

**Mini-Review**

Unexpected and widespread connections between bacterial glycogen and trehalose metabolism

Govind Chandra, Keith F. Chater and Stephen Bornemann

Departments of Biological Chemistry and Molecular Microbiology, John Innes Centre, Norwich Research Park, Norwich NR4 7UH, UK

Glycogen, a large \(\alpha\)-glucan, is a ubiquitous energy storage molecule among bacteria, and its biosynthesis by the classical GlgC-GlgA pathway and its degradation have long been well understood – or so we thought. A second pathway of \(\alpha\)-glucan synthesis, the four-step GlgE pathway, was recently discovered in mycobacteria. It requires trehalose as a precursor, and has been genetically validated as a novel anti-tuberculosis drug target. The ability to convert glycogen into trehalose was already known, so the GlgE pathway provides a complementary way of cycling these two metabolites. As well as containing cytosolic storage glycogen, mycobacteria possess an outer capsule containing a glycogen-like \(\alpha\)-glucan that is implicated in immune system evasion, so the GlgE pathway might be linked to capsular \(\alpha\)-glucan biosynthesis. Another pathway (the Rv3032 pathway) for \(\alpha\)-glucan biosynthesis in mycobacteria generates a methylglucose lipopolysaccharide thought to be associated with fatty acid metabolism. A comparative genomic analysis was carried out to evaluate the occurrence and role of the classical pathway, the new GlgE pathway and the Rv3032 pathway across bacteria occupying very different ecological niches. The GlgE pathway is represented in 14 % of sequenced genomes from diverse bacteria (about half as common as the classical pathway), while the Rv3032 pathway is restricted with few exceptions to mycobacteria, and the GlgB branching enzyme, usually presumed to be associated with the classical pathway, correlates more strongly with the new GlgE pathway. The microbiological implications of recent discoveries in the light of the comparative genomic analysis are discussed.

**Introduction: the classical view of the metabolism and functions of glycogen and trehalose**

This article first reviews recent discoveries in mycobacteria that have led to a substantial expansion in our understanding of the interlinked ways in which the carbohydrate storage compounds glycogen and trehalose are metabolized. We go on to show that a new pathway is unexpectedly widespread among diverse bacteria, and that the classical glycogen biosynthetic pathway is not as universal as was thought.

Glycogen is a largely soluble \(\alpha\)-linked glucose polymer (or \(\alpha\)-glucan) with \(~90\%~\alpha\)-1,4-links in its backbone and \(~10\%~\alpha\)-1,6-linked branches (reviewed by Preiss, 2009). Glycogen can comprise hundreds of thousands of glucose units and is generally synthesized in bacteria when there is an excess of carbon over another nutrient that limits growth (reviewed by Preiss & Romeo, 1989). As a carbon/energy storage molecule, it can make up 60 % of cell dry weight; it enhances the survival of cells (e.g. *Escherichia coli*) and it transiently accumulates prior to the onset of sporulation (e.g. in *Bacillus cereus*) or the production of exopolysaccharides (e.g. in *Streptococcus mutans*, which contributes to the formation of dental caries). Historically, bacteria have been considered to synthesize glycogen using the classical GlgC-GlgA pathway (Fig. 1) (reviewed by Preiss, 2006). This involves generating an activated glucose nucleotide diphosphate from glucose 1-phosphate by the action of nucleotide diphosphoglucose pyrophosphorylase (GlgC), and its subsequent polymerization by glycogen synthase (GlgA), to generate the linear glucan. This is turned into glycogen by the branching enzyme (GlgB)-mediated transfer of non-reducing-end oligoglucans to the 6-position of residues within a chain to generate side-branches. Throughout prokaryotes, glycogen has always been considered to be degraded by the highly conserved enzyme glycogen phosphorylase (GlgP) together with the debranching enzyme (GlgX), to generate glucose 1-phosphate, which feeds readily into primary metabolism.

The other well-known abundant bacterial storage carbohydrate, the disaccharide trehalose, is both an energy...
store and a stress-protectant, helping bacteria to survive desiccation, cold and osmotic stress (reviewed by Arguelles, 2000). Classically, trehalose (α,α-1,1-linked diglucose) is synthesized in bacteria from glucose phosphate intermediates via trehalose 6-phosphate, using the GalU-OtsA-OtsB system (reviewed by Elbein et al., 2003). Trehalose can constitute more than 10% of cellular dry weight, and can be the major storage carbohydrate in specialized developmental states such as spores and the bacteroids central to the Rhizobium–legume symbiosis. In mycobacteria (and the related corynebacteria), trehalose has attracted particular attention because of its incorporation into cell wall mycolic acids, which are involved in pathogenesis and immune system evasion (reviewed by Takayama et al., 2005). The degradation of trehalose to glucose has been relatively little studied, in part because trehalases appear to be poorly conserved (Carroll et al., 2007).

The main theme of this minireview is the extent to which recent results have revealed more of the intimate and complex interplay of glycogen and trehalose metabolism in diverse bacteria. A first hint of such interplay was the discovery 15 years ago of the TreY-TreZ pathway in Sulfolobus acidocaldarius (Fig. 1; Maruta et al., 1996). In this pathway, TreY inverts the reducing-end glucosyl residue of α-1,4 glucan to form an α,α-1,1-linked non-reducing trehalosyl disaccharide end, which yields free trehalose when it is cleaved from the glucan chain by TreZ. This process may be aided by glycogen debranching enzyme, TreX (also known as GlgX). Also in the 1990s, another enzyme, trehalose synthase (TreS) from Pimelobacter sp. R48, was discovered that could convert simple α-1,4-linked diglucose (maltose) to trehalose (Fig. 1; Nishimoto et al., 1996).

**The Rv3032 route: the first non-classical pathway to bacterial α-glucan**

The methylglucose lipopolysaccharides of mycobacteria typically comprise ~20 glucose units with an α-1,4-linked backbone, 10% α-1,6 branching and extensive acylation (reviewed by Jackson & Brennan, 2009). Despite such molecules being described 45 years ago (Lee, 1966), their role is not yet certain. They appear to be involved in chaperoning or regulating fatty acid biosynthesis within the cytosol by forming complexes with the products of fatty acid synthases. Recent studies in Mycobacterium tuberculosis have shown that a paralogue of GlgA, Rv3032, has α-(1,4)-glucosyltransferase activity that is associated with the generation of these specialized glucans (Kaur et al., 2009b; Stadthagen et al., 2007). Preliminary experiments suggested that Rv3032, unlike bacterial GlgA, might use UDP-glucose as well as ADP-glucose as the donor (Stadthagen et al., 2007). Branches are predicted to be introduced into the glucan by the branching enzyme Rv3031 (Stadthagen et al., 2007).

**GlgE mediates a new non-classical route to bacterial α-glucan**

A third pathway to bacterial α-glucans has recently been discovered in *M. tuberculosis* (Kalscheuer et al., 2010) and
identified in part in *Mycobacterium smegmatis* (Elbein et al., 2010). Rather than using a monosaccharide nucleotide diphosphate, the novel maltosyltransferase GlgE uses a disaccharide phosphate, α-maltose 1-phosphate, as the building block to extend glucan chains. This is reflected in GlgE belonging to the GH13 family rather than the GT4 family to which Rv3032 and GlgA belong, as defined by the CAZy database (http://www.cazy.org/) (Cantarel et al., 2009). A GlgB enzyme introduces branches reminiscent of the classical pathway. Remarkably, maltose 1-phosphate is generated from trehalose, in two steps: TreS isomerizes trehalose into maltose (Pan et al., 2004); and a maltokinase Pep2 (also known as Mak1) phosphorylates the maltose using ATP (Jarling et al., 2004). Trehalose synthase TreS was originally assumed exclusively to generate trehalose from maltose in vivo (Elbein et al., 2003), but the discovery of the GlgE pathway has now shown that the reverse reaction is also physiologically important (Kalscheuer et al., 2010). While it is possible that the source of trehalose for the GlgE pathway could be via the *de novo* GlgA, capsular glucan and immune system evasion of the mutant’s persistence in mice, implying a link between the *α*-glucan pathways has evidently yet to be fully deconvoluted.

Unexpectedly, disruption of *glgE* in *M. tuberculosis* is lethal in vitro and in a mouse model (Kalscheuer et al., 2010). Lethality is not due to the absence of the pathway product but rather to the accumulation of the GlgE substrate, maltose 1-phosphate, to presumably toxic levels. Maltose 1-phosphate accumulation and lethality are also observed with a *glgB* mutant (Kalscheuer et al., 2010; Sambou et al., 2008). Thus, both GlgE and GlgB have been genetically validated as novel anti-tuberculosis drug targets with a novel killing mechanism, although the former is the more attractive therapeutic target because GlgB homologues exist in mammals.

**Comparative genomics indicates considerable importance for the non-classical pathways across all bacterial classes but not Archaea**

Before being biochemically defined in mycobacteria (Kalscheuer et al., 2010), the GlgE pathway genes were first shown to be present in *Streptomyces coelicolor* (Schneider et al., 2000), while the Rv3032 pathway appeared to be confined to mycobacteria (Stadthagen et al., 2007) and their close relatives such as *Nocardia* species (Pommier & Michel, 1986). This raises the question as to how widely the two non-classical pathways are distributed. For the purposes of this article, we therefore carried out comparative genomic analyses. We first made reciprocal BLAST searches of over 1000 complete bacterial and archaeal genomes using *M. tuberculosis* and *S. coelicolor* genes of *α*-glucan metabolism as the query (for full technical details see the supplementary material available with the online version of this paper). In order to avoid bias towards Gram-positive gene sequences, we included *E. coli* genes as queries whenever they were known to exist (e.g. the classical glycogen pathway genes).

Unexpectedly, *glgE* was present in 21% of all completely sequenced bacterial genomes (including plasmids, where present), though only in one archaeal genome. When the distribution of the four genes of the *glgE* pathway (*treS*, *pep2*, *glgE* and *glgB*) was assessed, it turned out that the whole *glgE* pathway was present in 14% of genomes analysed, while 36% of bacterial and 81% of archaeal genomes had none of the four genes (See Fig. 2 for absolute numbers of occurrences).

The organisms possessing the GlgE pathway mostly had large, GC-rich genomes and complex life styles, and included many that interact with plants. Those lacking these genes were often animal pathogens or had small genomes. The entire GlgE pathway was particularly well represented in actinomycetes, which include high-GC *Streptomyces*, *Mycobacterium* and *Corynebacterium* species (Supplementary Fig. S1), while low-GC Gram-positives, such as *Clostridium* and *Bacillus* species, tended to lack at least *pep2* and *glgE*. In contrast, most Gram-positives had
the classical glycogen pathway genes. All four GlgE pathway genes were found sporadically in Gram-negative bacteria, including several Alphaproteobacteria (e.g. in Rhizobium, Rhodopseudomonas, Rhodobacter, Agrobacterium and Methylobacterium species), many Betaproteobacteria (especially in Burkholderia and Bordetella species), few Gammaproteobacteria (with the notable exceptions of xanthomonads and pseudomonads) and very few Deltaproteobacteria (but notably in Myxococcus xanthus). Other Gram-negatives such as Epsilonproteobacteria and Cyanobacteria tended not to have all four GlgE pathway genes. Among Gram-negatives, most Gammaproteobacteria and some Alphaproteobacteria but few others possessed the classical glycogen pathway genes. The only archaeal species with all four GlgE pathway genes was Picrophilus torridus DSM 9790, in which they were clustered. ABLASTN analysis did not provide evidence for the horizontal transfer of the cluster from the kingdom Bacteria. Most archaeal species, including P. torridus, lacked the genes for the classical glycogen pathway.

The GlgA pathway (as defined by the glgA and glgC genes) was roughly twice as common as the GlgE pathway and, as expected, the most common α-glucan biosynthetic pathway (Fig. 3a). Reciprocal BLAST searches using Rv3032 alone as the query seemed to suggest that this gene was as common as glgA. However, carbohydrate-active enzymes can evolve diverse functions with remarkably little sequence diversification (Cantarel et al., 2009). In order to ascribe function to Rv3032 homologues more robustly, the co-occurrence of two additional genes thought to be required for the Rv3032 pathway was determined (Rv3030, a putative S-adenosylmethionine-dependent methyltransferase gene, and Rv3031, a putative branching enzyme gene) (Fig. 3a). This analysis confirmed that the Rv3032 pathway was restricted to some actinomycetes: mycobacteria, rhodococci, Nocardia farcinica IFM 10152 and Actinosynnema mirum DSM 43827. The three defining Rv3032 pathway genes were always clustered when they co-existed, and both other α-glucan biosynthetic pathways were also usually present. Thus methylglucose lipopolysaccharides appear indeed to have a specialized role that is confined to very few bacteria. Somewhat surprisingly, 60% of bacteria and archaea had none of the three α-glucan biosynthetic pathways.

**Clustering and fusion of GlgE pathway genes**

The pep2 and glgE genes rarely occurred in the absence of the other two genes of the glgE pathway (Fig. 2). Indeed, pep2 and treS were more frequently fused than separate, often leading to their underannotation as just one or the other. Thus, pep2 and glgE are rather specifically associated with the glgE pathway. In contrast, both treS and glgB often occurred, on their own or together, without the other two
Fig. 3. Venn diagrams correlating the co-occurrence of genes encoding glucan biosynthetic and degradative enzymes. All values are expressed as a percentage of 1045 genomes analysed. (a) Genes defining the GlgA, GlgE and Rv3032 pathways (plasmids were devoid of these glucan pathway genes except for six occurrences of the GlgE pathway genes that amount to only 0.6% of genomes; see Fig. 2); (b) glgE, glgA and glgB; (c) glgE, glgA and glgB (i.e. those in panel b) co-occurring with Rv3031; (d) GlgA, GlgE and Rv3032 pathway genes (i.e. those in panel a) co-occurring with glgP; and (e) GlgA, GlgE and Rv3032 pathway genes (i.e. those in panel a) co-occurring with pgmA. 'None' refers to the number of genomes that have none of the relevant genes. Numbers next to a set refer to the total percentage associated with that set. See supplementary material for details.
genes. This may reflect roles in other contexts, such as the conversion of maltose to trehalose, and the GlgB-mediated branching of glucan generated by the classical pathway.

The clustering of the glgE pathway genes was assessed in each genome. Where all four functions were present, four alternative arrangements were found. If treS and pep2 were fused, the genes always formed a single operon-like cluster (Fig. 4a). This was the most common arrangement. When all four genes were present without a fusion they were equally likely to be in any one of three arrangements: fully clustered (Fig. 4b); in well-separated glgE-glgB and treS-pep2 pairs (Fig. 4c); or substantially unclustered. When one of the functions was absent and treS and pep2 were fused, the cluster was broken up about one-third of the time. The genes were never clustered when only three genes were present with no fusion. Overall, the tendency of the genes to cluster, either altogether or as two pairs, implies the existence of operons allowing their co-ordinated expression. Genetic evidence of an operon structure was provided for the two copies of the cluster in *S. coelicolor* (Schneider et al., 2000).

**Frequent association of glgB, glgP and both trehalose biosynthetic pathway genes with the GlgE pathway genes**

The GlgE, GlgA and Rv3032 α-glucan pathways are all completed by the action of a branching enzyme. To date, GlgB has always been assumed to be associated with the GlgA pathway, because of the proximity of their genes in the well-characterized glgPACXB cluster in *E. coli*. However, glgB was surprisingly absent from 30% of the genomes that possessed glgA (Fig. 3b). Furthermore, glgB was separated (by ≥10 genes) from glgA in 51% of cases where they coexist, making the clustering of the *E. coli* glg genes almost an exception rather than a rule. Unexpectedly, glgB was present in most (89%) glgE-containing genomes, and was usually clustered with glgE (in 80% of cases where they coexist). Only 8% of genomes had two copies of glgB according to reciprocal BLAST searches, reducing to 3% if gene annotations were also taken into account. Thus, it appears to be a rarity for the GlgA and GlgE pathways each to be associated with a dedicated glgB gene within a single organism.

The Rv3031 gene, predicted to code for the branching enzyme of the Rv3032 pathway (Jackson & Brennan, 2009), is sufficiently distinct from glgB that reciprocal BLAST searches do not confuse one with the other. This is a reflection of Rv3031 being a member of the GH57 family rather than the GH13 family that GlgB belongs to, with differences at the structural (Palomo et al., 2011) as well as primary sequence levels (Murakami et al., 2006). This raises the question as to whether a GH57 branching enzyme fulfils the role of GlgB in some organisms. Reciprocal BLAST searches using Rv3031 as the query revealed that 11% of genomes possessed a candidate gene for a GH57 branching enzyme, consistent with searches reported by others (Murakami et al., 2006). In 2% of genomes, the gene was clustered with Rv3030 and Rv3032 (Fig. 3a). However, the presence of a GH57 gene does not correlate with a lack of a glgB gene (Fig. 3c), so 25% of genomes that possess glgA are devoid of any known branching enzyme. This could reflect the operation of an unknown kind of branching enzyme, or the production of unbranched glucans. The latter is supported by the transient accumulation of granulose, an α,1,4-linked glucan with few or no branches, prior to spore formation in *Clostridium acetobutylicum* (e.g. Reysenbach et al., 1986). A putative glgB gene was also present in 3% of genomes that had none of the three glycosyltransferases, implying still undefined glucan pathways, other roles for GlgB, or the presence of orphaned genes.

The GlgP-mediated regeneration of glucose 1-phosphate could allow glucan cycling or provide activated glucose for other purposes. The glgP gene, next to a glgE-containing cluster in streptomycetes and the glgA-containing cluster in *E. coli*, was almost equally correlated with glgA and glgE (Fig. 3d). The presence of glgP in the absence of any of the α-glucan pathway genes implies either a role in catabolizing α-glucans from the environment or the existence of as-yet-undefined α-glucan pathways.

There are two potential sources of trehalose for the GlgE pathway: either *de novo* via GalU-OtsA-OtsB or through recycling of cytosolic α-glucan via (TreX)-TreY-TreZ. Comparative genomics of a representative gene for each source, otsA and treY, suggested that both sources coexist in most cases when glgE is present (data not shown).

**Developmental glycogen/trehalose cycling in streptomycetes**

The developmentally complex soil bacterium *S. coelicolor* experiences a very different environment to *M. tuberculosis*. The GlgE pathway genes are duplicated and separately developmentally regulated in *S. coelicolor*, being associated respectively with transient deposition of glycogen at the
initiation of aerial growth (phase I) and during the first stages of sporulation (phase II) (Plaskitt & Chater, 1995; Schneider et al., 2000; Yeo & Chater, 2005). Trehalose accumulates to high levels in spores at the expense of phase II glycogen deposits (Rueda et al., 2001), contributing significantly to resistance to heat and desiccation (McBride & Ensign, 1987), and is degraded during spore germination through the activation of a spore-associated trehalase proenzyme (McBride & Ensign, 1990). The discovery of the function of the GlgE pathway genes now provides an attractive hypothesis for the main metabolic processes underlying these observations. It is possible that the GlgE pathway is responsible, at least in part (Martin et al., 1997), for the formation of the glycogen deposits in both phases, and that recycling could involve either glucose phosphate or disaccharide intermediates.

**The GlgE pathway and non-**Mycobacterium** pathogens**

Although present in *M. tuberculosis* and *Pseudomonas, Burkholderia* and *Xanthomonas* animal and plant pathogens, the GlgE pathway is absent from most other pathogens. Nevertheless, experimental deletion of GlgE pathway genes and other loci associated with trehalose metabolism lowered *Pseudomonas syringae* fitness on host and non-host plant leaves (Freeman et al., 2010). Whether this was due to reduced virulence or increased susceptibility to hyperosmotic stress is not known. Interestingly, *glgE* appears in the list of genes that have undergone positive selection in the evolution of accidental virulence of *Burkholderia pseudomallei* (Nandi et al., 2010). The role of the GlgE pathway in pathogens requires further study.

**Some unexpected anomalies...**

The activity of phosphoglucomutase PgmA is ostensibly ultimately required to provide glucose 1-phosphate for all routes to the biosynthesis of both α-glucan and trehalose. The absence of a pgmA gene from ~11% of genomes that possess other genes needed for α-glucan synthesis is therefore problematical (Fig. 3e), and raises the possibility of either a different way of generating glucose 1-phosphate, or a novel form of phosphoglucomutase that is unrelated to PgmA. By contrast, the presence of pgmA in the absence of α-glucan and trehalose pathway genes is to be expected, given the requirement of many metabolites to be glucosidated.

**...and some immediate questions**

Additional questions are raised by the existence of three α-glucan pathways, often with two or even all three within the same organism. Which of the pathways are primarily responsible for which of the three α-glucan products and how much cross-talk is there between the pathways in different contexts? What is the nature of the α-glucan produced by the GlgE pathway and how does it compare with cytosolic glycogen and capsular glucan? The comparative genomic analysis supports the notion that the Rv3032 pathway is relatively specific for methylglucose lipopolysaccharide biosynthesis in a limited set of organisms, but evidence for redundancy with the GlgE pathway (Kalscheuer et al., 2010) suggests that one should not view the pathways independently. Could Rv3031 and Rv3032 make a glycogen-like molecule? Furthermore, what is the role of the many Rv3032 glycosyltransferase homologues when not associated with methylglucose lipopolysaccharide biosynthesis? Since GlgB branching enzyme is more often associated with the GlgE pathway than with the classical pathway, are branching enzymes shared between pathways, or is another branching enzyme yet to be discovered? Evidently there is a great deal more to discover about the biology of α-glucan metabolism in bacteria, providing new opportunities to tackle some important human and plant pathogens.

**Acknowledgements**

This work was supported by the United Kingdom Biotechnology and Biological Sciences Research Council through the MET Institute Strategic Programme Grant to the John Innes Centre. We thank Rainer Kalscheuer, William R. Jacobs, Jr and Alison M. Smith for their comments on the manuscript.

**References**


