Hypothesis

Glycogen synthase: towards a minimum catalytic unit?

Emili Cid[a,1], Roberto A. Geremia[b,1,*], Joan J. Guinovart[a], Juan C. Ferrer[a]

[a] *Departament de Bioquímica i Biologia Molecular, Universitat de Barcelona, Martí i Franquès 1, E-08028 Barcelona, Spain*
[b] *Plasticité et Expression des Génomes Microbiens, Université Joseph Fourier and Centre National de la Recherche Scientifique, FRE 2383, Bat. CERMO, BP 53X, 38041 Grenoble Cedex 9, France*

Received 25 July 2002; revised 16 August 2002; accepted 22 August 2002

First published online 2 September 2002

Edited by Robert B. Russell

Abstract Classically, α-1,4-glucan synthases have been divided into two families, animal/fungal glycogen synthases (GS) and bacterial/plant starch synthases (G(S)S), according to differences in sequence, sugar donor specificity and regulatory mechanisms. Detailed sequence analysis, predicted secondary structure comparison and threading analysis show that these two families are structurally related and that some domains of GSs were acquired to meet regulatory requirements. Archaeal G(S)S present structural and functional features that are conserved in one, the other or both families. Therefore, they are the link between GS and G(S)S and harbor the minimal sequence and structural features that constitute the minimum catalytic unit of the α-1,4-glucan synthase superfamily. © 2002 Published by Elsevier Science B.V. on behalf of the Federation of European Biochemical Societies.

Key words: Polysaccharide synthases; Glycosyltransferases; Glycogen; Starch; Evolution

1. Introduction

Almost all living organisms accumulate glucose as energy storage molecules, either in the form of glycogen or starch [1,2]. Glycogen refers to the polysaccharide localized in the cytoplasm, which is mainly used for bacterial [3], fungal [4] or animal glucose reserves [5]. Starch, which is the plant equivalent, is synthesized and stored inside plastids and is composed of two glucose polymers: amylopectin, which is the main component and is sufficient to form starch granules, and amylose [6]. Glycogen and amylopectin are ramified molecules formed by α-1,4-linked glucose units with α-1,6-branching points. The length and number of branches varies depending on the organism [2,5].

The first committed step in the biosynthesis of these polysaccharides, which is catalyzed by ADP- or UDP-glucose pyrophosphorylase, is the synthesis of the nucleotide-sugar ADP- or UDP-glucose from ATP or UTP and glucose α-1-phosphate. The α-1,4-glucan is synthesized by a single glycosyltransferase, which uses UDP-glucose or ADP-glucose as sugar donor, and a pre-existing α-1,4-glucan chain as acceptor. Based on sequence similarity, α-1,4-glucan synthases have been grouped in two families: glycogen synthases (GS, family 3 of glycosyltransferases) and glycogen (starch) synthases (G(S)S, family 5 of glycosyltransferases), which roughly correspond to the type of polysaccharide synthesized [7,8]. GS and G(S)S operate with retention of the configuration at the C-1 atom of the transferred glucose residue, but there is no significant overall sequence similarity between the two families.

The branches of mature polysaccharides are produced by the branching enzyme, which transfers a terminal α-1,4-oligosaccharide of the glucan chain to an internal glucose, forming an α-1,6-glycosidic bond. It should be noted that the structural variances between glycogen and starch are not due to intrinsic differences between GS and G(S)S activities. These glycosyltransferases are responsible only for the elongation step of glucan linear chain, and the final shape of the molecule depends on the action of other enzymes, specially the branching and debranching enzymes [2,5,6].

2. Yeast and mammalian GSs

The enzymes of this family not only share considerable sequence identity (45–50%), but also have two other key common features: (1) they use UDP-glucose as sugar donor, and (2) their activities are tightly regulated. Regulation occurs at two levels: inactivation by phosphorylation at serine/threonine sites and potent allosteric activation by glucose 6-phosphate (Glc 6-P), which can even overcome inhibition by phosphorylation [9]. The amino- and carboxy-termini are the most variable regions among the fungal/animal GS and harbor the regulatory phosphorylation sites [10].

Yeast and mammalian GS are the most studied enzymes of this group. There are two GS isoforms in *Saccharomyces cerevisiae*, Gsy1p and Gsy2p, the latter being the predominant and nutritionally regulated form [11]. These two isoenzymes present phosphorylation only at the carboxy-terminal region and their activities are controlled by extracellular conditions and cell cycle [12].

Mammalian glycogen metabolism is a paradigm of the regulation of enzyme activity [1]. Two GS isoforms, muscle [13] and liver [14], are present in mammals. They share complex regulation and high sequence similarity, and basically differ in their carboxy-termini. To date, nine serine residues have been shown to undergo phosphorylation in rabbit muscle GS [15–

*Correspondence author. Fax: (33)-476635663.
E-mail address: roberto.geremia@ujf-grenoble.fr (R.A. Geremia).

[1] The contributions from these authors to this work are equivalent.

Abbreviations: GS, glycogen synthase; G(S)S, glycogen starch synthase; *Hs*MGS, human muscle glycogen synthase; Glc 6-P, glucose 6-phosphate; GlcNAc, *N*-acetylglucosamine

0014-5793/02/$22.00
PII: S0014-5793(02)03313-6

17]. The phosphorylation of the enzyme occurs after glycogen loading and in fasting conditions. Allosteric and covalent regulations respond to metabolic, mainly glucose levels in blood, and hormonal signals, like insulin, glucagon and adrenaline [18–20].

In addition to the regulation of GS activity, glycogen synthesis may also be regulated through the control of the subcellular localization of the enzyme. In the absence of glucose, liver GS is homogeneously distributed in the cytoplasm and is directed to the hepatocyte periphery when glucose levels increase, and it is here that initial glycogen synthesis takes place [21,22]. In contrast, the muscle isoform is localized to the nucleus and translocates to the cytoplasm upon an increase in the concentration of glucose [23]. The molecular mechanisms that control these changes of localization have not yet been determined.

While several studies have focused on regulation, the catalytic mechanism of GS remains largely unknown [24], thereby precluding the understanding of the intramolecular arrangements that regulate catalytic activity. This lack of knowledge is mostly due to the inability to produce sufficient amounts of pure and functional protein for biochemical and structural studies [25] (J.C.F. and J.J.G., unpublished results). Comparison of GS with other retaining glycosyltransferases is an alternative to improve our understanding of the mechanism of action of this enzyme, but the identification of key residues is impaired by the high degree of identity within the members of family 3. Promising results have been obtained by comparison with other glycosyltransferases that act with retention of configuration at C-1 but present distinct substrate specificity and unrelated sequence features. A motif composed of two glutamic acid residues separated by seven amino acid residues (E-X_7-E), which is characteristic of family 4 of glycosyltransferases, is also conserved in GS [26]. This finding provided clues to predict some catalytic and structural features of GS. Although the exact function of the conserved Glu residues is not known, it has been suggested that they are involved in catalysis or recognition of the sugar donor [26]. In human muscle GS (*Hs*MGS), the invariant Glu residues correspond to Glu^{510} and Glu^{518}. Substitution of the former by Ala resulted in the inactivation of the enzyme, while replacement of the latter by Ala provided an enzyme with very low residual activity. The same 'biochemical phenotype' was found with AceA, a retaining mannosyltransferase from *Acetobacter xylinum* that belongs to family 4 of glycosyltransferases, thereby supporting the hypothesis that these residues play a similar role in the two families [27,28].

3. Eubacterial and plant G(S)S

G(S)S are involved in the synthesis of starch and bacterial glycogen. They present two remarkable differences with respect to the GS family: (1) they use ADP-glucose as sugar donor, and (2) neither inactivation by phosphorylation nor allosteric activation by Glc 6-P have been reported in this family. In fact, ADP-glucose availability is the triggering factor for starch biosynthesis, since ADP-glucose pyrophosphorylase is tightly regulated by allosteric effectors, both in plant and in bacterial enzymes [29]. At least for photosynthetic organisms, carbohydrate storage is not an essential requisite and starch could be considered a buffering pool for sugars and, therefore it is only synthesized when other requirements are fulfilled. Thus the use of ATP as nucleotide source instead of UTP has the theoretical advantage of coupling glycogen synthesis more tightly to the energetic state of the cell.

In most eubacteria the genes of the glycogenogenetic pathway, including that of G(S)S, are grouped in clusters [30–34]. Compared to bacterial G(S)S, the equivalent enzymes in plants form a more heterogeneous and complex group, which has acquired a higher specialization. This is mainly due to the requirements of a pluricellular organism, which usually implies the existence of several isoforms, and to the intraplastidial localization, which requires a targeting peptide.

Also in this case, most of the studies to date focus on the regulatory step, which is catalyzed by ADP-glucose pyrophosphorylase, rather than on catalysis of α-1,4-glucan elongation. Sequence analysis shows that G(S)Ss share a conserved region ('catalytic core'), which contains a variant of the motif E-X_7-E [26]. Interestingly, in this motif the Glu residue equivalent to Glu^{510} of *Hs*MGS (essential for activity) is strictly conserved, while that equivalent to Glu^{518} (not essential for activity) is not. Moreover, replacement of the first Glu residue of the motif (Glu^{391}) in maize starch synthase IIb-2 by an Ala residue leads to inactivation of the enzyme, indicating that the catalytic mechanisms of GS and G(S)S share some common elements [35].

4. Archaeal GSs

Systematic sequencing of archaea genomes has led to the finding of hypothetical proteins similar to G(S)S, by automatic annotation or by BLAST searches on the whole genome of *Pyrococcus abysii* (direct submission, Genoscope, 1999), *Pyrococcus furiosus* (unfinished genome sequence, Utah Genome Center), *Pyrococcus horikoshii* [36], *Methanococcus jannaschii* [37], *Sulfolobus acidocaldarius* [38], *Sulfolobus solfataricus* [39] and *Sulfolobus tokodaii* [40]. Indeed, glycogen is present in *Methanococcus methylutens* [41], and enzymes of glycogen metabolism are also detected in cell-free extracts of *Methanobacterium thermoautotrophicum* [42,43] and *Methanococcus maripaludis* [44].

In *P. abysii*, analysis of the G(S)S flanking regions has revealed the presence of three hypothetical proteins. However, none presents significant similarity to any known enzyme involved in glycogen metabolism. BLAST similarity searches on the whole genome predicted the presence of a glycogen phosphorylase in another region, but other genes usually associated to G(S)S in eubacterial glycogen operons were not identified. Identical results were obtained from the analysis of the genomes of *P. horikoshii* and *M. jannaschii*, which indicates that archaea have no eubacterial-like glycogen operons (E.C. and R.A.G., unpublished results).

Although the primary sequence of archaeal G(S)S is more similar to eubacterial G(S)S (30–40% similarity), which has led to their classification in the same glycosyltransferase family, we found that all hypothetical archaeal GSs share the 'canonical' E-X_7-E motif: both Glu residues are invariant and the intervening sequence is very similar to the family 3 consensus (Fig. 1). Furthermore, *S. acidocaldarius* G(S)S uses UDP-glucose as substrate [38] and *Thermococcus hydrothermalis* G(S)S can use either ADP- or UDP-glucose as sugar donors with similar apparent i_m values, although the V_{max} with ADP-glucose is approximately 10-fold higher [45]. We have also found that *P. abysii* G(S)S can use both nucleotide

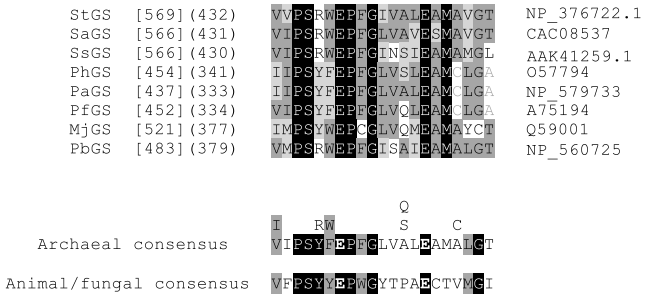


Fig. 1. Fragment of a multiple alignment of the predicted archaeal G(S)S sequences around the E-X$_7$-E motif. Alignments were obtained using CLUSTAL W locally [57]. The consensus sequence and the corresponding consensus of animal and fungal GS are also shown. The two conserved glutamic acid residues are shown in bold type in the consensus. The abbreviations used are: St, *S. tokodaii*; Sa, *S. acidocaldarius*; Ss, *S. solfataricus*; Ph, *P. horikoshii*; Pa, *P. abysii*; Pf, *P. furiosus*; Mj, *M. jannaschii* and Pb, *Pyrobaculum aerophilum*. The numbers between brackets show the total length and the numbers between parentheses indicate the first amino acid shown for each protein. Accession numbers are shown on the right.

sugars as substrates (C. Horcajada, E.C., J.C.F. and J.J.G., unpublished results). These two findings raise the exciting perspective that families 3 and 5 of glycosyltransferases may share common structural and catalytic features.

5. Predicted secondary structure features of α-1,4-glucan synthases

In addition to the E-X$_7$-E motif, the comparison of the predicted secondary structures of *P. abysii* G(S)S, *Escherichia coli* G(S)S and *Hs*MGS revealed the presence of five regions that exhibit similar features (Fig. 2).

Region I is located at the amino-terminus of the three proteins and the predictions suggest the presence of a β-strand followed by an α-helix and four additional β-strands. Interestingly, in the case of *Hs*MGS, the first β-strand and the α-helix are shorter, and the intervening sequence contains the two amino-terminal phosphorylation sites (2 and 2a) [17], which are absent in the other glycosyltransferases and also in yeast GS. This region might be an acquired feature of the mammalian enzyme, in order to meet regulatory requirements.

Region II is directly linked to region I in the two G(S)S and by an intervening peptide in *Hs*MGS. Another additional element of secondary structure, in this case a predicted α-helix that is absent in the other enzymes, is located between regions II and III in *Hs*MGS. In contrast to the two former regions, region III of *P. abysii* G(S)S presents a higher similarity to that of *Hs*MGS than to the corresponding zone of *E. coli* G(S)S.

Again an intervening peptide, which is absent in the bacterial G(S)S, connects regions III and IV in *Hs*MGS. This last region is very similar for the three enzymes. The peptide between regions IV and V shows a very similar predicted secondary structure for both G(S)S, but not for *Hs*MGS. Finally, region V constitutes the carboxy-terminus of the two G(S)S, but is followed in the mammalian GS by an additional non-structured segment that harbors the carboxy-terminal phosphorylation sites, again suggesting that this is an acquired feature.

Each of the three proteins analyzed in this study has an Arg-rich cluster in an equivalent position to region V. In mammalian and yeast GS, the residues in this cluster are involved in the intramolecular transduction of Glc 6-P regulation [46]. Although only the region of clustering and not the exact location of the Arg residues is conserved, this region may have already been present in a hypothetical ancestor protein and have been partially modified to fulfil the regulatory requirements of animal/fungal GS.

Other striking similarities are observed when comparing regions IV and V with the secondary structure of the glycosyltransferases T4βGT (not classified) [47], GtfB (family 1) [48] and MurG (family 28) [49] (only T4βGT is shown in Fig. 2). These proteins are functionally distant from GS and G(S)S. All are inverting enzymes and their specificities for

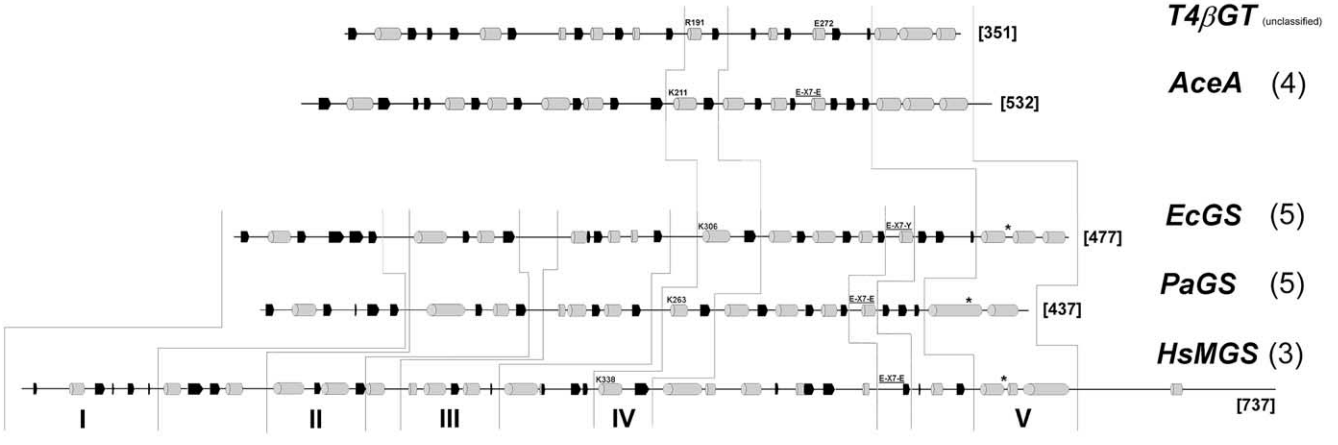


Fig. 2. Alignment of predicted secondary structures of *E. coli* (*Ec*) G(S)S, *P. abysii* (*Pa*) G(S)S and the muscle isoform of *Homo sapiens* GS (*Hs*MGS), the predicted secondary structure of *A. xylinum* mannosyltransferase (AceA) and the known secondary structure of the β-glycosyltransferase from bacteriophage T4 (T4βGT). The secondary structure prediction was obtained from the Jpred^2 server [58] and was performed on the basis of an automatic BLAST analysis [59,60], after removal of redundant sequences. The secondary structure prediction represents at least 99, 49, 8 and 80 sequence-related proteins for each enzyme, respectively. The position of the Lys residue in region IV analogous to Arg^{191} in T4βGT was confirmed by HCA plot analysis [61]. Gray barrels represent predicted α-helices and dark arrows β-sheets and for T4βGT only canonical α-helices and β-sheets are shown. The Arg-rich cluster in region V is marked with asterisks. The numbers between brackets show the number of amino acids of each protein and the numbers between parentheses indicate the glycosyltransferase family.

substrate are very diverse: T4βGT glucosylates DNA at 5-hydroxymethylcytosine residues [50], MurG transfers *N*-acetylglucosamine (GlcNAc) from UDP-GlcNAc to undecaprenol pyrophosphate-linked *N*-acetylmuramic acid [51] and GtfB uses UDP-glucose to glucosylate the vancomycin aglycone [52]. These enzymes, which have been crystallized and structurally solved, fold into two domains: the carboxy-terminal domain is involved in sugar donor binding and the amino-terminal domain is thought to participate in the binding of the acceptor. A conserved Lys residue in region IV of both G(S)S and *Hs*MGS is located just before or at the start of a predicted α-helix, in such a way that it resembles the Arg^{191} residue of T4βGT involved in binding the α-phosphate of the sugar donor [47] and the G-loop 3 of MurG, which has the same proposed function [49]. This Lys residue (Lys^{325}) and a nearby Arg (Arg^{319}) have been mutated to Ala in Gsy2p, and the resulting double mutant enzyme had only 0.2% of the activity of the wild-type protein [46].

The predicted secondary structures of region V of both G(S)S and *Hs*MGS are also very similar to the known secondary structure of the corresponding segment of the above-mentioned inverting glycosyltransferases. Interestingly, in T4βGT, MurG, and GtfB the third helix of this region interacts with the amino-terminal domain. Specifically, in MurG this helix makes contact with α1, β4, β5, β6 and β7 strands of the amino-terminal domain. If a similar arrangement were present in *Hs*MGS, it would bring together the amino- and the carboxy-terminal regulatory phosphorylation sites.

In addition to the similarities in the predicted secondary structure, a peptide resembling the E-X_7-E motif found in all the GS and G(S)S [26] is also present in T4βGT and in MurG. In both cases, the conserved amino acid is the second Glu residue. The role of this residue, established only for T4βGT (Glu^{272}), is to form an H-bond with the ribose of the sugar donor. The predicted secondary structure around the motif differs for *Hs*MGS, but this is due to the presence of a Pro just before the first Glu residue. This feature is also present in all archaeal G(S)S (Fig. 1). However, owing to the sequence similarity between archaeal enzymes and a larger number of glycosyltransferases, the program used for the prediction of secondary structure generates a consensus without this Pro and identical to the *E. coli* G(S)S sequence. Therefore, although the structural similarity of this zone between archaeal G(S)S and animal/fungal GS is not clearly shown by the prediction, it is indeed present, once again supporting our hypothesis.

It should be mentioned that the E-X_7-E motif and a high similarity of regions IV and V are also found in family 4 of glycosyltransferases, for which there is no structure available yet. The predicted secondary structure of AceA, as a representative of this family, is also included in Fig. 2.

Table 1
Threading analysis was performed using the PSSM algorithm [62]

Sequence	Hits	PSSM score
E. coli G(S)S	UDP-GlcNAc 2-epimerase	8.38 e^{-06}
	Proline Imino peptidase	3.05 e^{-01}
	GtfB	6.03 e^{-01}
P. abysii G(S)S	UDP-GlcNAc 2-epimerase	6.01 e^{-10}
	MurG	2.83 e^{-03}
	GtfB	4.07 e^{-03}
	T4βGT	1.91 e^{-01}
*Hs*MGS	Type I DNA topoisomerase	5.16 e^{-03}
	UDP-GlcNAc 2-epimerase	1.81 e^{-02}
	Haloperoxidase	9.34 e^{-01}
Δ*Hs*MGS	UDP-GlcNAc 2-epimerase	3.82 e^{-03}
AceA	UDP-GlcNAc 2-epimerase	1.72 e^{-07}
	MurG	3.06 e^{-03}
	T4βGT	5.83 e^{-02}
	GtfB	7.07 e^{-01}

The query sequences were obtained from the databases except for Δ*Hs*MGS. This sequence is an edited version of the *Hs*MGS sequence, in which the fragments composed by the amino acids 86–131, 285–334, 379–423, 617–737 were deleted (see text).

6. Threading analysis

Since the predicted secondary structure analysis supported the hypothesis that GS and G(S)S share structural features, we performed threading analysis using each of the three sequences studied as seed. The best and most significant scores for all the proteins, except *Hs*MGS, were obtained with the UDP-GlcNAc 2-epimerase (Table 1). The 3-D structure of this protein [53] has been shown to share structural homology with T4βGT and glycogen phosphorylase. The two best hits for *Hs*MGS were type I DNA topoisomerase and again UDP-GlcNAc 2-epimerase. The removal of the hypothetic intervening sequences in *Hs*MGS between regions I–II, III–IV, IV and the E-X_7-E motif, and the carboxy-terminus, found by secondary structure prediction, resulted in the increase of the score for UDP-GlcNAc 2-epimerase and the loss of similarity with type I DNA topoisomerase. This observation further reinforces the hypothesis that these intervening peptides were acquired during the course of evolution and may be involved in regulation of *Hs*MGS activity.

7. Phylogenetic analysis

The phylogenetic relationship between G(S)S and GS was analyzed by a CLUSTAL W alignment of representative enzymes from each family studied. The GS are clustered in a compact group while the G(S)S are basically distributed in groups that correspond to the organisms in which the enzymes are expressed: plants and algae, bacteria and archaea. The G(S)S from archaea are the closest group to the GS, which suggests that these enzymes are the link between the two glycosyltransferase families (Fig. 3). The sisterhood of the domains eukarya and archaea has been proposed in recent years and is supported by many gene trees [54].

8. Hypothesis

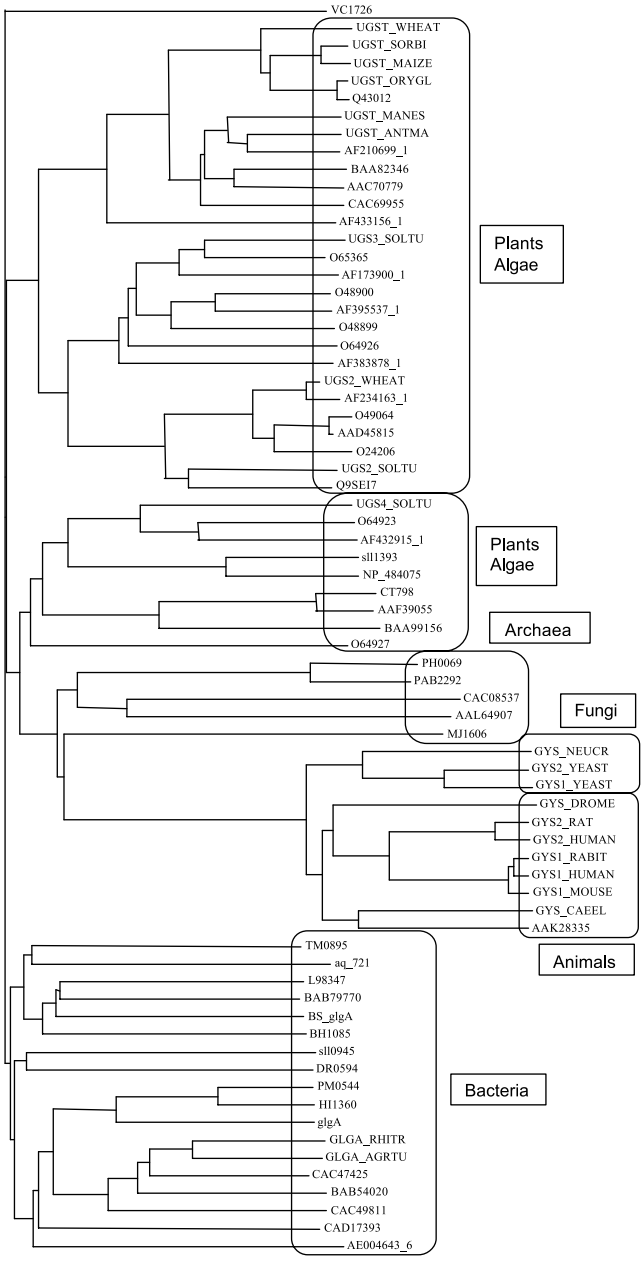
Our results, together with other information, indicate that GS and G(S)S belong to the same structural family and share a common catalytic mechanism. The main difference between the two families is the presence in the animal/fungal GS of additional peptides at both termini and intervening sequences between the regions I–II, III–IV, and IV–E-X_7-E motif. The results from threading analysis strongly suggest that these peptides are organized outside the 'catalytic unit'. Since they are absent from non-regulated GS, we propose that they are related with the intramolecular rearrangements involved in regulation by effectors. The similarity of the predicted secondary structure, the presence of conserved amino acids which directly participate in substrate binding or catalysis and the fact that all these enzymes catalyze the same biochemical re-

action strongly support the grouping of families 3 and 5 into a 'superfamily'. For many years starch synthases and bacterial GSs have been treated separately from mammalian or yeast GSs. Our results shed new light on the relations of these two families.

Family 4 of retaining glycosyltransferases is a heterogeneous group with respect to their function and comprises more than 300 proteins. Although they catalyze distinct glycosyl transfer reactions, they are also closely related to families 3 and 5. The predicted secondary structure of AceA is very similar to that of family 5 proteins (Fig. 2) and conserves many of the features mentioned above. While in families 3 and 5 the basic function (α-1,4-glucan synthesis) has been preserved, in family 4 the same overall structure has been used to deal with different substrates and reactions. It would be interesting to further study the structural relationships of the latter with families 3 and 5.

Taking into account these similarities, it is reasonable to assume that GS acquired some of its regulatory functions from non-regulated proteins, by using and rearranging pre-existing structural elements. Interestingly, some domains involved in the allosteric regulation of animal/fungal GS by Glc 6-P are also present in the non-regulated GS/G(S)Ss. This raises the attractive hypothesis that the domains involved in intramolecular signal transduction are permanently switched 'ON' in G(S)S, while they are 'OFF' in GS. In the latter an activation signal (binding of Glc 6-P) is necessary to switch 'ON' enzymatic activity. In accordance with this hypothesis, mutations of the Arg-rich cluster of the Gsy2p completely abolish the allosteric regulation and render an enzyme that is permanently active even in the absence of Glc 6-P [46].

The catalytic core of the glycogen phosphorylase is structurally similar to that of T4βGT [55,56], and both enzymes might have a common ancestor: an UDP-glucose-dependent



←

Fig. 3. Phylogenetic tree of G(S)Ss and GSs. From the local alignment of 71 sequences using CLUSTAL W, the phylogenetic tree was calculated using the neighbor joining method. Sequences are named by their databases accession numbers: VC1726: *Vibrio cholerae* G(S)S; UGST_WHEAT: *Triticum aestivum* granule-bound G(S)S; UGST_SORBI: *Sorghum bicolor* G(S)S; UGST_MAIZE: *Zea mays* granule-bound G(S)S; UGST_ORYGL: *Oryza glaberrima* granule-bound G(S)S; Q43012: *Oryza sativa* granule-bound G(S)S; UGST_MANES: *Manihot esculenta* granule-bound G(S)S; UGST_ANTMA: *Antirrhinum majus* granule-bound G(S)S; AF-210699_1: *Perilla frutescens* granule-bound G(S)S; BAA82346: *Phaseolus vulgaris* granule-bound G(S)S I; AAC70779: *Astragalus membranaceus* granule-bound G(S)S; CAC69955: *Pisum sativum* granule-bound G(S)S; AF433156_1: *Chlamydomonas reinhardtii* granule-bound G(S)S I; UGS3_SOLTU: *Solanum tuberosum* G(S)S; O65365: *Ipomoea batatas* G(S)S; AF1739900_1: *Manihot esculenta* granule-bound G(S)S II; O48900: *Zea mays* zSSII-2 G(S)S; AF395537_1: *Oryza sativa* soluble G(S)S II-2; O48899: *Zea mays* zSSII-1 G(S)S; O64926: *Chlamydomonas reinhardtii* soluble G(S)S; AF383878_1: *Oryza sativa* soluble G(S)S II-1; UGS2_WHEAT: *Triticum aestivum* soluble G(S)S; AF234163_1: *Hordeum vulgare* G(S)S I; O49064: *Zea mays* G(S)S I; AAD45815: *Sorghum bicolor* soluble G(S)S; O24206: *Oryza sativa* soluble G(S)S; UGS2_SOLTU: *Solanum tuberosum* soluble G(S)S; Q9SEI7: *Arabidopsis thaliana* soluble G(S)S; UGS4_SOLTU: *Solanum tuberosum* soluble G(S)S; O64293: *Zea mays* G(S)S DULL1; AF432915_1: *Oryza sativa* G(S)S III; sll1393: *Synechocystis* ssp. G(S)S; NP_484075: *Nostoc* sp. PCC 7120 G(S)S; CT798: *Chlamydia trachomatis* G(S)S; AAF39055: *Chlamydia muridarum* G(S)S; BAA99156: *Chlamydophila pneumoniae* J138 G(S)S; O64927: *Chlamydia reinhardtii* soluble G(S)S; PH0069: *P. horikoshii* G(S)S; PAB2292: *P. abysii* G(S)S; CAC08537: *S. acidocaldarius* G(S)S; AAL64907: *Pyrobaculum aerophilum* G(S)S; MJ1606: *M. jannaschii* G(S)S; GYS_NEUCR: *Neurospora crassa* GS; GYS2_YEAST: *S. cerevisiae* GS 2; GYS1_YEAST: *S. cerevisiae* GS 1; GYS_DROME: Drosophila melanogaster GS; GYS2_RAT: *Rattus norvegicus* liver GS; GYS2_HUMAN: *Homo sapiens* liver GS; GYS1_RABIT: *Oryctolagus cuniculus* muscle GS; GYS1_HUMAN: *Homo sapiens* muscle GS; GYS1_MOUSE: *Mus musculus* muscle GS; GYS_CAEEL: *Caenorhabditis elegans* GS; AAK28335: *Steinernema feltiae* GS; TM0895: *Thermotoga maritima* G(S)S; aq_721: *Aquifex aeolicus* G(S)S; L98347: *Lactococcus lactis* G(S)S; BAB79770: *Clostridium perfringens* G(S)S; BS_glgA: *Bacillus subtilis* G(S)S; BH1085: *Bacillus halodurans* G(S)S; sll0945: *Synechocystis* ssp. G(S)S; DR0594: *Deinococcus radiodurans* G(S)S; PM0544: *Pasteurella multocida* G(S)S; HI1360: *Haemophilus influenzae* G(S)S; glgA: *E. coli* G(S)S; GLGA_RHITR: *Rhizobium tropici* G(S)S; GLGA_AGRTU: *Agrobacterium tumefaciens* G(S)S; CAC47425: *Sinorhizobium meliloti* G(S)S; BAB54020: *Mesorhizobium loti* G(S)S; CAC49811: *Sinorhizobium meliloti* G(S)S (codified in the symbiotic plasmid pSymB); CAD17393: *Ralstonia solanacearum* G(S)S; AE004643_6: *Pseudomonas aeruginosa* G(S)S.

enzyme. In the evolution towards glycogen phosphorylase the ancestor acquired pyridoxal-phosphate-dependent catalysis and regulatory features, keeping the same catalytic core. Similarly, GS and G(S)S may have evolved from a common ancestor capable of forming α-1,4-glucans using both ADP- and UDP-glucose. At present, the products of divergent evolution are GS and G(S)S, which use two distinct strategies to regulate glycogen synthesis. On the one hand, the more complex GS acquired a molecular switch (Glc 6-P and phosphorylation) that allows direct control of the activity by critical metabolites. On the other hand, the bacterial/plant G(S)S are always active but adapted to the use of another sugar nucleotide, ADP-glucose, which is present only in conditions that are favorable for the production of glycogen. In the context of this hypothesis, it interesting to note that archaeal G(S)S can use either UDP- or ADP-glucose, which further points to them as being the link between the two families.

Without entering into a discussion about the phylogenetic origin of these proteins, it is reasonable to assume that archaeal G(S)Ss are the link between the GS and G(S)S families. They share catalytic properties of glycosyltransferase families 3 and 5 and some of their structural features are conserved in one, the other or both families. This leads us to propose that there is only one kind of structural core for the lineal α-1,4-glucan elongation and that *P. abysii* G(S)S, the smallest glycosyltransferase among this superfamily, represents this minimum catalytic unit.

Acknowledgements: This research was supported by Grant PB98-0992 from the Dirección General de Enseñanza Superior (Ministerio de Educación y Cultura, Spain) and by Grant ACI 99-16 from the Generalitat de Catalunya. E.C. was the recipient of a doctoral fellowship from the Generalitat de Catalunya (Comissió Interdepartamental de Recerca i Innovació Tecnològica). We thank Tanya Yates for assistance in preparing the English manuscript.

References

[1] Roach, P.J., Cheng, C., Huang, D., Lin, A., Mu, J., Skurat, A.V., Wilson, W. and Zhai, L. (1998) J. Basic Clin. Physiol. Pharmacol. 9, 139–151.
[2] Buleon, A., Colonna, P., Planchot, V. and Ball, S. (1998) Int. J. Biol. Macromol. 23, 85–112.
[3] Preiss, J. and Romeo, T. (1989) Adv. Microb. Physiol. 30, 183–238.
[4] Francois, J. and Parrou, J.L. (2001) FEMS Microbiol. Rev. 25, 125–145.
[5] Calder, P.C. (1991) Int. J. Biochem. 23, 1335–1352.
[6] Ball, S. et al. (1996) Cell 86, 349–352.
[7] Campbell, J.A., Davies, G.J., Bulone, V. and Henrissat, B. (1997) Biochem. J. 326, 929–939.
[8] Campbell, J.A., Davies, G.J., Bulone, V. and Henrissat, B. (1998) Biochem. J. 329, 719.
[9] Villar-Palasi, C. and Guinovart, J.J. (1997) FASEB J. 11, 544–558.
[10] Roach, P.J. et al. (1991) Adv. Enzyme Regul. 31, 101–120.
[11] Farkas, I., Hardy, T.A., Goebl, M.G. and Roach, P.J. (1991) J. Biol. Chem 266, 15602–15607.
[12] Hardy, T.A. and Roach, P.J. (1993) J. Biol. Chem. 268, 23799–23805.
[13] Zhang, W.M., Browner, M.F., Fletterick, R.J., DePaoli-Roach, A.A. and Roach, P.J. (1989) FASEB J. 3, 2532–2536.
[14] Bai, G., Zhang, Z.J., Werner, R., Nuttall, F.Q., Tan, A.W. and Lee, E.Y. (1990) J. Biol. Chem. 265, 7843–7848.
[15] Nakielny, S., Campbell, D.G. and Cohen, P. (1991) Eur. J. Biochem. 199, 713–722.
[16] Skurat, A.V. and Roach, P.J. (1995) J. Biol. Chem. 270, 12491–12497.
[17] Skurat, A.V., Wang, Y. and Roach, P.J. (1994) J. Biol. Chem. 269, 25534–25542.
[18] Ciudad, C., Camici, M., Ahmad, Z., Wang, Y., DePaoli-Roach, A.A. and Roach, P.J. (1984) Eur. J. Biochem. 142, 511–520.
[19] Lawrence Jr., J.C., Hiken, J.F., DePaoli-Roach, A.A. and Roach, P.J. (1983) J. Biol. Chem. 258, 10710–10719.
[20] Bollen, M., Keppens, S. and Stalmans, W. (1998) Biochem. J. 336, 19–31.
[21] Fernandez-Novell, J.M., Bellido, D., Vilaro, S. and Guinovart, J.J. (1997) Biochem. J 321, 227–231.
[22] Garcia-Rocha, M., Roca, A., De La Iglesia, N., Baba, O., Fernandez-Novell, J.M., Ferrer, J.C. and Guinovart, J.J. (2001) Biochem. J. 357, 17–24.
[23] Ferrer, J.C., Baque, S. and Guinovart, J.J. (1997) FEBS Lett. 415, 249–252.
[24] Mahrenholz, A.M., Wang, Y.H. and Roach, P.J. (1988) J. Biol. Chem. 263, 10561–10567.
[25] Zhang, W., DePaoli-Roach, A.A. and Roach, P.J. (1993) Arch. Biochem. Biophys. 304, 219–225.
[26] Cid, E., Gomis, R.R., Geremia, R.A., Guinovart, J.J. and Ferrer, J.C. (2000) J. Biol. Chem. 275, 33614–33621.
[27] Abdian, P.L., Lellouch, A.C., Gautier, C., Ielpi, L. and Geremia, R.A. (2000) J. Biol. Chem. 275, 40568–40575.
[28] Geremia, R.A., Petroni, E.A., Ielpi, L. and Henrissat, B. (1996) Biochem. J. 318, 133–138.
[29] Preiss, J. (1996) Biotechnol. Annu. Rev. 2, 259–279.
[30] Romeo, T., Kumar, A. and Preiss, J. (1988) Gene 70, 363–376.
[31] Kiel, J.A., Boels, J.M., Beldman, G. and Venema, G. (1994) Mol. Microbiol. 11, 203–218.
[32] Takata, H., Takaha, T., Okada, S., Takagi, M. and Imanaka, T. (1997) J. Bacteriol. 179, 4689–4698.
[33] Ugalde, J.E., Lepek, V., Uttaro, A., Estrella, J., Iglesias, A. and Ugalde, R.A. (1998) J. Bacteriol. 180, 6557–6564.
[34] Igarashi, R.Y. and Meyer, C.R. (2000) Arch. Biochem. Biophys. 376, 47–58.
[35] Nichols, D.J., Keeling, P.L., Spalding, M. and Guan, H. (2000) J. Biochem. 39, 7820–7825.
[36] Kawarabayasi, Y. et al. (1998) DNA Res. 5, 55–76.
[37] Bult, C.J. et al. (1996) Science 273, 1058–1073.
[38] Cardona, S., Remonsellez, F., Guiliani, N. and Jerez, C.A. (2001) Appl. Environ. Microbiol. 67, 4773–4780.
[39] She, Q. et al. (2001) Proc. Natl. Acad. Sci. USA 98, 7835–7840.
[40] Kawarabayasi, Y. et al. (2001) DNA Res. 8, 123–140.
[41] Maitra, P.K., Bhosale, S.B., Kshirsagar, D.C., Yeole, T.Y. and Shanbhag, A.N. (2001) FEMS Microbiol. Lett. 198, 23–29.
[42] Jansen, K., Stupperich, E. and Fuchs, G. (1982) Arch. Microbiol. 132, 355–364.
[43] Fuchs, G., Winter, H., Steiner, I. and Stupperich, E. (1983) Arch. Microbiol. 136, 160–162.
[44] Yu, J.P., Ladapo, J. and Whitman, W.B. (1994) J. Bacteriol. 176, 325–332.
[45] Gruyer, S., Legin, E., Bliard, C., Ball, S. and Duchiron, F. (2002) Curr. Microbiol. 44, 206–211.
[46] Pederson, B.A., Cheng, C., Wilson, W.A. and Roach, P.J. (2000) J. Biol. Chem. 275, 27753–27761.
[47] Morera, S., Lariviere, L., Kurzeck, J., Aschke-Sonnenborn, U., Freemont, P.S., Janin, J. and Ruger, W. (2001) J. Mol. Biol. 311, 569–577.
[48] Mulichak, A.M., Losey, H.C., Walsh, C.T. and Garavito, R.M. (2001) Structure 9, 547–557.
[49] Ha, S., Walker, D., Shi, Y. and Walker, S. (2000) Protein Sci. 9, 1045–1052.
[50] Tomaschewski, J., Gram, H., Crabb, J.W. and Ruger, W. (1985) Nucleic Acids Res. 13, 7551–7568.
[51] Crouvoisier, M., Mengin-Lecreulx, D. and van Heijenoort, J. (1999) FEBS Lett. 449, 289–292.
[52] Aoki, H., Shiroza, T., Hayakawa, M., Sato, S. and Kuramitsu, H.K. (1986) Infect. Immun. 53, 587–594.
[53] Campbell, R.E., Mosimann, S.C., Tanner, M.E. and Strynadka, N.C.J. (2000) Biochemistry 39, 14993–15001.
[54] Brown, J.R. and Doolittle, W.F. (1997) Microbiol. Mol. Biol. Rev. 61, 456–502.
[55] Holm, L. and Sander, C. (1995) EMBO J. 14, 1287–1293.
[56] Artymiuk, P.J., Rice, D.W., Poirrette, A.R. and Willett, P. (1995) Nat. Struct. Biol. 2, 117–120.

[57] Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) Nucleic Acids Res. 22, 4673–4680.
[58] Cuff, J.A. and Barton, G.J. (1999) Proteins 34, 508–519.
[59] Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) J. Mol. Biol. 215, 403–410.
[60] Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Nucleic Acids Res. 25, 3389–3402.
[61] Lemesle-Varloot, L., Henrissat, B., Gaboriaud, C., Bissery, V., Morgat, A. and Mornon, J.P. (1990) Biochimie 72, 555–574.
[62] Kelley, L.A., MacCallum, R.M. and Sternberg, M.J.E. (2000) J. Mol. Biol. 299, 499–520.