


Molecular basis for GTP recognition by light-activated guanylate cyclase RhGC

Agata Butryn^{1,2} , Hadeeqa Raza^{1,2}, Heather Rada³, Isabel Moraes^{2,4,*}, Raymond J. Owens³ and Allen M. Orville^{1,2}

1 Diamond Light Source Limited, Didcot, UK

2 Research Complex at Harwell, Didcot, UK

3 Protein Production UK, Research Complex at Harwell, Didcot, UK

4 Membrane Protein Laboratory, Diamond Light Source Limited, Didcot, UK

Keywords

cGMP; cyclic GMP; guanylate cyclase; guanylyl cyclase; retinylidene photoreceptor

Correspondence

A. Butryn and A. M. Orville, Diamond Light Source Ltd, Harwell Science and Innovation Campus, Didcot OX11 0DE, UK

Tel: +44 (0)1235 778788 (AB); +44 (0)1235 567505 (AMO)

E-mails: agata.butryn@diamond.ac.uk (AB); allen.orville@diamond.ac.uk (AMO)

Present address

*National Physical Laboratory, Teddington, Middlesex, TW11 0LW, UK

(Received 3 September 2019, revised 23 November 2019, accepted 5 December 2019)

doi:10.1111/febs.15167

Cyclic guanosine 3',5'-monophosphate (cGMP) is an intracellular signalling molecule involved in many sensory and developmental processes. Synthesis of cGMP from GTP is catalysed by guanylate cyclase (GC) in a reaction analogous to cAMP formation by adenylate cyclase (AC). Although detailed structural information is available on the catalytic region of nucleotidyl cyclases (NCs) in various states, these atomic models do not provide a sufficient explanation for the substrate selectivity between GC and AC family members. Detailed structural information on the GC domain in its active conformation is largely missing, and no crystal structure of a GTP-bound wild-type GC domain has been published to date. Here, we describe the crystal structure of the catalytic domain of rhodopsin–GC (RhGC) from *Catenaria anguillulae* in complex with GTP at 1.7 Å resolution. Our study reveals the organization of a eukaryotic GC domain in its active conformation. We observe that the binding mode of the substrate GTP is similar to that of AC–ATP interaction, although surprisingly not all of the interactions predicted to be responsible for base recognition are present. The structure provides insights into potential mechanisms of substrate discrimination and activity regulation that may be common to all class III purine NCs.

Database

Structural data are available in Protein Data Bank database under the accession number [6SIR](#).

Enzymes

[EC 4.6.1.2](#).

Introduction

Guanylate cyclases (GCs) convert guanosine 5'-triphosphate (GTP) into one of the most important intracellular messengers, cyclic guanosine 3',5'-monophosphate (cGMP). cGMP is central to many transduction pathways where it propagates signals in processes that include neurotransmission, blood

pressure regulation, bone growth, lipolysis or muscle contraction [1]. All GCs belong to the class III nucleotidyl cyclase (NC) family that also includes many adenylate cyclases (ACs), which perform an analogous function of converting adenosine 5'-triphosphate (ATP) to cyclic adenosine 3',5'-monophosphate

Abbreviations

AC, adenylate cyclase; ATP, adenosine-5'-triphosphate; cAMP, cyclic 3',5'-adenosine monophosphate; cGMP, cyclic 3',5'-guanosine monophosphate; GC, guanylate cyclase; GTP, guanosine-5'-triphosphate; NC, nucleotidyl cyclase; sGC, soluble guanylate cyclase.

(cAMP). Class III ACs and GCs are widely distributed through all kingdoms of life and are closely related in primary sequence [2]. They have evolved as dimeric head-to-tail wreath-like assemblies with the active site (s) formed at the interface between the catalytic subunits/subdomains [3].

Mechanistic insights into the activity of class III NCs have been derived from dozens of crystal structures of AC domains in a ligand-free form or bound to substrate-based inhibitors, substrates or ATP analogues in various conformations [3–11]. Structural analysis suggests that the AC dimer undergoes a functional conformational change that involves ‘closing’ of the active site region that brings the catalytic residues distributed across the subunit interface into the proper register. Such stimulatory dynamic changes can be governed *in trans* by factors directly binding to the catalytic domain such as G proteins [3] and bicarbonate [4,7], or *in cis* in prokaryotes and lower eukaryotes [12–14].

Despite many structural insights into AC action, our understanding of mechanisms of the reaction catalysed by GCs is largely missing. In particular, the current literature does not provide a sufficient explanation for the NC substrate selectivity. Most of the GC structures available to date represent ligand-free, atypical head-to-head or canonical head-to-tail dimers that are inactive due to misaligned active site residues [15–20]. A few reports describe crystal structures of the GC domain in the active conformation; however, they exploit enzymes of altered specificity [10,11]. Important new information has recently emerged from the 3.8 Å resolution cryo-EM study of the human soluble GC (sGC) in its inactive and GMPCPP-bound NO-activated state [21], suggesting that all type III NCs undergo dynamic structural changes as part of the catalytic cycle.

An aquatic fungus, *Blastocladiella emersonii*, uses cGMP-mediated signalling to control zoospore phototaxis [22–25]. *B. emersonii* genome sequencing data revealed the presence of a unique gene fusion comprising a microbial (type I) rhodopsin fused to a GC domain [26]. Additionally, orthologues were identified in the genomes of related species, *Catenaria anguillulae* and *Allomyces macrogynus*, but not in other fungi [26]. The rhodopsin-GC protein fusion (hereafter referred to as RhGC) was directly demonstrated to function as a light-activated GC that has a great potential as a new optogenetic tool linking to reversible cGMP manipulation in the cell [27,28].

Crystal structures of the soluble GC domain of *B. emersonii* RhGC (BeGC) [20] and a calcium- and ATP α S-bound dimeric form of the *C. anguillulae*

RhGC GC domain with several mutations that transform substrate specificity into an AC (CaAC \cdot ATP α S \cdot Ca $^{2+}$) [11] have been recently described. Here, we report the crystal structure of the wild-type GC domain of RhGC from *C. anguillulae* in its GTP- and calcium-bound form (CaGC \cdot GTP \cdot Ca $^{2+}$). This is the first study that reveals the organization of a native GC domain in its substrate-bound state. The structural analysis presented here provides important insights into substrate binding and selectivity of class III NCs.

Results

Structure of the CaGC \cdot GTP \cdot Ca $^{2+}$ complex

Rhodopsin–GC (RhGC) contains an N-terminal autoregulatory element, type I rhodopsin domain, coiled-coil region (signalling helix) and catalytic GC domain on the C terminus (Fig. 1A). We cloned, expressed and purified the C-terminal part of the *C. anguillulae* RhGC (residues 442–626, further denoted as CaGC) in *Escherichia coli*. Although RhGC is anticipated to act as a dimer like all other GCs [29,30], the isolated GC domain is monomeric in solution (Fig. 1B and [20]). In many of our crystallization conditions, we obtained crystals with unit cells comprising monomeric or unconventional head-to-head disulphide bond-stabilized dimeric assemblies that have been reported by others [11,20]; these will not be discussed further.

In both ACs and GCs, magnesium is the physiological active site ion [5,31]. Despite its monomeric nature in solution, our isolated CaGC was active in the presence of magnesium or manganese (Fig. 1C), confirming previous results [11,20]. As observed in other GCs, CaGC was inactive when calcium was used [16]. This led us to postulate that the addition of GTP and calcium might promote the stabilization of the active dimer for use in crystallization trials. Indeed, in the presence of GTP and calcium chloride, we obtained distinct orthorhombic crystals, which did not grow in the absence of these additives. These crystals diffracted to 1.7 Å resolution, and we solved the structure of CaGC \cdot GTP \cdot Ca $^{2+}$ by molecular replacement as described further below. The atomic model was refined to $R_{\text{work}}/R_{\text{free}}$ of 18.2/21.1% with good statistics with two dimers in the asymmetric unit (Table 1).

The crystal structure of the *C. anguillulae* RhGC GC domain engineered with mutations into an AC (CaAC) and bound to ATP α S has been described previously [11]. Primary sequences of CaAC and CaGC differ only in two positions (E153K/C566D), which are critical for substrate selectivity [11]. The atomic

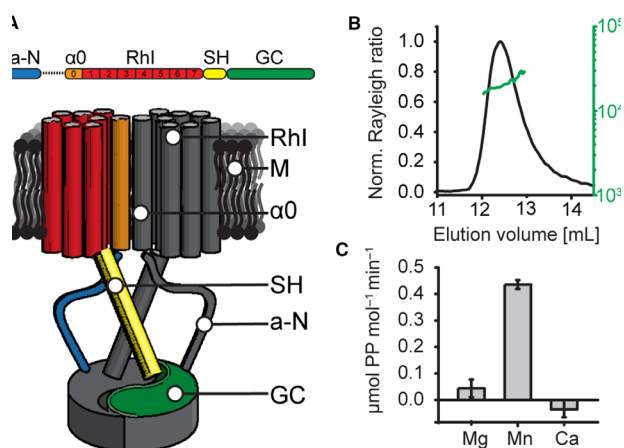


Fig. 1. Characterization of the crystallized construct. (A) Schematic domain organization of the full-length RhGC (top) and its potential dimeric assembly (bottom). M: membrane. From the N terminus: a-N: N-terminal autoinhibitory region. $\alpha 0$: predicted additional transmembrane helix. Rhl: type I rhodopsin domain. SH: signalling helix. GC: GC domain. On the bottom panel, one of the monomers is colour-coded, whereas the second one is shown in grey. (B) SEC-MALS analysis of the CaGC construct. The chromatogram displays normalized Rayleigh ratio (black) and molar mass calculated by MALS (20 kDa, green). (C) Activity of the CaGC construct in solution in the presence of different divalent cations measured as concentration of generated pyrophosphate. Activity was measured in triplicates. Depicted values represent the mean \pm SD.

models of CaGC monomer presented herein and CaAC are almost identical (RMSD of 0.3–0.9 Å across all atoms, Fig. 2A) with the largest differences found only in the ‘tongue’ regions. In CaAC·ATP α S·Ca²⁺, all $\beta 7/\beta 8$ tongues are similar and are located on the ventral side of the dimer, whereas $\beta 4/\beta 5$ tongues (dorsal side) are found in two different conformations in the dimer (distal and proximal, Fig. 2B). On the contrary, $\beta 7/\beta 8$ tongues of CaGC monomers are somewhat less well defined and the electron density of chains A and C suggests the presence of two coexisting conformations from which only the more prominent was modelled. Unlike in CaAC·ATP α S·Ca²⁺, in CaGC·GTP·Ca²⁺ $\beta 4/\beta 5$ tongues are entirely uniform and all adopt the distal conformation. Although in both structures conformations of β tongues are affected by the crystal packing, these differences might suggest that these regions play a functional role (see Discussion).

The CaGC·GTP·Ca²⁺ atomic model presented here represents the ‘active’ state of the wreath-like type III NC assembly. The monomers are related to each other by a translocation and rotation around the centre of the dimer interface in comparison with an inactive ligand-free form of GC (Fig. 2C [21]). Despite some differences, the CaGC·GTP·Ca²⁺ dimer conformation

is almost identical as in CaAC·ATP α S·Ca²⁺ (RMSD between the dimer conformations \sim 0.4 Å) and in the reconstruction of the activated human sGC·GMPCPP·2 \times Mg²⁺ complex (RMSD 1.4 Å, Fig. 2D [21]). The differences account for the slight relative tilt of the subunits around the axis parallel to the dimer interface. The ‘closed’ conformation was likely induced by the presence of two electron-rich atoms (tentatively assigned as calcium) and two GTP molecules that occupy the active sites located at the interface between the monomers (Fig. 2E, discussed further below). GTP and calcium binding induces closure of the ventral side of the dimer and overall decrease in the interface area between the subunits from predicted 1600 to 1300 Å², as calculated by PISA [32].

CaGC·GTP·Ca²⁺ active site

This CaGC·GTP·Ca²⁺ structure is the first to visualize a wild-type GC domain dimer bound to its natural substrate. The phosphate groups of GTP are coordinated by the backbone of Asn460*, Phe461*, Thr462* and side chain of Arg545*, which is believed to aid the exit of the product pyrophosphate (residues from the monomer that binds the triphosphate tail are marked with an asterisk, Figs 3A and 4). The other subunit chain contributes two residues to the active site. Lys612 interacts with the GTP γ phosphate, and side chain of Arg577 is positioned appropriately to help stabilize a transition state in the reaction cycle [5,33]. The density of these two side chains is less well defined in chains A and C that display two conformations of the $\beta 7/\beta 8$ tongue.

Two divalent cations, typically magnesium or manganese, are postulated to facilitate substrate binding (ion B) and catalytic turnover (ion A) [5,30]. With a few exceptions, in ligand- and calcium-bound NC structures, only the high-affinity ion B site is occupied [4,6–8,10,11,21,34,35]. This is most probably due to the significantly larger atomic radius of calcium compared to magnesium or manganese. In our CaGC·GTP·Ca²⁺ model, calcium also only binds in the position analogous to ion B and is coordinated by a pair of aspartic acids Asp457* and Asp501*, α -, β - and γ -phosphate, as well as the backbone carbonyl of Ile458*. Our electron density maps are most consistent with a water molecule occupying the ion A position, which explains our observation that calcium does not support catalytic activity. Despite the absence of low-affinity catalytic ion A, the formation of NC dimers in its fully active conformation is well supported [4,11,36].

In contrast to the polyphosphate tail, in most of the NC domain structures ribose moiety makes no or only

Table 1. Data collection and refinement statistics. Statistics for the highest-resolution shell are shown in parentheses

	CaGC-GTP·Ca ²⁺ , PDB 6SIR
Data	
Space group	<i>P</i> 2 ₁ 2 ₁ 2 ₁
Unit cell <i>a</i> , <i>b</i> , <i>c</i> , α , β , γ (Å, °)	58.65, 95.17, 138.66, 90, 90, 90
Resolution range (Å)	36.30–1.70 (1.76–1.70)
Completeness (%)	98.8 (97.8)
Total reflections	1 121 287 (59 894)
Unique reflections	85 087 (8312)
Multiplicity	13.2 (13.5)
Mean <i>I</i> /sigma(<i>I</i>)	20.3 (2.2)
<i>R</i> -merge	0.073 (1.27)
<i>R</i> -meas	0.076 (1.32)
<i>R</i> -pim	0.021 (0.36)
CC1/2	1 (0.79)
Refinement	
Resolution range (Å)	36.30–1.70 (1.76–1.70)
No. reflections (work)	85 085 (8311)
No. reflections (free)	4316 (415)
<i>R</i> -work	0.182 (0.273)
<i>R</i> -free	0.211 (0.303)
No. of non-H atoms	6546
Protein	5809
Ligand	162
Solvent	575
R.m.s. deviations	
Bonds (Å)	0.007
Angles (°)	1.06
Average <i>B</i> factors (Å ²)	28.9
Protein	27.9
Ligand	34.7
Solvent	37.5
Ramachandran plot	
Favoured regions (%)	98.2
Additionally allowed (%)	1.8
Outliers (%)	0.0

a few direct contacts with the protein. These contacts vary among NC structures and are likely to change in the presence of the transiently bound A site metal during the catalysis. The conserved Asn and Ser (α 4) residues were implicated as a catalytically important whereby they help to orient the sugar in the active site of ACs by lowering the activation energy of the transition state [5,7,33,37]. Neither in our CaGC·GTP·Ca²⁺ model, nor in CaAC·ATP α S·Ca²⁺ or sGC·GMPCPP·2 \times Mg²⁺ structures does the sugar moiety form hydrogen bonds to topologically equivalent Asn (Fig. 3B). However, a hydrogen bond is formed between the Ser residue via ribose O3' in CaGC·GTP·Ca²⁺, via O2'/O3' in sGC·GMPCPP·2 \times Mg²⁺ or via O4' in CaAC·ATP α S·Ca²⁺ (Fig. 3C).

In NCs, the purine ring binds in a hydrophobic pocket within the active site with few polar interactions that serve to discriminate between ATP and GTP. Catalytic

specificity for ATP in ACs is predominantly determined by two residues: Lys and Asp (or Thr), which have been shown to directly interact with N1 and N6 atoms of adenine base, respectively [5]. A few of the AC structures contain unmodified ATP [6,8,10,14]. Most of the ligand-bound ACs were cocrystallized in the presence of a substrate analogue. They provide useful information, but also have significant limitations, for example, often contain substrate modifications that eliminate interactions critical to catalysis [4,5,33]. The *Mycobacterium avium* Ma1120_{CHD} [10] and *Arthrospira platensis* CyaC [4] structures are good examples of wild-type ACs crystallized with the ligand in a conformation that represents a state well aligned for the nucleophilic attack and in which both direct interactions between the specificity-determining residues and substrate are present. In GCs, Lys and Asp present in ACs are replaced by Glu and Cys (or Ser) that are believed to mediate recognition of the exocyclic amine and carbonyl group of the guanine, respectively [38]. The second guanine-specifying residue shows lower conservation indicating that its interaction varies in different members of the GC family [16] (Fig. 5). In BeGC and CaGC, residues Glu497 and Cys566 control substrate specificity [11,39]. As expected, the guanine in the CaGC·GTP·Ca²⁺ complex is positioned as GMPCPP in the active site of sGC and this orientation is equivalent to adenine in most of the AC structures. In particular, Glu497 appears to form a hydrogen bond to N1/N2 of guanine. However, in all monomers Cys566 side chains clearly adopt rotamer conformations that face away from the base and are inconsistent with hydrogen bond formation (Fig. 3B). Instead, the guanine moiety is indirectly stabilized by water molecules that also interact with backbone atoms of Leu567 and Gly500. Although the 3.8 Å resolution of the sGC·GMPCPP·2 \times Mg²⁺ cryo-EM model limits the accuracy of the substrate–enzyme interactions, it does provide insight into GTP recognition [21]. For example, Glu473 is also well positioned to form a hydrogen bond to the guanine N1 or N2 atoms. The Cys541 residue interaction with the base is less obvious because it faces away from the active site; nevertheless, it could form a hydrogen bond with the carbonyl group if it assumed a different rotamer (Fig. 3C). The lack of the second canonical interaction was also reported in the structural analysis of the dual-specificity triple mutant of Ma1120_{CHD}, Ma1120_{CHD}^(KDA→EGY), in which the base is flipped by 180° and the observed mode of interaction is different than that predicted for GCs.

Discussion

Despite progress in our knowledge of the structure and regulation of ACs, the molecular mechanisms that

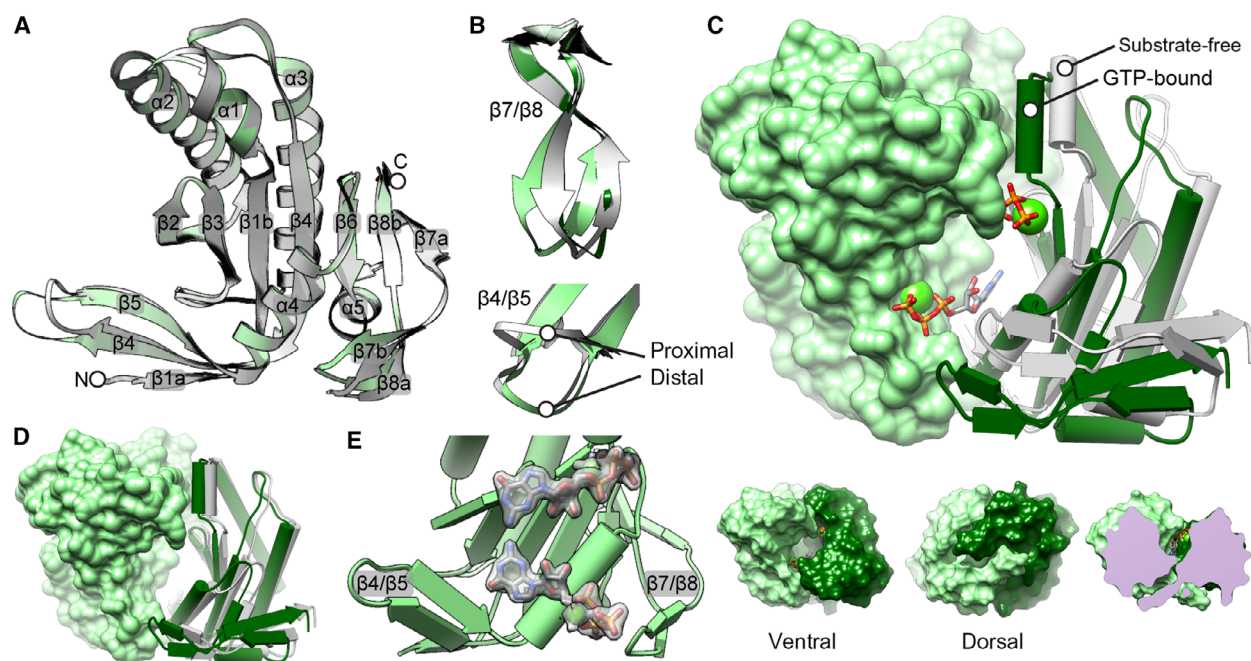


Fig. 2. CaGC-GTP-Ca²⁺ structure overview. (A) Cartoon representation of the CaGC (light green, PDB 6SIR) and CaAC monomer (grey, PDB 5OYH [11]). Secondary structure elements are depicted and labelled β 1– β 8 for the strands and α 1– α 5 for the helices. N and C terminus are labelled N and C, respectively. (B) Comparison of the tongue regions in CaGC and CaAC. Top: conformation of the β 7/ β 8 tongues differs between CaGC monomers (light and dark green) whereas it is same for all CaAC monomers (grey). Grey and dark green conformations are almost perfectly overlapping. Bottom: conformation of the β 4/ β 5 tongues differs between CaAC monomers within each dimer (both grey), whereas it is distal for all CaAC monomers (light green). (C) Top: comparison of the CaGC dimer (light and dark green) and its ligand-free form modelled based on the inactive sGC dimer (PDB 6JT1 [21]). View from the ventral side. Structures were superimposed via the light green monomer shown as single surface. Calcium ions are depicted as green spheres. GTP molecules are shown as sticks. Bottom: CaGC-GTP-Ca²⁺ dimer along its twofold axis viewed from the ventral (left) and dorsal side (middle) represented as solvent accessible surface. Right bottom panel shows cross section of the dimer with the view on one of the substrate binding pockets. (D) Comparison of CaGC-GTP-Ca²⁺ dimer and its model based on the sGC-GMPCPP-2 \times Mg²⁺ assembly (PDB 6JT2 [21]). Structures were superimposed via the light green monomers shown as single surface. Second subunit in CaGC-GTP-Ca²⁺ (dark green cartoon) is in almost identical relative orientation as in sGC-GMPCPP-2 \times Mg²⁺ (grey cartoon). (E) $2F_o - F_c$ electron density map carved within 2 Å distance around GTP and calcium, displayed at 1.5 σ level. For clarity, second monomer was omitted. Molecular graphics and analyses were performed with the UCSF CHIMERA package [52].

determine GC substrate specificity and control its activity are still poorly understood. Recent cryo-EM structures of the human sGC provided important new insights into the mode of enzyme–substrate interaction of a wild-type GC [21]. In our work, we extend the interpretation of the cryo-EM results by providing a high-resolution crystal structure for a GTP complex in a eukaryotic light-activated GC, RhGC. Together, these results support the conclusion that GTP binding and recognition by GCs are very similar to classical AC–ATP interaction(s). This explains why the specificity of type III NCs is achieved not upon substrate binding, but rather during catalysis [8,16,40]. Surprisingly, in our CaGC-GTP-Ca²⁺ structure only one canonical, guanine-specific hydrogen bond is formed via Glu497 located on β 5. Therefore, our data support

the hypothesis that, although in ACs both specificity-determining residues seem to form direct hydrogen bonds with ATP, in GCs Glu–N1/N2 interaction plays the major role in GTP specificity. The second base-specific residue could serve an alternative function; rather than forming a hydrogen bond to guanine, it could provide shape complementarity with the nucleotide base [10,16]. The organization of the active site appears to be sensitive to local rearrangements. Therefore, the Cys566–base interaction(s) could impact regulation of protein activity. This is reinforced by the observation that this residue is not located within the subunit chain of the (pseudo)dimer to which GTP triphosphates are tightly anchored via strong protein and metal ion interactions with highly conserved residues [30]. It is also very likely that perturbations in the

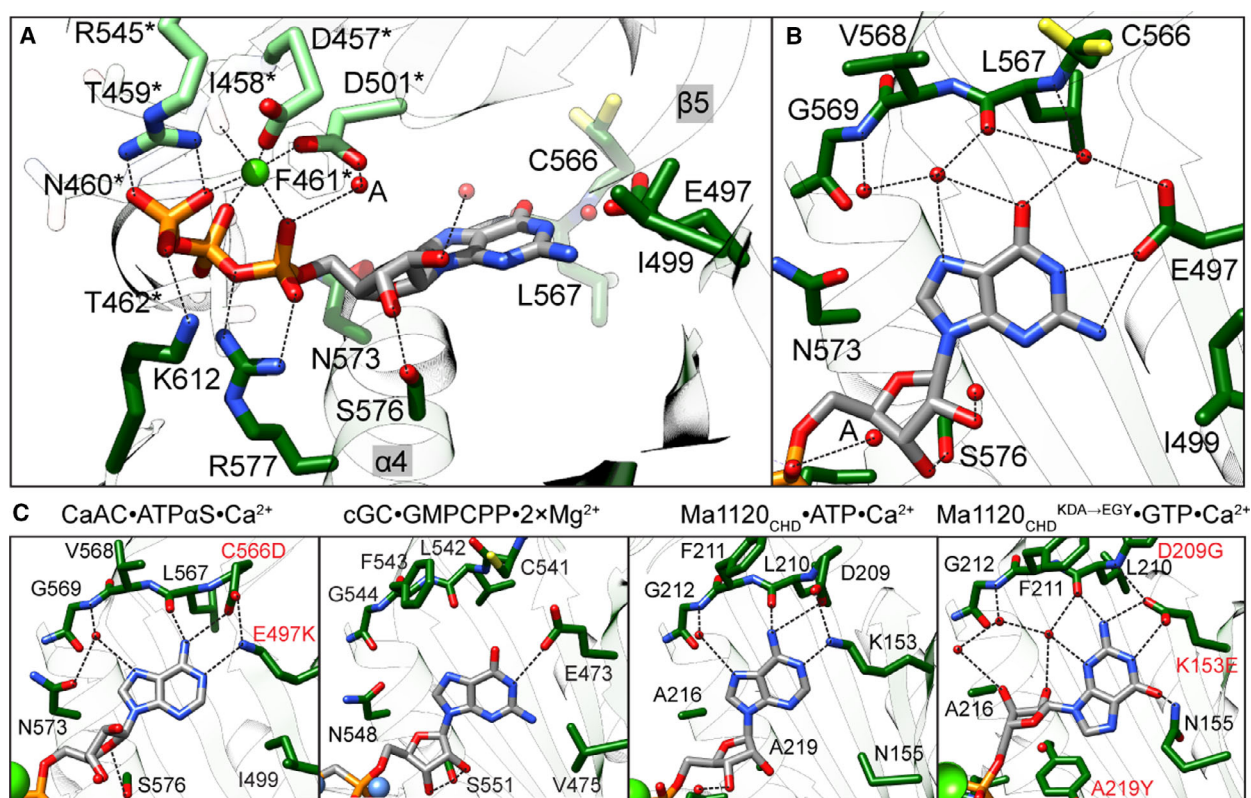


Fig. 3. Active site of CaGC-GTP- Ca^{2+} complex. (A) Active site in CaGC-GTP- Ca^{2+} structure. Chains A and B are in light and dark green, respectively. The calcium ion is shown as a green sphere. The GTP, residues crucial for substrate binding and turnover are shown as sticks. Water molecules are shown as red spheres. The metal coordination and selected hydrogen bonds between protein, triphosphate and ribose are shown as black dashed lines. (B) Close-up view of the protein-guanine interaction in the CaGC-GTP- Ca^{2+} structure. (C) From left to right: close-up view of the protein-base interaction in CaGC-ATP α S- Ca^{2+} (PDB 5OYH [11]), sGC-GMPCPP-2 \times Mg $^{2+}$ (PDB 6JT2 [21]), Ma1120_{CHD}-ATP- Ca^{2+} (PDB 5D15 [10]) and Ma1120_{CHD}^(KDA→EGY)-GTP- Ca^{2+} (PDB 5D0G [10]) structures. Introduced mutations are marked in red. In all panels, maximum acceptor-donor distance is 3.5 Å and colour coding as in A. Molecular graphics and analyses were performed with the UCSF CHIMERA package [52].

active site geometry originate from the presence of noncatalytic ions or substrate modifications. For example, in many AC structures of nonproductive enzyme-substrate complexes, one or even both base interactions are missing [6–8,34,35]. In the CaGC-GTP- Ca^{2+} structure, the catalytic ion (site A) is unoccupied. Because our complex is inactive, our atomic model does not represent a reaction-competent state. We speculate that the coordination mode we observe, wherein the guanine-Cys566 hydrogen bond formation is hindered by the presence of calcium, is the principal reason our complex does not support catalysis. Nevertheless, hypothetical modelling of the CaGC active site suggests that all canonical interactions could be supported in the presence of two catalytic ions (Fig. 6).

RhGC activity is regulated by light via the type I rhodopsin domain located towards the N-terminal

region the GC domain. The question remains open as to how light induces structural changes in the rhodopsin sensor and how these are transmitted to the catalytic domain. A parallel coiled-coil element formed by two signalling (S)-helices is a distinct integral element involved in the dimerization of diverse protein fusions [41]. In RhGC, such an element is thought to bridge the rhodopsin and GC domain. In addition to playing a structural role, the S-helix has also been proposed to be involved in the signal transduction from the N-terminal sensory to various C-terminal catalytic domains, including type III NCs [19,21,41–43]. According to this model, the activation signal propagates through the S-helix and a short ‘handle’ helix (also described as cyclase transducer element) towards the catalytic domain. We favour the idea that the enzyme catalysis is modulated through the changes in the N-terminal dimer

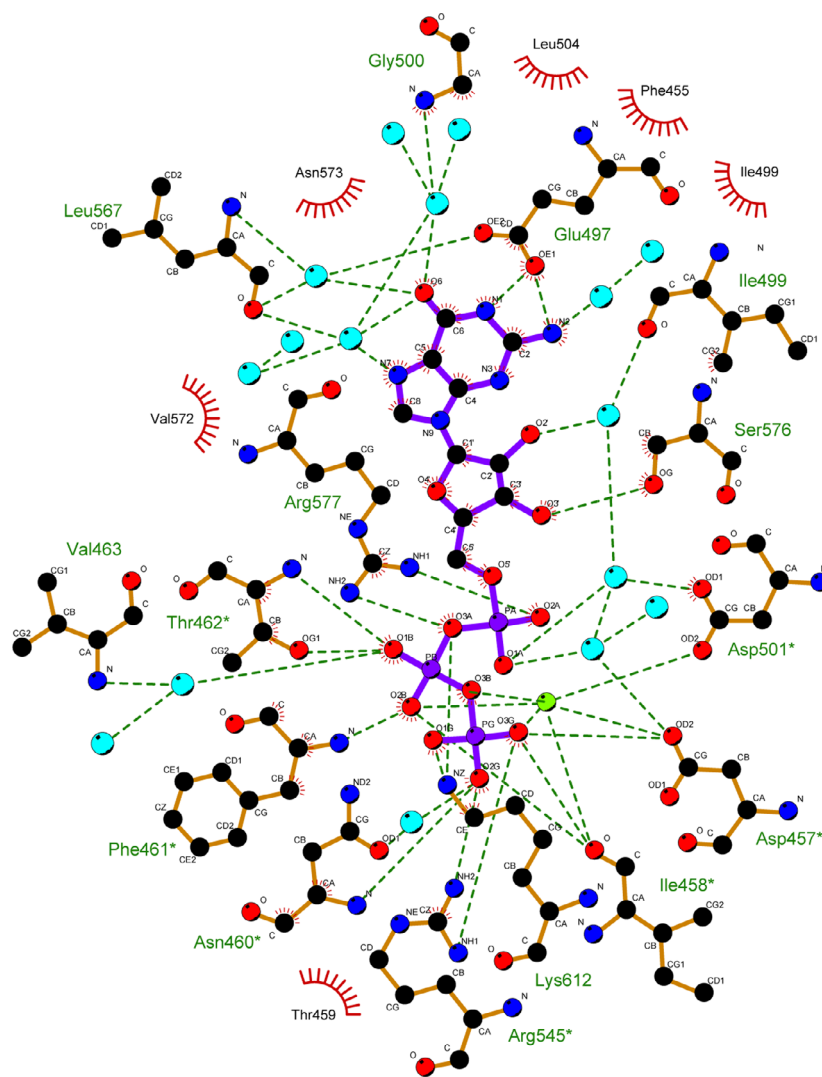


Fig. 4. GTP–CaGC interaction in one of the active sites. Residues involved in hydrophobic contacts are shown as a curved comb. Ionic and hydrogen bond interactions (maximal donor–acceptor distance = 3.5 Å) are shown as dark green dotted lines. Calcium and waters are represented by light green and light blue spheres, respectively. Figure generated using LIGPLOT+ [53].

interface, which subsequently affect the conformation of $\beta 4/\beta 5$ tongues and thereby affect the active site region where the base discrimination takes place [11,14,42]. Although a primary signal transduction mechanism may be common to type III NCs, the distinct regulatory mechanisms are likely to be protein-specific. For example, in the photoactivatable AC from *Beggiatoa* sp. (bPAC), the isolated catalytic domain is inactive and light stimulation promotes a conformational change to a catalytically competent state of the full-length protein [14]. Unlike bPAC, isolated GC domain of RhGC is constitutively active. This observation suggests that in RhGC the light signal might release the constraints on an otherwise intrinsically active enzyme [11,20]. Apart from the cyclase transducer element, the GC domain interacts with other regions of the protein through distinct interfaces providing an additional level of control.

This includes conserved N terminus that precedes the type I rhodopsin domain [11,27]. Truncations of this part of the protein, which is unique to RhGC, either disrupt light-dependent regulation of activity [27,28] or turn the protein entirely inactive [39]. Clearly, further studies on the full-length RhGC and other NCs in various physiological states will be required to understand how the interplay between different interactions is responsible for the regulation of the catalytic activity and how it differs between ACs and GCs.

Materials and methods

Cloning and protein purification

DNA encoding *C. anguillulae* RhGC was obtained from Integrated DNA Technologies (Leuven, Belgium) as a synthetic

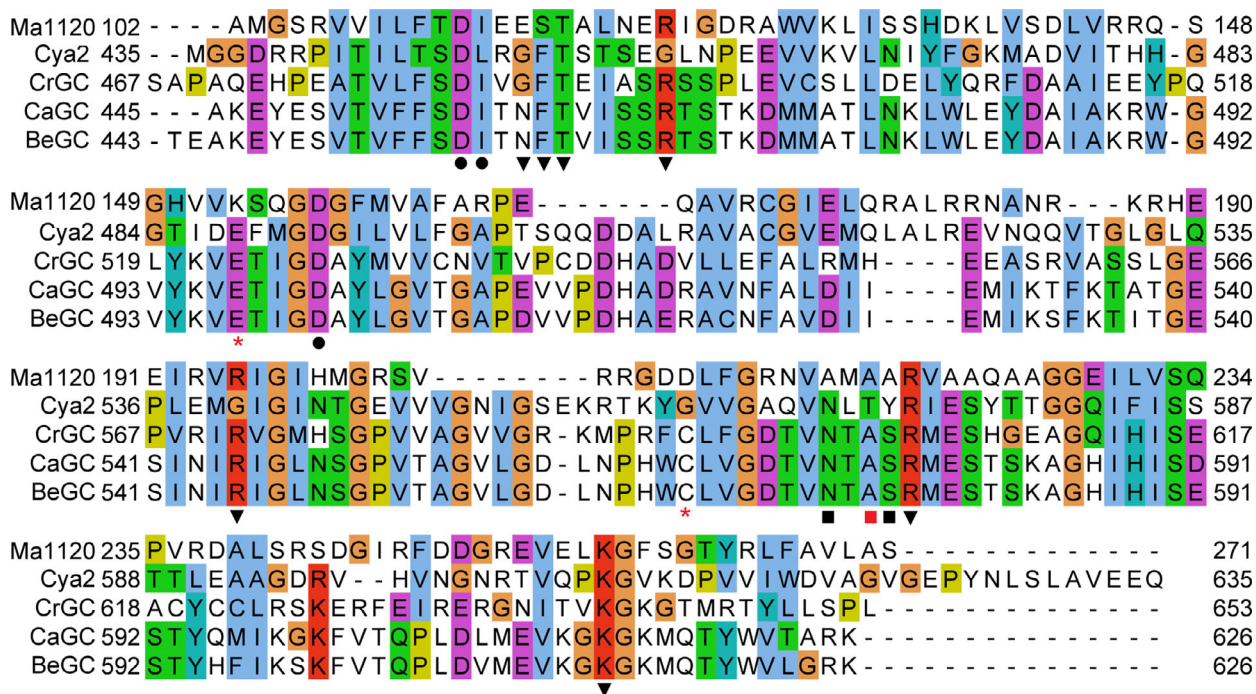


Fig. 5. Alignment of sequences of cyclases: *Catenaria anguillulae* CaGC (PDB 6SIR, this work), *Blastocladiella emersonii* BeGC (PDB 6A0B [20]), *Mycobacterium avium* Ma1120 (PDB 5D15 [10]), *Synechocystis* sp. Cya2 (PDB 2W01 [16]), *Chlamydomonas reinhardtii* GC (PDB 3ET6 [15]). Amino acids that coordinate to the catalytic ion (●), phosphate (▼), ribose (■) or base (*) are marked. Positions marked with red (*) are mutated in CaAC. Positions marked with red (*) and with red (■) are mutated in Ma1120_{CHD}^(KDA→EGY) mutant. Alignment was created using CLUSTAL OMEGA [54] and JALVIEW [55].

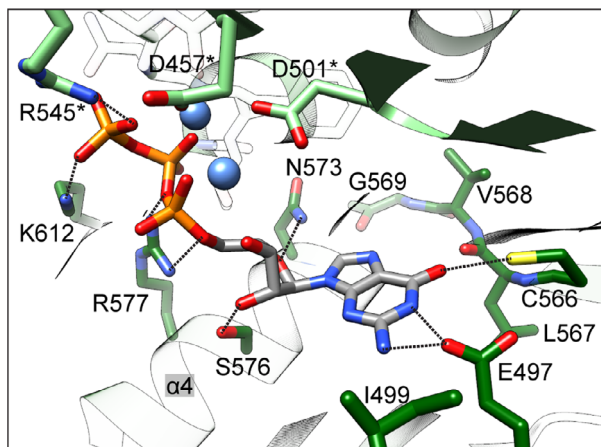


Fig. 6. Hypothetical model of the CaGC active site in the presence of two catalytic Mg²⁺ cations. Model was based on the Ma1120_{CHD}-ATP-Ca²⁺ crystal structure (PDB 5D15) [10]. Magnesium ions are shown as blue spheres. Selected hydrogen bonds between the protein and GTP are shown as black dashed lines. Colour coding is as in Fig. 3. Molecular graphics and analyses were performed with the UCSF CHIMERA package [52].

gene with a codon distribution optimized for expression in *E. coli*. The DNA construct corresponding to the catalytic domain (amino acid residues 442–626) was PCR-amplified

and cloned into pOPINM vector [44] with an N-terminal His₆-MBP fusion and HRV3C protease cleavage site. His₆-MBP-CaGC was transformed into *E. coli* Lemo21(DE3) cells (New England Biolabs, Hitchin, UK) and grown in TB medium (Merck Millipore, Watford, UK) at 37 °C until OD_{600 nm} reached 1.0. Expression was induced by the addition of IPTG at 1 mM final concentration and carried out at 18 °C overnight. After cell harvesting, the pellet was resuspended in lysis buffer (50 mM Tris 7.5, 500 mM NaCl, 10 mM imidazole, 2 mM β-mercaptoethanol) and sonicated for 15 min on ice. After centrifugation, supernatant was loaded onto HisTrap HP column (GE Healthcare, Little Chalfont, UK) equilibrated in the lysis buffer. The bound fraction was washed extensively with lysis buffer and eluted with elution buffer (50 mM Tris 7.5, 500 mM NaCl, 250 mM imidazole, 2 mM β-mercaptoethanol). HRV3C protease was added to the elution fractions to cleave off the N-terminal tag. Sample was then dialysed overnight against 20 mM Tris pH 7.5, 50 mM NaCl, 2 mM β-mercaptoethanol and loaded onto HiTrap Q HP column (GE Healthcare). Flow-through fractions were then concentrated and loaded onto HiLoad 16/600 Superdex 75 pg column (GE Healthcare) equilibrated in 5 mM Tris pH 7.8, 50 mM NaCl, 2 mM β-mercaptoethanol. Protein fractions were pooled and concentrated to 20 mg·mL⁻¹, flash-frozen in liquid nitrogen and stored at -80 °C.

SEC-MALS

Size exclusion chromatography coupled with multiangle light scattering (SEC-MALS) was performed on an Agilent 1100 system at 22 °C. Hundred microlitres of the protein solution at 5 mg·mL⁻¹ was injected onto Superdex 75 10/300 Increase column connected to ÄKTA pure (GE Healthcare) equilibrated in 50 mM TRIS pH 7.8, 50 mM NaCl, 0.5 mM TCEP and run at 1.3 mL·min⁻¹. The chromatography system was connected to a multiangle light scattering detector DAWN HELEOS-II (Wyatt Technology, Haverhill, UK) and Optilab T-rEX refractometer (Wyatt Technology) connected in-line. Data acquisition and processing were performed using ASTRA software (Wyatt Technology).

Pyrophosphate assay

To quantitate the level of generated cGMP, Phosphate Assay Kit (MAK168, Sigma-Aldrich, Gillingham, UK) was used to determine the amount of pyrophosphate. 4.7 µM CaGC was mixed with 0.8 mM GTP in 50 mM HEPES 7.5, 100 mM NaCl and 2.7 mM MgCl₂/MnCl₂/CaCl₂. After 1 h of incubation at 22 °C, the reaction was terminated by adding Master Reaction Mix, followed by 20-min incubation at 22 °C. Subsequently, fluorescence measurement (excitation 316 nm, emission 456 nm) was taken on a Spectramax M5 plate reader (Molecular Devices, Wokingham, UK). All measurements were done in triplicates.

Protein crystallization

Crystallization trials were done using CaGC at 20 mg·mL⁻¹ premixed with GTP and CaCl₂ (both at 1 mM final concentration). Several screening conditions yield protein crystals, most of which contained only protein and revealed no substrate bound and/or crystals of the protein–substrate complex that diffracted poorly. Well-diffracting crystals of the CaGC·GTP·Ca²⁺ complex in an orthorhombic space group were grown in 0.1 M MIB pH 4.4 and 22–24% PEG 2000 at 20 °C. Crystals appeared after one or two days as 300 µm × 20 µm × 20 µm rods. Crystals grew exclusively in the presence of GTP and CaCl₂. For data collection, crystals were cryo-cooled in mother liquor supplemented with 25% glycerol by plunging rapidly into liquid nitrogen.

Data collection and processing

Data were collected at I24, Diamond Light Source, UK, from a single crystal held at 100 K. Indexing and integration were done in DIALS to 1.7 Å resolution in space group *P*₂₁₂₁ [45]. Intensities were scaled and merged in AIMLESS [46,47]. The structure was solved by molecular replacement using PHASER [48] with the cyanobacterial GC Cya2 monomer as the search model (PDB 2W01) [16]. Distinctive electron density peaks in the 2*F*_o–*F*_c and *F*_o–*F*_c maps that

corresponded to one GTP and one Ca²⁺ ion in each active site of the CaGC·GTP·Ca²⁺ complex were fitted with these ligands and included in the structure refinement. Structure was refined using PHENIX [49] and COOT [50] and using the PDB_REDO server [51] to further optimize the model.

Acknowledgements

We thank beamlines I04, I24 and VMXi at Diamond Light Source where crystals were tested, and data were collected under proposals nr16818, nr19737 and mx19485. AB and AMO were supported in part by Strategic Award from the Wellcome Trust and the Biotechnology and Biological Sciences Research Council grant (102593), as well as Wellcome Investigator Award in Science (210734/Z/18/Z) and a Royal Society Wolfson Fellowship (RSWF\R2\182017) to AMO. RJO and HRad acknowledge support from the UK Medical Research Council (MR/K018779/1). IM acknowledges support from the Membrane Protein Laboratory Wellcome Trust grant (20289/Z16/Z).

Conflict of interest

The authors declare no conflict of interest.

Author contributions

AB, IM, RJO and AMO planned research; AB, HRad and HRad performed experiments; AB and AMO analysed data and wrote the paper; and all authors commented on the final manuscript.

References

- Potter LR (2011) Guanylyl cyclase structure, function and regulation. *Cell Signal* **23**, 1921–1926.
- Bassler J, Schultz JE & Lupas AN (2018) Adenylate cyclases: receivers, transducers, and generators of signals. *Cell Signal* **46**, 135–144.
- Tesmer JJ, Sunahara RK, Gilman AG & Sprang SR (1997) Crystal structure of the catalytic domains of adenylyl cyclase in a complex with G_sα.GTPγS. *Science* **278**, 1907–1916.
- Steebhorn C, Litvin TN, Levin LR, Buck J & Wu H (2004) Bicarbonate activation of adenylyl cyclase via promotion of catalytic active site closure and metal recruitment. *Nat Struct Mol Biol* **12**, 32.
- Tesmer JJ, Sunahara RK, Johnson RA, Gosselin G, Gilman AG & Sprang SR (1999) Two-metal-ion catalysis in adenylyl cyclase. *Science* **285**, 756–760.
- Mou TC, Masada N, Cooper DM & Sprang SR (2009) Structural basis for inhibition of mammalian adenylyl cyclase by calcium. *Biochemistry* **48**, 3387–3397.

- 7 Kleinboelting S, Diaz A, Moniot S, van den Heuvel J, Weyand M, Levin LR, Buck J & Steegborn C (2014) Crystal structures of human soluble adenylyl cyclase reveal mechanisms of catalysis and of its activation through bicarbonate. *Proc Natl Acad Sci USA* **111**, 3727–3732.
- 8 Kleinboelting S, van den Heuvel J & Steegborn C (2014) Structural analysis of human soluble adenylyl cyclase and crystal structures of its nucleotide complexes-implications for cyclase catalysis and evolution. *FEBS J* **281**, 4151–4164.
- 9 Steegborn C (2014) Structure, mechanism, and regulation of soluble adenylyl cyclases – similarities and differences to transmembrane adenylyl cyclases. *Biochim Biophys Acta* **1842**, 2535–2547.
- 10 Bharambe NG, Barathy DV, Syed W, Visweswariah SS, Colavarsigao M, Misquith S & Suguna K (2016) Substrate specificity determinants of class III nucleotidyl cyclases. *FEBS J* **283**, 3723–3738.
- 11 Scheib U, Broser M, Constantin OM, Yang S, Gao S, Mukherjee S, Stehfest K, Nagel G, Gee CE & Hegemann P (2018) Rhodopsin-cyclases for photocontrol of cGMP/cAMP and 2.3 Å structure of the adenylyl cyclase domain. *Nat Commun* **9**, 2046.
- 12 Tews I, Findeisen F, Sinning I, Schultz A, Schultz JE & Linder JU (2005) The structure of a pH-sensing mycobacterial adenylyl cyclase holoenzyme. *Science* **308**, 1020–1023.
- 13 Ohki M, Sugiyama K, Kawai F, Tanaka H, Nihei Y, Unzai S, Takebe M, Matsunaga S, Adachi S, Shibayama N *et al.* (2016) Structural insight into photoactivation of an adenylate cyclase from a photosynthetic cyanobacterium. *Proc Natl Acad Sci USA* **113**, 6659–6664.
- 14 Lindner R, Hartmann E, Tarnawski M, Winkler A, Frey D, Reinstein J, Meinhart A & Schlichting I (2017) Photoactivation mechanism of a bacterial light-regulated adenylyl cyclase. *J Mol Biol* **429**, 1336–1351.
- 15 Winger JA, Derbyshire ER, Lamers MH, Marletta MA & Kuriyan J (2008) The crystal structure of the catalytic domain of a eukaryotic guanylate cyclase. *BMC Struct Biol* **8**, 42.
- 16 Rauch A, Leipelt M, Russwurm M & Steegborn C (2008) Crystal structure of the guanylyl cyclase Cya2. *Proc Natl Acad Sci USA* **105**, 15720–15725.
- 17 Seeger F, Quintyn R, Tanimoto A, Williams GJ, Tainer JA, Wysocki VH & Garcin ED (2014) Interfacial residues promote an optimal alignment of the catalytic center in human soluble guanylate cyclase: heterodimerization is required but not sufficient for activity. *Biochemistry* **53**, 2153–2165.
- 18 Allerston CK, von Delft F & Gileadi O (2013) Crystal structures of the catalytic domain of human soluble guanylate cyclase. *PLoS ONE* **8**, e57644.
- 19 Ettl S, Lindner R, Nelson MD & Winkler A (2018) Structure-guided design and functional characterization of an artificial red light-regulated guanylate/adenylate cyclase for optogenetic applications. *J Biol Chem* **293**, 9078–9089.
- 20 Kumar RP, Morehouse BR, Fofana J, Trieu MM, Zhou DH, Lorenz MO & Oprian DD (2017) Structure and monomer/dimer equilibrium for the guanylyl cyclase domain of the optogenetics protein RhoGC. *J Biol Chem* **292**, 21578–21589.
- 21 Kang Y, Liu R, Wu JX & Chen L (2019) Structural insights into the mechanism of human soluble guanylate cyclase. *Nature* **574**, 206–210.
- 22 Vieira AL, Linares E, Augusto O & Gomes SL (2009) Evidence of a Ca²⁺-(*-NO-cGMP signaling pathway controlling zoospore biogenesis in the aquatic fungus *Blastocladiella emersonii*. *Fungal Genet Biol* **46**, 575–584.
- 23 Vale MR, Gomes SL & Maia JC (1975) Independent cAMP and cGMP phosphodiesterases in *Blastocladiella emersonii*. *FEBS Lett* **56**, 332–336.
- 24 Silverman PM & Epstein PM (1975) Cyclic nucleotide metabolism coupled to cytodifferentiation of *Blastocladiella emersonii*. *Proc Natl Acad Sci USA* **72**, 442–446.
- 25 Silverman PM (1976) Regulation of guanylate cyclase activity during cytodifferentiation of *Blastocladiella emersonii*. *Biochem Biophys Res Commun* **70**, 381–388.
- 26 Avelar GM, Schumacher RI, Zaini PA, Leonard G, Richards TA & Gomes SL (2014) A rhodopsin-guanylyl cyclase gene fusion functions in visual perception in a fungus. *Curr Biol* **24**, 1234–1240.
- 27 Gao S, Nagpal J, Schneider MW, Kozjak-Pavlovic V, Nagel G & Gottschalk A (2015) Optogenetic manipulation of cGMP in cells and animals by the tightly light-regulated guanylyl-cyclase opsin CyclOp. *Nat Commun* **6**, 8046.
- 28 Scheib U, Stehfest K, Gee CE, Korschen HG, Fudim R, Oertner TG & Hegemann P (2015) The rhodopsin-guanylyl cyclase of the aquatic fungus *Blastocladiella emersonii* enables fast optical control of cGMP signaling. *Sci Signal* **8**, rs8.
- 29 Lucas KA, Pitari GM, Kazerounian S, Ruiz-Stewart I, Park J, Schulz S, Chepenik KP & Waldman SA (2000) Guanylyl cyclases and signaling by cyclic GMP. *Pharmacol Rev* **52**, 375–414.
- 30 Sinha SC & Sprang SR (2006) Structures, mechanism, regulation and evolution of class III nucleotidyl cyclases. *Rev Physiol Biochem Pharmacol* **157**, 105–140.
- 31 Zimmermann G, Zhou D & Taussig R (1998) Mutations uncover a role for two magnesium ions in the catalytic mechanism of adenylyl cyclase. *J Biol Chem* **273**, 19650–19655.
- 32 Krissinel E & Henrick K (2007) Inference of macromolecular assemblies from crystalline state. *J Mol Biol* **372**, 774–797.

- 33 Hahn DK, Tusell JR, Sprang SR & Chu X (2015) Catalytic mechanism of mammalian adenylyl cyclase: a computational investigation. *Biochemistry* **54**, 6252–6262.
- 34 Saalau-Bethell SM, Berdini V, Cleasby A, Congreve M, Coyle JE, Lock V, Murray CW, O'Brien MA, Rich SJ, Sambrook T *et al.* (2014) Crystal structure of human soluble adenylate cyclase reveals a distinct, highly flexible allosteric bicarbonate binding pocket. *ChemMedChem* **9**, 823–832.
- 35 Ramos-Espiritu L, Kleinboelting S, Navarrete FA, Alvau A, Visconti PE, Valsecchi F, Starkov A, Manfredi G, Buck H, Adura C *et al.* (2016) Discovery of LRE1 as a specific and allosteric inhibitor of soluble adenylyl cyclase. *Nat Chem Biol* **12**, 838–844.
- 36 Barathy DV, Bharambe NG, Syed W, Zaveri A, Visweswariah SS, Colavarsigmao M, Misquith S & Suguna K (2015) Autoinhibitory mechanism and activity-related structural changes in a mycobacterial adenylyl cyclase. *J Struct Biol* **190**, 304–313.
- 37 Yan SZ, Huang ZH, Shaw RS & Tang WJ (1997) The conserved asparagine and arginine are essential for catalysis of mammalian adenylyl cyclase. *J Biol Chem* **272**, 12342–12349.
- 38 Tucker CL, Hurley JH, Miller TR & Hurley JB (1998) Two amino acid substitutions convert a guanylyl cyclase, RetGC-1, into an adenylyl cyclase. *Proc Natl Acad Sci USA* **95**, 5993–5997.
- 39 Trieu MM, Devine EL, Lamarche LB, Ammerman AE, Greco JA, Birge RR, Theobald DL & Oprian DD (2017) Expression, purification, and spectral tuning of RhoGC, a retinylidene/guanylyl cyclase fusion protein and optogenetics tool from the aquatic fungus *Blastocladiella emersonii*. *J Biol Chem* **292**, 10379–10389.
- 40 Ryu MH, Moskvina OV, Siltberg-Liberles J & Gomelsky M (2010) Natural and engineered photoactivated nucleotidyl cyclases for optogenetic applications. *J Biol Chem* **285**, 41501–41508.
- 41 Anantharaman V, Balaji S & Aravind L (2006) The signaling helix: a common functional theme in diverse signaling proteins. *Biol Direct* **1**, 25.
- 42 Ziegler M, Bassler J, Beltz S, Schultz A, Lupas AN & Schultz JE (2017) Characterization of a novel signal transducer element intrinsic to class IIIa/b adenylate cyclases and guanylate cyclases. *FEBS J* **284**, 1204–1217.
- 43 Vercellino I, Rezabkova L, Olieric V, Polyhach Y, Weinert T, Kammerer RA, Jeschke G & Korkhov VM (2017) Role of the nucleotidyl cyclase helical domain in catalytically active dimer formation. *Proc Natl Acad Sci USA* **114**, E9821–E9828.
- 44 Berrow NS, Alderton D, Sainsbury S, Nettleship J, Assenberg R, Rahman N, Stuart DI & Owens RJ (2007) A versatile ligation-independent cloning method suitable for high-throughput expression screening applications. *Nucleic Acids Res* **35**, e45.
- 45 Winter G, Waterman DG, Parkhurst JM, Brewster AS, Gildea RJ, Gerstel M, Fuentes-Montero L, Vollmar M, Michels-Clark T, Young ID *et al.* (2018) DIALS: implementation and evaluation of a new integration package. *Acta Crystallogr D* **74**, 85–97.
- 46 Evans PR & Murshudov GN (2013) How good are my data and what is the resolution? *Acta Crystallogr D Biol Crystallogr* **69**, 1204–1214.
- 47 Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AG, McCoy A *et al.* (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* **67**, 235–242.
- 48 McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC & Read RJ (2007) Phaser crystallographic software. *J Appl Crystallogr* **40**, 658–674.
- 49 Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213–221.
- 50 Emsley P, Lohkamp B, Scott WG & Cowtan K (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486–501.
- 51 Joosten RP, Long F, Murshudov GN & Perrakis A (2014) The PDB_REDO server for macromolecular structure model optimization. *IUCr J* **1**, 213–220.
- 52 Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC & Ferrin TE (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605–1612.
- 53 Laskowski RA & Swindells MB (2011) LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. *J Chem Inf Model* **51**, 2778–2786.
- 54 Sievers F & Higgins DG (2014) Clustal Omega, accurate alignment of very large numbers of sequences. *Methods Mol Biol* **1079**, 105–116.
- 55 Waterhouse AM, Procter JB, Martin DM, Clamp M & Barton GJ (2009) Jalview version 2 – a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191.