Structural and Functional Studies on IroB

A Pathogen-Associated C-glycosyltransferase

Daniel Foshag

A Thesis

In the Department

of

Chemistry and Biochemistry

Presented in Partial Fulfillment of the Requirements

For the Degree of Master of Science (Chemistry) at

Concordia University

Montreal, Quebec, Canada

April 2012

CONCORDIA UNIVERSITY
School of Graduate Studies

This is to certify that the thesis prepared

By: Daniel Foshag

Entitled: Structural and Functional Studies on IroB, a Pathogen-Associated

C-glycosyltransferase

and submitted in partial fulfillment of the requirements for the degree of

Master of Science (Chemistry)

complies with the regulations of the University and meets the accepted standards with

respect to originality and quality.

Signed by the final examining committee:

_____ Chair
Dr. Joanne Turnbull

_____ Examiner
Dr. Joanne Turnbull

_____ Examiner
Dr.Paul Joyce

_____ Supervisor
Dr. Peter Pawelek

Approved by   _____
              Chair of Department or Graduate Program Director

_____   _____
     Date   Dean of Faculty

Abstract
Structural and Functional Studies on IroB

Daniel Foshag

Bacterial iron acquisition by the means of enterobactin (ENT) is constrained in mammalian hosts due to ENT-binding proteins such as siderocalin and serum albumin. To evade sequestration by these proteins, ENT can be modified by the C-glycosyltransferase IroB, which is located in the *iroA* locus of *Salmonella* and certain extraintestinal *E. coli* strains such as uropathogenic *E. coli* CFT073. The glycosylation of ENT has been reported to be a bacterial evasion mechanism to restore the iron scavenging ability of ENT in the presence of mammalian ENT-binding proteins by the installation of a steric impediment.

The C-glycosyltransferase IroB catalyses the transfer of a glucose moiety to the DHB subunit of ENT under formation of a C-C bond between the anomeric C_1 of the glucose moiety and the C_5 of the 2,3-DHB subunit of ENT. The formation of mono-, di- and triglycosylated Ent (MGE/DGE/TGE) products where observed *in vitro*. The formation of a C-C bond is remarkable because of its chemical stability and resilience against enzymatic degradation.

In this M.Sc. thesis, we initially identified the *iroB* gene product in the *iroA*-harbouring *E. coli* strain Nissle 1917 on transcriptional and translational level and expressed and purified IroB recombinant. Then, we investigated the mechanism of the C-C bond formation catalysed by IroB *in vitro*. Based on the hypothesis that deprotonation of the catechol 2-hydroxyl renders the catechol C_5 *para* to the 2-hydroxyl nucleophilic, the C-C bond would then be formed in a general S_N2 reaction between the attacking nucleophile and the anomeric carbon of glucose, which is further facilitated by the excellent phosphate leaving group of the UDP-glucose donor. By the means of homology modelling and superposition strategies, we were able to identify the binding sites of the glycosyl <u>donor</u>

UDP-glucose and the glycosyl <u>acceptor</u> ENT and to locate residues that could potentially act as base catalysts to increase the phenolate anionic character of the 2,3-DHB subunit during catalysis.

We established an activity assay for IroB, separated products arising from IroB activity by reversed phase chromatography and compared so the activity of wild-type IroB and several variants. Additionally, all variants were characterized biophysically, mainly to confirm that the structural integrity was not impaired by mutations. Ultimately, our results enable us to propose a mechanism for C-glycosylation of IroB that is consistent with other glycosyltransferases found in nature.

Acknowledgments

I would like to thank my supervisor Dr. Peter Pawelek for his guidance, support and excellent advice throughout the last two years. Furthermore I am thankful for his financial support to overcome the burden of international tuition fees and also by encouraging and helping me with applications for scholarships.

I'm also grateful for having Dr. Joanne Turnbull and Dr. Paul Joyce on my committee. Their contribution and invaluable suggestions advanced my knowledge and gave me good hints throughout my studies at Concordia University. Also the support of Dr. Joanne Turnbull during bureaucratic crises is greatly appreciated.

Additionally, I would like to thank Dr. Christopher Wilds and Marc Ouellet for taking the time to discuss organic chemistry and technical issues, respectively.

For financial support I thankful to Groupe de Recherche Axé sur la Structure des Protéines (GRASP) from McGill University and Concordia University for awards that helped me to finance my studies here in Montreal.

The foundation provided by Drs. Anke Burger-Kentischer and Hans Weber turned out to be invaluable throughout my Master studies at Concordia University.

Finally, I would like to thank my family in Germany and my girlfriend Debora for their moral support throughout my studies far away from home.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

(v/v)	volume/volume
(w/v)	weight/volume
°C	degrees Celsius
μL	microliter
aa	amino acid
attλ	λ-attachment-site
BME	2-mercoaptoenthanol
BSA	bovine serum albumin
cDNA	complementary deoxyribonucleic acid
ddH$_2$O	double distilled water
DTT	DL-dithiothreitol
E. coli	*Escherichia coli*
ECL	enhanced chemiluminescence
EDTA	ethylendiamintetraacetate
ENT	enterobactin
FA	formic acid
g	gram
g	gravitational acceleration
GdmCl	guanidinium-hydrochloride
h	hour
HCl	hydrochloric acid
HRP	horseradish peroxidase

ib	inclusion body
IEP	isoelectric point
IPTG	Isopropyl-β-D-thiogalactopyranosid
kDa	kilodalton
LB	Luria-Bertani
min	minute
mL	milliliter
mRNA	messenger ribonucleic acid
MS	Mass spectrometry
MSA	multiple sequence alignment
n/a	not available
NaCl	sodium chloride
NaOH	sodium hydroxide
NGAL	Neutrophil Gelatinase-Associated Lipocalin
o/n	over night
OD_{600}	optical density at 600 nm
OTG	n-octyl-B-D-thioglucopyranoside
PAGE	polyacrylamide gel electrophoresis
PIC	protease inhibitor cocktail
RT-PCR	reverse transcription polymerase chain reaction
s	second
SD	Shine-Dalgarno
SDS	sodium lauryl sulphate

TCA	trichloracetic acid
TCEP	Tris(2-carboxyethyl)phosphine
TDP	thymidine diphosphate
TEMED	N,N,N' N' – tetramethylethylenediamine
TEV	*Tobacco Etch Virus*
TFA	trifluoroacetic acid
TMAO	trimethylamine N-oxide
Tris	2-Amino-2-hydroxymethyl-propane-1,3-diol
UDP	uridine diphosphate
UPEC	uropathogenic *E. coli*

Amino acid symbols

Ala (A)	Alanine	Leu (L)	Leucine
Arg (R)	Arginine	Lys (K)	Lysine
Asn (N)	Asparagine	Met (M)	Methionine
Asp (D)	Aspartic acid	Phe (F)	Phenylalanine
Cys (C)	Cysteine	Pro (P)	Proline
Gln (Q)	Glutamine	Ser (S)	Serine
Glu (E)	Glutamic acid	Thr (T)	Threonine
Gly (G)	Glycine	Trp (W)	Tryptophan
His (H)	Histidine	Tyr (Y)	Tyrosine
Iso (I)	Isoleucine	Val (V)	Valine

1   General introduction

Iron is an essential nutrient for most organisms, representing an important cofactor needed by approximately one-quarter to one-third of all proteins including those involved in primary and secondary metabolism (*1, 2*). These so-called metalloproteins can bind iron either directly using amino acid side chains, but mostly binding is mediated by heme groups or iron-sulphur clusters (*2*). Metalloproteins are an integral part of metabolic processes such as photosynthesis, respiration and nitrogen fixation (*3*). The reversible redox pair Fe^{2+} (ferrous iron) and its oxidized form Fe^{3+} (ferric iron) are the physiologically relevant iron sources. They are involved in broad spectrum of redox reactions, mediate electron chain transfer and serve as cofactor for structural and functional integrity of proteins. The bacterial uptake of iron is limited to these two physiologically relevant forms (*1, 3, 4*).

The reduced form Fe^{2+} is soluble in aqueous solution and thus readily available to bacterial metal transport systems. However, in bacterial habitats, Fe^{2+} is mostly oxidized to Fe^{3+} by molecular oxygen or enzymatic processes in mammalian hosts and Fe^{3+} is mostly insoluble (*5, 6*). The hard Lewis acid Fe^{3+} forms preferably adducts with hard Lewis bases such as water or hydroxides and is hexacoordinated at pH <1 by six water molecules. However, with increasing pH, the ferric coordinating water molecules eventually ionize to form a ferric-hydroxide complex. The ferric-hydroxide complex is prone to undergo a process named olation and oxolation, where oxygen and hydroxides act as bridging molecules and induce the formation higher order insoluble polymers between different ferric-hydroxide complexes (*7*). Ultimately, these processes lead to formation of ferric oxide hydrate complexes (Fe_2O_2 x nH_2O). The solubility of ferric-

oxide hydrate complexes is limited to a concentration of 10^{-18} M in aqueous solution at neutral pH. Thus, even though iron is one of the most abundant elements in the terrestrial environment, its availability is highly constrained by its low solubility and is far from the concentration needed by bacterial cells to maintain metabolism (the threshold metabolite concentration within the cells is in the micro molar range (*6*)). In mammalian hosts, the habitat of important pathogens, the availability of iron is even more limited due to sequestration of free Fe^{3+} to storage and transport proteins such as ferritin and transferrin. Iron chelation mechanisms in mammals result in a free serum iron(III) concentration of about 10^{-24} M. For this reason, serum is considered bacteriostatic (*8-10*).

In order to overcome these limitations and satisfy their need for iron, bacteria evolved different iron acquisition strategies. Evolution has led to two different strategies: *direct* and *indirect uptake* of iron. The first strategy incorporates *direct uptake* from iron sources such as the mammalian iron binding protein transferrin. One species that employs this strategy is the Gram-negative bacterium *Neisseria gonorrhoeae* that harbours the TbpA and the co-receptor TbpB. The Tbp-system allows binding of human transferrin and extraction of the iron(III) bound to transferrin. The crystal structure of receptor and co-receptor in complex with human transferrin was solved recently (*11*). The disadvantage of this strategy is that a specific receptor for each iron source is needed – in such a way the TbpA/B system is restricted to transferrin. The more widespread *indirect uptake* will be discussed in the next chapter.

1.1 Indirect uptake of iron

In principle, the *indirect uptake* strategy is quite diverse in its appearance. Some Gram-negative bacteria, for example, harbour an iron-acquisition system known as *has*, which is capable of exploiting different heme sources through the use of secreted hemophores that extract iron from heme (*12, 13*). However, this type of iron acquisition is quite similar to the *direct uptake* strategy and possesses similar disadvantages – it involves specialized molecules highly specific for a single iron source (*e.g.,* heme) and is therefore limited to particular compartments where these sources are available (*1*).

The most distributed, and possibly the most successful, strategy is based on highly efficient iron-chelating molecules called siderophores (greek for: "iron carrier") (*14*). These small (usually <1 kDa) molecules demonstrate a high affinity for ferric iron and are capable to "steal" iron from almost every source. Siderophores can by divided into three main categories on the basis of the functional group that is involved in Fe^{3+} coordination (*15*).

1.2 Siderophores

There are three major functional groups found in siderophores responsible for coordination of iron(III): catecholate (includes catecholate and phenolate (aka "aryl caps"), hydroxamate, and carboxylate (Figure 1) (*1, 15*). As mentioned elsewhere, Fe^{3+} is considered a hard Lewis acid that forms adducts with hard Lewis bases such as oxygens, water molecules or hydroxides in solution, preferably in a hexacoordinated manner with octahedral geometry. The functional groups in siderophores (catechol, carboxylate and hydroxamate) described to date all possess strong oxygen donor to coordinate iron(III).

On the contrary, for coordination of the intermediate Lewis acid Fe^{2+}, donor atoms like nitrogen or sulfur are preferred (*16*).
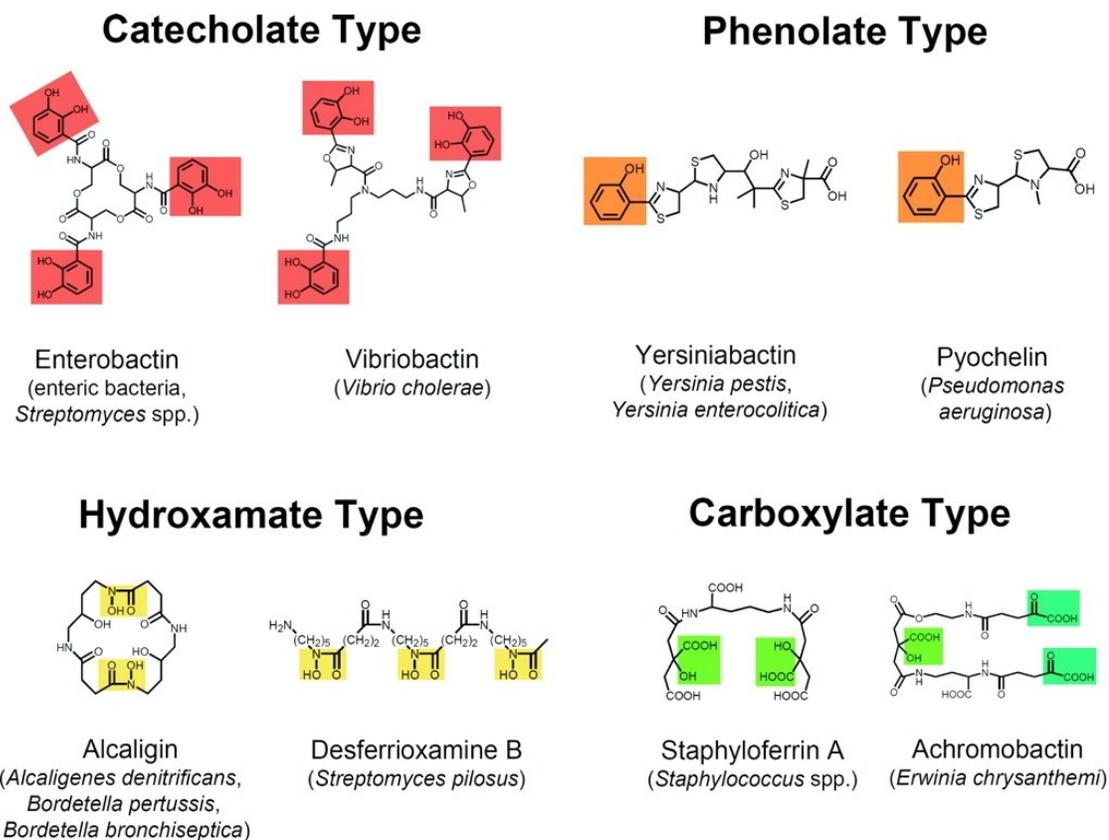
Figure 1: various types of siderophores with different functional groups and the corresponding natural producer. They are classified by the chemical nature of their functional group. The catechol group is highlighted is red, the phenolate group in orange, hydroxamate in yellow and carboxylate group in green (*1*).

The hexacoordination with octahedral geometry is thermodynamically very stable and is imitated by iron chelators. The functional groups of siderophores, for instance, generally supply six oxygen donors to coordinate Fe^{3+} in a highly stable hexacoordinated state that can be found for ENT (six hydroxyl groups) or desferrioxamine (three hydroxamate groups, each with two oxygen donors). Siderophores providing six donor atoms to form complexes with a 1:1 stochiometry, whereas if less than six donors are available (*e.g.,*

pyochelin from *Pseudomonas aeroginosa*), then complexes with higher stoichiometry may be formed or, alternatively, vacancies may be occupied by water molecules (*17*).

The efficiency of catechol siderophores is also dependent on the protonation state of their coordinating hydroxyl groups. Therefore, the affinity for Fe^{3+} is pH dependent and the physiological context must be considered. Many fungal species, for instance, live under acidic conditions, where the catechol hydroxyl groups would be fully protonated, making this type of siderophore less efficient under the acidic conditions of such a habitat. Therefore organisms living under these conditions prefer siderophores of the carboxylate type, with pK_a values of coordinating groups ranging between 3.5-5.0. Thus, carboyxylate type siderophores are more effective under acidic conditions (*1, 18*). However, the strongest siderophore under physiological conditions is known to be the catecholate siderophore enterobactin (ENT), which is described in the following Section.

1.2.1   The catecholate siderophore ENT

The catecholate siderophore ENT was first isolated from *Escherichia coli* and *Salmonella typhimurium.* ENT features threefold symmetry with three functional catecholate groups connected *via* an amide linkage to a trilactone macrocycle backbone (*19*) as shown in Figure 2.

As discussed briefly in the previous Section, ENT is an extremely strong iron chelator with an estimated equilibrium dissociation constant for Fe^{3+} to ENT of $K_D \sim 10^{-50}$ M (*20*); however, this estimated value refers to the rather hypothetical protonation state, where all six hydroxyl groups of the three catechol subunits are deprotonated. Since the backbone is easily hydrolyzed and thus unstable at high pH at which all hydroxyls could be

deprotonated, this value does not reflect physiological conditions. The experimentally

determined equilibrium dissociation constant, however, remains impressive: the K_D for

Fe^{3+} binding at pH 7.4 was determined to be $10^{-35.5}$ M, based on a competition assay with

EDTA (*19*). The high affinity of Fe^{3+} to ENT is largely due to the fact that the metal ion

is coordinated in the thermodynamically favoured hexacoordination sphere with

octahedral geometry, as shown in Figure 2, lower panel right (*21*).
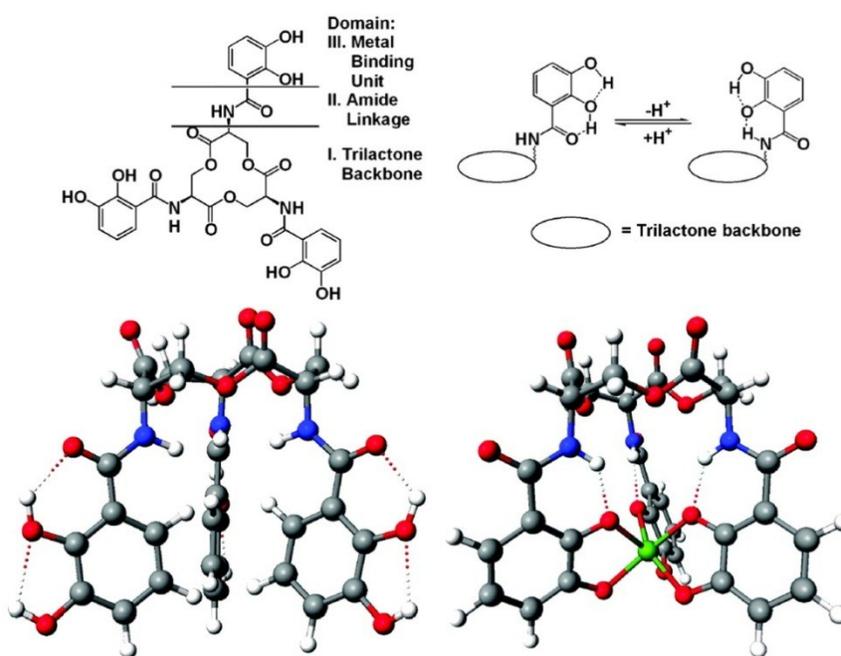
Figure 2: Structural representation of ENT (upper left). ENT possesses three catecholate groups connected *via* an amide linkage to a trilactone macrocycle backbone. The conformation of the catecholate group is charge-driven (upper right). Spacefilling representation (red oxygen, blue nitrogen, white hydrogen, grey carbon, green ferric iron) of ENT can be seen in the lower figure. The figure (lower left) shows the space filling representation of the protonated and uncomplexed ENT. Upon deprotonation, the functional hydroxyl groups point to the center and coordinates the metal ion in a hexacoordinated state with octahedral geometry (*21*).

Not only the equilibrium dissociation constant for iron(III), but also the conformation of

ENT is charge-driven and depends upon the pH. The *ortho*-hydroxyl groups (C_2) with

respect to the amide linked C_1 exhibits a pK_a of 6.5 to 8 and the *meta*-hydroxyl group

6

(C_3) possesses a pK_a of 11.5. At pH values below its pK_a, the *ortho*-hydroxyl group is bonded to the amide oxygen and the hydroxyl groups are pointing away from the center. Upon deprotonation or metal binding the *ortho*-hydroxyl group changes to the *trans* position, whereas the *ortho*-hydroxyl group is hydrogen bonded to the nitrogen of the amide linkage (*1, 21, 22*).

1.2.2   Biosynthesis of ENT

Proteins involved in the biosynthesis, export and reuptake of ENT are clustered on the *ent/fep* locus (Figure 3), which can be found in a large number of Gram-negative and Gram-positive bacteria. The gene expression for iron acquisition must be tightly regulated to avoid the accumulation of free iron, which can be toxic for the cell, but also to maintain an intracellular iron concentration at approximately 10^{-6} M for metabolic processes (*6, 15*). To achieve a tight regulation of the expression of genes involved in iron acquisition by ENT, bacteria have evolved two different transcription factors. In Gram-positive bacteria iron acquisition is controlled by DtxR (diphtheria toxin regulator) and in Gram negative (enteric bacteria) the control is accomplished by Fur (Ferric Uptake Regulator). Interestingly, the DtxR is found often in Gram-negative bacteria in addition to Fur and controls manganese transport (*23*).

Even though these regulators show low sequence identity, structural features are well conserved. Both regulators possess a structural binding site for zinc that mediates dimerization and a regulatory binding site for Fe^{2+}, which enables Fur to bind to its DNA recognition site (*24*). The DNA recognition site is known as the Fur box, and features the following inverted repeat consensus sequence: 5'-GATAATGATAATCATTATC-3'

(*23*). The *ent/fep* locus is negatively regulated by Fur depending on the concentration – expression is repressed when Fe^{2+} is bound to the regulatory binding site on Fur and derepressed under low iron availability due to Fe^{2+} dissociation from Fur (*24, 25*). Proteins involved in biosynthesis and uptake of iron are almost exclusively regulated at the transcriptional level (*26*).
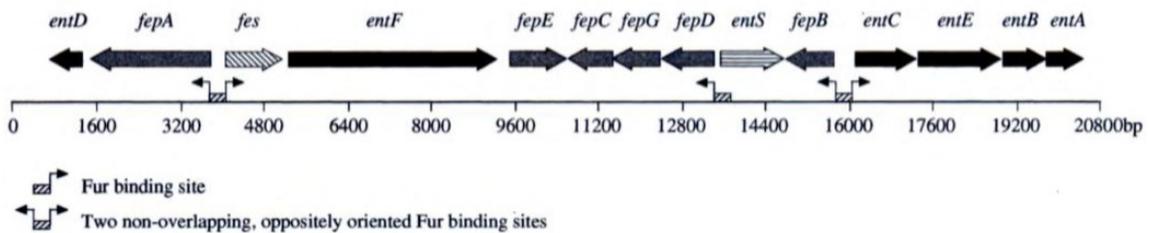
Figure 3: Fep/Ent gene cluster regulated by Fur (ferric uptake repressor) (*19*)

The *ent/fep* cluster encodes for proteins EntA-F, EntS, FepA-G and Fes. EntCBA (Figure 3) are involved in the synthesis of the functional DHB subunit of ENT and EntDEF assemble three DHB subunits and three L-serines to the final ENT molecule. EntS is responsible for secretion of ENT. The proteins FepAB and FepCDG import *holo*-ENT, where FepAB mediate transport through the outer membrane and FepCDG is the inner membrane permease complex. Finally, Fes is required for the release of Fe^{3+} from its ligand following import to the cytoplasm. The ENT biosynthetic and synthetic pathway is shown in Figure 4 (*19*).
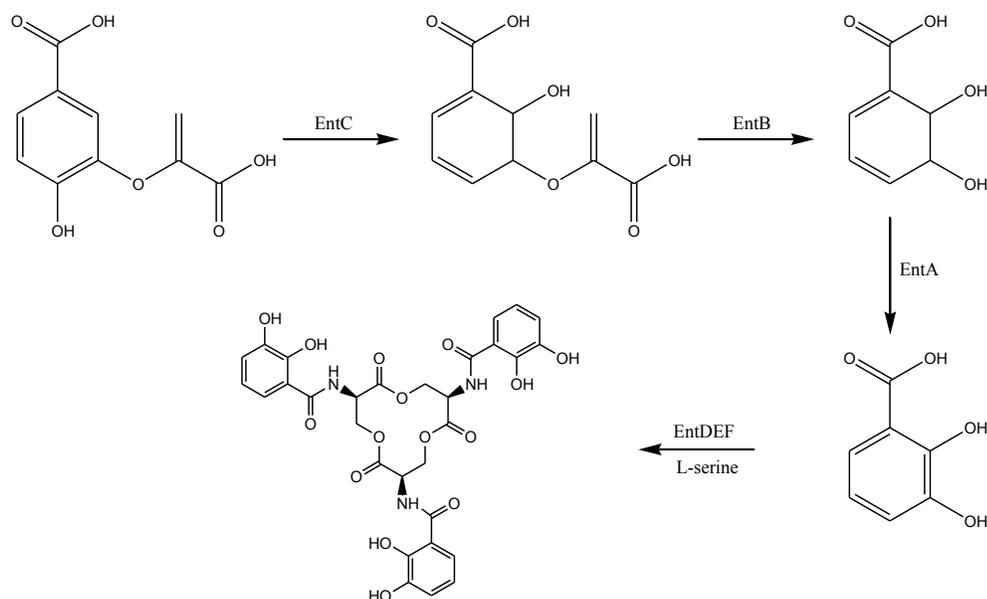
Figure 4: Biosynthetic and synthetic pathway for ENT. EntCBA are catalysing the conversion of chorismate to the final functional group 2,3-dihydroxybenzoic acid (DHB). EntDEF assembles the final ENT molecule using three L-serine molecules and connecting it *via* an amide linkage to three DHB subunit groups and enzymatic tethering of the three DHB-S units.

The biosynthesis of ENT can be divided in two modules, where the DHB module involves the activities of EntCBA (DHB subunit) and the non-ribosomal peptide synthesis (NRPS) module that involves the activities of EntDEF (assembly of three DHB subunit and three L-serine). The DHB module starts from an offshoot of the aromatic amino acid pathway. Firstly, chorismic acid is converted to isochorismate in the presence of magnesium by the isochorismate synthase EntC. In the next step, the isochorismatase activity of EntB converts isochorismate into 2, 3-dihydro-2,3-dihydroxybenzoate under removal of pyruvate. This intermediate is then oxidized to 2,3-dihydroxybenzoic acid in the presence of cofactor NAD$^+$ by EntA (*21*).

EntB is bifunctional protein that possesses an N-terminal isochorismate lyase domain involved in DHB biosynthesis and a C-terminal ArCP (aryl carrier protein domain) that also plays an essential role in the NRPS module of ENT assembly. During NRPS, the

EntB ArCP domain is phosphopantetheinylated by EntD to produce a *holo*-ArCP domain that facilitates transfer of adenylation-activated 2, 3-DHB onto *holo*-EntB by EntE (*27*). EntF is a multidomain (C-A-PCP-TE) protein and the central enzyme for final assembly of ENT in cooperation EntB. The A (adenylation) domain of EntF activates L-serine also by adenylation, which is then covalently attached to the PCP (peptide carrier protein) domain of EntF. The C (condensation) domain of EntF catalyses the formation of an amide linkage between L-serine using *holo*-EntB as the DHB donor. The DHB-serine product is then transferred to the TE domain of EntF, which makes the PCP domain available for another round to condense DHB and serine. In the course of this 'assembly line' process, three DHB-Ser are formed, transferred to the TE (termination) domain of EntF and covalently tethered to produce the final ENT molecule (*19, 21, 27, 28*).

1.3   Mammalian Defense Mechanisms against Bacterial Iron Acquisition

As described elsewhere, iron is extremely important in a variety of metabolic processes of bacteria, including pathogens, but its availability in mammalian host is limited due to sequestration by iron binding proteins such as transferrin or ferritin or prosthetic heme groups (*6, 12, 13, 29*). By the means of sequestration by transferrin, mammals limit the available free iron to a bacteriostatic concentration of 10^{-24} M in serum (*30*). However, by secretion of siderophores, many bacteria possess highly efficient molecular mechanisms to extract iron from different sources including transferrin or ferritin, thus enabling them to overcome limitations imposed by these mammalian iron binding proteins (*30*). In order to control siderophore-based iron acquisition, the innate immune system of mammals also directly interferes with this mode of iron-acquisition through the use of proteins known as

lipocalins (*31*). A lipocalin that has been demonstrated to interfere directly with ENT-mediated iron acquisition is NGAL (neutrophil gelatinase-associated lipocalin) (*32*); the surface representation is shown in Figure 5. It is secreted upon induction of an inflammation response by neutrophil granulocytes (white blood cells) and has been shown to bind *holo*-ENT (*i.e.*, ferric ENT) and also other siderophores with high affinity ($K_D = 0.4 \pm 0.1$ nM) (*31*).
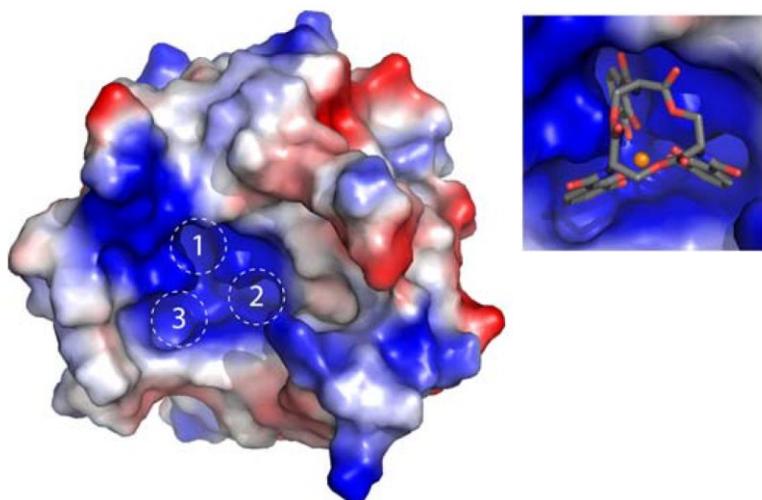
Figure 5: Siderocalin (surface representation), indicated with numerals 1,2 and 3 are the binding pockets that accommodate the DHB subunit upon binding of FeENT (left) and FeENT bound siderocalin (right) (*31*).

Lipocalins are quite diverse in terms of their sequence, but they share a characteristic fold: they possess an eight-stranded antiparallel β-barrel that surrounds a binding pocket called calyx; protein-ligand specificity is defined by residues in the calyx (*33*).

NGAL features a high specificity for *holo*-ENT, the calyx is defined by three binding pockets (1, 2 and 3; Figure 5) pockets accommodate the catechol subunits of ENT as shown in the right panel in Figure 5. In these binding pockets the positively charged residues Arg81, Lys125 and Lys134 define ligand specificity (*31*). ENT possesses a net

charge of 0, the FeENT complex however has a charge net charge of -3 (*34*). Thus, the interaction between the residues in the calyx and Fe-loaded ENT is composed of electrostatic and cation-π interactions between the catechol group of ENT and the positively-charged residues in the binding pocket of NGAL (*31*). The physiological relevance of the NGAL-FeENT interaction has been shown in *in vitro* and *in vivo*. For instance, NGAL-deficient mice are susceptible to infection by *E. coli* and growth of *E. coli* is limited in the presence of NGAL (*34, 35*).

Since the mammalian innate immune features an excellent countermeasure against ENT-based iron acquisition *via* NGAL, certain pathogens carry an additional "second layer" iron acquisition system called *iroA*. This system depends on the *ent/fep* system and facilitates glycosylation of ENT and additionally provides specialized uptake and reuptake systems to restore the iron-scavenging ability of ENT in the presence of mammalian NGAL (*30, 36*). An essential protein of the *iroA* system, IroB, is the focus of this thesis.

1.4   *iroA* – a pathogen-associated iron acquisition system

In addition to sequestration of ENT by NGAL that was described in the previous chapter, the efficiency of the *ent/fep* system in mammalian hosts is also impaired by the hydrophobic character of ENT. Its hydrophobicity results in partitioning into lipid bilayers (*37*) and/or association with serum proteins such as albumin that binds hydrophobic ligands (*38*). These factors together limit the bioavailability of Fe-ENT, thus affecting iron homeostasis of pathogens in the presence of mammalian host proteins (*30*).

To overcome these limitations, important pathogens such as uropathogenic *E. coli* (CFT073), *Salmonella* species and certain *Klebsiella* strains harbor the "second layer" iron acquisition system *iroA*. The *iroA* locus, comprised of a five-gene cluster that will be described in more detail later, allows for glycosylation of ENT and thus alters its physical and steric properties (*36*). In Table 1 the presence of different iron acquisition systems in intestinal and extraintestinal *E. coli* isolates are tabulated, which demonstrates that the *iroA* cluster is exclusively found in extraintestinal isolates.

Table 1: The presence of various iron-acquisition systems in intestinal/extraintestinal *E. coli* isolates. Data taken from (*19*).

	Presence of iron acquisition system				
	Ent/Fep[1]	Iuc/Iut[2]	Ybt[3]	Fhu[4]	Iro[5]
E. coli					
Intestinal					
K-12	+	-	-	+	-
EPEC E2348/69	+	-	-	+	-
EAggEC O42	+	-	+	+	-
EHEC O157:H7	+	-	+	+	-
Extraintestinal					
UPEC (CFT073)	+	+	+	+	+
NMEC O18ac:H7:K1	+	-	+	+	+

[1]ENT [2]aerobactin [3]yersiniabactin [4]hydroxomate type siderophores [5]salmochelin

While *E. coli* is a mutualistic organism in the human intestinal tract and thus contributes to human health, it can cause infections when present at extraintestinal locations such as the urinary tract or in wounds and other tissues. The ability to survive in extraintestinal environments depends on several virulence factors, including the ability to scavenge iron when competing with the mammalian immune system (*39*). When comparing iron acquisition systems present in intestinal and extraintestinal isolates in Table 1, one might

notice that the *iroA* system is not found in any listed intestinal *E. coli* isolate, but in every extraintestinal isolate. This suggests that the *iroA* cluster plays a significant role as a virulence factor. In fact, this notion is supported by a recent study that identified salmochelin and ENT followed by aerobactin in *E. coli* 83972 (asymptomatic bacteriuria related to UPEC) to be consistently important as a fitness factor in comparison with other iron acquisition systems under different growth conditions (*40*). But also other studies clearly demonstrated the importance of the *iroA* system to evade sequestration by host protein NGAL due to the altered steric properties of glycosylated ENT (*30, 33, 36, 41, 42*). Furthermore, glycosylated ENT features advantageous physical properties such as higher solubility and lowered membrane partition coefficients, thus maximising the efficiency in membrane-rich environments (*37*).

The *iroA* cluster consists of five genes: *iroB, iroC, iroD, iroE, and iroN*. Expression of these genes, like the *ent/fep* system, is under the control of the iron-sensing Fur transcription regulator, and both the *ent* and *iroA* systems are up regulated at the same time. The *iroA* cluster is found in pathogens on chromosomal pathogenicity islands or plasmids, suggesting that these genes are mobile and can be transferred by horizontal gene transfer. As mentioned previously, the function of the *iroA* system relies on the *ent/fep* system and involves the synthesis of IroBCDEN proteins to (*i*) produce salmochelin, (*i.e.*, glycosylated ENT), (*ii*) export and linearize *apo*-salmochelin, (*iii*) uptake *holo*-salmochelin, and (*iv*) degrade of *holo*-salmochelin in the cytoplasm to release iron (Figure 6) (*30*).

The first enzyme synthesized from the *iroA* operon is **IroB**, a glycosyltransferase that catalyses the transfer of glucose moieties onto C_5 carbons of ENT catechol groups, thus altering the physical and steric properties of the siderophore (*43*). Since ENT possesses three catechol groups, this reaction can occur up to three times under formation of mono-glycosylated ENT (MGE), di-glycosylated ENT (DGE) and triglycosylated ENT (TGE). IroB is the main topic of this work and will be described in Section 1.6 in detail.

IroC, like FepBDGC, is predicted to be an ABC (ATP-binding cassette) transporter based on sequence similarities (*44*). In fact, it was reported that IroC might play similar role as FepBDGC, since a FepB-deficient strain was still able to take up linearized salmochelins, suggesting that IroC could rescue and import linearized salmochelin (*45*) However, these results were not confirmed in a recent study that identified IroC as being able to facilitate excretion of salmochelin arising from IroB activity and ENT (*30, 44*). Furthermore, is was shown that IroC deficient strains gene are no longer able to export significant amounts of cyclic DGE (salmochelin S4) or linearized derivates, but only monomeric breakdown products such as salmochelin SX (glycosyl-DHB) (*46*); taken together, these results imply that IroC is indeed the salmochelin/ENT exporter.
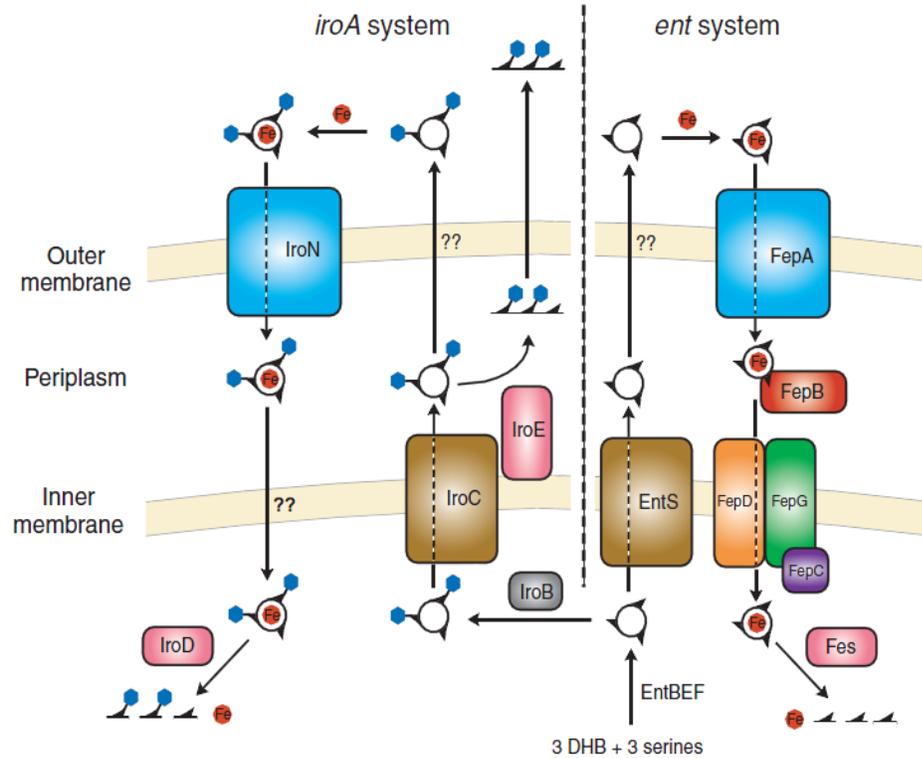
Figure 6: Schematic view of the the *ent/fep* and *iroA* system working in tandem (*30*).

IroE is periplasmic esterase associated with the inner membrane (*45, 47*). Recent findings support the hypothesis that IroE plays a significant role for export of *apo*-salmochelin rather than import of *holo*-salmochelin. Firstly, *in vitro* studies have shown that IroE shows higher affinity for *apo*-salmochelin and cleaves the trilactone backbone of salmochelin only once. This observation goes hand in hand with the relative high abundance of linearized derivates of salmochelin in culture medium of *iroA* harbouring strains – IroE-deficient strains, in contrast, secrete only minimal amounts of linearized salmochelin (*30, 41, 48*). The physiological significance of the linearization is possibly given by a lowered membrane partitioning coefficient, thus making linearized salmochelin more efficient in a membrane-rich environment (*37*).

IroN shows a high sequence similarity to the TonB-dependent ENT receptor FepA and has been demonstrated to be necessary for *holo*-salmochelin import, since salmochelin cannot be imported by any other transport system (*30*). Besides acting as *holo*-salmochelin receptor, Feldman *et al.* even attributed IroN to be an internalization factor as demonstrated by an *in vitro* invasion assay (*49*). Once *holo*-salmochelin has been imported into the periplasm, it is further processed by the common inner-membrane receptor complex FepBCDG and transferred into the cytoplasm.

There, hydrolysis of *holo*-salmochelin is carried out by **IroD.** In contrast to IroE, IroD cleaves salmochelin multiple times to yield DHB-Ser monomers to facilitate iron release. Additionally, IroD shows higher affinity for *holo*-salmochelin as opposed to IroE (*30*).

Now that we have examined the physiological context in which the glycosyltransferase IroB is found, a more general introduction to the protein will be provided in the next section. The general structural and mechanistic features of glycosyltransferases will be examined, followed by a detailed description of the current state of research on IroB.

1.5   Glycosyltransferases

Glycosyltransferases (EC 2.4.x.y.) are enzymes that catalyse the transfer of a carbohydrate moiety from a substituted phosphate leaving group to an acceptor molecule. The sugar donor substrate comprises a rather limited range of molecules and is usually a nucleoside di-phosphate such as UDP-glucose or GDP-mannose, but also mono-phosphate and lipid phosphate-activated carbohydrate moieties can be utilized by this

class of enzymes (*50*). As opposed to the narrow range of donor sugar substrate accepted by glycosyltransferases, a large variety of acceptor molecules can be glycosylated ranging from proteins, lipid, nucleic acids, antibiotics and oligosaccharides (*51, 52*). Owing to the variety of acceptor molecules, the glycosylation reaction is prevalent in a wide range of biological contexts. Glycosylation plays not only an important role for energy storage (Amylum (*50*)) and the structural integrity of cell walls (cellulose and peptidoglycan) in plants and bacteria and proteins (*53*), but also cell adhesion and signalling in immunological contexts (*e.g.,* pathogen recognition by mannose binding receptors), glycoprotein trafficking and many more (*54*). Therefore, it is not surprising that glycosyltransferases are encoded in 1-2% of the whole genome sequence; this is true for all three domains namely archaea, bacteria and eukarya (*50*).

Glycosyltransferases are classified based on sequence and hydrophobic cluster similarities by the database CAZy (Carbohydrate-active enzymes) (www.cazy.org). The sequence based classification results in useful predictions regarding the stereochemistry of the substrate and products in either inverting or retaining glycosyltransferases. In addition, information about fold characteristics are provided (*55*).

1.5.1   Glycosyltransferases: Fold characteristics

The classification of glycosyltransferases into either a GT-A or GT-B fold was suggested in 2001 by Bourne *et al.* after nine structures of glycosyltransferases had been solved. These structures were found to conform to only two distinct folds, named GT-A and GT-B (*56, 57*). A representative structure for the GT-A fold is that of SpsA from *Bacillus subtilis* (PDB code: 1QGQ) (*58*), which is characterized by central, twisted parallel beta

sheets that are surrounded by alpha helical structures conforming to a Rossmann fold (Figure 7, A). The overall subunit structure is comprised of two distinct N- and C-terminal domains that are responsible for donor- and acceptor substrate binding (*57*). However, since there is only one central beta strand and the two domains are closely associated, this fold is often considered a one-domain fold. Additionally, the GT-A fold usually possesses a DXD motif, where the carboxylate groups of aspartic acid are involved in divalent metal coordination (*50*).

The GT-B fold is represented by the first reported three-dimensional structure for a glycosyltransferase, the β-glycosyltransferase of the bacteriophage T4 (PDB code: 1JG7), which was solved in 1994 (*59*). In contrast to the GT-A fold, the GT-B consists of two domains of a Rossmann fold-like architecture with central parallel twisted beta sheets and helical structures above and below the plane of the central beta strand (Figure 7, B) (*50, 60*). The two Rossmann-like domains, namely the N- and C-terminal domains, are connected *via* a flexible linker. Localized between the two domains in a cleft is the active site (*50*). Despite high variability in the primary amino acid sequence, the secondary and tertiary structure is highly conserved in the GT-B fold (*57, 61, 62*). As in the GT-A fold, the C-terminal domain is the donor substrate-binding domain and the N-terminal domain accommodates the acceptor substrate (*50*). Not surprisingly, the N-terminal domain (acceptor-binding domain) is more variable than the C-terminal domain (donor-binding domain), which accounts for the high similarity of the donor substrates and large variety of acceptor molecules that can be utilized by glycosyltransferases (*63*). Notably, the GT-B fold does usually not exhibit a DXD motif for divalent metal coordination, in contrast

to the GT-A fold. Recently, a third fold, GT-C, has been predicted but not yet experimentally proven (*50*).

A   GT-A   **B**   GT-B
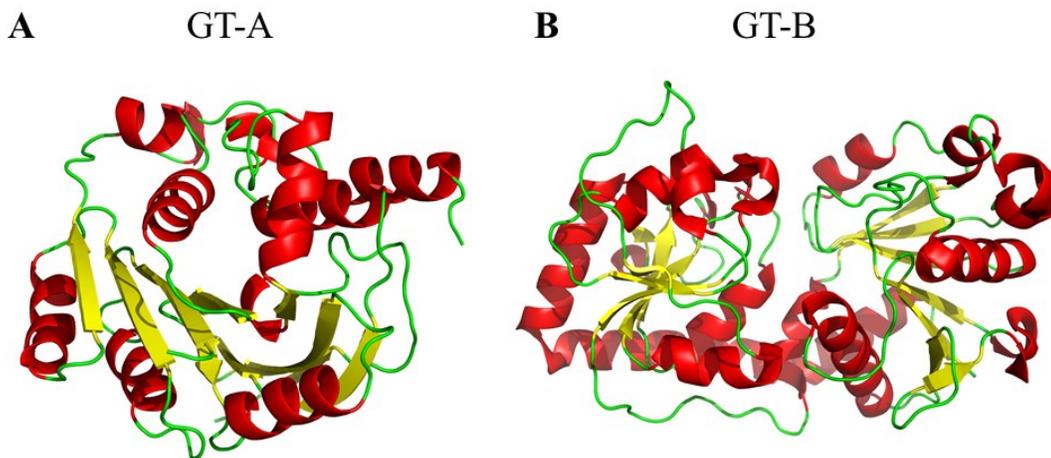
Figure 7: Cartoon representation of different folds of glycosyltransferases. (A) illustrates the GT-A fold of β-glycosyltransferase from bacteriophage T4 (PDB code: 1JG7) and (B) shows the GT-B fold of SpsA from *Bacillus subtilis* (PDB code: 1QGQ). Figures generated using PyMol (Schrödinger, LLC).

1.5.2 Glycosyltransferases: Mechanism

Glycosyltransferases catalyse the formation of a glycosidic bond between a variety of acceptor molecules with nucleophilic groups such as oxygen (O-glycosidic bond), sulfur (S-glycosidic bond), nitrogen (N-glycosidic bond), and even carbon (C-glycosidic bond) (*50*). The formation of a C-glycosidic bond is remarkable because of its resilience against hydrolysis reactions, enzymatic cleavage, as well as for its rarity among glycosyltransferases found in nature (*43, 64*).

During formation of the glycosidic bond, the configuration of the anomeric carbon of the donor sugar substrate is either inverted or retained in the final glycosylated product. Accordingly, glycosyltransferases are classified as retaining or inverting glycosyltransferases (*50*). The mechanism adopted by glycosyltransferases is independent of the fold (GT-A or GT-B): both folds can provide the structural basis for retaining and

inverting mechanisms (*65*). Figure 8 illustrates the differences between the stereochemical outcomes of inverting and retaining glycosyltransferases.
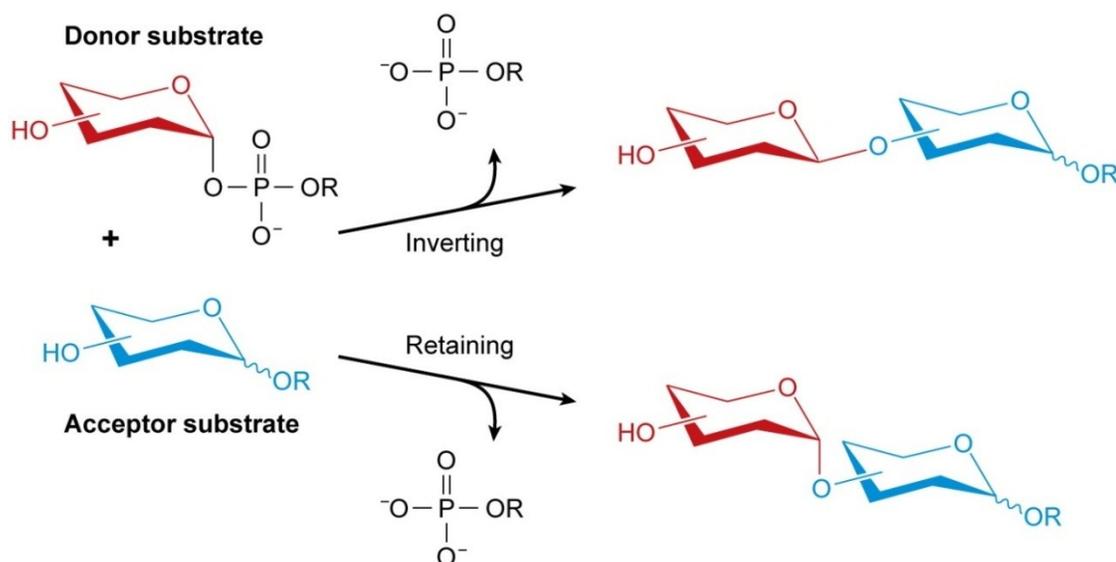
Figure 8: Illustration of the difference between retaining and inverting glycosyltransferases regarding the configuration at the anomeric carbon of the glycosyl donor substrate before and after the glycosylation reaction (*50*).

To date, the mechanism of retaining glycosyltransferases has not been completely elucidated (*63*). Two mechanisms have been suggested: (*i*) a double displacement mechanism (S_N2), with a transitional covalently linked glycosyl intermediate in the active site (*66*) and (*ii*) an S_Ni-like mechanism, where a oxocarbenium ion intermediate is formed and stabilized in the active site. The phosphate leaving group then activates and positions the incoming nucleophile for an attack of the anomeric carbon conserving the stereochemistry with respect to the anomeric carbon (Figure 9, A) (*50, 67*). However, despite extensive studies performed since the S_N2 mechanism was proposed in 1953, an intermediate has not been trapped. In contrast, a S_Ni-like mechanism is now supported by quantum mechanical calculations to be energetically favoured (*50, 68*). This does not

exclude the possibility that both mechanisms can be adopted by retaining glycosyltransferases (*50*).

The mechanism of inverting glycosyltransferases is well established to be a straightforward bimolecular nucleophilic substitution (S_N2) (Figure 9, B) (*69*). The incoming nucleophile of the acceptor is deprotonated by a base catalyst located in the active site of the glycosyltransferase. The nucleophile then attacks the anomeric carbon of the carbohydrate moiety of the donor opposite the pyrophosphate group and "substitutes" it. This results in the observed inversion of the configuration at the anomeric carbon from an α- to a β-configuration, hence glycosylating enzymes employing this mechanism are referred to as inverting glycosyltransferases (*50*).

Even though inverting glycosyltransferases with either the GT-A fold or the GT-B fold use the same S_N2 mechanism during catalysis, the method to facilitate the departure of the negatively-charged pyrophosphate leaving group is different for these two folds (*63*). As mentioned earlier, the GT-A fold generally features a DXD motif to coordinate divalent metal ions. In β-1,4-galactosyltransferase, the metal ion (usually Mn^{2+}) controls the sequential order of events, where Mn^{2+} first binds and then facilitates binding of the sugar donor. The binding of the sugar donor is followed by loop closure, subsequently creating the binding site for the acceptor. Ultimately, during the bimolecular nucleophile substitution reaction, the metal ion stabilizes the developing negative charge of the phosphate leaving group (*70*). In contrast, the active site of GT-B glycosyltransferases is different. Here the glycosylation reaction is catalysed in a metal-independent manner. The departure of the leaving group in these enzymes is stabilized by a helix dipole or a positively charged amino acid side chain (*71*).
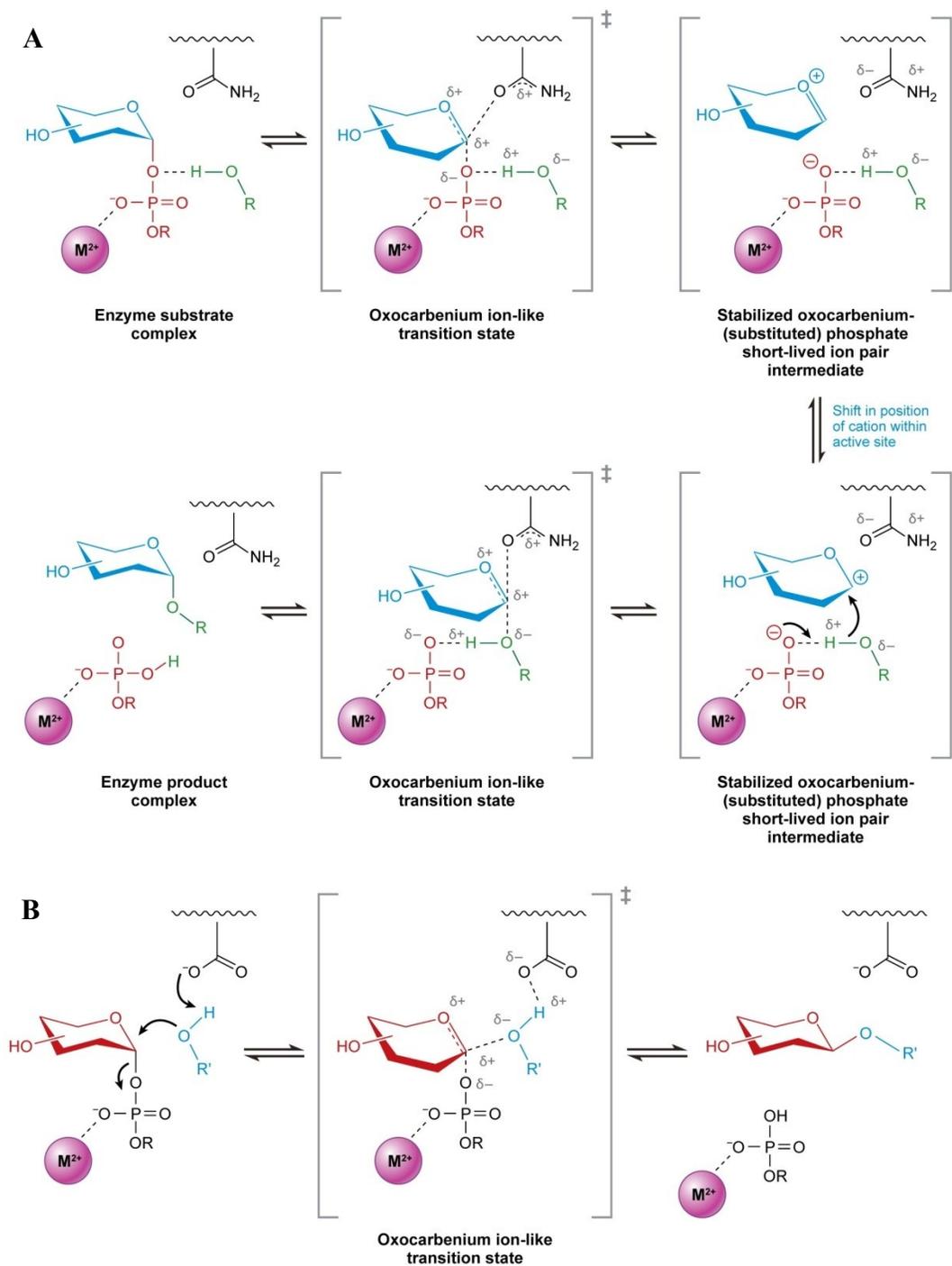
Figure 9: (A) S_ni like mechanism that is most likely utilized by the majority of the retaining glycosyltransferases and **(B)** illustrates the S_N2 mechanism employed by inverting glycosyltransferases (*50*).

Central to the investigation of a S_N2 mechanism in glycosyltransferases is (*i*) the elucidation of the identity of the base catalyst that deprotonates the incoming nucleophile and, (*ii*) elucidation of the mechanism involved in stabilization of the pyrophosphate leaving group of the donor (*50*).

1.6   IroB – a pathogen-associated C-glycosyltransferase

IroB is a C-glycosyltransferase that is a member of the large Family 1 (GT1) along with glycosyltransferases from many bacteria, fungi, mammals, plants and viruses. The database CAZy (www.cazy.org) identifies IroB (EC 2.4.1.x) as an inverting glycosyltransferase that transfers NDP-activated sugars. The fold of IroB has been predicted to be that of the GT-B subfamily (*55*).

IroB shows significant sequence identity (30%) to the glycosyltransferase CalG3 from *Micromonospora echinospora* (PDB code: 3D0R). This enzyme is involved in the glycosylation of calicheamicin, a prominent antitumor agent (*72, 73*). IroB also exhibits similarity (sequence identity = 23%) to UrdGT2 (PDB code 2P6P), a glycosyltransferase involved in biosynthesis of the urdamycin (*74*). Both of these IroB orthologues have been crystallized and have been shown to possess a GT-B fold with characteristics consistent with the inverting class of glycosyltransferases.

It has been clearly demonstrated by *in vitro* and *in vivo* studies that IroB alone is sufficient and necessary for ENT glycosylation (*43, 48*). IroB catalyses the sequential transfer of glucose moieties from donor UDP-glucose to ENT on a given 2,3-DHB (catechol) subunit, resulting in mono-, di- or triglycosylated ENT. During catalysis, a

C-glycosidic bond between the anomeric carbon of the glucose moiety and C_5 of an ENT

catechol is formed. The reaction catalyzed by IroB is shown in Figure 10.
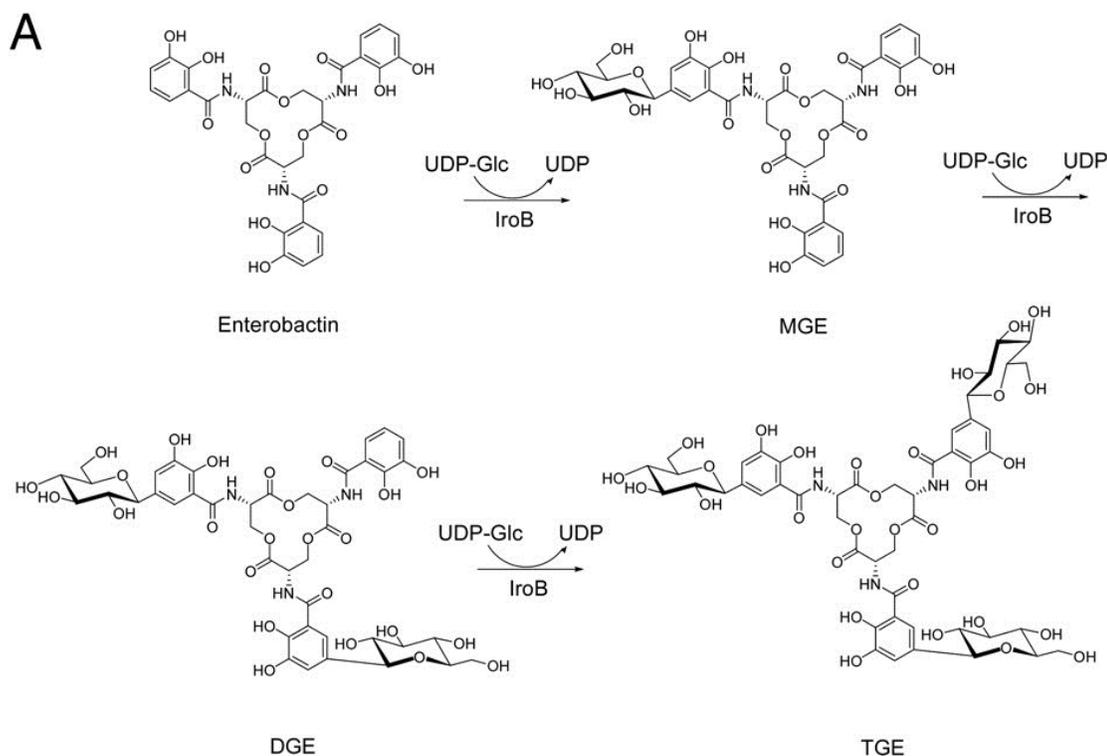
Figure 10: IroB glycosylates ENT in the presence of UDP-glucose by formation of a C-C bond between the anomeric carbon of glucose and C_5 of catechol group of ENT resulting in mono-, di- or triglycosylated ENT (*43*).

The identity of the C-glycosidic bond as well as the exact position in the DHB subunit

has been confirmed by extensive NMR studies by two different groups (*43, 48*).

Fischbach *et al.* also found a coupling constant (J = 10 Hz) between C1′ and C2′ protons

consistent with a β-configuration of the glycoside linkage. This means that the

configuration of UDP-α-glucose was inverted to a β-configuration during catalysis,

suggesting that IroB is indeed an inverting glycosyltransferase (*43*).

The glycosylation mechanism likely to be adopted by this type of glycosyltransferases is

that of an S_N2 mechanism. The mechanism requires the formation of: (*i*) a carbon

nucleophile that is able to attack the anomeric carbon of the glucose moiety and is referred to as direct glycosylation, or (*ii*) an O-glycosylation at the hydroxyl of the DHB subunit followed by a O-C$_5$ rearrangement (*43*). Both mechanisms for C-glycosylation are supported by chemical models, where mechanism (*ii*) has been proposed mainly for *ortho*-substituted C-glycosylated products. The latter proposal is supported by chemical models involving a Lewis acid catalyst that have shown O-C migration to mainly *ortho* and minorly *para* positions with regard to the hydroxyl group (*75-77*). The site of glycosylation in ENT, however, is *para* and *meta* with respect to hydroxyl groups of 2,3-DHB. Mechanism (*ii*) would thus require a regiospecific, non-enzymatic step that directs the glucose moiety *para* to 2-hydroxyl of 2,3-DHB. This seems unlikely compared to direct C-glycosylation that would occurs in mechanism (*i*) (*43*).

Direct C-glycosylation is similar to a Friedel-Crafts electrophilic, aromatic substitution (*78*) and is enhanced by *ortho-* and *para*-directing aromatic substituents such as electron-rich oxygens of a hydroxyl group that donate electrons to the aromatic system (*79-81*). Fischbach *et al.* suggested that the *ortho/para*-directing 2-OH of ENT is deprotonated in the IroB active site (Figure 11). Thus the oxygen could donate electrons to the aromatic system, resulting in formation of a phenolate anion that is partially delocalized to the *para* C$_5$, rendering this position nucleophilic. In the IroB active site, this C$_5$ position of the DHB subunit might be located close to the glucose moiety facilitating attack of the anomeric glucose carbon. Such a mechanism would be consistent with the common S$_N$2 displacement mechanism of inverting glycosyltransferases (*43, 50*). However, attempts to capture the predicted nucleophile in D$_2$O in the absence of the glucose moiety of UDP glucose have not been successful (*43*). A deuterium exchange at C$_5$ would have proven

the existence of the C_5 carbon anion (*43*). The lack of exchange does not exclude the possibility that the carbon anion is being formed in the presence of the electrophilic glucose moiety of UDP-glucose (*43, 78*).

In conclusion, the work of Fischbach *et al.* (*43*) predicts the presence of a base catalyst residue in the active site of IroB that deprotonates 2-OH of DHB subunit of ENT, thus rendering nucleophilic the C_5 located *para* to the 2-OH group of DHB (Figure 11).
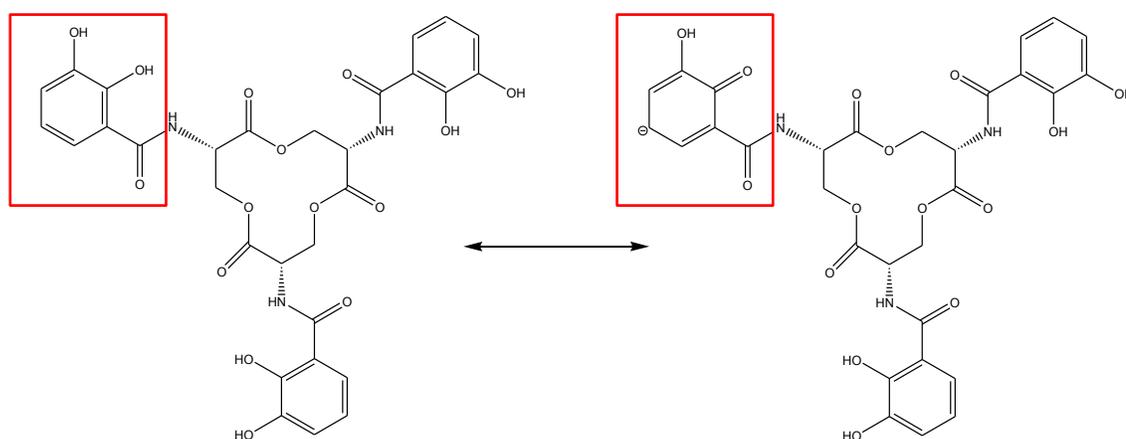
Figure 11: Fischbach *et al.* suggested that ENT is deprotonated by active site residue in IroB. Deprotonation could lead C_5 carbanion by the *para*-quinonoid resonance contributor.

2   Research objective

The overall major aim of this research project is the identification of possible base catalysts in the active site of IroB. In order to achieve this aim, it is essential to predict the three dimensional structure of IroB to locate the potential base catalysts in the active site. The research described in this thesis can be described by four major goals outlined below:

(1) The first major goal was to predict the three dimensional structure of IroB using homology modelling, based on the recently published crystal structure of UrdGT2 (PDB code: 2P6P) (*74*). The primary amino acid sequence of 2P6P shares ~22% sequence identity with IroB, and is thus a promising template to predict the three dimensional structure of IroB. The predicted structure will then be used as a guide and possible base catalysts will be identified.

(2) The second major goal was the expression and purification of IroB, as well as the establishment of an *in vitro* activity assay for IroB using the published methods of Fischbach *et al.* as a guide (*43*). Mutagenesis of predicted base catalysts, identified in the homology model of IroB, could then be performed to assess the influence of the identified residues on catalysis in comparison with wild-type IroB. These experiments provided valuable information about the role of various residues during IroB catalysis.

(3) The third major goal involved the biophysical characterization of wild-type IroB and mutants, and is thus highly complementary to first and second goals outlined above. Experimental verification for the homology model was attained *via* complementary biophysical methods such as circular dichroism and fluorescence spectroscopy.

Additionally, the overall structural integrity of mutagenized IroB variants was confirmed by these methods.

(4) The fourth and final major goal involved the development of an acceptor (ENT) binding assay to assess the effects of mutations performed on IroB. The apparent equilibrium binding affinity constant (K_D') for ENT binding to IroB (wild-type and variant proteins) was determined using a published method for this purpose (*82*). These data provided further information about the structural integrity of the variants and their ability to bind ENT, as well as further information about the role of mutagenized residues for binding and catalysis.

The biochemical and biophysical characterizations of IroB described in this thesis advances our understanding of the mechanism of an uncommon yet important class of glycosyltransferases that transfers sugar moieties *via* formation of a C-glycosidic bond.

As mentioned elsewhere, this type of bond is not susceptible to chemical and enzymatic hydrolysis and thus considered stable. These studies comprise an unprecedented investigation into structure and mechanism of a C-glycosyltransferase from a clinically relevant extraintestinal pathogen.

3 Materials

In this chapter all reagents used in the course of this work are listed. Additionally, composition of all buffers and solutions is described.

3.1 Reagents, kits and enzymes

In the following table (Table 2) all chemicals, reagents and kits are listed under specification of the manufacturer and catalog number and thus defining their purity.

Table 2: important reagents, kits and enzymes are listed with manufacturer and catalog number. If the country is not mentioned the company is located in Canada.

Product	Manufacturer	CAT#
2-Propanol	Thermo Fisher, Waltham, MA, USA	A416-4
a, a, Dipyrdiyl	Thermo Fisher, Waltham, MA, USA	D-95
Acetic acid	Thermo Fisher, Waltham, MA, USA	351271-212
Acetonitrile	Thermo Fisher, Waltham, MA, USA	A998-4
Acrylamide-Bis-Acrylamide 19-1 Mixture	Quantum Biotechnologies, Montreal, PQ	BIAC11
Agarose	BioShop Canada Inc., Burlington, ON	AGA001.500
Ammonium Persulfate	BioShop Canada Inc., Burlington, ON	AMP001.25
Ammonium Sulfate	Thermo Fisher, Waltham, MA, USA	BP212-212
Bio-Tryptone	BioShop Canada Inc., Burlington, ON	TRP402.1
Blotting Grade Blocker, Non-Fat Dry Milk	Biorad, Hercules, CA, USA	170-6404
BME	BioShop Canada Inc., Burlington, ON	MER002.100
Bromophenol Blue	Biorad, Hercules, CA, USA	161-0404
$CaCl_2 \times 2H_2O$	BioShop Canada Inc., Burlington, ON	CCL444
Centrifugal Devive Nanosep 10K Omega	PALL Life Science, Ann Arbor, MI, USA	OD010C34
Coomassie Brilliant Blue G-250	Biorad, Hercules, CA, USA	161-0406
Desalting Columns Pac 10 DG	Biorad, Hercules, CA, USA	732-2010
D-Glucose	BioShop Canada Inc., Burlington, ON	GLU501.1

Product	Manufacturer	CAT#
DMSO	Caledon Laboratories, Georgetown, ON	4100-1
DTT	BioShop Canada Inc., Burlington, ON	DTT001.1
EDTA	BioShop Canada Inc., Burlington, ON	EDT001.500
ENT	Sigma Aldrich, St. Louis, MO, USA	M3910-1MG
Ethanol	n/a	n/a
Ferric Chloride	BioShop Canada Inc., Burlington, ON	FER002.250
Formic acid	Sigma Aldrich, St. Louis, MO, USA	94318
Glycerol	Thermo Fisher, Waltham, MA, USA	BP229-4
Glycine	BioShop Canada Inc., Burlington, ON	GLN001.5
Goat Anti-Rabbit IgG (H+L)-HRP Conjugate	Biorad, Hercules, CA, USA	170-6515
HEPES	BioShop Canada Inc., Burlington, ON	HEP001.1
iBlot Gel Transfer stacks	Invitrogen, Carlsbad, CA, USA	IB3010-02
Imidazole	BioShop Canada Inc., Burlington, ON	IMD508.500
iProof HighFidelity	Biorad, Hercules, CA, USA	172-5301
IPTG	BioShop Canada Inc., Burlington, ON	IPT001.10
Kaleidoscope, prestained standards	Biorad, Hercules, CA, USA	161-0324
Kanamycin	BioShop Canada Inc., Burlington, ON	KAN201.10
KCl	n/a	n/a
KH_2PO_4	BioShop Canada Inc., Burlington, ON	PPM666.1
Magensium Cloride	BioShop Canada Inc., Burlington, ON	MAG520.500
Magnesium Sulfate	BioShop Canada Inc., Burlington, ON	MAG522
Methanol	Caledon Laboratories, Georgetown, ON	6700-3
Na_2HPO_4	BioShop Canada Inc., Burlington, ON	SPD579.1
NaCl	n/a	n/a
NAP5	GE Healthcare, Little Chalfont, Buckinghamshire, UK	17-0853-02
NH_4Cl	BioShop Canada Inc., Burlington, ON	AMC303
Ni-NTA Superflow	Quiagen, Hilden, NRW, Germany	30430
OTG	BioShop Canada Inc., Burlington, ON	OGP202.5
PfuTurbo	Agilent, Santa Clara, CA, USA	600250-52
Plasmid DNA Kit	Bio Basic Inc., Markham, ON	BS414/100Preps
Protease Inhibitor Cocktail	BioShop Canada Inc., Burlington, ON	PIC002

Product	Manufacturer	CAT#
Restriction Enzymes	NEB, Ipswich, MA, USA	n/a
Reverse Transcriptase	NEB, Ipswich, MA, USA	M0253L
RP column		
SDS, 20% solution	BioShop Canada Inc., Burlington, ON	SDS002.500
Sodium Bicarbonate	BioShop Canada Inc., Burlington, ON	SOB999.500
Sodium Hydroxide	n/a	n/a
Spectra/Por Dialysis Membrane MWCO 25 kDa	Spectrum Laboratories Inc., Rancho Dominguez, CA, USA	132554
Sucrose	BioShop Canada Inc., Burlington, ON	SUC507.5
SV Total RNA Isolation Kit	Promega, Madison, WI, USA	Z3100
SYBR Safe DNA gel stain	Invitrogen, Carlsbad, CA, USA	S33102
TCEP	BioShop Canada Inc., Burlington, ON	TCE101.10
TEMED	BioShop Canada Inc., Burlington, ON	TEM001.25
TEV protease	Promega, Madison, WI, USA	V6051
TFA	Acros Organics, Geel, Antwerp, Belgium	13972-1000
Tris	Bio Basic Inc., Markham, ON	TB0196
Tween 20	BioShop Canada Inc., Burlington, ON	TWN510.250
Uridine5'-diphosphoglucose disodium salt (UDP)	Sigma Aldrich, St. Louis, MO, USA	U4625-25MG
Yeast Extract	BioShop Canada Inc., Burlington, ON	YEX401.1
ECL	Thermo Fisher, Waltham, MA, USA	32109

3.2 *E. coli* strains

The strain DH5α (Invitrogen) bears following genotype and was used for cloning:

fhuA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17

Because of *recA1* and *endA1* mutations internal recombination events and endonuclease activity are reduced resulting in high plasmid yield.

For the expression of IroB the *E. coli* strain BL21 (DE3) was used. BL21 (DE3) possesses following genotype:

F- *ompT hsdS*B (rB-mB-) *gal dcm* (DE3)

This strain codes for the T7-polymerase under control of a *lac*UV5 promoter. It is localized on a Lambda prophage in the *att*λ. Furthermore, BL21 (DE3) is well suited for protein expression because of its minimized protease activity.

The strain Nissle 1917 (Mutaflor) is listed under DSM No. 6601 in the "German Collection for Microorganisms and Cell Cultures" (http://www.dmsz.de) and was kindly provided by Klaus Hantke (University of Tübingen, Germany). Nissle 1917 is non-pathogenic due to lack of virulence factors such as alpha-hemolysin, P-fimbrial adhesins, and the semirough lipopolysaccharide phenotype, but harbours the iron acquisition system *iroA* (*83*).

The *E. coli* strain AG1, a common cloning strain distributed by Stratagene, was used as negative control (*iroA*⁻) and harbours the following genotype:

recA1 endA1 gyrA96 thi-1 hsdR17 (*r*K− *m*K+) *supE44 relA1*

3.3 Plasmid

The expression vector pET24b (Novagen) was used throughout these studies and was adapted accordingly. This vector possesses a f1 origin of replication and kan^R selection marker. Genes encoding recombinant proteins were cloned under control of a T7 promoter in-frame with C-terminal His$_6$ tag.

3.4 Genomic DNA CFT073

Genomic DNA from uropathogenic *E.coli* CFT073 (ATCC700928D-5) was obtained from the ATCC (ATCC, Manassas, VA) and stored at -20 °C.

3.5 Polyclonal anti-IroB antibody (rabbit)

The polyclonal antibody anti-IroB was raised using the services of Open Biosystems (Open Biosystems Inc., Huntsville, AL). The antibody was first raised in a 90-day protocol in rabbit PA4061 (Project number: PFS221108) and then further affinity-purified against purified, recombinant IroB. The purification resulted in ~ 5 mL purified antibody solution (concentration 0.464 mg×mL^{-1}, Titer 5 ng), which was then stored in 50% (v/v) glycerol with a final concentration of 0.232 mg×mL^{-1} at -80 °C.

3.6 Primer

Primers were obtained from BioCorp DNA Inc, Canada. Lyophilized primers were resuspended in dH$_2$0 to a final concentration of 100 µM and stored at -20 °C. Primer sequences and modifications are listed in Table 3.

Table 3: List of primers used for plasmid construction including sequence and modifications. Restriction recognition sites are underlined.

Primer	Sequence (5'-3')
nMCS(p1) 5' phosphorylated	TATGCATCATCATCATCATCACAGCAGCGGCAGAGAAAAC TTGTATTTCCAGG
nMCS(p2) 5' phosphorylated	GATCCCTGGAAATACAAGTTTTCTCTGCCGCTGCTGTGAT GATGATGATGATGCA
IroB (NH6/TEV-387) FW	GGATCCATGCGTCGACTGCCTGATTTAGATCGTC
IroB (NH6/TEV) RV	CTCGAGTTACCCTTTCTGTACCATTGTGATCAGGC
IroB (NH6/TEV-371) FW	GGATCCATGCGTATTTTGTTTGTAGGCCCGC
IroB (CH6) FW	GGAATTCCATATGCGTATTTTGTTTGTAGGCCCGC
IroB (CH6) RV	GATCGAATTCGGCCCTTTCTGTACCATTGTGATCAGG
W264L FW	GTTCGCCTGGTTGACTTGATACCCATGGGTGTC
W264L RV	GACACCCATGGGTATCAAGTCAACCAGGCGAAC
H138L FW	GCCATACCCCCTGGCTTATCAGGGGTGTAAC
H138L RV	GTTACACCCCTGATAAGCCAGGGGGTATGGC

Primer	Sequence (5'-3')
E60L FW	CCCGGTCTGGATTCCTTAGCTGGTTACCGCCA
E60L RV	TGGCGGTAACCAGCTAAGGAATCCAGACCGGG
D304N FW	GGCCAGGGAGCCAATCGCCCGGTAA
D304N RV	TTACCGGGCGATTGGCTCCCTGGCC
H6566A FW	CCGAAGCTGGTTACCGCGCCGCTGAGGCACAGCGGAAAA
H6566A RV	TTTTCCGCTGTGCCTCAGCGGCGCGGTAACCAGCTTCGG
E67A FW	GTTACCGCCACCATGCGGCACAGCGGAAAAA
E67A RV	TTTTTCCGCTGTGCCGCATGGTGGCGGTAAC

3.7 Culture Media

As necessary, kanamycin (concentration 30-50 µg/ml) was added after liquid media was cooled down to ~50 °C after autoclaving. The pH of culture media was adjusted by the addition of 1 M NaOH. The Inductor solution was sterilized using a 0.2 µm syringe filter.

Inductor solution

IPTG	1 M	in dH_2O

LB-Medium (pH 7.0)

Peptone	10 g×L^{-1}	
Yeast extract	5 g×L^{-1}	
NaCl	5 g×L^{-1}	in dH_2O

LB-Agar (50 µg×mL^{-1} Kanamycin)

Agar	1.5 % (w/v)	in LB-Media

2x YT-Medium (pH 7.0)

Peptone	16 g×L^{-1}	
Yeast extract	10 g×L^{-1}	
NaCl	5 g×L^{-1}	in dH_2O

M9 Medium

Na_2HPO_4	44.74 mM
KH_2PO_4	22.04 mM
NaCl	8.56 mM
NH_4Cl	18.7 mM
$MgSO_4$	2 mM
$C_6H_{12}O_6$	22 mM
$CaCl_2$	0.1 mM

3.8   Buffers and solutions

In this chapter all buffers and solutions are defined under specification of component, their final concentration and solvent. Their use is described in Chapter 4.

3.8.1   Purification buffer

Preparation and adjustment of pH was carried out at RT. Buffer were then sterile filtered, degassed and stored at 4 °C, not longer than 3 days.

Cell Disruption/Binding Buffer

Tris-Cl (pH 8.5)	20 mM	
PIC (100x)	1x	
NaCl	500 mM	
Imidazole	20 mM	
TCEP	2 mM	
Sucrose	500 mM	in dH_2O

Wash Buffer

Tris-Cl (pH 8.5)	20 mM	
KCl	25 mM	
Imidazole	40 mM	
TCEP	1 mM	
Sucrose	1 M	in dH_2O

Elution Buffer

Tris-Cl (pH 8.5)	20 mM	
KCl	25 mM	
Imidazole	250 mM	
TCEP	1 mM	
Sucrose	1 M	in dH$_2$O

EDTA

EDTA (pH 8.0)	100 mM	in dH$_2$O

Dialysis Buffer

Tris-Cl (pH 8.5)	20 mM	
KCl	25 mM	
TCEP	1 mM	
Sucrose	1 M	in dH$_2$O

Dialysis Membrane Washing Solution

CHNaO$_3$	25 mM	
EDTA	25 mM	in dH$_2$O

3.8.2   Buffers for Activity Assay

The pH of buffers was adjusted at RT. Buffer were degassed, sterile filtered using a filter

with 0.2 µm pore size and stored at 4 °C.

<u>A</u>ctivity <u>A</u>ssay <u>B</u>uffer (AAB1) (used in LC-MS)

Tris-Cl (pH 7.5)	75 mM	
TCEP	2.5 mM	
MgCl$_2$	5 mM	in dH$_2$0

Activity Assay Buffer (AAB2) (used for comparison of variants)

HEPES-NaOH (pH 7.5)	75 mM	
TCEP	2.5 mM	
Glycerol	50% (v/v)	in dH_2O

UDP-Glucose Stock

Tris-Cl (pH 7.5)	75 mM	
UDP-Glucose	100 mM	in dH_2O

Enterobactin Stock

Enterobactin	14.9 mM	in DMSO

3.8.3 Buffers for Fluorescence-Based Quenching Assay

FBQA-Buffer

HEPES-NaOH (pH 7.5)	75 mM	
TCEP	2.5 mM	
$MgCl_2$	5 mM	in dH_2O

Enterobactin

Enterobactin (Sigma)	2.98 mM	in DMSO

3.8.4 SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE)

In this chapter buffers are listed that were used in the course of electrophoretic separation of proteins using SDS-PAGE. Colloidal Coomassie Blue (Blue Silver (*84*)) stain was used for detection of proteins in a polyacrylamide gel.

Running Buffer (10x)

Tris	30.3 g×L^{-1}	
Glycine	144 g×L^{-1}	
SDS	1% (w/v)	in dH_2O

Sample Buffer (5x)

Tris (pH 6.8)	250 mM	
BME	250 mM	
BPB	0.25% (w/v)	
SDS	5% (w/v)	
Glycerol	25% (v/v)	in dH$_2$O

Blue Silver Stain

(NH$_4$)$_2$SO$_4$	100 g×L^{-1}	
Coomassie G-250	1.2 g×L^{-1}	
H$_3$PO$_4$	10% (v/v)	
MeOH	20% (v/v)	in dH$_2$O

3.8.5   Immunodetection

Buffers are listed that were used in the course of immunodetection on nitrocellulose membrane.

BugBuster

Tris-Cl (pH 8.0)	75 mM
DTT	5 mM
OTG	1% (w/v)
PIC (10x)	1x
DNase/RNase	3 µg×mL^{-1}
Lysoyzme	30 µg×mL^{-1}

PBS (10x) (pH 7.4)

NaCl	80 g×L^{-1}	
KCl	2 g×L^{-1}	
Na$_2$HPO$_4$	14.4 g×L^{-1}	
KH$_2$PO$_4$	2.4 g×L^{-1}	in dH$_2$O

PBST (pH 7.4)

Tween 20	0.05% (v/v)	in PBS (1x)

Blocking Solution

    Skim milk powder   5% (w/v)   in PBST

3.9   Software

For the purpose of modelling of protein structures the software Modeller v9.9 (San Francisco, CA) was used (*85-88*). Protein structures were visualized using PyMol (The PyMOL Molecular Graphics System, Schrödinger, LLC) and superposition of protein structures was performed using the software LSQMAN (LSQMAN, Uppsala, Sweden) (*89, 90*). Model structures were validated using the VADAR software suite (*91*) and ProSA-web (*92, 93*). Data plotting and fitting was carried out using the software OriginLab v8.6 (OriginLab, Northampton, MA). Topology maps were produced by TopDraw (*94*). The molar extinction coefficient of protein from the primary amino acid sequence was determined using the ProtParam tool from the ExPASy Server (*95*). Sequence alignments carried out using T-Coffee (*96*) or PSI-Coffee (*97*) and alignments were processed with Jalview (*96, 98*)

4 Methods

In this chapter all the methods are described in detail, underlying functional principles are described as necessary.

4.1 Plasmid construction

The parental vector for all constructs was pET24b. All cloning steps were carried out according to standard cloning methods described in Sambrook *et al.* (*99*). For all cloning steps the *E. coli* strain DH5α was used and final expression construct was transformed into BL21 (DE3).

4.1.1 pET24b-IroB-CH6

The expression construct pET24b-IroB-CH6 was constructed as described by Fischbach *et al*. (*43*). Briefly, the *iroB* gene was amplified from genomic DNA (CFT073) using primer IroB (CH6) FW and IroB (CH6) RV, where the forward primer introduced an *Nde*I site and the reverse primer a *Eco*RI site, respectively. The amplification was performed using iProof DNA polymerase (Biorad) according to standard protocols. After digestion of the PCR product (insert) and the parental vector pET24b (vector) with *Nde*I and *Eco*RI, the insert was ligated into the vector under control of a T7 promoter in-frame with a C-terminal His$_6$ affinity tag. The resulting vector, namely pET24b-IroB-CH6, was transformed into DH5α and sequence identity was verified by sequencing. The protein product will be henceforth referred as IroB-CH6.

4.1.2 pET24b-nMCS

For the constructs pET24b-IroB-NH6/TEV-373 and -389, respectively, the parental vector pET24b had to be modified first to introduce an N-terminal His$_6$ affinity tag, as well as a TEV protease recognition site to allow cleavage of the tag in the final, recombinant protein. For this purpose we designed two complementary primers, namely nMCS(p1) and nMCS(p2), with 5' phosphorylation resembling a *Nde*I and *Bam*HI digested cutting site with corresponding 5' overhang (sticky ends). After annealing of these primers by mixing equimolar amounts, heating to 95 °C followed cooling to RT, the resulting double stranded insert was directly ligated into *Nde*I and *Bam*HI digested pET24b and transformed into DH5α. The intermediate pET24b-nMCS was isolated and served as the parental vector for the constructs described in Section 4.1.3 and 4.1.4.

4.1.3 pET24b-IroB-NH6/TEV-373

The *iroB* gene was amplified from genomic DNA (CFT073) using primer IroB (NH6/TEV-371) FW and IroB (NH6/TEV) RV, which introduced *Bam*HI and *Xho*I restriction sites. For the amplification iProof DNA Polymerase was used. The PCR product (insert) was digested with *Bam*HI and *Xho*I, as well as the parental vector pET24b-nMCS (construction described in 4.1.2). The digested insert and vector were ligated, resulting in the expression construct pET24b-IroB-NH6/TEV-373 with the *iroB* gene in frame with a N-terminal His$_6$ and a TEV protease recognition site. The protein product has a length of 373 aa after removal of the N-terminal hexahistidine and linker region, and is henceforth referred as IroB-NH6/TEV-373. The construct was transformed into *E. coli* DH5α and the sequence identity was verified by sequencing.

4.1.4   pET24b-IroB-NH6/TEV-389

For the construct pET24b-IroB-NH6/TEV-389 the *iroB* gene was amplified with iProof

DNA polymerase using the forward primer IroB (NH6/TEV-387) FW and IroB

(NH6/TEV) RV. The primer FW introduced a *Bam*HI site and the primer RV introduced

an *Xho*I site, respectively. After digestion of the PCR amplified *iroB* gene (insert) with

*Bam*HI and *Xho*I, it was ligated into pET24b-nMCS (construction described in 4.1.2) that

was digested correspondingly. The resulting expression vector pET24b-IroB-NH6/TEV-

389 contained the *iroB* gene with an N-terminal extension in frame with the N-terminal

His$_6$ tag cleavable by TEV resulting in a protein length after cleavage with TEV protease

of 389 aa, and is thus henceforth referred as IroB-NH6/TEV-389. The vector was

transformed into *E. coli* DH5α and verified by sequencing.

4.1.5   Site-directed mutagenesis

To introduce mutations into the *iroB* gene, mismatch PCR was used. To achieve the

desired mutations, 1-3 nucleotides in the middle of the primer were changed to alter the

corresponding codon encoded in the PCR-template. Site-directed mutagenesis was

carried out according to the QuikChange system (Stratagene) using the following primer

pairs W264L FW and RV (Tryptophan 264 to Leucine, H138L FW and RV (Histidine

138 to Leucine), E60L FW and RV (Glutamic acid 60 to Leucine), D304N FW and RV

(Aspartic acid 304 to asparagine), H6566A FW and RV (Histidine 65 and 66 to alanine),

E67A FW and RV (Glutamic acid 67 to alanine). PCR reactions were conducted using

PfuTurbo DNA polymerase according to the standard protocol for this polymerase in a 2

step PCR (18 cycles) with primer annealing and extension at 72 °C for 10 min and using

pET24b-IroB-CH6 as a template. Prior to transformation into DH5α the parental vector was digested with *Dpn*I. All mutagenized expression constructs were verified by sequencing.

4.2   Protein expression

For expression of IroB, the expression vectors pET24b-IroB-NH6/TEV-373, pET24b-NH6/TEV-389, pET24b-IroB-CH6 or mutagenized derivatives were used. These expression constructs contain the gene encoding *E. coli* IroB under control of a T7-phage promoter and were transformed into the expression strain *E. coli* BL21 (DE3). This strain carries the gene encoding T7-polymerase on a λ-prophage under control of a *lac*UV5 promoter; therefore, expression of the transgene can be induced by addition of the synthetic inducer IPTG. BL21 (DE3) cells harbouring pET24b-IroB-CH6 or mutagenized derivatives were streaked on LB-agar plates containing 50 µg×mL^{-1} kanamycin.

An overnight culture grown in LB medium (supplemented with 50 µg×mL^{-1} kanamycin) was inoculated from a single colony from a LB-agar plate and grown to stationary phase at 37 °C and 220 rpm. The expression was then conducted in 3 L Fernbach flasks filled with 1 L 2x YT media (supplemented with 30 µg×mL^{-1} kanamycin), where the initial culture was diluted 1:100 in 2x YT medium. The expression culture was grown at 37 °C and 220 rpm until it reached an $OD_{600\ nm} = 0.4\text{-}0.6$. The expression of the transgene was then induced with 400 µM IPTG and was allowed to proceed for 20 h at 20 °C and 220 rpm. The cells from 1 L culture were then pelleted by centrifugation $6000×g$ for 20 min at 4 °C and stored at -20 °C until use.

4.3 Protein purification/storage

Cells from 1 L culture were resuspended in 30 mL cold Disruption/Binding Buffer (supplemented with PIC prior to use) using a tissue homogenizer and lysed through three passes French Press operated at 16000 to 18000 lb$\times$in^{-2}. During the whole procedure the cell lysate was kept cold. Insoluble matter and cell debris were removed by ultracentrifugation operated at 130000$\times g$, 4 °C for 1 h. Using a BioRad Biologic DuoFlow FPLC, the clarified supernatant was immediately applied to a 10 mL Ni^{2+}-NTA Superflow (Qiagen) column, pre-equilibrated with Disruption/Binding Buffer, at a flow rate of 1 mL$\times$min^{-1} and washed at a flow rate of 3 mL$\times$min^{-1} with 100 mL Disruption/binding Buffer. In the next step the buffer was switched to Washing Buffer (40 mM Imidazole) using a gradient at a flow rate of 3 mL$\times$min^{-1} in a volume of 20 mL and then washed at the same flow rate with 20 mL. The elution was conducted in one step in Elution buffer (250 mM Imidazole) at a flow rate of 3 mL$\times$min^{-1}. Fractions were pooled, immediately placed on ice and EDTA pH 8 to a final concentration of 1 mM was added to chelate any Ni^{2+} leached from the column.

The dialysis membrane (Spectra/POR 6; MWCO: 25 kDa) was in the first instance washed in dH$_2$O and subsequently in 50% (v/v) EtOH. Afterwards the membrane was incubated under slow stirring in Dialysis Membrane Washing Solution for 30 min at RT and finally the membrane was washed in dH$_2$O under slow stirring for 10 min.

The pooled fractions from protein purification were dialysed using the primed dialysis membrane against 4 x 1 L Dialysis Buffer for at least 4 h each under constant stirring at 4 °C. After finishing the dialysis procedure the protein solution was sterile filtered

(0.2 µm pore size) and directly dripped into liquid N_2. The beads followed from this procedure were collected and stored at -80 °C.

4.4 Determination of protein concentration

The protein concentration was determined spectroscopically at 280 nm using molar extinction coefficient of 55,920 $L\times mol^{-1}\times cm^{-1}$ and extinction coefficient 1.316 $mL\times mg^{-1}\times cm^{-1}$, respectively. Extinction coefficients were predicted from primary amino acid sequences using the ProtParam tool (http://web.expasy.org/protparam/). These predicted extinction coefficients were also used for IroB-CH6-E60L, IroB-CH6-E67A, IroB-CH6-D304N, IroB-CH6-H6566A, IroB-CH6-D304N, IroB-NH6/TEV-373 and -389. The predicted molar extinction coefficient of 50420 $L\times mol^{-1}\times cm^{-1}$ and extinction coefficient 1.188 $mL\times mg^{-1}\times cm^{-1}$ was used for IroB-CH6-W264L, reflecting the change in primary sequence. All extinction coefficients were calculated using ProtParam tool from the ExPASy server (*95*).

4.5 Identification of *iroB* gene product

To identify the *iroB* gene product in *E. coli* Nissle 1917 we designed experiments to measure *iroB* expression at the mRNA and protein levels.

4.5.1 RNA-isolation and transcription to cDNA

E. coli Nissle 1917 and AG1 were first grown to saturation in 2x YT medium at 37 °C and 220 rpm overnight. The cells were then pelleted at 2000×*g* at 4 °C and 10 min and washed twice with M9 minimal media to remove excess nutrients from rich medium

before. Finally, Nissle 1917 and AG1 were cultivated at 37 °C and 220 rpm. The OD_{600} at the beginning of the cultivation was 0.83 and was aborted after 5 h when an OD_{600} of 2.41 was reached. Under presumption that 1 A_{600} ~ $8×10^8$ cells×mL^{-1}, ~ $1×10^9$ cells (~ 0.5 mL) were pelleted and used in RNA-isolation procedure (*100*).

The total RNA was isolated from the Nissle 1917 pellet obtained using the SV Total RNA Isolation System (Promega). The standard procedure described in the Quick protocol of the manufacturer was applied and the RNA was finally eluted in 100 µL RNase-free H_2O. The isolated mRNA was stored at -80 °C, whereas 10 µL were directly used in the reversed transcription reaction to obtain cDNA from mRNA. The reaction was set up according to the standard protocol M-MuLV Reverse Transcriptase (NEB), where 10 µL of eluent were mixed with primer IroB (CH6) RV and dNTP mix and heated to 65 °C for 5 min. After this incubation period 400 U of Reverse Transcriptase and the corresponding buffer supplied by the manufacturer were added to a final volume of 20 µL and the reaction allowed to proceed for 1 h at 42 °C in the absence of RNase inhibitor. Subsequently the enzyme was inactivated at 90 °C for 10 min and the cDNA stored at -20 °C until further use.

4.5.1.1 PCR reaction to amplify cDNA

In order to identify the *iroB* gene product, the cDNA resulting from the isolation procedure described in 4.5.1 was amplified with primer pair that was specific for either the *iroB* gene product with a length of 1113 bp or 1161 bp. For this purpose the reactions described in Table 4 were set up. The PCR reaction was conducted according to the

standard protocol for iProof DNA polymerase (Biorad) with a primer annealing temperature of 72 °C, 45 s extension time and 25 cycles.

Table 4: composition of PCR-reactions to identify *iroB* gene product from cDNA and corresponding control reactions to demonstrate primer specificity.

Component	Template DNA			
	cDNA (Nissle 1917)	cDNA (Nissle 1917)	gDNA (CFT073)	gDNA (CFT073)
Template	5 µL	5 µL	1 ng×µL^{-1}	1 ng×µL^{-1}
IroB (NH6/TEV-387) FW	-	0.3 µM	-	0.3 µM
IroB (NH6/TEV) RV	0.3 µM	0.3 µM	0.3 µM	0.3 µM
IroB (NH6/TEV-371) FW	0.3 µM	-	0.3 µM	-
DMSO (100%)	3% (v/v)	3% (v/v)	3% (v/v)	3% (v/v)
dNTP (10 mM)	0.2 mM	0.2 mM	0.2 mM	0.2 mM
iProof DNA-polymerase	0.02 U×µL^{-1}	0.02 U×µL^{-1}	0.02 U×µL^{-1}	0.02 U×µL^{-1}
iProof Buffer (5x)	1x	1x	1x	1x
ddH$_2$O	*ad* 50 µL	*ad* 50 µL	*ad* 50 µL	*ad* 50 µL

The amplified product after exposing the reaction mix to typical PCR temperature cycles was analyzed on a 1% (w/v) agarose gel in TAE buffer according to standard procedure from Sambrook *et al. (99)* and DNA was detected with SYBR Safe DNA stain.

4.5.1.2 Detection of IroB using Western Blot

E. coli Nissle 1917 harbouring the *iroA* cluster and *E.coli* AG1 as the negative control (*iroA*⁻) were grown first grown to saturation in LB media (no antibiotics) at 37 °C and 220 rpm. The cells were then pelleted at 2000×*g* at 4 °C for 10 min and washed once with M9 minimal medium (supplemented with 100 µM 2,2-dipyridyl). The cells were then resuspended in M9 minimal media supplemented with 100 µM 2,2'-dipyridyl, and

cultivated under constant shaking at 220 rpm at 37 °C for 20 h. The OD_{600} at beginning of the cultivation was 0.18 for Nissle 1917 and 0.29 for AG1. After 20 h cultivation Nissle 1917 reached an $OD_{600}=1.43$ under the cultivation conditions, whereas the AG1 cells reached an $OD_{600}=0.75$. The cells were harvested by centrifugation and stored at -20 °C.

For SDS-PAGE and the following immunodetection 100 mg (wet weight) AG1 and Nissle 1917 cells were lysed in BugBuster for 1 h at RT and samples were prepared and separated using 8% polyacrylamide gel according to Laemmli *et al.* (*101*). For comparison IroB NH6/TEV-373 and IroB NH6/TEV-389 were digested with TEV protease according to the manual supplied by the manufacturer and also prepared and separated on the same gel. Subsequently to separation proteins were transferred onto nitrocellulose membrane using iBlot Transfer system (Invitrogen). For immunodetection, the membrane was initially blocked in Blocking solution for 30 min at constant shaking and then incubated with the polyclonal anti-IroB (rabbit) antibody (final antibody concentration 4.64 ng×μL^{-1}) in Blocking solution for 45 min. After the incubation in the primary antibody the membrane was washed 3 times for 10 min with PBST under constant shaking. Then the secondary antibody Anti-Rabbit IgG (goat) conjugated with HRP was applied in a dilution of 1:3000 in Blocking Solution for 1 h under constant shaking. The membrane was finally washed 3 times for 10 min in PBST and subsequently incubated for 2 min in entry-level peroxidase substrate for enhanced chemiluminescence (ECL). Protein bands were visualized on autoradiography film using a short exposure of 1-2 s.

4.6 Homology modeling using Modeller v9.9

Due to the absence of any experimental structural data for IroB, we decided to predict the three-dimensional structure of IroB by homology modeling. The procedure consists basically of fold assignment, target-template alignment, model building, and model evaluation (*85*). A suitable template, the C-glycosyltransferase UrdGT2 from *Streptomyces fradiae* (PDB code: 2P6P), was identified using the FFAS03 server (http://ffas.ljcrf.edu/ffas-cgi/cgi/ffas.pl) (*42-44)* by comparison with the IroB primary amino acid sequence (NCBI reference: ACA34771.1) as published in Tivendale *et al*. (*102*). Target- and template alignment and model building was carried out using Modeller v9.9 according to a protocol published in Current Protocols in Protein Science (*85*). Structural superposition of IroB-model with UrdGT2, as well as the superposition with CalG1 (PDB code: 3OTH) to identify the donor- and acceptor binding site, was performed using LSQMAN; this software was also used to determine RMSD values of superposition (*89, 90*).   The integrity of the IroB model structure was evaluated using the implemented tools of Modeller v9.9 (*29-31)*, VADAR (*91*) and ProSA-web (*92, 93*) tools.

4.7 Activity assay

The activity assay described in the next two sections (4.7.1 and 4.7.2) differs only minimally in terms of the buffer that was used. During establishment of this activity assay in our lab, based on the method described in Fischbach *et al*. (*43*), we realized that IroB heavily precipitates in the absence of glycerol (Section 4.7.1). Therefore we

stabilized IroB in the course of this project and used a different buffer for the comparison of IroB WT and mutants (Section 4.7.2).

4.7.1   LC-MS to detect MGE/DGE/TGE

IroB-CH6 prepared and stored according to method described in 4.3 was thawed and exchanged AAB1 using an equilibrated NAP5 DNA desalting column according to the manual supplied by the manufacturer. Briefly, 0.5 mL of IroB containing protein solution was applied to the column and allowed to enter the resin. The protein was then eluted in 1 mL of AAB1 buffer. Approximately, 5 µM IroB were mixed with 100 µM ENT and 600 µM UDP-glucose and the conversion to MGE/DGE and TGE was allowed to proceed at RT. After 5 min, 40 min and 80 min time points, aliquots were taken, quenched with 10% (v/v) ACN and supernatant subjected to LC-MS analysis after centrifugation (1 min at 21100×g at RT). Unbound sample was removed by washing for 10 min at 0% ACN in 0.05% (v/v) FA in ddH$_2$0 at a flow rate of 300 µL×min^{-1}. ENT and the conversion products MGE/DGE/TGE were then separated in a gradient from 0-80% ACN in 0.05% FA in ddH$_2$O over 30 min at a flow rate of 300 µL×min^{-1} and directly injected in ESI-MS Qtof2 (CBAMS) and mass spectrometric analysis of ENT and IroB products (MGE, DGE, TGE) was performed in positive mode.

4.7.2   Comparison of IroB mutants

IroB-CH6 and corresponding mutants were exchanged in buffer AAB2 using a equilibrated DG10 desalting column according to the minimal dilution protocol supplied by the manufacturer. Specifically, 0.1 to 1.0 mL of protein solution, purified and stored

as described in Section 4.3, was loaded on the column and the liquid was allowed to enter the resin. Dependent on the volume loaded buffer AAB2 was then added to 3 mL and also allowed to reach the surface of the resin. The sample was then eluted with 1.5x volume of the sample volume in AAB2 buffer, and the protein concentration was determined spectroscopically as described in Section 4.4. The enzymatic reaction, in which ENT is eventually converted to MGE/DGE/TGE by IroB, was prepared in PCR tubes in a volume of 100 µL in containing 2 µM IroB, 100 µM ENT, 600 µM UDP-glucose and 1 mM $MgSO_4$ in AAB1 buffer and was allowed to proceed for 1 h at 25 °C. After this time, 80 µL was directly separated by HPLC (Agilent) using a reversed-phase C18 column at a flow rate of 1 $mL \times min^{-1}$. The sample was loaded and unbound material was washed for 10 min in 5% (v/v) ACN in ddH_2O and then separated in a gradient from 5-80% (v/v) ACN in 0.05% (v/v) TFA in ddH_2O over 20 min and was monitored at 254 nm. The chromatogram was exported and reconstructed using the software OriginLab v8.6.

4.7.3   Activity assay in the absence of Mg^{2+} and presence of Fe^{3+}

Divalent metal ions such as Mg^{2+} ions are often needed by glycosyltransferases due to mechanistic requirements to stabilize positively charged leaving groups during catalysis. We also tested the activity in the presence of Fe^{3+} to occupy the hydroxyl groups of the catechol groups of ENT. The ENT molecule binds ferric ions with high affinity and the coordination of iron is characterized by a certain conformation, where ferric ion is coordinated in the center of the three catechols of ENT. Thus, testing the ability of IroB

to glycosylate ENT in the presence of ferric iron provides valuable information about the conformation of ENT during catalysis.

IroB-CH6 was exchanged into AAB2 buffer using NAP5 columns and the protein concentration was determined spectroscopically at 280 nm. The reaction was then set up according to the composition described in Table 5 and incubated for 1 h in bench top centrifugation device with a MWCO of 10 kDA at 25 °C. After one hour the centrifugal device was centrifuged until no liquid was left in the top of the device. It was assumed that the ultrafiltration membrane retained all IroB-CH6 and the reaction was terminated by this step. The Fe-bound ENT is not retained during RP-HPLC separation and thus ferric iron has to be released prior to analysis. For consistency we acidified all samples with 0.5 μL of 4 M HCl to release ferric iron from ENT and analysed 80 μL of the supernatant using RP-HPLC in a ddH$_2$O/ACN gradient acidified with 0.05% (v/v) TFA. The absorbance of ENT was followed at 330 nm.

Table 5: Composition of the reaction mix for the activity assay to test the IroB activity in the presence of Fe^{3+} and in the absence of Mg^{2+}.

Component	Control	Presence of Fe^{3+}	Absence of Mg^{2+}
MgSO$_4$	500 μM	500 μM	-
FeCl$_3$	-	200 μM	-
EDTA	-	-	200 μM
UDP-glucose	600 μM	600 μM	600 μM
ENT	100 μM	100 μM	100 μM
IroB-CH6 (WT)	5 μM	5 μM	5 μM
Buffer AAB2	*ad* 250 μL	*ad* 250 μL	*ad* 250 μL

4.8   Structural integrity of IroB and mutants

The structural integrity of IroB was assessed experimentally by far-UV circular dichroism, by thermal denaturation followed by far-UV circular dichroism, and by fluorescence spectroscopy.

4.8.1   Far-UV circular dichroism

Far-UV CD spectra of IroB were collected using a Jasco-815 spectropolarimeter equipped with a Peltier heating/cooling temperature control system. Spectra were recorded at 20 °C in the far-UV region from 260 nm to 200 nm. In detail, IroB-CH6 or mutants from freeze stock were thawed and adjusted to a concentration to 0.2 mg×mL^{-1} with Dialysis Buffer based on the method described in Section 4.4. The sample was then loaded into a rectangular cell with 0.1 cm path length and spectra were recorded using the following parameters: sensitivity: standard (100 mdeg), data pitch: 0.2 nm, scanning mode: continuous, scanning speed: 20 nm×min^{-1}, band width: 1 nm. For each measurement 5 scans were averaged. The final spectrum was corrected for buffer contributions, smoothed and converted to molar ellipticity using the integrated software package. Ultimately, the spectra were exported and plotted using OriginLab v8.6.

4.8.2   Thermal denaturation

Thermal denaturation was conducted in a Jasco-815 spectropolarimeter, where the same sample that was used in Section 4.8.1 was subjected to thermal denaturation by following the loss of signal at 222 nm with increasing temperature (20 °C to 98 °C). The parameters used were set as follows: data pitch: 0.2 °C, sensitivity: standard (100 mdeg), response

time: 16 s, band width: 1 nm, temperature slope: 30 °C×h^{-1}. The spectrum was then normalized in terms 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 using the 1st derivative function of the instrument's software package, where the maximum between 20 °C and 60 °C determines the T_m.

4.8.3 Fluorescence spectroscopy

Fluorescence spectroscopy was used to assess the integrity of the tertiary structure of IroB-CH6 and mutants. Spectra were recorded at RT using a Varian Cary Eclipse spectrofluorimeter with the Scan Software version 1.1. Protein samples from freeze stock at -80 °C in Dialysis Buffer were thawed and the protein concentration was adjusted to 0.05 O.D. in Dialysis Buffer at 280 nm. Then, 600 μL were transferred in a 1×1 cm Varian cuvette and spectra were recorded from 300 to 400 nm for the excitation wavelength of 280 nm and 295 nm for tryptophan/tyrosine and tryptophan emission, respectively. All spectra were recorded in CAT mode (average of 10 scans), excitation and emission slit width each set to 5 nm, the voltage and scan speed adjusted to 600 V and 600 nm×min^{-1}, respectively. Spectra were then corrected for buffer contribution using the instrument's calculation software, exported and analyzed using Origin software.

4.9 Fluorescence based quenching assay

A fluorescence-based quenching assay was used to determine binding affinity constants for ENT and IroB. IroB-CH6 and mutants were thawed and exchanged into FBQA Buffer

using a DG10 desalting column, which was equilibrated with FBQA Buffer prior to use. IroB-CH6 and mutants were exchanged according to the minimal dilution protocol supplied by the manufacturer. The protein concentration was determined spectroscopically (according to Section 4.4) and was diluted as fast as possible to a concentration of 100 nM and 600 μL were transferred to a 1×1 cm Varian fluorescence cuvette. The fluorescence was recorded using the simple read mode at an excitation wavelength of 280 nm and emission set to 340 nm with an averaging time of 5 s. The excitation and emission slit widths were set to 10 nm, and detector voltage was set to 600 V. The initial fluorescence intensity of 100 nM IroB was always approximately 100 arbitrary units under these settings. Upon titration of 2, 4, 6, 8, 10, 20, 30 and 40 μM ENT, dissolved in DMSO, the fluorescence intensity was measured and corrected for buffer contribution. Correction for inner filter effect was also performed using the following equation (1):

$$F_{corr} = F_{obs} 10^{\frac{A_{exc} + A_{em}}{2}} \tag{1}$$

To obtain apparent equilibrium dissociation constants (K$_D$') an adapted equation based on (2) and (4) was as described in detail recently in van de Weert *et al.* (*103*). The definition of the binding constant K$_D$' is shown in equation (2), where [P] and [L] are the free receptor and free ligand, respectively, and [PL] represents the complex of ligand and receptor.

$$K_D' = \frac{[P] \cdot [L]}{[PL]} \tag{2}$$

By substitution of the free [P] with $[P]_{tot} - [PL]$ and the free [L] with $[L]_{tot} - [PL]$ we can rewrite equation (2) as follows:

$$K_D' = \frac{([P]_{tot} - [PL]) \cdot ([L]_{tot} - [PL])}{[PL]} \tag{3}$$

By rearrangement and expansion of equation (3) we obtain through the intermediate steps (3b) and (3c) the quadratic equation (3d) in the form $ax^2 + bx + c = 0$.

$$K_D' \cdot [PL] = ([P]_{tot} - [PL]) \cdot ([L]_{tot} - [PL]) \tag{3a}$$

$$K_D' \cdot [PL] = [P]_{tot}[L]_{tot} - [L]_{tot}[PL] - [P]_{tot}[PL] + [PL]^2 \tag{3b}$$

$$0 = [P]_{tot}[L]_{tot} - [L]_{tot}[PL] - [P]_{tot}[PL] + [PL]^2 - K_D'[PL] \tag{3c}$$

$$0 = [PL]^2 - (K_D' + [L]_{tot} + [P]_{tot})[PL] + [P]_{tot}[L]_{tot} \tag{3d}$$

Solving of the quadratic equation (3d) for [PA] using a standard solving formula for quadratic equations leads finally to equation (3e):

$$[PL] = \frac{(K_D' + [L]_{tot} + [P]_{tot)}) \pm \sqrt{(K_D' + [L]_{tot} + [P]_{tot)})^2 + 4[P]_{tot}[L]_{tot}}}{2} \tag{3e}$$

The following formula (4) describes the assumption that that the fluorescence we observe F is a linear response depending on the fraction of protein in the bound state $[PA] \times [P]^{-1}$. Furthermore, it represents the assumption that the observed fluorescence F consists of

two components, where one is the free protein with F_0 and the protein in complex with a ligand F_c. Equation (4) can be rearranged to obtain equation (4a).

$$F = F_0 \left[1 - \frac{[PL]}{[P]_{tot}} \right] + F_c \left[\frac{[PL]}{[P]_{tot}} \right] \qquad (4)$$

$$\frac{F}{F_0} = 1 - \left[1 - \frac{F_c}{F_0} \right] \left[\frac{[PL]}{[P]_{tot}} \right] \qquad (4a)$$

Ultimately, equation (3e) is used to substitute [PL] in equation (4a) and results in the equation (5).

$$\frac{F}{F_0} = 1 - \left[1 - \frac{F_c}{F_0} \right] \frac{(K_D' + [L]_{tot} + [P]_{tot)}) \pm \sqrt{(K_D' + [L]_{tot} + [P]_{tot)})^2 + 4[P]_{tot}[L]_{tot}}}{2[P]_{tot}} \qquad (5)$$

Data obtained in the fluorescence quenching experiment were imported into OriginLab v8.6 and fitted to equation (5) to obtain the variables F_c and K_D'.

5   Results

The Results section is divided into three main parts: (*i*) the purification of IroB and variants including the establishment of an activity assay, (*ii*) the identification of the *iroB* gene product in *E. coli* Nissle 1917 and (*iii*) the investigation of the mechanism of IroB leading to the formation of the C-C bond observed during catalysis. In part (*i*) the final purity of IroB and variants was assessed using SDS-PAGE. Enzymatic activity was demonstrated *in vitro* using RP-HPLC to separate and identify conversion products arising from IroB activity. The probiotic, *iroA*-harbouring strain *E. coli* Nissle 1917 was used in part (*ii*) to identify the *iroB* gene product on transcriptional and translational levels using RNA and Western Blotting techniques, respectively. In part (*iii*), a homology model for IroB was generated. The model was used to identify residues potentially involved in catalysis. Such residues identified by our model were mutated, and the enzymatic activities of IroB variants were compared to that of IroB-CH6 (WT). In addition, wild type and IroB variants were characterized biophysically. Apparent equilibrium binding affinity constants (K_D') of ENT binding were determined in order to verify that the structural integrity of IroB each variant was maintained upon mutagenesis.

5.1   Expression and purification of IroB and variants

The purifications of IroB-CH6, IroB-NH6/TEV-373, IroB-NH6/TEV-389 and all variant proteins were carried out as described in the Methods. The purification scheme reported in the Methods section represents the final, optimized expression and purification procedure. In the course of this project the expression was optimized in terms inductor concentration and expression temperature and time to avoid the formation of inclusion

bodies and to maximize the yield of soluble IroB. The purification of IroB according to the published protocol in (*43*) led to heavy precipitation at high concentrations of protein. Thus, we optimized the purification protocol and tested the solubility in the presence of different stabilizers such as amino acids (e.g., L-arginine, L-tryptophan, L-tyrosine, and L-phenylalanine), osmolytes (e.g. glycerol, sucrose, glucose, trehalose), salts (e.g. KCl and NaCl) and by altering buffer pH. We observed that IroB undergoes conformational changes in the absence of stabilizers such as glycerol and sucrose, losing helical structure and adopting a conformation enriched in beta sheet as determined by far-UV circular dichroism (data not shown). This result suggests that unfolding followed by beta-sheet aggregation leads to precipitation of IroB. Ultimately, we maintained the structural integrity of IroB in the presence of glycerol (30-50% (v/v)) or 1 M sucrose or 1 M trehalose at pH 8.5 and low KCl concentrations of 25 mM with no visible precipitation even at relatively high protein concentrations up to 3 mg×mL^{-1}. The far-UV circular dichroism spectrum under these conditions was stable for at least 72 h. We decided on sucrose as our preferred stabilizer since it can be easily removed by dialysis or gel filtration, is less viscous than glycerol, and is the least expensive stabilizer.

The final purification protocol yielded about 20 mg of IroB per liter culture. The final purity of WT IroB and variants after the final dialysis step of the purification procedure can be seen in Figure 12. In lane 1 and 9, the MWM is shown and the height of the 37 kDa band is indicated. IroB-CH6 (MW: 42,498.6 Da) was loaded in lane 2 and 10, IroB-NH6/TEV-373 and IroB-NH6/TEV-389 (MW: 42,572.6 Da and 44,586.0 Da) were loaded in lane 7 and 8, respectively, and in lane 3-6 and 11-16 all the variants are shown (MW: ~ 42,500 Da) that were purified over the course of this project. All of the purified

protein samples, except IroB-NH6/TEV-389, appear as a single dominant band above the

37 kDa Marker band, consistent with the molecular weight calculated based on the

primary amino acid sequence. The variant IroB-NH6/TEV-389 appears above the height

of the 37 kDa band of the MWM as a doublet.
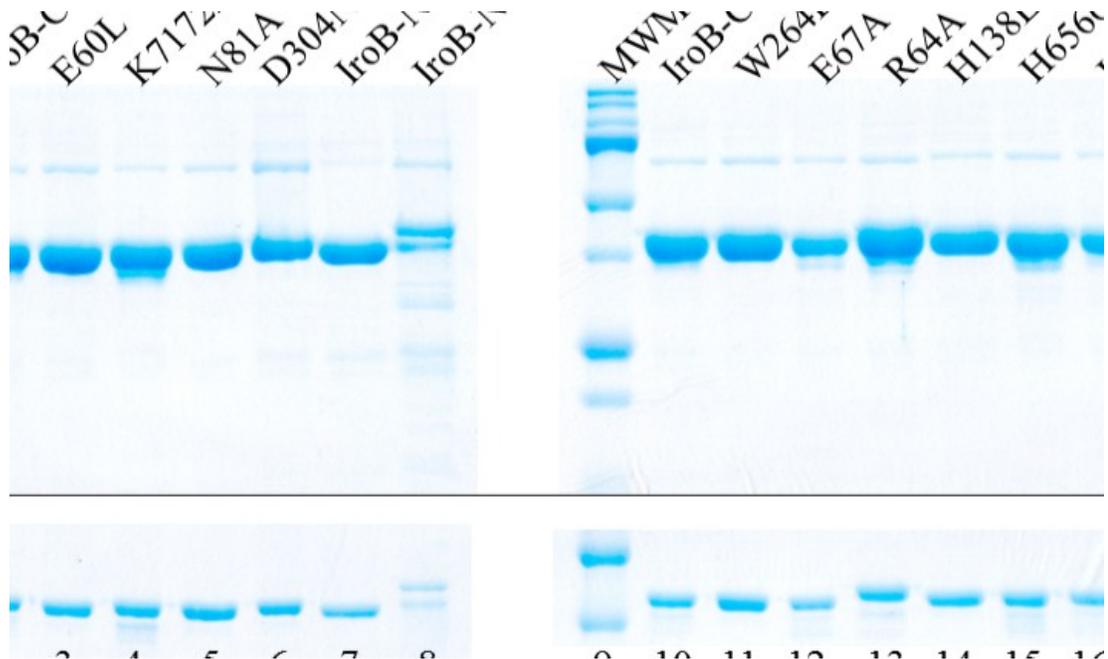
Figure 12: SDS-PAGE (10%) stained with Blue Silver Coomassie-based stain. The upper panel shows an overloaded gel and broader separation range to demonstrate the overall purity of the sample and the lower panel more stringent range providing information for migration properties of different IroB mutants and variants. In lane 1 and 9 the Kaleidoscope Molecular Weight Marker (MWM) was loaded and the height of the 37 kDa band was indicated. IroB-CH6 (lane 2 and 10), IroB-NH6/TEV-373, -389 (lane 7 and 8) and all IroB variants (lane 3-6 and 11-16) were loaded.

The migration properties of the upper band of this doublet is more consistent with the

molecular weight calculated for this variant (44,586.0 Da), whereas the lower band is on

the almost same height with the lighter IroB-NH6/TEV-373 (MW: 42,572.6 Da)

suggesting that the upper band represents the full length IroB-NH6/TEV-389 and the

lower band of the doublet a degradation product of the full-length protein. The purity of

IroB-CH6 and variants is comparable and was estimated to be ≥85% based on

densitometry (data not shown). Purified IroB-CH6 and variants were stored as frozen beads at -80 °C.

5.1.1 IroB enzymatic activity assay and product identity

In order to provide evidence that IroB maintained its structure during the purification and storage procedure, an RP-HPLC activity assay was established to follow formation of MGE/DGE/TGE during IroB catalysis (see 4.7.1). Briefly, IroB was incubated in the presence of UDP-glucose, $MgSO_4$, ENT and the supernatant was analyzed using a reversed phase (RP) column that was directly coupled to ESI/MS to confirm the identity of conversion products.
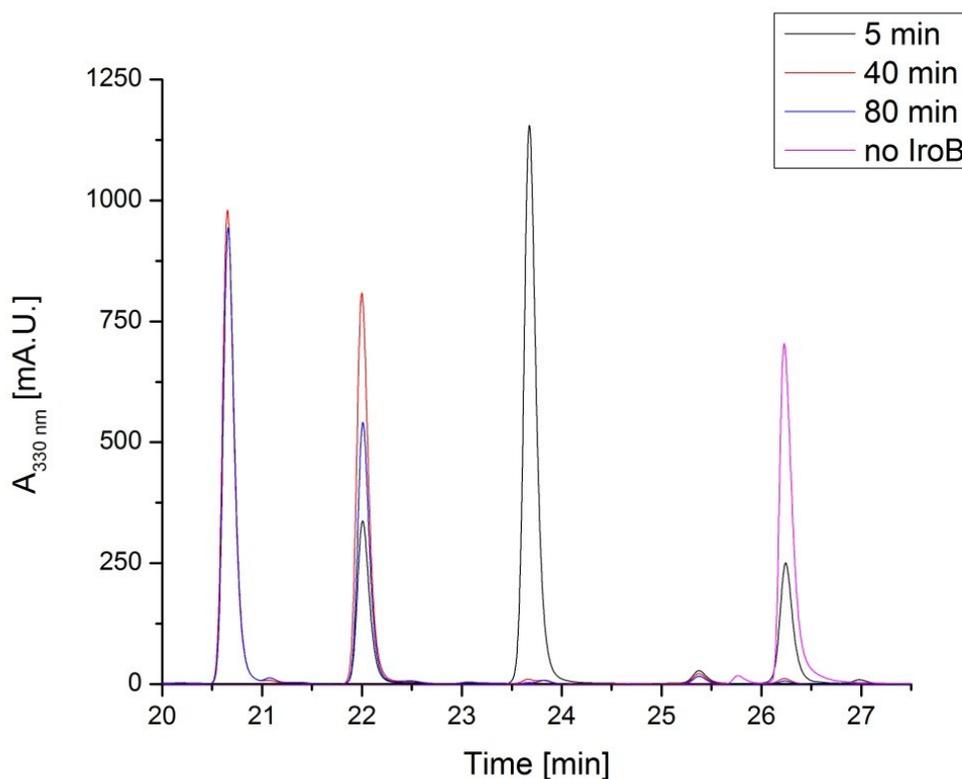
Figure 13: RP-HPLC chromatogram showing the separation of ENT and MGE/DGE/TGE. The separation was conducted on a C18 resin and the absorption of ENT was followed at 330 nm. Supernatant

was analyzed after 5, 40 and 80 min of incubation in the presence of IroB and also in the absence of IroB (IroB⁻).

The RP-HPLC chromatogram in Figure 13 shows the analysis of the reaction mixture at different incubation times upon addition of IroB after 5 min (black line), 40 min (red line) and 80 min (blue line) as well as a control reaction (IroB⁻) in the absence of IroB (pink line). The negative control IroB⁻ contains all the assay components (Mg^{2+}, UDP-glucose and ENT) except IroB. The analysis of the negative control (pink line) shows a single peak with a retention time of 26.25 min representing ENT. No other major peaks are detectable in retention range from 20-27.5 min suggesting that ENT is the only compound of the reaction mix that binds to the RP-column. After incubation time of 5 min (Figure 13, black line) in the presence of IroB, the peak representing ENT (26.25 min) decreases significantly and a major peak with a retention time of 23.75 min along with a smaller peak with the retention time of 22 min appear, where the peak at 23.75 min represents MGE and the peak at 22 min DGE. The reaction was the allowed to proceed for another 40 min (Figure 13, red line). The analysis after this incubation time reveals that substrate ENT (26.25 min) and MGE (23.75 min) were no longer detectable and the height of the peak representing DGE (22 min) significantly increased along with that of a new conversion product, namely TGE, represented by a new peak with a retention time of 20.75 min. An additional incubation of 40 min (T = 80 min; blue line) led to a decrease of the peak with retention time of 22 min (DGE) but no significant increase of the peak at 20.75 min (TGE).

The analysis shows that the recombinant IroB used in these studies was properly folded and enzymatically active. It also established that IroB-catalyzed conversion of ENT to MGE/DGE/TGE can be followed using RP-HPLC. The compound ENT and its

63

conversion products can be efficiently separated by this technique based on the order hydrophobicity from ranging from ENT>MGE>DGE>TGE with ENT being the most hydrophobic molecule due to the lack of hydrophilic glycosyl groups. The conversion appears to be sequential with an initial build-up of MGE, which is then converted to DGE, and accumulation of DGE is followed by conversion to TGE. The conversion of ENT to DGE appears to occur relatively quickly (within 40 min) whereas the conversion from DGE to TGE is not completed after an additional 40 min, suggesting that triglycosylation is less rapid than mono- and diglycosylation.

The RP-HPLC was directly coupled to ESI/MS to confirm the identity of the substrate ENT and the conversion products MGE/DGE/TGE. The results of the LC-MS analysis are shown in Table 6.

Table 6: Summary of the LC-MS results to confirm the identity of the substrate ENT and the conversion products MGE/DGE/TGE.

Compound	Expected Mass [M+H$^+$ in g×mol^{-1}]	Observed Mass [M+H$^+$ in g×mol^{-1}]	Retention time[1] [min]
ENT	670.6	670.2	26.25
MGE	832.7	832.3	23.75
DGE	994.8	994.4	22
TGE	1156.9	1156.4	20.75

[1] indicates the retention time under the applied conditions in RP-HPLC

The result shows that there is good correspondence between expected and observed mass for substrate ENT and the enzymatic products of IroB (MGE/DGE/TGE) in the LC-MS analysis. These results clearly demonstrate that RP-HPLC is well suited to follow conversion of ENT to MGE/DGE/TGE and the compounds can be separated based on the hydrophobic characteristics.

5.1.2 Activity in the presence of Fe^{3+} and absence of Mg^{2+}

The level of IroB activity in the presence of Fe^{3+} and in the absence of Mg^{2+} can provide valuable information about the conformation of ENT during catalysis and mechanistic requirement of IroB, respectively. In the presence of ferric iron the substrate ENT is fixed in a conformation, where the 3 catechol hydroxyls of ENT are involved in coordinating one ferric iron ion. In the iron-bound form, the DHB hydroxyl groups are not accessible for deprotonation by a potential base catalyst. A magnesium ion is often needed by glycosyltransferases of the GT-A fold to stabilize the negatively charged leaving group and glycosyltransferases of the GT-B fold often require a divalent metal ion for full activity. We therefore tested the ability of IroB to catalyse the glycosylation of ENT under three different conditions as shown in Figure 14. The control reaction contained Mg^{2+} (black line), the -Mg^{2+} reaction (blue line) contained no Mg^{2+} but did contain EDTA to chelate any metal ion present in the reaction mix. And finally the reaction in the presence of Fe^{3+} (red line) contained Mg^{2+} and Fe^{3+}. The analysis shows that IroB in the presence of Mg^{2+} (Figure 14, black line) is able to convert ENT (retention time 36.25 min) fully to DGE (retention time 31 min) and TGE (retention time 29.25 min). There are no significant amounts of other compounds detectable. However, in the absence of Mg^{2+} (Figure 14, blue line) there is still significant amount of ENT (retention time 36.25 min) present in the supernatant after an incubation time that would have allowed the full conversion of ENT to DGE/TGE in the presence of Mg^{2+}. The enzyme, though, was partially active in the absence of Mg^{2+} since a minor peak for MGE (retention time 33.5 min) was detectable, suggesting that Mg^{2+} is required for full activity of IroB. In the presence of Fe^{3+} there is absolutely no conversion detectable (Fig. 14, red line) with only

the substrate ENT (retention time 36.25 min) being visible. This results shows that Fe-loaded ENT is not a substrate for IroB.
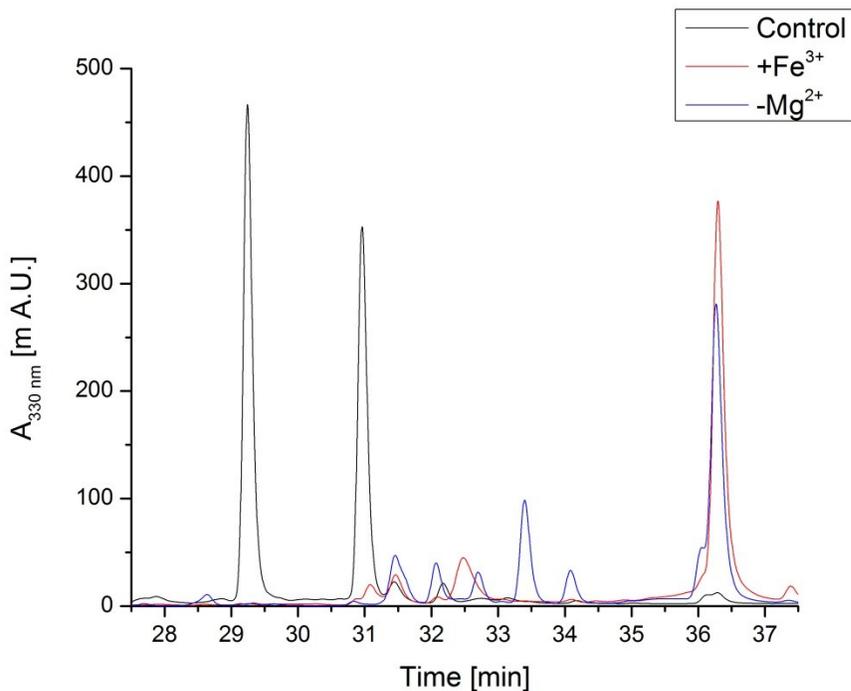
Figure 14: RP-HPLC chromatogram showing the conversion of ENT to MGE/DGE/DGE by IroB-CH6 in the presence of Fe^{3+} and absence of Mg^{2+}. The separation was conducted on a C18 resin and the absorption of ENT and conversion products was followed at 254 nm. The control reaction (Control, black line) contained Mg^{2+}, but no Fe^{3+}. The analysis of the reaction carried out in the presence of Fe^{3+} (+Fe^{3+}) is shown in red and the analysis of the reaction performed in the absence of Mg^{2+} (-Mg^{2+}) is depicted in blue.

5.2 Identification of *iroB* gene product in *E. coli* CFT073

When examining the complete genome sequence of *E. coli* CFT073 (NCBI reference: NC_004431.1), one notices that the *iroB* gene (NCBI reference: NP_753168.1) is annotated with a length of 1164 bp, and a predicted protein product length of 387 aa. In contrast, in a recent *in vitro* characterization (*43*), *iroB* was cloned using an alternative start codon resulting in a protein product length of 371 aa (1113 bp) lacking 16 aa residues on the N-terminus. To assess the question which version naturally occurs in

iroA-harbouring *E. coli* we decided to identify the gene product in the *iroA*-harbouring, non-pathogenic *E. coli* strain Nissle 1917 on the transcriptional and translational levels.

5.2.1 Identification of the *iroB* gene product on the transcriptional level

In order to detect the *iroB* gene product on the transcriptional level, we cultivated *E. coli* Nissle 1917 under low iron conditions to induce the expression of genes involved in iron acquisition. These genes are located on the *ent/fep*-operon and *iroA*-cluster under control of Fur (Ferric Uptake Regulator). The Nissle 1917 cells were harvested and RNA was isolated, transcribed to cDNA using RT-PCR and finally amplified using PCR. In the last PCR step we had two reaction mixes, each using primer pairs that were either specific for gene products with lengths of 1164 bp (387 aa) or 1113 bp (371 aa). The primers for amplification contained also a 5' overhang to introduce restriction sites that allow cloning of PCR products into a vector for sequencing. This 5' overhang introduces 12 additional nucleotides into the PCR amplification products that were not considered when indicating the length of the PCR product length. The specificity of our primers was demonstrated using genomic DNA of *E. coli* (CFT073) as a template in a PCR reaction, where we applied the same PCR conditions as in the final PCR step to amplify *iroB* cDNA. The resulting PCR products were separated on an agarose gel (1% w/v) and DNA bands were detected in UV light after staining with SYBR Safe DNA gel stain (Figure 15). In lanes 4 and 5 the products of the PCR amplification using genomic DNA from *E. coli* CFT073 as a template were identified between the 1 kb and 2 kb marker bands, consistent with the molecular weights expected after PCR amplification of products with lengths of 1164 bp and 1113 bp, respectively. The PCR product in lane 5 with an expected lengths of 1113

bp migrates closer to the 1 kb band found in lane 3 as the PCR product in lane 4 with an expected length of 1164 bp, consistent with the expected difference in length of this products of Δ=51 bp. In lane 1 there was no PCR product detectable after amplification of cDNA with the expected length of 1161 bp, implying that 'long' *iroB* mRNA, with a product length of 387 aa, does not occur in *E. coli* Nissle 1917.
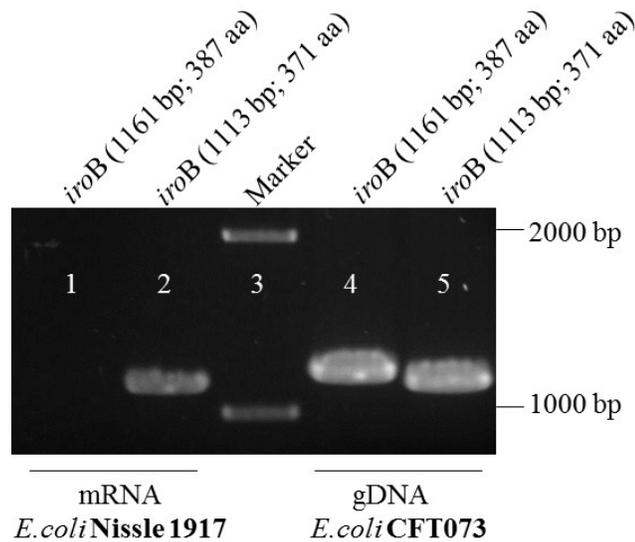
Figure 15: Agarose gel (1% w/v) stained with SYBR Safe DNA gel stain, followed by visualization of DNA bands in UV light at 332 nm. In lane 1 and 2 the PCR-amplification products using cDNA from *E.coli* Nissle 1917 as template were loaded, where in lane 1 a primer pair specific for the mRNA with a protein product length of 387 aa and in lane 2 a primer pair specific for the protein product length of 371 aa was used in the PCR reaction. Lane 3 shows the DNA molecular weight marker and lane 4 and 5 the positive control demonstrating the specificity of the primer pairs under the PCR conditions used in the reaction shown in lane 1 and 2, but using *E.coli* CFT073 gDNA as a template.

On the contrary, an amplification product for cDNA with the expected length of 1113 bp was clearly detectable after amplification migrating on the same height as the corresponding positive control in lane 5. This implies that mRNA encoding IroB with a protein product length of 371 aa is present in *E. coli* Nissle 1917 under low-iron conditions.

5.2.2 Identification of IroB gene product on translational level

For the purpose of detecting IroB on the translational level in *E. coli* Nissle 1917, we overexpressed IroB-NH6/TEV-373 and IroB-NH6/TEV-389 in BL21 (DE3) and purified these proteins with an in-frame N-terminal hexahistidine tag. Both proteins contained a TEV protease cleavage site between the N-terminal hexahistidine tag and the start of the protein that could be used to remove the tag, resulting in proteins with lengths of 373 aa and 389 aa, respectively. Upon cleavage, additional two residues (GS) remained before the start methionine due to the nature of the cleavage site. These recombinant proteins were then used as standards to identify endogenous IroB using Western Blotting techniques with affinity-purified polyclonal anti-IroB antibody, allowing for the discrimination of proteins with lengths of either 371 aa or 387 aa in *E. coli* Nissle 1917 lysate. The lysate from *E. coli* AG1 cells were used as a negative control (*iroA⁻*). The resulting Western Blot is shown in Figure 16.
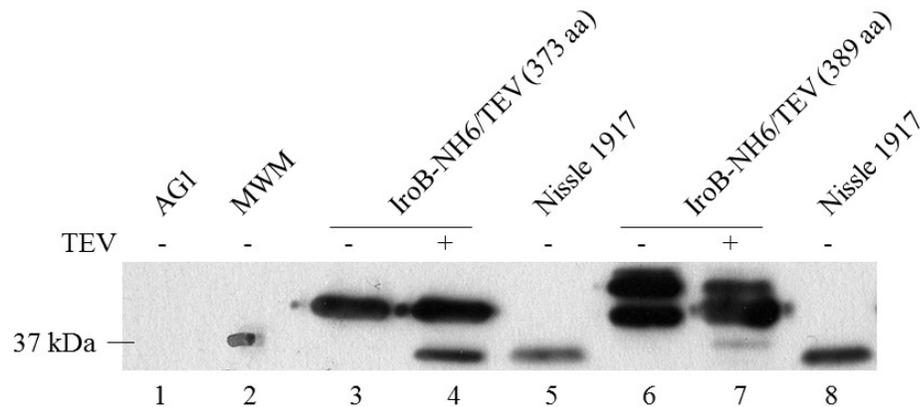
Figure 16: Western Blot analysis to identify IroB gene product on the translational level. Detection was carried out using an IroB-specific affinity-purified polyclonal antibody. In lane 1 lysate of *E. coli* AG1 cells was loaded as a negative control (*iroA⁻*). Lane 2 shows the MW marker where the 37 kDa band is highlighted. In lane 3 and 4 recombinant IroB-NH6/TEV (373 aa) was loaded, where lane 3 shows IroB-NH6/TEV before (-) and lane 4 after (+) cleavage with TEV protease. Lane 5 and 8 represents *E. coli* Nissle 1917 lysate. In lane 6 and 7 recombinant IroB-NH6/TEV with a length of 389 aa is loaded before (-) and after (+) cleavage with TEV protease.

Lane 1 contains proteins from *E. coli* AG1 (*iroA⁻*) lysate, indicating that there is no protein band visible around 37 kDa. This demonstrates that the polyclonal anti-IroB antibody is specific to IroB since it was not observed to bind to *E. coli* proteins of similar size to IroB. The non-cleaved recombinant IroB-NH6/TEV-373 in lane 3 migrates above the 37 kDa band of the MWM in lane 2. After adding TEV protease (lane 4), which removes the N-terminal hexahistidine tag and the linker region (MHHHHHHSSGRENLYFQ), the cleaved and the non-cleaved recombinant IroB-NH6/TEV-373 bands were visible. The cleaved IroB-NH6/TEV-373 (calculated mass = 40379.3 Da) migrates below the 37 kDa band of the MWM (lane 2). The migration distance of cleaved and non-cleaved IroB in lane 4 is in agreement with the expected difference in MW of $\Delta=2193.3$ Da between cleaved and not cleaved IroB-NH6/TEV-373. The cleaved IroB-NH6/TEV-373 below the 37 kDa band of the MWM (lane 2) migrates to the same height as endogenous IroB that was detected in *E. coli* Nissle 1917 lysate (lane 5).

The recombinant purified IroB-NH6/TEV-389 was loaded in lane 6 and was detected as a doublet, where one band migrated higher and one lower than non-cleaved IroB-NH6/TEV-373 (lane 3). This observation suggests that the N-terminus is unstable and undergoes degradation. The upper band is presumably non-degraded full length IroB-NH6/TEV-389 with a molecular weight of 44586.0 Da. After cleavage with TEV protease the doublet is still visible (lane 7), but the intensity of the upper band is significantly lowered. Furthermore, a band appears approximately at the same height as the 37 kDa band of the MWM (lane 2), representing cleaved IroB with a length of 389 aa and a calculated molecular weight of 42392.6 Da after removal of the hexahistidine and

linker region (MHHHHHHSSGRENLYFQ). The intensity of this band is weak and cannot account for the loss of intensity in the upper band of the doublet suggesting that the protein represented by the top-most band in lane 7 was further degraded during incubation with TEV protease. This is also supported by the fact that the intensity of the lower doublet band increased. The TEV cleaved IroB-NH6/TEV-389 migrates higher compared to the band representing endogenous IroB in the *E.coli* Nissle 1917 lysate in lane 8, suggesting that endogenous IroB of *E.coli* Nissle 1917 has a lower molecular weight and implying that the N-terminal sequence MRRLPDLDRQAERDFL of IroB-NH6/TEV-389 is not present in endogenous IroB of *E.coli* Nissle 1917. Instead, endogenous IroB has a size consistent with IroB-NH6/TEV-373.

For the remainder of the project, the IroB protein with a length of 371 aa was used for mutagenesis studies and biophysical characterizations.

5.3 Identification of active site residues in IroB

The next part of the thesis is divided into four subchapters (*i*) generating a homology model (*ii*) enzymatic activity of IroB-CH6 and variants (*iii*) biophysical characterization of IroB-CH6 and variants and finally (*iv*) determination of binding affinity constants for binding of ENT to IroB-CH6 and variants.

5.3.1 Bioinformatic analysis

The following results involve computational methods such as homology modelling, superposition approaches and sequence alignment to gain structural and functional information of IroB in the absence of any crystal structure.

5.3.1.1 Homology modelling

In order to localize the IroB active site in the absence of any crystal structure, we used homology modelling to predict the three-dimensional structure of IroB. Homology modelling involves three steps: the identification of homologs that have been crystallized, model building, and evaluation of model structure. We initially identified crystallized homologs using the FFAS03 server (*104-106*) and HHpred server (*107*). The search algorithms of these alignment servers consider, besides the sequence identity, also structural features using profile hidden Markov models and are well established to detect homologues even below 20% sequence identity. Both servers identified the C-glycosyltransferase UrdGT2 (PDB code: 2P6P) to be the best available homolog with a sequence identity between 22-23% suggesting that the structure of both proteins is similar. Thus, we used 2P6P as an input for Modeller v9.9 and used the integrated algorithm to align the primary amino acid of IroB and 2P6P resulting in the alignment shown in Figure 17, where identical residues are shaded in blue.

Figure 17: Alignment of primary amino acid sequence of IroB and C-glycosyltransferase UrdGT2 from *Streptomyces fradiae* (PDB code: 2P6P) as determined by Modeller v9.9 (*85*) and adapted in Jalview (*98*). Identical residues are shaded in blue and gaps are indicated by a hyphen.

The two sequences aligned well over the whole length with an overall sequence identity of 24.74% as determined by Jalview and was then used for model building in Modeller v9.9. After model building using a standard Modeller v9.9 scheme that includes optimization by molecular dynamics, we obtained a model with a minimal DOPE (Discrete Optimized Protein Energy) score of -40909 and a good GA341=1 score (cut off ≤ 0.7). The GA341 score indicates that the predicted fold is with $\geq 95\%$ probability correct, where a fold is considered correct when 30% of the C_α carbons of target and template structure have an RMSD of ≤ 3.5 Å (*108*). The DOPE score is an atomic distance-dependent statistical potential model and is used to distinguish between 'good' and 'bad' models; here the following is applicable: the more negative the value, the better the model (*109*). We therefore selected this as our IroB homology model, and evaluated its predicted structure independently from Modeller v9.9 by superimposing the backbone of the IroB model and the template structure 2P6P using the software LSQMAN. In addition, we used web-based evaluation server VADAR and ProSA-web to assess the overall quality of our model. The superposition of the IroB model and the template 2P6P (Figure 18) shows an overall excellent structural alignment as determined by LSQMAN with an RMSD of 0.547 Å and a normalized RMSD (100) of 0.331 Å. These excellent RMSD values demonstrate that IroB can be modelled based on spatial restraints derived from 2P6P.

The VADAR structural evaluation software suite generates an output file indicating the overall quality of the stereochemistry/packing quality and 3D profile quality index. Our IroB model was evaluated using the VADAR package, where it scored excellent overall. The Ramachandran analysis of phi/psi angles indicated showed that 93% of the residues

are found in the core region, 6% residues were detected in the allowed region, 1% of the residues are located in the generous and no residue was identified in the disallowed region. In comparison, 94% of the residues of the template structure 2P6P can be found in the core region and 3% of the residues in the allowed region. No residues were detected in the generous and disallowed region in 2P6P.
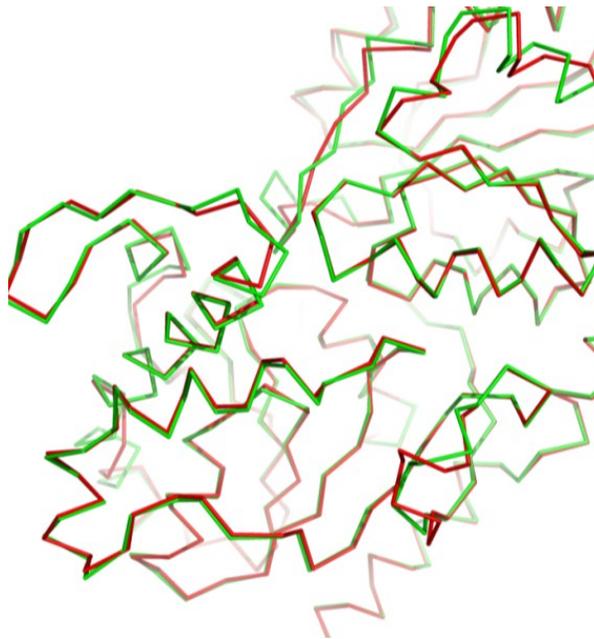
Figure 18: Overlay of IroB model and the template structure 2P6P by LSQMAN. The model IroB is shown in red and the template 2P6P is shown in green.

The ProSA-web server rated the IroB model with an overall excellent z-score of -8.57, which is well in the range of z-score determined from experimentally verified structures such as X-ray and NMR structures. The z-score of our template structure 2P6P, by comparison, was determined to be -10.98.

The predicted structure of our IroB homology model exhibits a two-domain fold known as GT-B (Figure 19) with the N-terminal and C-terminal domain. Both domains possess a

similar architecture with parallel central beta-sheets packed in between alpha-helical structures.
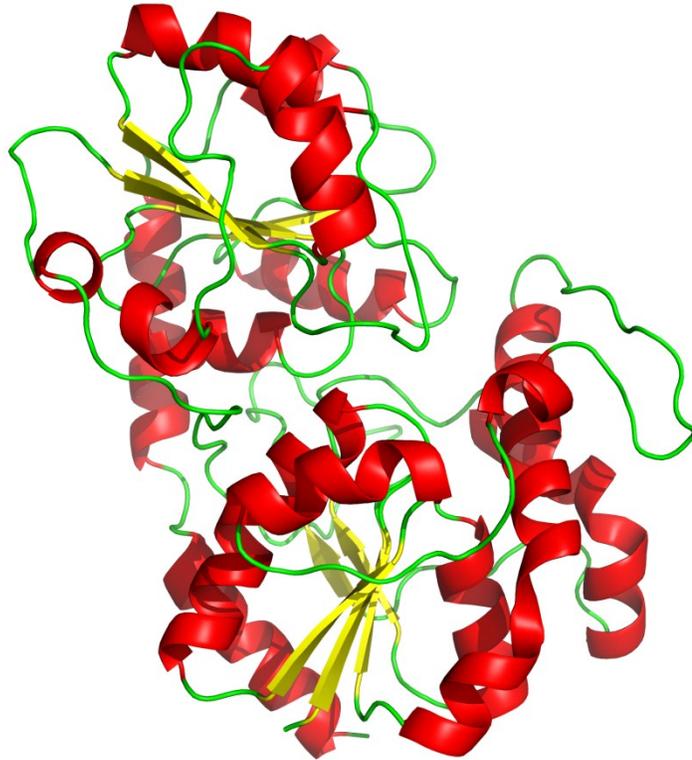
Figure 19: Cartoon 3D structure of IroB as predicted by Modeller v9.9 based on template 2P6P. Alpha-helical structures are shown in red, beta sheets in yellow and unordered structures in green. Figure was generated using PyMol (Schrödinger, LLC)

A topology map of the IroB homology model was created with TopDraw (*94*) on the basis of its predicted 3D structure (Figure 20). The topology map of IroB demonstrates the two-domain fold with the N-terminal and C-terminal domains. Both domains possess 6-7 central parallel beta-sheets (b1-7 and b8-13, respectively) that are surrounded by alpha helical structures (h1-13). The N- and C-terminal domains are connected by a long linker region. Ultimately, the C-terminal domain folds back to the N-terminal domain *via*

an envelope helix (h14 and h15). This type of fold is typical for nucleotide binding protein and known as a Rossmann-fold (*110*).
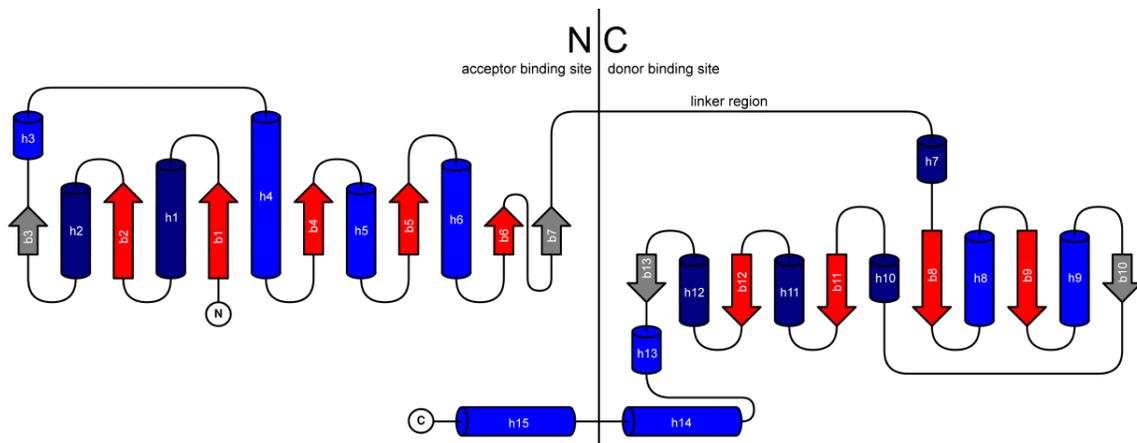
Figure 20: Topology map based predicted 3D structure of IroB generated using TopDraw. N and C indicates the N-and C-terminus, beta-sheets are shown as arrow. Red arrows indicate that this beta sheet was predicted by Modeller v9.9 and beta sheets in grey are assumed. Alpha-helical structures are shown as cylinders, where light blue indicates helical structure above and dark blue below the plane of the central beta sheet core, respectively.

5.3.1.2   Identification of active site in IroB model

After confirming that we generated a high-quality homology model, we then used it to identify the glycosyl donor binding site, which was expected to be highly conserved among glycosyltransferases (as opposed to the acceptor binding site due to the variable nature of acceptor molecules across glycosyltransferases in nature). The server HHpred (*107*) was used to identify structures homologous to our IroB model. A high-scoring hit was that of CalG1 (3OTH), an O-glycosyltransferase from *Micromonospora echinospora* with an overall sequence identity of 28%. The crystal structure of CalG1 exhibited a GT-B two-domain fold with Rossmann-like domains, similar to our predicted IroB structure. CalG1 was crystallized in the presence of TDP, an anologue of UDP, thus similar to the UDP-glucose co-substrate of IroB. Using LSQMAN, we therefore superimposed the

predicted IroB structure with TDP-bound CalG1 (3OTH) to predict the UDP-glucose

binding site in IroB. The overall RMSD of superposition of our IroB model onto the

CalG1 structure was 1.547 Å and the normalized RMSD(100) was 0.987 Å. Figure 21

shows the TDP binding site of 3OTH (A) and the predicted donor binding site of IroB (B)

based on this superposition.
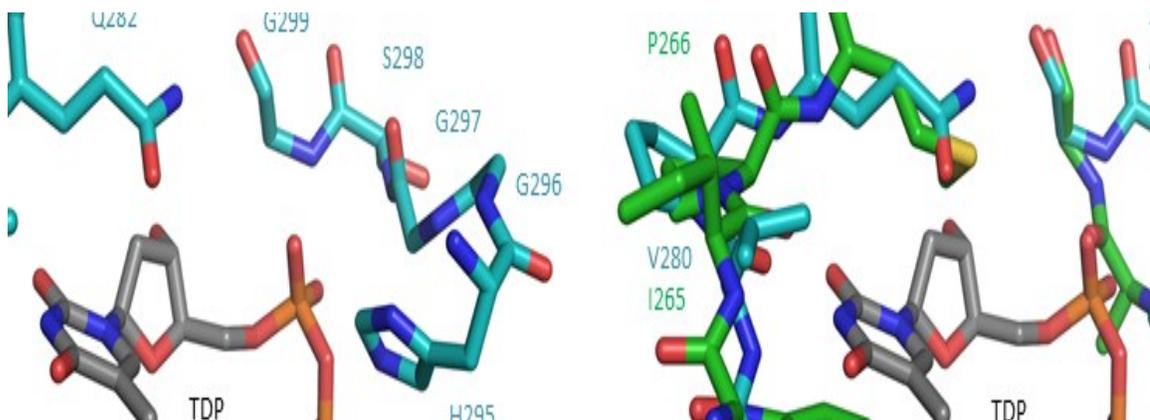
Figure 21: Residue involved in TDP coordination in 3OTH (A) and superimposed glycosyl donor binding site of 3OTH and IroB model (B). TDP is shown with grey carbons in A and B. The side chains carbons and corresponding labels of 3OTH are coloured in blue (A). In picture B the glycosyl donor binding site of 3OTH and IroB were superimposed and the residues positioned suitable for UDP coordination in IroB are coloured with green carbons as well as the labels.

The 3OTH binding site for TDP (Figure 21A) indicates the residues that are involved in

coordination of the thymidine base, the pentose sugar ribose group and pyrophosphate

group of TDP. Residue W279 is involved in a stacking interaction with the nucleotide

base. Going in the C-terminal direction, W279 is followed by the hydrophobic residue

V280 and then P281. The side chain of V280 possibly interacts with the hydrophobic side

of the nucleotide base opposite to W279, and the amide oxygen of V280 acts as a

hydrogen bond acceptor for thymidine. The residue P281 presumably plays a structural

role. This is then followed by a Q282, a potential hydrogen bond acceptor that interacts

with the ribose moiety of TDP. The negatively charged pyrophosphate group appears to be interacting with a H295 and a very Gly-rich binding motif.

The 3OTH binding site superimposes very well with the predicted glycosyl donor binding site of IroB (Figure 21B) and similar residues at equivalent locations can be found. Residue W264 in IroB is located at the same position as 3OTH W279, and can form a stacking interaction with the uridine base portion of UDP-glucose. At position 265 the hydrophobic residue Ile can be found, where in 3OTH a Val is located at the same position. I265 is well positioned to interact with the hydrophobic side of the nucleotide base, with the amide oxygen of this particular residue acting as a hydrogen bond acceptor. Proline at position 266 is well conserved (equivalent to 3OTH P281) and this residue could play a structural role by causing a turn that positions the following M267 to act a hydrogen bond acceptor for the pentose ribose sugar. The pyrophosphate group is stabilized by H280 (equivalent to 3OTH H295) and a Gly-rich motif, similar to the pyrophosphate-binding motif found in 3OTH.

Our computational studies suggest that the glycosyl donor binding site is well conserved among these homologs and was modelled in IroB with high accuracy. The high level of conservation enabled us ultimately to model TDP in the binding site of IroB (Figure 22). The prediction of the TDP binding site in IroB reveals that the UDP-glucose binding site is located in the C-terminal domain of IroB and binds such that it is deeply buried in the interior of the protein and covered by the loop connecting b8 and h8 (Figure 20). The binding motif 264-WIPM-267 (Figure 21B) interacts with the nucleotide base and pentose ribose sugar moieties of TDP and is located on the loop between b10 and h10, which then folds to the central b11 (Figure 20). Followed by this beta sheet the 280-

HGGAG-284 motif (Figure 21B) is positioned such that it could interact with pyrophosphate group of TDP.
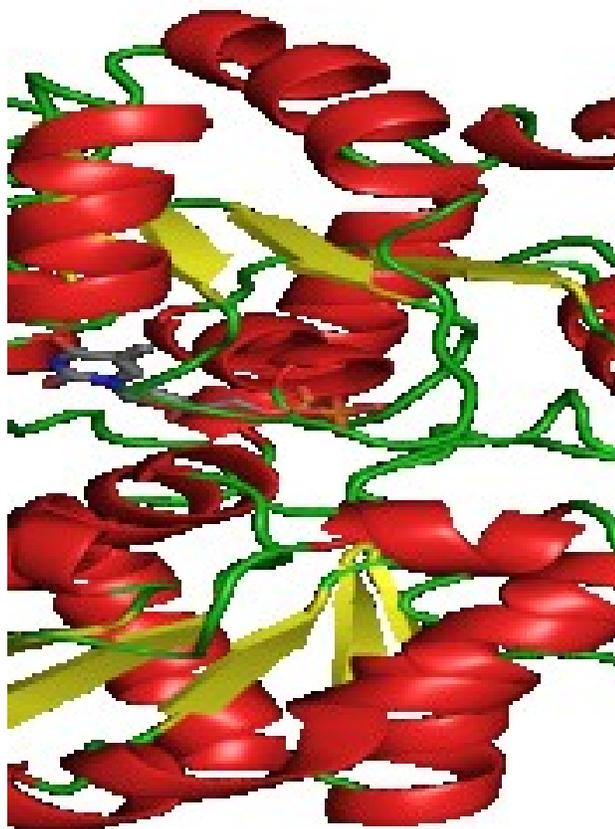
Figure 22: Predicted 3D structure of IroB and TDP modelled in the glycosyl donor binding site. Helical structures are shown in red, beta sheets as an yellow arrow and unordered structure in green. Carbons of TDP are coloured in grey.

Based on the position of TDP positioned into the IroB model, it can be presumed that the acceptor molecule ENT binds on the N-terminal domain close to the UDP-glucose pyrophosphate group in the cleft between the N- and C-terminal domains. Therefore, we identified residues in the vicinity of the pyrophosphate group of TDP that could be involved in binding of ENT (Figure 23). A very hydrophobic cluster consisting of Leu and Phe that could accommodate the hydrophobic catechol of ENT is marked with light

blue oval shape. Adjacent to this cluster the several potential base catalyst residues are located.
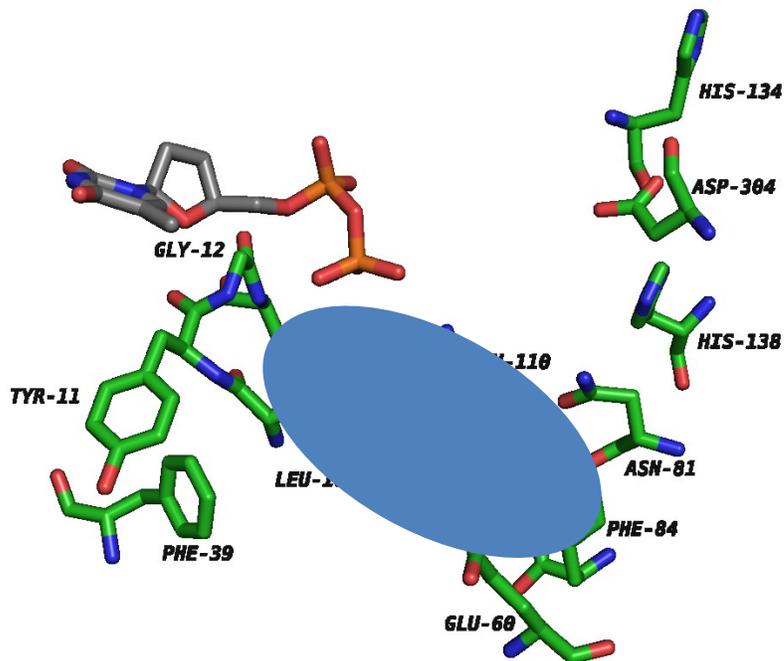
Figure 23: predicted ENT binding site in IroB. Residues around the pyrophosphate group of TDP (grey carbons) are highlighted. The residues located in the predicted IroB binding site are illustrated with green carbons. A hydrophobic cluster is highlighted in light blue.

These residues include: D304, H134 and H138 and E60. Any of these residues could potentially be involved in proton abstraction during catalysis, whereas D304 and H134 could also coordinate the glucose moiety of UDP-glucose.

Furthermore our homology model predicts that the region in IroB between residues ~60-70 is a long loop located proximal to the predicted binding site shown in Figure 23. This loop can possibly function as a "lid" that closes over ENT upon binding. We hypothesize that the loop could also contain catalytic residues. Potential base catalysts can be identified in this loop region, namely H65, H66 and E67. The identified residues located in the rigid ENT binding site and the mobile part adjacent to the binding site of

IroB represent promising targets for site directed mutagenesis studies. They were therefore mutated to gain insight in the role of these residues during IroB catalysis.

5.3.1.3   Multiple Sequence Alignment (MSA) of IroB and homologs

Multiple sequence alignment (MSA) was conducted with top hits of HHpred and IroB to gain more information about conserved motifs and structural features of IroB (Figure 24). The MSA was done using PSI-Coffee (*97*), a sequence alignment algorithm that considers also structural information in addition to the standard T-Coffee alignment procedure. Thus, the MSA tool was well suited for alignment of the rather distantly related sequences we were working with. The primary amino acid sequence of IroB (Figure 24, top) was aligned with the primary amino acids sequence of 11 different homologs with an averaged alignment score of 82 as determined by PSI-Coffee. As noted earlier IroB clearly features a fold with two similarly structured domains, where the N-terminal domain (residues 1-190) is the acceptor (ENT) binding domain and C-terminal (residues 190-371) is the donor (UDP-glucose) binding domain. The first 50-55 residues of the N-terminal domain align very well and represent the core region of this Rossmann-fold like domain. This conserved region is followed by a rather diverse region from residue 55-90 with respect to numbering of IroB and is boxed in black. This region corresponds to the h3-loop-h4 motif (see Figure 20) and was predicted to be the ENT binding site based on the predicted location of the pyrophosphate group of TDP in IroB (see Figure 23). The sequence diversity in this range possibly accounts for the variety of acceptor molecules that can be glycosylated by glycosyltransferases.

```
IroB/1-371        1  MRILFVGPPLYGLLYPVLSLAQAFRVNGHEVLIASGGQFAQKAAE--AGLVVFDAAPG-  56
2P6P/1-384        1  MRILFVAAGSPATVFALAPLATAARNAGHQVVMAANQDMGPVVTG--VGLPAVATTDL-  56
3IAA/21-416      21  AHLLIVNVASHGLILPTLTVVTELVRRGHRVSYVTAGGFAEPVRA--AGATVVPYQSE-  76
2IYA/13-424      13  RHISFFNIPGHGHVNPSLGIVQELVARGHRVSYAITDEFAAQVKA--AGATPVVYDSI-  68
3IA7/5-402        5  RHILFANVQGHGHVYPSLGLVSELARRGHRITYVTTPLFADEVKA--AGAEVVLYKSE-  60
3H4T/1-404        1  MGVLITGCGSRGDTEPLVALAARLRELGADARMCLPPDYVERCAE--VGVPMVPVGRA-  56
2IYF/8-430        8  AHIAMFSIAAHGHVNPSLEVIRELVARGHRVTYAIPPVFADKVAA--TGPRPVLYHST-  63
3OTG/21-412      21  MRVLFASLGTHGHTYPLLPLATAARAAGHEVTFATGEGFAGTLRK--LGFEPVATGMP-  76
1RRV/1-416        1  MRVLLSVCGTRGDVEIGVALADRLKALGVQTRMCAPPAAEERLAE--VGVPVHVPVGLP-  56
3OTV/21-398      21  MRVLFVSSPGIGHLFPLIQLAWGFRTAGHDVLIAVAE-HADRAAA--AGLEVVDVAPD-  75
3TSA/2-391        2  MRVLVVPLPYPTHLMAMVPLCWALQASGHEVLIAAPPELQATAHG--AGLTTAGIRGN-  57
2C1Z/8-456        8  PHVAVLAFPFSTHAAPLLAVVRRLAAAAPHAVFSFFSTSQSNASIFHDSMHTMQCNIKS  66

IroB/1-371       57  LDSEAGY-RHH-----------EAQRKKSNIGTQMGNFSFF-SEEMADHL---VEFAG   98
2P6P/1-384       57  PIRHFIT-TDREG--R----PEAIPSDPVAQARFTGRWFARM-AASSLPRM---LDFSR  104
3IAA/21-416      77  IIDADAA-EVFG-----------SDDLGVRPHLMYLRE-NVSVLRAT---AEALD      115
2IYA/13-424      69  LPKESNP-EESW-----------PEDQE-SAMGLFLDE-AVRVLPQL---EDAYA      106
3IA7/5-402       61  FDTFHVP-EVVK-----------QEDAETQLHLVYVRE-NVAILRAA---EEALG       99
3H4T/1-404       57  VRAGARE-PGEL-----------PPGAAEVVTEVVAE-WF---DKV---PAAIE        91
2IYF/8-430       64  LPGPDAD-PEAW-----------GSTLL-DNVEPFLND-AIQALPQL---ADAYA      101
3OTG/21-412      77  VFDGFLA-ALRIR--F-DTDSPEGLTPEQLSELPQIVFGRVIPQRVFDEL--QPVIE    127
1RRV/1-416       57  QHMMLQE-GMPPP-----------PPEEEQRLAAMTVEM-QF---DAV---PGAAE      93
3OTV/21-398      76  YSAVKVF-EQVAK--DNPRFAETVATRPAIDLEEWGVQIAAV-NRPLVDGT---MALVD 127
3TSA/2-391       58  DRTGDTG-GTTQLRFPNPAFGQRDTEAGRQLWEQTASNVAQS-SLDQLPEY---LRLAE 111
2C1Z/8-456       67  YDISDGVPEGYV---------------FAGRPQEDIELFTRA-APESFRQCMVMAVAET 109

IroB/1-371       99  HWRPDLIIYP--PLGV-IGPLIAAKYDIPVVMQTVGFGHTPWHIRG------------- 141
2P6P/1-384      105  AWRPDLIVGG--TMSY-VAPLLALHLGVPHARQTWDAVDADGIHPG------------- 147
3IAA/21-416     116  GDVPDLVLYD--DFPFIAGQLLAARWRRPAVRLSAAFASNEHYSFSQ------------ 160
2IYA/13-424     107  DDRPDLIVYD--IASW-PAPVLGRKWDIPFVQLSPTFVAYEGFEEDV------------ 150
3IA7/5-402      100  DNPPDLVVYD--VFPFIAGRLLAARWDRPAVRLTGGFAANEHYSLFK------------ 144
3H4T/1-404       92  GCD-AVVTTGLLPAAV-AVRSMAEKLGIPYRYTVLSPDHLPSEQSQA------------ 136
2IYF/8-430      102  DDIPDLVLHD--ITSY-PARVLARRWGVPAVSLSPNLVAWKGYEEEV------------ 145
3OTG/21-412     128  RLRPDLVVQE--ISNY-GAGLAALKAGIPTICHGVGRDTPDDLTRS------------- 170
1RRV/1-416       94  GCA-AVVAVGDLAAAT-GVRSVAEKLGLPFFYSVPSPVYLASPHLPPAYDEPTTPGVTD 150
3OTV/21-398     128  DYRPDLVVYE--QGAT-VGLLAADRAGVPAVQRNQSAWRTRGMHRS------------- 170
3TSA/2-391      112  AWRPSLVLVD--VCAL-IGRVLGGLLDLPVVLHRWGVDPTAGPFSDR------------ 155
2C1Z/8-456      110  GRPVSCLVAD--AFIW-FAADMAAEMGVAWLPFWTAGPNSLSTHV------------- 151

IroB/1-371      142  --------------------------VTRSLTDAYRRHNVGA-----             157
2P6P/1-384      148  --------------------------ADAELRPELSELGLER-----             163
3IAA/21-416     161  ---DMVTLAGTID------------PLDLPVFRDTLRDLLAEHGLSRS----        193
2IYA/13-424     151  ---PAVQDPTADRGEEAAAPAGTGDAEEGAEAEDGLVRFFTRLSAFLEEHGVDTP----  202
3IA7/5-402      145  ---ELWKSNGQRH------------PADVEAVHSVLVDLLGKYGVDTP---         177
3H4T/1-404      137  ---ERDMY-----------NQ-GADRLFGDAVNSHRASIGLPPVEHLY          169
2IYF/8-430      146  ---AEPMWREPRQ------------TERGRAYYARFEAWLKENGITEH-----       178
3OTG/21-412     171  -----------------------IEEEVRGLAQRLGLDL-----              186
1RRV/1-416      151  IRVLWEER-----------AA-RFADRYGPTLNRRRAEIGLPPVEDVF           186
3OTV/21-398     171  -----------------------IASFLTDLMDKHQVSL-----              186
3TSA/2-391      156  ------------------------AHELLDPVCRHHGLTG-----             171
2C1Z/8-456      152  ------------------------YIDEIREKIGVSGIQG-----             167

IroB/1-371      158  ----TPR------------DMAW--------I----DVTPPSMSILEN-           177
2P6P/1-384      164  ----LPA------------PDLF--------I----DICPPSLRPAN--           182
3IAA/21-416     194  ----VVD------------CWNHVEQLN--------L----VFVPKAFQIAG-       217
2IYA/13-424     203  ----ATE------------F-LIAPNRC--------I----VALPRTFQIKG-       225
3IA7/5-402      178  -----VKE------------YWDEIEGLT--------I----VFLPKSFQPFA-      201
3H4T/1-404      170  DYGYTDQ------------PWLA--------A----DPVLSPLRPTD-            192
2IYF/8-430      179  -----PDT------------F-ASHPPRS--------L----VLIPKALQPHA-      201
3OTG/21-412     187  -----PPGR------------ID-GFGNPF--------I----DIFPPSLQEPEFR    212
1RRV/1-416      187  GYGHGER------------PLLA--------A----DPVLAPLQPD---           208
3OTV/21-398     187  ----PEP------------VAT--------I----ESFPPSLLLEA-           204
3TSA/2-391      172  ----LPT------------PELI--------L----DPCPPSLQASD-           190
2C1Z/8-456      168  ----REDELLNFIPGMSKVRFRD-LQEGIVFGNLNSLFSRMLHRMGQVLPKATAVFI-- 219

IroB/1-371      178  DGEPI-------IPMQ--------YV-PYNGGAVWEPWWE----------          201
2P6P/1-384      183  -AAPA-------RMMR--------HV--ATSRQCPLEPWMY----------         205
3IAA/21-416     218  -DTFD-------DRFV--------FVGPCFDDRRFLGEWTR---------          242
2IYA/13-424     226  -DTVG-------DNYT--------FVGPTYGDRSHQGTWEG---------          250
3IA7/5-402      202  -ETFD-------ERFA--------FVGPTLTGRDGQPGWQP---------          226
3H4T/1-404      193  -L-------------------GTVQTGAWILPDQRPLSAELE                214
2IYF/8-430      202  -DRVDE-------DVYT--------FVGACQGDRAEEGGWQR---------         227
3OTG/21-412     213  ARPRR-------HELR--------PV--PFAEQGDLPAWLS---------         236
1RRV/1-416      209  -V-----------------DAVQTGAWLLSDERPLPPELE                  230
3OTV/21-398     205  -EPEG-------WFMR--------WV-PYGGGAVLGDRLP---------          227
3TSA/2-391      191  -APQG-------APVQ--------YV-PYNGSGAFPAWGA---------          213
2C1Z/8-456      220  -NSFEELDDSLTNDLKSKLKTYLNIGPFNLITPPPVVPNTTGCLQWLK---------- 266
```
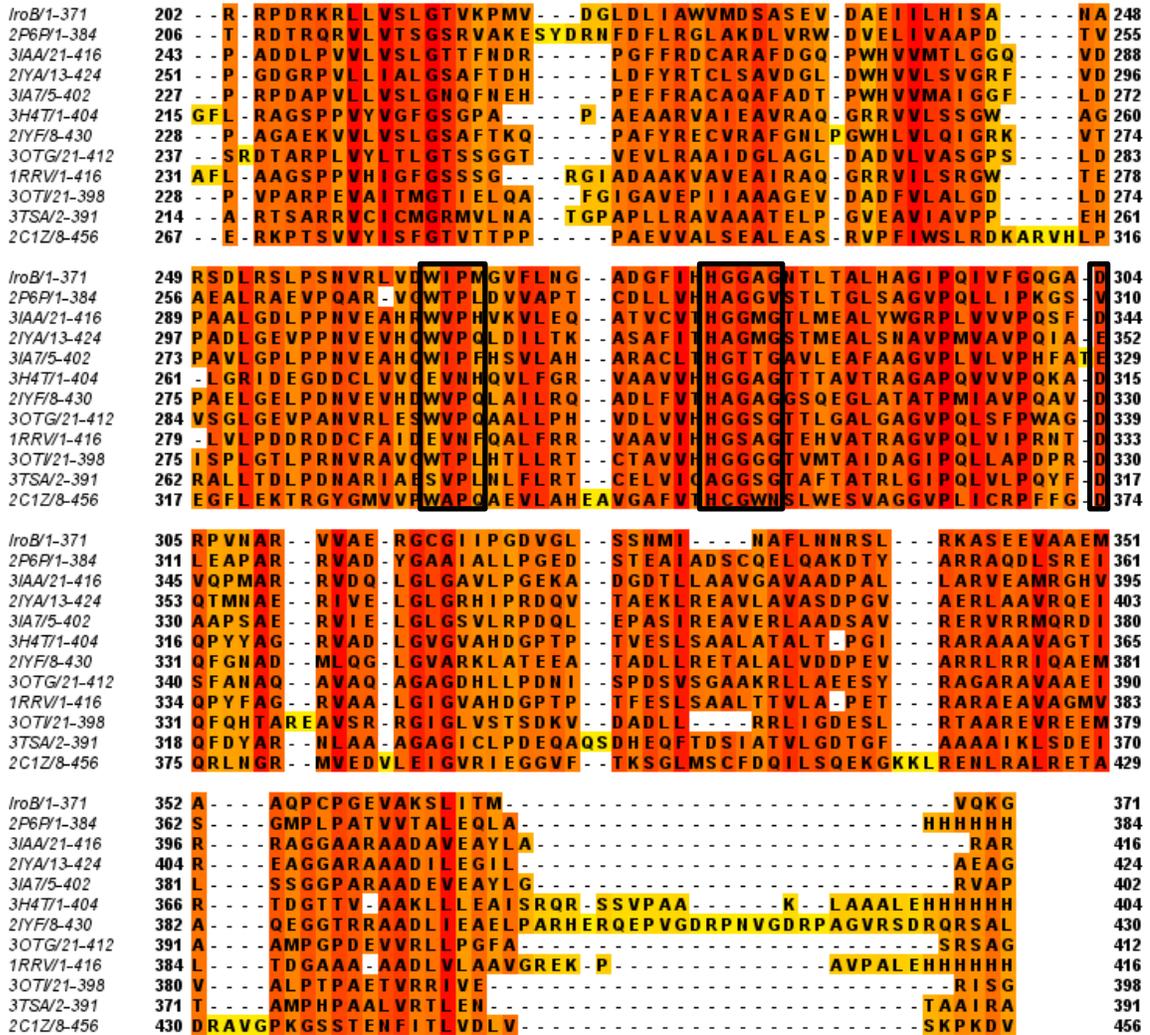
Figure 24: Multiple sequence alignment by PSI-COFFEE and further processed with Jalview of homologs detected by HHpred. Residues are shaded according to the degree of conservation from yellow to red. The conserved binding motif for the glycosyl donor is boxed in black. 2P6P: UrdGT2 from *Streptomyces 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*, 3OTG: 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.*

The protein then folds back into the hydrophobic core of the N-terminal domain represented by a conserved region of very hydrophobic residues (IroB residues 102-130) and then crosses over to the C-terminal domain *via* the linker region from IroB residues 180-200. There are several homologs such as 2C1Z or 2IYA causing large gaps in the

alignment before and within the linker region, suggesting that this region is highly variable in terms of loops connecting parallel beta sheets and the linker region itself. Ultimately, the C-terminal domain of IroB aligns very well overall between residues 200-371, illustrating that the diversity in the donor binding domain is low, consistent with the similarity of glycosyl donor molecules used by glycosyltransferases. Remarkable are the regions from 264-267 and 280-284 and representing the highly conserved binding motif for glycosyl donor molecule that was identified in the IroB model and 3OTH (compare Figure 21, A and B). The sequence of 3OTH is also included in the multiple sequence alignment under the name 3OTG, which was crystallized in the absence of TDP. The alignment overall strongly supports the identity of the predicted acceptor and donor binding site in IroB due to high diversity and high conservation, respectively, expected for these binding sites.

5.3.2   Comparison of enzymatic activities: IroB WT and variants

Based on our homology modelling and bioinformatics analyses, we predicted that E60, E67, H134, H138 and D304 could act potentially as base catalysts during catalysis. We therefore individually mutated each of these residues to assess the effects of substitution on IroB enzymatic activity. Additionally, W264 was mutated to experimentally confirm that this residue is essential for UDP-glucose binding. Briefly, we incubated wild-type IroB-CH6 and variants under assay conditions and terminated the reaction after 1 h at 25 °C by applying the reaction mix directly to the RP column. In contrast to the RP-HPLC analysis described in Section 5.1.1 we used a different column, which is the reason that

the retention times for ENT, MGE, DGE and TGE are shifted (17.0 min, 14.8 min, 13.25

min and 12 min for ENT, MGE, DGE and TGE, respectively).

We initially tested the activity of wild-type IroB-CH6 under assay conditions. The

resulting chromatogram is shown in Figure 25. A major peak with a retention time of

13.22 and a minor peak with a retention time of 11.97 min are visible representing DGE

and TGE, respectively. There is almost no MGE (14.8 min retention time) and no ENT

(retention time 17 min) detectable, suggesting that ENT and MGE were fully converted

under these assay conditions. The integrated area of peaks between 10-20 min were

determined using OriginLab v8.6 and the peak areas were measured, resulting in an area

of 200.56 for DGE and 28.64 for TGE. This means that 100% ENT was converted to

~12% TGE and 88% DGE.
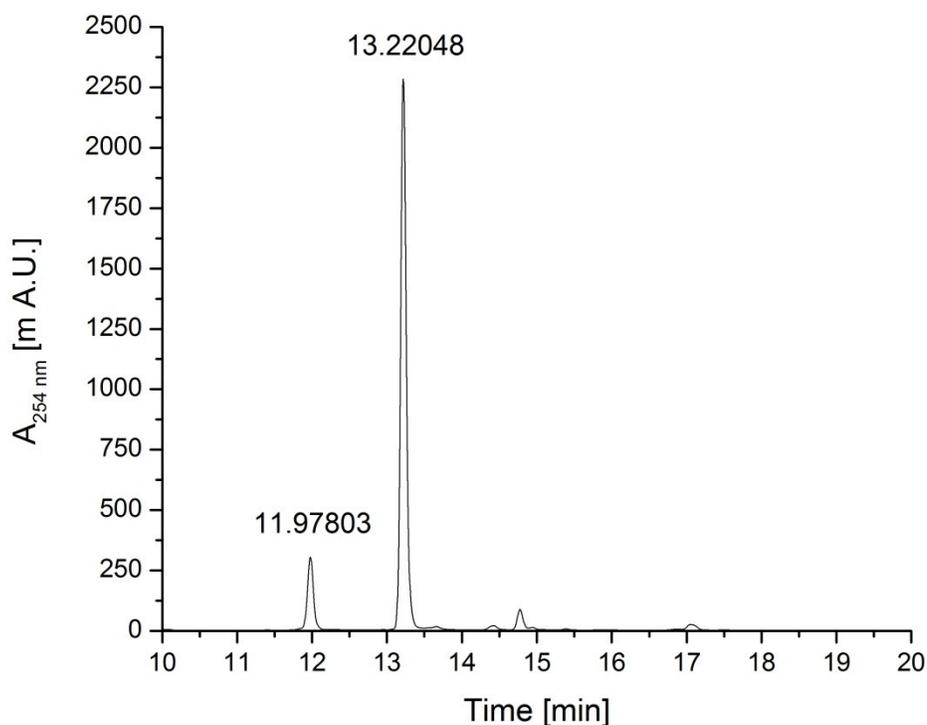
Figure 25: RP-HPLC chromatogram showing the conversion of ENT to MGE/DGE/DGE by IroB-CH6 The separation was conducted on a C18 resin and the absorption of ENT and conversionsyl products was followed at 254 nm.

Next we mutated W264L, which was predicted on the basis of our model to stack with the nucleotide base of the glycosyl donor UDP-glucose (see Figure 21). The leucine substitution disrupts the ability to stack with the nucleotide base, but still maintains the hydrophobic character of the residue. We applied exactly the same assay conditions to W264L as for IroB-CH6, and analyzed the reaction mix for the presence of ENT and conversion products after allowing the enzymatic reaction to occur for 1 h at 25 °C. The resulting RP-HPLC chromatogram is shown in Figure 26. The analysis shows the presence of three different compounds namely ENT, MGE and DGE represented by the peaks with the retention times of 17.0 min, 14.8 min and 13.2 min, respectively. The DGE peak is only minor with a relative abundance and an area of 28.98.
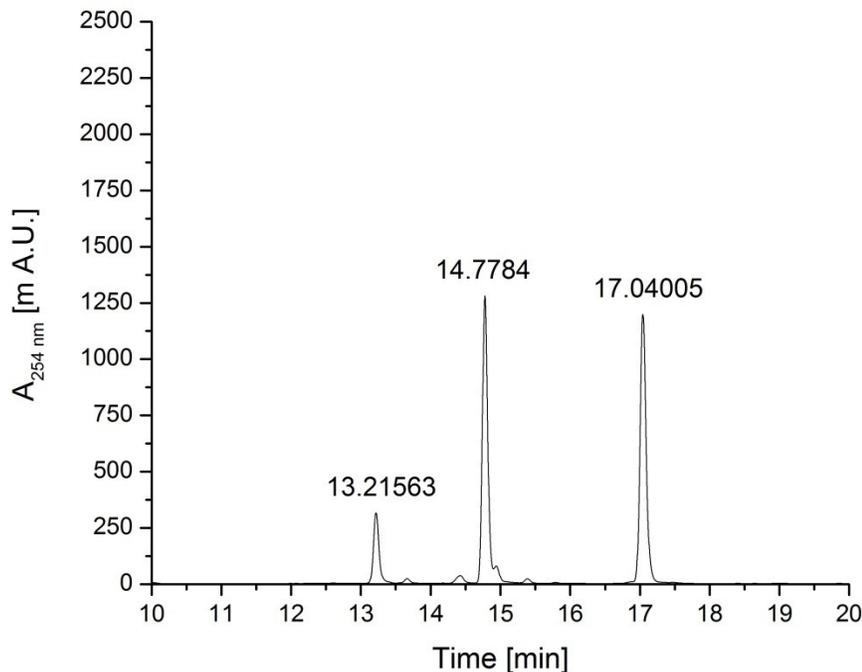
Figure 26: RP-HPLC chromatogram showing the conversion of ENT to MGE/DGE/DGE by IroB-CH6-W264L. The separation was conducted on a C18 resin and the absorption of ENT and conversion products was followed at 254 nm.

The peaks representing MGE and ENT have almost the same height and correspondingly the area as determined by OriginLab v8.6 is 124.90 for MGE and 123.24 for ENT. Taken together this result implies that 10% DGE, 45% MGE and 45% ENT are present and thus only 55% of the ENT was converted by W264L as opposed to 100% by IroB-CH6 suggesting that the enzymatic activity is significantly reduced in W264L, consistent with our predicted role of this residue in nucleotide base stacking interaction.
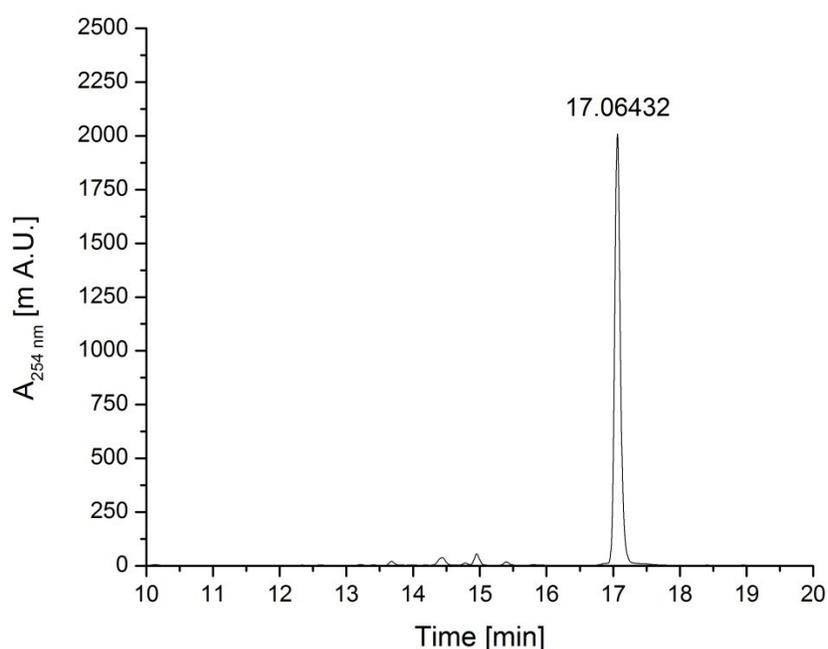
Figure 27: RP-HPLC chromatogram showing the conversion of ENT to MGE/DGE/DGE by IroB-CH6-D304N. The separation was conducted on a C18 resin and the absorption of ENT and conversion products was followed at 254 nm.

The residue D304 in IroB is highly conserved among glycosyltransferases (see Figure 24) and is thus an interesting candidate for mutagenesis. It is located on the C-terminal domain (glycosyl donor binding domain) and is well positioned to either coordinate the glycosyl moiety of UDP-glucose or act as base catalyst to activate a nucleophile during catalysis. We mutated D304 to asparagine to disrupt its ability to act as a base catalyst but

still maintain the ability to act as an isosteric hydrogen bond acceptor. Incubating D304N

under assay conditions resulted in the RP-HPLC chromatogram illustrated in Figure 27.

The analysis shows that after incubation period a single peak with the retention time of

17.06 min can be detected, which corresponds to ENT. No conversion products can be

identified, suggesting that D304 plays an important during catalysis. The area under the

peak was determined to be 205.75; since no conversion products have been identified this

value equates to 100% ENT.
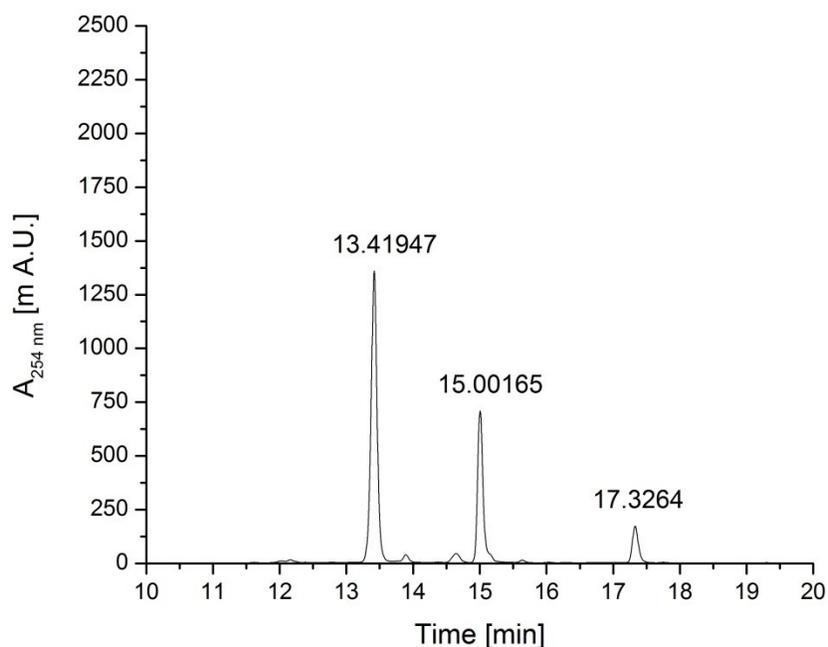
Figure 28: RP-HPLC chromatogram showing the conversion of ENT to MGE/DGE/DGE by IroB-CH6-H134A. The separation was conducted on a C18 resin and the absorption of ENT and conversion products was followed at 254 nm.

The residue H134 is located rather distant from the predicted ENT binding site. However,

considering that some residue might be modelled incorrectly H134 is also a reasonable

candidate to be a base catalyst. We therefore decided to mutate H134 to alanine, thus

eliminating its ability to act as a base catalyst. The analysis of this mutant by our activity

assay resulted in the chromatogram shown in Figure 28. The analysis revealed three major peaks with a retention times of 13.4 min (DGE), 15.0 min (DGE) and 17.3 min (ENT). The retention time for all compounds is slightly shifted around 0.2 min to higher retention times, but the overall pattern is retained. The relative abundances based on peak areas were calculated to be: 7.6% (ENT), 29.4% (MGE) and 63% (DGE). This result suggests that the catalytic efficiency of H134A is slightly affected due to the introduced mutation.
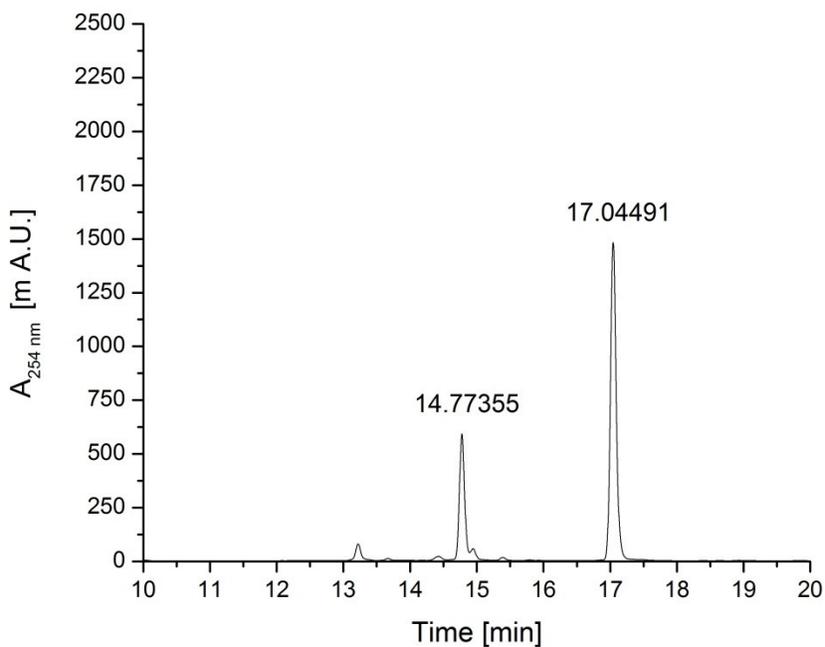
Figure 29: RP-HPLC chromatogram showing the conversion of ENT to MGE/DGE/DGE by IroB-CH6-E67A. The separation was conducted on a C18 resin and the absorption of ENT and conversion products was followed at 254 nm.

As discussed earlier, the loop located adjacent to the predicted ENT binding site (see Figure 20: h3-loop-h4) may possibly close over ENT upon binding and play a catalytic role. Therefore, we also identified IroB residue E67 as a possible base catalyst on this solvent-exposed loop. We mutated E67 to alanine and subjected this variant to our

89

activity assay. We obtained the RP-HPLC chromatogram shown in Figure 29. The chromatogram reveals two peaks typifying ENT (17.0 min) and MGE (14.8 min), where majority constitutes ENT. The calculated area for ENT is 150.93 and for MGE is 62.09, meaning that the relative abundance is 29.1% and 70.9% for MGE and ENT, respectively, demonstrating that the E67A mutation significantly reduces the catalytic activity of IroB. In addition we also prepared the following IroB variants: E60L, H65A/H66A, and H138L. We subjected these variants to our activity assay and calculated the relative abundance of ENT, MGE, DGE and TGE in the supernatant based on the RP-HPLC analysis (chromatogram not shown). The summary of all results with respect to the relative abundance of ENT/MGE/DGE/TGE of all our mutants is tabulated in Table 7.

Table 7: Summary of relative abundances [%] of ENT/MGE/DGE/TGE after assaying IroB-CH6 and variants. The results are derived from RP-HPLC chromatograms by peak integration.

| | Relative Abundance [%] | | | |
	ENT	MGE	DGE	TGE
IroB-CH6 (WT)	0	0	88	12
W264L	45	45	10	0
D304N	100	0	0	0
H134A	8	29	63	0
H138L	0	0	100	0
E60L	0	22	78	0
H65A/H66A	8	60	32	0
E67A	71	29	0	0

The summary in Table 7 shows that IroB-CH6 (WT) is the most active enzyme, catalyzing the full conversion from ENT and MGE to mainly DGE and some TGE. The mutation W264L, predicted to be in the UDP-glucose binding site, renders IroB less

active, consistent with the predicted role of W264 to stack with the nucleotide base of UDP-glucose. Our studies demonstrate that IroB residue D304 plays an important role during catalysis. Substitution to asparagine cannot be tolerated at this position. This residue is predicted to either help coordinate the glucose moiety of the UDP-glucose donor, or to function as a base catalyst. The residues H134, H138 and E60 are located in the rigid part of the ENT binding site in IroB, as predicted on the basis of the homology model. The influence of mutations on these residues on catalysis is moderate, and conversion to large amount of DGE can still occur even though the rate appears to be lower, as seen for H134A. The residues located on the predicted mobile loop adjacent to the ENT binding site have a high impact on catalysis as seen for H65/H66 and E67. The double variant H65A/H66A is only able to convert 32% of ENT to DGE and for E67A there is no DGE detectable. However, H65A/H66A is still able to catalyse the formation of considerable amount of MGE (60%) and only a small amount of the substrate ENT is left. The variant E67A was also significantly impaired in its ability to catalyze the formation of MGE leaving most of the substrate ENT (71%) in its unglycosylated form.

5.3.3 Biophysical characterization of IroB-CH6 and variants

Biophysical characterizations were conducted to confirm that the overall structural integrity of variant IroB proteins was not disturbed due to the introduction of mutations in the primary amino acid sequence. To assess overall fold quality, we subjected IroB and variants several biophysical methods that are sensitive to secondary structure (far-UV circular dichroism), overall stability (thermal denaturation followed by far-UV circular

dichroism) and tertiary structure (fluorescence spectroscopy to probe the environment of intrinsic fluorophores (Tyr, Trp)).

5.3.3.1 Far-UV Circular Dichroism

Far-UV circular dichroism was used to gain information about the secondary structure content of IroB-CH6 and to compare it to mutants. The overall shape, intensity and characteristics of the CD spectra provide information regarding the relative abundance of distinct secondary structure elements such as beta-sheets, alpha helices and random coil. CD is thus well suited for studies comparing the structural qualities of IroB variants relative to the wild-type protein.

We recorded CD spectra in the far-UV region from 260-200 nm at 25 °C in a 0.1 cm cuvette at 25 °C. IroB-CH6 and variants were directly taken from freeze stock measured in storage buffer containing 1 M sucrose and spectra were corrected for buffer contribution. We compared the CD-spectra of IroB-CH6 (WT), W264L, E67A and D304N and H65A/66A. Figure 30 shows the far-UV CD spectra of IroB-CH6 (WT) and selected variants. The spectrum of IroB-CH6 (WT) (black line) is typified by a positive band at around 200 nm. The minimum at 208-210 nm is skewed towards the 220 nm leading to a broadened minimum at 220 nm. At the 245 nm there is no absorption of circularly polarized light observed. The spectra of D304N (red line) displays similar characteristics of broad minimum at 220 nm and a positive ellipticity below ~203 nm. The spectra for E67A (blue line) and loop substitutions H65A/H66A (magenta line) show the same characteristics, with a positive band below ~203 nm and a broad minimum at

220 nm, suggesting that the mutations in the loop did alter the secondary structure content of IroB-CH6 (WT).
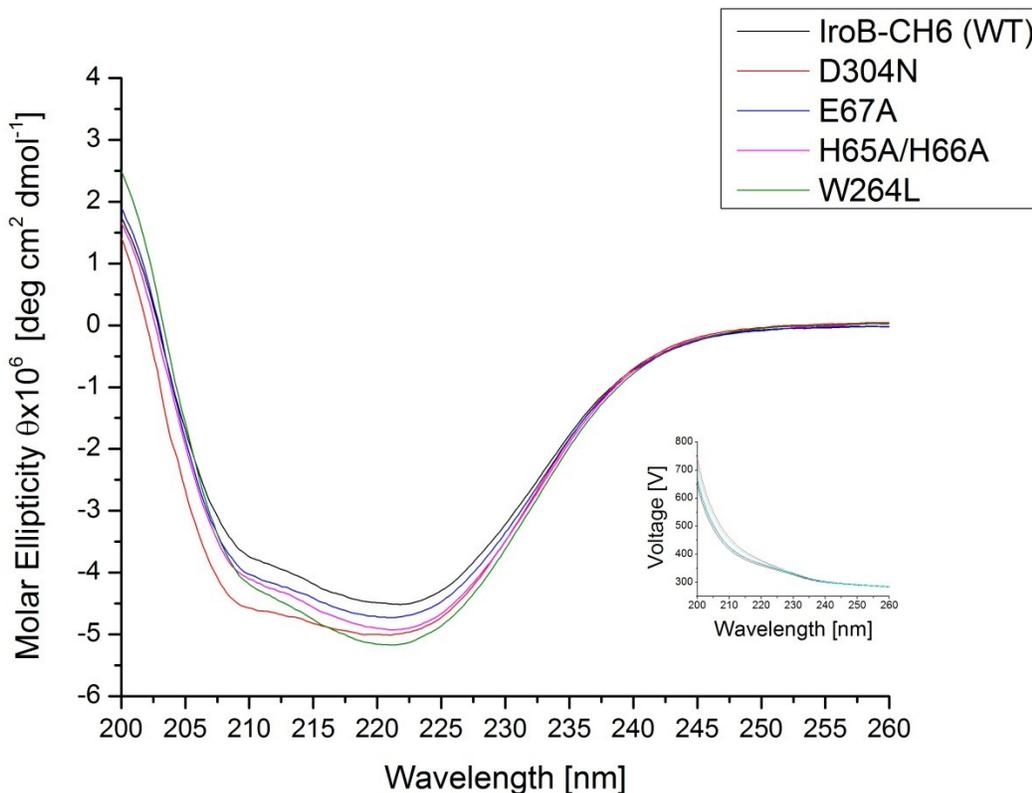
Figure 30: Far-UV circular dichroism spectra of IroB-CH6 (WT) variants. IroB-CH6 (WT) (black line), D304N (red line), E67A (blue line), H65A/H66A (pink line) and W264L (green line) were analysed. The inlet illustrates the trend of voltage during CD measurement.

The mutations D304N and W264L are located in the interior of the protein, where both induced minor changes in the properties of the CD spectrum. The mutant W264L (green line) shows similar characteristics as WT, with a slight variation in the 210-212 nm region. On the contrary, the mutant D304N possesses a less pronounced minimum at 220 nm and the minimum in the 210 nm region is clearly visible and more pronounced, suggesting slight perturbation of the structure of IroB as a result of this mutation. In summary, the far-UV CD analysis shows that IroB-CH6 (WT) and variants are well

folded. The observed spectral differences are minor and should not be over-intrepreted. All of the proteins analyzed show comparable CD-signature indicating a α+β or α/β fold.

5.3.3.2 Thermal denaturation

Thermal denaturation was performed to assess the overall thermal stability of IroB-CH6 (WT) and variants, and to test if protein stability was affected by mutations. The thermal denaturation was followed by far-UV CD spectroscopy at 222 nm with a rate of 30 °C×h^{-1}.
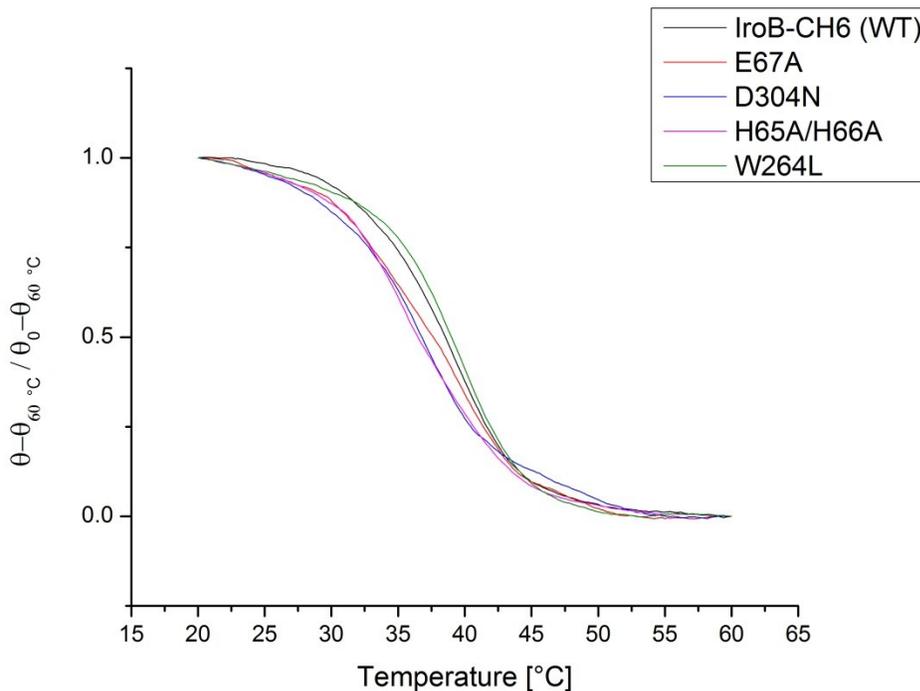
Figure 31: Thermal denaturation followed by far-UV circular dichroism spectroscopy at 222 nm shown as percent folded between 20 °C and 60 °C. Transitions of IroB-CH6 (WT) (black line), E67A (red line), D304N (blue line), H65A/H66A (pink line) and W264L (green line) are illustrated.

The raw thermal denaturation curves (data not shown) showed a transition at around 40 °C and a loss of about 40% of the signal intensity over the course of this transition,

followed by a stable signal at 60 °C - 85 °C. In a few cases a second transition occurred at around 85 °C, however, this transition was not included in the analysis, because a stable plateau was not achieved when raising the temperature to the maximum of ~95 °C. Therefore, the stable signal at 60 °C (60% of the initial CD signal) was assumed to be 0% folded (100% unfolded) and the denaturation curves were normalized accordingly and plotted using OriginLab v8.6 (Figure 31). The data shows that the mid-point of the transition from 100% folded to 0% folded I located for IroB-CH6 (WT) at around 37 °C (black line). The mid-point of transition is the nearly the same for all mutants suggesting that they have stabilities similar to the wild-type protein, and that the overall fold is not affected by these mutations. To determine the exact mid-point, 1st derivatives of the melting curves were calculated, where the maximum of the 1st derivative indicated the inflection point and therefore the melting temperature (T_m). The results are summarized in Table 8 showing that all T_m values are centred around 37 °C, thus demonstrating that the mutations did not alter thermal stability of IroB.

Table 8: Summary of T_m values of IroB-CH6 (WT) and variants. Values were determined using the maximum between 20 °C and 60 °C of the 1st derivative.

	T_m [°C]
IroB-CH6 (WT)	39.4
W264L	39.6
D304N	37.2
H65A/H66A	35.2
H138L	37.8
E67A	39.6

5.3.3.3 Fluorescence spectroscopy

The excitation and emission properties of fluorophores such as tryptophan are highly dependent on their environment and are thus sensitive to environmental changes. By the means of fluorescence spectroscopy, the integrity of the tertiary structure of IroB-CH6 (WT) and mutants were probed.

The protein concentration of IroB-CH6 (WT) and mutants was adjusted such that the absorbance at 280 nm was below 0.05 O.D. to avoid inner filter effects. The emission of the intrinsic fluorescence was measured from 300-400 nm after excitation of the sample at 280 and 295 nm. We first probed variant W264L for any significant changes in the tertiary structure by comparing its emission spectra to IroB-CH6 (WT) (Figure 32).
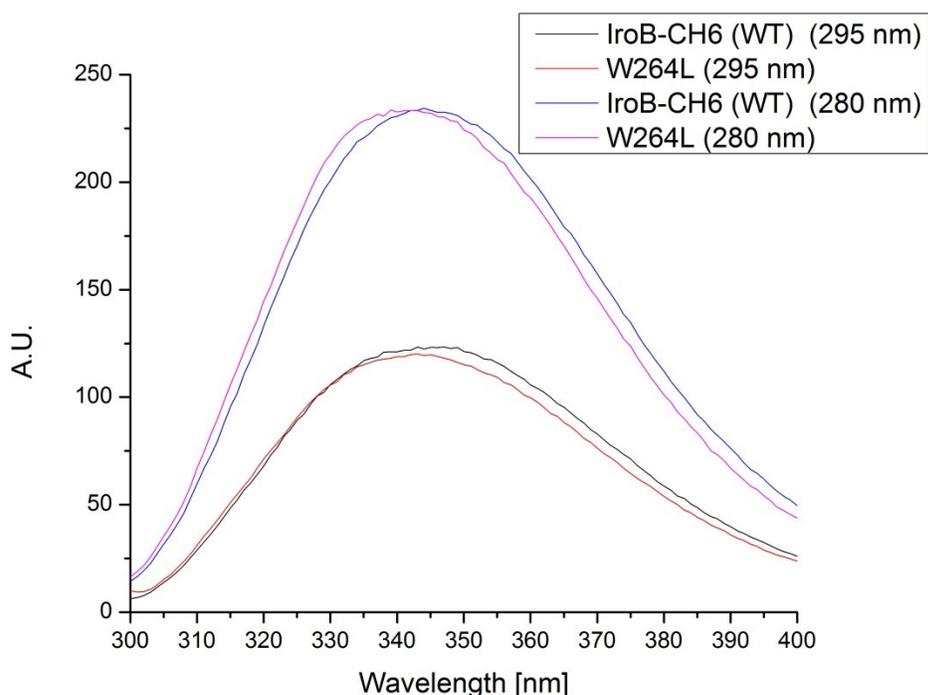
Figure 32: Fluorescence spectra of IroB-CH6 (WT) and W264L. The emission spectrum of IroB-CH6 (WT) after excitation at 280 nm and 295 nm is displayed in blue and black, respectively. The emission spectrum of W264L after excitation at 280 nm is shown in pink and after excitation at 295 nm in red.

The emission spectrum of IroB-CH6 (WT) after excitation at 280 nm (blue line) exhibits a higher intensity as after excitation at 295 nm accounting for the fact that the excitation wavelength of 295 nm specifically excites tryptophan, whereas at 280 nm both, tryptophan and tyrosine are excited (black line). The maximum emission of both spectra is at 345 nm. This maximum represents the red end of emission spectrum of tryptophan and tyrosine, thus implying that the eight tryptophan residues in IroB-CH6 (WT) are relatively solvent-exposed and located in a more polar environment. Interestingly, mutating one of the eight tryptophans to a leucine did not decrease the overall intensity significantly (red and pink line), even though the number of protein molecules used was constant as determined by UV/VIS spectroscopy using the corresponding molar extinction coefficient. However, a slight blue shift is detectable in both spectra for W264L (red and pink) with respect to IroB-CH6 (WT). Since the emission peak of tryptophan is solvatochromic, a blue shift indicates that W264, mutated to a leucine in W264L, was located in a more polar environment on the surface of IroB, contributing significantly to the red-shifted emission spectrum of IroB-CH6 (WT). In conclusion, such as shift is expected when mutating tryptophan and does not necessarily implicate a disturbance of the overall tertiary structure.

In Figure 33 the fluorescence spectrum of IroB-CH6 (WT) in comparison with those of the IroB variants D304N, E67A and H65A/H66A after excitation at 280 nm and 295 nm is shown. All spectra overlap very well with a maximum emission at ~345 nm. No solvatochromic shifts were detectable, demonstrating that the tertiary structures of these variant proteins, as reflected here by fluorophore environements, were not significantly affected by the substitutions.
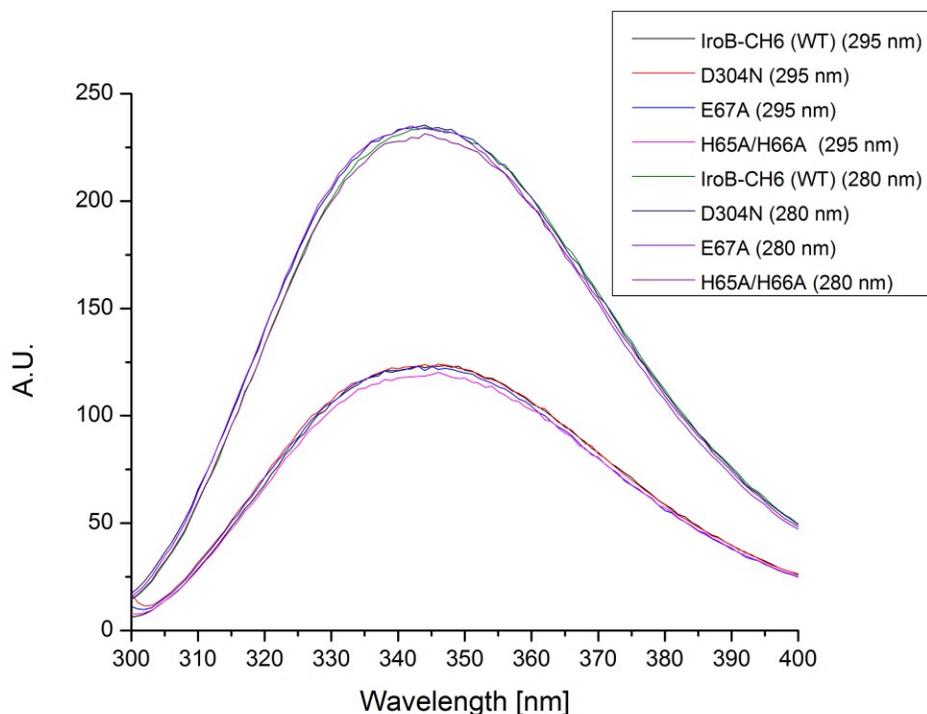
Figure 33: Fluorescence spectrum of IroB-CH6 (WT) and various mutants. The legend indicates the colour the emission spectrum is displayed in, the name of the mutants and in brackets the excitation wavelength.

5.3.3.4 Determining Binding Affinity Constants

In order to determine apparent binding affinity constants ($K_D^{'}$) of IroB to ENT, we used a quenching assay that was originally developed by Lin *et al.* for the same purpose (*82*). Briefly, the initial fluorescence intensity of IroB was measured, which is represented as F_0. The emission spectrum of intrinsic IroB fluorescence and the excitation spectrum of ENT possess substantial overlap. Thus, IroB can act as a FRET donor and transfer excitation energy to the FRET acceptor ENT. Fluorescence resonance energy transfer (FRET) is only possible if donor and acceptor are in close proximity, as it occurs during a binding event. This phenomenon leads to a concentration-dependent apparent quenching

effect of intrinsic IroB fluorescence by bound ENT that can be used to determine apparent binding affinity constants (K_D').
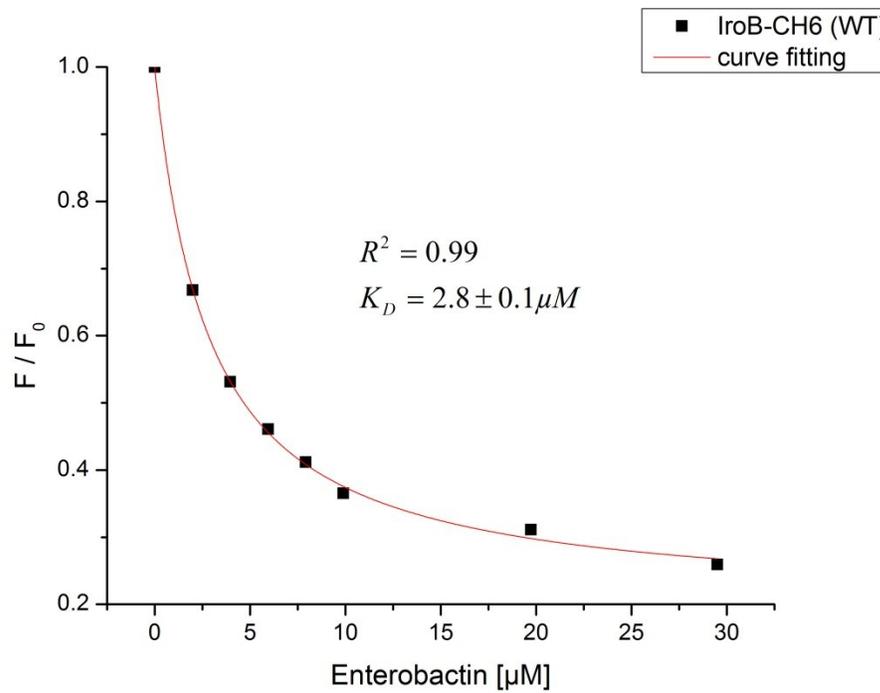
Figure 34: Plot of F/F$_0$ vs. ENT concentration and determination of binding affinity constants by fitting the data to equation (5).

The plot of F/F$_0$ vs. ENT concentration (Figure 34) resulted in a binding curve that converged against ~30% of the initial fluorescence, suggesting that several tryptophan residues are proximal to the ENT binding site in IroB. The data fitted well to equation (5) with high R^2 value of 0.99, the apparent K_D value derived was 2.8 μM with an error of the fitting of 0.1 μM. The determination of the apparent K_D was performed twice independently and resulted in similar apparent K_D values. The apparent K_D values were also determined for a number of the variants studied in this project using the same method. The determined values are summarized in Table 9. The binding assay for each mutant was performed twice and resulted always in similar values. The apparent K_D

values were all found to be on the same order of magnitude suggesting that all mutants were still able to bind ENT with high affinity. These binding data therefore provide an independent confirmation of fold preservation in the IroB variants examined.

Table 9: Apparent K_D values derived from fluorescence based quenching assay tabulated for selected mutants.

	K_D [μM]
IroB-CH6 (WT)	2.8 ± 0.1^1
W264L	4.5 ± 0.3^1
D304N	9.1 ± 0.6^1
E67A	4.8 ± 0.2^1

[1] the error of the fitting to equation (5)

6   Discussion

The Discussion is structured in a manner parallel to the Results Section: expression/purification, identification of the natural IroB gene product, and identification of IroB active site catalytic residues.

6.1   Expression/purification of IroB

The expression and purification of IroB was constraint by the high tendency to form insoluble inclusion bodies and also the insolubility of IroB at concentrations above ~0.3 mg×mL^{-1} during purification. To address the problem of the inclusion body formation, we lowered the expression temperature and decreased the inductor concentration to avoid the formation of insoluble aggregates during expression. The formation of inclusion bodies during overexpression is based on the aggregation of folding intermediate exposing hydrophobic patches (*111*). By lowering the temperature and inductor concentration to 20 °C and 400 μM, respectively, the yield of soluble IroB significantly increased presumably because IroB had more time to mature (*111*). However, an estimated 90% of IroB was still found in the insoluble fraction, supporting the notion that IroB is a slow-folding protein or/and IroB forms aggregates even in the folded state at high concentrations *in vivo*. IroB is a protein that naturally occurs in *E. coli* (CFT073). Therefore the expression strain *E. coli* BL21 (DE3) should be a good host and should provide a good environment to allow proper folding of IroB in terms of a reducing environment, chaperones, pH and ionic strength.

The second problem, the tendency to precipitate during purification, was addressed by optimization of various parameters such as ionic strength, pH, and additives known to

stabilize proteins conformationally (*e.g.,* glycerol) or shield hydrophobic regions of proteins (*e.g.,* L-arginine) (*112*). IroB was most soluble at a relatively high pH (8.5) and low salt concentration (25 mM). The combination of high pH and low salt could be beneficial due to the repelling overall negative charge of IroB at this pH, which is minimally shielded at low salt concentrations (*113*). More important, however, was the role played by osmolytes during purification. Not only the solubility was greatly improved in the presence of 30-50% (v/v) glycerol, 1 M trehalose and 1 M sucrose, but also conformational stability as determined by far-UV circular dichroism. Far-UV circular dichroism showed that IroB underwent conformational changes such as loss of alpha helical structure and a gain in beta sheet structures (data not shown) shortly after exchanging IroB in buffer without any osmolytes. Accompanied with the observation that IroB precipitates heavily at higher concentrations under these conditions, we concluded that partial unfolding followed by beta sheet aggregation could be responsible for the low solubility.

Osmolytes can be classified as stabilizing and destabilizing. Stabilizing osmolytes are preferentially excluded from the protein backbone (*e.g.,* glycerol, sucrose) and destabilizing accumulates at the backbone (*e.g.,* urea, GdmCl), resulting in an unfavourable interaction of stabilizing osmolytes with the unfolded state as (*114, 115*). Interestingly, the synthesis of osmolytes (*e.g.,* glycerol, trehalose) is induced *in vivo* under stress to protect proteins from denaturation (*116*). The fact that osmolytes had an especially beneficial effect on IroB stability and solubility further supports the notion that IroB is intrinsically conformationally unstable. The stabilizing effect of osmolytes is dependent on the polar surface area of the osmolyte – the smaller the fractional polar

surface area, the more stabilizing. The fractional polar surface area of trimethylamine N-oxide (TMAO) is significantly smaller compared to sucrose, trehalose and glycerol and could therefore be beneficial with regard to future experiments (*114*). Ultimately, we decided to use sucrose as a stabilizer, since it can be easily removed by gel filtration, and is commercially available for a low price. Once the constraints regarding solubility were addressed, IroB was able to be expressed, purified and stored in large amounts.

6.1.1   IroB enzymatic activity assay and product identity

To prove that IroB is enzymatically active and thus properly folded after expression, purification and storage, we tested the ability of IroB to catalyse the formation of MGE/DGE/TGE *in vitro* using the strategy previously published by Fischbach *et al.*, where they detected products arising from IroB activity by LC-MS (*43*). We used a relatively high concentration of IroB (5 µM) to achieve fast product formation and observe all possible products including TGE (Figure 13). After 5-min incubation in the presence of IroB we observed almost full conversion of ENT to MGE and only minor peaks representing DGE and ENT. The conversion appears sequential, since we initially observe accumulation of MGE. We allowed the reaction to proceed for additional 40 min and detected no more ENT and MGE, but approximately equal amounts of DGE and TGE. Additional 40 min incubation led to a decrease of the peak representing DGE, but no increasing amount of TGE. The successive conversion of ENT to first MGE and then MGE to DGE was also observed by Fischbach *et al*. (*43*). This could be an indication that IroB has different affinities for ENT and glycosylated derivatives of the order ENT<MGE<DGE<TGE, where the affinities are lowered through the additional bulk of

the glucose moieties attached to the catechol of ENT. This hypothesis could be easily tested using the fluorescence-based binding assay to determine K_D' with enzymatically prepared and purified MGE/DGE/TGE used as ligands instead of ENT.

Direct comparison of the enzyme activity of our recombinant IroB with published information on IroB (*43*) was not possible. Steady-state enzyme kinetics data has to be obtained for this purpose. However, such experiments proved to not be straightforward since ENT has a lower stability at physiologically relevant temperatures ($t_{1/2}$ = 30 min at 37 °C in Tris buffer at pH 8 (*19*)). This instability of the ENT substrate could be reason for the observed decrease in the amount in DGE with no increase in TGE upon HPLC separation after 80 min incubation (Figure 13). Qualitative analysis of the chromatogram in Figure 13 supports the hypothesis that the initial glycosylation reactions leading to MGE/DGE formation are catalysed with higher efficiency compared to the third glycosylation to produce TGE. This conclusion is drawn on the basis that almost full conversion to MGE is catalysed within 5 min, and full conversion to DGE is achieved some time before 45 min, as opposed to the time period between 45 min and 80 min, in which only minor decrease of the amount of DGE was detected and no more accumulation of TGE. Higher catalytic efficiencies for the first and second glycosylation steps have also been described by Fischbach *et al.* (*43*). These *in vitro* observations correlate very well with *in vivo* data demonstrating a high abundance of DGE in the supernatant of *iroA*-harbouring *Salmonella enterica* (*41*). However, a final conclusion on that matter cannot be drawn owing to the qualitative data, and also the effect of protein precipitation over time course of the assay cannot yet be excluded.

Fischbach *et al.* always included 5 mM MgCl$_2$ in their reaction mix during their *in vitro* characterizations of IroB, although never commented if a divalent metal ion is actually required (*43*). We therefore tested the activity of IroB in the presence and absence of Mg^{2+} and additionally supplemented our Mg^{2+} minus control with EDTA to chelate any metal ion present in our buffer or already bound to IroB. Furthermore, we tested the ability IroB to catalyse the glycosylation reaction in the presence of Fe^{3+}, thus fixing ENT in the defined iron-binding mode, where 2,3-OH of the three DHB subunits of ENT can hexacoordinate one ferric ion in an octahedral geometry (*21*). Analysing the resulting chromatogram revealed that the catalytic activity of IroB is significantly reduced in the absence of Mg^{2+} with only minor formation of MGE. This is in contrast to the positive control in the presence of Mg^{2+} where full conversion of ENT and MGE to DGE and TGE was observed under the same conditions. In the presence of Fe^{3+} no conversion to MGE/DGE/TGE was detectable under assay conditions.

The drastically reduced catalytic activity in the absence of Mg^{2+} was at first surprising to us. IroB was predicted to adopt a GT-B fold (www.cazy.org) and glycosyltransferases of that fold type usually do not require a divalent metal, since they stabilize the pyrophosphate leaving group during the S$_N$2 mechanism by helix dipole interaction or positively charged amino acid side chains (*63, 117*). In addition, the typical DXD metal-binding motif observed in metal-dependent GT-A is not present in IroB (*50*). However, it has been reported that most glycosyltransferases need a divalent metal ion for full activity except for GtfB (*50, 57, 118*). Interestingly, the β-Glucosyltransferase from phage T4 was crystallized in the presence of UDP and the full donor substrate UDP-glucose. Only in the UDP-bound form was a divalent metal ion found at the position where the glucose

moiety was located before. Therefore they assumed that the metal ion could play a role in product release rather than leaving group departure (*71*). Since IroB is still active in the presence of EDTA and the absence of Mg^{2+} this role of the metal ion can also be proposed.

The incapability of IroB to catalyse the glycosylation of *holo*-ENT (*e.g.*, ferric-ENT) has previously demonstrated by Lin *et al.* and could be confirmed by our experimental approach (Figure 14). The authors attributed this observation to the inaccessibility of the ENT catechol hydroxyls for binding to IroB (*82*). Noteworthy in this context, are the distinct physical and steric properties of *apo*-ENT (net charge: neutral) and *holo*-ENT (net charge: -3; fixed conformation with hydroxyl groups of catechols coordinating one iron ion) (*32*). The calyx of NGAL for *holo*-ENT, for example, is adapted to these distinct properties, and binding is mediated by NGAL positively charged side chains (lysine/arginine) based on cation-π and electrostatic interaction. Furthermore, the calyx of NGAL allows distinct binding of all three catechol groups of ENT (*30, 32*). Neither positively charged side chains nor distinct binding pockets for catechol subunits are found in IroB model (Figure 23), suggesting completely different binding modes for *holo*- and *apo*-ENT based on steric and physical properties. Regrettably, the approach to measure the apparent equilibrium binding affinity constant of *holo*-ENT and IroB was not successful, because of heavy precipitation of IroB in the presence of Fe^{3+} and also because of quenching of intrinsic IroB fluorescence by Fe^{3+} (data not shown). In conclusion, the notion that the hydroxyl group of the DHB subunit might be recognized by IroB is oversimplified, because the steric and physical differences between *holo*- and *apo*-ENT are significant and could simply prevent *holo*-ENT from being recognized by

IroB, independent from hydroxyl groups of DHB. Therefore, this interpretation is rather limited and allows only the conclusion that *holo*-ENT is not a substrate of IroB.

6.2 Identification of iroB gene product on the transcriptional and the translational level in Nissle 1917

As discussed in the previous section, when IroB was expressed according to Fischbach *et al.* (*43*) with a length of 371 aa (without linker and affinity tag), it appeared conformationally unstable during expression and purification. Interestingly, the *iroB* gene in *E. coli* CFT073 (NCBI reference: NP_753168.1) is annotated with 387 aa and contains additional N-terminal aa (MRRLPDLDRQAERDFL). Therefore, to exclude the possibility that these additional N-terminal aa are necessary for the conformational integrity of IroB, we identified the natural *iroB* gene product in *E. coli* Nissle 1917 on the transcriptional and the translational level. The probiotic, non-pathogenic strain *E. coli* Nissle 1917 (DSM6601) harbours the *iroA* gene cluster on the Genomic Island I (*83*). This region is fully sequenced and deposited under code AJ586887.1 in GenBank (http://www.ncbi.nlm.nih.gov). The *iroA* locus of Nissle 1917 on Genomic Island I (AJ586887.1) and *E. coli* CFT073 (NC_004431.1) were compared using BLAST (http://www.ncbi.nlm.nih.gov). The comparison clearly demonstrated that up- and downstream, as well as the *iroB* nucleotide sequence itself is identical (data not shown). Thus, *E. coli* Nissle 1917 is well suited to identify the *iroB* gene product in CFT073. Firstly, we analyzed the upstream sequence of the *iroB* gene to identify regulatory elements such as the translation initiation Shine-Dalgarno (SD) sequence. For optimal gene expression the SD sequence must provide good base pairing abilities with the anti-

SD sequence of 16S rRNA and must be located 8-10 base pairs upstream of start codon (*119*). Indeed, we found an optimally-spaced subsequence (GAGG) 10 bp upstream of the start codon coding for IroB with a length of 371 aa that could represent a potential degenerated SD core sequence (GGAGG) similar to what is usually found in *E. coli* (*119*). In contrast, no such sequence was found upstream of the start codon coding for IroB with a length of 387 aa. Additional information provides the start codon itself. The start codon of IroB (371 aa) is a strong ATG, whereas the start codon of IroB (387 aa) is a weak GTG. To compensate for a weak start codon one would expect a strong SD sequence (*119*). Therefore, IroB (371 aa) is more likely the gene product of *iroB* since it possesses the common ATG start codon and a SD sequence.

To ultimately exclude the possibility of the existence of IroB (387 aa), we identified the *iroB* gene product on the transcriptional level. Here we cultivated *E. coli* Nissle 1917 cells under iron starvation to induce Fur regulated genes. We then isolated whole RNA and transcribed mRNA for IroB to cDNA using a specific primer for IroB. The cDNA was amplified using specific primer pairs for IroB (371 aa) and IroB (387 aa), respectively. By this approach, we were only able to prove the existence for mRNA for IroB (371 aa), but not the existence of mRNA for IroB (387 aa). Nevertheless, we also identified IroB on translational level. We cloned expression constructs for IroB (371 aa and 387 aa) with TEV cleavable affinity tag. The second residues in either IroB (371 aa) and IroB (387 aa) is arginine, therefore the N-terminal methionine is not excised by methionine aminopeptidase and was included in all our constructs (*120*). After cleavage of the expressed proteins, they would only differ by an additional GS sequence due to the nature of TEV cleavage. Subsequently referred as IroB (373 aa and 389 aa), these

recombinant proteins were used for direct comparison with endogenous IroB from Nissle 1917 by SDS-PAGE followed by Western Blotting. Our results (Figure 16) showed that recombinant IroB (373 aa) is similar in MW as endogenous IroB (Nissle 1917), whereas the MW of IroB (389 aa) is significantly larger. Thus, we assume that IroB (371 aa) is the *iroB* gene product on translational level. Interestingly, IroB (389 aa) was prone to degradation in contrast to IroB (373 aa), indicating that the N-terminal sequence (MRRLPDLDRQAERDFL) is solvent exposed and flexible, providing further support that this sequence is not part of an IroB fold.

In summary, the bioinformatic analysis of the *iroB* gene sequence, as well as our experimental data clearly demonstrates that IroB (371 aa) is naturally expressed in *iroA*-harbouring *E. coli* strains such as Nissle 1917 and CFT073 and that the version of IroB with 387 aa in NP_753168.1 is likely an annotation error.

6.3   Identification of active site residues in IroB

The overall major objective of this project was to establish a homology model as structural framework to allow formulation and testing of mechanistic hypotheses for the glycosylation of ENT. The only available, experimentally determined, three dimensional structure of a C-glycosyltransferase, UrdGT2 (PDB code: 2P6P), proved to be well suited for this purpose. It allowed us to generate a homology model based on 24% sequence identity that passed all bioinformatics-based evaluation tests. Superposition of the IroB model with CalG2 (PDB code: 3OTH) furthermore allowed us to identify the conserved UDP-glucose (donor) binding site and provided us in addition with valuable information about location and identity of residues in the ENT (acceptor) binding site.

In the following sections the role of several residues involved in either UDP-glucose binding or residues located in the ENT binding site will be discussed. Relevant residues that were elucidated during these studies will be discussed individually taking into account all available information from the Results section, such as multiple sequence alignment, biophysical characterizations and apparent binding affinity constants. After that, a potential mechanism for the glycosylation of ENT will be suggested based on the information revealed by site directed mutagenesis and the current state our research into this protein.

6.3.1 W264 – π-stacking interaction with uridine

W264 is part of the well conserved 264-WIPM-267 binding (see Figure 24) motif located between b10 and h10 (see Figure 20) that interacts with nucleotide base and the ribose moiety. The W264 residue is predicted to interact with the nucleotide base by a stacking interaction (Figure 21, B). The W264L variant showed reduced activity and was only able to convert 45% of ENT to MGE and 10% DGE (Table 7). Biophysical characterization using far-UV circular dichroism revealed W264L is well folded and the structure is not significantly disturbed compared to wild-type IroB. Fluorescence spectroscopy showed a slight blue shift compared wild-type IroB showing that W264 is solvent exposed. The apparent equilibrium binding affinity constant was determined to be K_D'= 4.5 µM. The intention to mutate W264L was to experimentally verify the prediction that W264 interacts indeed with the nucleotide base as revealed by our homology model. We observed significantly reduced activity. The variant, however, was still able to bind ENT with affinity comparable to wild-type IroB. Therefore, we assume that the ability to

bind UDP-glucose is impaired by this mutation, which is in accordance with our prediction.

The observation that TDP is covered by a loop (Figure 22) implies that the loop connecting b8 and h8 exists in an "open" conformation to accept the glycosyl donor and a "closed" conformation once the donor is bound. The residue W264 is located on this mobile loop in the IroB model, and interacts presumably *via* π-stacking with the nucleotide base once the loop is in its "closed" conformation. This mode of binding is well conserved and has been described for various other glycosyltransferases such as GTf3 (*121*), VvGT1 (*122*) and the CalG1-4 family (*123*) and was suggested to protect the UDP-glucose against hydrolysis (*57*). It is notable that the crystal structures of CalG1-4 in the presence of TDP demonstrate how the binding motif 264-WIPM-267 might interact with TDP, the UDP-analog. The Trp residue interacts *via* π-stacking with the nucleotide base, as predicted in our IroB homology model. Based on the interaction of the binding motif 279-WVPQ-282 in CalG1, the IroB motif 265-WIPM-267 might predominantly interact with uridine and ribose moieties by hydrogen bonding of amide oxygen and nitrogen of the backbone. The 295-HGGSG-299 motif in CalG1 interacts similarly with α- and β-phosphate group *via* hydrogen bonding (*123*). Since this motif is also well conserved in IroB (see Figure 24) the same role can be predicted for 280-HGGAG-284.

In summary, we achieved our goal by showing that W264 is indeed involved in uridine coordination, thus experimentally confirming the accuracy of our homology model. In addition, the UDP analogue TDP could be modelled in the IroB active site and showed essentially the same interactions of well conserved motifs with TDP as shown in experimentally determined three dimensional structures such as CalG1-4. It is remarkable

that none of these structures served as template for homology modelling of IroB, demonstrating the high degree of conservation of the structural framework for TDP/UDP coordination.

6.3.2 D304 – coordination of the sugar donor moiety

D304 is located on the C-terminal donor-binding domain on N-terminus of helix h12 (see Figure 20). It is positioned on the interface between the N- and C-terminal domains, where it could participate in coordinating glucose moiety of UDP-glucose, or act as a base catalyst to abstract protons from ENT catechol that is bound in the N-terminal domain. Furthermore, D304 is part of D/E motif that is highly conserved among homologues (see Figure 24 (black box)). We mutated D304 to an asparagine, which completely abolished detectable glycosylation activity under our assay conditions − no MGE/DGE/TGE at all was identified. However, the variant D304N was well folded and the thermal stability was comparable to wild-type IroB as determined by far-UV circular dichroism. Also the tertiary structure was not significantly disturbed according to fluorescence emission; no change in the environment of Trp was detectable (Figure 33). Most importantly, this mutant was still able to bind ENT with an apparent K_D' on the same order of magnitude with wild-type IroB (9.1 µM as compared to 2.8 µM for wild-type) (Table 9). We initially mutated this residue under the assumption that it could potentially act as a base catalyst. In fact, it was shown for GtfB (PDB code: 1IIR) that mutating the conserved D332 significantly reduced k_{cat} 250-fold without affecting K_m for UDP-glucose. The authors concluded that D332 is part of the catalytic apparatus and could act as a general base (*118*). However, it should be noted that no crystal structure of

GtfB bound to UDP-glucose is available. The role of the D/E-Q motif was further elucidated by Offen *et al*. demonstrating that D374 in VvGT1 (PDB code: 2C1Z) is involved in coordinating the glucose moiety of the sugar donor by hydrogen bonding to 3-OH and 4-OH *via* the carboxyl group. The D374A mutation in VvGT1 also abolished the catalytic activity to non-detectable level, consistent with our IroB D304N mutant (*122*). The amino acid sequence of 2C1Z is included in the multiple sequence alignment shown in Figure 24, and it aligns very well in this region thus demonstrating that the residue D374 in 2C1Z is equivalent position as D304 in IroB. Based on this information, we conclude that D304 is involved in glucose orientation rather than acting as a base catalyst. Since the catalytic activity is absent or below detectable level in GtfB, VvGT1 and IroB when D/E motif is altered, the exact positioning of the glucose moiety is apparently essential for inverting glycosyltransferases employing an S_N2 mechanism.

In summary, this result provides us with valuable information about the exact position of the glucose moiety, and therefore the location where the nucleophile attacks the anomeric carbon of glucose. This knowledge allows us to infer the position of the C_5 of the ENT catechol during catalysis.

6.3.3   E67 – a potential base catalyst

The potential base catalyst E67 is localized in a region of IroB that is predicted to be a loop and is rather far away from the ENT binding site. Accordingly, the region of the ENT binding site is not conserved at all among homologs, suggesting that it is indeed a loop (see Figure 24). However, this region is clearly proximal to the ENT binding site and possibly contributes to the diversity of acceptor molecules that can be glycosylated.

Of all mutants we made with regard to potential base catalysts, this residue had the most severe impact on IroB catalysis, despite being far away from the active site in the model. After 1 h incubation only 30% MGE was detected, whereas most of ENT (70%) remained unprocessed. In comparison, IroB-CH6 (WT) completely converted ENT under the same conditions to DGE (88%) and TGE (12%) (see Table 7). At the same time, the E67A mutation had the least impact on protein fold and apparent binding affinity constant. E67A showed the same far-UV circular dichroism signature as wild-type IroB (Figure 30), the T_m was even slightly higher than with 39.6 °C as compared to 39.4 °C for wild-type (Table 8). The fluorescence spectroscopy analysis showed no disturbance of the tertiary structure – the spectra of wild-type and E67A were superimposable (Figure 33). Furthermore, the apparent binding affinity constant K_D' was minimally affected (K_D' (E67A) = 4.8 µM compared to K_D' (WT) = 2.8 µM) (Table 9).

Since the E67A mutation does not result in any disturbance of the secondary or tertiary structure, and the affinity for ENT is not impaired, it provides further support that E67 is indeed located in a loop as predicted by our homology model. Since E67 is rather distant from the active site, we therefore assume that this loop might be mobile and possibly closes over ENT upon binding. The template structure for our homology model UrdGT2 (PDB code 2P6P) possesses also a long, presumably mobile loop adjacent to the acceptor binding site (*74*). Regrettably, the three-dimensional structure of the C-glycosyltransferase UrdGT2 was solved in the absence of glycosyl-donor and -acceptor molecules; however, this did not prevent the authors from suggesting a mechanism for the C-glycosylation of the acceptor molecule UWM6 (University of Wisconsin Madison compound no. 6). They proposed the presence of two base catalysts in the UrdGT2 active

site. Since only one was found in the rigid part of UrdGT2, they suggested that a mobile loop (62-72 or 219-228) closes over UWM6 upon binding so that the second base catalyst, namely E66, E71 or E224, can fulfil its function (*74*). Since E67 is similarly positioned in a mobile loop between residues 60-75 on IroB, it is possible that the region between residues 62-72 in UrdGT2 also contains important catalytic residues. Further support for this perception comes from a recent publication in which the authors rationally designed a C-glycosyltransferase from an O-glycosyltransferase using segments from UrdGT2 (*64*). Remarkably, they achieved the desired switch in activity and identified segment 52-62 to be responsible, however, they did not identify specific protein-ligand interactions. The C-glycosylating activity, though, was low compared to wild-type UrdGT2 leading to the conclusion that residues upstream of 62 are also important for catalysis. Our experiments with the UrdGT2 homolog IroB strongly support this hypothesis.

6.3.4   Other residues

In the course of this project several other variants were generated and their activity was tested. Among these are E60L, R64A, double mutant H65A/H66A, double mutant K71A/K72A, N81A, H134A, H138L and triple mutant H65A/H66A/H138L. All of them were tested for activity and most of them were on a similar level with wild-type as determined by our RP-HPLC assay. Activity was slightly affected in H134A and H65A/H66A. However, due to time limitations, these mutants were not fully characterized, especially H134A for which no biophysical data is available. The K_D' was not determined for any of them.

Since most of these residues are apparently not involved in catalysis we decided to focus on H134A and H65A/H66A. The double mutation H65A/H66A is located adjacent to E67 on the presumably mobile loop that closes over ENT upon binding. Biophysical characterization of this double mutant revealed no disturbance, neither of the secondary nor the tertiary structure as determined by far-UV circular dichroism, thermal denaturation and fluorescence spectroscopy. The impact on catalysis compared to E67 is moderate, and this variant is still capable of producing MGE (60%) and DGE (32%), consuming almost all substrate ENT. Given the position in the loop adjacent to the important residue E67, H65/H66 could polarize E67 or play a structural role and therefore enhance the catalytic function of E67.

The residue H134 was, along with residue H138, among the most promising initial targets for base catalysts. Firstly, both histidine residues are positioned in the rigid part of the binding pocket close to the site of glycosylation, and secondly, in this region was the base catalyst Asp137 in UrdGT2 located (*64, 74*). None of them had an effect severe enough to be an essential catalytic residue. The variant H138L was fully active. However, an interesting pattern was observed in the RP-HPLC analysis of the supernatant of H134A (Table 7). The analysis shows the same relative abundance for ENT (8%) compared to double mutant H65A/H66A, however, the values for MGE (29%) and DGE (63%) are inverted compared to H65A/H66A. This was not observed previously; usually ENT is completely consumed when DGE is accumulated. This indicates that H134 could have a rate enhancing effect on catalysis. It should be noted, though, that this experiment was done only once and has to be reproduced before further hypotheses can be generated. Furthermore, the biophysical characterization (CD, fluorescence) of this variant is

missing, and therefore it cannot yet be excluded that the structural integrity was impaired by this mutation.

6.3.5 Proposed mechanism for C-glycosylation based on results

The authors who performed the first *in vitro* characterization of IroB proposed that a base catalyst in the active site of IroB deprotonates the 2-OH of the 2,3-DHB subunit of ENT, thus rendering C_5 *para* to the 2-OH nucleophilic (see Section 1.6, Figure 11) (*43*). This nucleophile, then, attacks the electrophilic anomeric carbon of the glucose moiety of the donor molecule UDP-glucose resulting in the observed C-C bond. We extend this mechanistic hypothesis by two major points: (*i*) the deprotonation of the 2-OH is not essential, but is a rate-determining step, and (*ii*) we predict the existence of a second base catalyst that helps removing the hydrogen at the site of glycosylation *after* the glycosylation occurred to re-establish the aromatic system.

To (*i*) we can offer several reason why deprotonation is not essential, but rate determining. The catechol subunit of ENT possesses two strongly activating hydroxyl substituents (*124*). Hydroxyl substituents on benzene rings are known to be *ortho/para* directing groups resulting from unshared pairs of electrons of the oxygen that can be donated in the ring system (*125*). Consequently, the resonance structures can be drawn as shown in Figure 35.
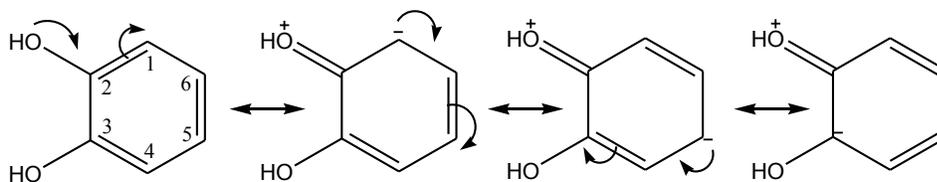
Figure 35: The catechol and its resonance contributor. On the left the most stable Lewis structure is shown followed by dipolar contributors (figure adapted from (*125*)). Position 1 is connected *via* an amide linkage to the trilactone backbone.

Since the site of glycosylation is *para* to 2-OH on C_5, the 2-OH must therefore be the donating group. Because of this highly activating 2-OH substitution, the ring system is already much more nucleophilic in the *para* position and thus more reactive than a non-substituted benzene ring (*125*). An even more striking effect, though, is given by the *ortho* carbonyl group (of the amide linkage in ENT) of 2-OH of the 2,3-DHB subunit of ENT, which significantly lowers the pK_a of the 2-OH group due to its electron with-withdrawing effect, in addition to the delocalizing effect of the aromatic ring system (*125*). The pK_a values for the non-substituted catechol are 9.25 and 13 for the first and second hydroxyl, respectively (*126*). However, upon ENT substitution, the catechol pK_a values change to between 6.5 and 8 for the dissociation of the first hydroxyl group and 11.5 for the second hydroxyl group (*18*). Therefore, under assay conditions at pH 7.5 the 2-OH of the DHB subunit of ENT can be expected to be partially deprotonated, thus making a base catalyst non-essential, but possibly rate enhancing. The absence of this base catalyst should therefore not completely prevent glycosylation at C_5. An appropriate base catalyst to carry out this function could be IroB E67, since we still observed some activity (the formation of MGE).

This notion is further supported by the hydrophobic binding pocket in IroB predicted by our model, consisting of residues F84, F85, L10, L13 and L110. The site of glycosylation is C_5, the hydrophobic hydrocarbon side of the catechol, whereas the hydroxyl groups of the catechol are hydrophilic and thus pointing away from these hydrophobic residues towards the solvent. There, the 2-OH is accessible to the predicted loop closing over ENT upon binding where E67 is located. Accordingly, the residue that accommodates the proton from C_5 after glycosylation might be located in the rigid part of the binding

pocket. Interestingly in this context, Fischbach *et al.* did not observe any deuterium exchange at C_5 of ENT in the presence of IroB and UDP, when trying to prove the existence of carbanion formation (*43*) and to explain why brominated ENT at position C_5 bound more tightly to IroB (*82*). This could indicate that C_5 is indeed located in a hydrophobic binding pocket and therefore not accessible to the solvent, providing further support for our model-based hypotheses.

With regard to point (*ii*), it is generally assumed that the aryl-C-glycosylation follows the mechanism of a Friedel-Crafts alkylation, which is shown in Figure 36 (*43, 78, 79*).
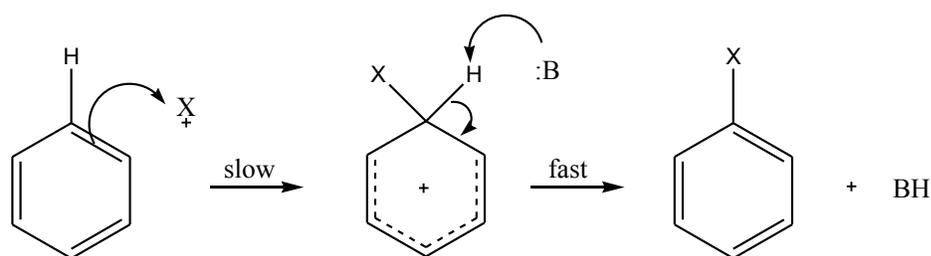
Figure 36: Mechanism of an aromatic electrophilic substitution. During Friedel-Crafts alkylation the benzene ring is substituted with an acyl group (X = [R-(C=O)-]) (figure adapated from (*124*))

In this mechanism the nucleophile attacks an electrophile under formation of a so-called σ-complex. The aromatic character is abrogated in the resulting σ-complex. However, the cation is stabilized by resonance contribution. In order to stabilize the carbocation and avoid loss of electrophile, the site of substitution must be deprotonated, and thus the aromatic system restored (*124, 125*). Accordingly, we predict the existence of *another* base catalyst in the active site of IroB, which abstracts a proton from the site of glycosylation to restore aromaticity after glycosylation has occurred. The existence of a second base catalyst was also predicted for the C-glycosyltransferase UrdGT2 for the same purpose (*74*) and was recently identified as S10 (*64*). Despite our efforts, we did not identify an appropriate base catalyst in IroB yet. However, IroB possesses a tyrosine at

position 11 - almost exactly the same position as the IroB homolog UrdGT2 (S10) - and

is therefore likely the required residue for this function. This is further supported by the

conservation of this residue among homologs (Figure 24), where usually at this position a

serine, tyrosine or histidine can be found, followed by a highly conserved glycine.

Based on the results we can propose the mechanism for C-glycosylation of ENT as

illustrated in Figure 37.
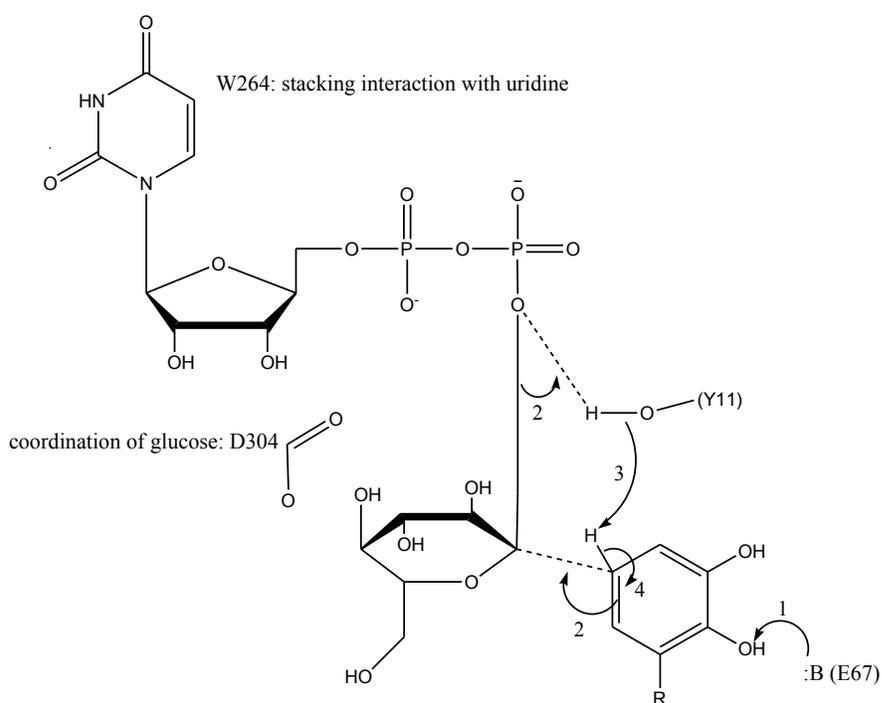
Figure 37: Proposed mechanism for the C-glycosylation of ENT.

The phenolic character of the 2,3-DHB subunit is increased by (position 1) depronotation

of the 2-OH by E67, which results in higher electron density at C_5 *para* to the 2-OH

group thus rendering this position nucleophilic. This position then attacks the anomeric

carbon of the glucose moiety (position 2) in a S_N2 displacement mechanism supported by

the phosphate leaving group. Simultaneously, Y11 acts as an acid catalyst and donates a

proton (position 2) to the phosphate leaving group thus helping to stabilize the

developing negative charge. The tyrosine then acting as a base catalyst and accommodates the proton of the site of glycosylation (position 3) to re-aromatize the 2,3-DHB subunit of ENT (position 4).

6.4   Summary and Future Work

In the course of this work we expressed and purified the C-glycosyltransferase IroB to near-homogeneity (>85%). In addition, we established an activity assay that allowed us to detect conversion products MGE/DGE/TGE arising from IroB activity, and to quantitate, in a limited manner, relative amounts of product formation. We found that divalent metal ion (Mg^{2+}) is needed for efficient catalysis, presumably to facilitate product release. We then produced a homology model of IroB that provided us with a structural framework to generate robust mechanistic hypotheses. We identified several residues that play an important role during catalysis; among these are D304, W264 and E67. Based on the information provided by our experimental data such as activity assay and biophysical characterizations, and in addition information given in the literature, we assume that W264 and D304 are likely involved in coordinating the nucleotide base and glucose moiety, respectively. The other residue that had a severe impact on catalysis was E67, which is an excellent candidate to act as a general base catalyst. It is predicted to be located on a mobile loop that could close over ENT upon binding. To confirm this prediction, we propose mutating position A61, G62 and A68 to prolines, thus introducing more rigidity into this loop. These mutations should immobilize the loop and significantly reduce activity, similar to what we observed for the E67A variant, if this

loop indeed closes over ENT upon binding. An interesting alternative could be the introduction of a Trp on the loop to follow closure and opening by a FRET-based assay. Despite our efforts we couldn't find an appropriate second base catalyst that could accommodate the proton at the site of glycosylation. However, a very recent finding with the IroB C-glycosyltransferase homologue UrdGT2 suggests that Y11 in IroB is an excellent candidate to fulfil this function. This notion is further supported by the relatively high conservation of residues at this position among glycosyltransferases. We propose substituting Y11 by a phenylalanine to remove the proton donor/acceptor that may mediate acid/base function during catalysis in our proposed mechanism (Figure 37).

Further issues that have to be addressed are the qualitative nature of our activity assay as well as the way the enzymatic reaction is terminated. Fischbach *et al.* were able to derive kinetic parameters such as k_{cat} and K_m (*43*). However, due to the unstable nature of ENT with a $t_{1/2}$ = 30 min under physiological conditions (*19*) we observed breakdown of ENT during this incubation period. To address this problem, Fe^{3+} could be added to stabilize ENT (*127*). The addition of Fe^{3+}, however, would also terminate the reaction since *holo*-ENT is not a substrate for IroB; it could also lead to precipitation of IroB consistent with our previous observations. It is notable that ferric-ENT does not bind to reversed phase column and has to be acidified *shortly* before separation is performed to avoid breakdown of ENT. However they could be obtained, kinetic parameters would be useful in determining the catalytic efficiencies of the first, second and third glycosylation steps could be compared. For this purpose, the intermediate substrates MGE/DGE would have to be purified by preparative reversed phase chromatography, and then used as a substrate

in steady-state kinetic assays. In addition, the apparent equilibrium binding affinity constant can be determined for each substrate if they were available in a purified form. Therefore, we would gain more insight into why DGE is mostly found in supernatants of *iroA* harbouring strains, but not MGE/TGE. Lastly, our polyclonal anti-IroB antibody opens the gates to perform co-immunoprecipitation experiments to identify possible protein-protein interaction partner. An interesting candidate is EntF, since an IroB-EntF interaction could facilitate channelling of ENT directly to IroB to prevent it from being exported before glycosylation occurs, or to be partitioned into the *E. coli* inner membrane upon dissociation from EntF. Straightforward experiments such as co-immunoprecipitation using Nissle 1917 (iron starved) lysate followed by 2D gel electrophoresis and MS analysis of resolved spots could identify potential IroB-interacting protein partners in a relatively short period of time.

References

1. Miethke, M., and Marahiel, M. A. (2007) Siderophore-based iron acquisition and pathogen control, *Microbiol Mol Biol Rev 71*, 413-451.
2. Waldron, K. J., and Robinson, N. J. (2009) How do bacterial cells ensure that metalloproteins get the correct metal?, *Nature Reviews Microbiology 7*, 25-35.
3. Vraspir, J. M., and Butler, A. (2009) Chemistry of Marine Ligands and Siderophores, *Annual Review of Marine Science 1*, 43-63.
4. Carrondo, M. A. (2003) Ferritins, iron uptake and storage from the bacterioferritin viewpoint, *The EMBO journal 22*, 1959-1968.
5. Carrano, C. J., and Raymond, K. N. (1978) Coordination chemistry of microbial iron transport compounds: rhodotorulic acid and iron uptake in Rhodotorula pilimanae, *Journal of bacteriology 136*, 69-74.
6. Braun, V., and Killmann, H. (1999) Bacterial solutions to the iron-supply problem, *Trends in biochemical sciences 24*, 104-109.
7. (2003) Periodicity of aqueous chemistry II: d-block chemistry, in *Inorganic Chemistry in Aqueous Solution* (Barrett, J., Ed.), pp 124-159, The Royal Society of Chemistry.
8. Aisen, P., Leibman, A., and Zweier, J. (1978) Stoichiometric and site characteristics of the binding of iron to human transferrin, *Journal of Biological Chemistry 253*, 1930-1937.
9. Martin, R. B., Savory, J., Brown, S., Bertholf, R. L., and Wills, M. R. (1987) Transferrin binding of Al3+ and Fe3+, *Clinical Chemistry 33*, 405-407.
10. Kretchmar, S. A., Reyes, Z. E., and Raymond, K. N. (1988) The spectroelectrochemical determination of the reduction potential of diferric serum transferrin, *Biochimica et biophysica acta 956*, 85-94.
11. Noinaj, N., Easley, N. C., Oke, M., Mizuno, N., Gumbart, J., Boura, E., Steere, A. N., Zak, O., Aisen, P., Tajkhorshid, E., Evans, R. W., Gorringe, A. R., Mason, A. B., Steven, A. C., and Buchanan, S. K. (2012) Structural basis for iron piracy by pathogenic Neisseria, *Nature 483*, 53-58.
12. Stojiljkovic, I., and Perkins-Balding, D. (2002) Processing of heme and heme-containing proteins by bacteria, *DNA Cell Biol 21*, 281-295.
13. Wandersman, C., and Stojiljkovic, I. (2000) Bacterial heme sources: the role of heme, hemoprotein receptors and hemophores, *Current Opinion in Microbiology 3*, 215-220.
14. Neilands, J. B. (1995) Siderophores: structure and function of microbial iron transport compounds, *J Biol Chem 270*, 26723-26726.
15. Sandy, M., and Butler, A. (2009) Microbial iron acquisition: marine and terrestrial siderophores, *Chemical reviews 109*, 4580-4595.
16. Bertini, I. (2007) *Biological inorganic chemistry : structure and reactivity*, University Science Books, Sausalito, Calif.
17. Tseng, C. F., Burger, A., Mislin, G. L., Schalk, I. J., Yu, S. S., Chan, S. I., and Abdallah, M. A. (2006) Bacterial siderophores: the solution stoichiometry and coordination of the Fe(III) complexes of pyochelin and related compounds, *Journal of biological inorganic chemistry : JBIC : a publication of the Society of Biological Inorganic Chemistry 11*, 419-432.

18.   Dertz, E. A., and Raymond, K. N. (2003) 8.6 - Siderophores and Transferrins, in *Comprehensive Coordination Chemistry II* (Editors-in-Chief: , J. A. M., and Meyer, T. J., Eds.), pp 141-168, Pergamon, Oxford.

19.   Crosa, J. H., Mey, A. R., and Payne, S. M. (2004) *Iron transport in bacteria*, ASM Press, Washington, D.C.

20.   Loomis, L. D., and Raymond, K. N. (1991) Solution equilibria of enterobactin and metal-enterobactin complexes, *Inorganic Chemistry 30*, 906-911.

21.   Raymond, K. N., Dertz, E. A., and Kim, S. S. (2003) Enterobactin: an archetype for microbial iron transport, *Proc Natl Acad Sci U S A 100*, 3584-3588.

22.   Hider, R. C., and Kong, X. (2010) Chemistry and biology of siderophores, *Nat Prod Rep 27*, 637-657.

23.   Hantke, K. (2001) Iron and metal regulation in bacteria, *Curr Opin Microbiol 4*, 172-177.

24.   Althaus, E. W., Outten, C. E., Olson, K. E., Cao, H., and O'Halloran, T. V. (1999) The ferric uptake regulation (Fur) repressor is a zinc metalloprotein, *Biochemistry 38*, 6559-6569.

25.   Bagg, A., and Neilands, J. B. (1987) Molecular Mechanism of Regulation of Siderophore-Mediated Iron Assimilation, *Microbiol Rev 51*, 509-518.

26.   Andrews, S. C., Robinson, A. K., and Rodriguez-Quinones, F. (2003) Bacterial iron homeostasis, *FEMS Microbiol Rev 27*, 215-237.

27.   Gehring, A. M., Bradley, K. A., and Walsh, C. T. (1997) Enterobactin biosynthesis in Escherichia coli: isochorismate lyase (EntB) is a bifunctional enzyme that is phosphopantetheinylated by EntD and then acylated by EntE using ATP and 2,3-dihydroxybenzoate, *Biochemistry 36*, 8495-8503.

28.   Crosa, J. H., and Walsh, C. T. (2002) Genetics and assembly line enzymology of siderophore biosynthesis in bacteria, *Microbiology and molecular biology reviews : MMBR 66*, 223-249.

29.   McKenna, W. R., Mickelsen, P. A., Sparling, P. F., and Dyer, D. W. (1988) Iron uptake from lactoferrin and transferrin by Neisseria gonorrhoeae, *Infection and immunity 56*, 785-791.

30.   Fischbach, M. A., Lin, H., Liu, D. R., and Walsh, C. T. (2006) How pathogenic bacteria evade mammalian sabotage in the battle for iron, *Nat Chem Biol 2*, 132-138.

31.   Clifton, M. C., Corrent, C., and Strong, R. K. (2009) Siderocalins: siderophore-binding proteins of the innate immune system, *Biometals 22*, 557-564.

32.   Abergel, R. J., Clifton, M. C., Pizarro, J. C., Warner, J. A., Shuh, D. K., Strong, R. K., and Raymond, K. N. (2008) The siderocalin/enterobactin interaction: a link between mammalian immunity and bacterial iron transport, *Journal of the American Chemical Society 130*, 11524-11534.

33.   Fischbach, M. A., Lin, H., Zhou, L., Yu, Y., Abergel, R. J., Liu, D. R., Raymond, K. N., Wanner, B. L., Strong, R. K., Walsh, C. T., Aderem, A., and Smith, K. D. (2006) The pathogen-associated iroA gene cluster mediates bacterial evasion of lipocalin 2, *Proc Natl Acad Sci U S A 103*, 16502-16507.

34.   Goetz, D. H., Holmes, M. A., Borregaard, N., Bluhm, M. E., Raymond, K. N., and Strong, R. K. (2002) The Neutrophil Lipocalin NGAL Is a Bacteriostatic

Agent that Interferes with Siderophore-Mediated Iron Acquisition, *Mol Cell 10*, 1033-1043.

35. Flo, T. H., Smith, K. D., Sato, S., Rodriguez, D. J., Holmes, M. A., Strong, R. K., Akira, S., and Aderem, A. (2004) Lipocalin 2 mediates an innate immune response to bacterial infection by sequestrating iron, *Nature 432*, 917-921.

36. Muller, S. I., Valdebenito, M., and Hantke, K. (2009) Salmochelin, the long-overlooked catecholate siderophore of Salmonella, *Biometals 22*, 691-695.

37. Luo, M., Lin, H., Fischbach, M. A., Liu, D. R., Walsh, C. T., and Groves, J. T. (2006) Enzymatic tailoring of enterobactin alters membrane partitioning and iron acquisition, *ACS chemical biology 1*, 29-32.

38. Konopka, K., and Neilands, J. B. (1984) Effect of serum albumin on siderophore-mediated utilization of transferrin iron, *Biochemistry 23*, 2122-2127.

39. Sharma, S., Bhat, G. K., and Shenoy, S. (2007) Virulence factors and drug resistance in Escherichia coli isolated from extraintestinal infections, *Indian J Med Microbiol 25*, 369-373.

40. Watts, R. E., Totsika, M., Challinor, V. L., Mabbett, A. N., Ulett, G. C., De Voss, J. J., and Schembri, M. A. (2012) Contribution of siderophore systems to growth and urinary tract colonization of asymptomatic bacteriuria Escherichia coli, *Infection and immunity 80*, 333-344.

41. Hantke, K., Nicholson, G., Rabsch, W., and Winkelmann, G. (2003) Salmochelins, siderophores of Salmonella enterica and uropathogenic Escherichia coli strains, are recognized by the outer membrane receptor IroN, *Proc Natl Acad Sci U S A 100*, 3677-3682.

42. Valdebenito, M., Muller, S. I., and Hantke, K. (2007) Special conditions allow binding of the siderophore salmochelin to siderocalin (NGAL-lipocalin), *FEMS Microbiol Lett 277*, 182-187.

43. Fischbach, M. A., Lin, H., Liu, D. R., and Walsh, C. T. (2005) In vitro characterization of IroB, a pathogen-associated C-glycosyltransferase, *Proc Natl Acad Sci U S A 102*, 571-576.

44. Crouch, M. L., Castor, M., Karlinsey, J. E., Kalhorn, T., and Fang, F. C. (2008) Biosynthesis and IroC-dependent export of the siderophore salmochelin are essential for virulence of Salmonella enterica serovar Typhimurium, *Molecular microbiology 67*, 971-983.

45. Zhu, M., Valdebenito, M., Winkelmann, G., and Hantke, K. (2005) Functions of the siderophore esterases IroD and IroE in iron-salmochelin utilization, *Microbiology 151*, 2363-2372.

46. Caza, M., Lepine, F., Milot, S., and Dozois, C. M. (2008) Specific roles of the iroBCDEN genes in virulence of an avian pathogenic Escherichia coli O78 strain and in production of salmochelins, *Infect Immun 76*, 3539-3549.

47. Lin, H., Fischbach, M. A., Liu, D. R., and Walsh, C. T. (2005) In vitro characterization of salmochelin and enterobactin trilactone hydrolases IroD, IroE, and Fes, *J Am Chem Soc 127*, 11075-11084.

48. Bister, B., Bischoff, D., Nicholson, G. J., Valdebenito, M., Schneider, K., Winkelmann, G., Hantke, K., and Sussmuth, R. D. (2004) The structure of salmochelins: C-glucosylated enterobactins of Salmonella enterica, *Biometals 17*, 471-481.

49. Feldmann, F., Sorsa, L. J., Hildinger, K., and Schubert, S. (2007) The salmochelin siderophore receptor IroN contributes to invasion of urothelial cells by extraintestinal pathogenic Escherichia coli in vitro, *Infect Immun 75*, 3183-3187.

50. Lairson, L. L., Henrissat, B., Davies, G. J., and Withers, S. G. (2008) Glycosyltransferases: structures, functions, and mechanisms, *Annual review of biochemistry 77*, 521-555.

51. Breton, C., Mucha, J., and Jeanneau, C. (2001) Structural and functional features of glycosyltransferases, *Biochimie 83*, 713-718.

52. Yonekura-Sakakibara, K., and Hanada, K. (2011) An evolutionary view of functional diversity in family 1 glycosyltransferases, *Plant J 66*, 182-193.

53. Rose, N. L., Palcic, M. M., and Evans, S. V. (2005) Glycosyltransferases A and B: Four Critical Amino Acids Determine Blood Type, *Journal of Chemical Education 82*, 1846.

54. Taylor, M. E., and Drickamer, K. (2006) *Introduction to glycobiology*, 2nd ed., Oxford University Press, Oxford.

55. Campbell, J. A., Davies, G. J., Bulone, V. V., and Henrissat, B. (1998) A classification of nucleotide-diphospho-sugar glycosyltransferases based on amino acid sequence similarities, *The Biochemical journal 329 (Pt 3)*, 719.

56. Bourne, Y., and Henrissat, B. (2001) Glycoside hydrolases and glycosyltransferases: families and functional modules, *Current opinion in structural biology 11*, 593-600.

57. Ünligil, U. M., and Rini, J. M. (2000) Glycosyltransferase structure and mechanism, *Current Opinion in Structural Biology 10*, 510-517.

58. Charnock, S. J., and Davies, G. J. (1999) Structure of the Nucleotide-Diphospho-Sugar Transferase, SpsA from Bacillus subtilis, in Native and Nucleotide-Complexed Forms†,‡, *Biochemistry 38*, 6380-6385.

59. Campbell, R. E., Mosimann, S. C., Tanner, M. E., and Strynadka, N. C. J. (2000) The Structure of UDP-N-Acetylglucosamine 2-Epimerase Reveals Homology to Phosphoglycosyl Transferases†,‡, *Biochemistry 39*, 14993-15001.

60. Kikuchi, N., and Narimatsu, H. (2006) Bioinformatics for comprehensive finding and analysis of glycosyltransferases, *Biochimica et biophysica acta 1760*, 578-583.

61. Hu, Y., and Walker, S. (2002) Remarkable Structural Similarities between Diverse Glycosyltransferases, *Chemistry & Biology 9*, 1287-1296.

62. Zhang, Z., Kochhar, S., and Grigorov, M. (2003) Exploring the sequence-structure protein landscape in the glycosyltransferase family, *Protein science : a publication of the Protein Society 12*, 2291-2302.

63. Breton, C., Snajdrova, L., Jeanneau, C., Koca, J., and Imberty, A. (2006) Structures and mechanisms of glycosyltransferases, *Glycobiology 16*, 29R-37R.

64. Harle, J., Gunther, S., Lauinger, B., Weber, M., Kammerer, B., Zechel, D. L., Luzhetskyy, A., and Bechthold, A. (2011) Rational design of an aryl-C-glycoside catalyst from a natural product O-glycosyltransferase, *Chemistry & biology 18*, 520-530.

65. Coutinho, P. M., Deleury, E., Davies, G. J., and Henrissat, B. (2003) An evolving hierarchical family classification for glycosyltransferases, *Journal of Molecular Biology 328*, 307-317.

66. Zechel, D. L., and Withers, S. G. (2000) Glycosidase mechanisms: Anatomy of a finely tuned catalyst, *Accounts Chem. Res. 33*, 11-18.

67. Persson, K., Ly, H. D., Dieckelmann, M., Wakarchuk, W. W., Withers, S. G., and Strynadka, N. C. (2001) Crystal structure of the retaining galactosyltransferase LgtC from Neisseria meningitidis in complex with donor and acceptor sugar analogs, *Nature structural biology 8*, 166-175.

68. Tvaroška, I. (2004) Molecular modeling insights into the catalytic mechanism of the retaining galactosyltransferase LgtC, *Carbohydrate Research 339*, 1007-1014.

69. Murray, B. W., Takayama, S., Schultz, J., and Wong, C. H. (1996) Mechanism and specificity of human alpha-1,3-fucosyltransferase V, *Biochemistry 35*, 11183-11195.

70. Ramakrishnan, B., Boeggeman, E., Ramasamy, V., and Qasba, P. K. (2004) Structure and catalytic cycle of beta-1,4-galactosyltransferase, *Current Opinion in Structural Biology 14*, 593-600.

71. Larivière, L., Gueguen-Chaignon, V., and Moréra, S. (2003) Crystal Structures of the T4 Phage β-Glucosyltransferase and the D100A Mutant in Complex with UDP-glucose: Glucose Binding and Identification of the Catalytic Base for a Direct Displacement Mechanism, *Journal of Molecular Biology 330*, 1077-1086.

72. Zhang, C., Bitto, E., Goff, R. D., Singh, S., Bingman, C. A., Griffith, B. R., Albermann, C., Phillips, G. N., Jr., and Thorson, J. S. (2008) Biochemical and structural insights of the early glycosylation steps in calicheamicin biosynthesis, *Chem Biol 15*, 842-853.

73. Bernstein, F. C., Koetzle, T. F., Williams, G. J., Meyer, E. F., Jr., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., and Tasumi, M. (1978) The Protein Data Bank: a computer-based archival file for macromolecular structures, *Arch Biochem Biophys 185*, 584-591.

74. Mittler, M., Bechthold, A., and Schulz, G. E. (2007) Structure and action of the C-C bond-forming glycosyltransferase UrdGT2 involved in the biosynthesis of the antibiotic urdamycin, *J Mol Biol 372*, 67-76.

75. Matsumoto, T., Katsuki, M., and Suzuki, K. (1988) New approach to C-aryl glycosides starting from phenol and glycosyl fluoride. Lewis acid-catalyzed rearrangement of O-glycoside to C-glycoside, *Tetrahedron Letters 29*, 6935-6938.

76. Hosoya, T., Ohashi, Y., Matsumoto, T., and Suzuki, K. (1996) On the stereochemistry of aryl C-glycosides: Unusual behavior of bis-TBDPS protected aryl C-olivosides, *Tetrahedron Letters 37*, 663-666.

77. Palmacci, E. R., and Seeberger, P. H. (2001) Synthesis of C-Aryl and C-Alkyl Glycosides Using Glycosyl Phosphates, *Organic Letters 3*, 1547-1550.

78. Bililign, T., Griffith, B. R., and Thorson, J. S. (2005) Structure, activity, synthesis and biosynthesis of aryl-C-glycosides, *Natural product reports 22*, 742-760.

79. Bililign, T., Hyun, C. G., Williams, J. S., Czisny, A. M., and Thorson, J. S. (2004) The hedamycin locus implicates a novel aromatic PKS priming mechanism, *Chemistry & biology 11*, 959-969.

80. Hayashi, M., Nakayama, S.-z., and Kawabata, H. (2000) HF-pyridine promoted Friedel-Crafts type arylation of 2-acetoxy--glucal. Stereoselective synthesis of 1-arylhex-3-enopyranosiduloses, *Chemical Communications*, 1329-1330.

81. Matsumoto, T., Katsuki, M., and Suzuki, K. (1989) Cp2ZrCl2-AgClO4: Efficient promoter for the Friedel-Crafts approach to C-aryl glycosides, *Tetrahedron Letters 30*, 833-836.

82. Lin, H., Fischbach, M. A., Gatto, G. J., Jr., Liu, D. R., and Walsh, C. T. (2006) Bromoenterobactins as potent inhibitors of a pathogen-associated, siderophore-modifying C-glycosyltransferase, *Journal of the American Chemical Society 128*, 9324-9325.

83. Grozdanov, L., Raasch, C., Schulze, J., Sonnenborn, U., Gottschalk, G., Hacker, J., and Dobrindt, U. (2004) Analysis of the genome structure of the nonpathogenic probiotic Escherichia coli strain Nissle 1917, *Journal of bacteriology 186*, 5432-5441.

84. Candiano, G., Bruschi, M., Musante, L., Santucci, L., Ghiggeri, G. M., Carnemolla, B., Orecchia, P., Zardi, L., and Righetti, P. G. (2004) Blue silver: a very sensitive colloidal Coomassie G-250 staining for proteome analysis, *Electrophoresis 25*, 1327-1333.

85. Eswar, N., Webb, B., Marti-Renom, M. A., Madhusudhan, M. S., Eramian, D., Shen, M. Y., Pieper, U., and Sali, A. (2007) Comparative protein structure modeling using MODELLER, *Curr Protoc Protein Sci Chapter 2*, Unit 2 9.

86. Fiser, A., Do, R. K., and Sali, A. (2000) Modeling of loops in protein structures, *Protein Sci 9*, 1753-1773.

87. Marti-Renom, M. A., Stuart, A. C., Fiser, A., Sanchez, R., Melo, F., and Sali, A. (2000) Comparative protein structure modeling of genes and genomes, *Annu Rev Biophys Biomol Struct 29*, 291-325.

88. Sali, A., and Blundell, T. L. (1993) Comparative protein modelling by satisfaction of spatial restraints, *J Mol Biol 234*, 779-815.

89. Kleywegt, G. J., and Jones, T. A. (1997) Detecting folding motifs and similarities in protein structures, *Methods in enzymology 277*, 525-545.

90. Kleywegt, G. J. (1999) Experimental assessment of differences between related protein crystal structures, *Acta crystallographica. Section D, Biological crystallography 55*, 1878-1884.

91. Willard, L., Ranjan, A., Zhang, H., Monzavi, H., Boyko, R. F., Sykes, B. D., and Wishart, D. S. (2003) VADAR: a web server for quantitative evaluation of protein structure quality, *Nucleic acids research 31*, 3316-3319.

92. Wiederstein, M., and Sippl, M. J. (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins, *Nucleic acids research 35*, W407-410.

93. Sippl, M. J. (1993) Recognition of errors in three-dimensional structures of proteins, *Proteins 17*, 355-362.

94. Bond, C. S. (2003) TopDraw: a sketchpad for protein structure topology cartoons, *Bioinformatics 19*, 311-312.

95. Wilkins, M. R., Gasteiger, E., Bairoch, A., Sanchez, J. C., Williams, K. L., Appel, R. D., and Hochstrasser, D. F. (1999) Protein identification and analysis tools in the ExPASy server, *Methods in molecular biology 112*, 531-552.

96. Notredame, C., Higgins, D. G., and Heringa, J. (2000) T-Coffee: A novel method for fast and accurate multiple sequence alignment, *Journal of Molecular Biology 302*, 205-217.

97.   Di Tommaso, P., Moretti, S., Xenarios, I., Orobitg, M., Montanyola, A., Chang, J. M., Taly, J. F., and Notredame, C. (2011) T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension, *Nucleic acids research 39*, W13-17.

98.   Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M., and Barton, G. J. (2009) Jalview Version 2--a multiple sequence alignment editor and analysis workbench, *Bioinformatics 25*, 1189-1191.

99.   Sambrook, J. (2001) *Molecular cloning : a laboratory manual / Joseph Sambrook, David W. Russell*, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. :.

100.   Monika, J. (2006) *Gentechnische Methoden: Eine Sammlung von Arbeitsanleitungen für das molekularbiologische Labor*, Elsevier, Spektrum Akad. Verl.

101.   Laemmli, U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4, *Nature 227*, 680-685.

102.   Tivendale, K. A., Noormohammadi, A. H., Allen, J. L., and Browning, G. F. (2009) The conserved portion of the putative virulence region contributes to virulence of avian pathogenic Escherichia coli, *Microbiology 155*, 450-460.

103.   van de Weert, M., and Stella, L. (2011) Fluorescence quenching and ligand binding: A critical discussion of a popular methodology, *Journal of Molecular Structure 998*, 144-150.

104.   Jaroszewski, L., Rychlewski, L., Li, Z., Li, W., and Godzik, A. (2005) FFAS03: a server for profile--profile sequence alignments, *Nucleic Acids Res 33*, W284-288.

105.   Jaroszewski, L., Rychlewski, L., and Godzik, A. (2000) Improving the quality of twilight-zone alignments, *Protein Sci 9*, 1487-1496.

106.   Rychlewski, L., Jaroszewski, L., Li, W., and Godzik, A. (2000) Comparison of sequence profiles. Strategies for structural predictions using sequence information, *Protein Sci 9*, 232-241.

107.   Soding, J., Biegert, A., and Lupas, A. N. (2005) The HHpred interactive server for protein homology detection and structure prediction, *Nucleic acids research 33*, W244-248.

108.   Melo, F., Sanchez, R., and Sali, A. (2002) Statistical potentials for fold assessment, *Protein science : a publication of the Protein Society 11*, 430-448.

109.   Shen, M. Y., and Sali, A. (2006) Statistical potential for assessment and prediction of protein structures, *Protein science : a publication of the Protein Society 15*, 2507-2524.

110.   Brändén, C.-I., and Tooze, J. (1999) *Introduction to protein structure*, 2nd ed., Garland Pub., New York.

111.   Fink, A. L. (1998) Protein aggregation: folding aggregates, inclusion bodies and amyloid, *Folding and Design 3*, R9-R23.

112.   Bondos, S. E., and Bicknell, A. (2003) Detection and prevention of protein aggregation before, during, and after purification, *Analytical biochemistry 316*, 223-231.

113.   Scopes, R. K. (1994) *Protein purification : principles and practice*, 3rd ed., Springer-Verlag, New York.

114. Street, T. O., Bolen, D. W., and Rose, G. D. (2006) A molecular mechanism for osmolyte-induced protein stability, *P Natl Acad Sci USA 103*, 13997-14002.

115. Zou, Q., Bennion, B. J., Daggett, V., and Murphy, K. P. (2002) The molecular mechanism of stabilization of proteins by TMAO and its ability to counteract the effects of urea, *Journal of the American Chemical Society 124*, 1192-1202.

116. Cho, S. S., Reddy, G., Straub, J. E., and Thirumalai, D. (2011) Entropic stabilization of proteins by TMAO, *The journal of physical chemistry. B 115*, 13401-13407.

117. Imberty, A., Wimmerova, M., Koca, J., and Breton, C. (2006) Molecular modeling of glycosyltransferases, *Methods Mol Biol 347*, 145-156.

118. Mulichak, A. M., Losey, H. C., Walsh, C. T., and Garavito, R. M. (2001) Structure of the UDP-glucosyltransferase GtfB that modifies the heptapeptide aglycone in the biosynthesis of vancomycin group antibiotics, *Structure 9*, 547-557.

119. Ma, J., Campbell, A., and Karlin, S. (2002) Correlations between Shine-Dalgarno sequences and gene features such as predicted expression levels and operon structures, *Journal of bacteriology 184*, 5733-5745.

120. Frottin, F., Martinez, A., Peynot, P., Mitra, S., Holz, R. C., Giglione, C., and Meinnel, T. (2006) The proteomics of N-terminal methionine cleavage, *Molecular & cellular proteomics : MCP 5*, 2336-2349.

121. Zhu, F., Erlandsen, H., Ding, L., Li, J., Huang, Y., Zhou, M., Liang, X., Ma, J., and Wu, H. (2011) Structural and functional analysis of a new subfamily of glycosyltransferases required for glycosylation of serine-rich streptococcal adhesins, *The Journal of biological chemistry 286*, 27048-27057.

122. Offen, W., Martinez-Fleites, C., Yang, M., Kiat-Lim, E., Davis, B. G., Tarling, C. A., Ford, C. M., Bowles, D. J., and Davies, G. J. (2006) Structure of a flavonoid glucosyltransferase reveals the basis for plant natural product modification, *The EMBO journal 25*, 1396-1405.

123. Chang, A., Singh, S., Helmich, K. E., Goff, R. D., Bingman, C. A., Thorson, J. S., and Phillips, G. N., Jr. (2011) Complete set of glycosyltransferase structures in the calicheamicin biosynthetic pathway reveals the origin of regiospecificity, *P Natl Acad Sci USA 108*, 17649-17654.

124. Smith, M., and March, J. (2007) *March's advanced organic chemistry : reactions, mechanisms, and structure*, 6th ed., pp 675-758, Wiley-Interscience, Hoboken, N.J.

125. Carey, F. A., and Giuliano, R. M. (2011) *Organic chemistry*, 8th ed., Ch. 14 and 24, McGraw-Hill, New York.

126. Schweigert, N., Zehnder, A. J. B., and Eggen, R. I. L. (2001) Chemical properties of catechols and their molecular modes of toxic action in cells, from microorganisms to mammals, *Environ Microbiol 3*, 81-91.

127. Payne, S. M. (1994) [25] Detection, isolation, and characterization of siderophores, in *Methods in Enzymology* (Virginia L. Clark, P. M. B., Ed.), pp 329-344, Academic Press.