**STRUCTURAL BASIS OF GLYCOGEN BIOSYNTHESIS REGULATION IN BACTERIA**

**Javier Cifuente1,2,3,§, Natalia Comino1,2,3,§, Julene Madariaga-Marcos2,3,§, Sonia López-Fernández2,3, Mikel García-Alija1,2,3, Jon Agirre4, David Albesa-Jové1,2,3,5,\*, Marcelo E. Guerin1,2,3,5,\***

From the 1Structural Biology Unit, CIC bioGUNE, Bizkaia Technology Park, 48160 Derio, Spain, 2Unidad de Biofísica, Centro Mixto Consejo Superior de Investigaciones Científicas - Universidad del País Vasco/Euskal Herriko Unibertsitatea (CSIC,UPV/EHU), Barrio Sarriena s/n, Leioa, Bizkaia, 48940, Spain, 3Departamento de Bioquímica, Universidad del País Vasco, Spain, 4York Structural Biology Laboratory, Department of Chemistry, The University of York, YO10 5DD, UK, 5IKERBASQUE, Basque Foundation for Science, 48013, Bilbao, Spain.

Running title: *Regulation of Glycogen Biosynthesis in Bacteria*

\*To whom correspondence may be addressed: Dr. David Albesa-Jové and Prof. Marcelo E. Guerin, Structural Biology Unit, CIC bioGUNE, Bizkaia Technology Park, 48160 Derio, Spain, Tel +34 944 061 309, E-mail: David Albesa-Jové, [dalbesa@cicbiogune.es](mailto:dalbesa@cicbiogune.es); Marcelo E. Guerin (lead contact), [mrcguerin@cicbiogune.es](mailto:mrcguerin@cicbiogune.es)

**§** These authors equally contributed to this work

**SUMMARY**

ADP-glucose pyrophosphorylase (AGPase) catalyzes the rate-limiting step of bacterial glycogen and plant starch biosynthesis, the most common carbon storage polysaccharides in nature. A major challenge is to understand how AGPase activity is regulated by metabolites in the energetic flux within the cell. Here we report crystal structures of the homotetrameric AGPase from *Escherichia coli* in complex with its physiological positive and negative allosteric regulators, fructose-1,6-bisphosphate (FBP) and adenosine 5'-monophosphate (AMP), and sucrose in the active site. FBP and AMP bind to partially overlapping sites located in a deep cleft between glycosyltransferase-A like and left-handed β-helix domains of neighboring protomers, accounting for the fact that sensitivity to inhibition by AMP is modulated by the concentration of the activator FBP. We propose a model in which the energy reporters regulate *Ec*AGPase catalytic activity by intra-protomer interactions and inter-protomer crosstalk, with a Sensory Motif (SM) and two regulatory loops playing a prominent role.

**INTRODUCTION**

Glycogen is a very large branched glucose homopolymer containing about 90% α-1,4-glucosidic linkages and 10% α-1,6 linkages (Ball et al., 1996; Roach et al., 2012). Glycogen localizes as discrete cytoplasmic granules of less than 50 nm found in most living organisms, ranging from archaebacteria, bacteria, fungi and higher eukaryotes. Eukaryotes utilize UDP-glucose as the activated nucleotide donor for glycogen biosynthesis, whereas archaebacteria and bacteria, have selected ADP-glucose (Leloir and Cardini, 1957; Recondo and Leloir, 1961). This defines two different pathways with distinct regulatory mechanisms and rate controlling steps (Preiss 1984; Roach 2012; Ball and Morell, 2003). In bacteria, the basic glycogen biosynthetic pathway involves the action of three enzymes: ADP-Glc pyrophosphorylase (AGPase), glycogen synthase (GS) and branching enzyme (BE). The first step is carried out by the AGPase, which catalyzes the biosynthesis of ADP-Glc (Ballicora et al., 2003). The second step generates linear α-1,4-linked glucose chains, a reaction catalyzed by GS (Buschiazzo et al., 2004), whereas the third step produces α-1,6-linked glucan branches in the polymer, a reaction catalyzed by BE (Feng et al. 2015). In contrast, glycogen degradation is carried out by the glycogen phosphorylase, which functions as a depolymerizing enzyme, and the debranching enzyme that catalyze the removal of α-1,6-linked ramifications.

ADP-Glc biosynthesis, mediated by AGPase, is considered the main regulatory step in glycogen and starch production in these organisms (Preiss, 1978). Specifically, AGPase catalyzes the reaction between ATP and α-D-glucose-1-phosphate (G1P) to bring forth pyrophosphate and ADP-Glc in the presence of the divalent metal ion Mg2+ (scheme 1; Figure S1).

ATP + G1P ↔ ADP-Glc + PPi (scheme 1)

AGPase activity displays cooperative behavior and a bi-bi mechanism with sequential binding of ATP and G1P, followed by ordered release of pyrophosphate and ADP-Glc (Gentner and Preiss, 1968; Paule and Preiss, 1971). The hydrolysis of pyrophosphate by the action of inorganic pyrophosphatases results in a global irreversible and energetically expensive reaction *in vivo* (Lahti, 1983). Interestingly, the expression of AGPase is highly regulated in response to fluctuating carbon/energy levels in the cell. In the case of the paradigmatic bacterial AGPase from *Escherichia coli* (*Ec*AGPase), the enzyme is encoded by a single gene (*glgC*) located inside an operon together with the genes that code for GS (*glgA*), GP (*glgP*), BE (*glgB*) and phosphoglucomutase (*pgm*) (Preiss and Romero, 1989). The resulting *Ec*AGPase protomers build into a physiological and functional homotetrameric structure. In contrast, plants AGPases consist of heterotetramers displaying two large subunits and two small subunits encoded by different genes (Crevillén et al. 2003; Georgelis et al. 2007; Petreikov et al. 2010; Ventriglia et al. 2007). To date, two crystal structures of AGPases have been reported, that of the bacterial AGPase from *Agrobacterium tumefaciens* (*At*AGPase; Cupp-Vickery et al., 2008) and the photosynthetic potato tuber AGPase (*St*AGPase; Jin et al., 2005).

Strikingly, evolution also led AGPase to acquire allosteric properties to control this key rate-limiting step by essential metabolites in the energetic flux within the cell ⁠(Preiss, 1978). In general, AGPase activators are metabolites that represent signals of high carbon and energy content of a particular bacteria or tissue, while inhibitors of the enzyme indicate low metabolic energy levels ⁠(Ballicora et al., 2003, Ball, 2011). Based on the specific positive or negative allosteric regulators, AGPases have been grouped into 9 different classes (Ballicora et al., 2003, Preiss, 1978). Glycolytic intermediates as fructose-6-phosphate (F6P), FBP and/or pyruvate, acts enhancing the activity of bacterial AGPases, whereas AMP, ADP and/or Pi display inhibitory properties. In contrast, AGPases from photosynthetic organisms, as plants and cyanobacteria, prefer 3-PGA as positive signal produced by the photosynthetic activity, and Pi as the inhibitory signal⁠. Thus, the understanding of the regulatory mechanism at the molecular level by which AGPase modulates catalysis represents a major challenge. Here, by using X-ray crystallography, we report the first crystal structures of an AGPase in complex with their physiological negative and positive allosteric regulators. In combination with biophysical data we provide unprecedented insight into the molecular mechanisms of AGPase allosteric regulation.

**RESULTS**

**Overall structure of *Ec*AGPase**

The crystal structures of the paradigmatic *Ec*AGPase were solved by molecular replacement using a tetramer of *At*AGPase (pdb code 3BRK) in two different states, including the complexes with its naturally occurring and preferred allosteric negative regulator AMP (*Ec*AGPase•AMP•SUC) and positive regulator FBP (*Ec*AGPase•FBP; Preiss et al., 1966). In addition, the *Ec*AGPase•AMP•SUCstructure displayed sucrose (SUC) located in the active site of the enzyme. *Ec*AGPase•AMP•SUC and *Ec*AGPase•FBP forms crystallized in space group *P* 21 with 16 molecules (431 residues each) in the asymmetric unit and diffracted to a maximum resolution of 2.67 Å and 3.04 Å, respectively (Table 1). *Ec*AGPase crystallized as a homotetramer with each protomer (48,7 kDa) composed of two domains, the N-terminal glycosyltransferase A-like domain (GT-A like; residues 1-315), containing the active site, and the C-terminal regulatory domain (residues 316 to 431) comprising a left-handed parallel β-helix (LβH; residues 316-396; Figure 1A). The GT-A like domain consists of one Rossmann fold domain (residues 1-315; Pelissier et al., 2010). The core is composed of a central β-sheet comprising seven β-strands (β5,β4,β1, β8, β14, β10, β15of which β14 is antiparallel) flanked on both sides with several α-helices (Figure 1A). In contrast, the LβH domain (residues 316 to 396; Raetz and Roderick, 1995) is built of short β-strands (β17-30), oriented parallel to each other, and describing a triangular prism (Figure 1A). The *Ec*AGPase protomers build into a physiological and functional homotetrameric structure (194.8 kDa) that can be viewed as a dimer of dimers(Figure 1B). The most important contribution to the dimer interface is the triangular base of the LβH prism of each protomer (strands β17, β18 and β19), resulting in two anti-parallel β-sheets. In contrast, the tetramer assembles mainly by interactions between the N-terminal GT-A like domains from different dimers. Specifically, as depicted in Figures 1C, three adjacent α-helices, α1 (residues 10 to 19), α6 (residues 93 to 95) and α8 (residues 149 to 159) from the GT-A like domain of protomer A interact with the equivalent structural elements of the GT-A like domain of protomer D. Moreover, α5 (residues 78 to 87), α7 (117 to 131) and β5 (98 to 103) interact with the equivalent structural elements of the GT-A like domain of protomer C, strongly contributing to anchoring both dimers in a competent tetramer configuration (Figure 1C-D). The resulting architecture allows the protomers to communicate with each other, from which cooperativity emerges.

**The active site of *Ec*AGPase**

The active site is located in a deep cleft of the GT-A like domain, as observed in other nucleotide sugar pyrophosphorylases (Jin et al., 2005; Cupp-Vickery et al., 2008; Figure S2 and S3). The crystal structure of *Ec*AGPase•AMP•SUC revealed the presence of SUC in the C-terminal region of the active site (residues 163-315). SUC is clearly visible in the electron density maps and is present in all four active sites of the *Ec*AGPase homotetramer. The glucose moiety binds to a deep pocket (the ‘sugar binding pocket’ according to Brito et al., 2011) defined by three β-strands, β11 (residues 179 to 183), β12 (residues 189 to 194) and β13 (residues 208 to 212), and four loops: β12- β13 (residues 194 to 208), α9-β11 (residues 177 to 179), α9-α12 (residues 261 to 280) and β8-α8 (residues 140 to 149). The O2 and O4 atoms of the glucose ring make hydrogen bonds with the side chain of Glu194 and the main chain carbonyl atom of Ser212, respectively. The glucose O6 is hydrogen bonded with the lateral chain of His143. Importantly, the replacement of His143, Glu194, and Ser212 per alanine displayed lower apparent affinity for G1P compared with the *Ec*AGPase wild type(Bejar et al., 2006). Several aromatic residues, including Phe178, Phe192, Tyr216, and Trp274 are important to constitute the walls of the cavity. Interestingly, the structural comparison of the *Ec*AGPase•AMP•SUC complex with that of the glucose-1-phosphate thymidylyltransferase RmlA from *Pseudomonas aeruginosa* in complex with G1P (pdb code 1G0R; Blankenfeldt et al., 2000) revealed that the glucose moieties superimpose very well (Figure S3). Taken into account all that experimental data, we clearly defined the location of G1P in the active site of *Ec*AGPase.

The ATP binding site is located in the N-terminal region of the GT-A like domain (residues 20-162; Figure S3). The crystal structures of *Ec*AGPase•AMP•SUC and *Ec*AGPase•FBP revealed the presence of PO4 and SO4 ions, respectively, in the ATP binding site. Both anions are bound in equivalent positions, making strong interactions with the lateral and main chains of Arg32 and the main chain of Thr31. The structural comparison of both *Ec*AGPase complexes with that of the N-acetylglucosamine-1-phosphate uridyltransferase GlmU from *Mycobacterium tuberculosis* in complex with ATP (pdb code 4K6R; Vithani et al., 2014) and the GDP-Man pyrophosphorylase from *Thermotoga maritima* in complex with GTP (pdb code 2X60; Pelissier et al., 2010) revealed that the anions superimpose with the ATP γ-PO4. According to this configuration, ATP accommodates into the active site of *Ec*AGPase, in close contact to the essential catalytic Lys42 and favorably positioned to receive the G1P (Ballicora et al., 2005; Ballicora et al., 2007; Figure S1 and S3).

**The AMP allosteric binding site**

The identification of the physiological positive and negative regulatory sites at the molecular level in AGPases has been a long-standing question and the matter of intense research in the field of glycogen/starch biosynthesis/regulation (Figure 2). In the *Ec*AGPase•AMP•SUC crystal structure, AMP is clearly visible in the electron density maps and present in all four allosteric sites located in the corresponding clefts between the N-terminal GT-A like and C-terminal LβH domains of neighboring protomers from different dimers (Figure 2A-C). Specifically, AMP is deeply buried into a cleft mainly defined by (i) the N-terminal β2-β3 hairpin (residues 46-52), α5 and the connecting loop α2-α3(residues 37-42), (ii) the C-terminal α15 (residues 419-425) and the connecting loops β28-β29 (residues 384 to 388) and β25-β26 (residues 367 to 371), and (iii) the N-terminal α7 from a neighbor protomer. The α-PO4 group occupies a cavity rich in positively charged residues and including Arg40 (α3), His46 and Arg52 (β2-β3 hairpin), Thr79 (α5), and Arg386 (LβH; Figure 2C). The adenine heterocycle is stabilized by a strong stacking interaction with Arg130 (α7) from the GT-A like domain of the neighbor protomer, and van der Waals interactions with Arg419 (α15) and Arg386 from the LβH domain. The side chain of Glu420 (α15) forms an important salt bridge with Lys39, communicating the α2-α3 loop with the α15. In addition, a strong hydrogen bonding interaction of the adenine N6 nitrogen with the side chain carboxylate group of Glu270 (LβH domain) may account for the nucleotide specificity. Finally, the ribose O2 atom makes a hydrogen bond with the side chain of Arg130 being also at van der Waals distance of Lys39 side chain.

Interestingly, a crosstalk event between protomers of same and different dimers, suggests AMP interactions might lead to the stabilization of the quaternary structure of *Ec*AGPase in solution. Supporting this notion, thermal unfolding followed by the far-UV CD signal at 222 nm indicated important differences in protein stability between the *apo* *Ec*AGPase and the *Ec*AGPase•AMP complex. The apparent melting temperatures (*Tm*) of *Ec*AGPase and *Ec*AGPase•AMP were 71.2 °C and 75.8 °C respectively, indicating that the AMP complexed form is *ca.* 4.6 °C more stable than the unliganded form (Figure 2GI).

**The FBP allosteric binding site: partial overlapping with the AMP binding site**

The *Ec*AGPase•FBP crystal structure reveals that FBP binds into the same cleft than AMP, but located in a more solvent exposed environment, with no evident interactions with neighbor protomers (Figure 2D-F). However, the structural comparison of the *Ec*AGPase•AMP•SUC and *Ec*AGPase•FBP crystal structures revealed that the AMP and FBP binding sites partially overlap (Figure 3).The FBP binding site comprises the last C-terminal residues of the enzyme (residues 420 to 431), with the FBP making important interactions with positively charged residues located in one side of α15 (Figure 2F). Supporting the relevance of the C-terminus in the recognition of FBP, a protein chimera containing the N-terminus (271 residues) of *At*AGPase, activated by fructose 6-phosphate and pyruvate, and the C-terminus (153 residues) of *Ec*AGPase, retained the selectivity for FBP (Ballicora et al., 2002). Moreover, a variant of *Ec*AGPase in which the two last C-terminal residues, Glu430 and Arg431, were removed, became less sensitive to FBP activation (Wu and Preiss, 2001).

FBP binding does promote important local conformational changes in the allosteric site when compared to the AMP complex (Figure 3B). Specifically, Lys39 side chain coordinates the O1 atom of the FBP PO4 group at position 6, whereas the side chain of Glu420 makes a hydrogen bond with the O3 of the fructose ring. Biochemical studies demonstrated the important role of Lys39 in the binding and the mechanism of activation of *Ec*AGPase by FBP (Gardiol and Preiss, 1990). Interestingly, Lys39 showed protection to the covalent modification with pyridoxal-PO4 (PLP), by reduction with NaBH4, in the presence of FBP (Parsons and Preiss, 1978). The modification of Lys39 with PLP resulted in an enzyme with a permanently enhanced activity, even in the absence of FBP. This result suggests that the Schiff base formed between Lys39 and PLP might result in binding of the PLP phosphate group to the allosteric site, mimicking FBP binding and contributing to permanently lock the enzyme in the activated state. In addition, the side chain of Arg423 positions its guanidinium group in close contact with the O5 atom of the FBP phosphate group at position 1, whose O6 atom makes a strong hydrogen bond with the main chain of Gln429. This structural configuration allows the fructose to be positioned in close proximity to the α15, allowing the side chain of Arg419 to make an important hydrogen bond with the sugar ring O2 atom. Finally, the last two residues, Glu430 and Arg431 become structured in the *Ec*AGPase•FBP complex.

Interestingly, in the *Ec*AGPase•FBP complex, the side chain of Arg130 from the neighbor protomer, completely change its conformation, suggesting FBP interactions might not lead to the stabilization of the quaternary structure of *Ec*AGPase in solution. The *Tm* value of *Ec*AGPase•FBP was 72.0 °C, indicating the formation of a less stable complex than that observed for *Ec*AGPase•AMP (Figure 2H). Moreover, the addition of FBP to the *Ec*AGPase•AMP complex triggered a clear reduction in the *Tm* values as revealed by the CD experimental data, indicating that FBP not only is able to compete with AMP but also to modify the structural arrangement of the *Ec*AGPase•AMP complex, leading to the occurrence of a less stable structure (Figure 2I). Altogether the structural configuration of the *Ec*AGPase regulatory site, in which the AMP and FBP binding sites partially overlap, account for the fact that sensitivity to inhibition by AMP is modulated by the concentration of the activator FBP (Preiss, 1978). In addition, the experimental data indicate that the *Ec*AGPase•FBP complex is markedly less stable and more flexible/dynamic than the *Ec*AGPase•AMP complex (Figueroa et al., 2011).

**DISCUSSION**

**A model for the allosteric regulation of *Ec*AGPase**

The closed inspection of the *Ec*AGPase•AMP•SUC and *Ec*AGPase•FBP crystal structures revealed how the allosteric and active sites are connected each other. The AMP allosteric site communicates with the active site (Figure 3C) of the same protomer through a region comprised of 27 residues, that we have defined as the ‘Sensory Motif’ (SM thereafter), located between β1 and α4 (residues 26 to 52; Figure 3D). The SM is constituted by (*i*) the nucleotide-binding loop NBL (residues 26 to 33) including the GGxGxR consensus sequence involved in ATP binding; followed by (*ii*) a segment rich in short secondary structure elements (residues 34 to 52) including α2(residues 34 to 37), α3(residues 42 to 44), β2(residues 46 to 47) and β3(51 to 52), the latest arranged in the form of a β-hairpin (residues 46 to 52). In addition to the NBL loop, two side chain residues of the SM motif face the active site playing a prominent role in ATP recognition and catalysis: Arg32 (α2) interact with the γ-PO4, whereas Lys42 (α3) is proposed to participate in the electrostatic stabilization of the transition state (Figure S1B and S3; Führing et al., 2013; Ballicora et al., 2005). In contrast, Arg40 (α3), His46 and Arg52 (β2-β3 hairpin), face the AMP allosteric site, strongly interacting with the α-PO4. The flexible β4-α5loop (residues 73 to 77; Regulatory Loop 1; RL1 thereafter) interacts with both the NBL loop and the segment rich in secondary structure elements located on the same protomer, likely modulating their conformations. This loop also connects with the α-helix 5 (residues 78 to 87) of which Thr79 interacts with the α-PO4 of AMP. Strikingly, the AMP binding site not only connects with the active site by intra protomer interactions, but also through inter-protomer cross talk. The adenine heterocycle of AMP makes an important stacking interaction with the side chain of Arg130 from a neighbor protomer of a different dimer. This important residue is located inside the α7(residues 117 to 131), and further communicates with a long loop (residues 104 to 116; Regulatory Loop 2; RL2 thereafter) that flanks the ATP binding pocket in the active site (Figure S3). In addition, the β2-β3 hairpin directly interacts with the loop connecting the N- and C-terminal domains (292-315) of the neighbor protomer of the same dimer. In contrast, the FBP allosteric site communicates with the active site mainly through the SM of the same protomer, involving a key interaction of Lys39 with the O1 atom of FBP phosphate group at position 6. This interaction directly modulates the conformation of the catalytic Lys42, located in the same loop, which is essential for the reaction to take place (Figure 3A, S1 and S3; Ballicora et al., 2007). Altogether we propose a model in which the positive and negative energy reporters regulate AGPase catalytic activity via intra-protomer interactions and inter-protomer crosstalk, with the SM motif and two critical regulatory loops RL1 and RL2 flanking the ATP binding site, playing a prominent role.

**The allosteric sites are essentially preserved in AGPases**

The crystal structure of *At*AGPase has been solved in the presence of SO4 (pdb code 3BRK; Cupp-Vickery et al., 2008). *Ec*AGPase primary sequence shares a 55% identity with *At*AGPase. The overall fold between *Ec*AGPase and *At*AGPase is essentially preserved with (i) an r.m.s.d. of 2.77 Å for the monomer and (ii) an r.m.s.d. of 2.92 Å for the tetramer. Importantly, multiple amino acid sequence alignments among the bacterial AGPase family, weighted by structural alignment of *Ec*AGPase and *At*AGPase, strongly support a common mechanism for the regulation of the enzymatic activity (Figure 4 and S4). The positively charged residues Arg40, His46, Arg52 and Arg 386, involved in the binding of the α-PO4 moiety of the negative regulator AMP, are highly conserved. Interestingly, the closed inspection of the *At*AGPase crystal structure shows that SO4 superimposes well with the α-PO4 of AMP in EcAGPase (Cupp-Vickery et al., 2008; Figure S4). In addition, Thr79 and Arg130, involved in the nucleoside ring binding of AMP, are mostly conserved within the enterobacteria family. The C-terminal region 419RxMLRKLxxKQER431 involved in FBP binding and the key residue Lys39 are also conserved among enterobacteria AGPases that use FBP as a positive regulator. Importantly, critical residues that participate in the SM motif and RL1 and RL2 loops are also preserved.

*St*AGPase is composed of two α and two β subunits, also referred to as small and large subunits, respectively, to form an α2β2 heterotetramer. The α subunit of AGPases is highly conserved in higher plants (85–95% identity), whereas the β subunit is less conserved (50–60% identity). In the *St*AGPase, the α and β subunits share 53% identity (Jin et al. 2005). Importantly, the two subunits have different functions: α is the catalytic subunit whereas β is the regulatory subunit. The crystal structure of a non-physiological, truncated recombinant homotetrameric version of the small subunit (α4) of *St*AGPase was solved in the presence of (i) SO4 (pdb code 1YP2), (ii) ATP (pdb code 1YP3) and (iii) ADP-Glc (pdb code 1YP3; Jin et al. 2005). *Ec*AGPase primary sequence shares a 31% identity with *St*AGPase. Although the overall fold between *Ec*AGPase and *St*AGPase is preserved, with (i) an r.m.s.d. of 4.35 Å for the monomer and (ii) an r.m.s.d. of 5.20 Å for the tetramer, clear differences can be found both in the GTA-like and the LβH domains, as revealed by the structural-weighted alignment (Figure 4). The structural comparison of *Ec*AGPase with *St*AGPase, an enzyme that is negatively regulated by inorganic phosphate (Pi), shows that SO4 binds to equivalent residues Arg40 and Arg52, and to a lysine occupying an equivalent position to Arg486 (Jin et al. 2005). Thus, the positively charged pocket responsible for the binding of (i) the AMP α-PO4 in bacterial AGPases and (ii) the Pi in plant AGPases, seems to be conserved in both families being essential for the negative regulation of most AGPases. Interestingly, Lys39 was also observed in several plant AGPases that uses 3-PGA as positive regulator, suggesting that the PO4 groups might be coordinated in a similar manner than FBP in *Ec*AGPase. The implication of plant AGPase as a critical enzyme in the regulation of starch biosynthesis is well established (Stitt and Zeeman, 2012). It is worth noting that the transformation of plants with *E. coli* allosteric mutants on the *glgC* gene significantly increased starch content (Tuncel and Okita, 2013). Therefore, the information reported herein provides exciting possibilities for industrial/biotechnological applications. Finally, the structural comparison of *Ec*AGPase with other non-regulated nucleotide sugar pyrophosphorylases (Figure S4; Vithani et al., 2014) revealed that although the secondary structure elements that conform the SM motif are essentially preserved, residues involved in allosteric regulation mediated by PO4 groups in AGPases are not. Therefore, those residues seem to be evolutionary traits acquired by this motif, to recognize these key allosteric modulators in order to regulate the glycogen pathway.

**EXPERIMENTAL PROCEDURES**

***Ec*AGPase Cloning, Expression and Purification** -*Ec*AGPase was expressed in *E. coli* BLL21(DE3) cells and purified to apparent homogeneity. The purification protocol comprised three main steps, including anionic exchange, ammonium sulfate precipitation and hydrophobic interaction criterias. *Ec*AGPase construct has no additional amino acids when compared to the native enzyme.

***Ec*AGPase Crystallization and Data Collection** -Crystallization trials were carried out using the sitting drop method. *Ec*AGPase•AMP•SUC and *Ec*AGPase•FBP complete datasets were collected at DLS. *Ec*AGPase•AMP•SUC and *Ec*AGPase•FBP forms crystallized in space group *P* 21 with 16 molecules in the asymmetric unit and diffracted to a maximum resolution of 2.67 Å and 3.04 Å, respectively (Table 1).

***Ec*AGPase Structure Determination and Refinement** -The crystal structures of *Ec*AGPase•AMP•SUC and *Ec*AGPase•FBP were solved by molecular replacement essentially using a tetramer from *At*AGPase (pdb code 3BRK; Cupp-Vickery et a., 2008)⁠. Atomic coordinates and structure factors have been deposited with the Protein Data Bank, accession codes 5L6V (*Ec*AGPase•AMP•SUC) and 5L6S (*Ec*AGPase•FBP).

***Ec*AGPase Thermal unfolding Analysis** -Thermal unfolding transitions were recorded on a J-810 CD spectropolarimeter (Jasco Corp., Tokio, Japan) at 222 nm by using Hellma 110-QS quartz cuvettes with a 1 mm optical path.

Details on the Experimental Procedures and any associated references are available in the Supporting Information section of the paper.

**AUTHOR CONTRIBUTIONS**

J.O.C., D.A-J. & M.E.G., conceived the project. J.O.C., N.C., J.M.-M., S.L-F., J.A., M.G-A & D.A-J., performed the experiments. J.O.C., N.C., D.A-J. & M.E.G., analyzed the results. J.O.C., D.A-J. & M.E.G., wrote the manuscript.

**ACKNOWLEDGMENTS**

This work was supported by the European Commission contract HEALTH-F3-2011-260872, and the MINECO contract BIO2013-49022-C2-2-R (to M.E.G.). J.A. was supported by Biotechnology and Biological Sciences Research Council contract BB/K008153/1. We acknowledge Diamond Light Source (DLS) (beamline I04 and I04-1 under proposals 8302 and 10130) and SOLEIL. Access to structural biology facilities was supported in part by the EU FP7 infrastructure grant BIOSTRUCT-X (contract no. 283570).

**COMPETING FINANCIAL INTERESTS**

The authors declare no competing financial interests.

**REFERENCES**

Ball, S., Guan, H. P., James, M., Myers, A., Keeling, P., Mouille, G., Buléon, A., Colonna, P., Preiss, J. (1996) From glycogen to amylopectin: a model for the biogenesis of the plant starch granule. Cell *86*, 349–352.

Ball, S. G., and Morell, M.K. (2003). From bacterial glycogen to starch: understanding the biogenesis of the plant starch granule. Annu. Rev. Plant Biol. *54*, 207–233.

Ball, S., Colleoni, C., Cenci, U., Raj, J. N., Tirtiaux, C. (2011). The evolution of glycogen and starch metabolism in eukaryotes gives molecular clues to understand the establishment of plastid endosymbiosis. J. Exp. Bot. *62*, 1775–1801.

Ballicora, M. A., Sesma, J. I., Iglesias, A. A., Preiss, J. (2002). Characterization of chimeric ADP-glucose pyrophosphorylases of *Escherichia coli* and *Agrobacterium tumefaciens*. Importance of the C-terminus on the selectivity for allosteric regulators. Biochemistry *41*, 9431–9437.

Ballicora, M. A., Iglesias, A.A., and Preiss, J. (2003). ADP-glucose pyrophosphorylase, a regulatory enzyme for bacterial glycogen synthesis. Microbiol. Mol. Biol. Rev. *67*, 213–225.

Ballicora, M. A., Dubay, J., Devillers, C., Preiss, J. (2005). Resurrecting the ancestral enzymatic role of a modulatory subunit. J. Biol. Chem. *280*, 10189–10195.

Ballicora, M. A., Erben, E. D., Yazaki, T., Bertolo, A. L., Demonte, A. M., Schmidt, J. R., Aleanzi, M., Bejar, C. M., Figueroa, C. M., Fusari, C. M., et al. (2007). Identification of regions critically affecting kinetics and allosteric regulation of the *Escherichia coli* ADP-glucose pyrophosphorylase by modeling and pentapeptide-scanning mutagenesis. J. Bacteriol. *189*, 5325–5333.

Bejar, C.M., Jin, X., Ballicora, M. A., Preiss,J. (2006). Molecular architecture of the glucose 1-phosphate site in ADP-glucose pyrophosphorylases. J. Biol. Chem. *281*, 40473–40484.

Blankenfeldt, W., Asuncion, M., Lam, J. S. Naismith, J. H. (2000). The structural basis of the catalytic mechanism and regulation of glucose-1-phosphate thymidylyltransferase (RmlA). EMBO J. *19*, 6652–6663.

Brito, J. A., Borges, N., Vonrhein, C., Santos, H., Archer, M. (2011). Crystal structure of *Archaeoglobus fulgidus* CTP:inositol-1-phosphate cytidylyltransferase, a key enzyme for di-myo-inositol-phosphate synthesis in (hyper)thermophiles. J. Bacteriol. *193*, 2177–2185.

Buschiazzo, A., Ugalde, J. E., Guerin, M. E., Shepard, W., Ugalde, R. A., and Alzari, P. M. (2004). Crystal structure of glycogen synthase: homologous enzymes catalyze glycogen synthesis and degradation. EMBO J. *23*, 3196–3205.

Crevillén, P., Ballicora, M. A., Mérida, A., Preiss, J., and Romero, J. M. (2003). The different large subunit isoforms of *Arabidopsis thaliana* ADP-glucose pyrophosphorylase confer distinct kinetic and regulatory properties to the heterotetrameric enzyme. J. Biol. Chem. *278*, 28508–28515.

Cupp-Vickery, J. R., Igarashi, R. Y., Perez, M., Poland, M., Meyer, C. R*.* (2008). Structural analysis of ADP-glucose pyrophosphorylase from the bacterium *Agrobacterium tumefaciens*. Biochemistry *47*, 4439-4451.

Feng, L., Fawaz, R., Hovde, S., Gilbert, L., Chiou, J., and Geiger, J. H. (2015). Crystal Structures of *Escherichia coli* Branching Enzyme in Complex with Linear Oligosaccharides. Biochemistry *54*, 6207–6218.

Führing, J., Cramer, J. T., Routier, F. H., Lamerz, A-C., Baruch, P., Gerardy-Schahn, R., Fedorov R. (2013). Catalytic Mechanism and Allosteric Regulation of UDP-Glucose Pyrophosphorylase from *Leishmania major*. ACS Catalysis *3*, 2976–2985.

Gardiol, A., and Preiss, J. (1990). *Escherichia coli* E-39 ADP-glucose synthetase has different activation kinetics from the wild-type allosteric enzyme. Arch. Biochem. Biophys. *280*, 175–180.

Gentner, N., and Preiss, J. (1968). Biosynthesis of bacterial glycogen. VI. Differences in the kinetic properties of the *Escherichia coli* B adenosine diphosphate glucose pyrophosphorylase depending on whether Mg++ or Mn++ serves as divalent cation. J. Biol. Chem. *243*, 5882–5891.

Georgelis, N., Braun, E. L., Shaw, J. R., Hannah, L. C. (2007). The two AGPase subunits evolve at different rates in angiosperms, yet they are equally sensitive to activity-altering amino acid changes when expressed in bacteria. Plant Cell, *19*, 1458–1472.

Hill, M. A., Kaufmann, K., Otero, J., Preiss, J. (1991). Biosynthesis of bacterial glycogen. Mutagenesis of a catalytic site residue of ADP-glucose pyrophosphorylase from *Escherichia coli*. J. Biol Chem. *266*, 12455–12460.

Jin, X., Ballicora, M. A., Preiss, J., Geiger, J. H. (2005). Crystal structure of potato tuber ADP-glucose pyrophosphorylase. EMBO J. *24*, 694–704.

Lahti, R. (1983). Microbial inorganic pyrophosphatases. Microbiol. Rev. *47*, 169–178.

Leloir, L. F., Cardini, C. E. (1957). Biosynthesis of glycogen from uridine diphosphate glucose. J. Am. Chem. Soc. *79*, 6340–6341.

Parsons, T. F., and Preiss, J. (1978). Biosynthesis of bacterial glycogen. Incorporation of pyridoxal phosphate into the allosteric activator site and an ADP-glucose-protected pyridoxal phosphate binding site of *Escherichia coli* B ADP-glucose synthase. J. Biol. Chem. *253*, 6197–6202.

Paule, M. R., and Preiss, J. (1971). Biosynthesis of bacterial glycogen: X. The kinetic mechanism of adenosine diphosphoglucose pyrophosphorylase from *Rhodospirillum rubrum.* J. Biol. Chem. *246*, 4602–4609.

Pelissier, M. C., Lesley, S. A., Kuhn, P. Bourne Y. (2010). Structural insights into the catalytic mechanism of bacterial guanosine-diphospho-D-mannose pyrophosphorylase and its regulation by divalent ions. J. Biol. Chem. *285*, 27468–27476.

Petreikov, M., Eisenstein, M., Yeselson, Y., Preiss, J., and Schaffer, A. A. (2010). Characterization of the AGPase large subunit isoforms from tomato indicates that the recombinant L3 subunit is active as a monomer. Biochem J. *428*, 201–212.

Preiss, J., Shen, L., Greenberg, E., Gentner, N. (1966). Biosynthesis of bacterial glycogen. IV. Activation and inhibition of the adenosine diphosphate glucose pyrophosphorylase of *Escherichia coli* B. Biochemistry *5*, 1833–1845.

Preiss, J. (1978). Regulation of adenosine diphosphate glucose pyrophosphorylase. Adv. Enzymol. Relat. Areas Mol. Biol. *46*, 317–381.

Preiss, J. Bacterial glycogen synthesis and its regulation. (1984). Annu. Rev. Microbiol. *38*, 419–58.

Preiss, J., and Romero, T. (1989). Physiology, biochemistry and genetics of bacterial glycogen synthesis. Adv Microb Physiol. *30*, 183–238.

Raetz, C. R. and Roderick, S. L. (1995). A left-handed parallel beta helix in the structure of UDP-N-acetylglucosamine acyltransferase. Science *270*, 997–1000.

Recondo, E., and Leloir, L. F. (1961). Adenosine diphosphate glucose and starch synthesis. Biochem. Biophys. Res. Commun. *6*, 85–88.

Roach P. J., Depaoli-Roach, A. A., Hurley, T. D., Tagliabracci, V. S. (2012). Glycogen and its metabolism: some new developments and old themes. Biochem. J. *441*, 763–787.

Stitt, M. Zeeman, S. C. (2012). Starch turnover: pathways, regulation and role in growth. Curr. Opin. Plant Biol. *15*, 282–292.

Tuncel, A. and Okita, T. W. (2013). Improving starch yield in cereals by over-expression of ADP glucose pyrophosphorylase: expectations and unanticipated outcomes. Plant Sci *211*, 52-60.

Ventriglia, T., Ballicora, M. A., Crevillén, P., Preiss, J., Romero, J. M. (2007). Regulatory properties of potato-*Arabidopsis* hybrid ADP-glucose pyrophosphorylase. Plant Cell Physiol. *48*, 875–880.

Vithani, N., Bais, V., Prakash, B. (2014). GlmU (N-acetylglucosamine-1-phosphate uridyltransferase) bound to three magnesium ions and ATP at the active site. Acta Crystallogr. F Struct. Biol. Commun. *70*, 703–708.

Wu, M. X., and Preiss, J. (2001). Truncated forms of the recombinant *Escherichia coli* ADP-glucose pyrophosphorylase: the importance of the N-terminal region for allosteric activation and inhibition. Arch. Biochem. Biophys. *389*, 159–165.

**FIGURE LEGENDS**

**Figure 1.** **The crystal structure of *Ec*AGPase.** *A.*The overall structure of an *Ec*AGPase protomer. Two views of an *Ec*AGPase protomer showing the GT-A like domain (green) and the LβH domain (orange).*B*. Four views of the *Ec*AGPase dimer. *C*. The structure of an *Ec*AGPase homotetramer. *D.* Surface representation of an *Ec*AGPase tetramer. The orientation is similar to that observed in panel *C*. See also Figure S1.

**Figure 2. The location of the regulatory sites in *Ec*AGPase.** *A.* Electrostatic surface representation of the *Ec*AGPase tetramer in complex with AMP (yellow spheres). *B.* Electrostatic suface representation showing a close view of the AMP binding site. *C.* Closed view of the AMP binding site, showing key interactions with selected residues. *D.* Electrostatic surface representation of the *Ec*AGPase tetramer in complex with FBP (orange spheres). *E.* Electrostatic surface representation showing a close view of the FBP binding site. *F.* Closed view of the FBP binding site, showing key interactions with selected residues. *G-I*. *Ec*AGPase thermal unfolding transitions recorded at 222 nm between 20 °C and 90 °C. Native fraction of *Ec*AGPase plotted versus temperature for the *apo* state (green), and *Ec*AGPase at different concentrations of AMP (panel *G* in a blue scale), or FBP (panel *H* in orange and red scale), or AMP and FBP (panel *I* in a purple scale). In panel *I* the curve for *Ec*AGPase•AMP (where concentration of AMP = 0.5 mM) is indicated for reference. The corresponding fitted two-states sigmoidal curves of the unfolding events are also shown. See also Figure S2.

**Figure 3. A regulatory mechanism for *Ec*AGPase.** *A.*Cartoon representation of the overall structure of the *Ec*AGPase tetramer, showing the location of the regulatory and active sites. AMP, FBP, ATP and G1P molecules are shown in one protomer of the *Ec*AGPase tetramer. The location of the ATP binding site in *Ec*AGPase was determined by structural superposition with (i) the crystal structure of GlmU in complex with ATP and that of *St*AGPase in complex with ATP and ADPG. The G1P binding site in *Ec*AGPase was determined taking into account the location of the glucose moiety of sucrose in the *Ec*AGPase•AMP•SUC crystal structure (Figure S3). The four protomers are depicted in four different colors, green, orange, blue and yellow. *B.* Structural superposition of the *Ec*AGPase•AMP•SUC and *Ec*AGPase•FBP, showing the partial overlapping of AMP and FBP binding sites*. C.*Cartoon representing the key structural elements involved in *Ec*AGPase allosteric regulation.Protomers A and C of the *Ec*AGPase•AMP•SUC complex are shown in green and orange, respectively. AMP and the superimposed FBP molecules are shown into the allosteric site. The SM motif is shown in red whereas the RBL1 and RBL2 loops are shown in cyan. ATP and G1P are shown in the active site. *D.* The SM motif in detail. See also Figure S3.

**Figure 4.** **Structure-weighted sequence alignment of *Ec*AGPase with other AGPases.** Structural alignment between the crystals structures of *Ec*AGPase (pdb code: 5L6V; Uniprot code: P0A6V1), *At*AGPase (3BRK; P39669) and *St*AGPAse (1YP3; P23509). The secondary structure elements corresponding to the GT-A like domain are shown in yellow (α-helices) and orange (β-helices); and to the LβH domain in green (α-helices) and blue (β-helices). Residues with poor electron density are highlighted as full boxes. The SM motif and the RL1 and RL2 loops are highlighted in yellow. Catalytic residues are highlighted as dotted boxes. The r.m.s.d. value is shown for each residue. Amino acid sequences of selected AGPases were aligned to the structure alignment: *Mycobacterium smegmatis* (class II; A0R2E1), *Serratia marcescens* (class II); A0A0U6P844), *Rhodobacter sphaeroides* (class V); Q9RNH7), *Rhodospirillium rubrum* (class VI); Q9ZFN4), *Bacillus subtilis* (class VII); P39122), *Synechococcus sp.* (class VIII; Q2JU94), *Ostreocuccus tauri* (clss VIII; Q6PYZ7), *Spinacia oleracea* (VIII; Q43152), *Triticum aestivum* (class IX; P30523). See also Figure S4.