

The C-glycosyltransferase IroB from Pathogenic *Escherichia coli*: Identification of Residues Required for Efficient Catalysis[†]

[†] This work was supported by Discovery Grant 341983-07 from the Natural Sciences and Engineering Research Council of Canada to PDP.

Daniel Foshag[§], Cory Campbell and Peter D. Pawelek^{✧*}

Department of Chemistry and Biochemistry, Concordia University, 7141 Sherbrooke St. W.,
Montreal, Quebec, Canada, H4B 1R6

[§] Current address: Institute for Interfacial Engineering and Plasmatechnology IGVP, University of Stuttgart. Stuttgart, Germany.

[✧] Groupe de Recherche Axé sur la Structure des Protéines (GRASP)

*Correspondence should be addressed to: Peter D. Pawelek, Tel: 514-848-2424 ext. 3118; Fax: 514-848-2868; E-mail: peter.pawelek@concordia.ca

Running title: Identification of residues required for *E. coli* IroB catalysis.

Abbreviations:

The following abbreviations are used in this manuscript: CD: circular dichroism; DGE: di-glucosylated enterobactin; C-GT: C-glycosyltransferase; DHB: 2,3-dihydroxybenzoic acid; ENT: enterobactin; ESI-MS: electrospray ionization mass spectrometry; FPLC: fast protein liquid chromatography; GE: glucosylated enterobactin; GT: glycosyltransferase; MGE: mono-glucosylated enterobactin; NTA: Ni-nitrilotriacetic acid; OD: optical density; PDB: Protein Data Bank; RMSD: root-mean-square deviation; RP-HPLC: reversed-phase high-performance liquid chromatography; SDS-PAGE: sodium dodecylsulfate-polyacrylamide gel electrophoresis; TCEP: Tris(2-carboxyethyl)phosphine; TGE: tri-glucosylated enterobactin.

Abstract:

The *E. coli* C-glycosyltransferase IroB catalyzes formation of a C-C bond between enterobactin and the glucose moiety of UDP-glucose, resulting in the production of mono-, di- and tri-glucosylated enterobactin (MGE, DGE, TGE). To identify catalytic residues, we generated a homology model of IroB from aligned structures of two similar C-glycosyltransferases as templates. Superposition of our homology model onto the structure of a TDP-bound orthologue revealed residue W264 as a possible stabilizer of UDP-glucose. D304 in our model was located near the predicted site of the glucose moiety of UDP-glucose. A loop containing possible catalytic residues (H65, H66, E67) was found at the predicted enterobactin-binding site. We generated IroB variants at positions 65-67, 264, and 304 and investigated variant protein conformations and enzymatic activities. Variants were found to have T_m values similar to wild-type IroB. Fluorescence emission spectra of H65A/H66A, E67A, and D304N were superimposable with wild-type IroB. However, the emission spectrum of W264L was blue-shifted, suggesting solvent exposure of W264. While H65A/H66A retained activity (92% conversion of enterobactin, with MGE as a major product), all other IroB variants were impaired in their abilities to glucosylate enterobactin: E67A catalyzed partial (29%) conversion of enterobactin to MGE; W264L converted 55% of enterobactin to MGE; D304N was completely inactive. Activity-impaired variants were found to bind enterobactin with affinities within 2.5-fold of wild-type IroB. Given our outcomes, we propose that IroB W264 and D304 are required for binding and orienting UDP-glucose, while E67, possibly supported by H65/H66, participates in enterobactin/MGE/DGE deprotonation.

Keywords: siderophore, enterobactin, glycosyltransferase, salmochelin, UDP-glucose.

Introduction

Siderophore-mediated iron acquisition is required for virulence by many pathogenic bacteria.¹ The *Escherichia coli* catecholate siderophore enterobactin (Fig. 1a) is not commonly a virulence factor since mammalian hosts can sequester the siderophore *via* the innate immune system protein NGAL (also known as lipocalin 2 or siderocalin). Pathogenic extraintestinal *E. coli* such as uropathogenic *E. coli* (UPEC)², avian pathogenic *E. coli* (APEC)³, the probiotic *E. coli* strain Nissle 1917⁴, as well as *Salmonella enterica*⁵, harbor a five-gene cluster known as *iroA*, which is involved in the glucosylation and linearization of enterobactin prior to its secretion. It has been shown that glucosylated enterobactin (GE) cannot be efficiently bound by NGAL.⁶ Pathogens containing the *iroA* cluster can therefore continue to acquire iron from the host while evading innate immune system defenses. The *iroA* cluster is comprised of the genes *iroB*, *iroC*, *iroD*, *iroE*, and *iroN* that encode the proteins IroB, IroC, IroD, IroE, and IroN, respectively. IroB produces GE in the cytoplasm. The inner-membrane transporter IroC facilitates transport of GE to the periplasm, where it is linearized by the esterase IroE and then secreted from the cell by a currently unknown mechanism. Upon acquisition of extracellular Fe³⁺, Fe-GE is imported into the bacterial cell by the TonB-dependent outer membrane receptor IroN. Finally, the cytoplasmic esterase IroD catalyzes the hydrolysis of the triserine trilactone core of GE, resulting in the formation of glucosylated DHB-serine subunits. This cleavage of GE facilitates the release of iron in the cytoplasm in order to support bacterial growth.

As the first *iroA* gene product required for production of GE, the 42.3 kDa C-glycosyltransferase IroB catalyzes the attachment of one, two, or three glucose moieties to enterobactin, resulting in the formation of mono-glucosylated enterobactin (MGE) (Fig. 1b), di-glucosylated enterobactin (DGE) (Fig. 1c), or tri-glucosylated enterobactin (TGE) (Fig. 1d). In each round of glucosylation, IroB transfers the glucose moiety from a bound UDP-glucose donor co-substrate to one DHB subunit of the enterobactin acceptor co-substrate. The glucose moiety is attached to a DHB subunit *via* C-C bond formation between the anomeric C1' atom of the glucose portion of UDP-glucose and the C5 atom of DHB. It has been shown that DGE, also known as salmochelin S4, is the principal product of IroB catalysis *in vivo*.⁵ The mechanistic aspects of C-glycosyltransferases are now beginning to be understood. Recent studies on glycosyltransferases that catalyze O- and C-glycosylation of hedamycin⁷ and urdamycin⁸ have explored two possible mechanisms for glucose attachment to aryl groups on acceptor substrates: in the first mechanism, glucose is first directly attached to a deprotonated phenolic group on the acceptor, followed by rearrangement of the O-glycoside; in the second mechanism, deprotonation of a phenolic group either *ortho* or *para* to the site of glucose attachment results in aromatic delocalization conferring a nucleophilic character to the carbon atom at which direct glucose attachment occurs. Experimental work on urdamycin glycosylation by UrdGT2⁸ supported the latter mechanism, in which direct glycosylation of urdamycin occurred following deprotonation of a phenolic group *ortho* to the attachment site. This direct glucosylation mechanism was supported by subsequent research on IroB by the Walsh group in which nucleophilicity of the C5 atom of DHB was suggested to be caused by deprotonation of the phenolic group on the DHB C2 carbon *para* to the C5 position of glucose attachment.⁹

Glycosyltransferase enzymes are widely found in nature, and are grouped into two families of folds: GT-A and GT-B (for a recent review, see Chang *et al.*¹⁰). Proteins of the GT-A family are single-domain folds and generally have a requirement for metals such as Mg²⁺ or Mn²⁺. Proteins from the GT-B family possess two major Rossmann-like domains: a C-terminal donor-binding domain and an N-terminal acceptor-binding domain.¹¹ The products of glycosyltransferase enzymes can either retain anomeric configuration of the sugar or invert it; however, there is no correlation between whether a GT is retaining or inverting, and whether its fold is GT-A or GT-B.¹⁰ The three-dimensional structure of IroB has not yet been reported, but the enzyme is thought to belong to the inverting GT-B family of glycosyltransferases, consistent with other C-glycosyltransferase structures.⁹ Although the nature of acceptor molecules across the GT-B family is diverse, most donor molecules are nucleotide cofactors such as TDP-glucose or UDP-glucose. In the case of IroB, the acceptor molecule can be enterobactin, MGE, or DGE. UDP-glucose has been shown *in vitro* to be a donor molecule for IroB.⁹

Here we have used homology modeling to gain structural insights into IroB substrate binding and catalysis. Our model revealed five residues (H65, H66, E67, W264, and D304) as being potentially involved in IroB catalysis. The roles of these residues were experimentally verified by enzymological assays and biophysical approaches. Taken together, our results demonstrate that IroB residues E67, W264, and D304 are necessary for efficient IroB catalysis, likely *via* their roles in UDP-glucose binding or deprotonation of DHB moieties of bound acceptor molecules. Residues H65 and H66 also appear to play a role in IroB catalysis, perhaps in the recognition of GE substrates.

Materials and Methods

Reagents

All chemicals were purchased from Bioshop Canada, Inc. (Burlington, Ontario) unless otherwise indicated. Enterobactin was purchased from Sigma-Aldrich (St. Louis, Missouri). Genomic DNA from *E. coli* CFT073 (ATCC code: 700928D-5) was obtained from the Cedarlane Labs (Burlington, Ontario).

Homology modeling

To find suitable template structures for homology-based modelling of IroB, we employed the FFAS03 server¹² (<http://ffas.ljcrf.edu/ffas-cgi/cgi/ffas.pl>) to query the Protein Data Bank for homologous structures using the predicted IroB primary amino acid sequence (Uniprot ID: Q8GH22). A multiple alignment based on the FFAS03-generated pairwise structure-based alignments of the IroB primary sequence to those of the C-glycosyltransferases UrdGT2 from *Streptomyces fradiae* (PDB code: 2P6P) and SsfS6 from *Streptomyces sp.* (PDB code: 4G2T) was used as input for Modeller v9.13.¹³ Overall model quality was evaluated using ProSA-web.¹⁴ The stereochemical validity of the IroB model structure was evaluated using PROCHECK¹⁵ via the PDBSUM server (<http://www.ebi.ac.uk/pdbsum>). Structural superpositions of resulting IroB-model with UrdGT2, SsfS6, as well as the superposition with TDP-bound CalG1 (PDB code: 3OTH) were performed using LSQMAN.¹⁶

IroB-CH6 expression construct

The *iroB* gene was PCR-amplified from *E. coli* uropathogenic strain CFT073 genomic DNA (Cedarlane Laboratories) using a forward primer (5'-

GGAATTCCATATGCGTATTTTGTGGTAGGCCCGC-3') and a reverse primer (5'-GATCGAATTCGGCCCTTCTGTACCATTGTGATCAGG -3') that resulted in the introduction of NdeI and EcoRI restriction sites (underlined) upstream and downstream of the *iroB* open reading frame, respectively. Amplification was performed using iProof DNA polymerase (Bio-Rad Laboratories) according to the protocol supplied by the manufacturer. After digestion of the PCR product and expression vector pET24b (Novagen) with *NdeI* and *EcoRI*, the insert was ligated into the vector under control of a T7 promoter in-frame with a C-terminal hexahistidine affinity tag. The resulting vector, pET24b-*iroB*-CH6, was transformed into DH5 α and identity was verified by DNA sequencing (Genome Quebec Innovation Centre, McGill University).

Site-directed mutagenesis

Site-directed mutagenesis of *iroB* was performed using the QuikChange Site-Directed Mutagenesis Kit (Stratagene) using pET24b-*iroB*-CH6 as a template along with a pair of custom-synthesized mutagenic oligonucleotide primers for each mutant generated. The mutagenic oligonucleotides were designed to substitute the following IroB residues: (i) Trp264 to Leu (W264L): forward primer (5'-GTTCGCCTGGTTGACTTGATACCCATGGGTGTC-3'), reverse primer (5'-GACACCCATGGGTATCAAGTCAACCAGGCGAAC-3'); (ii) Glu67 to Ala (E67A): forward primer (5'-GTTACCGCCACCATGCGGCACAGCGGAAAAA-3'), reverse primer: 5'-TTTTTCCGCTGTGCCGCATGGTGGCGGTAAC-3'); (iii) Asp304 to Asn (D304N): forward primer (5'-GGCCAGGGAGCCAATCGCCCGGTAA-3'), reverse primer (5'-TTACCGGGCGATTGGCTCCCTGGCC-3'); (iv) His65 to Ala and His66 to Ala (H65A/H66A): forward primer (5'-CCGAAGCTGGTTACCGCGCCGCTGAGGCACAGCGGAAAAA-3'),

reverse primer (5'-TTTTCCGCTGTGCCTCAGCGGCGCGGTAACCAGCTTCGG-3').

Mutagenesis reaction mixtures (50 μ l) contained: pET24b-IroB-CH6 template DNA (~ 2 μ g/ μ l), mutagenic forward and reverse primers (0.5 μ M each), dNTP mixture (0.2 mM each of dATP, dTTP, dCTP, and dGTP), and PfuTurbo DNA polymerase (0.05 U/ μ l). The number, duration, and temperature of the PCR cycles were performed according to the manufacturer's instructions. Prior to transformation into DH5 α , parental vectors were digested with DpnI. Double-stranded plasmids obtained from site directed mutagenesis were purified and variant identities were verified by DNA sequencing (Genome Quebec Innovation Centre, McGill University).

Expression and purification of IroB-CH6 and variant proteins

Expression constructs (pET24b-IroB-CH6) containing DNA sequences encoding in-frame C-terminal hexahistidine-tagged wild-type IroB (henceforth referred to as IroB-CH6) or IroB-CH6 variants (W264L, E67A, D304N, H65A/H66A) were transformed into competent *E. coli* BL21-DE3 cells (Novagen). Transformants were grown in 2xYT broth containing 30 μ g/ml of kanamycin at 37 °C to an OD₆₀₀ of ~0.6. Protein expression was induced with the addition of IPTG to a final concentration of 0.4 mM followed by further incubation in a platform shaker (220 RPM) at 20 °C for 20 hours. Cells were pelleted by centrifugation (6,000 g) and stored at -20 °C until required. For purification, cell pellets from 1 L culture were resuspended in 30 ml ice-cold Buffer A (20 mM Tris (pH 8.5), 500 mM NaCl, 20 mM imidazole, 500 mM sucrose, 2 mM TCEP) supplemented with Protease Inhibitor Cocktail (BioShop, Inc.). Cell lysis was achieved after three passes in a French pressure cell operated at 16,000 to 18,000 lb in⁻². Cell lysates were kept cold during this procedure. Insoluble matter and cell debris were removed by ultracentrifugation at 130,000 g for 1 h at 4 °C. Clarified lysates were applied to a 10 ml Ni-NTA

Superflow (Qiagen) column (pre-equilibrated with Buffer A) connected to a BioLogic DuoFlow FPLC system (Bio-Rad Laboratories). The column was extensively washed using Buffer B (20 mM Tris (pH 8.5), 25 mM KCl, 40 mM imidazole, 1M sucrose, and 1 mM TCEP). The washed hexahistidine-tagged proteins were eluted using a stepwise imidazole gradient (40 mM to 250 mM) in Buffer B. Fractions containing the highest OD₂₈₀ values were analyzed by SDS-PAGE, pooled, and dialyzed against Buffer C (20 mM Tris (pH 8.5), 25 mM KCl, 1M sucrose, 1 mM TCEP) followed by storage at -80 °C after direct beading into liquid nitrogen. Protein concentrations in this study were determined spectrophotometrically by measuring OD₂₈₀ and using molar extinction coefficients predicted from primary amino acid sequences ($\epsilon_{280} = 56,000 \text{ M}^{-1} \text{ cm}^{-1}$ for IroB-CH6, D304N, E67A, H65A/H66A; $\epsilon_{280} = 50,500 \text{ M}^{-1} \text{ cm}^{-1}$ for W264L).

Circular dichroism spectroscopy

Far-UV CD spectra of purified IroB-CH6 and variant protein samples were collected on a Jasco J-815 circular dichroism spectropolarimeter. Purified proteins were dialyzed into Buffer C and then diluted to a concentration of 4.7 μM in the same buffer. Diluted samples were loaded into a rectangular cell with 0.1 cm path length and spectra were recorded at 20 °C by averaging five wavelength scans (1 nm bandwidth) from 260 nm to 200 nm (step size: 0.2-nm, scan rate: 20 nm/min). Spectra were corrected for buffer contributions, smoothed and signals were converted to molar ellipticities using the instrument software package. Final spectra were exported and plotted using OriginLab (version 8.6).

For thermal stability experiments, ellipticities at 222 nm were monitored between 20 °C and 98 °C (rate of temperature change: 0.5 °C min⁻¹) using a Jasco J-815 CD spectropolarimeter with a temperature-controlled cell holder. Thermal denaturation data were normalized in terms of

fraction of folded protein at each temperature value by assigning the ellipticity value at 20 °C to 100% folded (fraction folded = 1) and the value at 60 °C to 100% unfolded (fraction folded = 0). The melting temperature (T_m) was determined from thermal denaturation curves by using the first derivative function of the instrument's software package, where maxima between 20 °C and 60 °C was used to determine T_m values. All CD spectra of protein solutions were corrected for contributions of buffer components using spectra collected in the absence of protein.

Fluorescence spectroscopy

Fluorescence spectroscopy was used to assess the tertiary structure of purified IroB-CH6 and variants. Spectra were recorded at room temperature using a Varian Cary Eclipse spectrofluorometer with the Varian Scan program (version 1.1). For fluorescence experiments, protein concentrations were diluted to avoid inner filter effects such that OD_{280} values were always below 0.05. Purified IroB-CH6 and variant protein concentrations were adjusted to 1 μ M in Buffer A. Samples were loaded into a 1 cm Varian fluorescence cuvette, and fluorescence emission spectra were collected from 300 to 400 nm using excitation wavelengths of 280 nm or 295 nm. All spectra were recorded in CAT mode (reported spectra are averages of 10 scans each), excitation and emission slit width were set to 5 nm, the voltage and scan speed adjusted to 600 V and 600 nm min^{-1} , respectively. Spectra were corrected for buffer contribution by subtraction of signal from spectra collected on protein-free samples using the instrument's calculation software. Scan data were exported and analyzed using OriginLab.

IroB-CH6 activity assay

The HPLC-based IroB activity assay used in this study was adapted from Fischbach *et al.*⁹ IroB-CH6 and variant proteins were exchanged into Buffer D (75 mM HEPES (pH 7.5), 50% glycerol

(v/v), 2.5 mM TCEP) using a DG10 desalting column, and concentrations of exchanged proteins were determined spectrophotometrically at 280 nm using molar extinction coefficients predicted from primary amino acid sequences. Enzymatic reactions were performed in a 100 μ l volume containing 2 μ M IroB-CH6 or variant protein, 100 μ M enterobactin (Sigma), 600 μ M UDP-glucose (Sigma), in Buffer E (75 mM Tris (pH 7.5), 5 mM MgCl₂, 2.5 mM TCEP). Reactions were allowed to proceed for 60 minutes at 25 °C, followed by immediate quenching upon loading onto a reversed-phase C18 column equilibrated with 5% ACN (v/v) in dH₂O and connected to an Agilent HPLC (flow rate: 1 ml min⁻¹). After sample loading, unbound material was washed away for 10 min in 5% (v/v) ACN in dH₂O. Reaction products were monitored at 254 nm and separated in a gradient from 5-80% (v/v) ACN in 0.05% (v/v) TFA in dH₂O over 20 min. Assays were performed in duplicate. Chromatograms were exported and plotted using OriginLab software. Reactants and products (enterobactin, MGE, DGE, TGE) separated by RP-HPLC were analyzed by ESI mass spectrometry at the Centre for Biological Applications of Mass Spectrometry (Concordia University). Eluted species were directly injected into a CapLC-QToF2 mass spectrometer (Waters Micromass) and masses were analyzed in the positive mode.

Fluorescence-based enterobactin-binding assay

A fluorescence-based enterobactin-binding assay modified from Lin *et al.*¹⁷ was used to estimate binding affinity constants for enterobactin binding to IroB-CH6 and variant proteins. Purified proteins were thawed and rapidly diluted to a concentration of 100 nM into Buffer F (75 mM Hepes (pH 7.5), 5 mM MgCl₂, 2.5 mM TCEP) in a Varian fluorescence cuvette (path length = 1 cm). Fluorescence emission at 340 nm was recorded at 20 °C using the simple read mode upon excitation at 280 nm. Excitation and emission slit widths were set to 10 nm, and detector voltage

was set to 600 V. Enterobactin (3 mM in DMSO) was added to final concentrations of 2, 4, 6, 8, and 10 μM , followed by rapid mixing and equilibration for 30 sec. Fluorescence emission intensities at 340 nm were measured and corrected for buffer and DMSO contributions.

Fluorescence intensities were corrected for inner filter effects according to the equation:

$$F_{\text{corr}} = F_{\text{obs}} \times 10^{((\text{OD}_{\text{ex}} + \text{OD}_{\text{em}})/2)} \quad (1)$$

where F_{obs} is the observed fluorescence intensity, OD_{ex} is the observed absorbance at the excitation wavelength, OD_{em} is the observed absorbance at the emission wavelength.

The ratio of initial fluorescence intensities to fluorescence intensities at given enterobactin concentrations were plotted as a function of enterobactin concentration according to the Stern-Volmer¹⁸ equation:

$$F_0/F = 1 + K_{SV} [\text{ENT}] \quad (2)$$

where F_0 is the corrected fluorescence emission intensity at 340 nm of the protein in the absence of enterobactin, F is the corrected fluorescence emission intensity at 340 nm in the presence of enterobactin, K_{SV} is the Stern-Volmer constant, and $[\text{ENT}]$ is the molar concentration of enterobactin. Quenching experiments were performed in triplicate for each protein. Data obtained from fluorescence quenching experiments were imported into Kaleidagraph 4.5 and fits were obtained by linear regression. Slopes of linear fits were used to determine K_{SV} values.

Results

Homology modeling of *E. coli* IroB reveals candidate residues involved in catalysis

In order to locate the IroB active site in the absence of structural data, we used homology modeling to predict the three-dimensional structure of the protein. We identified crystallized orthologues as candidate templates for homology modeling *via* the PDB search function of the

FFAS03 server.¹² Our FFAS03 search reported the C-glycosyltransferase UrdGT2 (PDB code: 2P6P)¹⁹ and the C-glycosyltransferase SsfS6 from *Streptomyces sp.* (PDB code: 4G2T)²⁰ as the C-GT orthologues with the highest degree of structural alignment. We therefore used 2P6P and 4G2T as templates for homology modeling using Modeller v9.13 and obtained an IroB model that superimposed well with the structures of 2P6P and 4G2T (superposition RMSD values: 1.00 Å and 0.593 Å, respectively) (Fig. 3a). Evaluation by PROCHECK¹⁵ and ProSA-Web¹⁴ indicated that the model was of acceptable quality and stereochemically correct. The predicted structure of our IroB homology model (Fig. 3b) exhibits a two-domain GT-B fold characteristic of C-glycosyltransferases of known structure.^{19,20} The N- and C-terminal domains of the model possess a similar architecture with central β -sheets sandwiched by α -helices, and are connected by a linker region between residues 185 and 197. The C-terminal domain is typical of a Rossmann fold for binding nucleotide cofactors, and likely comprises the UDP-glucose binding site of IroB. In order to gain further insights into the substrate-binding regions of IroB, we compared our model to the O-glycosyltransferase CalG1 from *Micromonospora echinospora* (PDB code: 3OTH), which was co-crystallized with a bound TDP nucleotide cofactor. The CalG1 primary structure aligned to that of *E. coli* IroB with an overall sequence identity of 28% (Fig. S1). The crystal structure of CalG1 exhibits a two-domain GT-B fold with a Rossmann-like C-terminal domain, similar to our predicted IroB structure.²¹ The 3OTH structure is comprised of CalG1 in complex with TDP, a nucleotide cofactor that shares chemical similarities with UDP-glucose. We superimposed our IroB model structure onto that of the CalG1-TDP complex in order to locate the UDP-glucose binding site in IroB. LSQMAN superposition of the C-terminal domain of our IroB model onto that of the CalG1 structure resulted in an overall RMSD of 1.547 Å and a normalized RMSD(100) of 0.987 Å. Part of the TDP-binding region of CalG1 between

279-WVPQ-282 was found to structurally align with IroB residues 264-WIPM-267 (Fig. 3c). In the CalG1 structure, residue W279 is involved in a stacking interaction with the thymidine base of TDP. IroB residue W264 superimposes with CalG1 W279, suggesting that W264 can form a similar stacking interaction with the uridine base of UDP-glucose. Given the predicted position of the binding site for UDP-glucose in our IroB model, we were able to infer the location of the acceptor-binding site, which would be proximal to the glucose moiety of UDP-glucose, at the cleft between the N- and C-terminal domains. IroB residue D304 is found at this cleft in our model, at a position that suggests a possible role in stabilizing the glucose moiety of the nucleotide cofactor. Our model also indicated the presence of a hydrophobic cluster near D304, consisting of IroB residues L10, L13, and L110 (not shown in Fig. 3), that could be involved in a hydrophobic interaction with the bound acceptor. Near this hydrophobic cluster, we found a long loop between residues 56-69 that could potentially close over the predicted acceptor-binding site. Possible catalytic residues on this loop that could participate in deprotonation of acceptor DHB moieties were identified: H65, H66 and E67. These three residues, as well as the predicted UDP-glucose-binding residues D304 and W264, were changed by site-directed mutagenesis to gain insights into their roles in IroB catalysis and substrate binding. Three of the five residues were altered to produce single-site IroB variant proteins: E67A, W264L, D304N. The histidine residues at positions 65 and 66 were changed to produce a double-site IroB variant protein: H65A/H66A.

Expression and purification of IroB-CH6 and variant proteins

We initially attempted to purify IroB-CH6 according to a previously reported approach⁹, which led to heavy IroB-CH6 precipitation at higher concentrations. We were able to optimize IroB-

CH6 solubility and structural integrity by the addition of the following stabilizing osmolytes (with final concentrations following in parentheses): glycerol (30-50% (v/v)), sucrose (1 M), or trehalose (1 M). Under these conditions, no visible precipitation of IroB was observed up to a concentration of 3 mg ml⁻¹. Far-UV circular dichroism spectroscopy (not shown) of IroB solutions indicated that the presence of any of these osmolytes resulted in enhanced long-term stability and solubility of IroB-CH6. Our final purification protocol yielded approximately 20 mg of recombinant IroB-CH6 per liter of culture, and this optimized protocol was also used to purify the IroB-CH6 variants that we generated. Purities of IroB-CH6 and variant preparations were found to be comparable, and were estimated to be ≥85% based on densitometry following SDS-PAGE (not shown).

Structural integrity of IroB-CH6 and variant proteins

To assess overall protein fold qualities, we examined our purified IroB-CH6 and variant proteins using several biophysical methods that are sensitive to secondary structure (far-UV circular dichroism (CD)), overall stability (thermal denaturation measured by far-UV CD) and tertiary structure (fluorescence spectroscopy). Our far-UV CD analyses (Fig. 4a) indicate that IroB-CH6 and variant proteins were folded with similar secondary structure contents. The spectrum of IroB-CH6 (WT) (Fig. 4a, black line) has a positive peak at around 200 nm and double minima indicative of alpha-helical content at 208 nm and 222 nm, with the minimum at 222 nm being more prominent. Spectra of all IroB variant proteins exhibited similar characteristics in comparison with the IroB-CH6 spectrum, with minor variations observed in the 208-222 nm region. The variant D304N (Fig. 4a, magenta line) possesses a slightly more pronounced minimum at 208 nm compared to its 222-nm minimum.

Thermal denaturation experiments were performed to compare the overall thermal stabilities of variant proteins to that of IroB-CH6. Thermal denaturation curves (Fig. 4b) showed maximal spectral transitions around 35 °C - 39 °C, followed by a stable minimal ellipticity signal at 60 °C. Additional ramping of temperature from 60 °C to 85 °C did not further alter this signal, so we concluded that maximal denaturation of all proteins occurred by 60 °C. The transition midpoint of wild-type IroB (black line) occurred around 39 °C. To determine transition mid-points, first derivatives of the melting curves were calculated where the maximum of the first derivative was taken as the melting temperature (T_m). These results are summarized in Table 1, which shows that all T_m values are similar to the wild-type T_m of 39.4 °C, with the H65A/H66A double variant having the lowest T_m at 35.2 °C. We therefore concluded that the overall folds of the four variant proteins examined were not adversely affected in comparison with the wild-type protein.

Fluorescence emission spectra

Fluorescence spectroscopy was used to further probe the tertiary structures of IroB-CH6 and variant proteins. The fluorescence emission spectra of wild-type IroB-CH6 upon excitation at 280 nm (Fig. 5a, black dashed trace) and 295 nm (Fig. 5a, black solid trace) exhibited emission maxima at 345 nm, suggesting that at least some of the eight tryptophan residues in wild-type IroB-CH6 are found in a polar environment, and are likely solvent-exposed. The fluorescence emission spectra (dashed traces: $\lambda_{ex} = 280$ nm; solid traces: $\lambda_{ex} = 295$ nm) of the IroB variants D304N (Fig. 5a, magenta traces), E67A (Fig. 5a, red traces) and H65A/H66A (Fig. 5a, green traces) all superimpose with those of IroB-CH6. In contrast, peaks of the fluorescence emission

spectra of W264L (Fig. 5b, blue traces) exhibited a blue shift relative to those of IroB-CH6 (Fig. 5b, black traces), consistent with W264 occurring in a more polar environment.

Enzymatic activities of IroB-CH6 and variant proteins

Based on our homology modelling and bioinformatics analyses, we predicted that IroB residues H65, H66, and E67 could potentially function as catalytic residues involved in deprotonation, while D304 and W264 could potentially be involved in UDP-glucose binding. We therefore incubated wild-type IroB-CH6 and our four variant proteins under identical assay conditions to compare levels of enzymatic activities. Under these conditions, we expected the wild-type enzyme to completely convert enterobactin predominantly to DGE, which has been shown to be the predominant biological product.⁵ We employed an HPLC-based activity assay modified from Fischbach *et al.*⁹ to measure the relative activities of IroB-CH6 and variant proteins. Reactant and product (enterobactin, MGE, DGE, TGE) identities were confirmed by inline ESI-MS analysis of eluted species from a quenched IroB-CH6 reaction mixture separated by RP-HPLC. Good correspondences (within 0.5 Da) were observed between expected and observed masses for all species, thus confirming their identities (Table 2). Chromatographic traces of IroB-CH6 and variant assay reactions quenched at 60 min are shown in Figure 6, where the relative abundances of reactants and products were calculated by integration of areas and shown above respective chromatographic peaks. As expected, wild-type IroB-CH6 converted 100% of the enterobactin substrate to DGE (88%) and TGE (12%) after 60 min (Fig. 6a). W264L, which our model predicted to stack with the uridine base of the UDP-glucose donor, was only able to convert 55% of total enterobactin substrate to MGE (45%) and DGE (10%) (Fig. 6b). The D304N variant, which we predicted to be proximal to the enterobactin-binding site and the

glucose moiety of UDP-glucose, was found to be completely inactive in our assay (Fig. 6c). The variant E67A, which our model predicted to be located on a loop near the enterobactin-binding site, was able to partially convert enterobactin to MGE (29%), leaving 71% of the substrate unconverted (Fig. 6d). Finally, we found the double variant H65A/H66A to be active, converting 92% of enterobactin to glycosylated products; however, in contrast to IroB-CH6, the double variant produced more MGE (60%) than DGE (32%) (Fig 6e).

Enterobactin binds to IroB variants

IroB variants W264L, D304N, E67A and H65A/H66A exhibited partial or complete loss of enzymatic activity with respect to enterobactin conversion. We therefore wanted to determine if the observed decreases in activities were due to an inability of the variants to bind the substrate enterobactin. Enterobactin has been previously shown to statically quench IroB intrinsic fluorescence upon binding, such that the quenching effect can be used to measure enterobactin binding to IroB.¹⁷ Here we titrated IroB-CH6 and variant proteins with pure enterobactin up to a concentration of 10 μM . Over this concentration range, increases in F_0/F as a function of [enterobactin] were observed to be linear. Quenching data were fit by linear regression, with R^2 values of 0.988 or better being obtained in each case. For static quenching systems, such as the binding of enterobactin to IroB, the slope of the linear fit (*i.e.*, the Stern-Volmer constant (K_{SV})) correlates with quencher binding affinity.¹⁸ Using this approach, we found that IroB-CH6 and all variant proteins were able to bind enterobactin with affinities within 2.5-fold of IroB-CH6 (Fig. 7). Binding of enterobactin to wild-type IroB-CH6 resulted in a K_{SV} value of 0.16 μM^{-1} (Fig. 7a). The variants W264A and D304N, residues that we predict to be involved in UDP-glucose binding, were found to have the lowest K_{SV} values, and thus the lowest affinities: 0.080 μM^{-1} and

0.069 μM^{-1} , respectively (Fig. 7b-c). The variants E67A and H65A/H66A, predicted to be on a mobile loop that closes over the IroB active site, were found to bind enterobactin with affinities ($K_{SV} = 0.14 \mu\text{M}^{-1}$ and $0.12 \mu\text{M}^{-1}$, respectively) similar to IroB-CH6 (Fig 7d-e).

Discussion

The enterobactin-modifying enzyme IroB from *E. coli* belongs to a relatively rare class of glycosyltransferases that attach glucose moieties to acceptor molecules *via* formation of C-C bonds. Previous studies on IroB have shown that it catalyzes the formation of MGE followed by conversion of MGE to DGE, with relatively little TGE being produced *in vivo*.^{5,9} Little is currently known about the mechanism of this enzyme, including knowledge of specific residues involved in substrate binding and catalysis. This lack of data may be in part due to the difficult nature of isolating this protein under conditions optimal for *in vitro* studies. Our expression and purification of recombinant IroB-CH6 was initially hampered by its tendency to partition in the insoluble phase upon centrifugation of cell lysates, and to precipitate above $\sim 0.3 \text{ mg ml}^{-1}$ during purification. We found that IroB-CH6 solubility and conformational stability were greatly improved in the presence of a number of different osmolytes, including glycerol, trehalose, and sucrose. Since we found that these osmolytes enhanced the folded state of the protein, we used sucrose (0.5-1.0 M) or relatively high concentrations of glycerol ($\sim 50\%$) to assist in stabilizing purified IroB-CH6 and variant proteins *in vitro*.

Such difficulties inherent in the handling of IroB *in vitro* may be informative of the protein's intracellular context. For example, our homology model indicates a region of high surface hydrophobicity around Val 49 that is surrounded by basic residues, suggesting that IroB may associate with the *E. coli* inner membrane *via* hydrophobic and electrostatic interactions.

This is not unprecedented. The *E. coli* O-glycosyltransferase MurG, for example, is involved in peptidoglycan biosynthesis and has been shown to associate with the inner membrane by a surface hydrophobic region surrounded by basic residues.²² Glycosylated enterobactin (GE) may similarly be directly exported from the cytoplasm *via* direct interaction between a membrane-anchored IroB and the IroC transporter. Furthermore, the relatively low *in vitro* thermostability of IroB-CH6 that we found in our melting experiments ($T_m \sim 39$ °C) suggests that an important intracellular stabilizer of IroB may be removed upon purification of the protein. It has been reported that *E. coli* enterobactin biosynthetic enzymes such as EntC and EntF have enhanced activities in the presence of macromolecular crowding agents such as Ficoll, a sucrose polymer, suggesting that these proteins are responsive to the physiological environment of the cytoplasm.^{23,24} Perhaps proteins of the functionally-related *iroA* cluster, such as IroB, are similarly responsive to their intracellular contexts. We are now further exploring this possibility.

In order to find IroB catalytic residues, given that no structure of *E. coli* IroB has been reported, we constructed a homology model of IroB by using two recently solved bacterial C-glycosyltransferase structures as templates: UrdGT2 from *S. fradiae* (PDB code: 2P6P), and SsfS6 from *Streptomyces sp.* (PDB code: 4G2T). Although IroB has only ~ 25% identity to each template structure, our use of two structural templates in the modeling of IroB, coupled by rigorous model evaluation using various scoring algorithms, led us to conclude that our model had acceptable accuracy. Superposition of our IroB model with the *M. echinospora* O-glycosyltransferase CalG1 (PDB code: 3OTH) allowed us to locate the UDP-glucose (donor) binding site, which in turn provided us with valuable information regarding IroB residues likely at the enterobactin/MGE/DGE (acceptor) binding site. At the predicted donor-binding site, we found IroB residue W264 to be structurally conserved with CalG1 W279, which forms a stacking

interaction with the thymidine base of co-crystallized TDP in the 3OTH structure. We therefore hypothesized that IroB W264 was involved in stabilization of the uridine base of UDP-glucose. In support of this, our fluorescence spectroscopy experiments revealed a blue shift in the W264L emission maximum compared to wild-type IroB, suggesting that W264 is solvent-exposed, consistent with its predicted location at the surface of the UDP-glucose binding site. Our W264L variant exhibited reduced (45%) activity compared to IroB-CH6 even though our biophysical characterizations indicated that this variant was well folded and could bind enterobactin with wild-type affinity. Taken together, these data support the hypothesized role of W264 in binding/positioning UDP-glucose as predicted by our homology model, thus contributing to validation of our IroB model.

Based on the assumption that IroB W264 is involved in UDP-glucose binding, we followed the orientation of the superimposed TDP cofactor from 3OTH in order to gain insights into the binding site of the acceptor molecule. Proximal to the predicted glucose moiety of UDP-glucose lies IroB residue D304, which is located at the cleft of the N- and C-terminal domains. Given its location, D304 could participate in coordinating the glucose moiety of UDP-glucose, or it could participate in proton abstraction from a DHB group on the bound acceptor. Many structures of glycosyltransferase enzymes have been reported to have a catalytic aspartate residue at the interdomain cleft of the GT fold, with the aspartate participating in a catalytic dyad with a nearby histidine residue.²¹ Our IroB model indicates the presence of H138 within hydrogen-bonding distance of D304, but an H138A variant that we generated retained full enzymatic activity (data not shown). Furthermore, the UrdGT2 catalytic dyad comprised of Asp 137 and the backbone amide of Val139 is located within the UrdGT2 interdomain cleft, but the C_{β} - C_{β} distance between IroB D304 and UrdGT2 D137 in the superimposed structures is 8.86 Å,

demonstrating that these residues are therefore not at structurally-conserved positions. To investigate a possible role for D304 in IroB catalysis, we changed this residue to asparagine. We observed that the D304N variant had no detectable activity under our assay conditions, even though D304N was found to be well folded and could bind enterobactin. It is therefore likely that the role of IroB D304 is in stabilizing the glucose moiety of UDP-glucose during catalysis. In support of this, Offen *et al.* demonstrated that D374 in the O-glycosyltransferase VvGT1 (PDB code: 2C1Z) from *Vitis vinifera* is involved in coordinating the glucose moiety of the sugar donor by hydrogen bonding to 3-OH and 4-OH of glucose *via* the D374 carboxyl group.²⁵ The D374A mutation in VvGT1 reduced catalytic activity to a non-detectable level, consistent with what we observed in our IroB D304N variant activity assay. Amino acid sequence alignment of VvGT1 with IroB shows that VvGT1 D374 and IroB D304 share a position of sequence identity (Fig. S1). We therefore propose that D304 is involved in the orientation of the glucose moiety of UDP-glucose rather than in acceptor deprotonation. To investigate this further, we are currently working towards optimizing our purification and IroB stabilization methods to allow for higher protein concentrations that would facilitate direct calorimetric binding assays of UDP-glucose to IroB-CH6 in comparison with the W264L and D304N variants.

If D304 is indeed positioned to stabilize glucose, then our model predicts that IroB E67 would likely be a base catalyst participating in deprotonation of DHB groups on the bound acceptor. In our model, E67 is located on a loop between IroB residues 56 and 69. This region of IroB is not conserved among GT orthologues (Fig. S1), providing support that E67 is indeed found in a loop. The lack of sequence conservation is also consistent with its proposed role in acceptor molecule interaction, given the heterogeneity of GT acceptor molecules. Based on its position in our IroB model relative to the predicted acceptor-binding site, we hypothesize that the

56-69 loop containing E67 is mobile, being able to close over the acceptor molecule upon its binding. Prior experimental data reported on similar enzymes support a catalytic role for such a loop in IroB. Hoffmeister and coworkers identified a region between residues 52 and 82 that confers substrate specificity for the conserved C-glycosyltransferases UrdGT1b and UrdgGT1c.^{26,27} Furthermore, the UrdGT2 structure (PDB code: 2P6P) revealed a structurally conserved loop comprised of UrdGT2 residues 62-72 that is adjacent to the acceptor-binding site. This loop has been proposed to close over the acceptor molecule upon binding so that a base catalyst (either UrdGT2 E66 or E71) could fulfill its function.¹⁹ Similar to the position of UrdGT2 E71, the location of IroB E67 in a predicted mobile loop in our model supports its role as a base catalyst in the deprotonation of bound acceptor. We found that mutation of E67 to alanine had a severe impact on IroB catalysis: 70% of enterobactin substrate remained unconverted, and only MGE could be produced. As with D304N, we found the E67A variant to be properly folded and it could bind to enterobactin with wild-type affinity. This demonstrates that the E67A mutation exclusively affects catalytic function. Fischbach *et al.* reported that the IroB mechanism is distributive, with accumulation of MGE followed by subsequent accumulation of DGE, and then finally of TGE. This study further suggested that carbanion formation at the C5 of the acceptor DHB moiety, where C-glycosylation occurs, requires prior deprotonation of the DHB C2-OH group, which is *para* to C5.⁹ Residue E67 may be involved in deprotonation at the DHB C2-OH position since the C_β-C_β distance between E67 and D304 is 10.9 Å, with sufficient intervening space for a DHB moiety to be situated with its C2 atom closer to E67 and its C5 atom closer to D304. Contiguous to IroB E67 are the residues H65 and H66, also on the predicted mobile loop in our model. We generated the double variant H65A/H66A to test the roles of these residues in IroB catalysis. The impact of these mutations on catalytic activity was moderate in comparison

to E67A: the H65A/H66A variant is still able to produce MGE (60%) and DGE (32%) upon converting ~ 90% of the enterobactin substrate. Given these data, and the positions of H65 and H66 in our homology model, these residues could be involved in stabilizing E67 in a deprotonated state, or perhaps enhancing the catalytic function of E67 by formation of H-bonds with other IroB residues upon loop closure. Furthermore, since E67A retained partial activity, it is possible that the nearby H65 residue, which also has its side chain oriented toward the predicted acceptor-binding site, could substitute as a less efficient base catalyst for DHB deprotonation in the absence of E67.

In support of our proposed roles for IroB residues W264, D304, H65, and H66, and E67, we determined that variant proteins with mutations at these positions could all bind enterobactin (Fig. 7). We found that the poorest binders of enterobactin, W264L and D304N, had affinities that were approximately 2.5-fold lower than that of IroB-CH6. Interestingly, these are the two residues that we propose to be involved in binding UDP-glucose. It is possible that W264 and D304 may also directly interact with the enterobactin acceptor, especially D304 given its predicted location at the interdomain cleft. However, it is also possible that subtle conformational rearrangements caused by the W264L and D304N mutations at the UDP-glucose-binding site are linked to conformational rearrangements at the enterobactin-binding site, resulting in the lower observed acceptor-binding affinity. We are now further investigating possible conformational linkage between IroB donor and acceptor binding sites. In any event, the extent of quenching that we observed for IroB-CH6 is consistent with the previously reported K_D for enterobactin binding to IroB (2.3 μM).¹⁷ Since the concentration of enterobactin used in our activity assays was 100 μM , a saturating concentration given our experimentally-determined enterobactin-binding data, the losses of enzymatic activities that we observed for our IroB

variants (which all had K_{SV} values within ~2.5-fold of IroB-CH6) were not due to lowered affinities for enterobactin binding.

In conclusion, we have identified three residues as being necessary for efficient IroB catalysis: E67, W264, and D304 (Fig. 8). Mutations at these positions negatively affected enzymatic activities, but these losses of activity were not due to protein misfolding, nor to impaired enterobactin binding. Taken together, our homology model and experimental data indicate that E67 is likely involved in acceptor (enterobactin/MGE/DGE) deprotonation at the phenolic group on the C2 carbon of the DHB moiety *para* to the site of glucose attachment at C5, while W264 and D304 are likely involved in donor (UDP-glucose) orientation to facilitate IroB catalysis.

Acknowledgements

The authors thank Dr. Paul Joyce, Dr. Ekkehard Hiller, and Paknoosh Pakarian for helpful discussions regarding this manuscript. D.F. received financial support from a GRASP Recruitment Award. The authors acknowledge the Concordia BIOFINS platform for access to its CD spectropolarimeters, and CBAMS for access to its mass spectrometers.

References

1. M. Miethke, M.A. Marahiel, Siderophore-based iron acquisition and pathogen control, *Microbiol. Mol. Biol. Rev.* 71 (2007) 413-451.
2. R.E. Watts, M. Totsika, V.L. Challinor, A.N. Mabbett, G.C. Ulett, J.J. De Voss, M.A. Schembri, Contribution of siderophore systems to growth and urinary tract colonization of asymptomatic bacteriuria *Escherichia coli*, *Infect. Immun.* 80 (2012) 333-344.
3. M. Caza, F. Lépine, S. Milot, C.M. Dozois, Specific roles of the *iroBCDEN* genes in virulence of an avian pathogenic *Escherichia coli* O78 strain and in production of salmochelins, *Infect. Immun.* 76 (2008) 3539-3549.
4. M. Valdebenito, S.I. Müller, K. Hantke, Special conditions allow binding of the siderophore salmochelin to siderocalin (NGAL-lipocalin), *FEMS. Microbiol. Lett.* 277 (2007) 182-187.
5. B. Bister, D. Bischoff, G.J. Nicholson, M. Valdebenito, K. Schneider, G. Winkelmann, K. Hantke, R.D. Süssmuth, The structure of salmochelins: C-glucosylated enterobactins of *Salmonella enterica*, *Biometals* 17 (2004) 471-481.
6. M. A. Fischbach, H. Lin, L. Zhou, Y. Yu, R. J. Abergel, D. R. Liu, K. N. Raymond, B. L. Wanner, R. K. Strong, C. T. Walsh, A. Aderem, K. D. Smith, The pathogen-associated *iroA* gene cluster mediates bacterial evasion of lipocalin 2, *Proc. Natl. Acad. Sci. USA* 103 (2006) 16502-16507.
7. T. Bililign, C.G. Hyun, J.S. Williams, A.M. Czisny, J.S. Thorson, The hedamycin locus implicates a novel aromatic PKS priming mechanism, *Chem. Biol.* 11 (2004) 959-969.

8. C. Dürr, D. Hoffmeister, S.-E. Wohler, K. Ichinose, M. Weber, U. von Mulert, J.S. Thorson, A. Bechthold, The Glycosyltransferase UrdGT2 Catalyzes Both C- and O-Glycosidic Sugar Transfers, *Angew. Chem. Int. Ed.* 43 (2004) 2962-2965.
9. M.A. Fischbach, H. Lin, D.R. Liu, C.T. Walsh, *In vitro* characterization of IroB, a pathogen-associated C-glycosyltransferase, *Proc. Natl. Acad. Sci. USA* 102 (2005) 571-576.
10. A. Chang, S. Singh, G.N. Phillips Jr., J.S. Thorson, Glycosyltransferase structural biology and its role in the design of catalysts for glycosylation, *Curr. Opin. Biotechnol.* 22 (2011) 800-808.
11. J. Liu, A. Mushegian, Three monophyletic superfamilies account for the majority of the known glycosyltransferases, *Protein Sci.* 12 (2003) 1418-1431.
12. L. Jaroszewski, L. Rychlewski, Z. Li, W. Li, A. Godzik, FFAS03: a server for profile-profile sequence alignments, *Nucleic Acids Res.* 33 (2005) W284-W288.
13. A. Šali, T.L. Blundell, Comparative protein modelling by satisfaction of spatial restraints, *J. Mol. Biol.* 234 (1993) 779-815.
14. M. Wiederstein, M.J. Sippl, ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins, *Nucleic Acids Res.* 35 (2007) W407-W410.
15. R.A. Laskowski, M.W. MacArthur, D.S. Moss, J.M. Thornton, PROCHECK: a program to check the stereochemical quality of protein structures, *J. Appl. Cryst.* 26 (1993) 283-291.
16. G.J. Kleywegt, Use of non-crystallographic symmetry in protein structure refinement, *Acta Cryst. D: Biol. Cryst.* 52 (1996) 842-857.
17. H. Lin, M.A. Fischbach, G.J. Gatto, D.R. Liu, C.T. Walsh, Bromoenterobactins as potent inhibitors of a pathogen-associated, siderophore-modifying C-glycosyltransferase, *J. Am. Chem. Soc.* 128 (2006) 9324-9325.

18. J.R. Lakowicz, Principles of Fluorescence Spectroscopy, 3rd Ed. (2006) Springer.
19. M. Mittler, A. Bechthold, G.E.Schulz, Structure and action of the C-C bond-forming glycosyltransferase UrdGT2 involved in the biosynthesis of the antibiotic urdamycin, J. Mol. Biol. 372 (2007) 67-76.
20. F. Wang, M. Zhou, S. Singh, R.M. Yennamalli, C.A. Bingman, J.S. Thorson, G.N. Phillips, Crystal structure of SsfS6, the putative C-glycosyltransferase involved in SF2575 biosynthesis, Proteins 81 (2013) 1277-1282.
21. A. Chang, S. Singh, K.E. Helmich, R.D. Goff, C.A. Bingman, J.S. Thorson, G.N. Phillips, Complete set of glycosyltransferase structures in the calicheamicin biosynthetic pathway reveals the origin of regiospecificity, Proc. Natl. Acad. Sci. USA 108 (2011) 17649-17654.
22. S. Ha, D. Walker, Y. Shi, S. Walker, The 1.9 Å crystal structure of *Escherichia coli* MurG, a membrane-associated glycosyltransferase involved in peptidoglycan biosynthesis, Protein Sci. 9 (2000) 1045-1052.
23. M. Jiang, Z. Guo, Effects of macromolecular crowding on the intrinsic catalytic efficiency and structure of enterobactin-specific isochorismate synthase, J. Am. Chem. Soc. 129 (2007) 730-731.
24. Z.-F. Guo, M. Jiang, S. Zheng, Z. Guo, Structural change of the enterobactin synthetase in crowded solution and its relation to crowding-enhanced product specificity in nonribosomal enterobactin biosynthesis, Bioorg. Med. Chem. Lett. 20 (2010) 3855-3858.
25. W. Offen, C. Martinez-Fleites, M. Yang, E. Kiat-Lim, B.G. Davis, C.A. Tarling, C.M. Ford, D.J. Bowles, G.J. Davies, Structure of a flavonoid glucosyltransferase reveals the basis for plant natural product modification, EMBO J. 25 (2006) 1396-1405.

26. D. Hoffmeister, K. Ichinose, A. Bechthold, Two sequence elements of glycosyltransferases involved in urdamycin biosynthesis are responsible for substrate specificity and enzymatic activity, *Chem. Biol.* 8 (2001) 557-567.

27. D. Hoffmeister, B. Wilkinson, G. Foster, P.J. Sidebottom, K. Ichinose, A. Bechthold, Engineered urdamycin glycosyltransferases are broadened and altered in substrate specificity, *Chem. Biol.* 9 (2002) 287-295.

Table 1. Thermal stabilities of IroB-CH6 and variant proteins.

Protein	T _m (°C) ^a
IroB-CH6	39.4
W264L	39.6
E67A	39.6
D304N	37.2
H65A/H66A	35.2

^a T_m values were determined as the peak of the first derivative of thermal denaturation curves collected between 20 °C and 60 °C.

Table 2. LC-MS analysis of IroB-CH6 assay species eluted by RP-HPLC.

Compound ^a	Expected Mass ([M+H ⁺] in Da)	Observed Mass ([M+H ⁺] in Da)
ENT	670.6	670.2
MGE	832.7	832.3
DGE	994.8	994.4
TGE	1156.9	1156.4

^a ENT=enterobactin, MGE=mono-glucosylated enterobactin,
DGE=di-glucosylated enterobactin, TGE=tri-glucosylated enterobactin.

Figure Legends

Fig. 1. Chemical structures of enterobactin and IroB enzymatic products. (a) enterobactin; (b) MGE: monoglucosylated enterobactin; (c) DGE: diglucosylated enterobactin; (d) TGE: triglucosylated enterobactin.

Fig. 2. Multiple alignment of *E. coli* IroB to C-glycosyltransferases of known structure.

4G2T: SsfS6 from *Streptomyces sp.*, 2P6P: UrdGT2 from *Streptomyces fradiae*. Pairwise structure-based alignments of each protein sequence to that of IroB were initially obtained from the FFAS03 server and then combined to form the multiple alignment. Black background and white letters indicate positions of sequence identity; grey background and black letters indicate positions of sequence similarity. Alignment figure generated using BoxShade (http://ch.embnet.org/software/BOX_form.html).

Fig. 3. Homology-based modelling of *E. coli* IroB and identification of possible catalytic

residues. (a) Superposition of IroB homology model (red) onto template structures 2P6P (blue) and 4G2T (yellow). Protein mainchains shown as smooth ribbons; (b) IroB homology model shown in cartoon representation (coils: α -helices, flat arrows: β -sheets). Protein model is colored from N-terminus (blue) to C-terminus (red). Residues mutagenized in this study are indicated: lines point to positions of C α atoms for each residue; (c) superposition of CalG1 (green) TDP-binding site (PDB code: 3OTH) with predicted IroB (magenta) UDP-glucose binding site. Labels indicating residues and TDP ligand from CalG1 are terminated with a ' symbol. Carbonyl oxygen atoms are indicated by the suffix "-CO".

Fig. 4. Circular dichroism spectroscopy of IroB-CH6 and variant proteins. (a) Far-UV CD spectra of IroB-CH6 and variants: black: IroB-CH6, blue: W264L, red: E67A, magenta: D304N, green: H65A/H66A. Voltages are shown in inset using same color scheme. (b) Thermal denaturation curves of IroB-CH6 and variants: traces colored as in part (a).

Fig. 5. Fluorescence emission spectra of IroB-CH6 and variant proteins. (a) Fluorescence emission scans ($\lambda_{\text{ex}} = 280 \text{ nm}$): dotted traces, fluorescence emission scans ($\lambda_{\text{ex}} = 295 \text{ nm}$): solid traces. Black: IroB-CH6, red: E67A, magenta: D304N, green: H65A/H66A; (b) Fluorescence emission scans ($\lambda_{\text{ex}} = 280 \text{ nm}$): dotted traces; fluorescence emission scans ($\lambda_{\text{ex}} = 295 \text{ nm}$): solid traces. Black: IroB-CH6, blue: W264L.

Fig. 6. Enzymatic assays of IroB-CH6 and variant proteins. Assay mixtures were initiated with the addition of IroB-CH6 or variant protein (2.5 μM). All reactions proceeded for 60 min under identical assay conditions. Reactants and products were separated by RP-HPLC, and chromatographic traces are shown here. Reactant and product identities were assigned according to retention times in comparison with those of standard IroB-CH6 assay analyzed by LC-MS (see Table 2). Peak identities, determined by mass spectrometry, are labeled. Relative abundances determined by peak area integration are shown as percentages. (a) IroB-CH6; (b) W264L; (c) D304N; (d) E67A; (e) H65A/H66A.

Fig. 7. Binding of enterobactin to IroB-CH6 and variant proteins. Stern-Volmer plots indicate loss of protein intrinsic fluorescence due to enterobactin binding. For each experiment, 100 nM of protein was titrated with increasing concentrations of enterobactin. Samples were

excited at 280 nm and fluorescence emission was measured at 340 nm. F_0/F = the ratio of protein fluorescence emission in the absence of enterobactin to protein fluorescence emission at the indicated enterobactin concentration. Solid lines represent fits of quenching data by linear regression. K_{SV} = Stern-Volmer constant, corresponding to the slope of the linear fit. All binding experiments were performed in triplicate. Mean F_0/F values are shown, and error bars represent standard deviations from the mean values. (a) IroB-CH6; (b) W264L; (c) D340N; (d) E67A; (e) H65A/H66A.

Fig. 8. Schematic of IroB residues W264, D304, and E67 in relation to bound co-substrates.

Approximate positions of IroB W264 (blue), D304 (magenta), E67 (red) are shown. Arrow indicates hypothesized role of deprotonation of DHB C2-OH by E67. Colored arcs are oriented towards regions of UDP-glucose that we propose interact with respectively colored residues. Specific moieties on co-substrates predicted to directly interact with W264, D340, and E67 are shaded with respective coloring. DHB C2 and C5 are labeled, along with C1' of glucose. 'R' indicates the remaining atoms of the bound acceptor (enterobactin/MGE/DGE) molecule.

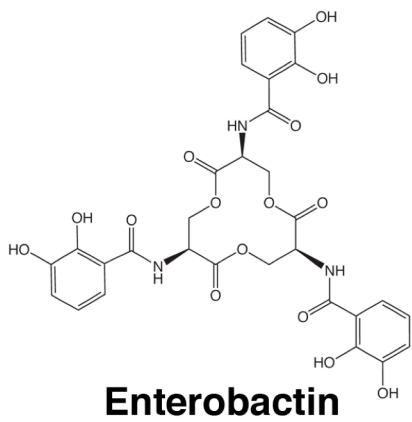
Fig. S1 Multiple sequence alignment of *E. coli* IroB to glycosyltransferase orthologues.

Alignment generated using PSI-COFFEE (<http://tcoffee.org.cat/apps/tcoffee/do:psicoffee>) and further processed with Jalview (<http://www.jalview.org>). IroB orthologues detected by HHpred (<http://toolkit.tuebingen.mpg.de/hhpred>). Residues are shaded according to the degree of conservation from yellow to red. The conserved binding motif for the glycosyl donor is boxed in black. 2P6P: UrdGT2 from *Streptomyces fradiae*, 3IAA: CalG2 from *Micromonospora echinospora*, 2IYA: OleI from *Streptomyces antibioticus*, 3IA7: CalG4 from *Micromonospora*

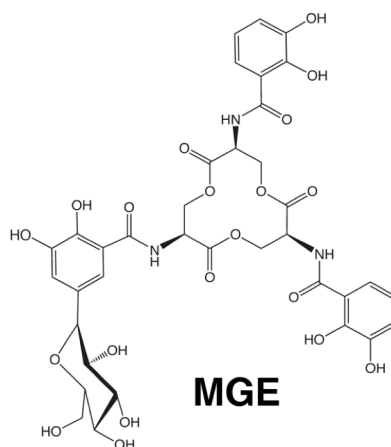
echinospora, 3H4T: chimeric protein engineered from GtfA (*Amycolatopsis orientalis*) and Orf1 (*Actinoplanes teichomyceticus*), 2IYF: OleD from *Streptomyces antibioticus*, 3OTH: CalG1 from *Micromonospora echinospora*, 1RRV: GtfD from *Amycolatopsis orientalis*, 3OTI: CalG3 from *Micromonospora echinospora*, 3TSA: SpnG from *Saccharopolyspora spinosa*, 2C1Z: flavonoid glucosyltransferase from *Vitis vinifera*.

Figure 1.

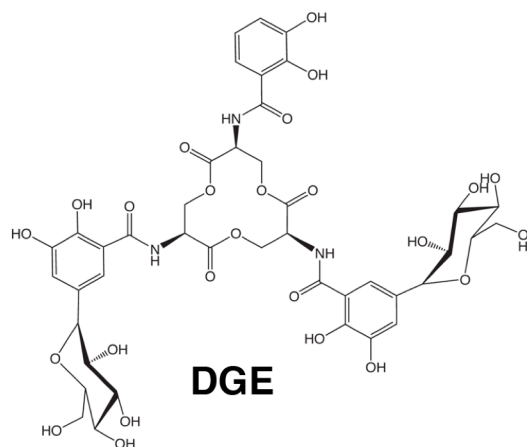
(a)



(b)



(c)



(d)

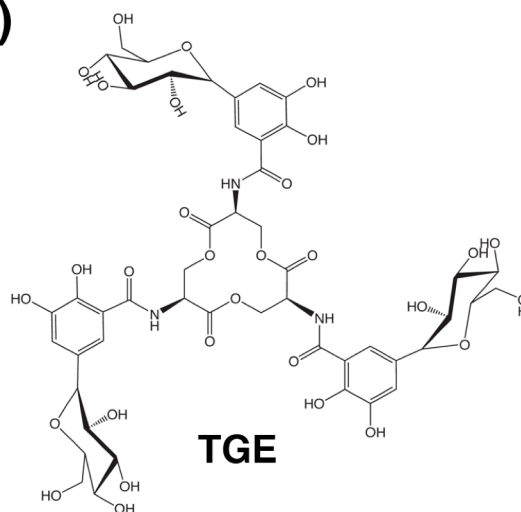


Figure 2.

```
IroB 1 -MRLFLVGPPLYGLLYPVLSLAQAFRVNGHEVLIASGGQFAQKAAEAGLVVFDAAFGLDSEA-----CYRHHEAQRKRSNIGTQMGNSFFSE
4G2T 1 HMRILVLAGCSSEGFVMPVPLSWALRAAGHEVLVAASENMGFTVTGAGLFFAFPCPSLDMPEVLSWDRECNRTMPREEKPLEHIGRQYGRVLV
2P6P 1 -MRLFLVAGSPATVFALAPLATAARNAGHQVMAANQDMGPFVVTGVGLFAVAATDLP-IRHFITTDREGRPEATPSDPVAQARFTGRWFARMAA

IroB 88 EMADHLVEFAGHWRPDLIIYPPLGVIGPLIAAKYDIEVVMQTVGFGHTPWHIRCVTRSLTDAYRRHNVGATPRDMAWIDVTPPSMSILENDGEP
4G2T 96 RMRDEALALAEERWKPDVLVLETYSLTGPLVAATLGIPIWIEQSIRLASPELTKSAGVGLAPELAELGLTDFPDPLLSIDVCPSPMEA-----GT
2P6P 94 SSLPRMLDFSRAWRPDLIVGGTMSYVAPLLALELGVPHARQTVDAVDADGIHPCADAEIRPELSELGLERLPAADLFIDICPPSLRPA--NAAPA

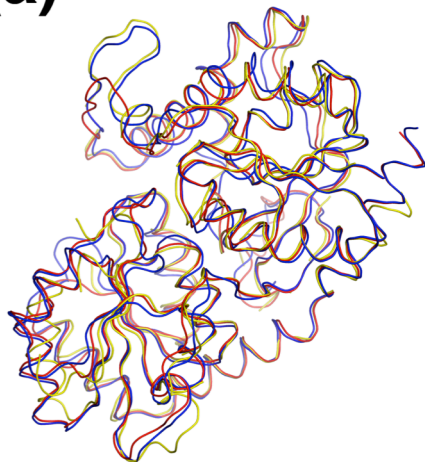
IroB 183 IPMQYVPYNGGA-VWEPWERRPDKRLVLSLCT---VKPMVDGLDLIAVMDSASEVDAEILHISANARSDLRSLPSNVRLVDWIFPMGVFLNG
4G2T 185 TKMRYVPYNGRNDQVPSWVFEERKQPRLCLEFCT-----RVPLL-LQALSQELPKLGFVAVVSDKLA----REGVLAAGQFPLSAIMPA
2P6P 187 RMRHVVATS-RQCPLFPWYTRDTRQRVLTSCSRVAKESYDRNFDLRLGLAKDLVRRWDVELIVAAPDVAEALRAEVPQA-RVGVTELDVVAPT

IroB 274 ADGFIHHGGAGNTLTALHAGHPQIVFGQADRPNARVVAERGCIP---GDVGLSSNMINAFLNRRSLRKASEEVAEMAAPCFGEVAKSL
4G2T 266 CDVVVHHGGHCTTLTCLSEGVQVSVPIAEVWDSARLLHAAAGAVEVP-----SVLAAQARIRDDSVVGNARRLAAEMATLPTPADIVRLI
2P6P 280 CDLLVHHAGGVSTLTCLSAQVQLLIPKGSVLEAPARRVADYGAATALLPGEDSTEATADSQELQAKDTVARRAQDLSREISGMPLPATVVTAL

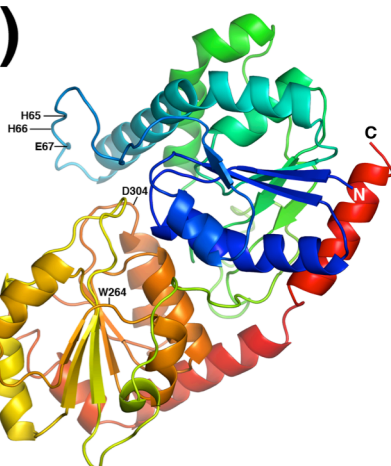
IroB 365 ITMVQKG-
4G2T 354 EQA-----
2P6P 375 EQLAHHHH
```

Figure 3.

(a)



(b)



(c)

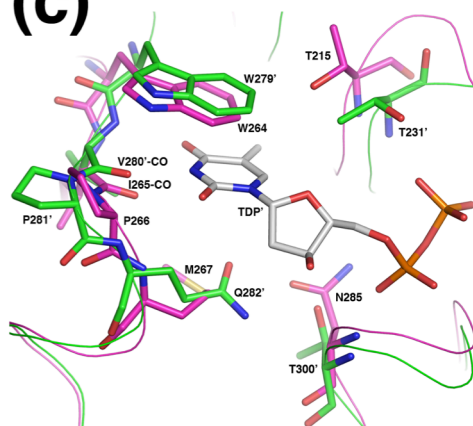


Figure 4.

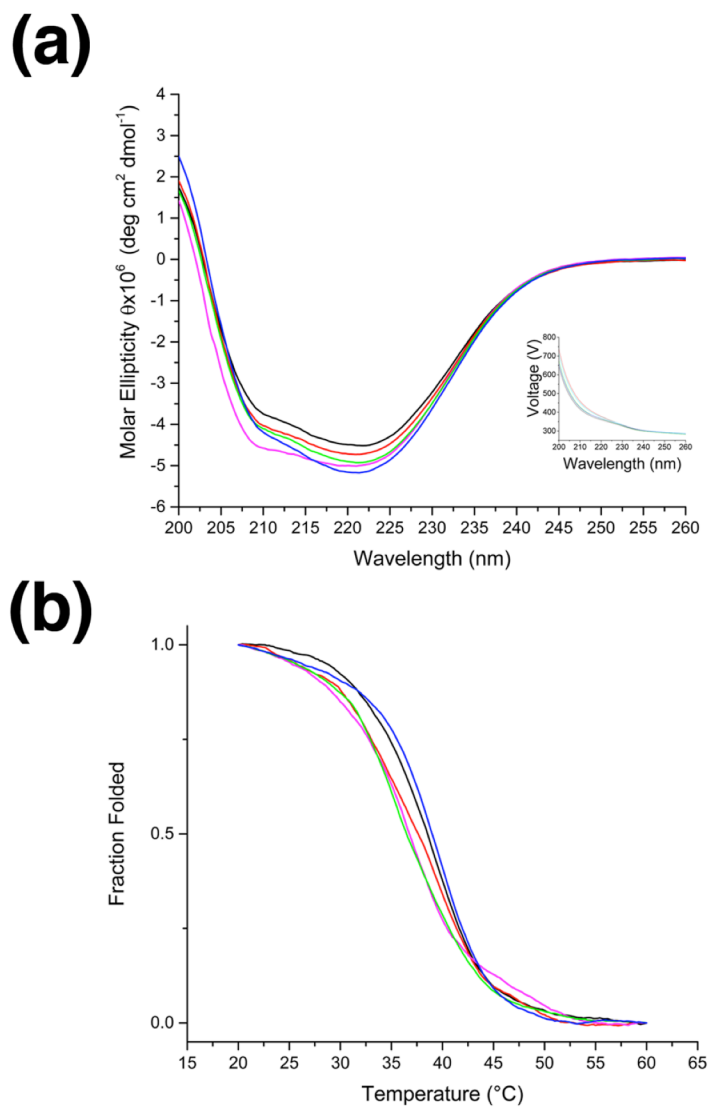
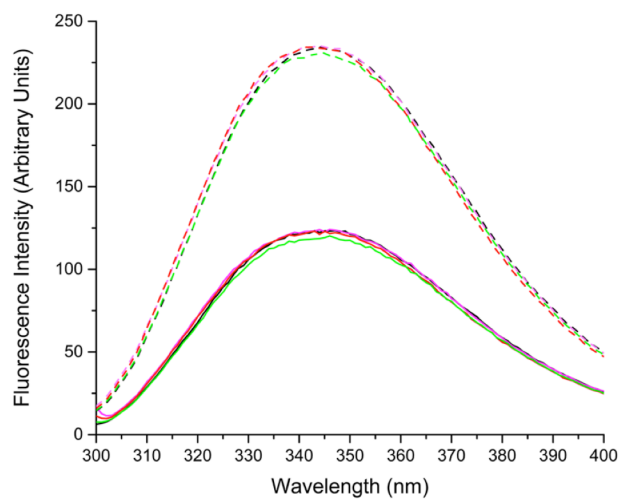


Figure 5.

(a)



(b)

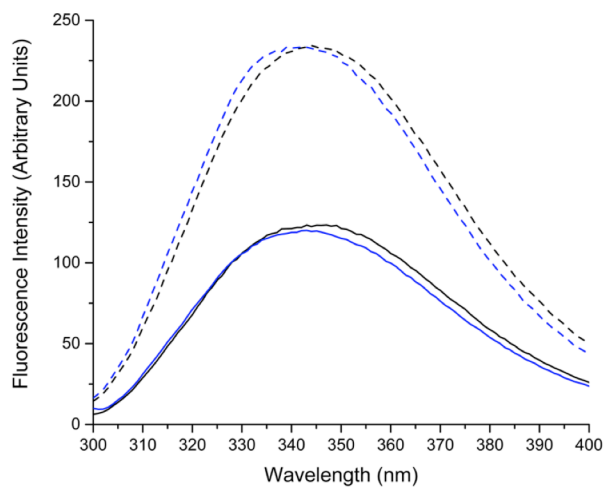


Figure 6.

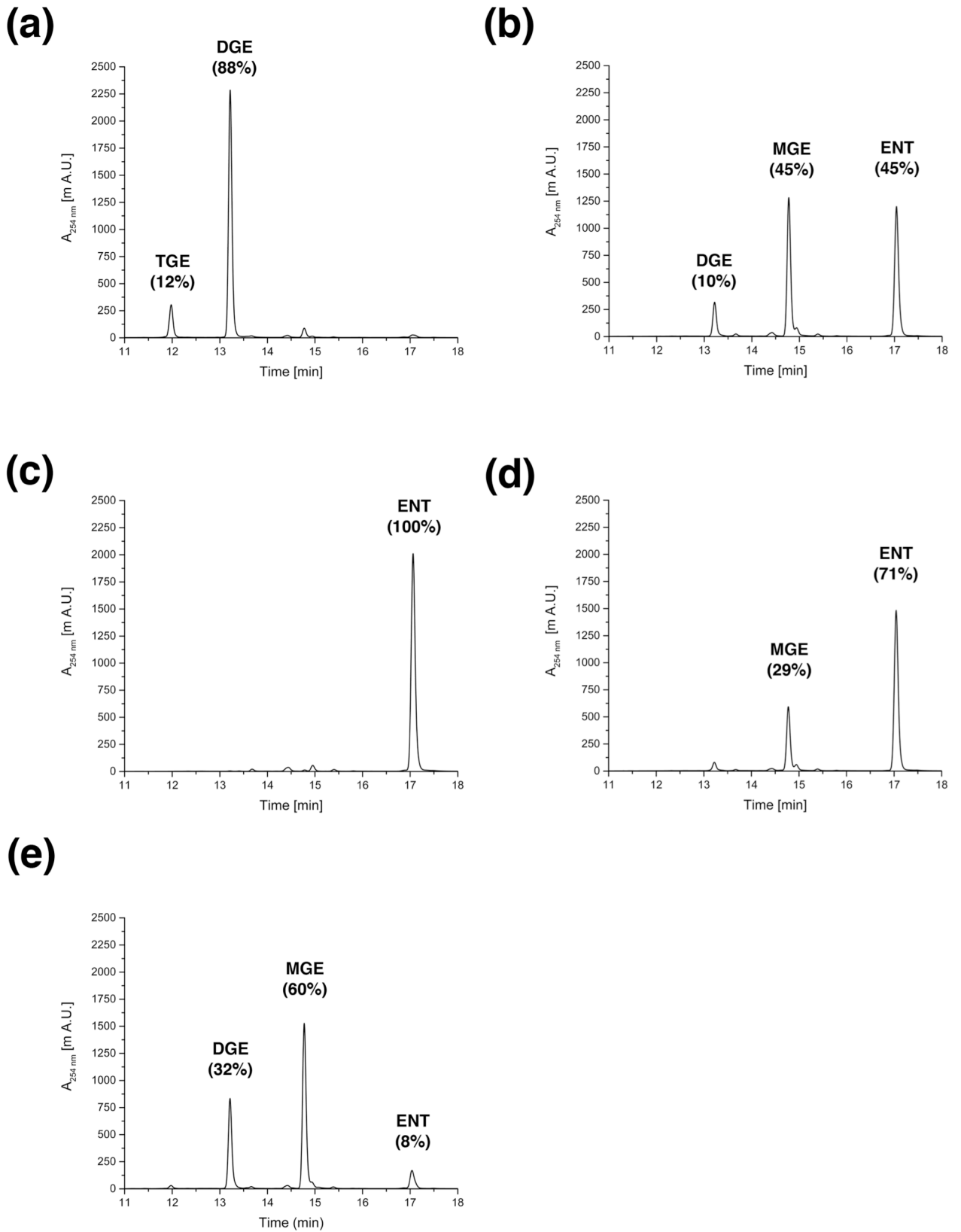


Figure 7.

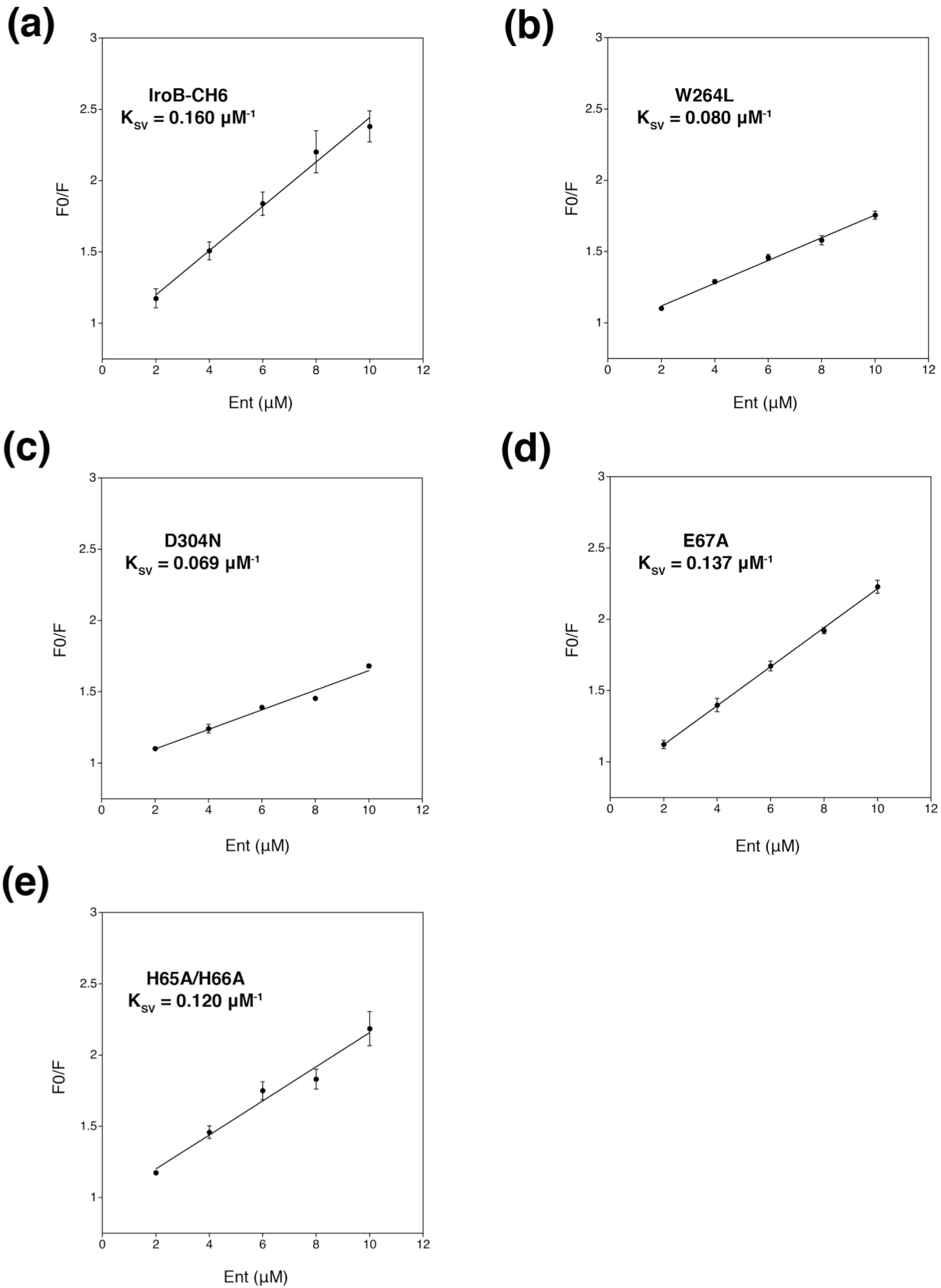


Figure 8.

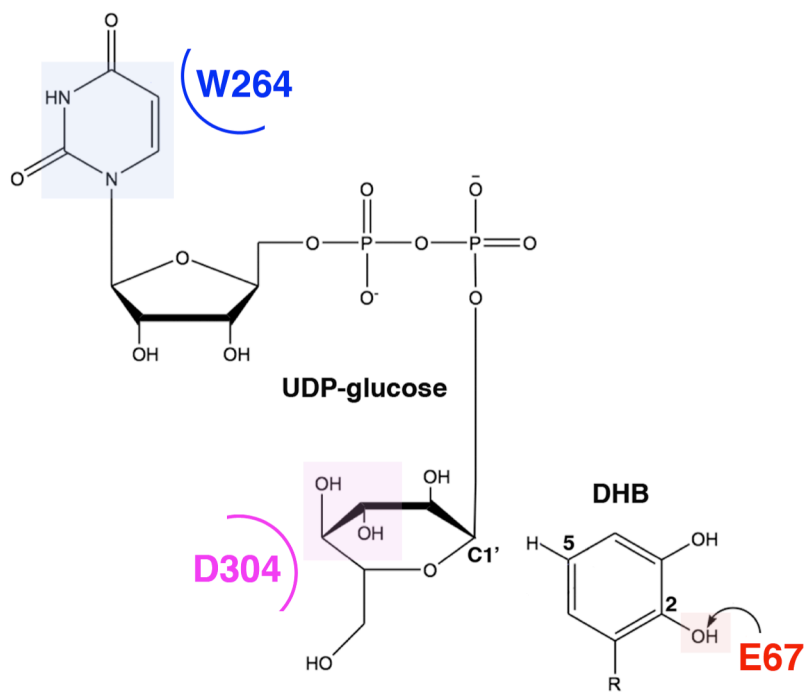


Figure S1.

<i>IroB</i> /1-371	1	MRILFVGPPLYGLLYPVLSLAQAFRVNGHEVLIASGGQFAQKAAE	--	AGLVVFDAAPG	-	56
2P6P/1-384	1	MRILFVAAGSPATVFALAPLATAARHAGHQVMAANQDMGPVVTG	--	VGLPAVATDDL	-	56
3IAA/21-416	21	AHLLIYHVASHGLILPTLTVVTELVRGRHRSYVTTAGGFAEPVRA	--	AGATVVPYQSE	-	76
2IYA/13-424	13	RHISFFNIPGHGHNPSLGI VQELVARGHRVSYAITDEFAAQVKA	--	AGATPVVYDSI	-	68
3IA7/5-402	5	RHILFANVQGHVYPSLGLVSELARRGHRITYVTTPLFADEVKA	--	AGAEVVLKSE	-	60
3H471-404	1	MGVLITGCGSRGDTEPLVALAARLRELGAARMCLPPDYVERCAE	--	VGVPMVPVGRA	-	56
2IYF/8-430	8	AHIA MFSAAHGHVNPSEVIRELVARGHRVYAI PPVFADKVA	--	TGPRPVLYHST	-	63
3OTG/21-412	21	MRVLFASLGTHTYPLLP LATAARRAGHEVTFATGEGFAGTLRK	--	LGFEVPATGMP	-	76
1RRV/1-416	1	MRVLLSVCGRGDVEIGVALADRLKALGVQTRMCAPPAAEERLAE	--	VGVPHVPVGLP	-	56
3OTV/21-398	21	MRVLFVSSPGIGHLFLPLQLAWGFRTAGHDVLI AVAEHADRAA	--	AGLEVVDVAPD	-	75
3TSA/2-391	2	MRVLVVPLPYPTHLMAMVPLCWALQASGHEVLI AAPP LQATAHG	--	AGLTTAGIRGN	-	57
2C1Z/8-456	8	PHVAVLAFPFSTHAAPLLAVVRRLLAAAAPHAVFSFFSTS QSNAS	I	FHDSMHTMQCNIKS	-	66
<i>IroB</i> /1-371	57	LDSEAGY	-	RHH	-	98
2P6P/1-384	57	PIRHFIT	-	TDREG	-	104
3IAA/21-416	77	IIDADAA	-	EVFG	-	115
2IYA/13-424	69	LPKESNP	-	EESW	-	106
3IA7/5-402	61	FDTFHVP	-	EVVK	-	99
3H471-404	57	VRAGARE	-	PGEL	-	91
2IYF/8-430	64	LPGPDAD	-	PEAW	-	101
3OTG/21-412	77	VFDGFLA	-	ALRIR	-	127
1RRV/1-416	57	QHMMQLQ	-	GMPPP	-	93
3OTV/21-398	76	YSAVKVF	-	EQVAK	-	127
3TSA/2-391	58	DRTGDTG	-	GTTQLRFPNPAFGQRDTEAGRQLEQASHVAQS	-	111
2C1Z/8-456	67	YDISDGVPEGYV	-	-	-	109
<i>IroB</i> /1-371	99	HWRPDLI	I	YPL	-	141
2P6P/1-384	105	AWRPDLV	I	GG	-	147
3IAA/21-416	116	GDVPDLV	I	LD	-	160
2IYA/13-424	107	DRPDLI	I	VYD	-	150
3IA7/5-402	100	DNPPDLV	I	VYD	-	144
3H471-404	92	GCD	-	AVVTTGLL	-	136
2IYF/8-430	102	DDIPDLV	I	LHD	-	145
3OTG/21-412	128	RLRPDLV	I	VQE	-	170
1RRV/1-416	94	GCA	-	AVVAVGDL	-	150
3OTV/21-398	128	DYRPDLV	I	VVE	-	170
3TSA/2-391	112	AWRPSVLL	I	VLD	-	155
2C1Z/8-456	110	GRPVSCL	I	VAD	-	151
<i>IroB</i> /1-371	142	-	-	-	-	157
2P6P/1-384	148	-	-	-	-	163
3IAA/21-416	161	-	-	-	-	193
2IYA/13-424	151	-	-	-	-	202
3IA7/5-402	145	-	-	-	-	177
3H471-404	137	-	-	-	-	169
2IYF/8-430	146	-	-	-	-	178
3OTG/21-412	171	-	-	-	-	186
1RRV/1-416	151	-	-	-	-	186
3OTV/21-398	171	-	-	-	-	186
3TSA/2-391	156	-	-	-	-	171
2C1Z/8-456	152	-	-	-	-	167
<i>IroB</i> /1-371	158	-	-	-	-	177
2P6P/1-384	164	-	-	-	-	182
3IAA/21-416	194	-	-	-	-	217
2IYA/13-424	203	-	-	-	-	225
3IA7/5-402	178	-	-	-	-	201
3H471-404	170	-	-	-	-	192
2IYF/8-430	179	-	-	-	-	201
3OTG/21-412	187	-	-	-	-	212
1RRV/1-416	187	-	-	-	-	208
3OTV/21-398	187	-	-	-	-	204
3TSA/2-391	172	-	-	-	-	190
2C1Z/8-456	168	-	-	-	-	219
<i>IroB</i> /1-371	178	-	-	-	-	201
2P6P/1-384	183	-	-	-	-	205
3IAA/21-416	218	-	-	-	-	242
2IYA/13-424	226	-	-	-	-	250
3IA7/5-402	202	-	-	-	-	226
3H471-404	193	-	-	-	-	214
2IYF/8-430	202	-	-	-	-	227
3OTG/21-412	213	-	-	-	-	236
1RRV/1-416	209	-	-	-	-	230
3OTV/21-398	205	-	-	-	-	227
3TSA/2-391	191	-	-	-	-	213
2C1Z/8-456	220	-	-	-	-	266

