The EMBO Journal vol.14 no.7 pp.1287–1293, 1995

Evolutionary link between glycogen phosphorylase and a DNA modifying enzyme

Liisa Holm and Chris Sander

EMBL, D-69012 Heidelberg, Germany

Communicated by I.Mattaj

We report here an unexpected similarity in three-dimensional structure between glucosyltransferases involved in very different biochemical pathways, with interesting evolutionary and functional implications. One is the DNA modifying enzyme β-glucosyltransferase from bacteriophage T4, alias UDP-glucose:5-hydroxymethyl-cytosine β-glucosyltransferase. The other is the metabolic enzyme glycogen phosphorylase, alias 1,4-α-D-glucan:orthophosphate α-glucosyltransferase. Structural alignment revealed that the entire structure of β-glucosyltransferase is topographically equivalent to the catalytic core of the much larger glycogen phosphorylase. The match includes two domains in similar relative orientation and connecting helices, with a positional root-mean-square deviation of only 3.4 Å for 256 C^{α} atoms. An interdomain rotation seen in the R- to T-state transition of glycogen phosphorylase is similar to that observed in β-glucosyltransferase on substrate binding. Although not a single functional residue is identical, there are striking similarities in the spatial arrangement and in the chemical nature of the substrates. The functional analogies are (β-glucosyltransferase–glycogen phosphorylase): ribose ring of UDP–pyridoxal ring of pyridoxal phosphate co-enzyme; phosphates of UDP–phosphate of co-enzyme and reactive orthophosphate; glucose unit transferred to DNA–terminal glucose unit extracted from glycogen. We anticipate the discovery of additional structurally conserved members of the emerging glucosyltransferase superfamily derived from a common ancient evolutionary ancestor of the two enzymes.

Key words: carbohydrate metabolism/DNA modification/protein evolution/structural alignment/T4 phage

Introduction

New enzyme functions may easily evolve as a result of tinkering with existing proteins. Structurally conserved superfamilies comprising muconate lactonizing enzyme/mandelate racemase (Neidhart *et al.*, 1990), at least three enzymes of the tryptophan biosynthesis pathway (Wilmanns *et al.*, 1990), or the amido-, imido- and amidinohydrolytic enzymes methionine aminopeptidase/creatinase/prolidase (Murzin, 1993) are just a few examples of the principle. With the structures of hundreds of different proteins currently available, more or less systematic searches for structural similarities frequently uncover remote evolutionary relationships in spite of mutational noise that makes detection by sequence comparison difficult (review: Holm and Sander, 1994a). Here, we describe a particularly fascinating case of comparative structural biochemistry that sheds new light on the recently solved structure of T4 β-glucosyltransferase (Vrielink *et al.*, 1994).

β-Glucosyltransferase (EC 2.4.1.27) is involved in the DNA modification process that protects the phage genome against its own nucleases and the host restriction endonuclease system. β-Glucosyltransferase transfers a β-D-glucosyl residue from UDP-glucose to hydroxymethylcytosine residues in double-stranded DNA. The genes for three enzymes catalyzing DNA glucosylation in T-even phages have been sequenced. Sequence comparisons show no obvious similarity between the three glucosyltransferases, suggesting that they may have different evolutionary roots, nor are any other sequence relatives known (Tomaschewski *et al.*, 1985). Topological comparison of β-glucosyltransferase with DNA methyltransferases of known structure showed only limited similarity (Vrielink *et al.*, 1994). Surprisingly, our database scan using the Cα coordinates of β-glucosyltransferase (Protein Data Bank datasets 1BGT and 1BGU) revealed an overwhelming similarity to glycogen phosphorylase, a classic allosterically controlled enzyme in carbohydrate metabolism.

Glycogen phosphorylase (1,4-α-D-glucan:orthophosphate α-glucosyltransferase, EC 2.4.1.1) catalyzes *in vivo* the phosphorolytic degradation of glycogen, generating glucose-1-phosphate that may enter the glycolytic cycle. The amino acid sequence is well conserved from rabbit to yeast to bacteria (>40% residue identity) although enzymes from different sources differ in their regulatory mechanisms and in their natural substrates which include glycogen, starch and dextrans (see, e.g. Fukui *et al.*, 1982; Hudson *et al.*, 1993; Kiel *et al.*, 1994). The mammalian enzyme's activity is controlled both by phosphorylation and by non-covalent binding of allosteric effectors (AMP, ATP, ADP, glucose-6-phosphate) which induce changes mainly in inter-subunit association. All α-D-glucan phosphorylases are dependent for their activity on a covalently bound pyridoxalphosphate (PLP) co-enzyme. The phosphorylase from rabbit muscle has been extensively studied by crystallography (e.g. Johnson *et al.*, 1990; Acharya *et al.*, 1991; Sprang *et al.*, 1991, 1992; Duke *et al.*, 1994) and was the largest 3-D enzyme structure when it was first solved. The structure of the yeast enzyme has recently been solved (Rath and Fletterick, 1994).

Here, we first describe the remarkable common architecture of β-glucosyltransferase and glycogen phosphorylase. We then analyze differences in the active site pocket that correspond to differences in enzyme specificity. Finally, the observations are rationalized in an evolutionary model.

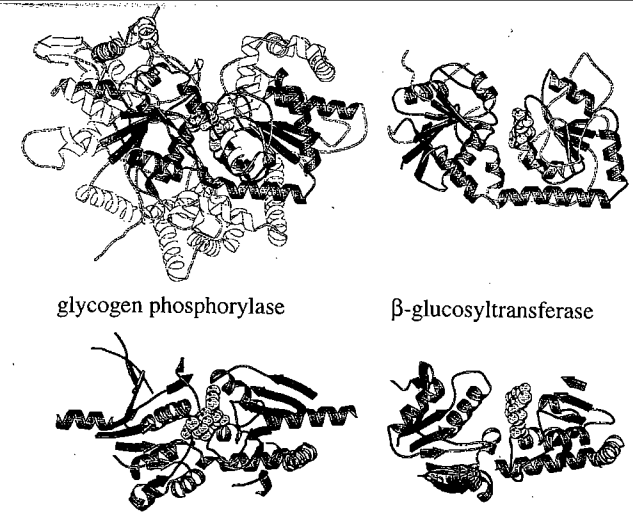

Fig. 1. Three-dimensional structural similarity of β-glucosyltransferase (right; Vrielink *et al.*, 1994) and glycogen phosphorylase (left; Acharya *et al.*, 1991) becomes apparent after 3-D alignment and presentation in equivalent orientations (domain 1, left; domain 2, right; active site, center). The top view is from the back of the active site with the hinge helix α12 (β-glucosyltransferase)/α28 (glycogen phosphorylase) horizontally in front. The lower view is rotated by 90°, with only the common parts shown. Structurally equivalent helices and strands (broad ribbons) are shaded grey. The common core comprises the catalytic core of glycogen phosphorylase and almost the entire structure of β-glucosyltransferase. The active sites are in the cleft between the two domains. The crystallographically located PLP co-enzyme and a glucose analog (heptulose-2-phosphate) in glycogen phosphorylase and UDP in β-glucosyltransferase are in the center of the molecule, shown as space filling. Structurally non-equivalent regions are white; many of them carry regulatory sites in glycogen phosphorylase. Plotted using Molscript (Kraulis, 1991).

Results

Conservation of structural core

3-D superimposition of glycogen phosphorylase and β-glucosyltransferase. Figure 1 shows how the fold of β-glucosyltransferase is completely embedded in the structure of glycogen phosphorylase. All nine α-helices and 13 β-strands of β-glucosyltransferase match elements of glycogen phosphorylase in sequential order (Figure 2). 256 pairs of Cα atoms of glycogen phosphorylase and β-glucosyltransferase can be superimposed with a root-mean-square (r.m.s.) positional deviation of 3.4 Å. The common core comprises two domains with remarkably similar relative orientations in the two structures. The quasi-rigid unit of the β-sheets at the center of each domain plus the connecting helices (α10−α13 in β-glucosyltransferase) stands out as particularly well conserved (Figure 3). The similarity score optimized by the Dali program (Holm and Sander, 1993) is 16 s.d. above the mean of the database background when normalized for protein size. Such high scores are usually obtained only between (remote) homologs (cf. Table I in Holm and Sander, 1994b).

The significance of the match is further supported by the fact that the first and second domains of the common core are architecturally distinct despite topographical similarity to the classical nucleotide binding fold. Glycogen phosphorylase ranks first in the database search also if the two domains of β-glucosyltransferase are separately used as the search structure: the similarity scores of the

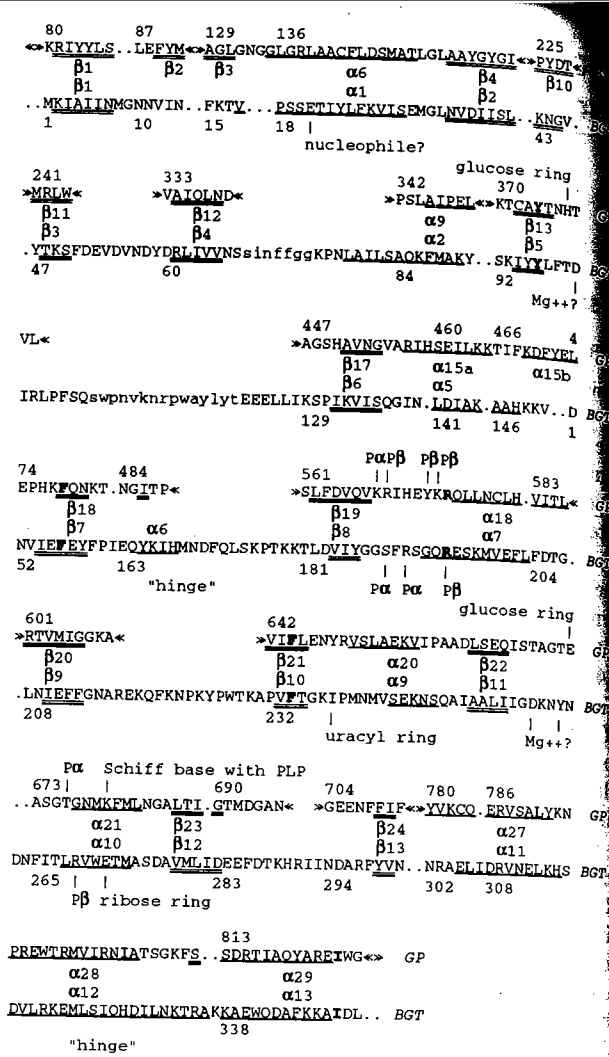

Fig. 2. Structural alignment presented in linear form [glycogen phosphorylase (GP) upper sequence; β-glucosyltransferase (BGT) lower sequence] highlights equivalent secondary structures and remarkable coincidence of functional sites in spite of the absence of sequence similarity. The full sequence of β-glucosyltransferase is shown (lower case residues were invisible in the crystal structure) but only the structurally equivalent residues of glycogen phosphorylase (≪≫ indicate segments not shown). Helix and strand elements (by the program DSSP, Kabsch and Sander, 1983) are named as in the crystallographic reports (α3, a4, and α8 in β-glucosyltransferase were not identified as α-helices by DSSP). Crystallographically determined binding sites (Vrielink *et al.*, 1994; Hwang and Fletterick, 1986) are annotated and indicated by vertical bars (? for inferred sites). For glycogen phosphorylase, the phosphate of PLP is denoted Pα and the orthophosphate Pβ. Bold type residues are invariant in glycogen phosphorylases from diverse sources and shared by β-glucosyltransferase.

two best hits, using the first domain, are 739 for glycogen phosphorylase (1gpb) versus 537 for the N-terminal domain of elongation factor Tu (1eft) and, using the second domain, 827 for glycogen phosphorylase versus 525 for glyceraldehyde-3-phosphate dehydrogenase (1gd1O). Comparison of the first domain of β-glucosyltransferase with the second domain in the same protein yields a similarity score of only .363. Taken together, these facts indicate that the two-domain core already existed as a structural entity in the common ancestor of β-glucosyltransferase and glycogen phosphorylase.

Fig. 3. Three-dimensional superimposition (stereo view) of glycogen phosphorylase (dashed, common core only; Acharya *et al.*, 1991) and β-glucosyltransferase (Vrielink *et al.*, 1994) shows near-perfect matching of the two β-sheets and of helices α10 (bold vertical helix in the centre), α11–α12 (bold helices at right angles in bottom right corner) and α13 (bold helix at left) of β-glucosyltransferase with equivalent elements in glycogen phosphorylase. Plotted using WhatIf (Vriend, 1990).

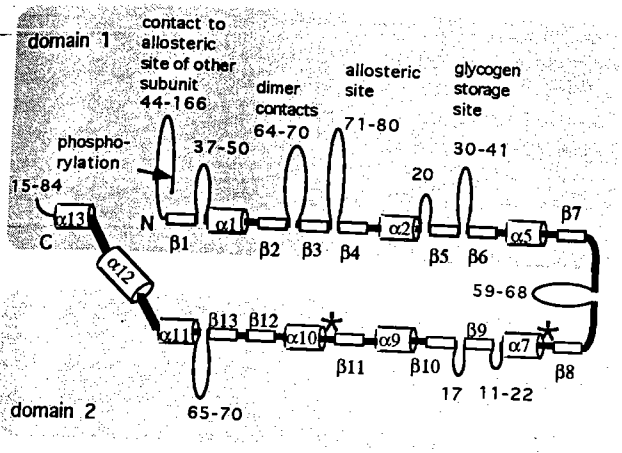

Fig. 4. The glycogen phosphorylase chain makes major excursions (shown as loops) relative to the common catalytic core (shown as linear sequence of helices and strands). Elements of the common core are shown as rectangles for strands and as cylinders for helices and are numbered as in β-glucosyltransferase (α1, β1, ..., cf. Figure 2) . The lengths of the excursions vary in the glycogen phosphorylase family, as indicated: e.g. '44–166' labels an insert of length 44 residues up to 166 residues. Excursions are also labelled by their approximate function in muscle glycogen phosphorylase. Note the absence of major insertions in the functionally important two main phosphate binding loops (*), in the α1–β2 unit which is close to the glucose binding site, and in the C-terminal hinge region. Both domains contain a parallel β-sheet, with strand order 3-2-1-- 4-5-6-7 in domain 1 and strand order 13-12-11-- 8-9-10 in domain 2 ('switch' topography).

Supporting evidence from the domain structure of glycogen phosphorylase. Glycogen phosphorylases are some of the largest known single chain enzymes (790–992 residues), and are more than twice as long as β-glucosyltransferase (351 residues). The structural alignment (Figure 2) is unusual for a pair of such unequal sizes in that the structurally equivalent segments are distributed over the entire chains of both proteins, except for the first 80 and last 16 residues of (rabbit) glycogen phosphorylase (Figure 4). The structural background to this lies in the domain architecture of glycogen phosphorylase. An automatic domain decomposition (Holm and Sander, 1994c) of glycogen phosphorylase yields compact folding units corresponding to the two domains of the catalytic core (domain 1: residues 62–90, 130–165, 190–485 and 812–841; domain 2: 166–177, 540–711 and 780–811) and three large excursions (domain 3: 14–61; domain 4: 486–539; domain 5: 712–779), in general agreement with the crystallographers' description of the structure (Acharya *et al.*, 1991). The domain decomposition is confirmed by the structural homology with β-glucosyltransferase, which quite precisely maps to domains 1 and 2 of glycogen phosphorylase. Several of the chain excursions in glycogen phosphorylase carry allosteric control sites (Figure 4). The excursions (white in Figure 1) surround the entrance to the catalytic site whereas the 'back side' of the enzyme is bare and similar to that of β-glucosyltransferase.

Conserved interdomain connection and motion. The chain crosses twice between the two domains of the catalytic core in both enzymes. The first hinge segment (after α6 in β-glucosyltransferase, after β18 in glycogen phosphorylase) is not structurally conserved. In the second connecting segment structural conservation is quite conspicuous, in particular around the α11−α12 (of β-glucosyltransferase) corner at the bottom right in Figure 3. A short loop between α12 and α13 (of β-glucosyltransferase) appears flexible, but the end of α13 again overlaps neatly.

Interestingly, similar interdomain rotations (~5°) have been reported for both glycogen phosphorylase and β-glucosyltransferase. In glycogen phosphorylase, the conformational change is related to the transition between the high-activity R state and the low-activity T-state (Sprang *et al.*, 1992). In β-glucosyltransferase, the change occurs on binding the UDP-glucose substrate (Vrielink *et al.*, 1994). The interdomain rotation might be required during the catalytic cycle to juxtapose groups attached to either domain 1 or domain 2 (see Withers *et al.*, 1982b). Thus, there may be a functional reason for the striking structural conservation in interdomain connecting elements.

Very low sequence similarity. Sequence identity between β-glucosyltransferase and glycogen phosphorylase in the structural alignment (Figure 2) is <10%. Inspection of sequence conservation in a multiple alignment of 12 glycogen phosphorylases and the single β-glucosyltransferase sequence reveals that not a single functional residue is identically conserved. However, there is one conserved buried hydrophobic site which might be structurally important: Tyr374, Val/Ile452 and Phe479 from the adjacent strands β13, β17 and β18 in glycogen phosphorylases are layered on top of one another in a way similar to Tyr96, Val134 and Phe157 in strands β5−β7 of β-glucosyltransferase.

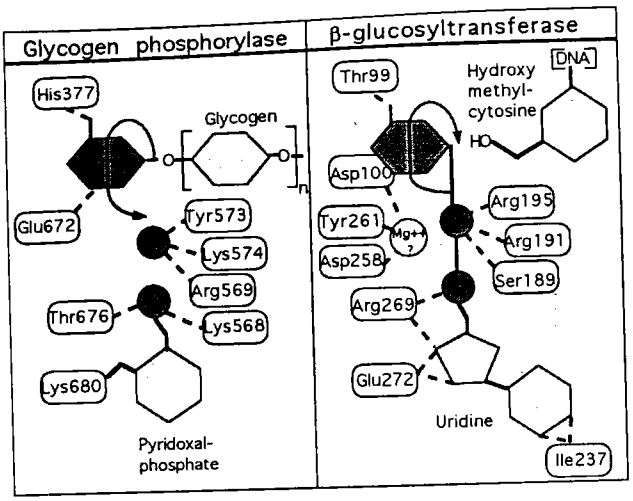

Fig. 5. Major evidence for evolutionary relatedness comes from the fact that the structural alignment of the polypeptide chains of β-glucosyltransferase and glycogen phosphorylase non-trivially leads to superimposition of the substrates: the phosphate groups in the catalytic site of β-glucosyltransferase overlap with those of glycogen phosphorylase and the pyridoxal ring in glycogen phosphorylase occupies a position structurally equivalent to that of the ribose ring of UDP in β-glucosyltransferase. Polar interactions between substrates and protein (side chain or backbone) are indicated by dashed lines. Hexagons with dark shading denote the glucose unit that is transferred (arrow) between glycogen and orthophosphate (glycogen phosphorylase) or between UDP and hydroxymethylated cytosine (β-glucosyltransferase). Circles with dark shading denote phosphate groups.

Conservation and modifications in the active site

In both enzymes, the active site is harbored in the cleft between the two domains of the common core. Surprisingly, superimposition of the polypeptide chains leads, non-trivially, to spatial coincidence of bound substrates and co-enzyme, i.e. UDP-glucose in β-glucosyltransferase, and PLP, orthophosphate and glucose in glycogen phosphorylase (Figure 3). The most important protein–substrate interactions, summarized schematically in Figure 5, involve structurally equivalent residues between β-glucosyltransferase and glycogen phosphorylase (Figure 2).

PLP/UDP site

The PLP ring in glycogen phosphorylase corresponds structurally to the ribose ring of UDP in β-glucosyltransferase. An intriguing feature of the global 3-D superimposition (Figure 3) of the two structures is exact conservation of the position of helix α10 of β-glucosyltransferase with respect to helix α21 of glycogen phosphorylase. There may be a functional reason: helix α21 of glycogen phosphorylase is the helix that carries the PLP co-enzyme in a Schiff base linkage to Lys680. Remarkably, the residue equivalenced in the structural alignment with Lys680 of glycogen phosphorylase is Glu272 in β-glucosyltransferase—precisely the residue that by hydrogen bonding anchors the ribose ring of UDP in place.

Phosphate binding sites

Multiple phosphate binding sites have been defined in the glycogen phosphorylase structure in crystallographic studies of complexes with different ligands (Sprang *et al.*, 1992). After superimposition of the two protein structures, there is clear overlap with the positions of the phosphate groups of UDP in β-glucosyltransferase. The phosphate binding loops in the two enzymes are structurally equivalent although their conformations differ slightly between glycogen phosphorylase and β-glucosyltransferase. In particular, the first turn of helix α7 (β-glucosyltransferase)/α18 (glycogen phosphorylase) is unwound in glycogen phosphorylase (cf. Figure 2). Surprisingly, the invariant Arg574 in glycogen phosphorylases points away from the active site whereas the equivalent Arg195 in β-glucosyltransferase points toward the active site and interacts with the β-phosphate of UDP. The conservation of these residues may therefore be a coincidence.

Glucose site

The site of glucose binding inferred by the β-glucosyltransferase crystallographers appears roughly equivalent to the site seen in the glycogen phosphorylase complex with glucose analogues. There are some architectural changes in this region. In particular, the end of helix α1 (β-glucosyltransferase)/helix α6 (glycogen phosphorylase) that points toward the active site is shifted ~5 Å. Nevertheless, if the known position of the glucose ring is transferred from glycogen phosphorylase (Johnson *et al.*, 1990) to the structurally equivalent region in β-glucosyltransferase, we can predict a testable conserved hydrogen bonding interaction with glucose by Thr99 of β-glucosyltransferase, which is equivalent to the invariant His377 of glycogen phosphorylases.

Binding of polymer substrates

Neither enzyme has been crystallized in a complex with bound polymer substrate (DNA or glycogen), but two at least qualitatively common features have been inferred from the crystal structures: (i) a solvent channel leads out of the active site in equivalent directions; (ii) a flexible loop covers part of the entrance to the active site (280s loop in glycogen phosphorylase; disordered loop 110–122 in β-glucosyltransferase). β-Glucosyltransferase modifies its polymer substrate with endo-activity and glycogen phosphorylase with exo-activity. There are obvious, and so far unsolved, steric problems in how the large substrate can reach the buried active site. Apart from radical conformational changes of the protein, one possibility for β-glucosyltransferase is that the hydroxymethylcytosine base flips out of the double-stranded DNA structure (cf. Vrielink *et al.*, 1994).

Similarities and differences in catalytic mechanism

The biochemical features of both enzymes are remarkably similar, in spite of their very different biological roles. Both β-glucosyltransferase and glycogen phosphorylase catalyze the transfer of glucosyl units between phosphate ester and *O*-glycosidic bonds (see Figure 5). The mechanism of glycogen phosphorylase has been more extensively studied than that of β-glucosyltransferase, but the nature of the contact between the substrate and co-enzyme phosphate groups and the mechanism by which it facilitates catalysis is not perfectly understood even in glycogen phosphorylase (Fukui *et al.*, 1982; Withers *et al.*, 1982a; Johnson *et al.*, 1990; Sprang *et al.*, 1992; Duke *et al.*, 1994).

The reaction catalyzed by glycogen phosphorylase is

reversible; the principle of mass action drives it in the direction of glycogen degradation *in vivo*. Interestingly, it has been demonstrated that glycogen phosphorylase can catalyze the cleavage of the glycosidic bond of the covalently bound co-enzyme–substrate analogue PLPP-α-Glc, transferring the glucose unit to glycogen in α-1,4-linkage (Withers *et al.*, 1982b). This experiment beautifully highlights the very close analogy of enzymatic action, as PLPP-α-Glc used in this glycogen phosphorylase reaction differs chemically from UDP-glucose, the substrate of β-glucosyltransferase, only in the substitution of pyridoxal for uridine. Thus, the diphosphate-glucose component containing the reactive bonds is functionally equivalent between glycogen phosphorylase and β-glucosyltransferase (dark shading in Figure 5).

The electronic environments of the reactive phosphate are similar but not identical between glycogen phosphorylase and β-glucosyltransferase (Figure 5). In glycogen phosphorylase, the PLP co-enzyme acts to polarize the reactive orthophosphate. β-Glucosyltransferase requires magnesium for activity. The Mg^{2+} ion, bound to the UDP phosphates, has been proposed to stabilize the negative charge of the leaving group phosphate (Vrielink *et al.*, 1994). The stereochemistry of the reactions also differs. Glycogen phosphorylase is specific for α-linkages while β-glucosyltransferase makes β-linkages. In other words, the glycosidic bonds are formed from below or from above the plane of the glucose ring, respectively, implying small shifts in the relative positions of either glucose or the polymer substrate.

Evolutionary implications

The 3-D structural similarities between β-glucosyltransferase and glycogen phosphorylase are extensive enough to leave no doubt about their evolutionary relatedness. In addition to the fold of the common catalytic core, the mechanics of interdomain motion and the catalytic principle of the two glucosyltransferases are conserved.

Phylogenetic background

Glycogen phosphorylase is a basic metabolic enzyme found in animals, plants and bacteria alike. All glycogen phosphorylases have similar sizes, strongly conserved sequences and most respond similarly to at least one allosteric inhibitor (glucose-6-phosphate; Hudson *et al.*, 1993). Thus, glycogen phosphorylases constitute a phylogenetically ancient enzyme family that probably already existed in close to its present form before the divergence of prokaryotes and eukaryotes. On these time scales, the host-dependent phage T4 is probably a newcomer.

A relatively recent origin of the phage DNA modification process is consistent with the fact that the first step, cytosine hydroxymethylation, is catalyzed by a phage T4 enzyme which is homologous to phage T4 (and *Escherichia coli*) thymidylate synthase (an enzyme essential for DNA synthesis). Apart from T-even phages, DNA glucosylation so far has only been detected in another parasite, the African trypanosome (Gommers-Ampt *et al.*, 1993). It seems plausible that the DNA modification systems of T4 phage and trypanosomes have evolved independently in these phylogenetically distant life forms and may be based on a completely different protein fold.

A second relative in phage T4?

The second DNA modification step in T4 phage, i.e. glucosylation of hydroxymethylcytosine residues, occurs both by α- and β-linkages. Although sequence similarity is weak (~16% identity), a possible divergent evolutionary relationship of the α-glucosyltransferase with the β-glucosyltransferase with conservation of DNA binding surfaces is supported by several facts: similar protein sizes (400 versus 351 amino acids); similar hydrophobicity patterns (atomic solvation preferences −167 for the crystallographic model and −132 for a 3-D model comprising 314 residues of α-glucosyltransferase mounted onto a β-glucosyltransferase backbone template; Holm and Sander, 1992a,b); conservation of the functionally important Glu272 of β-glucosyltransferase (Figure 5); and surprisingly similar electrostatic fields (calculated from α- and β-glucosyltransferase all-atom models as in Nicholls and Honig, 1991). We caution that the relationship between α- and β-glucosyltransferase is speculative and subject to experimental falsification.

Solving an evolutionary paradox

It appears paradoxical that the architecture of T4 β-glucosyltransferase is simpler and therefore appears more primitive than that of glycogen phosphorylase, which is the older enzyme by phylogenetic arguments. The paradox may be resolved by either of two evolutionary models that differ in the point of divergence between glycogen phosphorylase and T4 β-glucosyltransferase. The first model postulates that the gene for T4 β-glucosyltransferase diverged rather recently from a fully evolved glycogen phosphorylase and evolution in T4 rapidly simplified the structure of the protein down to the essential catalytic core. In the second model both descend along separate lineages from a very ancient common ancestor (Figure 6). Subsequently, glycogen phosphorylase gradually acquired structural decorations with regulatory function. In the other lineage, the T4 β-glucosyltransferase rather recently branched off from a probably UDP-dependent descendant of the ancient common ancestor. Such a UDP-dependent enzyme is probably still extant in bacteria, and constitutes a crucial missing link in this evolutionary scenario.

Evidence from a large number of distantly related protein families shows that protein folds tend to be very well conserved, i.e. structural modifications are much less likely than sequence changes or changes of enzymatic specificity. The second evolutionary model is therefore preferable as it requires only one round of structural modifications, while the first model requires first addition and subsequently removal of numerous structural excursions (Figure 4).

Recruitment of a co-enzyme

How can the presence of alternatively UDP- or PLP-dependent enzymes of identical fold and similar chemistry be explained in evolutionary terms? Nucleoside sugars are glycosyl donors in many biosynthetic pathways. We presume that use of non-covalently bound UDP-glucose substrate preceded recruitment of covalently bound co-enzyme in the glycogen phosphorylase lineage. The way in which PLP is exploited in glycogen phosphorylase is quite different from a wide range of other PLP enzymes (e.g. aspartate aminotransferase; Kirsch *et al.*, 1984) that

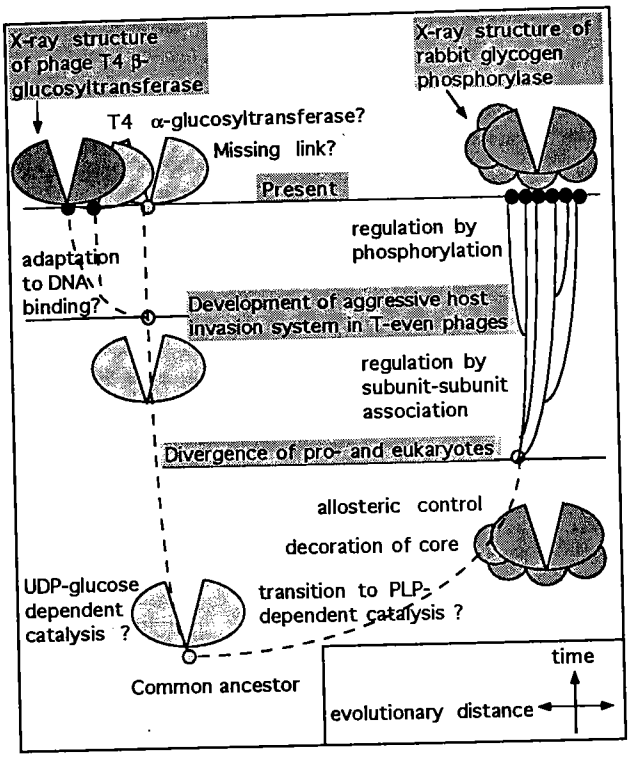

Fig. 6. Plausible evolutionary progression consistent with the structural similarity between β-glucosyltransferase and glycogen phosphorylase. Facts are indicated by dark shading and solid lines. Light shading, question marks and dashed lines denote conjecture. The evolution of regulation and allosteric control mechanisms in glycogen phosphorylases has been established from sequence comparisons (Fukui *et al.*, 1982; Hwang and Fletterick, 1986; Hudson *et al.*, 1993). We predict the presence of a missing link, an NDP-dependent bacterial enzyme with structural similarity to T4 phage β-glucosyltransferase.

catalyze amino acid transformations by reversibly exchanging the PLP–Schiff base between a lysine residue in the enzyme and amino acid substrates. The structural analysis of the present enzymes indicates that a transition to PLP-dependent catalysis of glycogen degradation may have been triggered by a point mutation (Glu272 of β-glucosyltransferase to Lys680 of glycogen phosphorylase, Figure 5). The mutation would have been fixed in the glycogen phosphorylase lineage because of its energetic efficiency (one ATP per glucose is saved compared with hydrolysis and subsequent phosphorylation).

Search for the missing link

The missing link enzyme in our evolutionary model is more likely to be involved in carbohydrate metabolism than in DNA modification. A later acquisition of non-specific DNA binding by β-glucosyltransferase could have been easily achieved by mutation of surface residues resulting in the currently observed enrichment of positive charges on one side of the molecule (Vrielink *et al.*, 1994). This led us to commence the search for the missing link by examining the UDP-glucose (in bacteria, ADP-glucose) dependent glycogen synthesis pathway, i.e. the reverse of glycogen degradation, for a possible candidate for the missing link enzyme. A novel superfamily comprising mammalian/yeast glycogen synthases, plant/bacterial glycogen (starch) synthases, and plant sucrose and sucrose-phosphate synthases emerged with a putative active site around an invariant glutamic acid residue [representative Swissprot (Bairoch and Boeckmann, 1991) sequences are ugst_human: Glu510, glga_ecoli: Glu378, susy_soltu: Glu674, sps_maize: Glu592]. However, secondary structure predictions based on the multiple sequence alignment were incompatible with a glycogen phosphorylase/β-glucosyltransferase-like fold (data not shown). So we have not found the missing link but propose that the quest for one might be an interesting case for sequence–structure alignment (threading) techniques as more and more unidentified protein sequences become available.

Conclusion

We have discovered a distant evolutionary relationship between the DNA-modifying enzyme β-glucosyltransferase from bacteriophage T4 and glycogen phosphorylase which is involved in carbohydrate metabolism. Analysis of the active sites revealed unexpected biochemical similarities that may inspire further studies of the catalytic mechanism in both enzymes. It is plausible, in our view, that β-glucosyltransferase and glycogen phosphorylase are only the first members of a larger glucosyltransferase superfamily characterized by a structurally conserved catalytic core. We would not be surprised if additional members, e.g. other NDP-dependent glucosyltransferases, are discovered soon. Our results suggest that the T4 phage DNA modification system, responsible for the extreme virulence of the phage, represents an evolutionary adaptation of basic metabolic pathways based on diphosphates and sugars.

Materials and methods

Three-dimensional structural equivalences were determined and optimized using the Dali algorithm which is based on alignment of two-dimensional distance maps (Holm and Sander, 1993). Glycogen phosphorylase was identified as distinctly the best hit in a scan of β-glucosyltransferase against a set of 420 representative protein structures (Hobohm and Sander, 1994). Atomic coordinates were retrieved from the Protein Data Bank (Bernstein *et al.*, 1977). Sequence numbers for glycogen phosphorylase used throughout this work are those of the rabbit muscle enzyme. Sequence database searches were performed on the GeneQuiz workbench (Scharf *et al.*, 1994). Multiple alignments were edited in the Genetic Data Environment (Smith, 1993). Secondary structure predictions were by the PHD program (Rost and Sander, 1994). A rough partial model of α-glucosyltransferase was generated by replacing side chains in the β-glucosyltransferase structure using the MaxSprout program (Holm and Sander 1992a). The sequence alignment of α- and β-glucosyltransferase was iteratively improved guided by inspection of atomic solvation preference profiles (Holm and Sander, 1992b). Electrostatic potential calculations were undertaken using the Delphi program (Nicholls and Honig, 1991).

References

Acharya,K.R., Stuart,D.I., Varvill,K.M. and Johnson,L.N. (1991) *Glycogen Phosphorylase b: Description of the Protein Structure.* World Scientific Publishing Co. Pte. Ltd, Singapore.

Bairoch,A. and Boeckmann,B. (1991) *Nucleic Acids Res.*, **19**, 2247–2250.

Bernstein,F.C., Koetzle,T.F., Williams,G.J.B., Meyer,E.F., Brice,M.D., Rodgers,J.R., Kennard,O., Shimanouchi,T. and Tasumi,M. (1977) *J. Mol. Biol.*, **112**, 535–542.

Duke,E.M.H., Wakatsuki,S., Hadfield,A. and Johnson,L.N. (1994) *Protein Sci.*, **3**, 1178–1196.

Fukui,T., Shimomura,S. and Nakano,K. (1982) *Mol. Cell. Biochem.*, **42**, 129–144.

Gommers-Ampt,J.H., VanLeeuwen,F., DeBeer,A.L.J., Vliegenhart,J.F.G., Dizdaroglu,M., Kowalak,J.A., Crain,P.F. and Borst,P. (1993) *Cell*, **75**, 1129–1136.

Hobohm,U. and Sander,C. (1994) *Prot. Sci.*, **3**, 522.
Holm,L. and Sander,C. (1992a) *Proteins*, **14**, 213–223.
Holm,L. and Sander,C. (1992b) *J. Mol. Biol.*, **225**, 93–105.
Holm,L. and Sander,C. (1993) *J. Mol. Biol.*, **233**, 123–138.
Holm,L. and Sander,C. (1994a) *Proteins*, **19**, 165–173.
Holm,L. and Sander,C. (1994b) *Nucleic Acids Res.*, **22**, 3600–3609.
Holm,L. and Sander,C. (1994c) *Proteins*, **19**, 256–268.
Hudson,J.W., Golding,G.B. and Crerar,M.M. (1993) *J. Mol. Biol.*, **234**, 700–721.
Hwang,P.K. and Fletterick,R.J. (1986) *Nature*, **324**, 80–84.
Johnson,L.N., Acharya,K.R., Jordan,M.D. and McLaughlin,P.J. (1990) *J. Mol. Biol.*, **211**, 645–661.
Kabsch,W. and Sander,C. (1983) *Biopolymers*, **22**, 2577–2637.
Kiel,J.A.K.W., Boels,J.M., Beldman,G. and Venema,G. (1994) *Mol. Microbiol.*, **11**, 203–218.
Kirsch,J.F, Eichele,G., Ford,G.C., Vincent,M.G., Jansonius,J.N., Gehring,H. and Christen,P. (1984) *J. Mol. Biol.*, **174**, 497–525.
Kraulis,P. (1991) *J. Appl. Crystallogr.*, **24**, 946–950.
Murzin,A.G. (1993) *Trends Biochem. Sci.*, **18**, 403–405.
Neidhart,D.J., Kenyon,G.L. Gerlt,J.A. and Petsko,G.A. (1990) *Nature*, 347, 692–694.
Nicholls,A. and Honig,B. (1991) *J. Comp. Chem.*, **12**, 435–445.
Rath,R.L. and Fletterick,R.J. (1994) *Struct. Biol.*, **1**, 681–690.
Rost,B. and Sander,C. (1994) *Proteins*, **19**, 55–72.
Scharf,M., Schneider,R., Casari,G., Bork,P., Valencia,A., Ouzounis,C. and Sander,C. (1994) In Altman,R., Brutlag,D., Karp,P., Lathrop,R. and Searls,D. (eds), *Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology.* AAAI Press, Menlo Park CA, pp. 348–353.
Smith,S. (1992) Genetic Data Environment version 2.0. Harvard University.
Sprang,S.R., Madsen,N.B. and Withers,S.G. (1992) *Protein Sci.*, **1**, 1100–1111.
Sprang,S.R., Withers,S.G., Goldsmith,E.J., Fletterick,R.J. and Madsen,N.B. (1991) *Science*, **254**, 1367–1371.
Tomaschewski,J., Gram,H., Crabb,J.W. and Rüger,W. (1985) *Nucleic Acids Res.*, **13**, 7551–7568.
Vrielink,A., Rüger,W., Driessen,H.P.C. and Freemont,P.S. (1994) *EMBO J.*, **13**, 3413–3422.
Vriend,G. (1990) *J. Mol. Graph.*, **8**, 52–56.
Wilmanns,M., Hyde,C.C., Davies,D.R., Kirschner,K. and Jansonius,J.N. (1991) *Biochemistry*, **30**, 9161–9169.
Withers,S.G., Madsen,N.B., Sprang,S.R. and Fletterick,R.J. (1982a) *Biochemistry*, **21**, 5372–5382.
Withers,S.G., Madsen,N.B., Sykes,B.D., Takagi,M., Shimomura,S. and Fukui,T. (1982b) *J. Biol. Chem.*, **256**, 10759–10762.

Received on November 28, 1994; revised on January 11, 1995

Note added in proof

Independently, Artymiuk *et al.* [Artymiuk,P.J., Rice,D.W., Poirette,A.R. and Willett,P. (1995) *Struct. Biol.*, **2**, 117–120] have detected this same structural similarity.