E. Coli MurG: A Paradigm for a Superfamily of Glycosyltransferases

Ha, S., Gross, B. and Walker, S.*

Chemistry Department, Princeton University, Princeton, NJ 08544, USA

Abstract: MurG is an essential bacterial glycosyltransferase that is involved in the biosynthesis of peptidoglycan. The enzyme is found in all organisms that synthesize peptidoglycan and is a target for the design of new antibiotics. A direct assay to study MurG was reported recently, followed shortly by the crystal structure of *E. coli* MurG. This first MurG structure, combined with sequence data on other glycosyltransferases, has revealed that MurG is a paradigm for a large family of metal ion-independent glycosyltransferases found in both eukaryotes and prokaryotes. A better understanding of MurG could lead to the development of new drugs to combat antibiotic resistant infections, and may also shed light on a broad class of glycosyltransferases.

THE PROBLEM OF ANTIBIOTIC RESISTANCE

Overuse of antibiotics in medicine and agriculture has led to extensive resistance against most classes of clinically used antibiotics [1]. Genes that confer resistance can be transferred readily from one bacterial strain to another, producing multi-drug resistant organisms. Multi-drug resistant bacteria are a particular threat in hospitals, and thousands of patients now die each year from hospital-acquired bacterial infections that are resistant to antibiotic treatment. We need to identify new antibacterial agents before particularly virulent strains of multi drug resistant bacteria - vancomycin-resistant staphylococcal stains, for example - spread from hospitals to the community at large. Structural and mechanistic information on essential bacterial enzymes could lead to the design of new antibiotics. Hence, scientists in both academia and industry have redoubled their efforts to identify and characterize key bacterial enzymes.

THE BACTERIAL CELL WALL AS A TARGET

Bacterial cell membranes are subjected to high osmotic pressures because bacterial cells generate high internal concentrations of ions and metabolites under normal growth conditions. Bacterial membranes are able to withstand these high pressures because concentric layers of a crosslinked polymer called peptidoglycan (Fig. (**1**)) surround them [2-4]. Compounds that damage the peptidoglycan layer cause bacterial cells to lyse. Therefore all of the enzymes involved in the biosynthesis of peptidoglycan are potential targets for the design of new antibiotics.

Although there are other essential metabolic pathways in bacteria - DNA, RNA, and protein synthesis, for example - peptidoglycan synthesis may have advantages as a target for inhibition because there are no parallels to this pathway in eukaryotes. Peptidoglycan is comprised of linear chains of a repeating -linked disaccharide unit held together by peptide crosslinks (Fig. (**1**)) [2]. One of the sugars in the repeating disaccharide, N-acetyl muramic acid, is not made in eukaryotic cells. Furthermore, the peptides in the crosslinks contain both D-amino acids and unusual backbone linkages, two additional features not found in eukaryotes. Therefore inhibitors of enzymes involved in many steps of peptidoglycan synthesis are likely to be specific for bacterial cells. Because the basic elements of the peptidoglycan layer are remarkably similar across bacterial strains, such inhibitors may well have broad-spectrum activity. Structural information on enzymes involved in peptidoglycan synthesis is an important first step toward the development of new inhibitors of this pathway.

THE THREE STAGES OF PEPTIDOGLYCAN BIOSYNTHESIS

The biosynthetic pathway to peptidoglycan occurs in three distinct stages. The first stage involves the conversion of UDP-N-acetyl glucosamine to UDP-N-acetyl muramyl pentapeptide [5] (Fig. (**2**)). This stage occurs in the cytoplasm of the bacterial cell and requires at least seven enzymes [5]. The first two enzymes, MurA and MurB, convert UDP-GlcNAc to UDP-MurNAc [6,7]. The rest of the enzymes are ligases that form the amide bonds of the peptide chain. The composition of the peptide chain varies somewhat between organisms [8]. For example, in *E. coli* the peptide chain contains *meso*-diaminopimelic acid in the third position, whereas in most gram positive organisms lysine is found in the third position. In some gram positive organisms, a string of glycines is attached to the lysine at position 3. Nevertheless, certain features of the peptide chain are broadly conserved. For example, the peptide chain typically contains five amino acids in the linear sequence; the terminal dipeptide is usually D-Ala-D-Ala; and the second amino acid is coupled through its side chain. The functional differences between peptides with different compositions are poorly understood.

*Address correspondence to this author at the Chemistry Department, Princeton University, Princeton, NJ 08544, USA; phone: (609) 258-1149; fax: (609) 258-2617; email: swalker@Princeton.edu
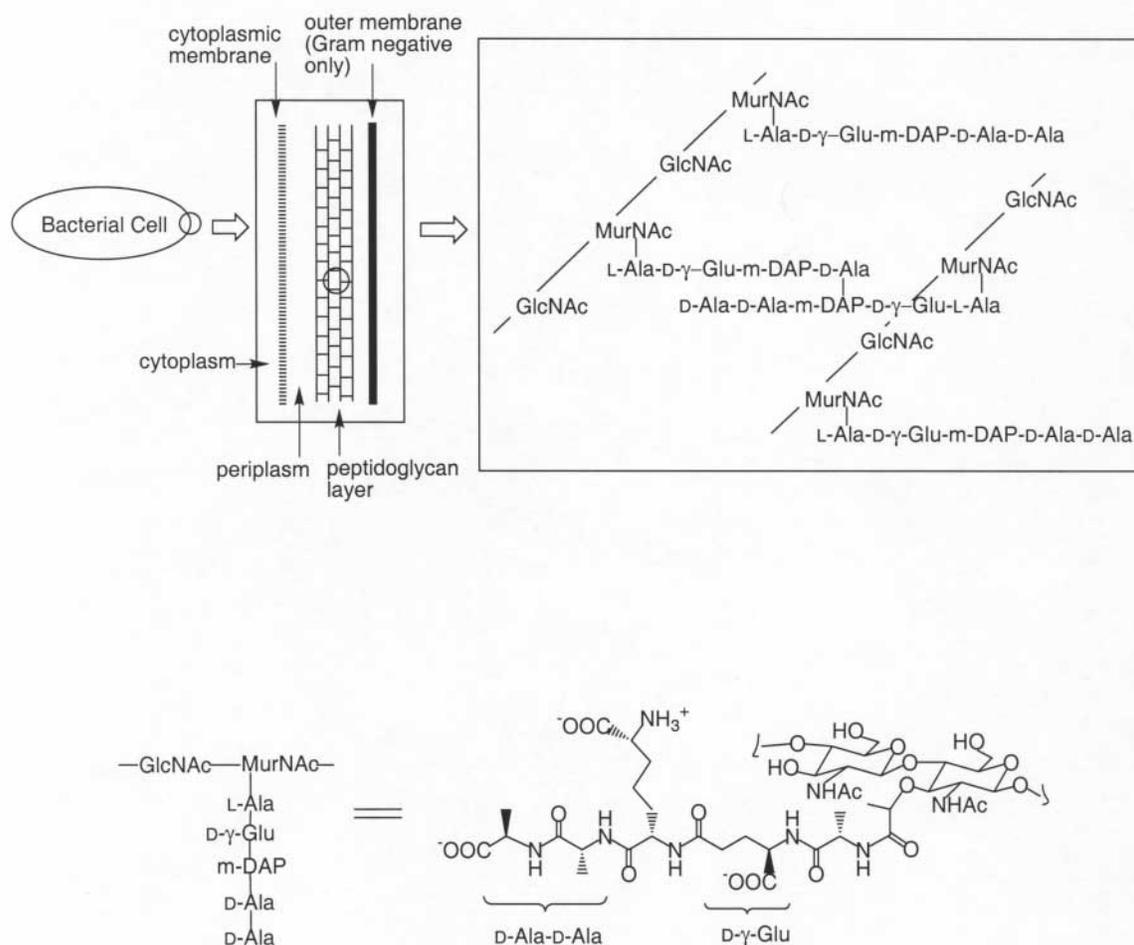
Fig. (1). Architecture of peptidoglycan.

The second stage of peptidoglycan synthesis occurs on the cytoplasmic surface of the bacterial membrane and involves two enzymes, MraY and MurG. MraY is an integral membrane protein that catalyzes a pyrophosphate exchange reaction between an undecaprenyl phosphate anchored in the bacterial membrane and UDP-MurNAc pentapeptide to form a lipid anchored MurNAc pentapeptide called Lipid I [9]. The next enzyme in the biosynthetic pathway, MurG, is a glycosyltransferase that catalyzes the transfer of GlcNAc from UDP to the C4 hydroxyl of Lipid I to produce a -(1,4)-linked disaccharide known as Lipid II [3].

Lipid II is the minimal subunit of peptidoglycan. Once it is formed, it is somehow translocated to the exterior surface of the bacterial membrane where Stage III of peptidoglycan synthesis occurs. This stage of the biosynthesis is quite complex, but generally involves the coupling of disaccharide units by transglycosylases to form linear carbohydrate polymers which are then crosslinked in a transamidation reaction [10,11]. In this transamidation activity, the Lys or

meso-DAP side chain amine of a pentapeptide on one carbohydrate strand attacks the D-Ala-D-Ala amide bond of another carbohydrate strand. There are a large number of enzymes involved in the process of polymerization and crosslinking. Some of the enzymes contain both transglycosylase and transpeptidase activities; others contain only one or the other function. Although the individual transglycosylase and transpeptidase domains have significant homology, the enzymes perform different functions during the bacterial life cycle. The specific tasks performed by individual enzymes during cell growth and division, and how an intact sacculus is maintained throughout the process, remains poorly understood [11].

THE ENZYMES THAT MAKE PEPTIDOGLYCAN

All of the enzymes involved in Stage I of peptidoglycan synthesis have been cloned, purified, and characterized. By 1995, there were crystal structures of MurA [12,13], MurB [14,15], MurD [16,17], and D-Ala-D-Ala ligase [18]. More
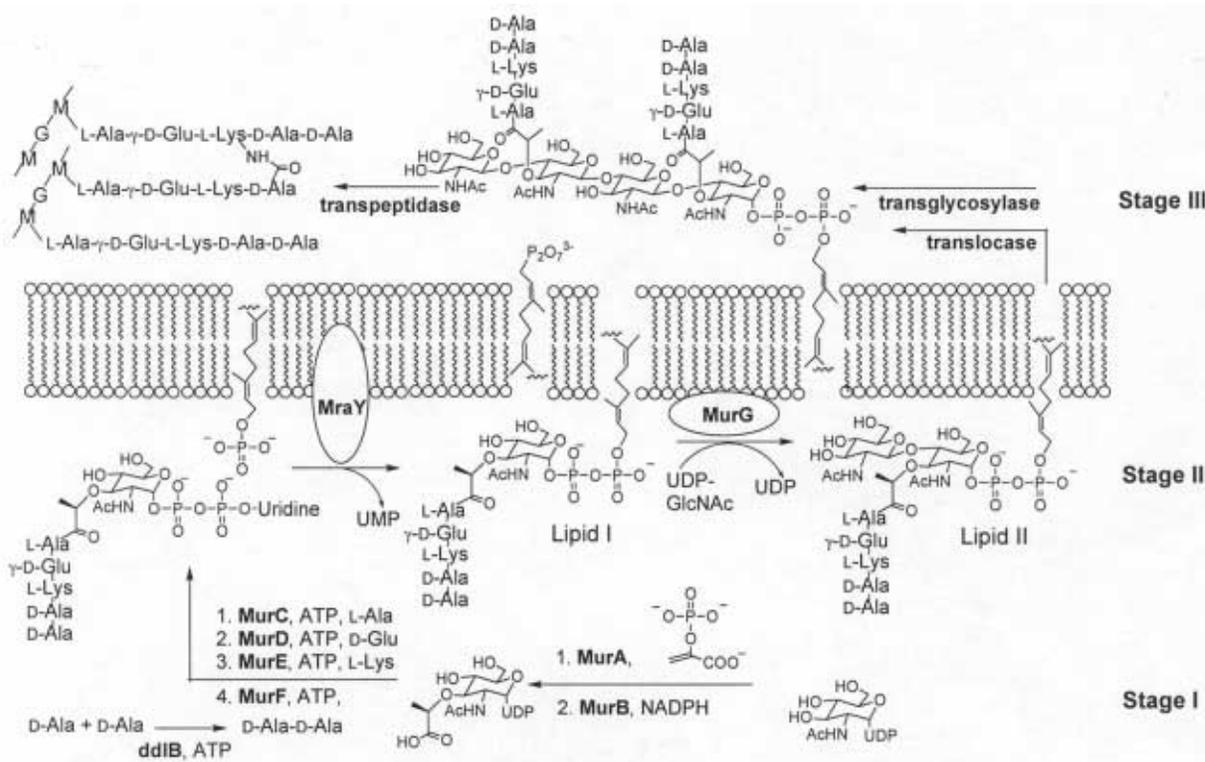
Fig. (2). Scheme of peptidoglycan synthesis. The composition of the pentapeptide can vary somewhat, as discussed in the text.

recently, the structure of MurF [19] has been reported. Hence, there is a large amount of structural and mechanistic information on the enzymes involved in the first stage of peptidoglycan synthesis. This information may prove useful in the design of new antibiotics. Good screens for evaluating inhibitors of the enzymes have also been developed, and it may not be long before promising antibiotics that target Stage I of peptidoglycan synthesis are reported [20].

Progress in characterizing the enzymes involved in the second and third stages of peptidoglycan synthesis has been slow compared to the advances in characterizing Stage I enzymes. Second and third stage enzymes are more difficult to study for two reasons. The first reason is that these enzymes are all membrane-associated, which makes them more difficult to overexpress and purify than the soluble, cytoplasmic Stage I enzymes. The second reason is that the substrates for these enzymes are difficult to obtain and to handle. As peptidoglycan synthesis progresses, the substrates, or intermediates, become larger and more complicated. Then, at the beginning of the second stage of peptidoglycan synthesis, the intermediate is anchored to the membrane through a fifty-five carbon undecaprenyl chain. The long hydrocarbon chain is carried throughout the rest of the pathway and renders all subsequent intermediates insoluble in water. An additional problem is that at least some of the intermediates involved in the third stage of peptidoglycan synthesis are polymers of indeterminate length

and composition. Finally, most of the substrates involved in the second and third stages of peptidoglycan synthesis are present in minute quantities in bacterial cells and thus cannot be isolated readily from natural sources [3,21]. Without suitable substrates to monitor activity, it is not possible to study purified enzymes. Only one structure of an enzyme involved in the late stages of peptidoglycan synthesis was reported prior to the year 2000. This structure was of PBP2x from *Streptococcus pneumoniae* [22]. Dideberg and co-workers, who solved the structure, assumed that the protein had been purified in its active conformation because it retained the ability to bind penicillin. Another penicillin binding protein has since been crystallized, with an antibiotic bound [23]. However, studies on other peptidoglycan synthesizing enzymes have had to await the development of methods to obtain suitable substrates. Below, we describe studies on MurG that grew out of chemical methods to make Lipid I analogues. It should be possible to adapt the approach described to study other enzymes involved in the membrane-linked stages of peptidoglycan synthesis.
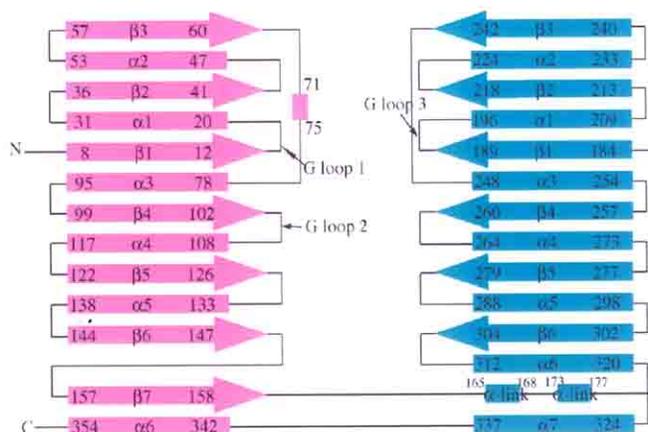
HISTORICAL BACKGROUND OF MurG

MurG is the second enzyme in stage II of peptidoglycan biosynthesis. It catalyzes the transfer of N-acetylglucosamine (GlcNAc) from UDP-GlcNAc to the C4 hydroxyl of the

A.
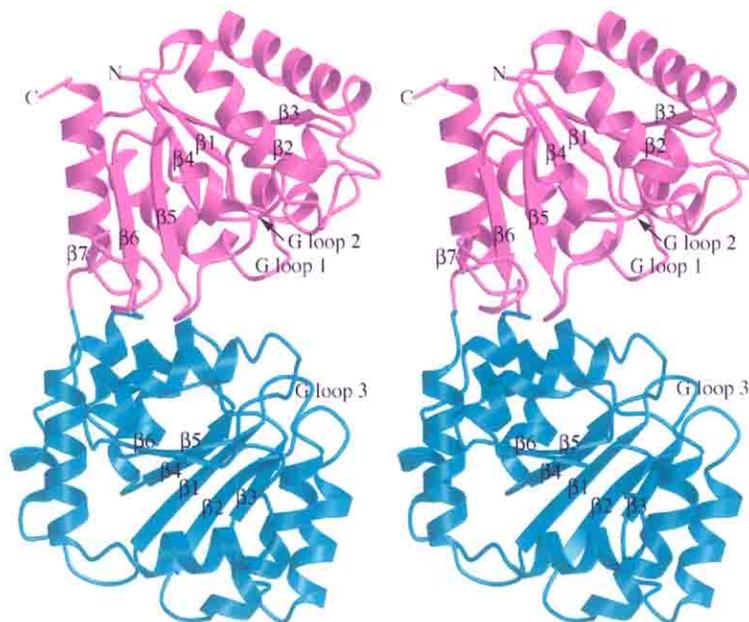
B.

Fig. (3). Overall architecture of MurG. (A) Stereoview of the MurG structure. The N-domain is shown in purple and the C-domain is shown in green. Generated with the programs MOLSCRIPT and RASTER3D. (B) Topology diagram of MurG. Color coding is the same as in (A).

membrane anchored undecaprenyl-pyrophosphate-MurNAc-pentapeptide (Lipid I) to form Lipid II. The study of MurG dates back to the 1960s when Strominger and co-workers first identified the lipid intermediates of cell wall synthesis [24,25]. However, the *murG* gene was not identified until 1980 when Salmond *et al.* cloned it [26]. It was speculated that the product of *murG* was involved in peptidoglycan metabolism based on alterations in the cell shape of a thermosensitive *murG* mutant strain. The sequence of the *murG* gene and the corresponding amino acid sequence were reported independently by two different groups in 1990 [27,28]. In 1991, Heijenoort and co-workers established for the first time that the *murG* gene codes for the GlcNAc transferase whose existence was first proposed by Strominger [29]. Almost at the same time, Matsuhashi *et al.* from Japan

independently reported MurG as a GlcNAc transferase [30]. Then, in 1993, MurG was localized to the inner surface of cell membrane [31].

Although the *murG* gene had been cloned and sequenced by 1990, purification of the protein was not reported until 1998 because there was simply no way to assay MurG activity except in crude membranes [30,32,33]. The lipid I substrate cannot be obtained for enzyme assays from natural sources because it is present in tiny quantities [34].

Recently, two alternative approaches to obtaining Lipid I analogues to study MurG have been reported. Auger *et al.* [35] have reported a semi-synthetic route to making Lipid I analogues from UDP-MurNAc pentapeptide, an intermediate

which can be isolated in large quantities from bacterial cells. In this approach, UDP-MurNAc pentapeptide was treated with snake venom phosphodiesterase to cleave the pyrophosphate linkage and the anomeric sugar phosphate was then coupled chemically to a thirty five carbon activated lipid phosphate. This 35 carbon Lipid I analogue was found to be a substrate for MurG, but has not been utilized in further studies.

We developed a convergent total synthesis of Lipid I and analogs [36-38]. Although time consuming, a fully synthetic approach has advantages over isolation or semi-synthesis. The chief advantage is that is possible to vary independently the three major components of Lipid I - the sugar, the peptide, and the lipid chain - in order to evaluate their roles [36-38]. Using synthetic substrate analogs, we have been able to monitor the activity of *E. coli* MurG during purification. Although it is normally associated with the cytoplasmic membrane [31], MurG can be readily removed from the membrane with Triton X-100. Furthermore, the solubilized enzyme is active and accepts Lipid I substrates containing hydrocarbon chains at least as short as ten carbons [38]. Hence, a membrane interface is not required to maintain the active conformation of MurG. Moreover, MurG can be concentrated to high levels (>10mg/mL) without precipitating. Although this behavior was unexpected, it afforded us the opportunity to try to grow crystals.

STRUCTURE OF *E. COLI* MurG

Crystals of *E. coli* MurG containing a C-terminal LEHHHHHH sequence were grown at room temperature using the hanging-drop vapor diffusion method in a NaMES buffer (pH 6.5) containing 0.5 M $(NH_4)_2SO_4$ and 0.2% Triton X-100. Triclinic crystals belonging to the P1 space group and having two molecules per asymmetric unit grew to a typical size of 0.2 mm × 0.1 mm × 0.1 mm within a week. The crystal structure was solved by a combination of anomalous scattering and multiple isomorphous replacement and refined to 1.9 Å. The structure is shown in Fig. (**3**) [39]. MurG contains two domains separated by a cleft that is approximately 20Å deep × 18Å wide at the widest point. Both domains adopt an / open-sheet motif, also known as a Rossmann fold. A common Rossmann motif contains six strands connected by helices. The strands form a twisted parallel sheet with the strand order reading 654123 [40]. As shown in the topology diagram (Fig. (**3**)) for MurG, both domains approximate a classic Rossmann fold and have high structural homology despite a lack of sequence identity.

The Rossmann fold was first identified in proteins that contain domains that bind to diphosphate-containing cofactors such as NADH [41]. In a typical Rossman domain, the diphosphates are bound near a P-loop, a glycine-rich stretch of amino acids located at the carboxyl terminus of one strand and its adjacent helix. It is believed that the positively charged helix dipole of the helix connecting the P-loop to the preceding element of sheet interacts favorably with the anionic substrate [42]. Thus, no metal ions or positively charged residues are necessary to stabilize the negative charges. Several facts suggest that the two domains in MurG not only look like Rossman domains, but perform similar functions with regard to binding diphosphates. For example, both substrates of MurG contain diphosphates that play a critical role in binding. Furthermore, sequence analysis (see below) shows that MurG contains three conserved glycine-rich loops near the cleft, two located in the N-domain and one in the C-domain. Finally, MurG is not dependent on metal ions for activity. There must be some mechanism by which the anionic substrates are stabilized for binding which does not involve metal ion coordination. The glycine-rich loops in the two Rossmann domains would provide such a mechanism.

The preceding analysis combined with the two-domain structure of *E. coli* MurG suggests a binding model in which each substrate binds in a separate domain with its diphosphate in contact with at least one glycine-rich loop. The substrates are brought together for reaction across the cleft. Although we do not yet have a co-complex of *E. coli* MurG with either substrate, a comparison of the MurG structure to that of another glycosyltransferase, T4 phage -glucosyltransferase has provided considerable insight into the donor binding domain [43,44].

TOPOLOGICAL SIMILARITY TO BGT

T4 phage -glucosyltransferase (BGT) was the first NDP-glycosyltransferase ever crystallized [43,44]. The BGT structure was reported five years before any other glycosyltransferase structures appeared, but its relevance to a broader understanding of glycosyltransferase structure and mechanism was not clear. BGT has no significant homology to any other glycosyltransferase. Although it is a TDP/UDP glycosyltransferase, it cannot be classified in any of the existing glycosyltransferases families. Moreover, certain functional differences between BGT and other glycosyltransferases made BGT appear to be a special case. For example, BGT is not membrane-associated like most other glycosyltransferases. In addition, the acceptor in the BGT reaction is a segment of duplex DNA, which is larger than a typical glycosyltransferase acceptor. BGT is proposed to clasp the double helix with loops extending from both domains. A hydroxymethylcytosine residue is proposed to flip out of the helix and into the cleft between the two domains where it is glycosylated [44].

In spite of the functional differences between *E. coli* MurG and BGT, a structural comparison of MurG and BGT has revealed striking similarities. Like MurG, BGT is a two-domain protein in which both domains have an / open sheet fold. The N-terminal Rossmann domain contains a helix that originates in the C-terminal domain just as in MurG (Fig. (**4**)). Thus, the two proteins are topologically almost identical even though they share little sequence homology (11%).

The structural homology between BGT and *E. coli* MurG is particularly good in the C-domains (RMSD = 2.218Å over 89 aligned C atoms), and the similarity is even more pronounced when the invariant residues in MurG are considered. It is possible to identify the residues most critical for binding and catalysis in MurG by aligning the available MurG sequences and identifying the invariant
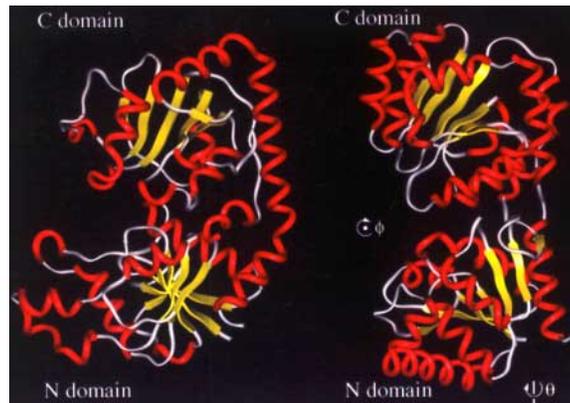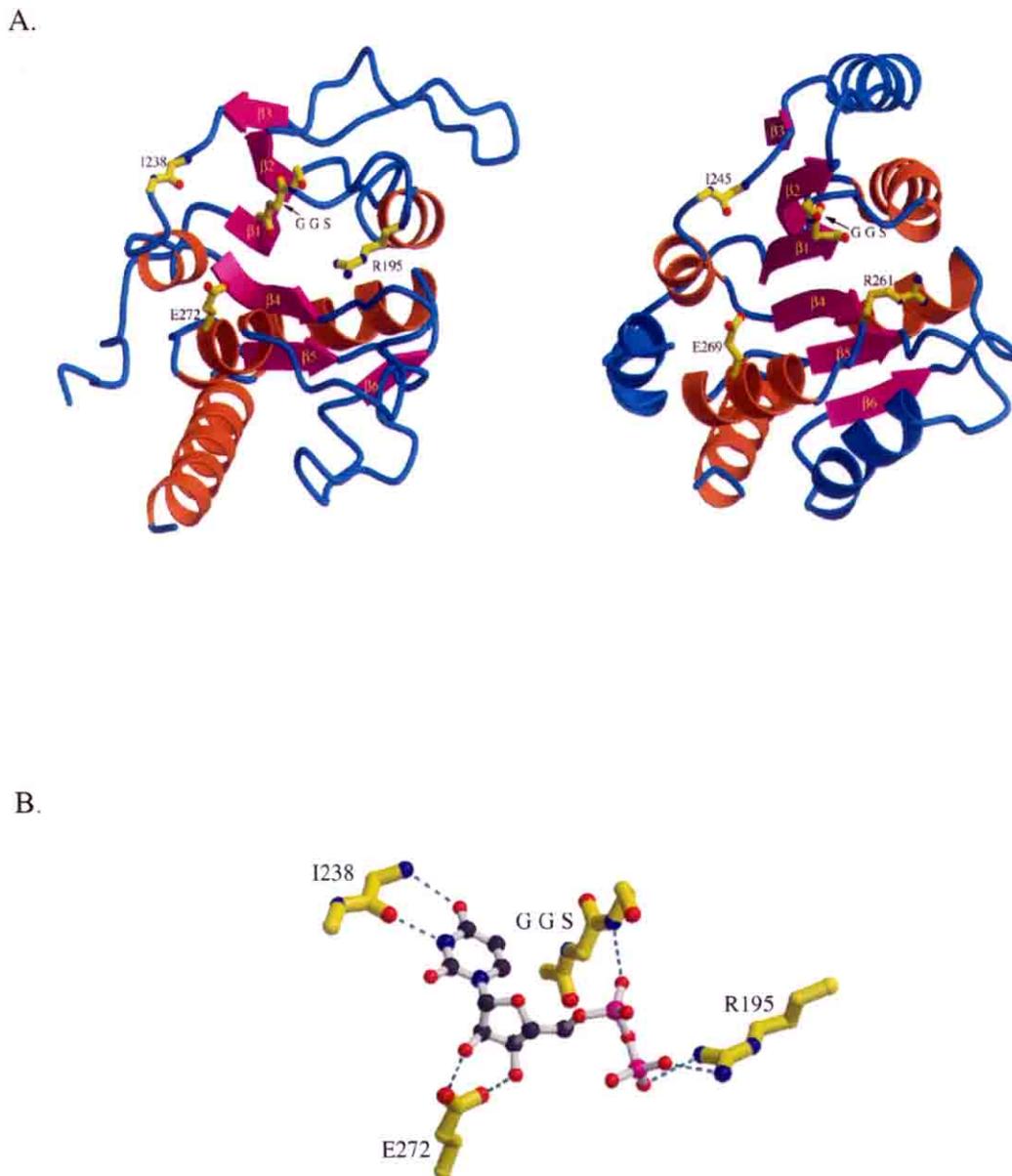
Fig. (4). Structures of T4 phage -glucosyltransferase (BGT) (left) and *E. coli* MurG (right). MurG is shown in the same orientation as BGT based on the superposition of the C-domains. In both structures, sheets are colored yellow, helices are colored red, and the rest of the structural elements are colored white. The BGT structure was solved as a co-complex with UDP (not shown). Although the two enzymes have a similar structure, BGT adopts a more closed conformation, which is facilitated by substrate binding. The MurG N-domain would require a 30° rotation and a 25° rotation to overlay with the BGT N-domain. Note that there is a long helical region at the C terminus of both proteins. This helix is continuous in BGT but interrupted between the N and C-domains in MurG.

residues. The alignment shows that a relatively small number of residues are invariant across the range of MurG homologs (Fig. (**5**)). These residues are confined to five different regions, three of which are the glycine-rich loops that are proposed to be involved in binding the diphosphates of the substrates. One of the other two regions containing invariant residues is located in the C-domain. Several of the

invariant residues in the C-domain of MurG have a counterpart at the same position in the C-domain of BGT. For example, E272 of BGT is located at the same site as invariant residue E269 in MurG; the GGS motif at the turn between 1 and 1 in BGT is found in the same location as the GGS motif in the third glycine rich loop of MurG.
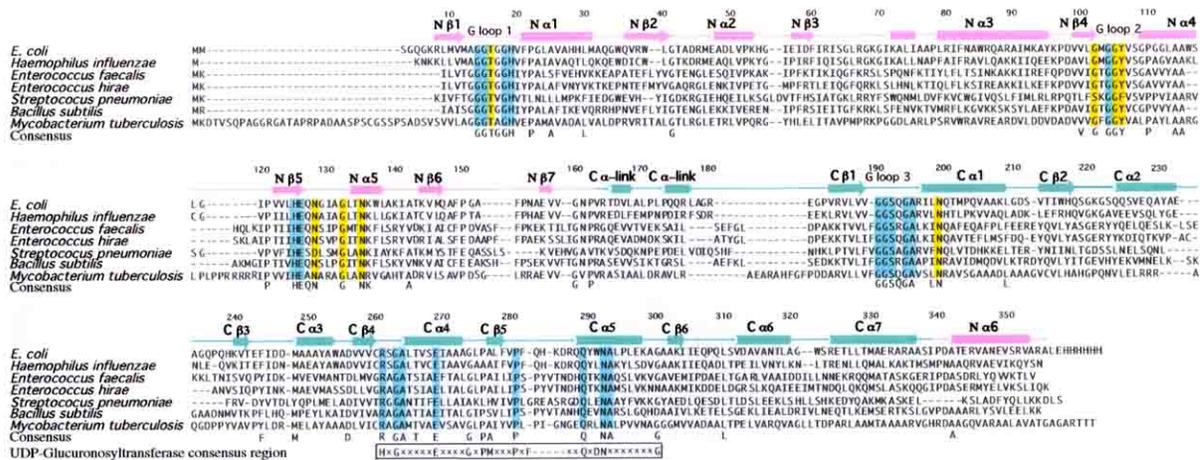
Fig. (5). Sequence alignment of *E. coli* MurG with homologs from six other bacterial strains, chosen for diversity. The secondary structure of *E. coli* MurG is shown above the alignment. Residues highlighted in yellow are identical in 85% of homologs, and in the remaining 15% only conservative amino acid substitutions are found. Residues highlighted in blue are invariant, and remaining conserved residues that do not meet the stringent criteria established for highlighting are shown in the consensus sequence. A UDP-glucuronosyltransferase consensus is also shown. Numbering corresponds to *E. coli* MurG residues.

A.

B.

Fig. (6). (A) Structures of the C-domain of T4 phage -glucosyltransferase (BGT) (left) and the C-domain of *E. coli* MurG. The aligned -strands are colored magenta, the aligned -helices are colored orange, and the other structural elements are colored blue. In BGT, key residues involved in UDP binding are highlighted in yellow. The analogous residues in MurG are also highlighted in yellow. (B) A close-up view of the binding interaction between UDP and BGT. Based on the spatially conserved residues, analogous binding interactions are proposed for MurG.

A co-complex of BGT with UDP shows that the nucleotide portion of the glycosyl donor binds near the aforementioned residues in the C-domain (Fig. (**6**)). For example, the 2' and 3' hydroxyls of the ribose sugar contact the side chain of E272. The   phosphate contacts the GGS motif. The -phosphate is located close to a positively charged residue, R195. Although there is no arginine at that position in *E. coli* MurG, there is an invariant arginine at position 261 which could perform an analogous function in contacting the   phosphate of MurG. Finally, the uracil base in BGT is held in place by hydrogen bonds from the backbone amide of isoleucine 245 in the extended strand connecting 3 to 3. There is an extended strand in the

same location in *E. coli* MurG, with an isoleucine that could hydrogen bond in a similar manner. Thus a structural comparison of the two proteins suggests a high degree of similarity in the binding pockets [39].

We have built a model of the C-domain of MurG with the nucleotide-sugar docked into a cleft in the domain and the UDP group making the same set of contacts as in BGT (Fig. (**7**)). This model provides clues to the function of some of the conserved residues in MurG. For example, as the Rossmann-like fold and the conserved glycine loop in the C domain suggest, the GGS motif is near the   -phosphate and probably plays a role in recognition of the anionic
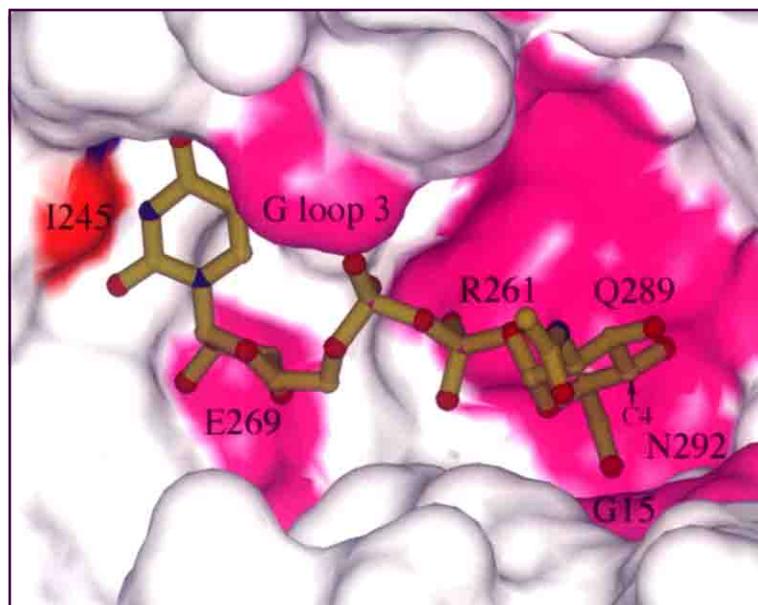
Fig. (7). A close-up view of the proposed donor binding pocket in the MurG C-domain with UDP-GlcNAc manually docked in place. The invariant residues are colored magenta. The carbonyl oxygen of I245 is shown in red and the backbone nitrogen is shown in blue.

diphosphate. R261 is close to the -phosphate and may function to stabilize the developing negative charge on the leaving group. E269 contacts the 2' and 3' hydroxyls of the ribose sugar and helps position the glycosyl donor.

Our model for how UDP binds to MurG is supported by experimental data. For example, MurG binds significantly better to UDP than to CDP, ADP or GDP. The purine nucleotides are presumably excluded from the nucleotide binding pocket on the basis of size, whereas CDP binding is disfavored because CDP presents an inappropriate pattern of hydrogen bond donors and acceptors to the binding pocket. Only pyrimidines containing a hydrogen bond acceptor at the 4 position and a donor at the 3 position are complementary to the backbone amide. Finally, mutational analysis indicates that E269 is an important residue in binding the nucleotide-sugar donor (unpublished results from this laboratory).

THE ACCEPTOR BINDING SITE

Our structural analysis suggests that the pyrophosphate portion of the acceptor contacts at least one of the two glycine-rich loops located in the N-domain near the cleft. The MurNAc sugar probably contacts residues in the loop spanning residues 125-133, which contains a number of invariant residues, including a conserved HEQN motif. With the pyrophosphate anchored near G-loop 2 and the sugar bound in the aforementioned loop, the C4 hydroxyl on the acceptor would then protrude into the cleft between domains. Examination of the substrate-free structure suggests, however, that the domains are too far apart for reaction to occur. Binding of one or both substrates may, therefore, facilitate a conformational change that brings the two domains closer together. Such a change is thought to occur in BGT [43]. The two domains in MurG are connected by a linker region (residues 158-165) and a long helix (residues

324-354). The helix is interrupted at the region between domains. A modest rearrangement of this linker could facilitate a large change in relative domain orientation.

THE MEMBRANE ASSOCIATION SITE

Many glycosyltransferases are anchored to membranes via a transmembrane spanning helix. MurG has no transmembrane helix but nevertheless associates with bacterial membranes. We have found that MurG can be removed from membranes with Triton X-100. Earlier reports suggest that membrane association is also decreased in the presence of high salt concentrations [45,46]. The crystal structure of MurG shows that the N-domain contains a concave hydrophobic patch surrounded by basic residues (Fig. (**8**)). We speculate that this hydrophobic and basic patch is the membrane association site. Consistent with this hypothesis, the crystal structure of *E. coli* MurG shows that the isooctylphenyl rings of three Triton X-100 molecules occupy the hydrophobic patch. An examination of the structure suggests that this patch could contact the membrane surface without impeding access of either the Lipid I acceptor substrate or the UDP-GlcNAc donor substrate to their respective binding sites [39]. In addition, the orientation of the Lipid I substrate with respect to entry into the binding site would be appropriate for reaction.

MurG AND OTHER GLYCOSYLTRANSFERASES

Glycosyltransferases are an enormous family of enzymes that catalyze the transfer of sugar moieties from activated donor molecules to specific acceptor molecules, forming glycosidic bonds. The most common activated donors are nucleotide sugars. The acceptors can be other sugars, lipids, proteins, DNA, or virtually any small molecule containing a hydroxyl. Glycosyltransferases are present in virtually every
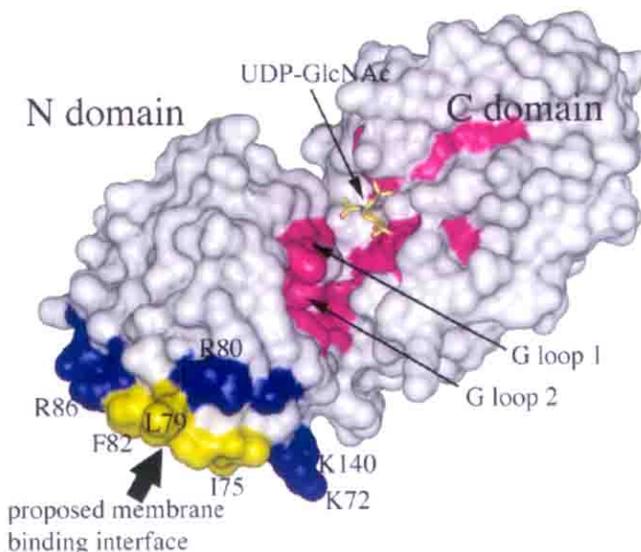
Fig. (8). The surface view of *E. coli* MurG. The G loops and other highly conserved residues are shown in magenta. The proposed membrane-binding interface is also highlighted with hydrophobic residues in yellow and positively charged residues in blue.

organism on earth. They are estimated to comprise more than 1% of proteins in cells, and they account for the bulk of the biomass produced. For a long time it has been known that glycosyltransferases play important roles in energy storage, as structural elements, in cell-cell recognition, and in intracellular trafficking [47-49]. Recently, however, it has also become clear that glycosylation events play important roles in signaling pathways [50]. This finding has intensified interest in understanding the mechanisms of glycosyltransfer and in developing strategies to inhibit glycosyltransferases.
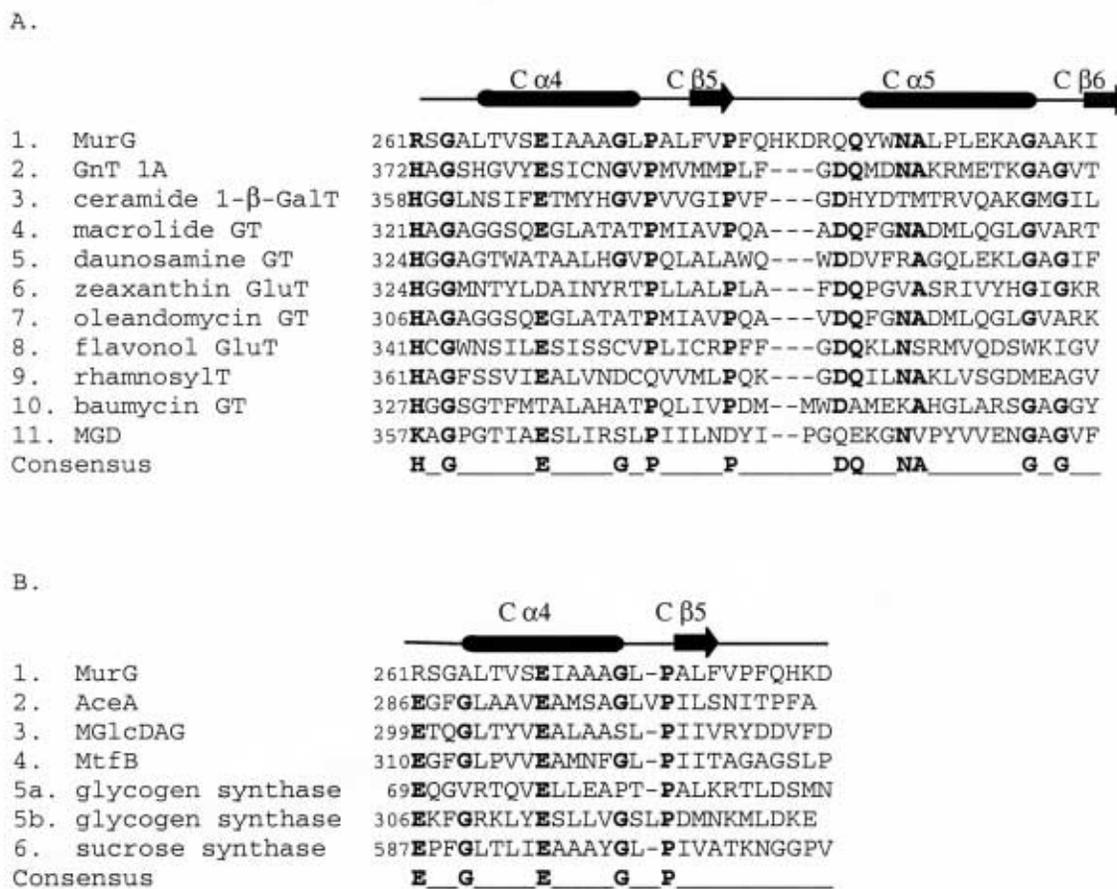
Most glycosyltransferases are membrane associated and have proven difficult to handle. Thus, structural studies on glycosyltransferases have lagged well behind studies on other large enzyme families. A dearth of structural information has impeded efforts to understand glycosyltransferases. Compounding this problem is the fact that sequence homology among glycosyltransferases is notoriously low. Therefore, it has been difficult to use sequence information to identify the active sites of glycosyltransferases or to predict their substrate specificities.

Campbell *et al.* have attempted to develop a comprehensive system to classify and compare glycosyltransferases based on sequence homologies and relative substrate/product stereochemistries [51]. In this classification scheme (also known as the CAZy scheme [52]), most glycosyltransferases have now been grouped into 52 different subfamilies. Glycosyltransferases within each family are presumed to be structurally and mechanistically similar. Unfortunately, because there were no crystal structures of any members of the 52 families until very recently, the categories were of limited utility. Kapitonov and Yu recently reevaluated the sequences of a large number of glycosyltransferases [53] and identified sets of conserved residues that allowed them to group many UDP/TDP glycosyltransferases into a large family which they called the NRD1 family (for *nucleotide recognition domain 1*). The

NRD1 glycosyltransferase family was further divided into   and   sub-groups, depending on whether the glycosyltransferases catalyze the transfer of an   -linked nucleotide sugar to an acceptor with retention of anomeric configuration or with inversion of configuration. Retaining glycosyltransferases are presumed to involve a double displacement with formation of a covalent intermediate, and inverting glycosyltransferases are presumed to involve a direct displacement of the nucleotide leaving group from the anomeric center of the donor. Kapitonov and Yu noted key similarities in the sequences of retaining and inverting glycosyltransferases. Because the acceptors utilized by the glycosyltransferases vary widely, Kapitonov and Yu concluded that the sequence homology reflects a structural similarity in the nucleotide-sugar binding domain. The region that defines the NRD1 domain is shown in Fig. (**9**). The key difference between NRD1   and NRD1 glycosyltransferases is that the former contains an EX_7E motif whereas the latter contains an R/HX_7E motif. Both NRD1   and NRD1   family members share a highly conserved pattern of prolines and glycines that contain these motifs.

According to the CAZy scheme, MurG belongs to glycosyltransferase family 28. It has been noted that family 28, which contains only inverting glycosyltransferases, bears some sequence similarities to families 3, 4, and 5, which contain only retaining glycosyltransferases.

Kapitonov and Yu have noted that MurG contains features of both the NRD1   and NRD1 glycosyltransferases. The NRD1   glycosyltransferases are primarily grouped in family 1 of the CAZy scheme. Hence, MurG shows sequence homology to members of glycosyltransferase families 1, 3, 4, 5, and other members of family 28. Our crystal structure of *E. coli* MurG provides insight into the structure of the motif that is conserved within each of these families. Fig. (**9**) shows that the NRD1

A.

```
                                    C α4            C β5             C α5           C β6
1.   MurG            261RSGALTVSEIAAAGLPALFVPFQHKDRQQYWNALPLEKAGAAKI
2.   GnT 1A          372HAGSHGVYESICNGVPMVMMPLF---GDQMDNAKRMETKGAGVT
3.   ceramide 1-β-GalT 358HGGLNSIFETMYHGVPVVGIPVF---GDHYDTMTRVQAKGMGIL
4.   macrolide GT    321HAGAGGSQEGLATATPMIAVPQA---ADQFGNADMLQGLGVART
5.   daunosamine GT  324HGGAGTWATAALHGVPQLALAWQ---WDDVFRAGQLEKLGAGIF
6.   zeaxanthin GluT 324HGGMNTYLDAINYRTPLLALPLA---FDQPGVASRIVYHGIGKR
7.   oleandomycin GT 306HAGAGGSQEGLATATPMIAVPQA---VDQFGNADMLQGLGVARK
8.   flavonol GluT   341HCGWNSILESISSCVPLICRPFF---GDQKLNSRMVQDSWKIGV
9.   rhamnosylT      361HAGFSSVIEALVNDCQVVMLPQK---GDQILNAKLVSGDMEAGV
10.  baumycin GT     327HGGSGTFMTALAHATPQLIVPDM--MWDAMEKAHGLARSGAGGY
11.  MGD             357KAGPGTIAESLIRSLPIILNDYI--PGQEKGNVPYVVENGAGVF
Consensus                H_G     E    G_P    P      DQ  NA       G_G
```

B.

```
                              C α4          C β5
1.   MurG            261RSGALTVSEIAAAGL-PALFVPFQHKD
2.   AceA            286EGFGLAAVEAMSAGLVPILSNITPFA
3.   MGlcDAG         299ETQGLTYVEALAASL-PIIVRYDDVFD
4.   MtfB            310EGFGLPVVEAMNFGL-PIITAGAGSLP
5a.  glycogen synthase  69EQGVRTQVELLEAPT-PALKRTLDSMN
5b.  glycogen synthase 306EKFGRKLYESLLVGSLPDMNKMLDKE
6.   sucrose synthase 587EPFGLTLIEAAAYGL-PIVATKNGGPV
Consensus                E__G    E    G_P
```

Fig. (9). (A) Sequence alignment of *E. coli* MurG with selected NDR1 family proteins. Conserved residues are in bold. The secondary structure of MurG is shown above the sequence. The Swiss-Prot number is given in parenthesis, except when noted. 1. MurG from *Escherichia coli* (P17443), 2. glucuronosyltransferase 1A from *Homo sapiens* (P22309), 3. ceramide 1-β-galactosyltransferase from *Homo sapiens* (O00196), 4. macrolide glycosyltransferase from *Streptomyces lividans* (Q54387), 5. daunosamine transferase from *Streptomyces peucetius* (Q54824), 6. zeaxanthin glucosyltransferase from *Erwinia ananus* (P21686), 7. oleandomycin glycosyltransferase from *Streptomyces antibioticus* (Q53685), 8. flavonol O3-glucosyltransferase from *Perilla frutescens* (O04114), 9. UDP rhamnose:anthocyanidin-3-glucoside rhamnosyltransferase from *Petunia hybrida* (Q43716), 10. baumycin glycosyltransferas from *Streptomyces sp.* C5 (Q53881), 11. monogalactosyldiacylglycerol synthase (MGD) from *Arabidopsis thaliana* (O82730). All sequences, except for MurG and MGD (family 28), are from CAZy family 1. (B) Sequence alignment of *E. coli* MurG with selected NDR1 family proteins. Conserved residues are in bold. The secondary structure of MurG is shown above the sequence. The Swiss-Prot number is given in parenthesis except when noted. 1. MurG from *Escherichia coli* (P17443), 2. AceA from *Acetobacter xylinum* (Q44571), 3. monoglucosyldiacylglycerol synthase (MGlcDAG) from *Acholeplasma laidlawii* (AF349769 - GeneBank), 4. mannosyltransferase MtfB from *Synechocystis sp.* (P74013), 5. glycogen synthase from *Homo Sapiens* (P13807) (Two possible NRD1 domains appear in this sequence.), 6. Sucrose synthase from *Spinacia oleracea* (P31928). With the exception of glycogen synthase (family 3), all members are from CAZy family 4.

motif, which is specified by a conserved pattern of prolines and glycines on which are grafted certain key residues, adopts an – – fold. The model we have presented for how the donor substrate binds suggests a function for some of the conserved residues of the NRD1 motif. In the inverting glycosyltransferases, for example, the R/H residue that precedes the first -helix of the – – fold plays a role in stabilizing the leaving group (through *e.g.*, electrostatic interactions or protonation). This interpretation is similar to that proposed by Kapitonov and Yu, who also suggested that the corresponding glutamate in retaining glycosyltransferases is the nucleophile that forms the covalent intermediate. Our model also suggests that the glutamate that follows eight residues later helps bind the

ribose sugar. In contrast, Kapitonov and Yu had proposed that this residue plays a catalytic role and deprotonates the incoming acceptor hydroxyl.

Analysis of MurG mutants supports the proposal that the glutamate in the R/HX_7E motif is involved in donor binding but not catalysis. Mutation of E269 to A or to D increases the K_m of the UDP-GlcNc substrate significantly but has only a modest effect on k_{cat} (unpublished results). Others have reported that mutation of the corresponding glutamate in AceA, a retaining UDP-mannosyltransferase which contains an EX_7E sequence characteristic of the NRD1 family, decreases but does not abolish activity [54]. Therefore, it was suggested that this second glutamate

probably plays a role in donor binding but not catalysis. In contrast, mutation of the first glutamate led to a complete loss of activity, consistent with its critical catalytic role.

One of the most surprising findings to come out of our crystallographic studies is that MurG has a very similar donor-binding site to BGT. BGT could not be grouped into any of the 52 families in the CAZy scheme. Furthermore, it does not contain a characteristic NRD1 motif. We have now shown, however, that BGT has both an overall fold and an - - subdomain similar to the conserved subdomain of MurG. We think that BGT evolved from an NRD1 glycosyltransferase, but lost the sequence identity as the loops separating the elements of secondary structure became longer. Several long loops in the N and C-domains of BGT are proposed to clasp the duplex DNA acceptor, drawing it towards the cleft between domains where the active site is located. The combination of the MurG and the BGT structures, and the sequence homologies between MurG and other NRD1 family members, is beginning to shed quite a bit of light on a large number of glycosyltransferases. (It should be noted that others have already used the MurG structure in predicting the three-dimensional fold of the mannosyltransferase AceA [54] and the bacterial lipid glycosyltransferase MGlcDAG synthase [55]).

THE DXD FAMILY OF GLYCOSYLTRANS-FERASES

Six other glycosyltransferase crystal structures have appeared in the last two years. These structures include a single member from CAZy families 2 [56], 6 [57], 7 [58], 8 [59], 13 [60], and 43 [61]. All bear remarkable structural similarity to each other, but not to MurG or BGT. These glycosyltransferases contain a characteristic DXD motif, which is believed to be part of a metal binding site that helps position the nucleotide diphosphate and which is essential for reaction. A DXD motif is also found in families 10, 12, 34, 44, and 49 [52]. All these families probably have important similarities in their mechanisms [62].

Despite the large numbers of glycosyltransferases and the limited sequence homology, the structural comparisons show that glycosyltransferases probably fall into a small number of superfamilies. So far, only two different superfamilies of glycosyltransferases have been identified. These two superfamilies evidently represent two different ways of binding nucleotide sugars and catalyzing glycosyltransfer. The superfamily to which MurG belongs is apparently not dependent on metal ions for catalysis whereas the DXD superfamily is. More structural and mechanistic information will be necessary to understand how these different families of enzymes catalyze glycosyltransfer.

FUTURE DIRECTIONS

The crystal structure of *E. coli* MurG has revealed that Nature has adapted a common nucleotide-sugar binding domain for a wide variety of different glycosylation reactions. Since a conserved supersecondary structure apparently exists

in the donor binding domain [39], and MurG is a two domain enzyme, it has been suggested that the MurG is a modular enzyme [63]. As defined by Khosla and Harbury, a modular device is a multicomponent system in which individual components can be interchanged with functionally distinct analogues from related systems. That MurG is modular is supported by the fact that eukaryotic glucuronosyltransferases genes can be spliced at different sites, resulting in a change of acceptor specificity [64,65]. The prospect of domain swapping these enzymes to mate different donor and acceptor specificity is tantalizing. If domain swapping is possible, the directed evolution of glycosyltransferases to accomplish different or unique glycosylation reactions could proceed rapidly. The ability to engineer glycosyltransferase selectivity could be useful for the synthesis of glycoconjugates and for understanding the roles of oligosaccharides in biological systems.

The structure of MurG also makes it possible to think about the rational design of glycosyltransferase inhibitors. There are very few examples of designed glycosyltransferase inhibitors, in part because the structural or mechanistic information needed to design selective glycosyltransferase inhibitors has not been available [66]. It is generally agreed that glycosyltransfer proceeds through an oxocarbenium intermediate like a glycosidase reaction [67-69], but there has not been sufficient detailed information to design transition state analogs. Glycosidase inhibitors that mimic oxocarbenium ion intermediates do not tend to inhibit glycosyltransferases very well [70]. One possible reason could be that the carboxylate present in the active site of glycosidases is not present in NDP-glycosyltransferases. Instead, the negatively charged NDP leaving group itself may stabilize the developing positive charge on the anomeric carbon. Consistent with this hypothesis, Wong and co-workers have demonstrated synergistic inhibition with a glycosidase inhibitor and a nucleotide diphosphate [71]. Inhibitors containing features that mimic the NDP group in the transition state as well as the oxocarbenium ion may be more effective.

Some efforts to design glycosyltransferase inhibitors that include pyrophosphate mimics have been reported. It has generally been assumed, however, that the diphosphate binds a divalent metal ion. For example, Wong and co-workers have investigated hexoses as pyrophosphate mimics on the assumption that the pyrophosphate-metal complex forms a chair-like structure in which the negative charges are neutralized by the metal [72]. Although one superfamily of glycosyltransferases may be metal-ion dependent, evidence is accumulating that the large group of glycosyltransferases represented by the MurG/BGT superfamily is not. The conformation of pyrophosphate in these glycosyltransferases is uncertain and may vary widely due to the inherent flexibility of the diphosphate linkage. Furthermore, in these glycosyltransferases, the charges are not neutralized by a metal ion. To design inhibitors of these glycosyltransferases, we need more structural information, including structures of co-complexes. Work on solving co-complexes of MurG with both substrates is currently going on in our laboratory and may provide information to use in inhibitor design soon.

REFERENCES

[1]   Walsh, C. *Nature*, **2000**, *406*, 775-781.

[2]   Park, J. T. *The Murein Sacculus,* in *Escherichia coli* and *Salmonella:* Cellular and Molecular Biology, ASM Press: Washington, D.C., **1996**.

[3]   van Heijenoort, J. *Murein Synthesis,* in *Escherichia coli* and *Salmonella:* Cellular and Molecular Biology, ASM Press: Washington, D.C., **1996**.

[4]   Ghuysen, J.-M. *Bacterial Cell Wall. New Comprehensive Biochemistry*, Elsevier: Amsterdam, **1994**; Vol. 27.

[5]   Bugg, T. D. H.; Walsh, C. T. *Natural Product Reports*, **1992**, 199-215.

[6]   Marquardt, J. L.; Siegele, D. A.; R., K.; Walsh, C. T. *J. Bacteriol.*, **1992**, *174*, 5748-5752.

[7]   Benson, T. E.; Marquardt, J. L.; Marquardt, A. C.; Etzkorn, F. A.; Walsh, C. T. *Biochemistry*, **1993**, *32*, 2024-2030.

[8]   Betina, V. *The Chemisty and Biology of Antibiotics*, Elsevier Scientific Publishing Company, **1983**.

[9]   Ikeda, M.; Wachi, M.; Jung, H. K.; Ishino, F.; Matsuhashi, M. *J. Bacteriol.*, **1991**, *173*, 1021-1026.

[10]   Holtje, J.-V. *Microbiol. Mol. Biol. Rev.*, **1998**, *62*, 181-203.

[11]   van Heijenoort, J. *Glycobiology*, **2001**, *11*, 25R-35R.

[12]   Schonbrunn, E.; Sack, S.; Eschenburg, S.; Perrakis, A.; Krekel, F.; Amrhein, N.; Mandelkow, E. *Structure*, **1996**, *4*, 1065-1075.

[13]   Skarzynski, T.; Mistry, A.; Wonacott, A.; Hutchinson, S. E.; Kelly, V. A.; Duncan, K. *Structure*, **1996**, *4*, 1465-1474.

[14]   Benson, T. E.; Filman, D. J.; Walsh, C. T.; Hogle, J. M. *Nat. Struct. Biol.*, **1995**, *2*, 644-653.

[15]   Benson, T. E.; Walsh, C. T.; Hogle, J. M. *Structure*, **1996**, *4*, 47-54.

[16]   Bertrand, J. A.; Auger, G.; Fanchon, E.; Martin, L.; Blanot, D.; van Heijenoort, J.; Dideberg, O. *EMBO J.*, **1997**, *16*, 3416-3425.

[17]   Bertrand, J. A.; Auger, G.; Martin, L.; Fanchon, E.; Blanot, D.; Beller, D. L.; van Heijenoort, J.; Dideberg, O. *J. Mol. Biol.*, **1999**, *289*, 579-590.

[18]   Fan, C.; Moews, P. C.; Walsh, C. T.; Knox, J. R. *Science*, **1994**, *266*, 439-443.

[19]   Yan, Y.; Munshi, S.; Leiting, B.; Anderson, M. S.; Chrzas, J.; Chen, Z. *J. Mol. Biol.*, **2000**, *304*, 435-445.

[20]   Wong, K. K.; Pompliano, D. L. *Adv. Exp. Med. Biol.*, **1998**, *456*, 197-217.

[21]   Kohlrausch, U.; Wientjes, F. B.; Holtje, J. V. *J. Gen. Microbiol.*, **1989**, *135*, 1499-1506.

[22]   Gordon, E.; Mouz, N.; Duee, E.; Dideberg, O. *J. Mol. Biol.*, **2000**, *299*, 477-485.

[23]   Lee, W.; McDonough, M. A.; Kotra, L.; Li, Z. H.; Silvaggi, N. R.; Takeda, Y.; Kelly, J. A.; Mobashery, S. *Proc. Natl. Acad. Sci. USA*, **2001**, *98*, 1427-1431.

[24]   Anderson, J. S.; Matsuhashi, M.; Haskin, M. A.; Strominger, J. L. *J. Biol. Chem.*, **1967**, *242*, 3180-3190.

[25]   Anderson, J. S.; Matsuhashi, M.; Haskin, M. A.; Strominger, J. L. *Proc. Natl. Acad. Sci. USA*, **1965**, *53*, 881-889.

[26]   Salmond, G. P. C.; Lutkenhaus, J. F.; Donachie, W. D. *J. Bacteriol.*, **1980**, *144*, 438-440.

[27]   Mengin-Lecreulx, D.; Texier, L.; van Heijenoort, J. *Nucl. Acid Res.*, **1990**, *18*, 2810.

[28]   Ikeda, M.; Wachi, M.; Jung, H. K.; Ishino, F.; Matsuhashi, M. *Nucl. Acid Res.*, **1990**, *18*, 4014.

[29]   Mengin-Lecreulx, D.; Texier, L.; Rousseau, M.; van Heijenoort, J. *J. Bacteriol.*, **1991**, *173*, 4625-4636.

[30]   Ikeda, M.; Wachi, M.; Matsuhashi, M. *J. Gen. Appl. Microbiol.*, **1992**, *38*, 53-62.

[31]   Bupp, K.; van Heijenoort, J. *J. Bacteriol.*, **1993**, *175*, 1841-1843.

[32]   Tamura, G.; Sasaki, T.; Matsuhashi, M.; Takatsuki, A.; Yamasaki, M. *Agr. Biol. Chem.*, **1976**, *40*, 447-449.

[33]   Axelrod *et al.* have developed an improved assay: Branstrom, A. A.; Midha, S.; Longley, C. B.; Han, K.; Baizman, E. R.; Axelrod, H. R. *Anal. Biochem.*, **2000**, *280*, 315-319.

[34]   van Heijenoort, Y.; Gomez, M.; Derrien, M.; Ayala, J.; van Heijenoort, J. *J. Bacteriol.*, **1992**, *174*, 3549-3557.

[35]   Auger, G.; Crouvoisier, M.; Caroff, M.; van Heijenoort, J.; Blanot, D. *Lett. Peptide Sci.*, **1997**, *4*, 371-376.

[36]   Men, H.; Park, P.; Ge, M.; Walker, S. *J. Am. Chem. Soc.*, **1998**, *120*, 2484-2485.

[37]   Ha, S.; Chang, E.; Lo, M.-C.; Men, H.; Park, P.; Ge, M.; Walker, S. *J. Am. Chem. Soc.*, **1999**, *121*, 8415-8426.

[38]   Ye, X.-Y.; Lo, M.-C.; Brunner, L.; Walker, D.; Kahne, D.; Walker, S. *J. Am. Chem. Soc.*, **2001**, *123*, 3155-3156.

[39]   Ha, S.; Walker, D.; Shi, Y.; Walker, S. *Protein Sci.*, **2000**, *9*, 1045-1052.

[40]   Branden, C.; Tooze, J. *Introduction to protein structure*, Gerland Publishing Inc.: New York, **1998**.

[41]   Rossmann, M. G.; Moras, D.; Olsen, K. W. *Nature*, **1974**, *250*, 194-199.

[42]   Hol, W. G.; van Duijnen, P. T.; Berendsen, H. J. *Nature*, **1978**, *273*, 443-446.

[43]   Vrielink, A.; Ruger, W.; Driessen, H. P.; Freemont, P. S. *The EMBO Journal*, **1994**, *13*, 3413-3422.

[44]   Morera, S.; Imberty, A.; Aschke-Sonnenborn, U.; Ruger, W.; Freemont, P. S. *J. Mol. Biol.*, **1999**, *292*, 717-730.

[45]   Taku, A.; Fan, D. B. *J. Biol. Chem.*, **1976**, *251*, 1889-1895.

[46]   Taku, A.; Fan, D. B. *J. Biol. Chem.*, **1976**, *251*, 6154-6156.

[47]   Gagneux, P.; Varki, A. *Glycobiology*, **1999**, *9*, 747-755.

[48]   Monsigny, M.; Midoux, P.; Mayer, R.; Roche, A. C. *Biosci. Rep.*, **1999**, *19*, 125-132.

[49]   Reuter, G.; Gabius, H. J. *Cell Mol. Life. Sci.*, **1999**, *55*, 368-422.

[50]   Wells, L.; Vosseller, K.; Hart, G. W. *Science*, **2001**, *291*, 2376-2378.

[51]   Campbell, J. A.; Davies, G. J.; Bulone, V.; Henrissat, B. *Biochem. J.*, **1997**, *326*, 929-939.

[52]   An excellent source for current glycosyltransferase classification. http://afmb.cnrs-mrs.fr/~pedro/CAZY/gtf.html

[53]   Kapitonov, D.; Yu, R. K. *Glycobiology*, **1999**, *9*, 961-978.

[54]   Abdian, P. L.; Lellouch, A. C.; Gautier, C.; Ielpi, L.; Geremia, R. A. *J. Biol. Chem.*, **2000**, *275*, 40568-40575.

[55]   Berg, S.; Edmanm, M.; Li, L.; Wikstroem, M.; Wieslander, A. **2001**, *in press*.

[56]   Charnock, S. J.; Davies, G. J. *Biochemistry*, **1999**, *38*, 6380-6385.

[57]   Gastinel, L. N.; Bignon, C.; Misra, A. K.; Hindsgaul, O.; Shaper, J. H.; Joziasse, D. H. *EMBO J.*, **2001**, *20*, 638-49.

[58]   Gastinel, L. N.; Cambillau, C.; Bourne, Y. *EMBO J.*, **1999**, *18*, 3546-3557.

[59]   Persson, K.; Ly, H. D.; Dieckelmann, M.; Wakarchuk, W. W.; Withers, S. G.; Strynadka, N. C. *Nat. Struct. Biol.*, **2001**, *8*, 166-175.

[60]   Unligil, U. M.; Zhou, S.; Yuwaraj, S.; Sarkar, M.; Schachter, H.; Rini, J. M. *EMBO J.*, **2000**, *19*, 5269-5280.

[61]   Pedersen, L. C.; Tsuchida, K.; Kitagawa, H.; Sugahara, K.; Darden, T. A.; Negishi, M. *J. Biol. Chem.*, **2000**, *275*, 34580-34585.

[62]   For two recent reviews on glycosyltransferase structure and mechanism, see: a) Breton, C.; Imberty, A. *Curr. Opin. Struct. Biol.*, **1999**, *9*, 563-571; b) Unligil, U. M.; Rini, J. M. *Curr. Opin. Struct. Biol.*, **2000**, *10*, 510-517.

[63]   Khosla, C.; Harbury, P. B. *Nature*, **2001**, *409*, 247-52.

[64]   Strassburg, C. P.; Oldhafer, K.; Manns, M. P.; Tukey, R. H. *Mol. Pharmacol.*, **1997**, *52*, 212-220.

[65]   Koiwai, O.; Hasada, K.; Yasui, Y.; Sakai, Y.; Sato, H.; Watanabe, T. *Biochem. Genet.*, **1995**, *33*, 111-122.

[66]   See Kim *et al.* and references therein: Kim, Y. J.; Ichikawa, M.; Ichikawa, Y. *J. Am. Chem. Soc.*, **1999**, *121*, 5829-5830.

[67]   Bruner, M.; Horenstein, B. A. *Biochemistry*, **2000**, *39*, 2261-2268.

[68]   Murray, B. W.; Takayama, S.; Schultz, J.; Wong, C.-H. *Biochemistry*, **1996**, *35*, 11183-11195.

[69]   Tvaroska, I.; Andre, I.; Carver, J. P. *J. Am. Chem. Soc.*, **2000**, *122*, 8762-8776.

[70]   Platt, F. M.; Neises, G. R.; Reinkensmeier, G.; Townsend, M. L.; Perry, V. H.; Proia, R. L.; Winchester, B.; Dwek, R. A.; Buttlers, T. D. *Science*, **1996**, *276*, 428-431.

[71]   Qiao, L.; Murray, B. W.; Shimazaki, M.; Schultz, J.; Wong, C.-H. *J. Am. Chem. Soc*., **1996**, *118*, 7653-7662.

[72]   Wang, R.; Steensma, D. H.; Takaoka, Y.; Yun, J. W.; Kajimoto, T.; Wong, C.-H. *Bioorg. Med. Chem.*, **1997**, *5*, 661-672.