q211, R212, E213, E214, R215, R216, K217, W218, I219, Q220, C221, F222, N223, D224	Molecular switches
G1 box motif	G46, A47, G48, E49, S50, G51, K52, S53	GTP binding signature
G2 box motif	T188	Involved in Mg2+ coordination
G3 box motif	D207, V208, G209, G210	Links the subsites for binding of Mg2+ and the γ phosphate of GTP (GTP γ- phosphate- binding site)
G4 box motif	N276, K277, Q278, D279	Recognizes the guanine ring
G5 box motif	C350, A351, V352	Buttresses the guanine base recognition site

Notes: In the *Schistosoma mansoni* protein Smp_059340.1, each of the residues contributing to the function of each domain in the complex network is presented. Most domains share residues with each other making the network very complex at atomistic level.

using the QMEAN score4. The QMEAN score4 measures the global score of the whole model, reflecting the predicted model reliability range from 0 to 1. The modeled Smp_059340.1 structure had QMEAN score4 value of 0.68, suggesting that it is a good, biologically informative model.

The Swiss model server estimate per residue inaccuracy was visualized based on secondary structural components. The color gradient ranges from blue (more reliable region) to red (potential unreliable regions; Fig. 4B). A greater proportion of unreliable regions were located around the loops and random coils, which are usually the most difficult parts to predict. The sequence alignment of the template and Smp_059340.1 (target) generated by HMM-HMM matching using the Phyre/Phyre2 server shows that the alpha helices and beta strands were predicted with approximate better confidence values than the ran-
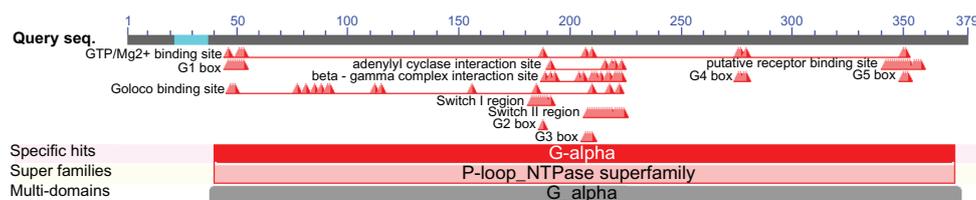
Figure 3. Visualization of the 12 functional domains of *Schistosoma mansoni* Smp_059340.1 protein.
Notes: The Conserved Domain Search Tool (CD-Search http://www.ncbi.nlm.nih.gov/cdd/) was used to determine functional domains in Smp_059340.1. The position and span of each domain unit across the protein are shown. The positioning indicates interdomain connections in a probable complex network. The 12 domain units regulate the function of the Smp_059340.1 protein.
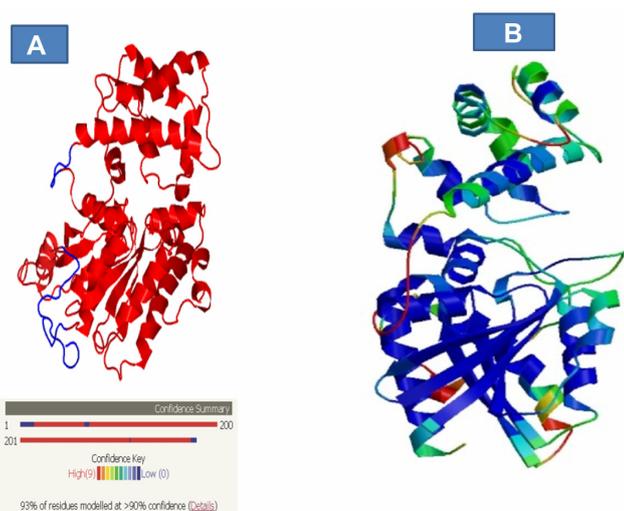
Figure 4. Visualization of the confidence level of the modeled residues and the error regions in the 3D modeled structure of Smp_059340.1. (**A**) The 3D modeled structure of Smp_059340.1 using 1TL7C (chain C) as template with 93% of the residues modeled at >90% confidence level analyzed with Phyre/Phyre2 server and confidence level range from red (high) to low (blue). (**B**) Residues with high score (good structure, highly conserved, deeply buried) are colored blue (low B-factor) while residue with low scores (poor structure, variable, exposed) are colored red (high B-factor).

Notes: The intermediate values may range from green to yellow to orange. This indicates that the model has problems with the loops and alpha helix but not affecting the beta sheet structures.

dom coil (Fig. 5A). Further comparison between the template and Smp_059340.1 (target) using structural superposition helped to establish the regions of the model, which agree and disagree (Fig. 5B). As with every model, the agree regions are trustworthy while disagree regions are generally treated with caution when making biological inferences.[25]

**Table 6.** Modeling Statistics of Smp_059340.1 derived from SwissModel server.

Model information	Value
Modeled residue aligned range	39 to 373
Target sequence length	378
Template sequence length	402
Template	1TL7C (2.80 Å)
Sequence identity	70.00%
Target aligned coverage	88.00%
E value	1.15e-133
Model quality	QMEAN score4: 0.68
Ligand (s) in template	GSP:1, MG:1

Notes: Residues 39 to 373 of Smp_059340.1 were aligned with the template, covering 88.00% of its entire sequence length with 70.00% sequence identity. A QMEAN score4 value of 0.68 suggests the model could be of good quality and can be biologically informative.
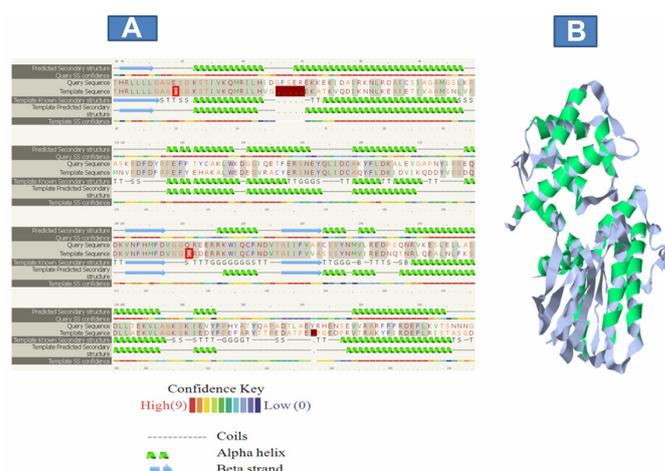
Figure 5. Sequence alignment and structural superposition of Smp_059340.1 (target) and the template. (**A**) Alignment generated by Phyre2 server based on HMM-HMM matching shows that the alpha helices and beta strands were predicted with approximate better confidence values than the random coil. Identical residues in the alignment are highlighted with a grey background. Green helices represent α-helices, blue arrows indicate β-strands and faint lines indicate coil. The 'SS confidence' line indicates the confidence in the prediction, with red being high confidence and blue low confidence. The orange, yellow and green indicate a weak prediction. (**B**) Comparison between the template and Smp_059340.1 (target; green) using structural superposition to establish the regions of the models, which agree and disagree. The agree regions are trustworthy while disagree regions are generally treated with caution when making biological inference.

We further evaluated the quality of the model using the Ramachandran plot on the web (2.0)[26] (Fig. 6A) and RAMPAGE server[27] (Fig. 6B). The Ramachandran plot had been widely used for assessing the quality of predicted models based on the phi-psi torsion angles of all the residues.[25,26,28–30] The fully-allowed and additionally-allowed regions from the Ramachandran plot on the web (2.0), together, are equivalent to the favored (core) region of the RAMPAGE server. In addition, the generously allowed region from the Ramachandran plot on the web (2.0) is equivalent to the allowed region from the RAMPAGE server (Fig. 6A and B). The Ramachandran plot analysis indicates that 95.72% of the model residues (76.56% fully allowed + 19.16% additionally allowed) are located in the core or favored region, 2.99% in the allowed region and 1.2% in the outside region, also known as the outlier or high energy region (Fig. 6C). Residues in the high energy region are considered unstable and can compromise the quality of a modeled structure.[31] A good quality model is expected to have over 90% of the residues in the core region.[26,29,30] This indicates that the Smp_059340.1 predicted model with 320 residues in the core region could be of good quality and biologically informative.
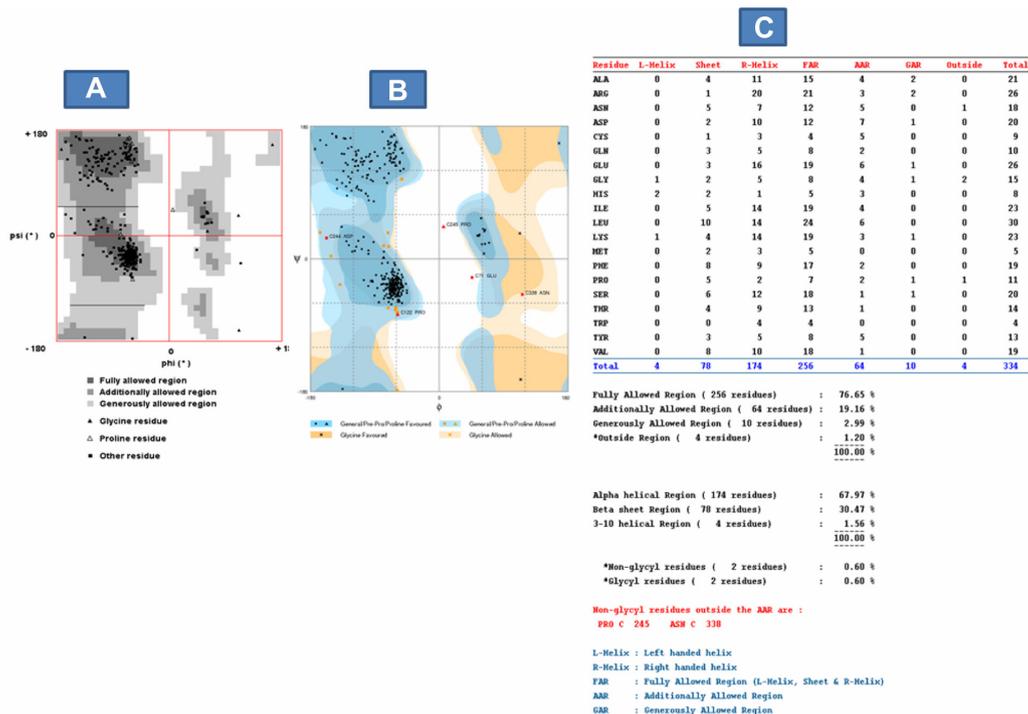
Figure 6. Model quality evaluation using Ramachandran plot. (**A**) Ramachandran plot based on Ramachandran plot on the web (2.0) server: More of the model residues are in the fully allowed and additional allowed regions (total 320 residues (95.72%)) with very few residues in the outside region (4 residues (1.2%)). The 95.72% indicates that the model is stable and expected to be of good quality based on the phi–psi torsion angles of all the residues in the model. (**B**) Ramachandran plot based on RAMPAGE server: This confirms and validates the Ramachandran plot on the web (2.0) server. It shows more model residues are in the favored region (total 316 residues (94.6%)) with very few residues in the outlier region (5 residues (1.5%)). It also shows that PRO 245 and ASN 338 residues are the non-glycyl residues in the additional allowed region (AAR). Image (**C**) Complete Ramachandran plot analysis: It shows the distribution of the whole model residues into various energy regions and on the α-helix, β-strand or coil based on the phi–psi torsion angles of all the residues in the model.

The Rasmol tool was used to visualize the modeled tertiary structure of the Smp_059340.1 protein. The secondary structure components are composed of 20 helices, 8 sheets, 35 turns and 3158 hydrogen bonds. The helices are represented in red, sheets in yellow and turns in blue. The white or grayish colored regions near the turns are other secondary structures (Fig. 7).
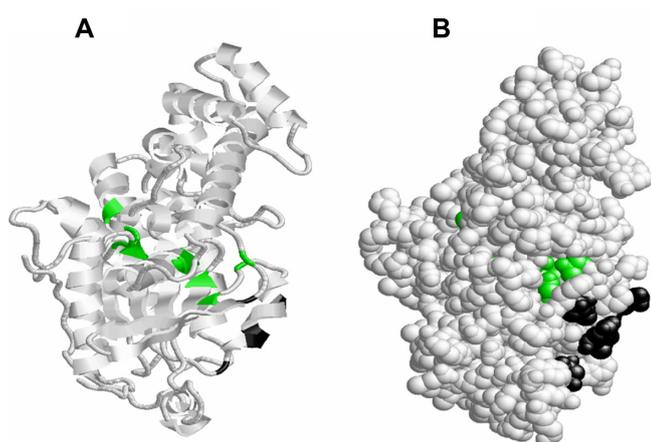
GTP/Mg2+ ion binding site domain

The distribution of the GTP/Mg2+ ion complex binding site residues (green) indicates that they are located on the loops, helices and the β sheets (Fig. 8A). On the protein folded structure these residues are situated in the active site (fold) of the protein that extends to the outer side, through a hollow cavity cutting across the protein internal structure (Fig. 8B). The GTP/Mg2+ ion binding site shares four common residues with the G1 box motif at Gly46, Gly51, Lys52 and Ser53 (Table 5). The GTP/Mg2+ ion binding site converges at two plastic regions of switch I (G2 box motif-threonine (Thr)188) and switch II (Gly210; Table 5).

The GTP/Mg2+ ion binding site also shares some conserved residues with the G3 box motif at aspartate (Asp)207, and with the G4 box motif at residue asparagine (Asn)276, Lys277 and Asp279 (Table 5).
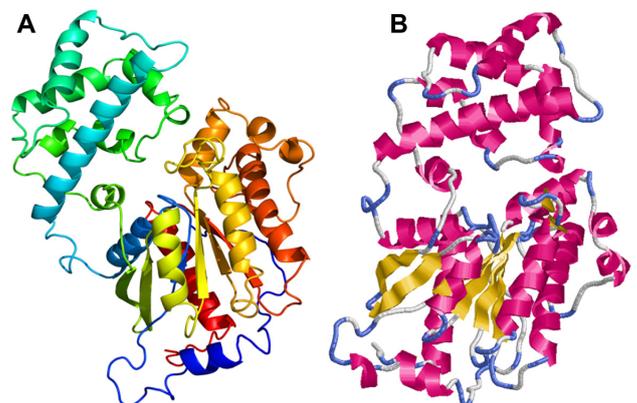
Figure 7. Homology model of *Schistosoma mansoni* protein Smp_059340.1. (**A**) The model colored in rainbow N → C terminus. (**B**) Rasmol tool visualization of model.
Notes: The helices are represented in red; beta sheets are yellow and turns are depicted in blue. The white or grayish colored regions near the turns are other secondary structures. The secondary structure components are composed of 20 helices, 8 sheets, 35 turns and 3158 hydrogen bonds.

Figure 8. The distribution of GTP/Mg2+ ion complex binding site an adenylyl cyclase interaction site residues on the 3-dimensional structure of *Schistosoma mansoni* Smp_059340.1. (**A**) The GTP/Mg2+ ion complex binding site residues (Gly46, Gly51, Lys52, Ser53, Thr188, Asp207, Gly210, Asn276, Lys277, Asp279, Cys350, Ala351; green) are located on the loops, helices and the β sheets. The adenylyl cyclase interaction site residues (Ile191, Arg216, Ile219, Gln220, and Asn223; black) are located on the loops and helices. (**B**) On the folded structure the GTP/Mg2+ ion complex binding site residues are situated in the active site (fold) of the protein that extends to the outer side through a hollow cavity cutting across the protein internal structure. The adenylyl cyclase interaction site residues are localized on the surface of the protein.

Adenylyl cyclase interaction site domain

The adenylyl cyclase interaction site consists of 5 residues made up of Ile191, Arg216, Ile219, glutamine (Gln)220, and Asn223 (black; Fig. 8A), which binds and stimulates adenylyl cyclase activities. The distribution of the adenylyl cyclase interaction site residues (black) indicates that they are located on the loops and helices only (Fig. 8A). On the protein folded structure these residues are localized on the surface of the protein (Fig. 8A). This approximate surface location might be a natural structural orientation to provide efficiency in recognition and binding of the adenylyl cyclase enzyme, leading to stimulation of the adenylyl cyclase pathway during the various developmental phases of *S. mansoni*. The adenylyl cyclase interaction site shared 4 out of its 5 residues (Arg216, Ile 219, Gln220 and Asn223) with the switch II region and three residues (Ile191, Gln220 and Asn223) with the beta-gamma complex binding site (Table 5). This interaction indicates that the adenylyl cyclase interaction site is totally coupled to the beta-gamma complex, and that it is almost completely embedded in the switch II region. As such, the adenylyl cyclase binding site might be contributing to the conformation changes that occur during functioning of the switch II region.

Beta-gamma complex interaction site and GoLoco binding site domains

The beta-gamma residues (green) are located on the helix, loop and β sheets, and they share common residues (Gly210, Trp218 and Phe222; blue) with the GoLoco binding site on the loop and helix (Fig. 9A). Within the protein folded structure, the beta-gamma complex site (green) is positioned on the surface (Fig. 9B), in close proximity to the location of the GTP/Mg2+ ion complex binding site through residue Gly210. At the adenylyl cyclase interaction site the beta-gamma complex site interconnect at residue Ile191, Gln220 and Asn223, while at the GoLoco binding site, the beta-gamma complex site shares residue Gly210, Trp218 and Phe222 shown in blue (Fig. 9B).

The GoLoco binding site (black) has residues located on the helices and the loops (Fig. 9A). It is situated in close proximity to the location of switch I, switch II and the beta-gamma binding site (Fig. 9B), where it can influence their activities independently due their common shared residues as explained above. The GoLoco
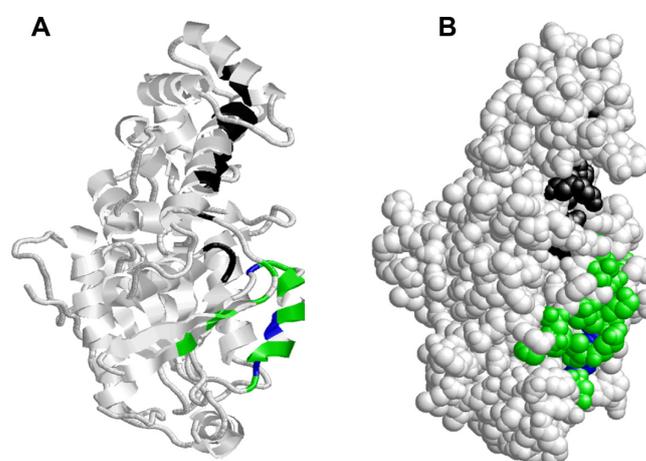
Figure 9. The distribution of beta-gamma an GoLoco binding site residues on the 3-dimensional structure of *Schistosoma mansoni* Smp_059340.1. (**A**) Beta-gamma residues (Ser189, Ile191, Glu193, His204, Phe206, Gly210, Gln211, Arg212, Glu214, Lys217, Trp218, Gln220, Cys221, Phe222, Asn223; green) are located on the helix, loop and β sheets. The GoLoco binding site residues (Ala47, Gly48, Glu49, Asp77, Lys81, Asp85, Cys88, Ala91, Gly92, Arg112, Ile115, Ile156, Arg185, Gly210, Trp218, Phe222; black) are located on the helices and the loops. The common residues shared between both domains (Gly210, Trp218 and Phe222) are depicted blue. (**B**) On the folded structure the beta-gamma complex site (green) is positioned on the surface in close proximity to the location of the GTP/Mg2+ ion complex binding site at residue Gly210 and adenylyl cyclase interaction site (Ile191,Gln220 and Asn223; Fig. 5). The GoLoco binding site is situated exposed on the surface with extension into the GTP/Mg2+ ion complex site at Gly210 (Fig. 8).
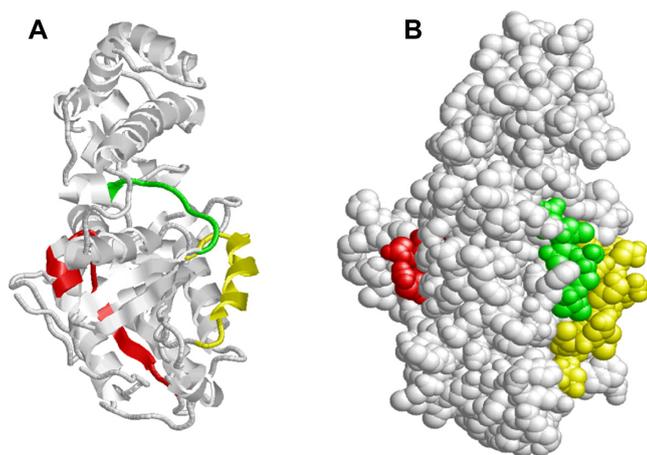
Figure 10. Distribution of putative receptor, switch I and switch II residues in *Schistosoma mansoni* Smp_059340.1. (**A**) The putative receptor conserved residues (red) are positioned on the loop and helix; switch I residues (green) are located on the loop and helix while switch II residues (yellow) are also situated on the loop and helix. (**B**) On the folded structure the putative receptor conserved residues (ted) stretches from within the inner active cavity were it shares two residues with the GTP/Mg2+ ion complex binding site at Cys350 and Ala351 to the surface of the protein. Both switch I (green) and switch II (yellow) regions are orientated on the surface close to each other. Switch I is shorter than the switch II region by 8 residues. These regions shares conserved residues with GTP/Mg2+ ion complex, adenylyl cyclase interaction site and beta–gamma complex, GoLoco binding site as shown above.

binding site also shares conserved residues with the GTP/Mg2+ ion complex site and the G3 box motif.

Putative receptor binding site domain

The putative receptor conserved residues (red) are positioned on a loop and helix (Fig. 10A). On the protein-folded structure, the receptor binding site (red) stretches from within the inner active cavity where it shares two residues with the GTP/Mg2+ ion complex binding site at Cys350 and Ala351 to the surface of the protein (Fig. 10B). It also shares all 3 conserved residues (Cys350, Ala351 and valine (Val)352) found in G5 box motif. Hence the G5 box motif is part of the putative receptor binding site, and functions as a buttress factor for the guanine base recognition site.

Switch I and switch II region domains

The switch I residues (green) are located on the loop and helix, while switch II residues (yellow) are also situated on the loop and helix. Both molecular switches cycle between an active state (GTP bound) and an inactive state (GDP bound; Fig. 8A). Our analysis shows that both switches are located on the outer surface of the protein, with switch I shorter than

the switch II region by 8 residues (Table 5, Fig. 10B). These regions share conserved residues with the GTP/Mg2+ ion complex, adenylyl cyclase interaction site, beta-gamma complex, GoLoco binding site and the G2 box motif (Table 5) as explained above. On the folded protein structure the switch I region (green) is orientated on the surface, very close to the switch II region (yellow; Fig. 10B). The switch II is another region responsible for the conformation changes between the GTP and the GDP as stated above. This switch II region shares conserved residues with GTP/Mg2+ ion complex, adenylyl cyclase interaction site, beta-gamma complex, GoLoco binding site and the G3 box motif (Table 5). We deduced that both the adenylyl cyclase binding site (4 out of 5 residues) (Arg216, Ile219, Gln220 and Asn223) and the G3 box motif (3 out of 4 residues) (Val208, Gly209, Gly210) are completely integrated into the switch II region in terms of their shared residues.

G1-G5 box motif domains

The G protein superfamily folds in different variations depending on its nucleotide binding folds.[32] There are five polypeptide loops that form the guanine nucleotide-
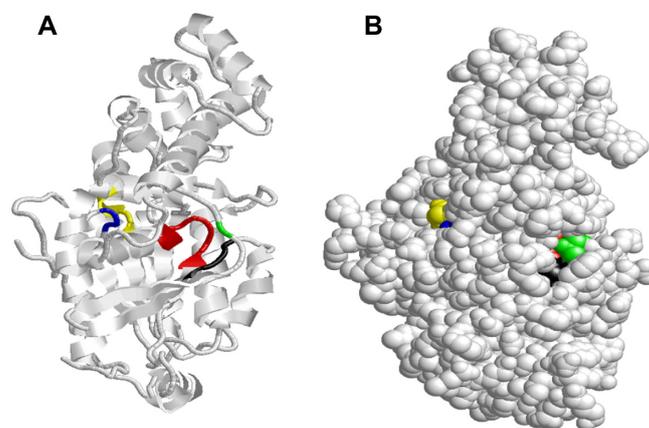

Figure 11. Distribution of G box motifs labeled G1-G5 residues of *Schistosoma mansoni* Smp_059340.1. (**A**) The G1 box motif (red) is located on the loop, sheet and the helix while the G2 box motif (green) is located on the loop and contains only one residue Thr188. The G3 box motif (black) is situated on the loop and sheet; G4 box motif (yellow) on the loop, helix and sheet and G5 box motif (blue) is located on the loop. (**B**) On the folded structure all five G motifs are located in the active site were of GTP binding. The G1 box motif share residue with GTP/Mg2+ ion binding site, G2 box motif (green) share with GTP/Mg2+ ion complex and the switch 1, G3 box motif (black) shares with Mg2+ ion/GTP complex, beta-gamma complex, GoLoco binding site and the switch 2 region. The G5 box motif shares residue (Ala351) with the GTP/Mg2+ ion binding site and all its 3 residues with the putative receptor binding site. The G2 box motif (green) is embedded in the active cavity of the protein very close to G3 (Black) with G4 box motif (yellow) situated very close to G5 box motif (blue).

binding site, and all have well-conserved elements in their domain. They are used for defining the G protein superfamily. These loops are labeled G1 box-G5 box.[33] All five loops were present in this protein structural view (Fig. 11A). The G1 box motif (GxxxxGK(S/K); red) has been mentioned above in connection with the GTP/Mg2+ ion binding site, and it is located on the loop, sheet and helix (Fig. 11A). It is used for defining GTP binding proteins in general.[33] The G2 box motif (green) is located on the loop and contains only one residue, being Thr188. Only the Thr188 is conserved throughout the superfamily, but surrounding residues are conserved within the families. The G2 box motif is part of GTP/Mg2+ ion complex and switch I region.

The G3 box motif (Black) is also known as the DxxG consensus sequence[34–36] and is situated on the loop and sheet, which function to link the sub-sites for binding the Mg2+ ion with the γ phosphate of the of GTP. It also shares some common conserved residues with the GTP/Mg2+ ion complex, beta-gamma complex, GoLoco binding site and the switch II region (Table 1) as mentioned above. The G3 box motif is also known as the Walker B motif[37] and forms an integral part of the switch II region. The G4 box motif (NKXD; yellow) is localized on the loop, helix and sheet. The main function of the G4 box motif is to recognize the guanine ring. The G5 box motif ([C/S]A[K/L/T]; blue) functions as a booster to the guanine base recognition site. It shares one conserved residue (Ala351) with the GTP/Mg2+ ion binding site and all 3 residues with the putative receptor binding site. Hence, it is part of the putative receptor binding site (Table 4). The protein folded structure indicates that all five loops (G1 box-G5 box) are located in the active fold (Fig. 11B). The G2 box motif (green) is embedded in the active cavity of the protein, very close to G3 box motif (black) with G4 box motif (yellow) situated very close to G5 box motif (blue).

Ligand binding site

The predicted ligand binding analysis shows that at the GTP/Mg2+ ion complex site, the Mg2+ ion binds to Ser53 and the metallic ligand is well embedded in the active cavity (Fig. 12). This ligand binding indicates that irrespective of the residues coordinating the GTP binding at GTP/Mg2+ ion complex site, Ser53 could be considered crucial in regulating the adenylyl cyclase pathway in *S. mansoni* developmental cycles.
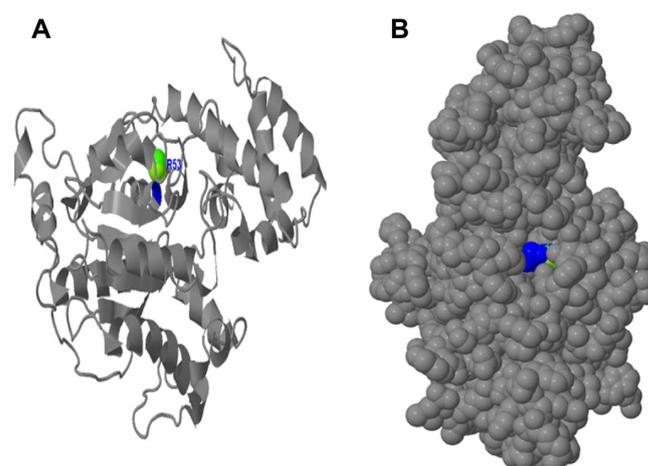
Figure 12. Binding of Mg2+ ion on *Schistosoma mansoni* Smp_059340.1 GTP/Mg2+ ion complex site. (**A**) The binding of Mg2+ ions on the secondary structure with only Ser53 (blue) located on the helix is involved in binding. (**B**) The Mg2+ ion is embedded deeper in the cavity binding with Ser53 and regulated by the G box motifs (Fig. 11) and other domains sharing residues with GTP/Mg2+ ion complex site.

Discussion
Physicochemical properties of the *S. mansoni* Smp_059340.1 protein

The Smp_059340.1 protein regulating adenylyl cyclase pathway in *Schistosoma mansoni* development can be described as moderately hydrophilic and basic in nature (Table 1). The hydrophobic residues are usually found in the core of most proteins and they help stabilize the proteins during their numerous *van der Waal* interactions.[38] The hydrophilic residues are located mostly at the surface active sites of the proteins, where they interact with other polar residues in the protein or with water molecules. The pI indicates the pH at which the protein surface is covered with charge,[39] and the net charge of Smp_059340.1 is positive. The high number of positive charged residues (Arg + Lys = 58) against the total number of negatively charged residues (Asp + Glu = 53) is the main contributing factor to the positive charge. pI proteins are generally stable and compact, thus this parameter will be useful for developing buffer systems for purification of this protein by isoelectric focusing techniques.[39] The high extinction coefficient of 43485 M^{-1} cm^{-1} at 280 nm wavelength computed for Smp_059340.1 was due to the individual contributions of Cys (2.9%), Trp (1.1%) and Tyr (3.7%) concentrations respectively. This observation suggests that the Smp_059340.1

protein can be analyzed using a UV spectrum assay protocol.[40] The computed protein concentration and the extinction coefficient could be important in the quantitative analysis of the protein-protein and protein-ligand interactions of this protein in solutions.[40,41]

The estimated half-life of this protein with Met as the N-terminal of the sequence was 30 (>20) hours. The high concentrations of Ala (6.1%), Leu (9.2%) and Val (5.3%) may be contributing to the stability of this protein.[42,43] The half-life of these 3 residues has been well documented in mammals with values of Ala (4.4 hours), Leu (5.5 hours) and Val (100 hours). In other organisms, these residues also contribute to the protein stability. In yeast, the half life was the same (>20 hours) for both Ala and Val, as well as in *Escherichia coli*, with values > 10.[44]

The instability index prediction using the ProtParam tool indicates that Smp_059340.1 may be unstable with a value of 48.15. This parameter was computed from the impact of dipeptides in the protein sequence.[45] However, Smp_059340.1 is shown to have high aliphatic index and half-life, and enough hydrogen atoms (3093) to form hydrogen bonds. Such hydrogen bonds are known to impact protein stability significantly, making them resistant to degradation.[46] Therefore, it appears that the formation of hydrogen bonds in Smp_059340.1 may override the impact of dipeptides. The calculated value is a measure of protein stability in a test tube.[45] A protein of instability index < 40 is considered as stable, while those with values > 40 are unstable.[45]

The aliphatic index (AI) of a protein is the relative volume occupied by the aliphatic side chains (Ala, Val, Ile, and Leu) and is considered a contributor to the increased thermal stability of globular proteins. The aliphatic index computed for Smp_059340.1 was 82.17, using a formulated rule.[47] This high aliphatic index indicates that Smp_059340.1 can be stable within a wide temperature range. Proteins with low thermal stability turn out to be more structurally flexible.

The Grand Average of Hydropathy (GRAVY) is the computed sum of hydropathy values of all the amino acids, divided by the number of residues in the sequence.[48] The very low GRAVY index (−0.458) of Smp_059340.1 indicates that there is good interaction between this protein and water.

The Hopp-Woods scale identified three regions on this polypeptide predicted to be highly hydrophilic. Hydrophilic regions are exposed on the surface and may possibly represent antigenic sites,[49] indicating that this protein can serve as a possible drug target for schistosomiasis.

Interactions of domain-specific amino acids at the Smp_059340.1 protein active sites

The presence of conserved domain active sites within the guanine nucleotide-binding proteins (G proteins) is responsible for many important biological functions within the cellular process of an organism.[36,50] These functions range from signal transduction, protein synthesis, cell proliferation, protein targeting, membrane trafficking and secretion, including cell skeletal organization and movement.[51–53] The sharing of functional unit residues among domains suggests that the Smp_059340.1 functions as a highly interconnected network of functional domains, as will be explained below.

GTP/Mg2+ ion binding site interaction

The distribution of the GTP/Mg2+ ion complex binding site residues and their orientation to the active fold may be seen as a native adaptation for efficient binding of the GTP molecule. The Mg2+ ion interacts with alpha subunits of the G proteins in the presence of GTP to form a complex, from which nucleotides dissociate slowly at multiple sites. Higher concentrations of the Mg2+ ion have been documented to promote the dissociation of GDP.[54] The dissociation of nucleotides at multiple sites could be accounted by the stretch of active fold observed in Figure 8A. The GTP is bounded as a complex with the Mg2+ ion through the coordination of one oxygen atom from the γ-phosphate, as Gαs proteins are known to be unstable in the absence of a bounded nucleotide.[55] The binding sites for both Mg2+ ion and GTP are strongly linked together.[54] We found that the GTP/Mg2+ ion binding site shares four common residues with the G1 box motif at residues Gly46, Gly51, Lys52 and Ser53 (Table 5). The G1 box motif is one of the five polypeptide loops that form the guanine nucleotide-binding site with well-conserved elements in the domain. It is also called the Walker A motif, and has been used to define the G protein

superfamily and its ability to bind guanine nucleotide such as GTP.[33] The shared residues between the GTP/Mg2+ ion binding site and the G1 box motif could account for the binding of GTP at the GTP/Mg2+ ion binding site domain.

We also found that the GTP/Mg2+ ion binding site converged at two plastic regions of switch I (G2 box motif-Thr188) and switch II (Gly210; Table 5). The coupling between the GTP-γ-phosphate site and the Mg2+ ion may be accounted for by the rigidity of the switch I/switch II interface.[56] The key residue could be Thr188, which binds the magnesium ions of the Mg-GTP complex, allowing it to sense the presence of the γ- phosphate found in GTP molecules. The GTP/Mg2+ ion binding site also shares some conserved residues with the G3 box motif at Asp207 (Table 5). The conserved Asp207 residue in the G3 box motif also coordinates the Mg2+ ion through a water molecule. This coordination could also be crucial for the tight linkage of both the Mg2+ ion and the GTP binding sites. In the GTP/Mg2+ ion complex, the ligand bonds are contributed to by both the β and γ phosphates of the GTP molecule, as well as by Thr188 and the water molecules.

The Asp207 in the G3 box motif binds to one of the water molecules in the process. There is also possible hydrogen bond formation with the α-phosphate oxygen atom from the GTP molecule. The GTP/Mg2+ ion binding site also shares some common residues with the G4 box motif (Asn276, Lys277 and Asp279; Table 5). The G4 box motif functions to recognize the guanine ring, suggesting that the Mg-GTP complex binding site can be regarded as the initiating point for the adenylyl cyclase metabolic pathway. The methylene group of the common Lys277 residue of both the GTP/Mg2+ ion complex binding site and the G4 motif might be providing the hydrophobic surface that lies over the purine ring[57] during the binding process.

Beta-gamma complex and GoLoco binding interaction site

On the protein folded structure, the beta-gamma complex site is positioned on the surface in close proximity to the location of the GTP/Mg2+ ion complex binding site, adenylyl cyclase interaction site and GoLoco binding site. This closeness helps both the beta-gamma complex and the GoLoco site in influencing the activity of the GTP/Mg2+ ion complex

binding site, the adenylyl cyclase interaction site and the switch II region respectively. However, the beta-gamma complex can stimulate and inhibit adenylyl cyclase activity.[58] This inhibition might be through the common shared residues (Ile191, Gln220 and Asn223) with the beta-gamma complex, although no model for this function had been established. The beta-gamma complex also shares Gly210 with the G3 box motif. There is a possible hydrogen bond formation between the γ-phosphate of GTP and the main chain amide of the conserved Gly210 in the G3 box. The stimulation or inhibition of adenylyl cyclase activity may be linked to Gly210 residue, considering its function during the binding of the GTP/Mg2+ ion complex. The beta-gamma complex also shares two residues (Ser189 and Ile189) with switch I, and 10 residues (Gly210, Gln211, Arg212, Glu214, Lys217, Trp218, Gln220, Cys221, Phe222 and Asn223) with switch II (Table 1). This residue sharing suggests that the beta-gamma complex might influence the conformation changes exhibited by both switches, which could affect adenylyl cyclase activity. The GoLoco binding site is known to act as a guanine nucleotide dissociation inhibitor (GDI),[59] and this could account for how it might inhibit adenylyl cyclase activity.

The beta-gamma complex interaction site shares conserved residues with many other domains (Table 5). Its functions had been extensively covered in the sections above regarding how it might influence the activity of the other domains together with the GoLoco motif binding site. The binding of extracellular hormone (5HT) on the S. mansoni serotonin receptor (Smp_126730) initiates the ejection of GDP from the G-alpha subunit and initiates the binding of GTP to the G-alpha subunit.[24] This binding causes the disassociation of G-alpha subunit from the G-alpha/beta-gamma complex. Within the protein folded structure the beta-gamma complex site (green) is positioned on the surface, suggesting that the beta-gamma complex is also in close proximity to the location of switch I and switch II, where it might influence their functions. This closeness might help both the beta-gamma complex and the GoLoco sites to influence the activity of both switches through their shared residues.

The GoLoco binding site (black) has residues located on the helices and the loops (Fig. 9A). It is situated in close proximity to the location of switch I, switch II and the beta-gamma binding site (Fig. 9B),

where it might influence their activities independently due their common shared residues, as explained above. The GoLoco binding site also shares conserved residues with GTP/Mg2+ ion complex site and G3 box motif. The GoLoco binding site is also known as the G protein regulatory (GPR) motif,[60] and it is known to prevent the spontaneous release of GDP by Gαs, thus acting as a guanine nucleotide dissociation inhibitor (GDI).[59] We have discussed in the previous sections how this domain influences the activities of other domains with shared residues, and we shall look at its impact with the rest of the uncovered individual domains.

Putative receptor, switch I and switch II interaction sites

On the protein folded structure, the receptor binding site stretches from within the inner active cavity were it shares two residues with the GTP/Mg2+ ion complex binding site at Cys350 and Ala351 to the surface of the protein (Fig. 10B). It also shares all 3 conserved residues (Cys350, Ala351 and Val352) found in G5 box motif. Hence, the G5 box motif is part of the putative receptor binding site and functions as a buttress factor for the guanine base recognition site. Ala351 could be contributing to this buttress factor function due to its presence at the GTP/Mg2+ ion complex binding site. The structural conformity of the putative receptor might be contributing to the binding stability of the Gαs (Smp_059340.1) protein, the GTP molecule and to the serotonin receptor (Smp_126730). Considering the function, conserved shared residues and location of the putative receptor binding site, this domain can be very crucial to the continuous flow of signal in the adenylyl cyclase pathway.

Switch I and switch II are both molecular switches cycling between an active state (GTP bound) and an inactive state (GDP bound), and switch I is shorter than switch II by 8 residues. These regions share conserved residues with the GTP/Mg2+ ion complex, adenylyl cyclase interaction site, beta-gamma complex, GoLoco binding site and the G2 box motif as explained above. In switch I, the transitional conformation changes mainly involve flips of some peptide units and reorientation of some side chains.[61,62] The key residue could be the Thr188 that coordinates the magnesium ions of the Mg-GTP complex, which can thus sense the presence of the γ phosphate.[63] Therefore,

the conformation transition in the switch I region is manifested in a reorientation of the conserved residue (Thr188) in both the Mg2+ ion ligand and the G2 box motif (Thr188), which is also an integral residue of the switch I region. The switch I region has also been identified as the GAP (GTP activation protein) binding region.[64,65] This region may be responsible for the tight coupling of the GTP/Mg2+ ion complex through their shared Thr188 residue.[56] As explained in the previous paragraphs, the function of the switch I region during conformation changes between GTP and GDP can be compromised by both the beta-gamma complex and the GoLoco binding sites, through their common shared conserved residues (Table 5). On the folded protein structure, the switch I region (green) is oriented on the surface, very close to the switch II region (yellow; Fig. 10B). As mentioned above, the beta-gamma complex or the GoLoco binding sites share common conserved residues with both switch I and switch II regions. Therefore, this close proximity of switch I and switch II region suggests that the conformational changes in both switches might be coupled with, and the functionality of both switches might be influenced simultaneously by, either the beta-gamma complex or the GoLoco binding sites.

The switch II shares conserved residues with the GTP/Mg2+ ion complex, adenylyl cyclase interaction site, beta-gamma complex, GoLoco binding site and the G3 box motif. It was deduced that both the adenylyl cyclase binding site (4 out of 5 residues) and the G3 box motif (3 out of 4) residues are completely integrated into the switch II region in terms of the their shared residues. There is considerable conformational flexibility in this region, especially in the GDP bound state.[61,62] As mentioned above, the adenylyl cyclase binding site may be contributing to these conformation changes. The same plausible reason as explained above for the switch II region could be the case here—the conformational changes between GTP and GDP can be greatly hampered by both the beta-gamma complex and the GoLoco binding sites through their common shared conserved residues. The switch II influences the tight coupling of the GTP/Mg2+ ion complex through the shared Gly210 with DxxD near the N-terminal end of the switch II helix.[66] In this superfamily, the binding energy of GTP is normally used to stabilize switch regions, so that a conformation is produced that favors its association

with an effector. The switch II region is relatively mobile at the effector binding region, particularly in the GTP bound form. Gly220 in switch II and DxxD (Table 5) might be the determining point for the reorientation and partial refolding of the helical region of switch II[61] due to its contribution to forming a hydrogen bond with the gamma-phosphate, as mentioned above.[61]

However, there is a slight difference in structural position between switch I and switch II regions (Fig. 10 A). Both are located on the protein surface, but the switch II (yellow) residues are oriented outwards, while the switch I (green) residues form part of the active site (fold) of the protein, and are localized at the gateway of the active site of the protein. The Thr188 shared between the GTP/Mg2+ ion complex and switch I region may account for differences in the active site location. The GTP/Mg2+ ion binding site residues are located in the active cavity of the protein (Fig. 8B).

G1-G5 box motif interaction sites

The G1 box motif is used to define GTP binding proteins in general.[33] Only the G2 box residue Thr188 is conserved throughout the superfamily, but surrounding residues are conserved within the families. As mentioned above, its main function is to coordinate the magnesium ions of the Mg-GTP complex, which can thus sense the present of the γ-phosphate. The function of the Thr188 had been elaborated extensively in both the GTP/Mg2+ ion complex and the switch I sections above. The conserved Asp207 residue in the G3 box motif coordinates the Mg2+ ion through a water molecule, and the conserved Gly220 in the switch II and in the G3 box motif determines the reorientation and partial refolding of the helical region of switch II[61] by forming a hydrogen bond with the γ-phosphate.[61] The G3 box motif is also known to form an integral part of the switch 2 region. All five loops (G1-G5) are located in the active fold, with the G4 box motif situated very close to G5 box motif. The plausible reason for this may be that the G5 box motif buttresses the guanine base recognition site. The residue Ala351 may be the booster component in guanine site recognition. The location of the five loops in the cavity indicates their importance in the binding of GTP molecules (Fig. 11B).

Conclusion

We have elucidated the distinctive characteristic features of the *S. mansoni* Smp_059340.1 protein, a predicted drug target, from its primary structure to learn how these features contribute to the functioning and regulation of the developmental process in the schistosome's life cycle. However, because these distinctive characteristic features are predictions, they should be considered with caution. The protein is basic and moderately hydrophilic, carrying a net positive charge with possible antigenic properties on its surface. Smp_059340.1 is predicted to be stable within a wide range of temperature, with a high possibility of being analyzed using a UV spectrum assay protocol. Further, Smp_059340.1 is a soluble protein consisting of mixed secondary structure features, with the transmembrane helices having an inside → outside orientation.

The quality of the modeled Smp_059340.1 structure, measured in terms of reliability using the QMEAN score4 range, was 0.68. The Ramachandran plot analysis indicates that 95.72% (320 residues) of the model residues are located in the core or are favored. As such, the predicted model can be assumed to be of good quality and biologically informative. With respect to the reliability of the modeled protein structure, the key features controlling GTP binding and effective pathway regulations were found to be Thr188, Gly210, Asp207 and Ser53. Analysis of the protein domain organization revealed that the pathway regulated by this protein is very complex, and that these domains are interconnected through shared conserved residues regulating the network.

These findings offer new insights into the dynamic and functional determinants of the Smp_059340.1 protein in regulation of the life cycle pathway. These conspicuous structural and functional biomarkers in the Smp_058340.1 encoded protein will add impetus in biological experimental design, particularly site–directed mutagenesis experiments regulating the pathway control by this gene during the developmental stages of the *S. mansoni*.

We envision that by following this type of approach, the distinctive functional features and the derived 3D model can provide a simplistic understanding of important biological issues, such as how protein domains interact at a molecular level in a network.

The modeled binding interfaces and their residues could be used as starting points to guide selective modulations of interactions within the pathway using small molecules, peptides, or mutagenesis.

Funding

Competing Interests

Author(s) disclose no potential conflicts of interest.

Author Contributions

Conceived and designed the experiments: ANM, RDI Analyzed the data: ANM, RDI. Wrote the first draft of the manuscript: ANM. Contributed to the writing of the manuscript: RDI, HLK, ORA. Agree with manuscript results and conclusions: ANM, RDI, HLK, ORA. Jointly developed the structure and arguments for the paper: ANM, RDI, HLK, ORA. Made critical revisions and approved final version: ANM, RDI, HLK, ORA. All authors reviewed and approved of the final manuscript.

Disclosures and Ethics

As a requirement of publication author(s) have provided to the publisher signed confirmation of compliance with legal and ethical obligations including but not limited to the following: authorship and contributorship, conflicts of interest, privacy and confidentiality and (where applicable) protection of human and animal research subjects. The authors have read and confirmed their agreement with the ICMJE authorship and conflict of interest criteria. The authors have also confirmed that this article is unique and not under consideration or published in any other publication, and that they have permission from rights holders to reproduce any copyrighted material. Any disclosures are made in this section. The external blind peer reviewers report no conflicts of interest.

References

1. Borch M, Kiernan M, Rust K, et al. Schistosomiasis: a case study. *Urol Nurs*. 2009;29(1):26–9.
2. Brindley PJ, Mitreva M, Ghedin E, Lustigman S. Helminth genomics: The implications for human health. *PLoS Negl Trop Dis*. 2009;3(10):e538.
3. Melman SD, Steinauer ML, Cunningham C, et al. Reduced susceptibility to praziquantel among naturally occurring Kenyan isolates of *Schistosoma mansoni*. *PLoS Negl Trop Dis*. 2009;3(8):e504.
4. Doenhoff MJ, Kusel JR, Coles GC, Cioli D. Resistance of *Schistosoma mansoni* to praziquantel: is there a problem? *Trans R Soc Trop Med Hyg*. 2002;96(5):465–9.
5. Berriman M, Haas BJ, LoVerde PT, et al. The genome of the blood fluke *Schistosoma mansoni*. *Nature*. 2009;460(7253):352–8.
6. Crowther GJ, Shanmugam D, Carmona SJ, et al. Identification of attractive drug targets in neglected-disease pathogens using an in silico approach. *PLoS Negl Trop Dis*. 2010;4(8):e804.
7. Gobert GN, Moertel L, Brindley PJ, McManus DP. Developmental gene expression profiles of the human pathogen *Schistosoma japonicum*. *BMC Genomics*. 2009;10:128.
8. Jolly ER, Chin CS, Miller S, et al. Gene expression patterns during adaptation of a helminth parasite to different environmental niches. *Genome Biol*. 2007;8(4):R65.
9. Patocka N, Ribeiro P. Characterization of a serotonin transporter in the parasitic flatworm, *Schistosoma mansoni*: cloning, expression and functional analysis. *Mol Biochem Parasitol*. 2007;154(2):125–33.
10. Pax RA, Siefker C, Bennett JL. *Schistosoma mansoni*: differences in acetylcholine, dopamine, and serotonin control of circular and longitudinal parasite muscles. *Exp Parasitol*. 1984;58(3):314–24.
11. Wood PJ, Mansour TE. *Schistosoma mansoni*: serotonin uptake and its drug inhibition. *Exp Parasitol*. 1986;62(1):114–9.
12. Boyle JP, Zaide JV, Yoshino TP. *Schistosoma mansoni*: effects of serotonin and serotonin receptor antagonists on motility and length of primary sporocysts in vitro. *Exp Parasitol*. 2000;94(4):217–26.
13. Ribeiro P, Webb RA. The occurrence, synthesis and metabolism of 5-hydroxytryptamine and 5-hydroxytryptophan in the cestode Hymenolepis diminuta: a high performance liquid chromatographic study. *Comp Biochem Physiol C*. 1984;79(1):159–64.
14. Mettrick DF, Rahman MS, Podesta RB. Effect of 5-hydroxytryptamine (5-HT; serotonin) on in vitro glucose uptake and glycogen reserves in Hymenolepis diminuta. *Mol Biochem Parasitol*. 1981;4(3–4):217–23.
15. Boyle JP, Hillyer JF, Yoshino TP. Pharmacological and autoradiographical characterization of serotonin transporter-like activity in sporocysts of the human blood fluke, *Schistosoma mansoni*. *J Comp Physiol A Neuroethol Sens Neural Behav Physiol*. 2003;189(8):631–41.
16. Gustafsson MK. Immunocytochemical demonstration of neuropeptides and serotonin in the nervous systems of adult *Schistosoma mansoni*. *Parasitol Res*. 1987;74(2):168–74.
17. Rothman RB, Glowa JR. A review of the effects of dopaminergic agents on humans, animals, and drug-seeking behavior, and its implications for medication development. Focus on GBR 12909. *Mol Neurobiol*. 1995;11(1–3):1–19.
18. Rahman MS, Mettrick DF, Podesta RB. *Schistosoma mansoni*: effects of in vitro serotonin (5-HT) on aerobic and anaerobic carbohydrate metabolism. *Exp Parasitol*. 1985;60(1):10–7.

19. Bayne CJ, Grevelding CG. Cloning of *Schistosoma mansoni* sporocysts in vitro and detection of genetic heterogeneity among individuals within clones. *J Parasitol*. 2003;89(5):1056–60.

20. Samuelson JC, Quinn JJ, Caulfield JP. Hatching, chemokinesis, and transformation of miracidia of *Schistosoma mansoni*. *J Parasitol*. 1984;70(3):321–31.

21. McNall SJ, Mansour TE. Novel serotonin receptors in *Fasciola*. Characterization by studies on adenylate cyclase activation and [3H]LSD binding. *Biochem Pharmacol*. 1984;33(17):2789–97.

22. Shiff C, Naples JM, Isharwal S, Bosompem KM, Veltri RW. Non-invasive methods to detect schistosome-based bladder cancer: is the association sufficient for epidemiological use? *Trans R Soc Trop Med Hyg*. 2010;104(1):3–5.

23. Kasschau MR, Mansour TE. Serotonin-activated adenylate cyclase during early development of *Schistosoma mansoni*. *Nature*. 1982;296(5852):66–8.

24. Weitmann S, Würsig N, Navarro JM, Kleuss C. A functional chimera of mammalian guanylyl and adenylyl cyclases. *Biochemistry*. 1999;38(11):3409–13.

25. Kelley LA, Sternberg MJ. Protein structure prediction on the Web: a case study using the Phyre server. *Nature Protoc*. 2009;4(3):363–71.

26. Gopalakrishnan K, Sowmiya G, Sheik SS, Sekar K. Ramachandran plot on the web (2.0). *Protein Peptide Lett*. 2007;14(7):669–71.

27. Lovell SC, Davis IW, Arendall WB 3rd, et al. Structure validation by Calpha geometry: phi,psi and Cbeta deviation. *Proteins*. 2003;50(3):437–50.

28. Hooft RW, Sander C, Vriend G. Objectively judging the quality of a protein structure from a Ramachandran plot. *Comput Appl Biosci*. 1997;13(4):425–30.

29. Porter LL, Rose GD. Redrawing the Ramachandran plot after inclusion of hydrogen-bonding constraints. *Proc Natl Acad Sci U S A*. 2011;108(1):109–13.

30. Zhou AQ, O'Hern CS, Regan L. Revisiting the Ramachandran plot from a new angle. *Protein Sci*. 2011;20(7):1166–71.

31. Maccallum PH, Poet R, Milner-White EJ. Coulombic attractions between partially charged main-chain atoms stabilise the right-handed twist found in most beta-strands. *J Mol Biol*. 1995;248(2):374–84.

32. Brändeén CI. Relation between structure and function of alpha/beta-proteins. *Q Rev Biophys*. 1980;13(3):317–38.

33. Jurnak F, McPherson A, Wang AH, Rich A. Biochemical and structural studies of the tetragonal crystalline modification of the Escherichia coli elongation factor Tu. *J Biol Chem*. 1980;255(14):6751–7.

34. Colicelli J. Signal transduction: RABGEF1 fingers RAS for ubiquitination. *Curr Biol*. 2010;20(15):R630–2.

35. Field CM, Kellogg D. Septins: cytoskeletal polymers or signalling GTPases? *Trends Cell Biol*. 1999;9(10):387–94.

36. Pandit SB, Srinivasan N. Survey for g-proteins in the prokaryotic genomes: prediction of functional roles based on classification. *Proteins*. 2003;52(4):585–97.

37. Walker JE, Runswick MJ, Saraste M. Subunit equivalence in Escherichia coli and bovine heart mitochondrial F1F0 ATPases. *FEBS Lett*. 1982;146(2):393–6.

38. Berezovsky IN, Trifonov EN. Van der Waals locks: loop-n-lock structure of globular proteins. *J Mol Biol*. 2001;307(5):1419–26.

39. Sillero A, Maldonado A. Isoelectric point determination of proteins and other macromolecules: oscillating method. *Comput Biol Med*. 2006;36(2):157–66.

40. Stoscheck CM. Quantitation of protein. *Methods Enzymol*. 1990;182:50–68.

41. Gill SC, von Hippel PH. Calculation of protein extinction coefficients from amino acid sequence data. *Anal Biochem*. 1989;182(2):319–26.

42. Bachmair A, Finley D, Varshavsky A. In vivo half-life of a protein is a function of its amino-terminal residue. *Science*. 1986;234(4773):179–86.

43. Gonda DK, Bachmair A, Wünning I, Tobias JW, Lane WS, Varshavsky A. Universality and structure of the N-end rule. *J Biol Chem*. 1989;264(28):16700–12.

44. Varshavsky A. The N-end rule pathway of protein degradation. *Genes Cells*. 1997;2(1):13–28.

45. Guruprasad K, Reddy BV, Pandit MW. Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. *Protein Eng*. 1990;4(2):155–61.

46. Chen YW, Fersht AR, Henrick K. Contribution of buried hydrogen bonds to protein stability. The crystal structures of two barnase mutants. *J Mol Biol*. 1993;234(4):1158–70.

47. Ikai A. Thermostability and aliphatic index of globular proteins. *J Biochem*. 1980;88(6):1895–8.

48. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. *J Mol Biol*. 1982;157(1):105–32.

49. Hopp TP, Woods KR. Prediction of protein antigenic determinants from amino acid sequences. *Proc Natl Acad Sci U S A*. 1981;78(6):3824–8.

50. Colicelli J. Human RAS superfamily proteins and related GTPases. *Sci STKE*. 2004;2004(250):RE13.

51. Yamanaka G, Eckstein F, Stryer L. Interaction of retinal transducin with guanosine triphosphate analogues: specificity of the gamma-phosphate binding region. *Biochemistry*. 1986;25(20):6149–53.

52. Gilman AG. G proteins: transducers of receptor-generated signals. *Annu Rev Biochem*. 1987;56:615–49.

53. Birnbaumer L. Transduction of receptor signal into modulation of effector activity by G proteins: the first 20 years or so… *FASEB J*. 1990;4(14):3178–88.

54. Higashijima T, Ferguson KM, Sternweis PC, Smigel MD, Gilman AG. Effects of Mg2+ and the beta gamma-subunit complex on the interactions of guanine nucleotides with G proteins. *J Biol Chem*. 1987;262(2):762–6.

55. Ferguson KM, Higashijima T, Smigel MD, Gilman AG. The influence of bound GDP on the kinetics of guanine nucleotide binding to G proteins. *J Biol Chem*. 1986;261(16):7393–9.

56. Hepler JR, Kozasa T, Smrcka AV, et al. Purification from Sf9 cells and characterization of recombinant Gq alpha and G11 alpha. Activation of purified phospholipase C isozymes by G alpha subunits. *J Biol Chem*. 1993;268(19):14367–75.

57. Gille A, Liu HY, Sprang SR, Seifert R. Distinct interactions of GTP, UTP, and CTP with G(s) proteins. *J Biol Chem*. 2002;277(37):34434–42.

58. Weitmann S, Schultz G, Kleuss C. Adenylyl cyclase type II domains involved in Gbetagamma stimulation. *Biochemistry*. 2001;40(36):10853–8.

59. Willard FS, Kimple RJ, Kimple AJ, Johnston CA, Siderovski DP. Fluorescence-based assays for RGS box function. *Methods Enzymol*. 2004;389:56–71.

60. Takesono A, Cismowski MJ, Ribas C, et al. Receptor-independent activators of heterotrimeric G-protein signaling pathways. *J Biol Chem*. 1999;274(47):33202–5.

61. Milburn MV, Tong L, deVos AM, et al. Molecular switch for signal transduction: structural differences between active and inactive forms of protooncogenic ras proteins. *Science*. 1990;247(4945):939–45.

62. Schlichting I, Almo SC, Rapp G, et al. Time-resolved X-ray crystallographic study of the conformational change in Ha-Ras p21 protein on GTP hydrolysis. *Nature*. 1990;345(6273):309–15.

63. Kraulis PJ, Domaille PJ, Campbell-Burk SL, Van Aken T, Laue ED. Solution structure and dynamics of ras p21.GDP determined by heteronuclear three- and four-dimensional NMR spectroscopy. *Biochemistry*. 1994;33(12):3515–31.

64. Calés C, Hancock JF, Marshall CJ, Hall A. The cytoplasmic protein GAP is implicated as the target for regulation by the ras gene product. *Nature*. 1988;332(6164):548–51.

65. Adari H, Lowy DR, Willumsen BM, Der CJ, McCormick F. Guanosine triphosphatase activating protein (GAP) interacts with the p21 ras effector binding domain. *Science*. 1988;240(4851):518–21.

66. Greasley SE, Jhoti H, Teahan C, et al. The structure of rat ADP-ribosylation factor-1 (ARF-1) complexed to GDP determined from two different crystal forms. *Nat Struct Biol*. 1995;2(9):797–806.