# ACCELERATED COMMUNICATION

# The 1.9 Å crystal structure of *Escherichia coli* MurG, a membrane-associated glycosyltransferase involved in peptidoglycan biosynthesis

SHA HA,<sup>1</sup> DEBORAH WALKER,<sup>1</sup> YIGONG SHI,<sup>2</sup> AND SUZANNE WALKER<sup>1</sup>

<sup>1</sup>Department of Chemistry, Princeton University, Princeton, New Jersey 08544 <sup>2</sup>Department of Molecular Biology, Princeton University, Princeton, New Jersey 08544

(RECEIVED April 21, 2000; FINAL REVISION April 21, 2000; ACCEPTED April 21, 2000)

# Abstract

The 1.9 Å X-ray structure of a membrane-associated glycosyltransferase involved in peptidoglycan biosynthesis is reported. This enzyme, MurG, contains two  $\alpha/\beta$  open sheet domains separated by a deep cleft. Structural analysis suggests that the C-terminal domain contains the UDP-GlcNAc binding site while the N-terminal domain contains the acceptor binding site and likely membrane association site. Combined with sequence data from other MurG homologs, this structure provides insight into the residues that are important in substrate binding and catalysis. We have also noted that a conserved region found in many UDP-sugar transferases maps to a  $\beta/\alpha/\beta/\alpha$  supersecondary structural motif in the donor binding region of MurG, an observation that may be helpful in glycosyltransferase structure prediction. The identification of a conserved structural motif involved in donor binding in different UDP-sugar transferases also suggests that it may be possible to identify—and perhaps alter—the residues that help determine donor specificity.

Keywords: glycosyltransferase; MurG; peptidoglycan biosynthesis; X-ray crystal structure

The increasing frequency of resistance to existing antibiotics represents a serious public health threat. Structural and mechanistic information on essential bacterial enzymes could lead to the development of antibiotics that are active against resistant microorganisms. Both gram positive and gram negative bacterial cells are surrounded by a cross-linked carbohydrate polymer, peptidoglycan, which protects them from rupturing under high osmotic pressures. Many of the best antibiotics function by inhibiting peptidoglycan synthesis, which ultimately causes cell lysis. In recent years, intense effort has been focused on determining the structures of the enzymes that synthesize peptidoglycan. Structures of several of the early enzymes in the biosynthetic pathway have been reported (Fan et al., 1994; Benson et al., 1995; Skarzynski et al., 1996; Bertrand et al., 1997); however, the later enzymes have proven more difficult to study because both they and their substrates are membrane-associated.

MurG is the last enzyme involved in the intracellular phase of peptidoglycan synthesis (Bugg & Walsh, 1992). It catalyzes the transfer of N-acetyl glucosamine (NAG) from UDP to the C4 hydroxyl of a lipid-linked N-acetylmuramoyl pentapeptide (NAM) to form a  $\beta$ -linked NAG-NAM disaccharide that is transported

across the cell membrane where it is polymerized and cross-linked (Fig. 1). In bacterial cells, MurG associates with the cytoplasmic surface of the membrane (Bupp & van Heijenoort, 1993). However, we have found that *Escherichia coli* MurG can be solubilized at high concentrations in active form (Ha et al., 1999), and we now report the first X-ray crystal structure of a MurG enzyme at 1.9 Å.

# **Results and discussion**

## Overall fold

The crystal structure of *E. coli* MurG was solved by a combination of multiple isomorphous replacement and anomalous scattering and refined to 1.9 Å resolution (Table 1). The structure consists of two domains separated by a deep cleft (Fig. 2A). Both domains exhibit an  $\alpha/\beta$  open-sheet structure and have high structural homology despite minimal sequence homology (root-mean-square deviation (RMSD) = 2.02 over 85 aligned C $\alpha$  atoms). The N-domain includes residues 7–163 and 341–357 and contains seven parallel  $\beta$ -strands and six  $\alpha$ -helices, the last of which originates in the C-domain (Fig. 2B). The C-domain comprises residues 164–340 and contains six parallel  $\beta$ -strands and eight  $\alpha$ -helices, including one irregular bipartite helix ( $\alpha$ -link) that connects the N-domain to the first  $\beta$ -strand of the C-domain. The  $\beta$ -strands in both domains are ordered as for a typical Rossmann fold. The N- and C-domains

Reprint requests to: Suzanne Walker, Department of Chemistry, Princeton University, Princeton, New Jersey 08544; e-mail: swalker@princeton.edu.

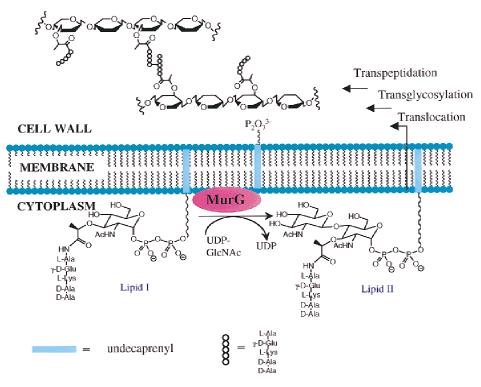


Fig. 1. Pathway for peptidoglycan biosynthesis.

are joined by a short linker between the seventh  $\beta$ -strand of the N-domain and the  $\alpha$ -link of the C-domain. This interdomain linker and the peptide segment that joins the last helix of the C-domain to the last helix of the N-domain define the floor of the cleft between the two domains. The cleft itself is about 20 Å deep and 18 Å across at its widest point. Contacts <4 Å across the cleft are limited primarily to interactions between residues from C- $\alpha$ 5 to the loop connecting N- $\beta$ 5 to N- $\alpha$ 5.

The  $\alpha/\beta$  open-sheet motif (Rossmann fold) adopted by both the N- and C-domains of MurG is characteristic of domains that bind nucleotides (Branden & Tooze, 1998). Classical Rossmann domains typically contain at least one conserved glycine rich motif, with the consensus sequence GXGXXG, located at a turn between the carboxyl end of one  $\beta$ -strand and the amino terminus of the adjacent  $\alpha$ -helix (Baker et al., 1992). This motif is involved in binding the negatively charged phosphates (Carugo & Argos, 1997). There are three glycine rich loops (G loops) in *E. coli* MurG (Fig. 3A) that may be variants on the phosphate binding loops found in other dinucleotide binding proteins (see below).

#### Sequence homology

Amino acid sequences for 18 MurG homologs are now available. The sequence similarity between *E. coli* MurG and homologs from other bacterial strains ranges from <30% to >90% depending on the evolutionary relationship between the organisms. In all MurG homologs, however, there are several invariant residues. Figure 3A shows a sequence alignment for a subset of MurG homologs with the invariant and highly conserved residues indicated. These residues, which include the three G loops, have been highlighted in the *E. coli* MurG structure (Fig. 3B). Almost all of the invariant

residues are located at or near the cleft between the two domains. Two of the G loops are found in the N domain (between N- $\beta$ 1/N- $\alpha$ 1 and N- $\beta$ 4/N- $\alpha$ 4) and one is found in the C-domain (between C- $\beta$ 1/C- $\alpha$ 1). The strict conservation of the highlighted residues among different bacterial strains, and their location as determined from the *E. coli* MurG structure, implicates them in substrate binding and catalytic activity.

#### Structural homology reveals the donor binding site

The three-dimensional backbone structure of E. coli MurG was compared to known protein structures, including the three other NDP-glycosyltransferase structures that have been reported (Vrielink et al., 1994; Charnok & Davies, 1999; Gastinel et al., 1999). The C-terminal domain was found to have significant structural homology (RMSD = 2.218 Å for 89 aligned C $\alpha$  atoms) to the C-terminal domain of phage T4  $\beta$ -glucosyltransferase (BGT), an enzyme that catalyzes the glucosylation of hydroxymethyl-cytosines in duplex DNA. A cocrystal structure of BGT with UDP bound in the C-terminal domain reveals the topology of the UDP binding pocket and also shows important contacts to the nucleotide (Vrielink et al., 1994; Moréra et al., 1999). These contacts include: (1) hydrogen bonds from the backbone amide of I238 to the N3 and O4 positions of the base; (2) hydrogen bonds between the carboxyl side chain of E272 and the O2' and O3' hydroxyls of the ribose ring; and (3) contacts from a GGS motif in the loop following the first  $\beta$ -strand of the C domain to the alpha phosphate of UDP. The structurally homologous C-domain of MurG contains a topologically similar pocket (Fig. 4A). Furthermore, even though the two domains share only 11% sequence identity overall, there are identical residues in the same spatial location in E. coli MurG

Table 1. Summary	of crystallographic and	l refinement data
------------------	-------------------------	-------------------

Data set	Native	HgCl <sub>2</sub> (form A derivative)	HgCl <sub>2</sub> (form B derivative)	$(NH_4)_2WS_4$	(NH <sub>4</sub> ) <sub>2</sub> OsBr <sub>6</sub>			
Resolution (Å)	1.9	2.0	1.9	2.4	2.3			
Observations	288,150	101,913	245,320	44,366	106,606			
Unique reflections	65,567	53,391	65,581	27,950	36,443			
$R_{sym}^{a}$ (last shell)	0.032 (0.187)	0.043 (0.200)	0.042 (0.296)	0.031 (0.080)	0.056 (0.302)			
$I/\sigma$ (last shell)	41.9 (7.0)	20.4 (2.9)	29.0 (3.7)	24.6 (8.2)	19.6 (2.5)			
Completeness (last shell)	97.7% (96.4%)	91.4% (66.6%)	97.4% (94.0%)	83.8% (62.0%)	94.3% (78.6%)			
MIR analysis (40.0–2.5 Å)								
Mean isomorphous difference <sup>b</sup>		0.163	0.130	0.068	0.134			
Phasing power <sup>c</sup> (last shell)		1.09 (0.73)	0.57 (0.50)	0.61 (0.24)	0.61 (0.58)			
$R_{\text{cullis}}^{d}$ (last shell)		0.81 (0.91)	0.94 (0.96)	0.92 (0.99)	0.94 (0.95)			
Anomalous $R_{\text{cullis}}^{d}$ (last shell)		0.96 (1.00)	0.95 (1.00)					
Refinement statistics								
Resolution	40.0–1.9 Å		RMSD <sup>e</sup>					
Reflections $( F  > 2\sigma)$	61,989		Bonds (Å)	0	.006			
Protein atoms (a. u.)	5,280		Angles (°)	1.29				
Water atoms	298		-					
Sulfate groups	1		Ramachandran plot <sup>f</sup>					
<i>R</i> -factor <sup>g</sup>	22.0%		Residues in most favored	region 94	94.6%			
R-free <sup>h</sup>	24.7%		Residues in additional allo	owed region 5	5.4%			

 ${}^{a}R_{sym} = \sum |I_i - \langle I \rangle| / \sum I_i$ , where  $I_i$  is the intensity of a reflection, and  $\langle I \rangle$  is the average intensity of that reflection.

<sup>b</sup>Mean isomorphous difference =  $\sum |F_{PH} - F_P| / \sum F_{PH}$ , where  $F_{PH}$  and  $F_P$  are the derivative and native structure factors, respectively.

<sup>c</sup>Phasing power is the ratio of the mean calculated derivative structure factor to the mean lack of closure error.

 ${}^{d}R_{cullis}$  is the mean residual lack of closure error divided by the dispersive or anomalous difference.

<sup>e</sup>RMSD, root-mean-square deviations from ideal bond lengths and bond angles.

<sup>f</sup>Calculated with program PROCHECK.

<sup>g</sup>*R*-factor =  $\sum ||F_{obs}| - |F_{calc}|| / \sum |F_{obs}|$ .

hR-free is the *R*-factor calculated using 10% of the reflection data chosen randomly and omitted from the start of refinement.

and in BGT. Based on this comparison, we have concluded that the C-domain of *E. coli* MurG is the UDP-GlcNAc binding site.

We have docked UDP-GlcNAc into the C-domain of E. coli MurG using the information on how UDP binds to BGT as a guide. As shown in Figure 4B, the uracil is held in place by contacts from the N3 and O4 atoms to the backbone amide of I245. The O2' and O3' hydroxyls on the ribose sugar are within hydrogen bonding distance of the invariant glutamate residue (E269) in the middle of helix C- $\alpha$ 4. The conserved GGS motif in G loop 3 is positioned to contact the alpha phosphate. When these contacts are made, the UDP-GlcNAc substrate fits nicely into a pocket in the C-domain, where it is surrounded by many of the invariant residues identified through sequence analysis of other MurG homologs. It is possible to propose roles for some of these invariant residues from the model. For example, the side chain of R261 can be rotated to contact the second phosphate; this contact may help stabilize the UDP leaving group. The side chain of Q289 is within hydrogen bonding distance of the C4 hydroxyl of the GlcNAc sugar. This contact may explain why MurG can discriminate between UDP-GlcNAc and its C4 axial isomer, UDP-GalNAc (Ha et al., 1999).

#### The acceptor binding site

Structural considerations suggest that the primary acceptor binding site is located in the N-terminal domain of MurG. This domain contains three highly conserved regions, two of which are glycinerich loops that face the cleft (Figs. 3A, 4C). These G loops are reminiscent of the phosphate binding loops found in other nucleotide binding proteins and are most likely involved in binding to the diphosphate on Lipid I. The N-termini of the helices following each G loop form opposite walls of a small pocket between the G loops. The helix dipoles create a positively charged electrostatic field in the pocket that can stabilize the negative charged diphosphates. When the diphosphate of the acceptor is anchored in the pocket created by the G-loops, the MurNAc sugar emerges into the cleft between domains and the C4 hydroxyl can be directed toward the anomeric carbon of the GlcNAc for attack on the face opposite the UDP leaving group. The third conserved region in the N domain spans the loop from the end of N- $\beta$ 5 to the middle of N- $\alpha$ 5. Kinetic analysis of mutants is required to evaluate the roles of these residues (Men et al., 1998; Ha et al., 1999).

#### Proposed membrane association site

MurG associates with the cytoplasmic surface of bacterial membranes where it couples a soluble donor sugar to the membrane anchored acceptor sugar, Lipid I. Analysis of the *E. coli* MurG structure shows that there is a hydrophobic patch consisting of residues I75, L79, F82, W85, and W116 in the N-domain, which is surrounded by basic residues (K72, K140, K69, R80, R86, R89). We propose that this is the membrane association site and that association involves both hydrophobic and electrostatic interactions with the negatively charged bacterial membrane. The location of this patch in MurG is also consistent with the proposed acceptor binding site: membrane association at this patch would

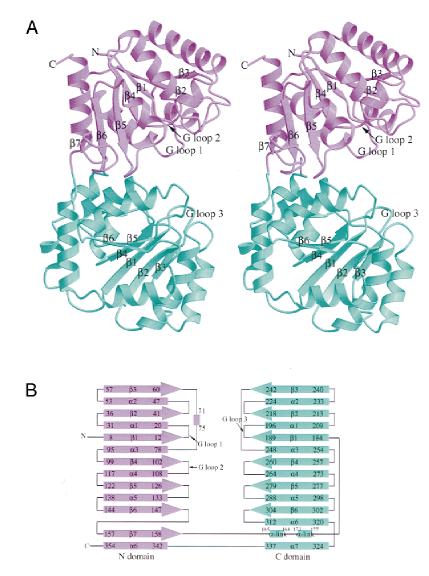


Fig. 2. Overall architecture of MurG. A: Stereoview of the MurG structure. The N domain is shown in purple; the C domain is shown in green. The figure was generated with the programs MOLSCRIPT (Kraulis, 1991) and RASTER3D (Merrit & Murphy, 1994). B: Topology diagram of MurG.

**Fig. 3** (*facing page*). Identification of critical residues in MurG and related glycosyltransferases. **A:** Sequence alignment of *E. coli* MurG with homologs from seven other bacterial strains, deliberately chosen to represent a disparate group of organisms. The secondary structure of *E. coli* MurG is shown above the sequences. Gaps mapping to the loop regions of *E. coli* MurG suggest that some sequences include other structural elements. Residues highlighted in blue are invariant among the 18 MurG sequences available. Residues highlighted in yellow are identical in 85% of the 18 homologs, while in the remaining 15%, only closely related amino acid substitutions are found. Highly conserved residues that do not meet the stringent criteria established for highlighting are shown in the consensus sequence. A consensus motif for UDP-glucuronosyltransferases is also shown. Numbering is with respect to the over-expressed *E. coli* MurG construct, which contains an additional N-terminal methionine. **B:** Mapping of the G loops and other highlighted residues from **A** in red on the MurG structure. Side chains for highly conserved residues are also shown. **C:** Model for the proposed UDP-binding subdomain found in many UDP-glycosyltransferases based on the *E. coli* MurG structure. Conserved residues in VDP-glucuronosyltransferases are highlighted in red. Side chains are shown for residues that are located near the cleft and may be involved in substrate binding. The glutamate residue is proposed to interact with the ribose sugar. The dotted loop varies in length within the MurG family and in other UDP-sugar transferases, but the N and Q on the following helix are invariant. Note that the UDP-glucuronosyltransferases contain a conserved D preceding the Q, which is not shown on this model.

мм					1 G loop 1	Ναί		Νβ2		<sup>50</sup> Ν α2	N β3			No	ພ	Ν β4	G loop	2 N
MK MK MK MK				QGKRLM KNKKLL IL KIV KII IA	VHAGGTGGH VHAGGTGGH VTGGGTGGH VTGGGTGGH FTGGGTVGH LVAGGTGGH LSGGGTGGH LAGGGTAGH	HVFPGLA HVFPAIA HIYPALS HIYPALA HVTLNLL HFFPAVA HIYPALA HVFPAMA	VAQTLQK FVEHVKK FVNYVKT LMPKFIE LGEELIK VEIKEVQR VADALVA	QEWDICW- EAPATEFL KEPNTEFM DGWEVH RGYTVHFI RHPNVEFL	-LGTKDRM YVGTENGL YVGAQRGL YIGDKRGI TDLRC YIGTENGL	EHQEILKSGL KKYINKDMK- EKKIVEREN-	-IPIRFJ -IPFKTJ -MPFRTL DVTFHSJ -IIFHIL -IPFRSJ	QISGLRG KIQGFKR EIQGFQR ATGKLRR NLKRF-SI EITGFKR	GIKALLNA SLSPQNFKT GLSLHNLKT FFSWQNMLD HIFLFLPIL GLSFENVKT	NPFAIFRAN FIYLFLTSI FIQLFLKSI DVFKVCWGI SITFLKSI FVMRFLKGN	'LQAKKI NKAKKI REAKKI VQSLFI RL 'KKSKSY	IQEEKPDAV IREFQPDVV LKEFKPDVV MLRLRPQTL IYNIKCCVI LAEFKPDAV LDDVDADVV	LGMGGYV9 IGTGGYV9 IGTGGYV9 FSKGGFV9 IGFGGYP9 IGTGGYV9 VGFGGYV9	SGPAGVA/ SGAVVYA/ SGAVVYA/ SVPPVIA/ /IAPMFA/ CGPVVYA/
12	<sup>10</sup> N β5	130 N	140 a 5	N β6	150	Nβ7	7 <sup>160</sup> C	α-link <sup>170</sup> C	α-link <sup>1</sup>	80			Cβ	81 G loop	3 20	<sup>00</sup> Cα1	<sup>210</sup> C	β <b>2</b> , <sup>220</sup>
E. coli Haemophilus influenzae Enterococcus faecalis Streptococus pirae Rickettsia prowazekii Bacillus subtilis LRIP II HEQNS YE GUNK FLARY VP VF HESDLSMG LANK TARKFARY LRIP II HEQNS YE GUNK FLARY LRIP II HEQNS YE GUNK FARFA	TC VLQA 'DKIAIC 'DRIALS TKMYST .KKIATS 'NKVAIC	F P TA F P DVAS F F E DAAP F F E QASS L S- Y E D I KNIL PE F E EAKS H	- FPHAE V - FPKEK T - FPAEK S K V E FAKSK I - FPSEK V	/VG N PM FILTG N PR SSLIG N PR /EHVG A VTI CVLTG G TV /VFTG N PR	REDLFEMP GQEVVTVE AQEVADMD KVSDQKNP RKNIRELD ASEVVSIK	N PDIR ESD K SAIL K SKIL E PDEL VDI S FMY S VSQ T GRSL	R - SEFGL QSHF HSLTKL TQTA - AEFKL	LTNTFNP	L V KGRNDI	EEKLRV DPAKKT DPEKKT NHKLPT FANSNIFT SEDKKT	VLVV-GGSQ TVVLFGGSR TVLFFGGSQ TVLFVGGSA TIFIFGGSQ TVLFFGGSQ VLLVFGGSQ	GARVLN GALKIN GALKIN GARVFN GAKLFS GAAPIN GAVSLN	HTLPKVVAQ QAFEQAFPL QAVTEFLMS QLVTDHKKE ELIPASIKI RAVIDMQDV	LADK-LEI FEEREYQ FDQ-EYQ LTER-YN LMKKQPSI LKTRDYQ	FRHQVGKI VLYASGEI VLYASGEI IINLTGD: LELNIIQI VLYITGE'			
	2 С 64	60 L C	270 α4	280 C β5		<sup>290</sup> <b>Cα5</b>	300 <b>Cβ6</b>	310	С аб <sup>32</sup>	0	<sup>330</sup> C α7	340	Να6 <sup>3</sup>	50				
EEVSQLYGE QELQESLK-LS KDIQTKVP-AC NELSQNL LDHQVKIKDIY EKVMNELKS LELRRR	NLÊ-QVH SEKKLTNIS ANVS SKLNITYI SKGAADNMY AQGDPPYY	KITEFID SVQPYID SIQPYIN RV-DYVT EFAEFFD VTKPFLH	)N-MAEAYA )K-MVEVMA (K-MAEVMA (DLYQPLME )N-IALQYK 10-MPEYLK	WADVVI NTDLMV SSDLLV LADIVV VANLVI AIDVIV	CRSGALTVO GRAGATSIA GRAGATSIA TRGGANTIF SRAGASTIE ARAGAATIA CRAGAMTVA R GA T	CEIAAVG AEFTALG AELTALG FELLAIA EELTYIG AEITALG AEVSAVG E G	AAAI FVP LPAILIP KLHVIVP KLHVIVP LPAIFIP JPSVLIP LPAIYVP F PA P	F-QH-KDR S-PYVTND S-PYVTND LGREASRGI L-PSAADN S-PYVTAN L-PI-GNG	QQYLNAKY HQTKNAQS HQTKNAMS DQLENAAY HQYYNAKL HQEVNARS EQRLNALP Q NA	LSDVGAAKII LVKVGAVEMI LVKNNAAKMI FVKKGYAEDL LEDNKAGWCL LGQHDAAIVL VVNAGGGMVV G	EQADLTP PDAELTQ KDDELDQ QESDLTL EQNNISS KETELSQ	E ILVNYLH ARLVAAII R SLKQAII D SLEEKLS EKLADKII E KLIEALI	SNLTREN DILLNNEK EIMINDQL HLLSHKED DLISNRQL DRIVLNEQT	ILLQMALKA RQQMATAS QKQMSLAS YQAKMKAS LEDAS LKEMSERT	KTMSMPI KGERIPI KQQGIPI KEL QNLLNRI KSLGVPI	NAAQRVAEV DASDRLYQV DASERMYEL KSLADF KKEGHVLLS DAAARLYSV	IKQYSN VKTLV VKSLIQK YQLLKKDL NLIEDTVF LEELKK	.5 L
											_			•				
	K	6			Glass	2					C	,	]	E				
	KK      MK      MK      MK      MK      CG	Ize    β5      L G    IP WL HEQ      C G    IP WL HEQ      C G    II HEQ      C G2    C      Q SVE QAYAE    AGQPQHI      E EVSQLYGE NLE-QVI    PLEQUYGE ANU      N ELSQNL    FI      L DHQVX IKDIYSKLNITY    FI	NK    NK      MK    MK      MK    NK      MR    NK      MR    NK      MKDTVSQPAGGRGATAPRPADAASS      120    N β5      S    N      CG	NK    MK      MK    MK      MK    MK      MK    MK      MR    MK      MKDTVSQPAGGRGATAPRPADAASPSCGSSPSA      120    N βS      N βS    N αS      140      CG	MK KIV    MK KIV    MK	$\label{eq:resonance} \begin{array}{c} K L V F I G G T G G K K K K K K K K$	MK	$ \begin{array}{c} K K K K K K K K$	$ \begin{array}{c} K_{N} & \qquad K_{N} V_{P} I G G G G G G H F P A A L L M K K K K K K K K$	MK	MK	MK  KIVFTGGG VGHVTLNLLLMVKTIEDGMEVIT-YIGDKRGLEHQELLKSGLOVTHST    MK  KILLVAGGTGGIFFPAALGELIKKVDKVHTT-YIGDKRGLEHQELLKSGLOVTHST    MKDTVSQPAGGRGATAPRPADAASPSCGSSPSADSVSVVLGGGTAGHVPAALQELKKVQRRHPNVETTJGTENGLEKKIVERKN-TPFRST    MKDTVSQPAGGRGATAPRPADAASPSCGSSPSADSVSVVLGGGTAGHVPAAMAVADALVALDPRVRITALGTLRGLETRLVPQRG-YHLELT    120  N β5  130  N σ5  140  N β6  N β7  160  Ca-link <sup>10</sup>	MX	NR	NR	$\frac{1}{120} \frac{1}{100} \frac{1}$		$\frac{1}{1000} - \frac{1}{1000} + 1$

Fig. 3. See caption on facing page.

E20

H125

G leop 2 GGY P

G

4122

Ten 6

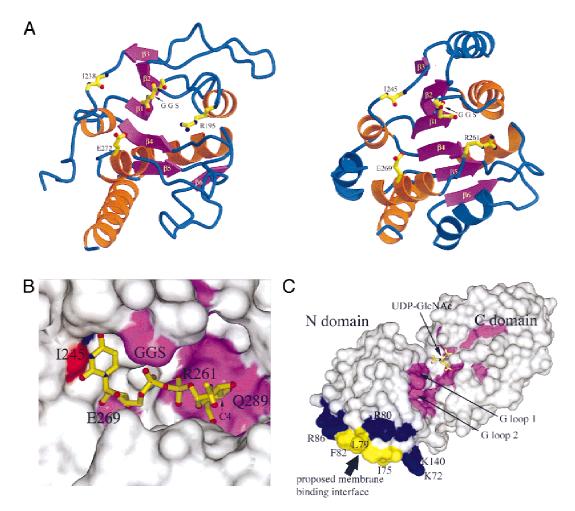


Fig. 4. Structural analysis of the substrate binding pockets in MurG. A: Structural comparison between the C-terminal domain of phage T4  $\beta$ -glucosyltransferase (left) and the C-terminal domain of *E. coli* MurG (right). The aligned six  $\beta$ -strands are magenta, the aligned  $\alpha$ -helices are orange, and the other structural elements are blue. In  $\beta$ -glucosyltransferase, 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 proposed donor binding pocket in the MurG C domain with the docked UDP-GlcNAc. Conserved residues in MurG are colored magenta. The carbonyl oxygen of residue 1245 is shown in red, and its backbone nitrogen is shown in blue. C: The surface of *E. coli* MurG. The G loops and other conserved residues in MurG are colored magenta. The proposed membrane binding interface is also highlighted with hydrophobic residues in yellow and positively charged residues in blue.

bring the two N-terminal G loops close to the membrane surface where the diphosphate portion of the acceptor is located (Fig. 4C). Moreover, the cleft between the two domains would remain accessible, consistent with the biochemical requirement that the soluble UDP-GlcNAc donor be able to find its binding site from the cytoplasm.

#### Implications for other glycosyltransferases

Glycosyltransferases that utilize an activated nucleotide sugar as a donor comprise a large family of enzymes in both prokaryotes and eukaryotes, and they play central roles in many important biological processes (Dennis et al., 1999; Koya et al., 1999; Verbert & Cacan, 1999). Glycosyltransferases are typically classified according to the nucleotide sugar they utilize, and it has frequently been noted that there is no significant sequence homology even among glycosyltransferases in the same family. This has made it difficult to identify common structural features and residues important in binding and catalysis. There are only three other glycosyltransferase structures available, and although none of them show any sequence homology to MurG, a structural comparison indicates that one of them, BGT, contains a related donor binding site.

In addition to this structural homology, we have identified a strikingly similar sequence motif in the MurG family and certain other UDP-glycosyltransferase families. This sequence motif spans about a 30 amino acid stretch in the C-domain of MurG and includes most of the invariant residues found in that domain. As shown in Figure 3A, a similar motif is found in the UDP-glucuronosyltransferases (Mackenzie, 1990; Kapitonov & Yu, 1999). Certain residues are identical, including a number of prolines and glycines, and the spacing between them is invariant. This suggests that the UDP-glucuronosyltransferases contain a region of  $\alpha/\beta$  supersecondary structure that is involved in a similar function as the corresponding region in MurG (Fig. 3C). This region binds the

donor sugar. By analyzing the similarities and differences between the conserved residues in this subdomain in the MurG family and other UDP-glycosyltransferase families, it may be possible to identify—and perhaps alter—residues that are involved in determining donor selectivity. We note that it would be useful to be able to manipulate donor specificity because it would extend the utility of glycosyltransferases as reagents for glycosylation of complex molecules, and for selective remodeling of cell surfaces (Saxon & Bertozzi, 2000).

### Conclusion

This first structure of a member of the MurG family of glycosyltransferases lays the groundwork for further mechanistic and structural investigations, which may lead to the design of inhibitors and perhaps even new antibiotics. The work also shows that there can be conserved subdomains even in very different glycosyltransferase families. Information on conserved subdomains will be useful for structure prediction and may help guide experiments directed toward changing substrate specificity.

#### Materials and methods

## Crystallization

*E. coli* MurG containing a C-terminal LEHHHHHH sequence was purified as described (Ha et al., 1999) and concentrated to 10 mg mL<sup>-1</sup> in 20 mM Tris-HCl, pH 7.9/150 mM NaCl/50 mM ethylenediaminetetraacetic acid (EDTA). The protein concentrate was mixed with UDP-GlcNAc in a 1:3 molar ratio. Crystals were grown at room temperature using the hanging-drop vapor-diffusion method by mixing equal volumes of protein with reservoir solution (0.1 M NaMES, pH 6.5/0.96 M (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>/0.4% Triton X-100/10 mM dithiothreitol). Triclinic crystals with a typical size of 0.2 mm × 0.1 mm × 0.1 mm grew within a week. The crystals belong to the P1 space group, with two molecules per asymmetric unit. The cell dimensions are a = 60.613 Å, b = 66.356 Å, c = 67.902 Å,  $\alpha = 64.294$ ,  $\beta = 83.520$ ,  $\gamma = 65.448$ .

### Data collection and processing

All data sets were collected at 100 K on previously flash frozen crystals. Crystals were equilibrated in a cryoprotectant buffer with 0.1 M NaMES, pH 6.5, 1.44 M (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 0.4% Triton X-100, and 20% glycerol. Heavy-atom soaks were carried out in the same buffer containing one of the following heavy-atom solutions: 2 mM HgCl<sub>2</sub>, 1 mM (NH<sub>4</sub>)<sub>2</sub>WS<sub>4</sub>, 1 mM (NH<sub>4</sub>)<sub>2</sub>OsBr<sub>6</sub>. Crystals were flash-frozen in liquid nitrogen. HgCl<sub>2</sub> (form A derivative) and (NH<sub>4</sub>)<sub>2</sub>OsBr<sub>6</sub> derivative data were collected at an R-AXISIIC imaging plate detector mounted on a Rigaku 200HB generator. Native, HgCl<sub>2</sub> (form B derivative), and (NH<sub>4</sub>)<sub>2</sub>WS<sub>4</sub> derivative diffraction data were collected at beam-line BioCARS-14B at the Advanced Photon Source, at wavelengths 1.0092, 0.9900, and 1.2147 Å, respectively. Collection of data on the HgCl<sub>2</sub> derivative was initially designed for MAD phasing; however, the mercury derivative proved to be unstable to X-rays, and after a 2 h exposure to synchrotron radiation the form A derivative metamorphosed into a different mercury derivative (form B) that was suitable for MIR phasing. All the data were reduced using DENZO and SCALEPACK (Otwinowski & Minor, 1997) and processed with CCP4 programs (CCP4, 1994).

#### Structure determination and refinement

The structure was solved by multiple isomorphous replacement combined with anomalous scattering of mercuric derivatives (Table 1). Initial multiple isomorphous replacement (MIR) phases calculated with program MLPHARE had a mean figure of merit of 0.44 to 2.5 Å and were improved by solvent flattening and histogram matching using density modification. A MIR map was generated that had continuous electron density for most regions of the protein. A model was built with the program O (Jones et al., 1991), and the structure was refined against 1.9 Å data using energy minimization, simulated annealing and *B*-factor refinement with the program CNS (Brünger et al., 1998). The N-terminal six residues and the C-terminal His-tag had no electron density and were not included in this model. There was no electron density for UDP-GlcNAc.

#### Acknowledgments

We thank Wayne Hendrickson, Fred Hughson, Xin Chen, and Jeff Lerman for advice and assistance. This work was supported by Incara Pharmaceuticals and the National Institutes of Health (S.W.) and start-up funds from Princeton University (Y.S.). Y. Shi is a Searle Scholar and a Rita Allen Scholar. The coordinates of the structure have been deposited in the Protein Data Bank (ID 1F0K).

#### References

- Baker PJ, Britton KL, Rice DW, Rob A, Stillman TJ. 1992. Structural consequences of sequence patterns in the fingerprint region of the nucleotide binding fold: Implications for nucleotide specificity. J Mol Biol 228:662–671.
- Benson TE, Filman DJ, Walsh CT, Hogle JM. 1995. An enzyme-substrate complex involved in bacterial cell wall biosynthesis. *Nat Struct Biol* 2:644–653.
- Bertrand JA, Auger G, Fanchon E, Martin L, Blanot D, van Heijenoort J, Dideberg O. 1997. Crystal structure of UDP-N-acetylmuramoyl-L-alanine:Dglutamate ligase from *Escherichia coli*. *EMBO J* 16:3416–3425.
- Branden C, Tooze J. 1998. Introduction to protein structure. New York: Gerland Publishing, Inc.
- Brünger AT, Adams PD, Clore GM, Delano WL, Gros P, Grosse-Kunstleve RW, Jiang J-S, Kuszewski J, Nilges N, Pannu NS, et al. 1998. Crystallography and NMR system (CNS): A new software system for macromolecular structure determination. Acta Crystallogr D 54:905–921.
- Bugg TDH, Walsh CT. 1992. Intracellular steps of bacterial cell wall peptidoglycan biosynthesis: Enzymology, antibiotics, and antibiotic resistance. *Nat Prod Rep* 199–215.
- Bupp K, van Heijenoort J. 1993. The final step of peptidoglycan subunit assembly in *Escherichia coli* occurs in the cytoplasm. *J Bacteriol* 175:1841– 1843.
- Carugo O, Argos P. 1997. NADP-dependent enzymes. I: Conserved stereochemistry of cofactor binding. *Proteins* 28:10–28.
- CCP4 (Collaborative Computing Project Number 4). 1994. The CCP4 suite: Programs for protein crystallography. Acta Crystallogr D 50:760–763.
- Charnok SJ, Davies GJ. 1999. Structure of the nucleotide-diphospho sugar transferase, SpsA from *Bacillus subtilis*, in native and nucleotide-complexed forms. *Biochemistry* 38:6380–6385.
- Dennis JW, Granovsky M, Warren CE. 1999. Glycoprotein glycosylation and cancer progression. *Biochim Biophys Acta 1473*:21–34.
- Fan C, Moews PC, Walsh CT, Knox JR. 1994. Vancomycin resistance: Structure of D-alanine:D-alanine ligase at 2.3A resolution. *Science* 266:439–443.
- Gastinel LN, Cambillau C, Bourne Y. 1999. Crystal structures of the bovine β4 galactosyltransferase catalytic domain and its complex with uridine diphosphogalactose. *EMBO J* 18:3546–3557.
- Ha S, Chang E, Lo M-C, Men H, Park P, Ge M, Walker S. 1999. The kinetic characterization of *Escherichia coli* MurG using synthetic substrate analogues. J Am Chem Soc 121:8415–8426.
- Jones TA, Zou J-Y, Cowan SW, Kjeldgaard M. 1991. Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallogr A 47:110–119.
- Kapitonov D, Yu RK. 1999. Conserved domains of glycosyltransferases. Glycobiology 9:961–978.
- Koya D, Dennis JW, Warren CE, Takahara N, Schoen FJ, Nishio Y, Nakajima T, Lipes MA, King GL. 1999. Overexpression of core 2 N-acetylglycosaminyl-

transferase enhances cytokine actions and induces hypertrophic myocardium in transgenic mice. FASEB J 13:2329-2337.

- Kraulis PJ. 1991. MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. J Appl Crystallogr 24:946–950.
- Mackenzie PI. 1990. Structure and regulation of UDP glucuronosyltransferases. In: Ruckpaul K, Rein H, eds. Frontiers in biotransformation: Principles, mechanisms and biological consequences of induction. Berlin: Akademie-Verlag, pp 211–243.
- Men H, Park P, Ge M, Walker S. 1998. Substrate synthesis and activity assay for MurG. J Am Chem Soc 120:2484–2485.
- Merrit EA, Murphy ME. 1994. Raster3D Version 2.0: A program for photorealistic molecular graphics. Acta Crystallogr D 50:869–873.
- Moréra S, Imberty A, Aschke-Sonnenborn U, Rüger W, Freemont PS. 1999. T4 phage β-glucosyltransferase: Substrate binding and proposed catalytic mechanism. J Mol Biol 292:717–730.

- Otwinowski Z, Minor W. 1997. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* 276:307–326.
- Saxon E, Bertozzi CR. 2000. Cell surface engineering by a modified Staudinger reaction. Science 287:2007–2010.
- Skarzynski T, Mistry A, Wonacott A, Hutchinson SE, Kelly VA, Duncan K. 1996. Structure of UDP-N-acetylglucosamine enolpyruvyl transferase, an enzyme essential for the synthesis of bacterial peptidoglycan, complexed with substrate UDP-N-acetylglucosamine and the drug fosfomycin. *Structure* 4:1465–1474.
- Verbert A, Cacan R. 1999. Trafficking of oligomannosides released during N-glycosylation: A clearing mechanism of the rough endoplasmic reticulum. *Biochim Biophys Acta* 1473:137–146.
- Vrielink A, Rüger W, Driessen HPC, Freemont PS. 1994. Crystal structure of the DNA modifying enzyme  $\beta$ -glucosyltransferase in the presence and absence of the substrate uridine diphosphoglucose. *EMBO J* 13:3413–3422.