Bacterial cell wall recycling in molecular detail

X-ray structural analysis of MurNAc-1P Uridylyltransferase MurU from Gram-negative *P.putida* and MurNAc-6P etherase MurQ from Gram-positive *B.subtilis*

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät

der Eberhard Karls Universität Tübingen

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

vorgelegt von

Michaela Gabriele Renner-Schneck

aus Tübingen

Tübingen

2015

Tag der mündlichen Qualifikation:	24.06.2015
Dekan:	Prof. Dr. Wolfgang Rosenstiel
1. Berichterstatter:	Prof. Dr. Thilo Stehle
2. Berichterstatter:	apl. Prof. Dr. Christoph Mayer

Acknowledgements:

With these words I want to thank some very special people for helping me to finally accomplish this PhD project:

Thilo for giving me the opportunity to be a PhD-student in his lab. For the trust he put in me, the patience, and persistent encouragement.
The scientific and personal experiences in your lab greatly strengthened my scientific and personal assertiveness but made me also enjoy the time in your lab.

Christoph Mayer, Amanda Duckworth, Jonathan Gisin, Isabel Hinderberger and Alexander Schneider for their cooperation in the MurQ and MurU projects.

Georg, Dirk, and Sonja for guidance when I was a rookie in terms of X-ray crystallography and linux systems.

Dirk, Sonja, Michael, Felix and Maunel for great times in our office and elsewhere.

Alex, Caro, Kerstin and Nici for the coffe-breaks and funny nights with the ‘girls gang’.

My best friend Michael, who suffered from the same stuff but in a different setting.

Allen Freunden außerhalb der Uni, für ihre Wertschätzung, Aufmunterung oder einfach nur für gemeinsame Trainingskilometer.

Meinen Eltern, für ihre Unterstützung – all die ganzen Jahre.

Meiner Schwester für ihre ehrliche ‘Systemkritik’ – und für’s Feiern.

JOE… !

1. X-ray crystallography, an outstanding tool to identify structure and function

The determination of the crystal structure of diamonds by William Henry Bragg and his son William Lawrence Bragg in 1913 can be viewed as the birth of X-ray crystallography. Thus, this method has only recently turned one hundred years old, and can now look back on a century in which it has often enriched and sometimes revolutionized scientific research. To mark the occasion, the year 2014 has been announced as the "International Year of Crystallography" (iycr2014, 2014) and the journals Nature and Science have launched dedicated special issues to review the historical milestones of X-ray crystallography, its achievements, developments and future prospects (NATURE, 2014, SCIENCE, 2014). It is beyond the scope of this work to give a similarly detailed overview over the first crystallographic century. It is clear, however that X-ray crystallography has come a long way from the very first diffraction images of a crystal recorded by Max von Laue in 1912 to today's landmark structures, many of which have been honored with Nobel Prizes (Sumner, 2014). Today, the Protein Data Bank (PDB, (www.pdb.org)) features over 100.000 structures and over 600.000 structures of organic and organometallic molecules are stored in the Cambridge Structural Database many of which have also been solved by X-ray crystallography. The timeline 'evolution of X-ray crystallography' shown in Fig. 1 attempts to capture at least some of the successes over the last 100 years.

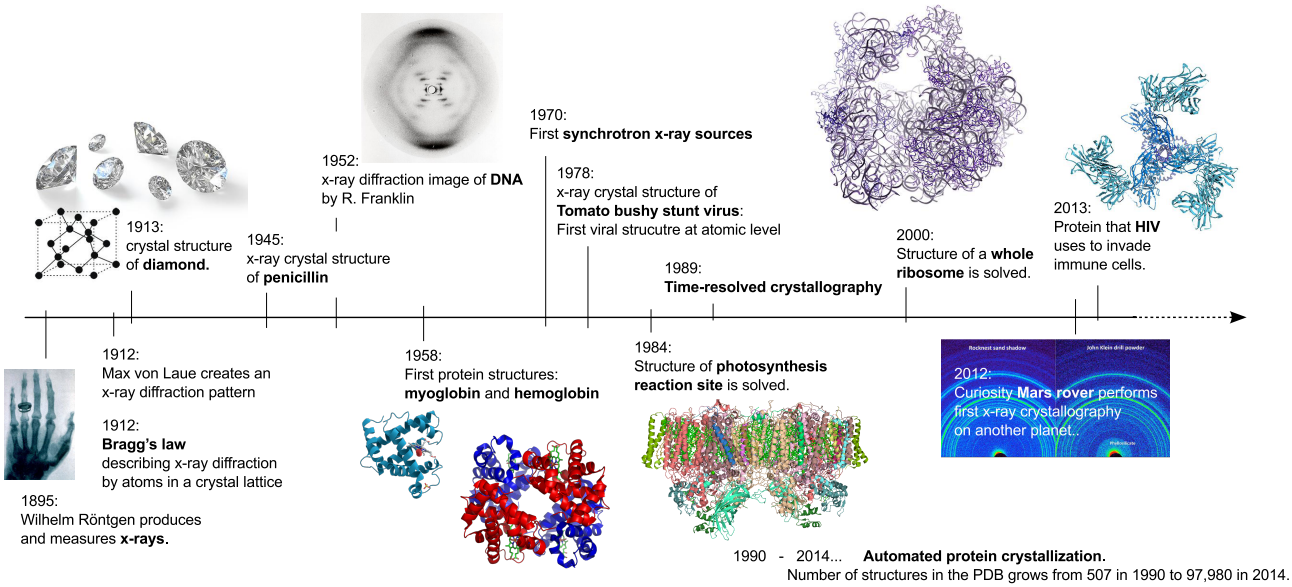


Fig. 1: The „evolution" of X-ray crystallography from 1895 until 2014

2. Biological background of this work

2.1 The bacterial PGN

The bacterial cell wall is an elastic macromolecule that defines the shape of the bacterium and prevents it from lysis due to its high intracellular osmotic pressure.

Bacterial cell walls can be divided into two major classes: Gram- positive and Gram-negative. Gram-positive cell walls consist of one lipid bilayer and a thick peptidoglycan (PGN) layer with associated proteins and phosphate-rich polymers called teichoic acids. In Gram-negative bacteria the PGN layer is significantly thinner, has no proteins associated and is sandwiched between the inner plasma membrane and an outer lipopolysaccharide rich cell envelope, both with associated proteins (see Fig. 2, (Silhavy *et al.*, 2010)). The PGN is an essential component of both Gram-positive and Gram-negative bacterial cell walls. It consists of elongated polysaccharide chains of alternating N-acetylglucosamine (GlcNAc) and N-acetylmuramic acid (MurNAc) sugar units that are cross-linked at the lactyl-moiety of MurNAc via short oligopeptide bridges, forming a huge mesh-like macromolecule (see Fig. 2, (Vollmer *et al.*, 2008)). The amino acid composition of these peptide linkers differs between Gram-positive and Gram-negative cell walls (reviewed by Schleifer and Kandler (1972) (Schleifer & O., 1972)). The PGN of Gram-positive bacteria generally contains l-lysine and a pentaglycine bridge between two connected peptide stems, whereas Gram-negative bacteria incorporate meso-1,6-diaminopimelic acid (DAP in Fig. 2) instead of lysine and link the peptide stems directly with each other. Moreover, numerous variations in the structure of the pentapeptide stem are known among different species (Johnson *et al.*, 2013). In Fig. 2 the most common core structure of the peptide linkers of Gram-positive and Gram-negative cell walls is depicted.

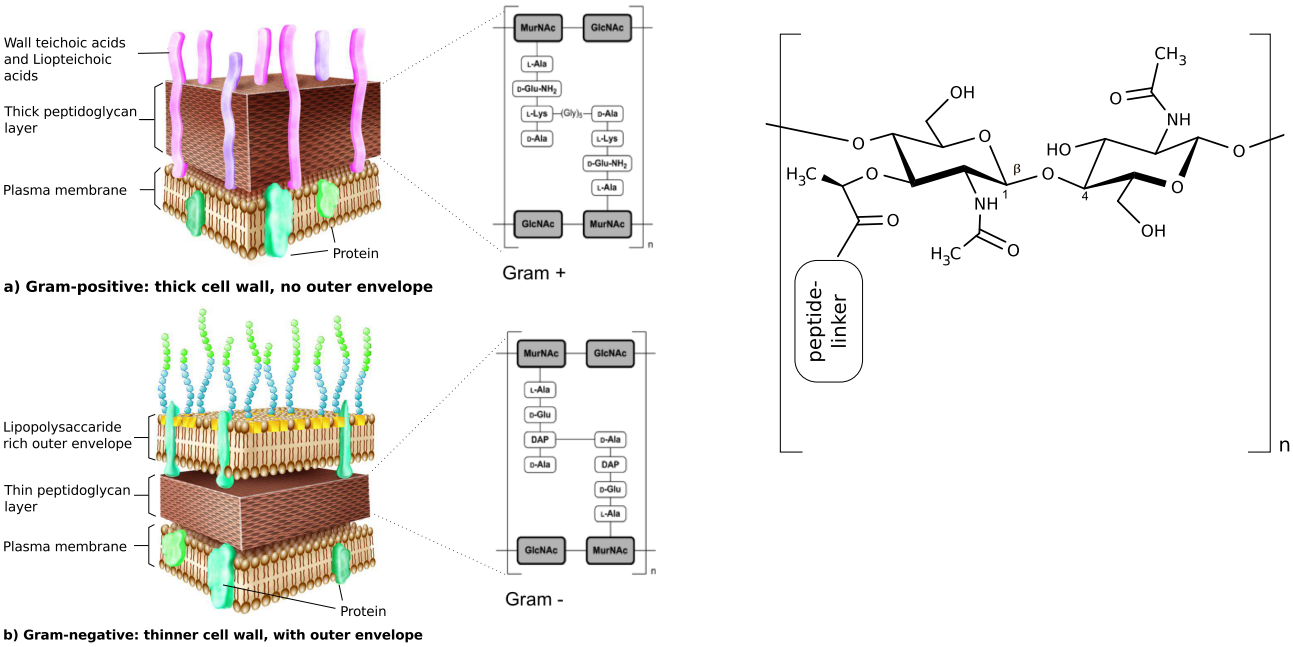


Fig. 2: Composition oft the bacterial cell wall (left) and the peptidoglycan layer (middle and right) in a) Gram positive and b) Gram negative organisms. Adapted from http://biology-forums.com/index.php?opic=5003.0

The term PGN was originally defined as a class of polymers composed of approximately equal amounts of amino acids and sugars, whereas the term murein was reserved for the PGN present in the cell envelopes of bacteria. However, as the only PGN polymer presently recognized in nature is the one found in bacterial cell walls, PGN has been used increasingly in place of murein (Park & Uehara, 2008) and will be used exclusively to refer to bacterial cell wall PGN in this work.

2.1.1 PGN biosynthesis

Biosynthesis oft he PGN meshwork starts in the cytosol with the formation of UDP-N-acetyl-muramyl-pentapeptide (UDP-MurNAc-pentapeptide) from UDP-MurNAc. This precursor is attached to a polyprenyl membrane anchor and the N-acetyl-glucosamine moiety is added before the whole building block is translocated across the plasma membrane. Polymerization occurs in a glycosyltransferase reaction and the peptidyl-bridges are cross-linked via transpeptidation (Fig. 3) (Lovering *et al.*, 2012).

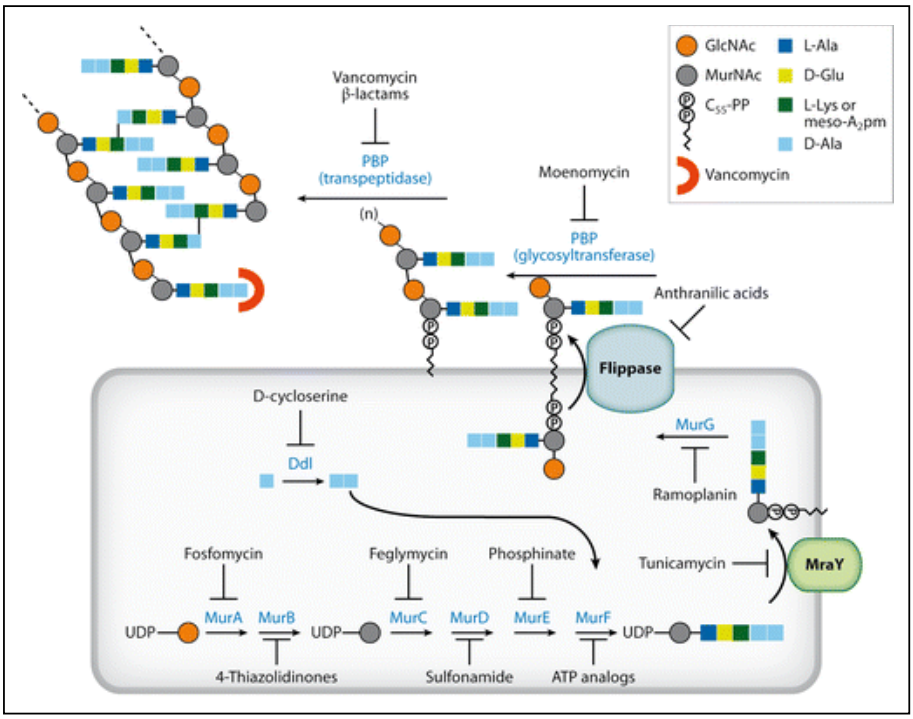


Fig. 3: Schematic representation of the bacterial peptidoglycan biosynthesis pathway. Enzymatic reactions (arrows) and some selected inhibitors (blunt arrows) are shown. Integral membrane proteins (MraY and flippase) are present at the transition between cytoplasmic synthesis of the uridine diphosphate (UDP)-linked precursor and outer leaflet utilization of the lipid II monomer. Abbreviations: C55-PP, undecaprenyl diphosphate; D-Ala, D-alanine; Ddl, D-alanyl-D-alanine ligase; D-Glu, D-glutamic acid; GlcNAc, *N*-acetylglucosamine; L-Ala, L-alanine; L-Lys, L-lysine; meso-A2pm, mesodiaminopimelic acid; MurA, MurB, MurC, MurD, MurE, MurF, MurG, enzymes involved in the cytoplasmic biosynthesis steps of peptidoglycan; MurNAc, *N*-acetylmuramic acid; PBP, penicillin-binding protein. Figure adapted from (Lovering *et al.*, 2012).

2.1.2 PGN biosynthesis as a prominent target pathway for antibiotics

The PGN layer present in the bacterial cell wall is the only PGN polymer currently recognized in nature (Park & Uehara, 2008). Its biosynthesis is a biochemical pathway that is unique for bacterial species and consequently represents a prominent target for antibiotic drug design. In particular, the cell wall sugar MurNAc is exclusively occurring in bacteria and functions as a structural branching point in the PGN macromolecule (Fig. 2) (Litzinger & Mayer, 2010, Jaeger & Mayer, 2008). Moreover, UDP-MurNAc is the first committed precursor in PGN biosynthesis (Brown *et al.*, 1995). Its formation occurs in the cytosol in two steps from phosphoenol pyruvate (PEP) and UDP-GlcNAc, catalyzed by the enolpyruvyl transferase MurA and the reductase MurB. The former enzyme can be blocked by the PEP-analog fosfomycin ((1*R*,2*S*)-1,2-epoxypropylphosphonic acid), which irreversibly inhibits MurA, resulting in a

reduced pool of UDP-MurNAc and decelerated biosynthesis (F.M. Kahan *et al.*, 1974) – see text below and Fig. 5+Fig. 6.

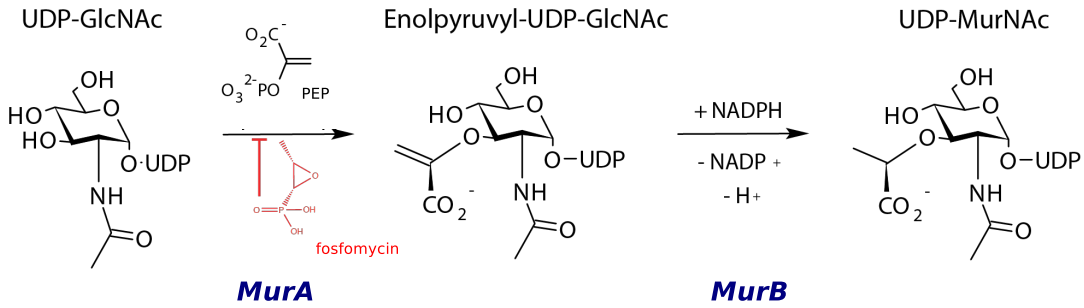


Fig. 4: Formation of UDP-MurNAc by the sequential action of MurA and MurB.
The enolpyruvyl transferase MurU can be irreversible inhibited by the PEP-analog fosfomycin (depicted in red). Adapted from M. Vorländer (Master thesis)

Another group of antibacterial drugs interfering with cell wall biosynthesis are the β-lactam antibiotics. They block the cross-linking of glycan strands by binding to the transpeptidase domain of bifunctional enzymes called penicillin-binding proteins (PBPs), which catalyze the polymerization of the glycan strand and cross-linking of the peptide stems. In this context, the relationship between cell wall recycling and β-lactamase induction is of particular interest. It has been known for many years that, in Gram-negative bacteria, the presence of β-lactam antibiotics is sensed by perturbations in the cytoplasmic pool of muropeptides (cell wall fragments of the composition $[MurNAc\text{-}GlcNAc]_n$-(anhydro)MurNAc – peptide linker). This results in de-repression of the gene that encodes the AmpC β-lactamase, which in turn renders the organism resistant to β-lactam antibiotics. Similar links between β-lactamase induction and cell wall recycling have been revealed in some Gram-positive organisms only recently (Johnson *et al.*, 2013). An important implication of the link between recycling and resistance is that inhibitors of cell wall recycling and biosynthesis pathways, such as the above-mentioned fosfomycin, might be combined with cell wall-targeting antibiotics. Such a strategy could prove useful in treating infections with multi-drug resistant pathogens such as methicillin-resistant Staphylococcus aureus (MRSA), which produces the β-lactam-resistant penicillin-binding protein PBP2a (Johnson *et al.*, 2013).

A further interesting aspect regarding bacterial cell walls is that muropeptides have important messenger functions in bacterial communication, as signal molecules, and immuno-evasins. In eukaryotes for example, the detection of muropeptides (e.g. via peptidoglycan- recognizing proteins and NOD receptors) initiates an immune response, and the recovery of cell-wall muropeptides suppresses this response (Johnson *et al.*, 2013).

2.1.3 Recycling of the bacterial PGN at the example of Gram-negative *E.coli*

The biosynthesis of the PGN polymer is a highly conserved multistep process that is carefully coordinated with the cell wall remodeling processes required for cell growth and division. In the bacterial life cycle the peptidoglycan polymer is constantly degraded and resynthesized explaining why cell wall lytic enzymes (autolysins) like carboxypeptidases, endopeptidases, muramidases, lytic transglycosylases or amidases are present in virtually all bacteria (Shockman *et al.*, 1996, Johnson *et al.*, 2013). Actually, the crystallographers favorite model protein, hen egg white lysozyme, which is possibly the best studied enzyme of all, is a muramidase that hydrolyzes the glycosidic bond between MurNAc and GlcNAc in the PGN (Jollès, 1996), however lysozyme is not involved in cell wall remodeling. Many bacteria remodel as much as half of their cell wall per generation. Thus, even though recycling of cell wall components is not essential for bacterial growth, liberation of cell wall fragments into the medium would represent a significant loss of resources if they were not recovered and recycled (Park & Uehara, 2008, Mayer, 2012, Johnson *et al.*, 2013). Indeed, it is known that many bacteria are able to reuse a huge proportion of the peptidoglycan catabolites.

Cell wall recycling has been extensively studied in the Gram-negative model organism *E.coli,* which possesses a large number of specifically dedicated enzymes. This is remarkable, since the peptidoglycan layer accounts for only 1-2 % of its cell dry mass of *E.coli*, and cell wall recycling is not essential for the organism's survival. In Gram-positive bacteria the relative amount of cell wall material is much higher and hence cell wall recycling might be of even greater significance in these organisms (Jaeger & Mayer, 2008). In the last years a much progress has been made in elucidating the unclear mechanisms of peptidoglycan recycling in Gram positive and Gram-negative organisms and in the characterization of the enzymes involved. However, up to date the information about these pathways in organisms other than *E.coli* are still rather incomplete.

Thus in the following, the *E.coli* model system will serve as a guideline to help the reader keep the overview when being introduced to cell wall biosynthesis and recycling of the two organisms, *B.subtilis* and *P.putida* that are in the focus of this work (see Fig. 5). In *E. coli* about 50 % of the PGN of the endogenous cell wall is recycled, and PGN fragments from the environment are utilized as well. GlcNAc – anhydroMurNAc (anhMurNAc)-tetrapeptides are released from the PGN by the action of lytic transglycosylases (LT) and endopeptidases (EP), whereas amidases release tripeptides consisting of alanine, glutamate and diaminopimelic acid (Ala-Glu-DAP). The PGN catabolites are imported into the cytoplasm where they are further degraded (Dahl *et al.*, 2004, Jaeger & Mayer, 2008). GlcNAc-anhMurNAc-tetrapeptides are taken up by the secondary transporter AmpG, whereas Ala-Glu-DAP tripeptides pass the inner membrane via the ABC transporter MppA/Opp. Additionally, the individual amino sugars, GlcNAc, MurNAc and anhMurNAc, of the endogenous cell wall as well as from the environment, can also be taken up separately by the phosphotransferase systems (PTS) NagE and

MurP. Concomitantly to their cytosolic import GlcNAc and MurNAc are phosphorylated, resulting in a net import of GlcNAc-6-phosphate and MurNAc-6-phosphate (Jaeger & Mayer, 2008).

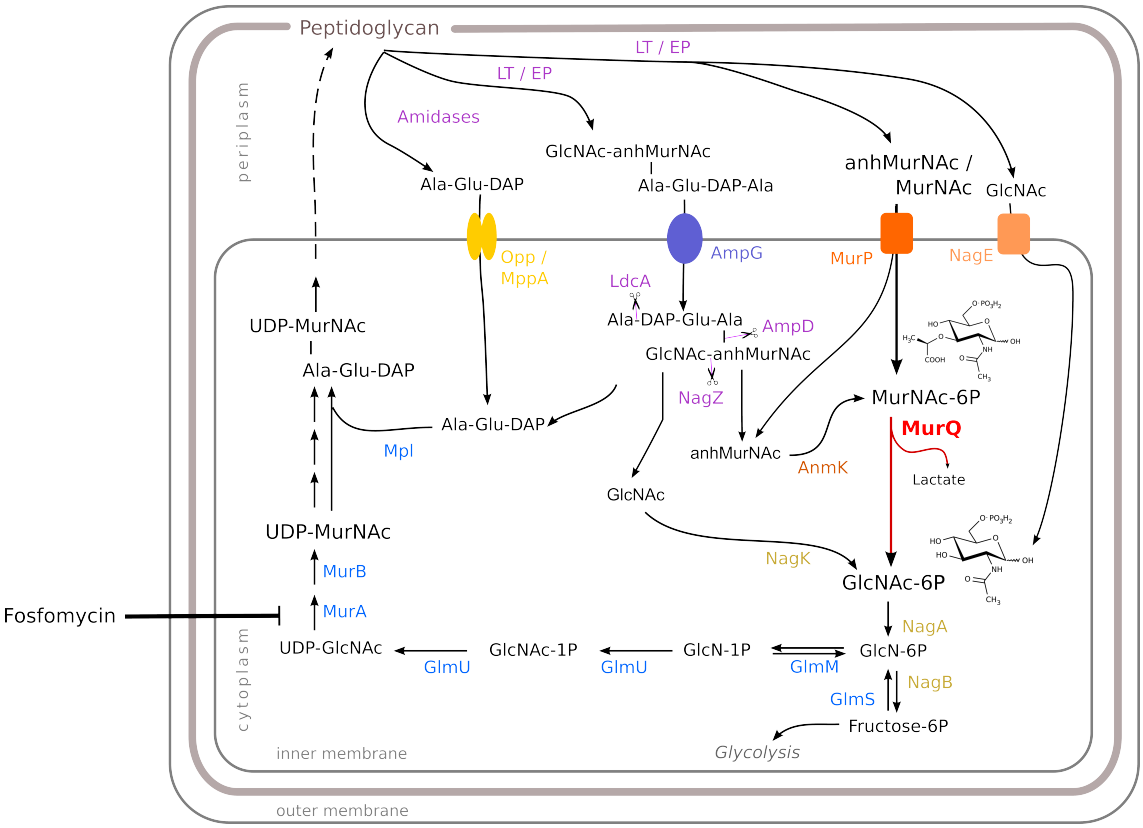


Fig. 5: Cell wall recycling in *E.coli*
For an integrative description please refer to the text. Names of catabolic enzymes are written in purple and those of enzymes contributing to PGN synthesis in blue. The sequential activity of the isomerases GlmS and GlmM, the bifunctional acetyltransferase/uridylyltransferase GlmU as well as the enolpyruvyl transferase MurA and the reductase Mur B account for the glucosamine phosphate pathway mentioned in the text. The reaction of the etherase MurQ is highlighted in red and the chemical structure of its sugar substrate and the reaction product are shown.
Further abbreviations and enzyme descriptions: LT: lytic transglycosylases; EP: endopeptidases; DAP: diaminopimelic acid; AmpG: a secondary transporter for GlcNAc-anhMurNAc-peptides; MppA/Opp: transporter for Ala-GLu-DAP; NagE: phosphotransferase systems (PTS) for GlcNAc; MurP: PTS for MurNAc and anhMurNAc; MurQ: etherase cleaving off the D-lactic acid substituent of MurNAc-6P; NagA: glucosamine-6-phosphate (GlcN-6P) deacetylase; NagZ: N-acetylglucosaminidase, AmpD: anhMurNAc-peptide amidases; LdcA, a D-Ala-carboxypeptidase; Mpl: muropeptide ligase; GlmS: glucosamine-6-phosphate synthase ; GlmM: glucosamine-phosphate isomerase; GlmU: bifunctional glucosamine-1P (GlcN-1P) acetyltransferase / N-acetylglucosamine-1-phosphate (GlcNAc-1P) uridylyltransferase.

The etherase MurQ, a distinctive enzyme of *E.coli* peptidoglycan recycling (Jaeger *et al.*, 2005) with orthologs in many other organisms (Jaeger & Mayer, 2008), cleaves off the D-lactic acid substituent of 6-OH phosphorylated MurNAc, generating GlcNAc-6P, the common metabolite that merges the catabolism of the two cell wall sugars. GlcNAc-6P is subsequently converted to glucosamine-6-phosphate (GlcN-6P) by the deacetylase NagA (Jaeger & Mayer, 2008, Hadi *et al.*, 2008, Park & Uehara, 2008), which can either be metabolized in glycolysis or go into PGN synthesis. Moreover, MurQ is

required for the utilization of anhMurNAc, which is released from anhMurNAc-peptides in the cytoplasm by the action of NagZ, an N-acetylglucosaminidase, AmpD, an anhMurNAc-peptide amidase, and LdcA, a D-Ala-carboxypeptidase. Recycling of the amino sugars of the cell wall for PGN synthesis proceeds through the glucosamine phosphate pathway (see Fig. 5) and recycling of the peptides released from the cell wall proceeds through the muropeptide ligase, Mpl. An overview of PGN biosynthesis and cell wall recycling in E.coli is given in Fig. 5.

The basal constitutive expression of MurQ is rather low in *E.coli*, since only a low level of etherase is needed for recycling purposes during normal growth. However, *E.coli,* and some other bacteria are able to utilize cell wall components such as GlcNAc and MurNAc as sole source of carbon, nitrogen and energy. Growth of *E.coli* on GlcNAc and MurNAc requires the *nag* gene products (NagE/B/A/C/D) as well as MurQ, the glycosidase/kinase AnmK and the MurNAc-specific transporter MurP (see Fig. 5). High level expression of both MurNAc etherase and the PTS has been reported when *E. coli* is grown on GlcNAc or MurNAc (Jaeger & Mayer, 2008). MurQ and MurP are encoded by two adjacent genes, *murQ* and *murP* that are regulated by the repressor MurR, which is encoded in the same gene cluster. MurR and MurQ are structurally related in that they both proteins contain a sugar-phosphate binding domain (SIS domain, see below) and bind MurNAc 6-phosphate. Binding of MurNAc-6P inactivates the repressor MurR by releasing it from the operator DNA (*murQP)* thereby facilitating transcription of the genes *murQ* and *murP*. Furthermore high level expression of MurQ depends on activation by CAP (cyclic AMP (cAMP) catabolite activator protein) (Jaeger & Mayer, 2008).

2.2 Project Description MurU:

2.2.1 PGN recycling in Gram-negative *P.putida*

Recent studies revealed that despite broadly relying on a similar set of enzymes (Shockman *et al.*, 1996), the occurrence of cell wall recycling enzymes differs among bacteria. Specifically all *pseudomonades* but also many other Gram-negative bacteria, including important pathogens, lack an ortholog of MurQ. These bacteria instead possess a shortcut from MurNAc-6P to UDP-MurNAc that has been described recently for *Ps. putida KT2440* and *P. aeruginosa PA1/PA14*. (Gisin *et al.*, 2013, Borisova *et al.*, 2014). This pathway bypasses peptidoglycan *de novo* biosynthesis and thus endows the bacterium with an intrinsic resistance against the antibiotic fosfomycin (Fig. 6). Three novel, sequentially acting enzymes were identified to account for this shortcut: a MurNAc-6P phosphatase, the kinase AmgK that re-phosphorylates the resulting MurNAc at its C1 hydroxyl-group and the nucleotidyltransferase MurU that generates UDP-MurNAc from UTP and MurNAc-1P (Fig. 6).

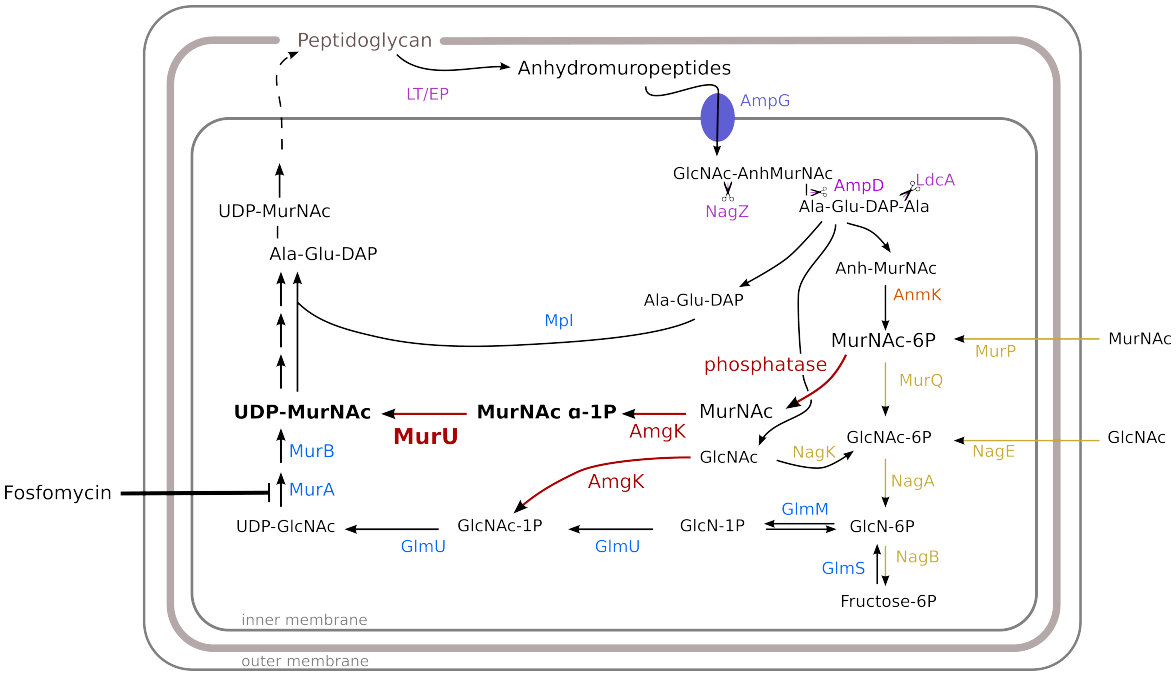


Fig. 6: Simplified scheme of PGN synthesis and recycling pathways of *E.coli* and *P.putida*. In both species, anhydromuropeptides are steadily released during growth by lytic transglycosylases and endopeptidases (LT/EP) and imported by AmpG. In the cytoplasm anhydromuropeptides are hydrolyzed by the β-N-acetylglucosaminidase NagZ, the anhMurNAc-L-alanine amidase AmpD and the L,D-carboxypeptidase LdcA yielding GlcNAc, anhMurNAc, Ala-Glu-DAP and D-Ala. The muropeptide ligase Mpl links the latter tripeptide to UDP-MurNAc thereby generating the PGN precursor UDP-MurNAc-Ala-Glu-DAP. (Jaeger & Mayer, 2008). In *P.putida*, the kinase AmgK, MurU and a phosphatase constitute a direct recycling pathway for cell wall amino sugars (red labels and arrows) providing a shortcut to UDP-MurNAc and thereby bypassing the MurA- and MurB-catalyzed de novo synthesis of UDP-MurNAc, that is susceptible to fosfomycin.
Enzymes that are exclusively present in E.coli are colored in orange whereas general recycling enzymes are depicted in purple. Enzymes represented in blue contribute to PGN biosynthesis. (adapted from (Gisin *et al.*, 2013))

Mutations or deletions of AmgK and/or MurU rendered the *Pseudomonas* strains more sensitive to fosfomycin (Gisin *et al.*, 2013, Borisova *et al.*, 2014), which brings the MurNAc-6P to UDP-MurNAc shortcut-pathway in a pharmacological focus of interest. Accounting for the recent rise of multidrug-resistant bacteria, fosfomycin, due to its low toxicity and low cross-resistance with other antibiotics, has regained significant therapeutic relevance especially in combination therapy with other antibiotics (Michalopoulos *et al.*, 2011). In this context x-ray structural analysis of MurU is of particular interest, since this enzyme is highly selective for MurNAc-α1P (Gisin *et al.*, 2013) and its crystal structure might thus help elucidating some key features for this selectivity. Eventually this might provide a rationale for the design of inhibitory drugs specifically targeting MurU and the above-described shortcut in cell wall recycling.

Moreover, even though some nucleotidyltransferases have already been structurally characterized, so far there is no structure available of a nucleotidyltransferase that specifically accepts MurNAc 1P as its sugar substrate (Singh *et al.*, 2012). This adds additional significance to elucidating the catalytic mechanism of MurU, in particularly with focus on the question about how its substrate specificity is determined on the molecular level. X-ray structural analysis of MurU provides one of the most suitable ways to address these issues.

2.2.2 Aim of the 'MurU-project'

The objective of this work was to solve the crystal structures of native MurU and MurU-ligand complexes at high resolution. A comparison of these structures with the structures of related transferases was supposed to provide first insights into mechanistic details and the narrow substrate specificity of the enzyme. Additionally the question of how MurU fits into the structural and physiological context of sugar-nucleotidyltransferases and which of its structural features it could serve as a rationale for inhibitory drug design was formative to this project.

2.3 Project Description YbbI (MurQ)

2.3.1 Cell wall recycling in Gram positive B.subtilis also involves a MurNAc-6P etherase

The MurNAc-6P etherase MurQ is a specific enzyme for the recycling of cell wall amino sugars and for their utilization in energy metabolism. MurQ-like MurNAc etherases are found mainly in Gram-positive bacteria (Jaeger & Mayer, 2008) supporting the notion that, in these organisms, cell wall recycling might as well be of physiological relevance. However in all key aspects the pathways and purposes of PGN recycling are less well understood in Gram-positive then in Gram-negative species. In 2010 Litzinger et al. presented the first evidence of a muropeptide rescue pathway in a Gram-positive organism (*B.subtilis*). This organism shared some orthologous enzymes with, but was nevertheless clearly distinct from cell wall recycling in *E.coli* and other Gram negative bacteria (Litzinger *et al.*, 2010). Litzinger et al. could show that this PGN recycling pathway in *B.subtilis* comprises a cluster of six genes, encoding an exo- β-N-acetylglucosaminidase (exo-GlcNAc'ase; YbbD/NagZ), a MurNAc-L-A-specific amidase (YbbE/AmiE), a MurNAc specific phosphotransferase system (YbbF/MurP), and a MurNAc specific transcriptional regulator as well as an ortholog of the *E.coli* MurNAc-6P etherase MurQ (YbbI/MurQ). A schematic representation of the sequential action of these enzymes in *B.subtilis* PGN recycling along with a short description of the respective catalyzed reaction is given in Fig. 7.

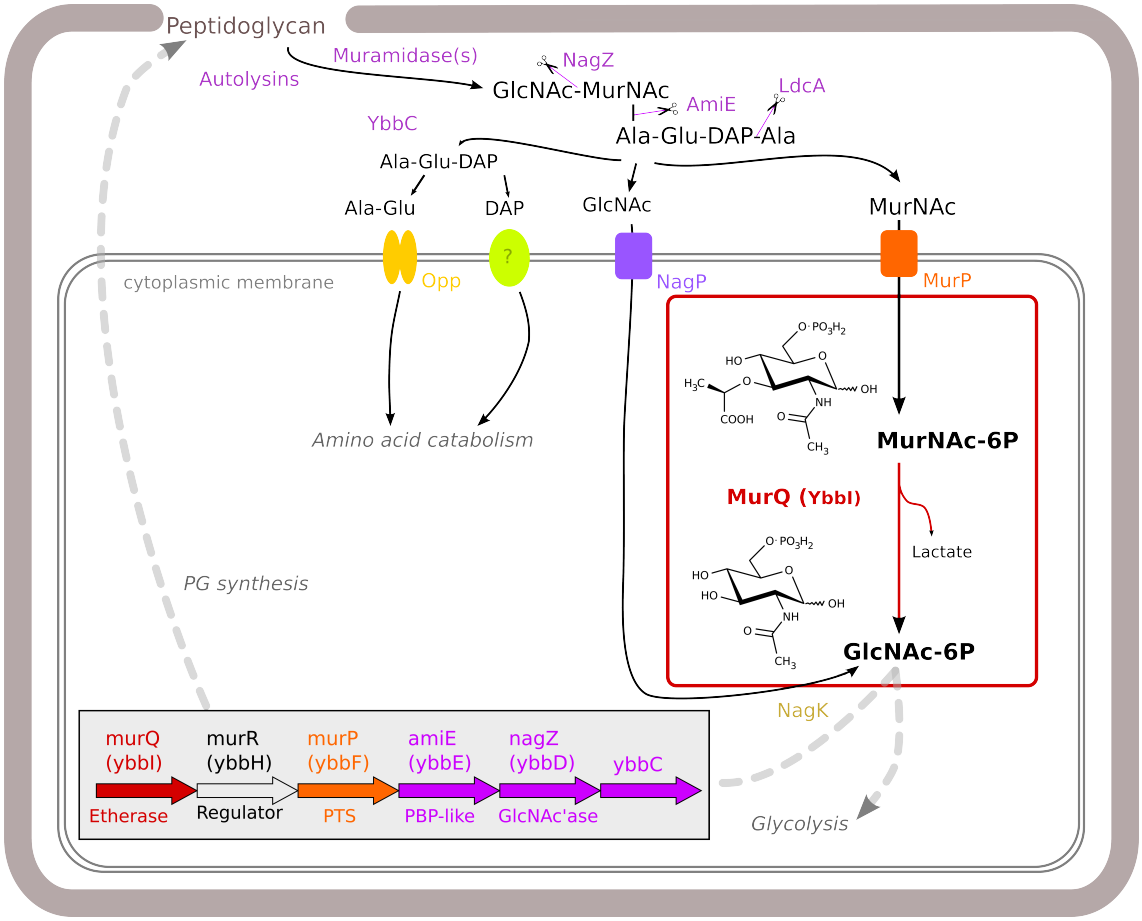


Fig. 7: Simplified scheme of PGN utilization and recycling in *B.subtilis* and organization of the corresponding gene cluster. Bacterial autolysins and muramidase(s) release approximately 50% of the endogenous cell wall in one generation during growth. Soluble muropeptides generated in this way are further processes by the exo-GlcNAc'ase NAgZ which noncovalently binds to peptidoglycan fragments upon their release from the cell wall and cleaves the 1,4-glycosidic bond between GlcNAc and MurNAc. Subsequently, the amidase AmiE cleaves the muramyl-L-Ala amide bond, releasing GlcNAc, MurNAc, and peptides. The carboxypeptidase LdcA, which is also present in the *B.subtilis* genome (Park & Uehara, 2008), cleaves off the terminal Ala from the resulting tetrapeptide, thereby generating the tripeptides Ala-Glu-DAP that might serve as a substrate for the yet poorly characterized putative peptidase YbbC. Recovery of MurNAc then proceeds like MurNAc dissimilation in E. coli; this process involves a MurNAc-specific phosphotransferase system (YbbF/MurP), by which MurNAc is taken up and simultaneously phosphorylated, yielding MurNAc-6-phosphate. In the cytoplasm, the etherase YbbI (MurQ) converts MurNAc-6-phosphate to GlcNAc-6- phosphate. YbbH (MurR) is a putative MurNAc-6-phosphate-specific transcriptional regulator, which is believed to function in a similar way as its ortholog in E.coli (see above).
The reaction catalyzed by YbbI/MurQ is highlighted (red frame) with the chemical structures of its product and educt depicted. The small inset (black frame) illustrates the organization of the PGN-recycling gene cluster in the *B.subtilis* genome.
PTS: phosphotransferase system; PBP-like: penicillin-binding protein like.

2.3.2 MurNAc etherases a subfamily of SIS domain proteins

In 2008 Mayer et al. generated a structural model for *E.coli* MurQ ($MurQ^{E.c.}$) (Jaeger & Mayer, 2008), which allowed the classification of $MurQ^{E.c}$ and other MurNAc etherases as a member of the SIS domain protein family (PF01380 according to www.pfam.xfam.org/family/PF01380 and (Bateman, 1999)). Structurally, SIS domains are characterized by a stretch of about 140 amino acids that adopts a characteristic α-β-α sandwich fold consisting of a five-stranded parallel β-sheet that is flanked on both sides by α-helices. Typically, the strands of the parallel β-sheet follow the order 2-1-3-4-5 and are connected via a loop-helix-loop motif (see Fig. 8). The abbreviation SIS domain stands for Sugar ISomarase domain and indicates the fact that these domains are wide-spread in sugar phosphate-binding proteins, such as isomerases, C-N lyases, and bacterial transcriptional regulators involved in the metabolism of sugar-phosphates. The above-mentioned MurR is an example of the latter case. However, in contrast to their common structural pattern, the level of sequence identity between SIS domains usually is rather low. Thus, the SIS domain should be considered a conserved structural rather than a conserved sequence domain (Jaeger & Mayer, 2008).

There exist two types of SIS domain proteins: double-SIS domain proteins that carry two SIS domains within one polypeptide chain and mono-SIS domain proteins hosting only one SIS domain per chain. From several crystal structures of SIS proteins (PDBids: 2VF5; 1TK9, 1X94, 1X9*), it is evident that the active biological molecules either form homodimers (mono-SIS domain proteins) or show a pseudo dyad symmetry that brings the two SIS domains in close spatial contact (double-SIS domain proteins). In both cases, the active site is usually located at the interface of two SIS domains (Jaeger & Mayer, 2008). MurQ-like etherases form a subfamily of mono-SIS domain proteins, which is characterized by additional helices at the N and C-termini that might be involved in dimer stabilization and/or catalysis (Jaeger & Mayer, 2008). Figure Fig. 8 schematically shows the topology of the SIS domain fold of MurQ-like etherases and illustrates how the physiologically important interface of two SIS domains is formed.

* 2VF5: Glcosamine-6-phosphate synthase from E.coli in complex with substrate, monomeric double-SIS domain protein; 1TK9: phosphoheptose isomerase 1 from Campylobacter jejuni, homo-tetramer of 4 mono-SIS domain protein chains; 1X94: phosphoheptose isomerase from Vibrio cholera, homo-dimer of 2 mono-SIS domain protein chains; 1X92: phosphoheptose isomerase from Pseudomonas aeruginosa, homo-tetramer of 4 mono-SIS domain protein chains

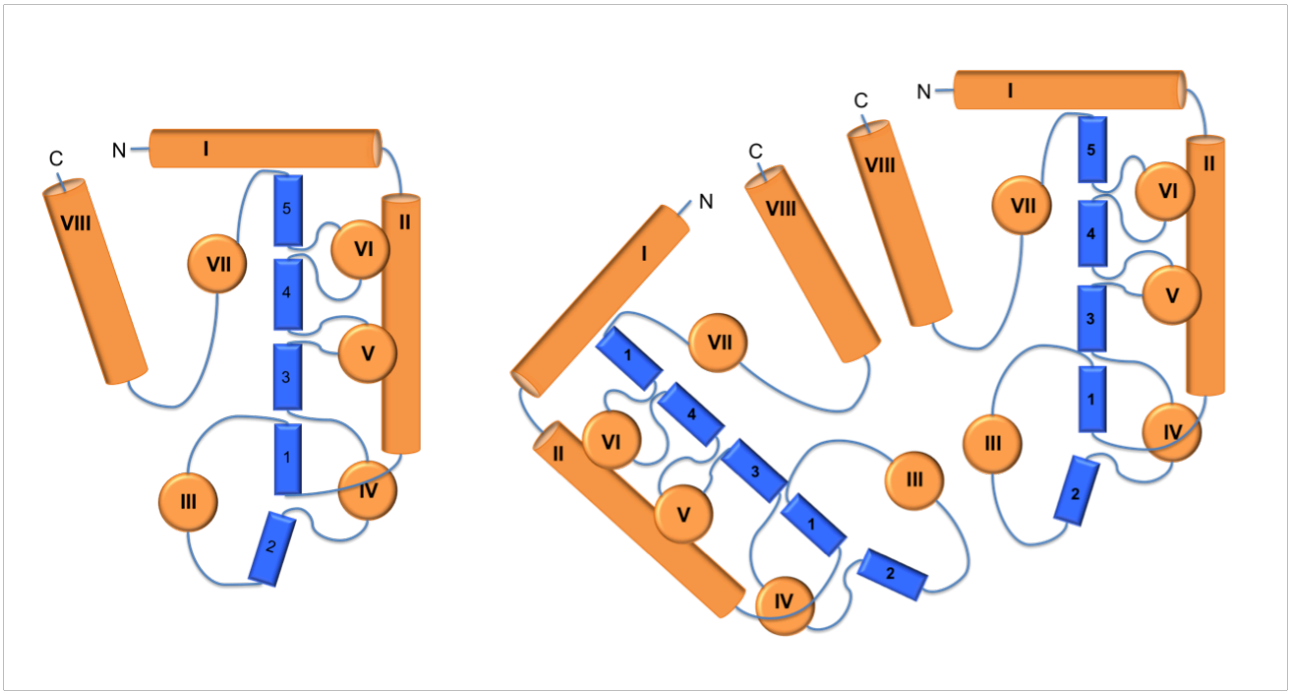


Fig. 8: Schematic representation of the SIS domain fold.
Depicted is the topology of the SIS domain fold as for MurQ-like etherases (right) and the dimerization mode of SIS domains making up the physiologically important domain interface (left) in both mono- and double-SIS domain proteins. The coloring reflects secondary structural elements: α-helices: orange; β-strands: blue; loops: gray. α-helices are numbered with roman and β-strands with arabic numbers.

2.3.3 Proposed reaction mechanism for MurNAc etherases

From a chemical point of view, enzymes that catalyze the scission of ether bonds are remarkable catalysts, since ether bonds (C-O-C) generally have a quite high bond energy which results in a considerable chemical inertness and a concomitant high capability of resistance to biodegradation (G. F. White *et al.*, 1996). The trivial name etherase classifies enzymes that catalyze the scission of such ether bonds exhibiting a wide variety of mechanisms: (1) oxygenative cleavage via monooxygenases; (2) oxidation of the carbon atom α-linked to the ether bond, followed by hydrolysis of the resulting ester; (3) hydroxyl shift mechanisms; (4) direct hydrolysis of the C-O bond; (5) anaerobic cleavage of methyl-aryl ethers ; (6) oxidative mechanisms ; (7) reductive mechanisms and, finally (8) carbon-oxygen lyase-mediated cleavage (G. F. White *et al.*, 1996).

The structural model of MurQ$^{E.c.}$ by Mayer et. al. (Jaeger & Mayer, 2008) is based on the crystal structure of a putative, at that time functionally uncharacterized, etherase from *Heamophilus influenzae* (PDBid: 1NRI) and reveals a significant structural similarity to the *E.coli* enzymes MurQ and GlmS (Glcosamine-6-phosphate synthase). This similarity allowed for the proposition of a reaction mechanism for MurQ-like etherases in analogy to the catalytic mechanism of GlmS, which is very well characterized (Teplyakov *et al.*, 1999). Accordingly, the cleavage of the ether bond by MurQ is

proposed to proceed in three steps (see Fig. 9). First, opening of the sugar ring of MurNAc-6P generates a C1 aldehyde, and consequently acidifies the hydrogen at C2 that is in α-position relative to the carbonyl group. This hydrogen is then removed by a general acid/base catalytic residue, such as a Glu-side chain (B^- in Fig. 9), generating a resonance stabilized enolate anion. The protonation of the alkoxy leaving group by an additional acidic residue (AH in Fig. 9) subsequently facilitates the β-elimination (deprotonation at C2, leaving group at C3) of lactic acid. The conjugate base A^- subsequently serves to deprotonate an incoming water molecule for the addition at C3 of the alkene intermediate (Δ-2,3-GlcNAc-6P), which is protonated at C2 by the conjugate general acid of the second catalytic residue (BH in Fig. 9). In a sequence mirroring the elimination of lactate, the resulting enolate is then converted to GlcNAc-6P, whereby the final ring closure yields both anomeres of the sugar (Jaeger & Mayer, 2008, Hadi *et al.*, 2008). For MurQ$^{E.c.}$ mutational studies identified residues E83 and E114 to be crucial for enzyme activity, which have thus been assigned the roles of B^-/BH and AH/A^- (Hadi *et al.*, 2008).

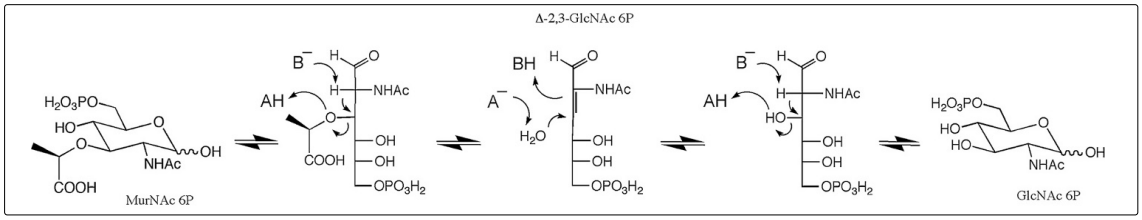


Fig. 9: Proposed reaction mechanism of MurQ-like MurNAc etherases.
Adapted from T.Jaeger and C.Mayer (Jaeger & Mayer, 2008). For a more detailed description please refer to the text.

Considering their proposed reaction mechanism, MurQ-like etherases can mechanistically be classified as C-O lyases (EC 4.2), which is a unique functionality among the members of the SIS domain protein family. Other proteins belonging to one of the other subfamilies of mono-SIS domain proteins mostly catalyze sugar isomerizations or carry an additional DNA-binding motif that endows them with the functionality of transcriptional regulators. Double-SIS domain proteins are also often involved in sugar isomerization reactions or function as C-N lyases (Jaeger & Mayer, 2008).

2.3.4 Aim of the 'MurQ-project'

When this project was started, the only structural model for MurQ-like etherases was the model of MurQ$^{E.c.}$ by Mayer et.al (Jaeger & Mayer, 2008) for Gram-negative *E.coli*. However, at that time, no experimental structural data were available for MurQ-like etherases. In particular for Gram-positive organisms, which have been shown to predominantly possess MurQ-like (Litzinger et al., 2010), no structural model existed at all. Consequently the experimental structure determination of MurQ from a Gram-positive organism – such as *B.subtilis*, which is often referred to as the Gram-positive model organism – was of obvious scientific interest. Moreover, apart from the interesting structural comparison of MurQ from Gram-positive and Gram-negative species, the crystal structure of MurQ was expected to provide objective experimental data to validate the proposed mechanism described above.

Consequently the determination of the crystal structure of MurQ from B.subtilis (MurQ$^{B.s.}$) at atomic resolution was the major target of this project.

Additionally, to address experimental validation of the postulated reaction mechanism, the determination of (active site) mutant structures as well as enzyme-substrate complex structures was planned. Unfortunately (from my point of view), in November 2013 the group of J. S. Blanchard published the structure of MurQ from *H.influenzae* (MurQ$^{H.i.}$), both natively and in complex with a transition-state analog (Hadi *et al.*, 2013). These structures sufficed for structural validation of the MurQ mechanism described above. Moreover, the overall structures of MurQ$^{B.s.}$ and MurQ$^{H.i.}$ were not significantly different (see below), which limits the insights gained from a structural comparison between MurQ-like etherases from Gram-positive and Gram-negative organisms. Nevertheless, the 'MurQ-project' still holds the potential to answer some interesting scientific questions.

3. Methodology of this work

3.1 Protein expression and purification

3.1.1 General methods

3.1.1.1 SDS PAGE

Sodium dodecylsulfate polyacrylamide gel-electrophoresis (SDS PAGE) was performed as described in (Schneck, 2008).

3.1.1.2 Analytical gel filtration

Analytical gel filtration was performed on a Äkta-Ettan system (GE healthcare®). For each run a test-sample of purified protein (25 µl of about 1.5 mg/ml) was analyzed on an analytical SD75 gel filtration column (GE healthcare®) using 30 mM Tris pH 7,6; 50 mM NaCl for MurU, or 30 mM Tris pH 7,7; 20 mM NaCl for MurQ (and MurQ point mutants) samples respectively, as gel filtration buffer at a flow rate of 0,05 ml/min.

3.1.2 MurU

3.1.2.1 Cloning of murU, protein expression and purification, Enzyme Assay

murU of *Pseudomonas putida* was cloned in vector pET29b (Novagen), heterologously overexpressed in *E. coli* as a C-terminally His_6-tag MurU-fusion protein, and purified by nickel affinity and gel filtration chromatography as described previously (Gisin *et al.*, 2013). The purified protein (yield c. 20 mg/l culture) was concentrated to 5 – 10 mg/ml and could be stored for several weeks at -20°C in gel filtration buffer (30 mM Tris-HCl pH 7.6, 50 mM NaCl) without significant loss of enzyme activity.

Based on its amino acid sequence the MurU-His_6 expression construct was analyzed and its molecular weight, theoretical isoelectric point (pI) and extinction coefficients at 280 nm were calculated using the ProtParam online tool on the ExPASY server (Gasteiger E.). The results of this analysis, along with the amino acid sequence of MurU-His_6 are attached in the appendix.

3.1.2.2 HPLC analysis of purified MurU:

High performance liquid chromatography (HPLC; Dionex BioLC) with a size-exclusion column (BioSep-SEC-S3000, 600x7, 800 mm, 5 m particle size; Phenomenex, Aschaffenburg) was used at a flow rate of 1 mg/ml (buffer: 20 mM Na_2HPO_4, 500 mM NaCl, pH 7.4) to assess the association state of native MurU.

Standard proteins used for gel filtration separation by HPLC were: albumin (molecular mass 66 kD), chymotrypsinogen (25.6 kD) and ribonuclease (13.7 kD).

3.1.2.3 Enzymatic preparation of MurNAc-α1P:

The enzymatic preparation of the MurU substrate MurNAc-α1P was conducted as described (Gisin *et al.*, 2013) and the concentration of MurNAc-α1P in preparations was determined by assaying the phosphate release with malachite green upon treatment with alkaline phosphatase according to previously established protocols (Baykov A.A. *et al.*, 1988).

3.1.2.4 Enzyme activity assay:

The malachite green assay also served to evaluate the enzymatic activity of purified MurU. No $MgCl_2$ had to be added to yield active enzyme, however, addition of EDTA (1 mM) led to complete loss of MurU activity, indicating the requirement of divalent cations, e.g. Mg^{2+}, which has been shown previously for other nucleotidyltransferases (Singh *et al.*, 2012).

3.1.2.5 Substrate specificity assays:

The nucleotidyl triphosphate specificity of MurU was analyzed using 5 µl of radiolabeled AmgK-reaction product (MurNAc-α1(^{32}P)-phosphate) in Tris-HCl buffer (pH 7,6) containing 0,25 U of baker's yeast inorganic pyrophosphate (Sigma-Aldrich) to drive the transferase reaction towards completion and 50 mM of either UTP, ATP, CTP, GTP, TTP nucleotidyl triphosphate (Sigma-Aldrich) in 25-µl reaction volumes at 37°C. The nucleotidyl transfer reactions were started in parallel by the addition of 1 µg of purified MurU and 5 µl of the reaction mixtures were spotted immediately and after 180 min of incubation on a TLC plate. Reaction products were separated in basic solvent with n-butyl alcohol/methanol/25% ammonium hydroxide/water (5:4:2:1). The radioactive products were detected using a Typhoon TRIO+ bimolecular imager (GE Healthcare). The ^{32}P-radiolabled MurNAc-α1P was prepared by adding 50 mM of MurNAc to a reaction mixture containing 100 mM Tris-HCl (pH 7,6), 100 mM ATP (pH 7,6), 10 mM $MgCl_2$ and γ-^{32}P-ATP (140 kBq) in a total volume of 100 µl. The reaction was started by the addition of 25 µg of purified AmgK and incubated for 3 hours at 25°C.

Cloning, expression, purification and HPLC analysis of MurU as well as the enzyme activity and specificity assays were performed by Jonathan Gisin, Alexander Schneider and Isabel Hinderberger (AG Mayer, IMIT Uni Tübingen).

3.1.3 MurQ

3.1.3.1 Cloning of murQ, and point-mutants

The *murQ* gene was amplified from *B.subtilis* 168 by PCR using chromosomal DNA and Phusion polymerase (Finnzymes). *murQ* was cloned into the *E. coli* expression plasmid pQE30 (Qiagen) generating a N-terminally His_6-tagged fusion construct (pQE30-murQ) for overexpression of His_6-MurQ that was amplified in *E. coli* DH5α. The mutagenesis polymerase incomplete primer extension method (M-PIPE; described by Klock & Lesley, 2009) was used to generate point mutations at positions E87 (E87Q), E118 (E118Q) and K237 (K237A) using pQE30-murQ as template. His_6-MurQ was overexpressed in *E. coli* BL21(DE3) which were grown in LB with ampicillin (100 µg/ml) at 37 °C. Expression was induced at an optical density at 600 nm of 0.8 by adding 1 mM of IPTG and cultivation was continued for 4 hours at 37 °C with shaking. Cells were harvested by centrifugation at 4400 g at 4 °C, resuspended in phosphate buffer (20 mM sodium phosphate, 500 mM sodium chloride, pH 7.4) and lysed using a French Press. Cell debris was removed by centrifugation at 37000 g and additional filtration through a syringe filter (Whatman). His_6-MurQ was isolated from the supernatant via Nickel affinity chromatography using a 1 ml HisTrap HP column (GE Healthcare) according to the manufacturer's instructions. The protein was eluted from the column with a linear gradient from 20 mM to 500 mM imidazole. Peak fractions were pooled and stored at 4 °C. After nickel affinity purification, MurQ peak fractions were further purified by gel filtration using a Superdex75 (SD75) gel filtration column (GE Healthcare). During gel filtration, a buffer exchange was performed from phosphate buffer to Tris buffer (30 mM Tris pH 7.7, 20 mM NaCl). MurQ point mutants were purified by the same method as MurQ.

3.1.3.2 Expression and purification of His_6-MurQ

Purified His_6-MurQ (MurQ from here on) was stored in 30 mM Tris pH 7,7; 20 mM NaCl at 4°C. For long term storage MurQ could be frozen at concentrations between 1 and 2 mg/ml without adding cryo-protecting additives.

As for MurU-His_6 the amino acid and atomic composition of the His_6-MurQ expression construct were analyzed based on its amino acid sequence and its molecular weight, theoretical pI and extinction coefficients at 280 nm were calculated using the ProtParam online tool on the ExPASY server (Gasteiger E.). See appendix.

Cloning, expression, and purification of MurQ and MurQ-mutants were performed by Amanda Duckworth (AG Mayer, IMIT Uni Tübingen).

3.2 Protein Crystallization

Diffraction quality crystals of both MurU and MurQ were grown using the sitting drop vapor diffusion method at room temperature. Crystallization trays were set up with the help of a tecan pipetting robot by mixing 300 nl of protein solution with 300 nl of a crystallization buffer containing varying additives.

Both MurU as well as MurQ (and MurQ point mutants) were crystallized as His-tagged fusion proteins. In the following, the expressions MurU-His_6 and MurU, or His_6-MurQ and MurQ respectively, will be used synonymously and interchangeably. For final quality assessment prior to crystallization, test-samples of the purified proteins were analyzed in an SDS-PAGE and analytical gel filtration (see above). Crystallization screens that yielded diffraction quality crystals are listed in the APPENDIX.

3.2.1 Crystallization of MurU

MurU crystallized in space group $P6_122$ with one monomer in the asymmetric unit and a solvent content of the crystals of around 50% (Matthews, 1968). For crystallization, concentrations of 7,5 – 10,5 mg/ml were required and well solutions contained 0,05 – 0,1 M buffer (MES pH 6 or 6,5; HEPES pH 7 or Tris pH 8,5) and 1,0 – 1,5 M $(NH_4)_2SO_4$.

For experimental phasing crystals were soaked with crystallization buffer containing 250 mM NH_4I for 10 min. Enzyme-substrate complexes were formed by soaking crystals with (i) 1.25 – 2.5 mM MurNAc-α1P and 3.0 – 3.5 mM of the β-γ non-hydrolysable UTP analog UppNHp for 75 min as well as with (ii) 1.25 – 2.5 mM MurNAc-α1P, 2.8 – 3.7 mM of the α-β non-hydrolysable UTP analog UpNHpp, and 20 mM $MgCl_2$ for 2 h. Crystals were transferred into a cryo-protecting solution containing 0.8 – 3.3 M $(NH_4)_2SO_4$ as well as 5-25 % glycerol and flash frozen in liquid nitrogen prior to data collection.

3.2.2 Crystallization of MurQ

Diffraction quality crystals of MurQ were grown at concentrations of 6,5 – 9,0 mg/ml with crystallization buffers consisting of 0.1 M citrate pH 4,5 – 5,5; 12 – 18% PEG 3350 and in some cases 10 mM Na_2HPO_4. All measured crystals belonged to the space group $P3_221$ (Space group no. 154) with 3 protomers in the asymmetric unit and an approximate solvent content of 52%. Enzyme-substrate complexes were formed by soaking native MurQ crystals with (i) 100 mM MurNAc, (ii) 100 mM MurNAc + 10 mM $MgCl_2$ or (iii) 100 mM GlcNAc-6P respectively for one hour. For data collection crystals were transferred to a cryo-protecting solution containing 28 % PEG 200 and flash frozen in liquid nitrogen. Soaking experiments with lower substrate concentrations and different soaking times as well as co-crystallization approaches were also conducted but were not successful. Crystals either dramatically lost diffraction power or no ligands was bound at any active site.

3.3 X-ray crystallography

The intention of the following chapter 3.3 is not only to summarize the experimental procedures applied in this work but to additionally provide the non-expert reader with the basic knowledge on X-ray crystallography that is needed for a profound understanding and critical evaluation of the results presented in this work. The major part of the chapter (sections 3.3.1 to 3.3.12) will therefore be a brief theoretical explanation of X-ray crystallography as a method, with the focus on some of its major challenges that have explicitly been relevant to the determination of the structures of MurQ and MurU. The last two sections will then give a detailed description of the experimental methods and computational programs that have been used to solve these structures.

This theoretical chapter is written significantly more detailed than the introduction into the biological background of the projects and their scientific integration into other fields of research, which would be inopportune for presenting this work in a scientific journal or to funding executives. However, since the genuine motivation for this PhD-project was the application of X-ray crystallography on a viable scientific problem, this emphasis seems appropriate in the context of this dissertation.

3.3.1 Overview

X-ray crystallography is not a direct imaging technique that focuses visible light scattered from objects through refractive lenses to create a magnified image of the object. Rather, it exploits the fact that X-rays with wavelengths between 0,05 and 5,0 nm (0,5 – 50 Å) are scattered by the electron shells of atoms and thus provide the possibility to obtain structural information of molecules at near-atomic resolution. However, due to the refractive index of X-rays in different materials, which is essentially equal and close to unity, it is not possible to obtain direct atomic resolution images of a single protein molecule (or other macromolecules) through simple focusing of scattered X-rays (Rupp, 2010, Sumner, 2014). Nevertheless, diffraction images can be obtained from protein crystals in X-ray diffraction experiments. These images carry information about the content of the crystal's unit cell (the protein of interest), but this information is encoded in intensity distributions of reflections in "reciprocal space" and thus not easily accessible. With the help of Fourier transformations, this information can be "translated" back into molecular, "real space", giving rise to an image of the crystallized molecule. The Fourier transformation is a straightforward mathematical operation that requires two terms as Fourier coefficients: (i) the structure factor amplitudes, which can be obtained from the measured and corrected diffraction spot intensities, and (ii) the relative phase angle corresponding to each observed diffraction spot (Rupp, 2010). Since these phase angles are not directly accessible by experimental methods they must be obtained in so-called phasing approaches, which can involve additional experiments or molecular replacement calculations with the help of the phases from related known

structures. This is generally known as the “phase problem” in crystallography, and it is one reason why X-ray crystallographic structure determinations remain challenging even today. Once initial phases are determined, an initial electron density map can be calculated that provides the basis for molecular model building and structural refinement. Fig. 10 illustrates the x-ray crystallographic workflow in a schematic diagram. In the following chapter the principles of X-ray crystallography will be explained in more detail.

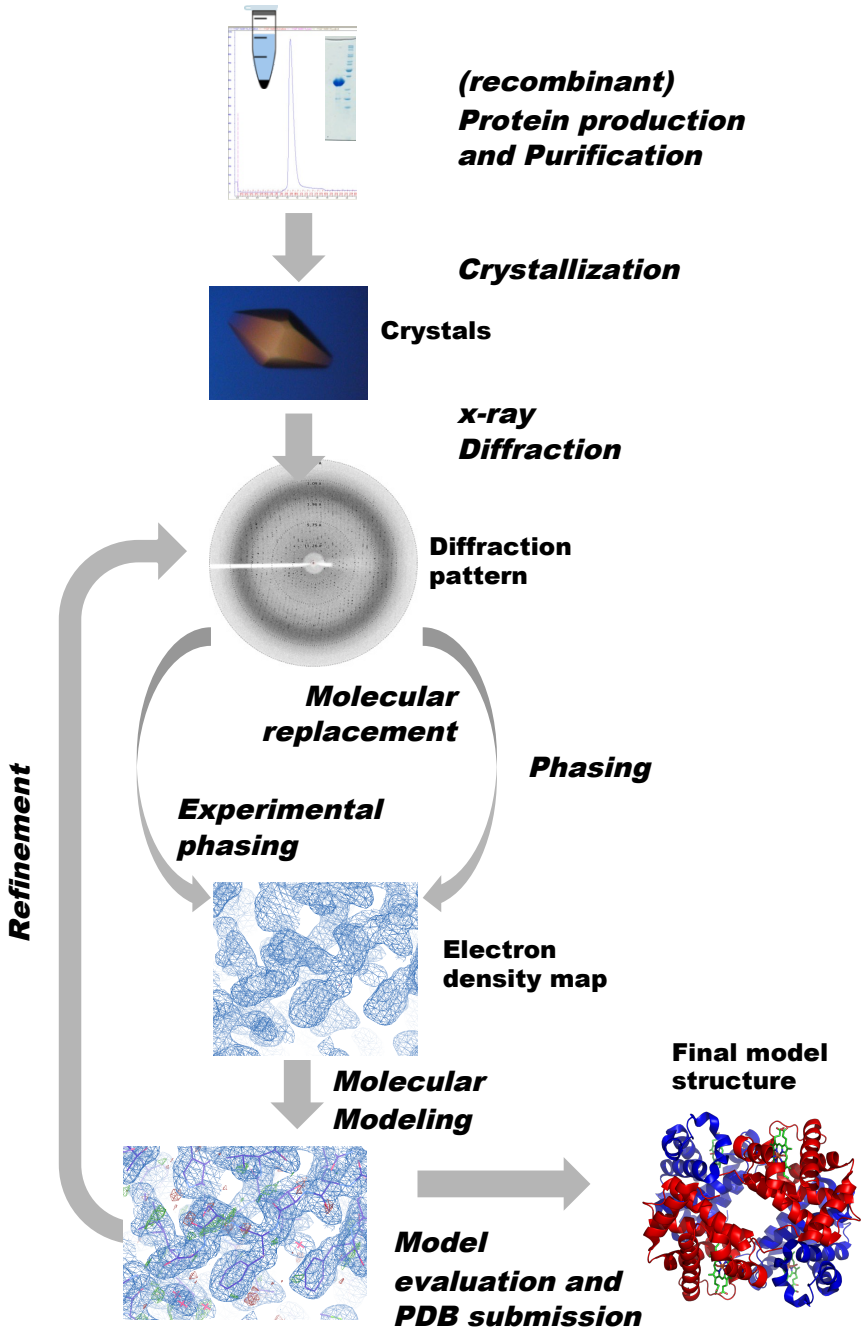


Fig. 10: X-ray crystallographic workflow
Schematic diagram showing the workflow for macromolecular structure determination by X-ray crystallography

3.3.2 Crystal geometry and space groups

Crystals are composed of identical units, the so-called unit cells, that are repetitive in all directions and form a continuous lattice. A crystal's unit cell is characterized by the distance vectors a, b and c defining its spatial extensions and the angles α, β and γ spanned by these vectors. The vectors a, b, c together with the angles α, β, γ, make up the so called Bravais lattice of a unit cell, which, based on its symmetry, can be characterized as either cubic, tetragonal, trigonal/hexagonal, orthorhombic, monoclinic or triclinic. Bravais lattices are further sub-classified as primitive- or non-primitive, whereby primitive lattices contain only one lattice point per unit cell (1/8th of a point at each corner point) and non-primitive lattices contain additional lattice points at either the delineating planes of the unit cell (face-centered lattices) or at its center (body-centered lattices). These options give rise to a total of 14 Bravais lattices, as visualized in Fig. 11 (Rupp, 2010).

A crystal lattice is made up by the spatially repetitive arrangement of its unit cells containing the crystallographic motif, which in turn is formed by the protein of interest. Typically a crystallographic motif consists not only of a single protein molecule but often of several copies, which are linked by crystallographic symmetry operations. In this context, from the generally possible symmetry operations, translation, rotation, inversion and mirrors, inversions and mirrors cannot occur for chiral molecules such as proteins. Moreover, only certain symmetry operations are compatible with the requirement for translational symmetry of the crystal lattice that is imposed by the need to form a continuous lattice. On the other hand, so-called screw axes can often be found in protein crystals, which can be described as the combination of two symmetry operations, rotation and translation. The rotational symmetry operators describing a given unit cell define the crystals point group, whereas the combination of Bravais lattices and symmetry operations (including screw axes) define the crystal's space group. For proteins, a total of 72 space groups exist, which define all possible ways in which a protein molecule can be packed to form a continuous three-dimensional lattice. The smallest unit that gives rise to the crystal through application of the crystallographic symmetry operations is called the asymmetric unit (asu). Often the asu contains not only one but several molecules that are symmetrically linked. Since this symmetry between protein molecules is not necessarily featured in the crystal packing it is referred to as non-crystallographic symmetry (NCS) (Rupp, 2010).

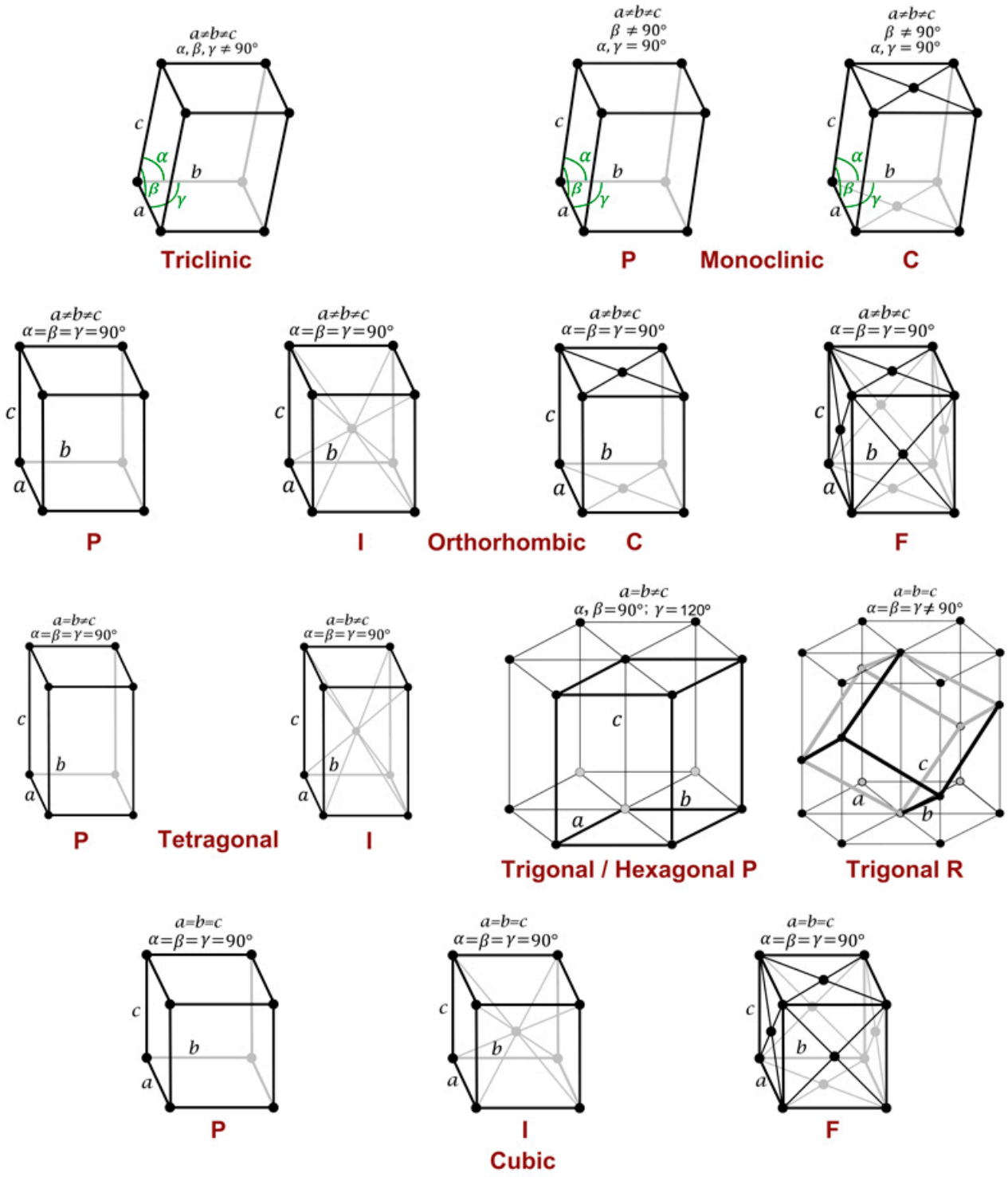


Fig. 11: The 14 Bravais lattices:

The Symbols P C I F R refer to the different lattice types:

P = primitive (there is only one reticular point inside the cell (⅛ of a point in each of the 8 corners of the unit cell)

C = centered in the faces perpendicular to the cell c axis

I = centered in the body of the cell

F = centered in all faces of the cell

R = primitive, identical cell axes and cell angles, or hexagonal two times body centered

adapted from: http://www.xtal.iqfr.csic.es/Cristalografia/parte_03_4-en.html

3.3.3 X-ray scattering

X-rays with wavelengths between 0,05 and 5,0 nm (0,5 – 50 Å) are scattered by the electron shells of atoms and thus provide the possibility to obtain structural information of molecules at near-atomic resolution. For (protein) crystallography, monochromatic X-rays with wavelengths of 1,0-1,7 Å are most frequently used, which coincides with the bond-lengths found in biomolecules allowing for molecules packed in a crystal to serve as a diffraction lattices. X-ray scattering, which is the basis of X-ray diffraction, is a discrete event that arises from the interaction of the electric field vector of X-ray photons with electrons of matter. However, x-ray scattering by electrons is weak and only a small fraction of incident radiation gets scattered. In the event of inelastic scattering, which is the general basis of diffraction and the prevailing mode of interaction for X-rays with matter, the energy of the scattered photon remains unchanged and a fixed phase relation is maintained, only the travelling direction of the photon changes. Generally the scattering process can be conceptualized as a two-step model: Firstly, the energy of the photon is absorbed by an atom thereby inducing oscillations to its electrons. Secondly, as these oscillating electrons represent an accelerated charge, they in turn emit electromagnetic waves (photons) that have the same wavelength as the incoming photon but that recombine to a resulting wave (or photon). In contrast to the wavelength, the spatial direction of this resulting scattered X-ray photon is not predetermined by the incoming X-ray photon. However, the probability of the photon being observed in a specific direction is not equal for all directions and the amplitude of the resulting wave in this direction is proportional to this probability. It should be emphasized at this point that these steps do not actually happen in a sequential and distinct manner but are to be understood as an explanation model. Nevertheless, the probability of a scattered photon to emerge from an atom in a given direction can be described mathematically by a cosine-dependent scattering function. The integrated scattering function yields the total scattering power of the electrons of an atom, defined as the atomic scattering factor. For scattering of electrons in atoms, the scattering function is further dependent on the location of the electrons in the atoms. The probability of a certain electron being at position r, represents its electron density ($\rho(\mathbf{r})$, known as atomic orbitals from quantum chemistry) and is given by the square of the wave function of the electron. In general, the approximation of a spherical electron density is sufficient, resulting in a spherical atomic scattering function (Rupp, 2010). Fig. 12 illustrates the scattering of X-rays by a single atom.

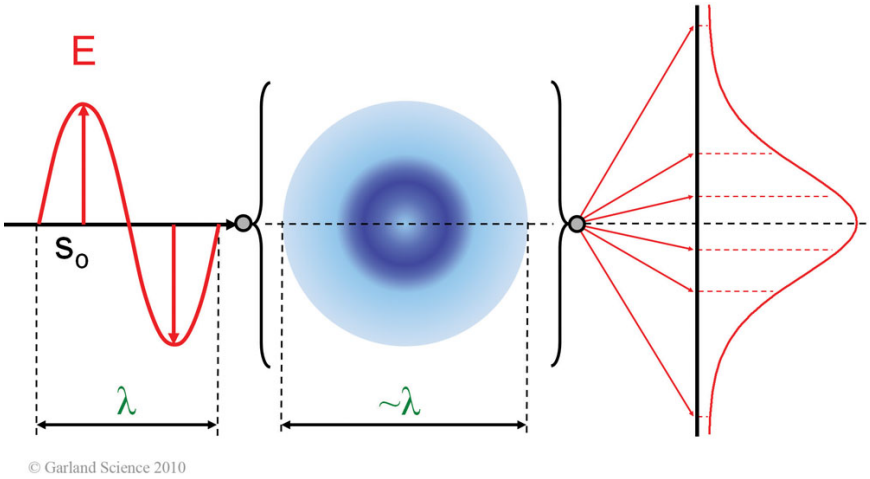


Fig. 12: Scattering of X-rays by a single atom.
An X-ray with direction S_0 is scattered by an atom with spherical, centro-symmetric electron distribution, producing a scattered wave. The scattered wave, shown on the right, is described by the atomic scattering function and is proportional to the square root of the intensity profile of a spot recorded a X-ray detector. Adapted from Biomolecular Crystallography by Bernhard Rupp, (Rupp, 2010).

3.3.3.1 The atomic scattering factor

Considering the net electromagnetic wave resulting from atomic x-ray scattering, which is the emitted photon, it is important to take the particle-wave dualism of photons into account. In this context, X-ray radiation can be conceptualized as photons travelling in wave-packets. This means that, within the coherence length of a given X-ray source (several micrometers for synchrotrons), all electrons will be induced to oscillate coherently, giving them a fixed phase relationship, and, in consequence, a fixed phase relation between all scattered waves. This relation represents the phase difference $\Delta\varphi$ of two waves scattered by two electrons and is given by

$$\Delta\varphi = 2\pi\vec{S}\cdot\vec{r}$$

Equation 1

with $\vec{S}$ being the scattering vector and $\vec{r}$ the distance vector between the electrons the scattered waves originate from. To obtain the wave F_S scattered by an atom, all the waves of the form $exp(i\varphi)$ originating from the volume element $\rho(r)$ within the volume of the atom (V_{atom}) are integrated, yielding

$$F_S = \int_r^{V_{atom}} \rho(\vec{r})\,\exp\left(2\pi i\vec{S}\vec{r}\right)\,d\vec{r}.$$

Equation 2

In this equation F_S has the form of a Fourier transform and describes the scattering of the entire atom as a function of the scattering direction and thus of the scattering angle. Consequently F_S corresponds to the Fourier transform (FT) of the atom's electron density (or its electron orbital) and is therefore called the atomic scattering factor, or atom form factor, which is annotated as f_s (Rupp, 2010).

$$F_S = f_S = \mathrm{FT}\,[\rho(\vec{r})]$$

Equation 3

3.3.4 Diffraction patterns and Structure factors

3.3.4.1 Crystal diffraction patterns

The above-described model explains x-ray scattering at a single atom. When scattering from more than one atom is considered, the resulting scattering function becomes more complex as it has to account for interference of the scattered waves originating from several individual atoms. For a two-atom molecule the resulting wave shows characteristic maxima and can still be nicely visualized (Fig. 13).

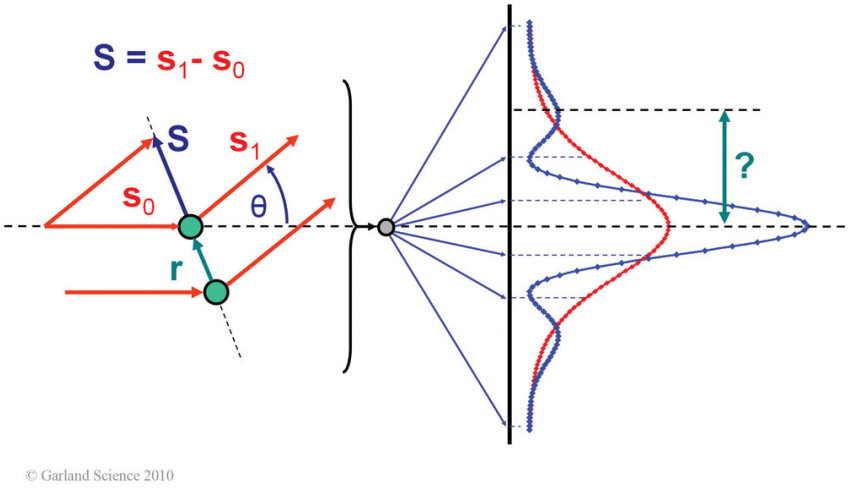


Fig. 13: Scattering by a two-atom molecule.
The scattered wave has a characteristic intensity profile with maxima spaced by a regular distance (indicated by the green arrow). S: scattering vector, r: atomic distance, θ: scattering angle. Adapted from Biomolecular Crystallography by Bernhard Rupp (Rupp, 2010)

In case of a more complex molecule, the overall molecular scattering function is irregularly modulated by the interference of photons scattered by neighboring atoms. But still, as deduced for single-atom scattering, the resulting total scattered wave F_S is the Fourier transform (*FT*) of the electron density. However, in this case, the electron density of the whole molecule has to be considered and its *FT* has thus to be expressed as a summation over all atomic scattering factors $f_{s,j}$ within the molecule.

$$F_S = \sum_{j=1}^{atoms} f_{s,j} \exp\left(2\pi i \vec{S}\vec{r}_j\right)$$

Equation 4

For a continuous electron density this sum becomes an integral and can be expressed as:

$$F_s = \int_r^{V_{molecule}} \rho(r) \exp\{(2\pi i \vec{S}\vec{r}_j)\} \, \mathrm{dr} = \mathrm{FT}\,[\rho(r)]$$

Equation 5

However, X-ray scattering of single molecules is much too weak for detection, which necessitates the regular arrangement of many molecules in a crystal lattice to get a diffraction signal. In a lattice, a scattered wave originates from each reticular point, and the waves scattered from individual molecules sitting at these points will interfere. This again modulates the resulting wave. The waves scattered by molecules that are spaced by an integer number of unit cells along one lattice axis have the same phase and their intensities will be increased as a result of constructive interference. Moreover, the maxima of the intensities will be sharpened by destructive interference of any waves that are not exactly in phase. Hence, through the periodic alignment of many molecules, continuous scattering is modulated by interference in a characteristic manner resulting in a discrete diffraction pattern with regularly spaced maxima. These maxima can be detected as reflection spots on an x-ray detector and their positions on the diffraction image can be described via the coordinate indices *h*, *k*, and *l*. From a mathematical point of view, the *FT* of a given function in real space represents the corresponding function in the so-called reciprocal space – a mathematical auxiliary tool. Accordingly, the *FT* of a real space lattice results in another lattice, yet in reciprocal space and hence the lattice function of a crystal defines its reciprocal lattice function. Considering this, the diffraction pattern of a protein crystal can be considered as a convolution of the molecular scattering function with its reciprocal lattice function (see Fig. 14). Consequently, the crystal's diffraction pattern can be formally described as the reciprocal space-representation of the electron density present in the (real space) crystal. In other words, the diffraction pattern of a given crystal samples the structure factors of the molecule at discrete spots the positions (*h,k,l*) of which are solely defined by the geometric parameters of the crystal lattice. The intensity of each spot, however, is dependent on the electron density present in the unit cell (Rupp, 2010).

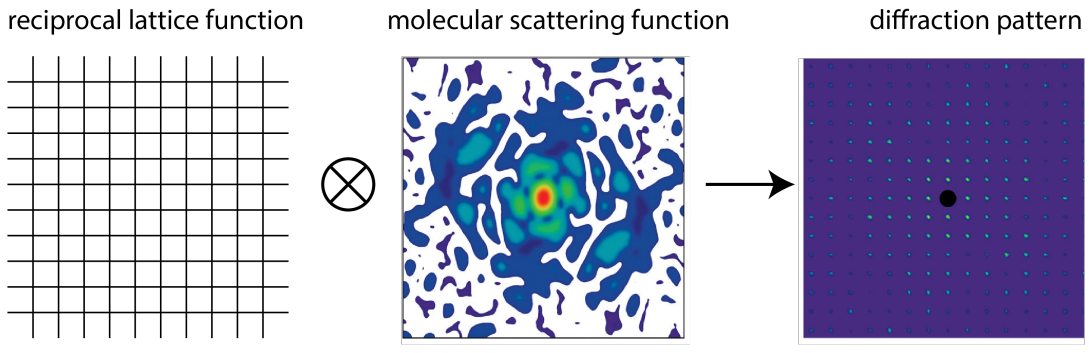


Fig. 14: Diffraction patterns of molecules arranges in a regular lattice
The diffraction pattern (right) can be described by the convolution of the reciprocal crystal lattice function (left) with the molecular scattering function (center). Adapted from Biomolecular Crystallography by Bernhard Rupp (Rupp, 2010) and M.Vorländer (Master-thesis, 2015)

3.3.4.2 The complex structure factor

The total scattering of a crystal in a given reciprocal lattice direction is proportional to the sum of all scattering contribution is the unit cell. Each atom j in the unit cell contributes a partial wave to every reflection h. The structure factor F_h* of spot h can thus be written as the summation of all partial waves of j atoms with the atomic scattering factor f_j at position x_j:

$$F_h = \sum_{j=1}^{atoms} f_{s,j}^0 \exp\left(2\pi i \vec{h}\vec{x}_j\right)$$

Equation 6

The exponential term of the structure factor equation describes the contribution of a partial wave scattered at atom *j.* It contains the relative phase angle of the partial wave, which depends solely on the direction of scattering, $\vec{h}$, and on the position of the atom j relative to the origin, given by the fractional coordinate x_j.

Equation 5 describes the atomic scattering factors as a reciprocal space representation (Fourier transformation) of the electron density of a single molecule. The same correspondence exists between the electron density of an entire crystal unit cell and its scattering function, the complex structure factor F_S. Here the integration extends over the entire unit cell volume:

$$F_S = \int_r^{V_{cell}} \rho(r) \exp\{(2\pi i S r)\}\, \mathrm{dr} = \mathrm{FT}\,[\rho(r)]$$

Equation 7

This equation (Equation 7) is fundamental to X-ray crystallography, since it constitutes the basis for the reconstruction of the electron density of a molecule in a crystal from the crystal's X-ray diffraction pattern. Accordingly the scattering function (or the complex structure factor F_S) represents the Fourier transform of the electron density present in a crystal's unit cell, which can be calculated as a summation over all structure factors F_h* sampled in all reflections with the reciprocal lattice index vectors $\vec{h}_i$ and the reciprocal lattice coordinates h,k,l.

$$\rho(\vec{r}) = 1/V_{unit\ cell} \sum_{\vec{h}} F\left(\vec{h}\right) \exp\left(-2\pi i \vec{h}\vec{r}\right) = \rho(xyz)$$
$$= \sum_h \sum_k \sum_l F(hkl) \exp\left(-2\pi i (hx + ky + lz)\right)$$

Equation 8

*The structure factor represents a wave. It is a common mathematical practice to express waves as complex numbers, since superposition of waves becomes very convenient using the vector representation of waves in the complex plane. See also below.

3.3.5 Diffraction conditions

In practice x-ray diffraction patterns of protein crystals are rather complex. Depending on the interference of the waves scattered by the molecules at the grid points of the diffraction lattice, there might be reflections missing or displaying attenuated intensity in the resulting diffraction image. Therefore the interpretation of diffraction images requires some theoretical background knowledge.

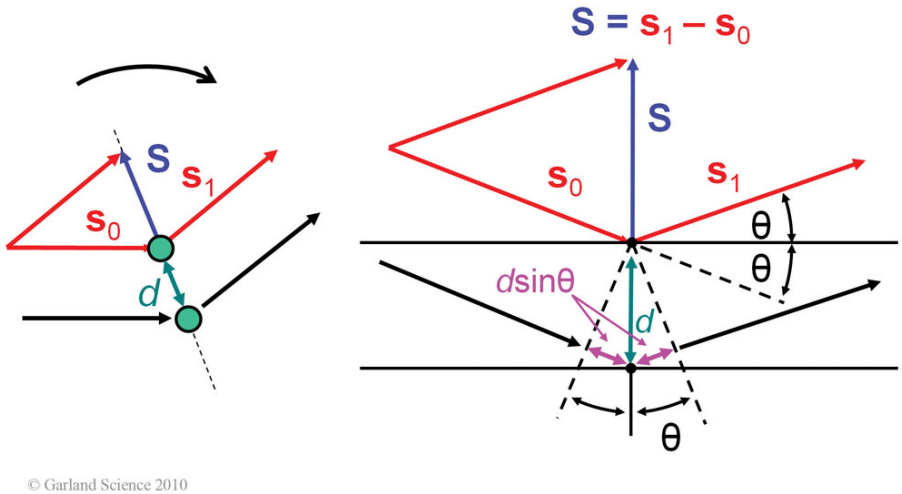
3.3.5.1 Bragg Equation

The graphical interpretation of the scattering diagram of two atoms (see Fig. 13) as the reflection of X-rays on a set of planes in a crystal to which the scattering vector S is normal has been introduced by Sir Willliam Bragg. This theoretical construct is also valid for large macromolecules and drastically simplifies the interpretation of X-ray diffraction patterns. It reveals the relation between the scattering angle θ and the interplanar distance d_{hkl} for a given set of reflecting lattices *hkl,* which can be directly deduced from the graphical interpretation of x-ray scattering depicted in Fig. 15 using trigonometric laws:

$$n\lambda = 2d \sin\theta \quad \text{or} \quad \frac{1}{d_{hkl}} = \frac{2\sin\theta}{n\lambda}$$

Equation 9

Equation 9- also known as the Bragg equation – states the fundamental diffraction condition. It describes that the total path difference $2d\ sin\theta$ between two scattered partial waves must equal a multiple of $n\lambda$ for maximum constructive interference (where n is a natural integer) (Rupp, 2010).



© Garland Science 2010

Fig. 15: Interpretation of X-ray diffraction as reflection on a lattice plane.
X-rays scattered by a set of planes with the spacing *d* interfere constructively when the difference in path length between the scattered waves - *2d sin θ* - is an integer multiple of the x-ray wavelength λ. This requirement allows for the graphical derivation of the Bragg Equation. S: scattering vector, θ: scattering angle. Adapted from Biomolecular Crystallography by Bernhard Rupp (Rupp, 2010).

In this context, the values *h,k,l* define the hypothetical planes of the Bragg construction (Fig. 15). They correspond to the inverse of the fractional intercept with the unit cell axis, multiplied by a common factor to give integer numbers and are known as the Miller indices of the plane. Any reflection that originates from diffraction at a plane *h,k,l* is also denoted with the Miller indices of the respective plane (Rupp, 2010).

Moreover, it gets obvious from Fig. 15 that, at a given wavelength, the scattered waves will cancel each other out due to destructive interference for certain values of interplanar distance d_{hkl}. For a crystal diffraction pattern this means that certain reflections will be missing depending on the crystal's space group, and these systematic absences of reflections are very helpful in spacegroup determination.

3.3.5.2 Ewald sphere

Another way of visualizing diffraction conditions is the Ewald construction. This model is based on the same geometric considerations as the Bragg equation and relates the real space crystal lattice, defined by d_{hkl}, to its counterpart in reciprocal space that is characterized by d^*_{hkl}.

$$d^*_{h,k,l} = \frac{1}{d_{h,k,l}}$$

Equation 10

The reflection condition of defined by the Bragg equation can thus be also expressed as:

$$\frac{1}{d_{hkl}} = \frac{2\sin\theta}{n\lambda} = d^*_{hkl}$$

Equation 11

Graphically, in the Ewald construction for every set of planes with the miller indices *h,k,l* a reciprocal lattice point *h,k,l* is constructed by spanning a vector with the length $d^*_{h,k,l}$ from the origin of the reciprocal lattice to the point *h,k,l*. The length of the vector equals the inverse plane spacing and it is normal to the set of planes from which it originates (Rupp, 2010). From Equation 11 it is obvious, that the reflection condition at a given wavelength λ is fulfilled, for any reciprocal lattices with the spacing *2d sinθ/n.* For a graphical interpretation this means that a reciprocal lattice point h, k, l is in diffracting condition when it intersects with a sphere with radius 1/λ around the origin of the real space lattice (see also Fig. 16).

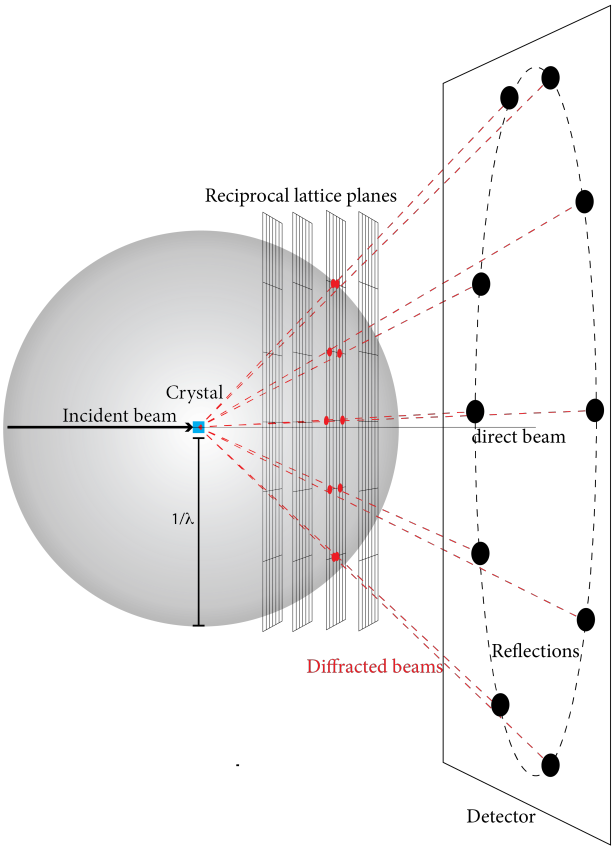


Fig. 16: The Ewald sphere construction
For one selected reciprocal lattice the points that intersect with the Ewald sphere and the corresponding diffraction image are indicated by red and black spots respectively. Reflections from the other reciprocal lattices are omitted for clarity. Courtesy of M. Vorländer (master thesis 2015).

The geometric construction depicted in Fig. 16 is also known as the Ewald sphere. It describes the reflection condition in reciprocal space and relates it to the underlying real space crystal lattice. Thus, the Ewald construction is very helpful for the interpretation of diffraction images particularly with respect to space group assignment. Moreover it nicely illustrates the considerations mandatory for successful crystallographic data collection: In order to reconstruct the electron density of a molecule from its diffraction pattern, it is necessary to measure each (unique) reflection at least once. The Ewald construction nicely shows that, for a given wavelength and at a given crystal orientation, this cannot be achieved at a time. Therefore, either the wavelength or the crystal orientation needs to be continuously altered to bring every reciprocal lattice point into reflection condition. In practice the latter option is applied most frequently by simply rotating the crystal or through the beam and thereby moving each reciprocal lattice point through the Ewald sphere (Rupp, 2010). See also section 3.3.9.

3.3.6 The phase problem

As illustrated above, the complex structure factor is the Fourier transform of the electron density present in the unit cell of a given protein crystal. Theoretically, this electron density can thus be reconstructed from the diffraction pattern of the crystal, which samples the structure factors of the crystallized molecule in its spot intensities. In practice however, this is not a straightforward process due to the so-called 'phase problem in X-ray crystallography', which will be described in the following section:

It is a common mathematical practice to express waves as complex numbers as superposition of waves becomes very convenient using the vector representation of waves in the complex plane. Since structure factors represent waves they can thus be also described as complex numbers (note the *i* in the structure factor equations Equation 4 and Equation 6).

The representation of the electron density present in a crystal unit cell described by Equation 8, which is a summation of complex numbers, can also be expressed as in dependency of the phase angle φ of the scattered wave and the structure factor amplitude $|F_h|$:

$$\rho(r) = \sum_{h=-\infty}^{+\infty} |F_h| \exp[2\pi i(hr) + i\varphi(h)]$$

Equation 12

To reconstruct the electron density of a molecule from its crystal diffraction pattern Equation 12, which represents the molecules electron density of the crystal unit cell in reciprocal space, needs to be transformed back into real space.

It is important to note that in Equation 12 the phase angle φ and the structure factor amplitude $|F_h|$ are separate factors. The values for $|F_h|$ can experimentally addressed, since they correspond to the square root of intensity I_{hkl} of a given reflection hkl (Equation 13), which is what we measure in a X-ray diffraction experiment.

$$I_{h,k,l} = \left|F_{h,k,l}\right|^2$$

Equation 13

However, information on the phase angles φ, which are also required for the back transformation, is completely missing on the diffraction images collected by an X-ray detector. In other words: In an X-ray diffraction experiment the structure factor amplitudes can be obtained by quantifying the intensities of the diffraction spots, whereas the phase angles cannot be directly measured and thus there is essential information lacking for the reconstruction of electron density. This dilemma, resulting from 'missing phases' is therefore is called the 'phase problem' (Rupp, 2010). Fig. 17 graphically illustrates this central issue in X-ray crystallography.

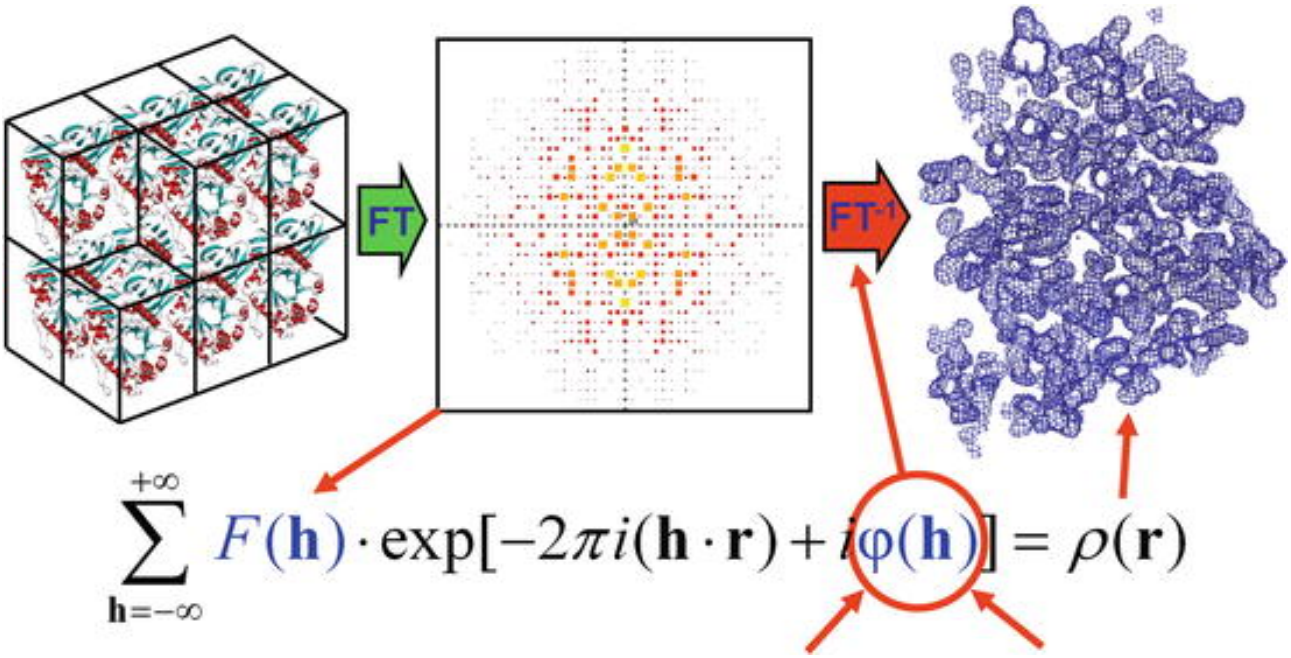


Fig. 17: The crystallographic phase problem
The measurable component of the Fourier transform of the crystal is only the scalar structure factor amplitude *F(h)*, which is proportional to the square roots of the reflection intensities *I(h)*. The missing phase angels *φ(h)* must be supplied by additional phasing experiments or in the form of model phases via molecular replacement (see below). The two necessary Fourier coefficients in the back-transformation formula are emphasized in blue. Adapted from Biomolecular Crystallography by Bernhard Rupp (Rupp, 2010).

3.3.7 Crystallographic phasing methods

For the reconstruction of the electron density of an unknown protein structure from its crystal diffraction pattern, the phase problem needs to be solved, since a phase value for each measured reflection is required. The available methods to obtain phase values can be roughly groups into four general areas:

- Marker atom substructure methods
- Density modification
- Molecular replacement
- Direct methods

3.3.7.1 Phases from marker atom substructures

In marker atom substructure methods, several marker atoms provide the source of electronic differences relative to an isomorphous reference structure. This electronic difference in turn implies different atomic scattering factors and thus different structure factor amplitudes and different intensities relative to the reference structure. The difference data are then used to determine the location of the source of the electronic difference, the so-called marker atom. This reduces the phasing problem to solving a substructure of only a few up to a few hundred atoms, compared to the many thousands of the whole macromolecule, which now can be addressed by direct methods or by Patterson search methods (see below).

The difference principle is generally applicable regardless of what underlying electronic difference gives rise to the differences in atomic scattering factors. In case of the above-described concept of isomorphous difference data, the electronic differences usually result from additional heavy atoms soaked into the crystal or from marker atoms replacing certain protein atom (e.g. a Se introduced as selenomethionine). Depending on the strategy applied to break phase ambiguity these techniques are referred to as MIRAS (multiple isomorphous replacement with anomalous scattering) or SIRAS (multiple isomorphous replacement with anomalous scattering) methods. In the former case several anomalous data sets need to be recorded, whereas in the latter case density modification techniques (see below) are additionally applied and thus only one anomalous dataset is sufficient (for more information see (Rupp, 2010)). Furthermore intensity differences can also be a result of scattering vector contributions from anomalously scattering atoms that are natively present in the crystal (e.g. S-atoms) and in the last 15 years it has become increasingly popular to also exploit the anomalous signals of halides such as bromides or iodides that have been soaked into the crystal.

Regarding scattering vector differences anomalous and dispersive contributions to the atomic scattering vector need to be distinguished. Anomalous scatterers provide electronic differences, the so-called Bijvoet differences, between centrosymmetrically related wedges of reciprocal space within a single data set, whereas dispersive differences manifest between data sets of the same crystal measured at different wavelengths. Dispersive differences result from the fact that, above certain element-specific wavelengths, X-ray absorption rapidly increases, leading to a significant, wavelength dependent, anomalous contribution to the atomic scattering factor in the vicinity of this wavelength. This specific wavelength at which X-rays are absorbed by a given element is also referred to as the 'absorption edge' of the respective element.

Once marker atom positions are determined, the phasing equations for the protein electron density can be solved and the initial protein phases can be determined. With these an electron density map can be reconstructed by Fourier synthesis (see Equation 12 and Fig. 17). Depending on the source of the difference signal and the data available, various methods for phase determination can be

distinguished and may also be combined to obtain an unambiguous solution for the respective phase equation. Generally, all these methods are referred to as anomalous diffraction (or anomalous dispersion) techniques and for all of them structural isomorphism between the native and the derivative (sub-) structure is a fundamental requirement. This can be difficult to achieve for heavy atom derivatives in isomorphous replacement (IR) methods, but is inherent in anomalous or dispersive data from one and the same crystal. While multi-wavelength anomalous diffraction (MAD) methods require exactly tunable wavelengths, and thus generally a synchrotron source, single-wavelength anomalous diffraction (SAD) methods can be recorded at any X-ray wavelength as long as anomalous data can be recorded*. Since marker atom substructure methods do not depend on prior structural information, they are also referred to as *de novo* or experimental phasing techniques. (Rupp, 2010)

3.3.7.2 Density modification

Density modification techniques are based on the fact that protein crystals consist to a large proportion (commonly $\sim$50%) of disordered solvent surrounding the protein molecules. In the so-called solvent flattening method, these solvent regions are determined and mathematically separated from protein regions. The solvent region is either flattened to a constant value or flipped after adjusting to a mean solvent density of zero. Further density modification methods include histogram matching, which is an image processing technique that compares and adjusts electron density pixel values with an electron density histogram for an idealized protein structure, and reciprocal space maximum likelihood density modification. Moreover, density averaging of multiple copies of the molecule in the asymmetric unit can be effectively used to improve map quality.

Density modification techniques are used after the substructure phasing methods in practically all de novo structure determinations and are powerful phase improvement methods. In cases of very high non-crystallographic symmetry (NCS, see above), as it occurs for example in some virus capsids, *de novo* phasing using only density modification and density averaging is possible. Moreover, density modification techniques are indispensable in the resolution of phase ambiguity occurring when phases are derived from a single set of difference data as well as for the discrimination between substructure enantiomorphs (see below). Density modification does not require an atomic model, and can thus be considered as an extension of experimental phasing (Rupp, 2010).

*This depends on the X-ray absorption edge oft the respective anomalous scatterer and the X-ray wavelength available. For a more detailed explanation see (Rupp, 2010).

3.3.7.3 Molecular replacement phasing

Molecular replacement (MR) methods require the availability of a structurally similar search model. Phases can then be obtained by placing this model into the unit cell in the same position and orientation as the actual crystallographic motif. "Replacement" is thus rather to be understood as "positioning" of the search model in the crystal structure than as "substitution". The phases of the correctly placed model can then be calculated and serve as initial phases for electron density reconstruction from the measured structure factor amplitudes. However, the resulting initial electron density map is heavily biased by the search model, because phase values dominate electron density reconstruction much more than structure factor amplitudes. Therefore for any structure determined by molecular replacement the initial map needs to be appropriately refined to remove (or at least reduce) model bias. (Rupp, 2010)

3.3.7.4 Direct phasing methods

Direct phasing methods exploit the fact that relations exist between certain sets of structure factors. These require high-resolution data (1.2 Å or better) for *ab initio* determination of a protein structure and have so far been limited to relatively small proteins. Direct phasing methods are, however, used successfully in the marker atom substructure determination required for *de novo* phasing techniques. (Rupp, 2010)

3.3.8 Patterson maps

A central tool for basically all crystallographic phasing techniques is the calculation of so-called Patterson maps directly from the measured structure factor amplitudes. The underlying Patterson function *P(u,v,w)* is the convolution of the electron density over the unit cell, which in turn is the value of the Patterson function at point u that can be calculated as the integrated product of the electron density at point *r* and the electron density at another point *r+u* over the unit cell. This is equivalent with an autocorrelation of the electron density ρ(r) with its reciprocal value ρ(-r).

$$P(u) = \int_R \rho(\vec{r})\rho(\vec{r}+\vec{u})\, d\vec{r} = \rho(\vec{r}) \otimes \rho(-\vec{r})$$

Equation 14

The value of the function *P(u)* (Equation 14) becomes large when the electron density ρ is high at both points r and *r+u*, which is the case when both points coincide with atomic positions. Peaks in a graphical plot of the Patterson function, the Patterson map, hence represent interatomic distance

vectors. Conveniently, the Fourier transform of the Patterson function *FT[P(u)]* can be obtained by using squared structure factor amplitudes, since these represent the product of the structure factor F_h with its complex conjugate F_h^*:

$$FT\ [P(u)] = FT[\rho(\vec{r})] \cdot FT[\rho(-\vec{r})] = \ F_h F_h^* = |F^2|$$

Equation 15

This means that Patterson maps can be calculated without any phase values. In theory, the interatomic distance vectors obtained through the Patterson function are sufficient to compute the atomic structure of the corresponding molecule, and this has already been done successfully for relatively small molecules. However, there are some major problems complicating direct structure solution of proteins using this method: Firstly, a molecule with *N* atoms will generate *N(N-1)* non-origin peaks, a insurmountable computational challenge in case of large molecules such as most proteins; secondly, Patterson maps are centrosymmetric irrespective of the structure symmetry and thus do not contain information on the handedness of the underlying structure; and lastly Patterson maps are typically very noisy.

3.3.8.1 Patterson maps in substructure determination

Nevertheless, Patterson maps are very useful for (heavy atom) substructure determination required for experimental phasing. Typically these substructures consist of only a few heavy atoms or anomalous scatterers that moreover are usually very electron rich and thus produce very high peaks that are well distinguishable from noise in Patterson maps. Patterson vector based methods for substructure solution exploit the fact, the autocorelation described by Equation 14 and search for a fit of substructure fragments that fit to the interatomic distance vector peaks. Automated programs, such as SHELX-D from the SHELX-program suite (Sheldrick, 2010), generate multiple possible heavy atom substructure models that are ranked according to a combined scoring function and often these programs employ a combination of Patterson and Fourier methods to test solutions for consistency. However, because of the centrosymmetry of Patterson space, there is a 50% chance that the best solution is either the correct substructure enantiomorph or needs to be inverted. The discrimination between correct and inverted substructure is usually also performed by automated programs. SHELX-E (Sheldrick, 2010) for example calculates the phases for both enantiomorphs and compares the contrast level of the resulting electron density maps. After several cycles of density modification, which are commonly executed by the same program, the contrast between solvent and molecule regions in the map of correct enantiomorph will become clearly distinguishable while the map of its inverted counterpart remains noisy.

3.3.8.2 Patterson maps for the calculation of self-rotation functions

Another important use of Patterson maps is the calculation of self-rotation functions. One use of these rotational matrices is to detect non-crystallographic symmetry (NCS) in a crystal's asymmetric unit. For this purpose, a Patterson function for the entire protein is calculated and superimposed with itself at discrete rotation angles. This generates distinct peaks for those rotation angles that superimpose two copies of the same molecule in the apart from that very noisy map.

3.3.8.3 Patterson maps in molecular replacement

Finally, Patterson maps are essential for phasing by molecular replacement (MR). The above-mentioned placement of the search model in the correct orientation into the unit cell is a 6-dimensional search problem (3 dimensions of rotation + 3 dimensions of translation). Since this usually exceeds the available computational power, MR searches are typically split into a 3D rotational search followed by a 3D translation search with the best rotation solution. By placing the MR search model into a large virtual box, any (for this purpose useless) intermolecular vectors in the Patterson map of the search model can be mathematically excluded during rotational search. The rotational space solutions are then sampled and a Patterson function for each orientation is computed. These functions are subsequently scored against the Patterson map of the experimental data, and the best orientation will show high correlation with the computed Patterson map because their intramolecular vectors match. Once a suitable rotational function is found, a translational search is carried out. By placing the MR-search model at pre-defined grid points of the unit cell, the best translation solution can be identified since intermolecular distance vectors of the search model and the actual protein coincide and produce high correlation scores. MR searches often produce multiple solutions, which are evaluated using three parameters. Firstly, obvious clashes between molecules are discarded with by calculating a packing function, which at the same time serves as a constraint for the translation function. Secondly, the log-likelihood gain (LLG) score indicates how much better the data can be explained using the oriented model versus a random model. The LLG score is an absolute score, which can also be used to compare molecular replacement solutions of different search models and has high values for good solutions. Third, the Z-score indicates the signal-to-noise ratio of a given solution, which is the number of standard deviations by which a given LLG score lies above the mean LLG of a set of randomly placed search models.

Generally, molecular replacement becomes increasingly difficult when multi-subunit proteins or multiple (NCS related) copies are present, because the signal of each individual correctly placed subunit becomes much weaker. This effect is also referred to as 'flattening out of the search landscape' (Rupp, 2010).

3.3.9 Data collection and processing

To collect a diffraction data set for a successful electron density reconstruction, each unique reflection within the diffraction limit of the crystal has to be recorded at least once, that is the corresponding reciprocal lattice point has to pass through the Ewald sphere. For that purpose, the crystal can either be rotated through an X-ray beam of a given wavelength (monochromatic diffraction) or diffraction data are collected by shooting a continuous X-ray spectrum at a stationary crystal (Laue diffraction). For the detection of the scattered X-ray photons several types of detectors exist, the most advanced being pixel-based single-photon counters with quasi-instantaneous read-out, allowing continuous data collection.

Once X-ray data from a crystal are available, the intensities of the reflections need to be extracted from the data images and processed further. Here, all spots recorded on the diffraction images are indexed according to the crystal's space group, and their intensities are subsequently integrated. The processed data (the "data set") then forms the basis for phase determination either by experimental phasing methods or with the help of molecular replacement models.

Data processing is performed with the help of dedicated programs such as XDS (Kabsch, 2010). This particular program assigns reciprocal space coordinates *hkl* to each reflection (indexing step) by extracting distance vectors between strong spots on a subset of images, based on known experimental parameters like wavelength, detector distance, rotation range and beam position. From these distance vectors, unit cell parameters are derived and scored against the list of possible lattices, assigning a "quality of fit" score to each of the 14 Bravais lattices. In the subsequent integrating step, the intensity of each spot is then estimated. For this, the local background around each reflection is determined and for reflections that are spread over two or more frames, a three-dimensional profile fitting is performed and the intensity of the spot is integrated over this profile. Together with the information on the unit-cell lattice from the indexing step, the integrated intensities are then used to predict the crystal's space group by looking for systematic absences of reflections according to Bragg's law. However, at this stage, screw axes, and much less their handedness, cannot be predicted reliably. This remains an issue of uncertainty until the final structure is solved – an aspect every crystallographer should be aware of and consider revision when facing difficulties with data interpretation even at later stages in the crystallographic pipeline. Nevertheless, the symmetry operations of the assigned space group can be used to merge symmetry-equivalent reflections, which can basically be considered as the redundant measurement of the same reflection. This merging step on the one hand increases the reliability of the measured data by adding redundancy to each measurement; on the other hand it requires the introduction of scaling factors to account for radiation damage, absorption effects and local variations of the detector sensitivity. Basically this step can be considered as kind of sampling several data sets that due to the crystal's space group symmetry contain redundant data. The

correlation of merged reflections has long been indicated by a merging R-factor (Diederichs & Karplus, 1997), which usually is still published along with a crystal structure. However in 2012 Karplus & Diederichs introduced a Pearson correlation coefficient ($CC½$) between the average intensities of the reflections in two random halves of the measurements of each unique reflection in the data set (Karplus & Diederichs, 2012), which nowadays has widely been accepted as a substitute for the merging R-values (see section 3.3.11.2).

3.3.10 Refinement and electron density maps

After an initial molecular model is obtained, this model needs to be refined to optimally explain the calculated electron density. This is executed in an iterative process of computationally refining the model against the experimental data and subsequent adjustment of the model. The computational part part of this process is executed by dedicated programs such as PHENIX (Adams *et al.*, 2010), which refine the model parameters (atom coordinates, B-factors and occupancies) using reciprocal space methods. The algorithms implemented in these programs have a wide radius of convergence* (Garman, 2014) and allow for manual adjustment of several refinement-parameters and the application of various refinement-strategies. Generally reciprocal space refinement aims to minimize a target function, which is the agreement of the structure factor amplitudes computed from the model (F_{calc}) with the experimentally observed structure factor amplitudes (F_{obs}). The convergence criterion of these target functions can either be based on maximum likelihood functions or least-square methods, whereby the former is more robust and therefore most frequently used today (Rupp, 2010).

However, even though software for automatically building atomic models into electron density maps is becoming more reliable, manual model building in real space with the help of graphical modeling programs is still state of the art. The most commonly used program for this purpose is COOT (crystallographic object-oriented toolkit, (Emsley *et al.*, 2010)), which has implemented manifold algorithms allowing the user to for example manually place α-helices, β-strands or single amino acids as well as nucleic acids and ligands and refining their position by means of rigid-body fitting, to execute real-space refinement with adjustable geometric restraints, or to manually rotate and translate objects.

While electron density maps constructed with experimentally determined phases will not be biased towards any structural model, phases obtained by molecular replacement are usually biased and the

* The radius of convergence for an optimization algorithm describes its ability to escape local minima and approach the global minimum.

map will strongly reflect the model features. To minimize model bias in electron density maps, likelihood based minimization coefficients, often combined with additional bias minimization measures, are applied prior to electron density map construction. Roughly spoken, a bias reduced electron density map is calculated based on the difference of the two-fold weighted F_{obs} and the F_{calc} values ($2F_{obs}$-F_{calc}). Moreover the respective refinement strategy should be tailored to initially remove model bias for example by the 'simulated annealing'-algorithm implemented in PHENIX (for more information see (Rupp, 2010)). Especially when refining parameters of ligands that have been soaked into a crystal or introduced by co-crystallization it is recommended to calculate a so-called difference omit density map. This is done by refining the model parameters without the ligand and then calculating a electron density map from the difference between F_{obs} and F_{calc} (F_{obs}-F_{calc}). Depending on the program used for electron density map calculation, both, $2F_{obs}$-F_{calc} and F_{obs}-F_{calc}, can be optimized by additional factors.

3.3.11 Evaluating crystal structures

A number of streamlined program packages are now available that can overcome many difficulties in data interpretation and phasing automatically, or with only minimal user adjustment and in some cases, crystallographic software packages are even capable of solving structures without human intervention. However, since data processing and phasing have a major impact on the resulting structural information while leaving room for dramatic misinterpretations at several stages, it is still useful to assess their outputs for biochemical plausibility. Moreover, the automated approaches typically fail when challenging macromolecules are analyzed, such as large complexes, poorly diffracting crystals, or complicated crystal packing arrangements. Here, the input of human intelligence, experience and creativity is still essential.

Perhaps not everybody needs to understand the intricate details and challenges of molecular structure determination that have only been briefly described above. However, this work is fundamentally based on the crystal structures of two bacterial enzymes that both play central roles in cell wall metabolism and recycling of their respective organism, *P.putida* (MurU) and *B.subtilis* (MurQ). It is thus indispensable to the critical reader that he or she is able to evaluate a given structural model. In the following paragraph, a brief overview of the most commonly used and descriptive parameters to assess the quality and reliability of crystal structures will be given. The intention hereby is not to summarize the current state of the art in X-ray crystallography, but to provide even the non-expert reader with a basic toolkit to critically review the structural models presented in this work.

3.3.11.1 Assessment of model quality

Geometrical Parameters

With very few exceptions, X-ray crystallographic structures are deposited in the Protein Data Bank (PDB) at www.rcsb.org. The PDB employs a set of quality controls that ensure that the deposited structures conform to standard geometrical parameters, such as bond distances, bond-angles or protein backbone torsion angles. Bond distances and bond angles are commonly reported using root mean square (rms)-deviations describing how far (in Å or degrees, respectively), on average, bonds or angles deviate from ideal values. A trustworthy structure might have rms-deviations of less than 0.015 Å and less than 1.5° for bonds and angles, respectively. Usually protein geometry is also evaluated in a Ramachandran diagram, which is a suitable indicator of the stereochemical quality of a structural model and where all torsion angles for all amino acids are plotted and so called favorable, allowed and disallowed torsion angles are defined (Ramachandran, 1963). Ideally, a structural model would have favorable torsion angles for over 90% of its residues and no residues in disallowed regions.

Refinement R-values

The standard indicator for assessing the agreement of a refined model with the diffraction data is the crystallographic refinement R-value, R_{work}, defined as:

$$R_{work} = \frac{\sum_{hkl} |F_{obs}(hkl) - F_{calc}(hkl)|}{\sum_{hkl} F_{obs}(hkl)}$$

Equation 16

R_{work} numerically quantifies the overall agreement between the observed diffraction amplitudes (F_{obs}) and "virtual" amplitudes calculated using the structural model (F_{calc}).

Values for R_{work} are reported as fractional- or %-values. A model that agrees well with the diffraction data will give rise to calculated amplitudes that are very similar to the observed ones, meaning that the R-value is 0 for perfect agreement of model and data and near 60 % for a random model. In practice, good values for R_{work} are in the range between 15 % to 25 %.

The so-called R_{free} is a cross-validation R-value that is needed to prevent over parameterization of the model during model building and refinement. It is basically calculated in the same way as R_{work} but only from a small subset of the experimental data that is separated from the working data set and not used during structural refinement. Reciprocal space refinement programs calculate R_{work} and R_{free} values for the molecular model for each iterative cycle of refinement. During model building and refinement both, R_{work} and R_{free}, should improve (become lower) progressively as the model becomes more complete and more parameters are introduced describing the experimental data better and better. At a certain stage of refinement the model will be optimal and introduction of further parameters will not improve the model. At this point R_{free} will stop improving, and with further (over)

fitting of model parameters will start to increase again, while R_{work} keeps dropping. Consequently the lowest possible R_{free} of a model marks its optimal refinement that can be achieved from the respective data set, whereas low values of both R_{work} and R_{free} indicate a good fit between model and experimental data. In this context 0,1 (or 10%) would be exceptionally good R-values and a trustworthy structure is generally expected to have R-values that are lower than 0,3 (or 30%). Generally, the R_{free} should close to the value for R_{work}. A small difference between them (~5%) is acceptable, while a difference > 10% might well indicate a problem with the structure determination, such as an over-parameterization ('overfit') of the model. Actually the numerical difference between R_{work} and R_{free} constitutes a more suitable criterion to evaluate model quality than the absolute number of the two R-values (Karplus & Diederichs, 2012).

3.3.11.2 Assessment of data quality

There is also a set of parameters that can be consulted for the evaluation of the quality of the collected data, and thus the reliability and significance of the resulting structural model. In serious works, these are published along with the model refinement statistics. For the sake of simplifying the question of how to asses the quality of diffraction data as objectively as possible, it is helpful to dissect this task into two categories: The evaluating the quality of a data set in which equivalent intensities have been merged on the one hand and the assessment of unmerged data on the other (Weiss, 2001).

Quality indicators for data sets containing averaged intensities

In the former case the resolution to which reflections could be indexed on the diffraction images is probably the most intuitively understandable criterion. Even though in X-ray crystallography there exist several different definitions for the resolution of a given data set, this value gives an impression on how distinct and reliable atomic details can be deduced from the electron density map and thus how significant conclusions on molecular interactions will be. Because the strength of diffraction decreases with resolution, a high-resolution cut-off is applied to discard data considered so noisy that their inclusion might degrade the quality of the resulting model. Unfortunately, the resolution of a given data set is not unambiguously defined by the diffraction data (Karplus & Diederichs, 2012). Even though there exist some helpful, robust and statistically informative quantities indicating the physical diffraction limit of the crystal (see below) the actual definition of the high-resolution cutoff remains a subjective decision. In this context it is helpful to also check the parameter I/σ, which describes the signal to noise ratio of the measured reflections over all resolution bins included in a data set to assess the respective reliability of the detected (or missing) reflection signals (Weiss, 2001). Good resolutions in x-ray crystallography are between 1,5 and 2,5 Å (1 Å = 0,1 nm) but resolutions of 1,0 Å and even lower have also been achieved. Most structures deposited in the PDB have a resolution between 1,5

and 2,5 Å (www.pdb.org), but there are as well some structures at lower resolutions (> 3 Å). In these 'lower-resolution structures' interpretation of details such as hydrogen bonds, contacts etc. must be done with caution (Weiss, 2001).

Another important number is the completeness of the data set. Any missing reflection, be it one that has not been measured or one that has been deliberately excluded, leads to a deterioration of the model parameters much in the same way as reduced resolution does (Weiss, 2001).

Quality indicators for unmerged data sets

Prior to merging equivalent reflections, there are several statistics providing a valuable source of information about the quality of a data set. The first criterion is the redundancy of the data. Since X-ray diffraction data collection is to some extent influenced by counting statistics, the averaged measurement should become more accurate as more individual measurements are made. A highly redundant data set will therefore be intrinsically of higher quality than a data set in which every reflection has only been measured once. Data collected from crystals exhibiting low space group symmetry typically have lower redundancies as more reflections need to be recorded to yield a complete data set (Weiss, 2001). Crystallographic data quality has long been assessed by the merging R-factor R_{merge} (originally R_{sym}), which measures the spread of n independent measurements of the intensity of a reflection, $I_i(hkl)$, around their average $I(hkl)$

$$R_{merge} = \frac{\sum hkl \sum_{i=1}^{n} |I_i(hkl) - I(hkl)|}{\sum hkl \sum_{i=1}^{n} I_i(hkl)}$$

Equation 17

Adjusting R_{merge} by a factor of $n/(n-1)$ yields values that are independent of the multiplicity and this multiplicity-corrected version, called R_{meas}, reliably reports on the consistency of the individual measurements (Diederichs & Karplus, 1997). A further variant, R_{pim}, reports on the expected precision of $I((hkl)$ and is lower by a factor of $1/\sqrt{n}$ compared with R_{meas} (Weiss, 2001, Karplus & Diederichs, 2012). Data typically have been used to be truncated at a resolution before the R_{merge} (or R_{meas}) value exceeds ~0.6 to 0.8 and before the empirical signal-to-noise ratio, I/σ, drops below ~2.0. However, in 2012, Karplus & Diederichs introduced the correlation coefficient $CC½$ as a new, robust, and statistically informative quantity for defining the high-resolution cut-off of a given data set replacing the hitherto-used value R_{meas}. It is calculated as a Pearson correlation coefficient between the average intensities of the reflections in two random halves of the measurements of each unique reflection in the data set. The application of $CC½$ as high resolution cut-off allows the inclusion of weak reflections from high resolution bins, which according to Karplus & Diederichs still contain structural information and contribute to improve model quality (Karplus & Diederichs, 2012).

3.3.11.3 B factors

Scattering is generally attenuated by destructive interference of scattered waves. In the context of x-rays scattered at atomic electron shells, this interference originates from path differences between the scattered waves emerging from different locations and with different scattering angels from the same atom. Hence, the atomic scattering function is dependent on the scattering angle θ and decreases with higher scattering angles. In the context of a crystal lattice, the atomic scattering factors are additionally attenuated by displacement of the atoms from their mean positions, resulting from thermal vibrations as well as from displacive disorder in the crystal lattice. The mean displacement Δr of an atom around its equilibrium position is described by its (isotropic) displacement parameter B_{iso} (also called the (isotropic) temperature-factor) which is given by Equation 18 and has the dimension Å^2.

$$B_{iso} = 8\pi^2 \, \Delta r^2$$

Equation 18

All these effects compound and are accounted for in an exponential, B-factor- and scattering angle-dependent, attenuation factor (called Debeye-Waller factors T_s) which moreover describes the wavelength dependency of the scattering factor attenuation. With this, the temperature-factor dependent atomic scattering factor becomes:

$$f_s^{B_{iso}} = f_s^0 \exp\left(-B_{iso}(sin\theta/\lambda)^2\right)$$

Equation 19

Equation 19 essentially describes the attenuation of the atomic scattering function as a function of the scattering angle (θ), expressed in terms of $sin\theta/\lambda$. (Rupp, 2010). From the definition given in Equation 18 is is obvious that B-factors provide information on the mobility of the crystallized protein and its associated components such as ligands or solvent. Considering this, B-factors are very helpful in evaluating crystal structures in terms of protein or ligand dynamics and flexibility. Depending on the quality of a given data set, B-factors can be described as isotropic average values for the whole molecule, parts of it or, in case of good data, by more elaborate anisotropic models that also consider the directionality and effective degrees of freedom of molecular movements. B-factors can also be calculated for single atoms of a molecule and are then specified as atomic B-factors. Especially in the case of ligand refinement, but also for all other atoms of a given structure, atomic B-factors need to be refined along with the corresponding occupancies of the respective atom in the crystal since the two parameters are mutually influencing each other. High atomic B-factor values (≥ ~ 80 Å^2) might indicate a problem, particularly if for example a ligand has a significantly elevated B-factor compared to surrounding amino acids. Such a case indicates that the ligand may not be bound well in the binding pocket, and might have high mobility and/or low occupancy.

The so-called Wilson B-factor provides an estimate for the overall B-factors and thus an approximate scale factor for the data. It is calculated on the assumption that the atoms in a given crystal are randomly distributed. However, since the largest B factors for atoms in the crystal are ignored in the

Wilson B calculation, it is not an estimate of the average of the individual atomic B-factors but rather an interesting property of a set of diffraction intensities. In practice this omission results in the Wilson B usually being lower than the mean of the individual B factors (Malito, Stamp Collecting Project).

3.3.12 Ligand structures

As an inherent methodological property of X-ray crystallography, crystallographic macromolecular structures are time and space averages over the many millions of macromolecules within the crystal. A "large" protein crystal is typically smaller than 100 mm in all three dimensions. For an average-sized 5- nm-diameter globular protein, such crystals would contain $\sim 10^{13}$ molecules (Garman, 2014). Consequently it is not unusual that in a co-crystallization or soaking experiment not each potential binding site in the crystal is occupied by a ligand. This might have steric reasons resulting from the crystal packing which can hinder ligand diffusion and thus lower its local concentration or might be the effect of kinetic factors such as binding constants or enzyme turnover rates. Adding the influence of flexibility of either the ligand itself or protein-side chains involved in ligand binding and even reorientations of whole domains upon ligand binding, it becomes comprehensible that it can be quite difficult to trap a ligand in its protein-bound state in a crystal. In the resulting electron density map these effects become evident in a reduced quality of the ligand's electron density. In the refinement of ligand parameters this is accounted for by optimizing the ligand's occupancy along with its atomic B-factors. In this context the ligand's occupancy is an average value describing the percentage of binding sites in the crystal that are occupied by the ligand, whereas its atomic B-factors (often also calculated as an average B-factor value for the whole ligand), roughly spoken, reflect the inherent flexibility or the thermal motion of the ligand molecule. A dynamic behavior of a given protein upon ligand binding can furthermore have severe impact on crystal packing, and, when soaked with ligand solution, sometimes result in crystal damage and/or a change of the crystal's space group. This might require a new round of screening for optimal crystallization conditions to co-crystallize the protein with its ligand(s). Both, in soaking and co-crystallization experiments, it is not unusual that the crystal's diffraction quality is also influenced by the ligand making it is mandatory to optimize the ligand concentration in the soaking or co-crystallization solution. In this respect, it is often useful to on the one hand reflect the protein's physiological ligand binding conditions as realistic as possible (e.g. by the addition of ionic cofactors) and on the other hand to preserve the crystals growing conditions.

3.4 X-ray diffraction of MurU and MurQ

3.4.1 Data collection

All data were collected at 100 K at the synchrotron radiation beamline ID29 of the European Synchrotron Radiation Facility (ESRF) in Grenoble, France as well as at the beamlines X06SA and X06DA of the Swiss Light Source in Villigen, Switzerland. Crystals used for data collection were preselected according to their diffraction power at a Cu-K_α X-ray radiation source (MAR345), the 'home source' of the Stehle-group in Tübingen.

Diffraction patterns of the non-liganded MurU as well as of the enzyme-ligand complexes were generated using an X-ray wavelength of 1.0 Å, whereas diffraction data of the NH_4I-soaked crystals for experimental phasing were obtained at 1.7 Å.

All diffraction experiments with MurQ-crystals were performed at a wavelength of 1.0 Å.

3.4.2 Structure Determination, model building and structure evaluation

3.4.2.1 Data processing, spacegroup determination and crystal packing

Indexing, integrating and scaling of all diffraction data was done with the XDS software package (Kabsch, 2010). Since the algorithm for space-group assignment implemented in XDS is not able to assign screw axes, space-groups were assigned using the POINTLESS software package (Evans, 2006), which serves to identify space group possibilities from unmerged, indexed and integrated data as derived from an XDS output-file (XDS_ASCII.HKL). Additionally, spacegroup assignment was cross-validated manually by checking for consistency with systematically missing reflections. Moreover the crystal packing (in this context defined as the number of monomers per asymmetric unit) and solvent content for the respective space-group was verified by calculating the crystal volume per unit of the protein's molecular weight, also known as Matthews coefficient V_M, and by analyzing the Matthews probability of the respective crystal using the Matthews Probability Calculator online tool (http://www.ruppweb.org/mattprob/default.html) (Kantardjieff & Rupp, 2003). Here the Matthews probability is calculated as the probability of the respective crystal packing to be correct when compared to the crystal packing of all structures deposited in the PDB (or to all PBD-structures of the same resolution yielding a resolution dependent Matthews probability).

3.4.2.2 Phasing

Experimental phases for non-liganded ("native") MurU were determined in a single anomalous dispersion (SAD) experiment, exploiting the anomalous signal of iodine in NH_4I-soaked crystals at a wavelength of 1.7 Å (Abendroth *et al.*, 2011, Dauter & Dauter, 2007). Initial phases were derived by determining the substructure of 4 iodine atoms per asymmetric unit with the help of the program suite SHELX C/D/E (Sheldrick, 2010) and were further extended and refined with SHARP / autoSHARP (Vonrhein *et al.*, 2007) and PHENIX (Adams *et al.*, 2010). Molecular replacement for MurQ and MurQ point mutants (structures not presented in this work) as well as for all ligand-bound structures of MurU was done with PHASER (McCoy *et al.*, 2007). For the MurU enzyme-substrate complexes and the MurQ point mutants the structure of the native enzyme served as a search model, whereas for MurQ the structurally related putative phosphosugar isomerase HI0754 from *Haemophilus influenzae* (PDBid: 1NRI) was used.

3.4.2.3 Model building

Model building was performed manually in COOT (Emsley *et al.*, 2010), and models were refined against experimental data using PHENIX/phenix.refine (Adams *et al.*, 2010).

The TLSMD ('translation, libration, screw' motion determination) web server (Painter & Merritt, 2006) was used for the determination of translation, libration (small movements) and screw-rotation (TLS) parameters of the protein. To remove model bias, simulated annealing was performed at least once for all structures. Unbiased difference electron density maps (calculated by PHENIX/phenix.maps as mF_o-DF_c) were calculated for the ligands in the substrate soaked complex structures by doing simulated annealing (with PHENIX) with a structural model in which no ligands had been modeled. The resulting mF_o-DF_c-map, which can be considered a 'simulated annealing omit difference density map', is thus suitable to verify if the ligands are actually present in the crystal. Coordinate and parameter files for MurNAc-α1P, UpNHpp and UppNHp were generated using PHENIX/phenix.elbow. All ligands were built manually using COOT, and their coordinates were further refined with PHENIX/phenix.refine. All structure figures were generated with PyMOL (DeLano, 2012). Comparative structure superimpositions were performed using the alignment algorithm implemented in PyMOL and electrostatic potentials were calculated with PyMOL-implementations (Baker *et al.*, 2001) as well.

3.4.2.4 Structure evaluation

Structural models were validated using the respective implementations in COOT (Emsley *et al.*, 2010), PHENIX, and the CCP4 suite (Winn *et al.*, 2011) as well as the 'PDB Data Validation and Deposition Portal' (http://deposit.pdb.org) and the MolProbity Web page (Chen *et al.*, 2010).

To distinguish between physiological interfaces and protein-protein contacts merely resulting from crystal packing – so called 'crystal contacts' – the refined models were analyzed with the EPPIC web interface at http://www.eppic-web.org (Duarte *et al.*, 2012).

3.5 Molecular dynamics simulations

Molecular dynamics (MD) simulations were calculated by Thomas E. Exner (Institute of Pharmacy, University of Tübingen). Depending on the intention of the respective simulation run, force field constants were defined differently to either predict the presence and positions of protein side chains, ligands or metal ions. By applying for example a high force field to the protein, its atoms were kept close to their crystallographically determined positions while optimizing the positions of the ligands. The presence and positions of divalent cations at the active site were predicted by keeping both, protein and substrate atoms in place. However, it is important to note that this MD-approach was not suitable to distinguish between Mg^{2+}, NH^{4+} or Na^{+} at these positions.

4. Results MurU

4.1 Characterization of MurU

MurU belongs to the enzyme family of sugar-nucleotidyltransferases (SNTs, also referred to as sugar nucleotide pyrophosphorylases, EC 2.7.7.-) whose members share a similar domain organization but do not necessarily display a high sequence identity (Singh *et al.*, 2012). MurU orthologs from Pseudomonas species, however, display sequence identities of over 70%, arguing for a general relevance of the insights gained from MurU of *P.putida* among these organisms. Moreover, previous studies revealed functional orthologs of *P.putida* MurU in *P. aeruginosa PAO1*, (70.8% amino acid sequence identity, 80.5% similarity), *Neisseria meningitidis MC58* (56.2% identity, 69.1% similarity) and in *Caulobacter crescentus CB15* (34.3% identity, 46.1% similarity) (Gisin *et al.*, 2013).

4.1.1 MurU has a narrow substrate specificity for MurNAc a-1P and UTP.

MurNAc-α1P, but not the related GlcNAc-α1P or Glc-α1P compounds, can function as sugar phosphate substrate for MurU, yielding UDP-MurNAc along with UTP (Gisin *et al.*, 2013). MurU also displays narrow specificity for UTP as no other tested nucleotide triphosphate (ATP, CTP, TTP, or GTP) is accepted as a substrate according to a ^{32}P-radioassay (Fig. 18A).

4.1.2 MurU behaves as a monomer in gel filtration

The theoretical molecular mass of the His_6-tag version of MurU is 24 879 Da. Analytical gel filtration of MurU by HPLC revealed an apparent molecular weight of a monomer, indicating that the physiologically active moiety of MurU is the monomer(Fig. 18B). In contrast, related nucleotidyl transferases have been reported to form multimers, tetramers (RmlA, (Blankenfeldt *et al.*, 2000), or trimers (GlmU, (Trempe *et al.*, 2011) However, within these, nucleotidyl multimerization is facilitated by the C-terminal domain, which is truncated in MurU.

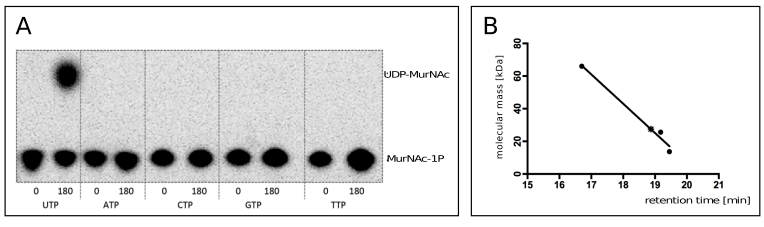


Fig. 18: Experimental characterization of MurU: A: Nucleoside triphosphate specificity of MurU. MurNAc- α -1-(32P)-phosphate (MurNAc-1P) was incubated with the indicated nucleoside triphosphates. Only the combination MurNAc- α -1-(32P)-phosphate and UTP yielded a product, UDP- α -(32P-Phosphate)-MurNAc (UDP-MurNAc). B: Recombinant MurU (star) appeared as monomer sized protein (His6-tag fusion protein, theoretical molecular mass 24.9 kD; apparent molecular mass 27.5 kD, 18.87 min). Standards used for gel filtration separation by HPLC (filled dots) were albumin (molecular mass 66.0 kD, retention time 16.70 min), chymotrypsinogen (25.6 kD, 19.18 min), ribonuclease (13.7 kD, 19,45 min). Experiments were performed by Jonathan Gisin and Alexander Schneider (AG Mayer).

4.2 Crystallization of MurU

For crystallization the uncleaved MurU-His_6-fusion construct was used. The homogeneity and purity of each crystallization sample was validated in a SD75 gel filtration and a SDS-PAGE. Fig. 19 shows an exemplary gel filtration profile and the corresponding SDS-gel confirming homogeneity, purity and molecular weight of the analyzed sample.

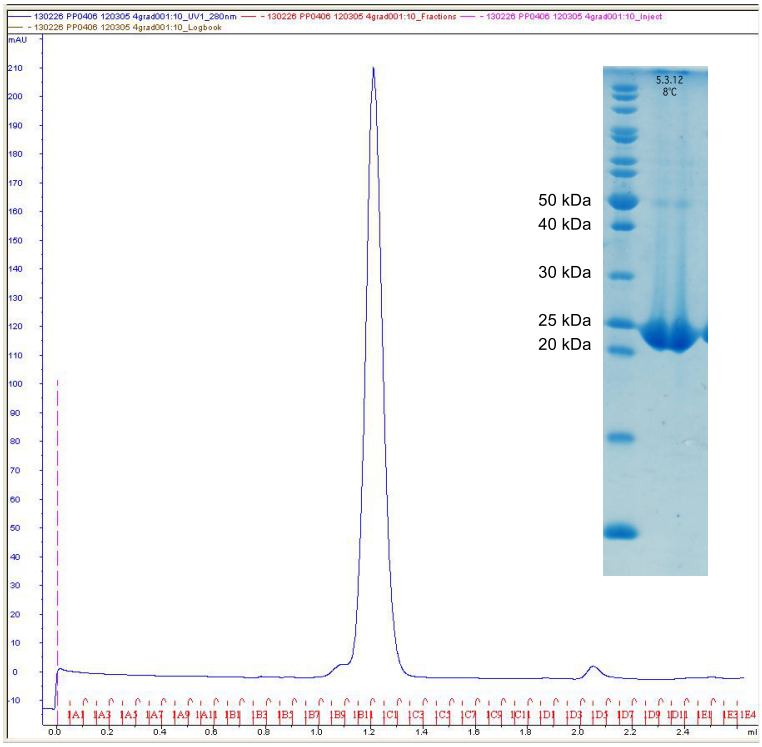


Fig. 19: Elution profile of MurU-His_6 in an analytical gel filtration run using SD75 column. Insert: SDS-PAGE analysis of MurU-His_6.

MurU required protein concentrations from 6,5 to 9,0 mg/ml for crystallization and 1.0 – 1.5 M $(NH_4)_2SO_4$ as precipitant tolerating a pH-rage from 6 to 8.5. Initial small non-diffracting crystals grew within 2 to 3 weeks at 20°C, however diffraction-quality crystals could be obtained approximately 15 weeks after tray setup at the earliest, and even after 8 months and neraly 2 years more new crystals emerged. Fig. 20 illustrates the crystallization behavior of MurU displaying some typical crystals with their corresponding crystallization conditions and approximate growing times.

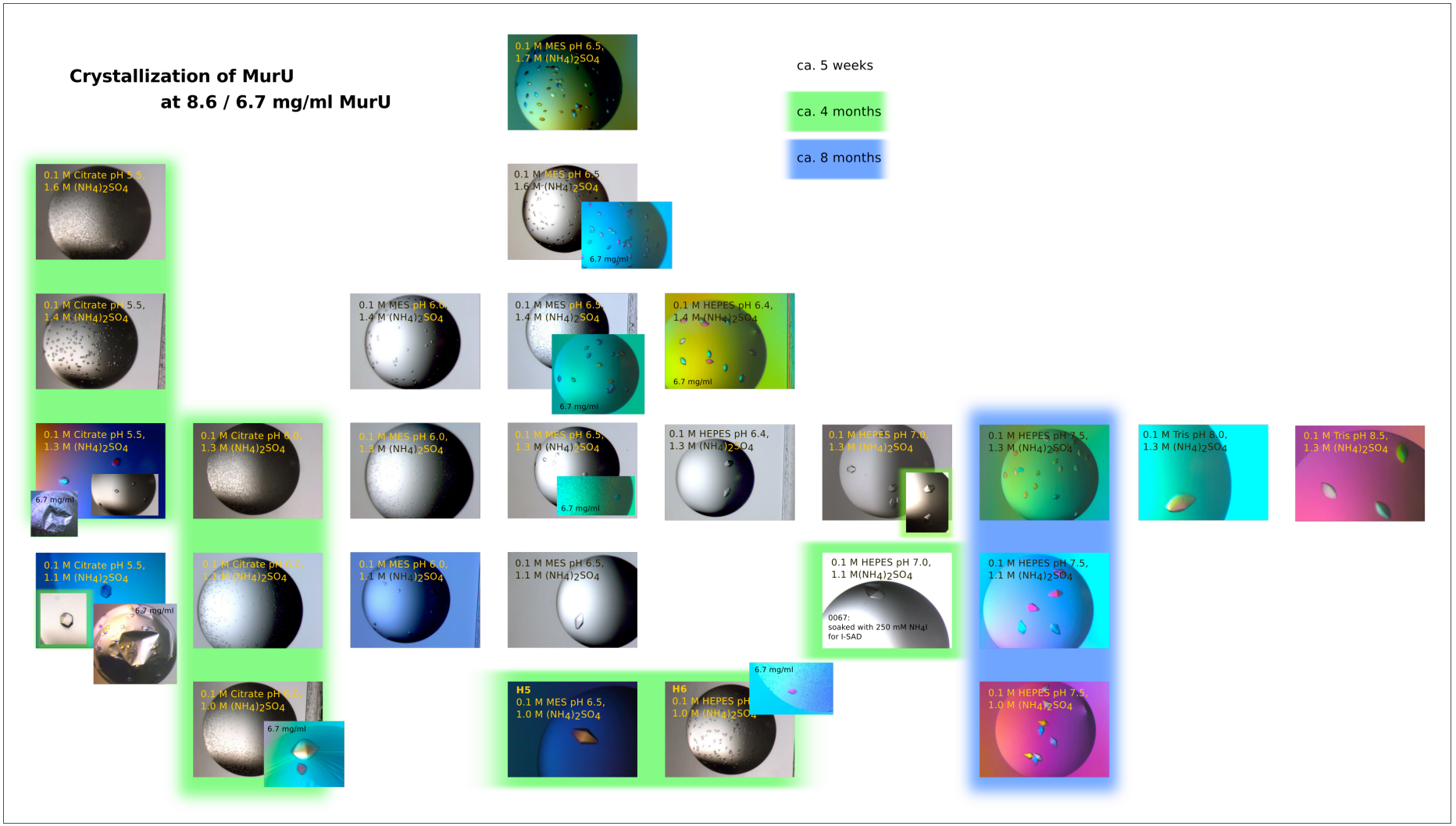


Fig. 20: Crystallization conditions for MurU

Depicted are some of the most typical crystal forms that grew in the respective conditions. The background color indicates the average time span the respective crystal from needed to grow. White: ca. 5 weeks, green: ca. 4 months, blue: ca. 8 months. Crystal pictures are arranges according to the respective crystallization condition with vertically descending concentrations of $(NH_4)_2SO_4$ (top to bottom: 1,7 M → 1,0 M) and horizontally ascending pH (left to right: pH 5,5 → pH 8,5).

4.3 Structure determination MurU

4.3.1 MurU - space group determination and crystal packing

Diffraction data from MurU crystals were processed in space group $P6_122$ as determined using XDS and POINTLESS. To further analyze crystal packing, the crystal volume per unit of the protein's molecular weight, the so-called Matthews coefficient V_m was calculated from the space group and unit cell parameters (axes and angels). The crystallized MurU-His_6 construct has a molecular mass of 24,88 kDa and the unit cell axes lengths determined by XDS were a = b = 72,61 Å; c = 158,47 Å with angles of α = β 0 90°; γ = 120°. For space group $P6_122$ this results in a Matthews coefficient of V_m = 2,46 $Å^3$/Da and an approximate crystal solvent content of 50,0 % with a Matthews probability of 1.0 (Matthews, 1968) (see). This corresponds to a crystal packing with one monomer per asymmetric unit (asu), which nicely reflects the monomeric nature of physiologically active MurU, even though the two aspects are not causally related and have nothing to do with each other.

```
+--------------------------------------------------------------------------+
| MATTPROB - MATTHEWS PROBABILITY and SOLVENT CONTENT ESTIMATES             |
+--------------------------------------------------------------------------+
| The selected method is parameter-free kernel estimator to 2013 data      |
| (C) 2013 C.X. Weichenberger and B. Rupp                                  |
+--------------------------------------------------------------------------+
| Unit cell dimensions in A, deg(o) and cubic A :                          |
+--------------------------------------------------------------------------+
|      a         b         c       alpha     beta      gamma      volume   |
|    72.88     72.88    159.70     90.00     90.00    120.00      734602.3  |
+--------------------------------------------------------------------------+
| Space group P 61 2 2   , space group number 178, Laue class 6/mmm, z= 12 |
+--------------------------------------------------------------------------+
| Unit cell volume is       :     734602.3 A**3                            |
| Asymmetric unit volume is :      61216.9 A**3                            |
+--------------------------------------------------------------------------+
+--------------------------------------------------------------------------+
| Found  1 possible solutions :                                            |
+--------------------------------------------------------------------------+
| Vm of protein (Matthews coeff.) is  2.46 A**3/Dalton                     |
| Estimated solvent content is       50.01 %                               |
+--------------------------------------------------------------------------+
+--------------------------------------------------------------------------+
|  N(mol)      Prob(N)     Prob(N)        Vm          Vs          Mw       |
|          for resolution  overall      A**3/Da   % solvent     Da          |
+--------------------------------------------------------------------------+
|   1          1.0000      1.0000        2.46       50.01    24879.00      |
+--------------------------------------------------------------------------+
```

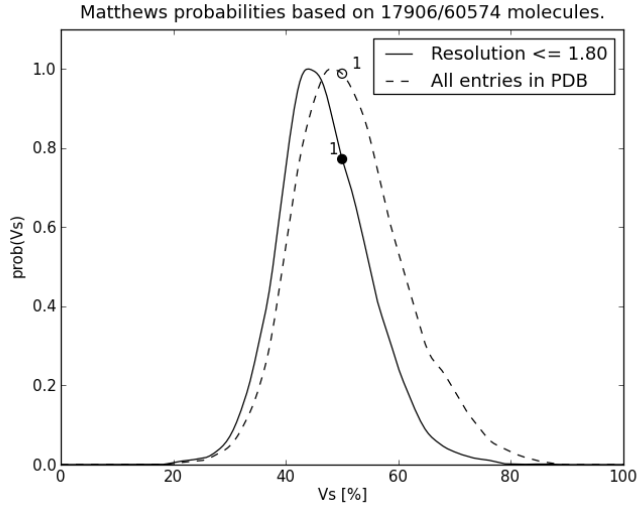


Fig. 21: Calculation of the Matthews coefficient V_m, the solvent content and the Matthews probability for a typical MurU crystal from its spacegroup ($P6_122$) and unit cell parameters using the Matthwes Probability Calculator online tool.

4.3.2 Phase determination for MurU

Experimental phases for non-liganded ("native") MurU were determined in a single anomalous dispersion (SAD) experiment, exploiting the anomalous signal of iodine in NH_4I-soaked crystals at a wavelength of 1.7 Å as described in (Abendroth *et al.*, 2011, Dauter & Dauter, 2007). For initial sub-structure determination, data from the derivatized crystal were used with a resolution ranging from 50.0 to 3.9 Å that was later on extended to a 2.4 Å (Fig. 22A+B). With the help of the C and D packages of the SHELX program suite (Sheldrick, 2010) the iodine substructure consisting of 4 I^--ions could be solved (Fig. 22C+D).

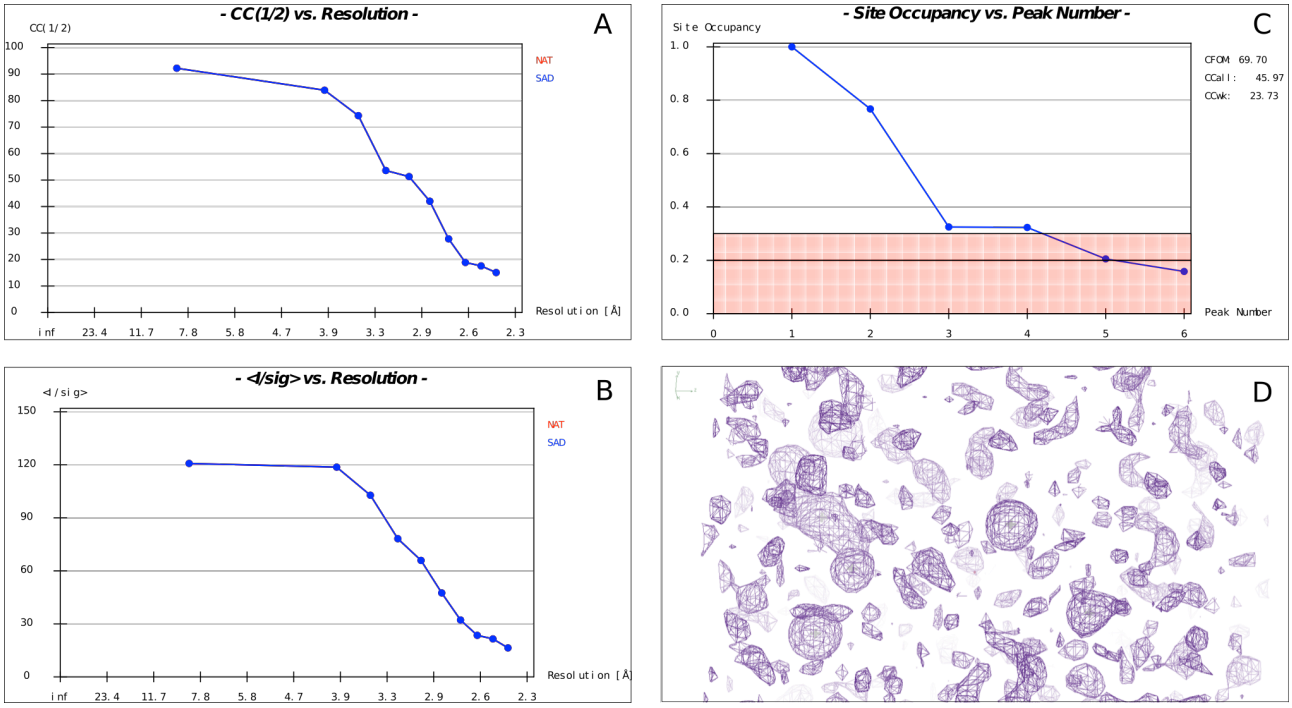


Fig. 22: Iodine substructure determination using SHELX C and D.
A and B: Plots of the resolution cut-off criteria *CC½* (A) and *I/σI* (B) vs. resolution as generated by SHELX C. Since only a derivatized data set was used for I-SAD (blue: SAD) no plot for the native data set (red: NAT) is shown.
C: Plot of site occupancy vs. peak number for the best substructure solution as generated by SCHLX D. From this analysis it gets obvious, that 4 electron density peaks are well above the threshold (shaded in red) and can therefore be attributed to I^--ions.
D: Electron density map contoured at 1,6 σ and coordinates of the I^- -substructure of the $P6_122$ enantiomorph as calculated for the derivatized data set.

Based on this substructure, initial phases for the experimental data were obtained and 50 cycles of density modification were iteratively executed for both space group enantiomorphs ($P6_122$ and $P6_522$) using the E package of the SHELX program suite. Based on the map contrast (contrast between protein regions and solvent regions in the map) and the map correlation coefficient together with the map connectivity in the protein region (Fig. 23A-C), the $P6_122$ enantiomorph turned out to be the correct one in this case. The phases obtained from SHELX-E were further extended, improved and

refined using the programs SHARP/autoSHARP (Vonrhein *et al.*, 2007) and PHENIX (Adams *et al.*, 2010). The initial model calculated this way from the derivatized data set was then used as a search model for MR-phasing of the native data set yielding the final molecular model for non-liganded MurU (Fig. 23D).

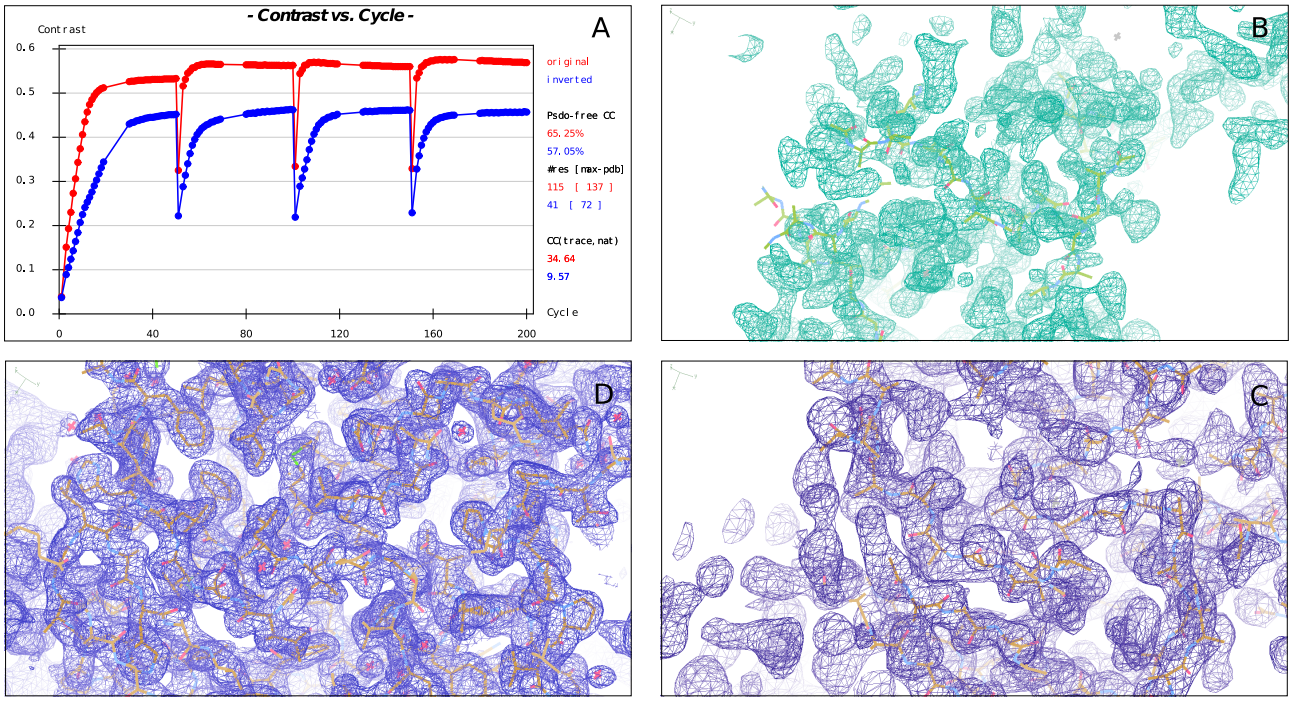


Fig. 23: Determination oft he correct space group enantiomorph.
A: Plot of the map contrast against density modification cycle (as generated by SHELX E). Note, that density modification was performed for both space group enantiomorphs ($P6_122$ and $P6_522$) in an iterative process until convergence executing 50 cycles of density modification per macrocycle.
B and C: Bias reduced electron density map contoured at 1,0 σ and initial structural model after density modification for the $P6_522$-enantiomorph (B, green map) and the $P6_122$ enantiomorph (A, purple-blue map). From the comparison of panels B and C it gets obvious, that the connectivity of the electron density map as well as the map contrast level is better for the $P6_122$ enantiomorph.
D: Bias reduced electron density map contoured at 1,0 σ and structural model for the finally refined non-liganded MurU. The red crosses indicate water molecules from the solvent. The comparison of panels C and D nicely visualizes how phases, and hence electron density maps, improve during refinement.

In this way the structure of non-liganded MurU could be solved at a nominal resolution of 1.8 Å (Fig. 23D and Table 1) and the model coordinates were deposited in the PDB and as been assigned with the PDB accession code (PDBid) 4Y7U.

The phases for the MurU enzyme-substrate complexes (see paragraph below as well as Table 1 and Table 2) were also determined using molecular replacement, this time with the final molecular model of the non-liganded enzyme as a search model. The coordinated of the two most relevant complexes were also deposited in the PDB (PDBid: 4Y7V and 4Y7T).

PDBid:	native dataset 4Y7T	NH_4I-soak 64	MurNAc-α1P + UppNHp 4Y7V complex #1	Mg^{2+} + MurNAc-α1P + UpNHpp 4Y7U complex #2	MurNAc-α1P + UppNHp complex #3	MurNAc+ UDP complex #4
Data collection						
Beamline	ID14-4, ESRF	X06DA, SLS PXIII	X06DA, SLS PXIII	X06DA, SLS PXIII	X06DA, SLS PXIII	X06DA, SLS PXIII
Wavelength [Å]	1,0	1,7	1,0	1,0	1,0	1,0
Spacegroup	P 6,22	P 6,22	P 6,22	P 6,22	P 6,22	P 6,22
Cell dimensions						
a; b; c [Å]	72,61; 72,61; 158,47	72,71; 72,71; 159,01	72,35, 72,35, 162,48	72,75; 72,75; 163,13	72,75, 72,75 159,43	73,08, 73,08, 158,74
α; β; γ [°]	90,0; 90,0; 120,0	90,0; 90,0; 120,0	90,0; 90,0; 120,0	90,0; 90,0; 120,0	90,0; 90,0; 120,0	90,0; 90,0; 120,0
Resolution (Å)	50,00 – 1,80 **(1.92 - 1.80)**	30,0 - 2,4 (2,46 – 2,40)	50,0 – 1,8 **(2,08 - 1,8)**	50,0 - 1,7 **(1,97 - 1,70)**	50,0 - 1,9 (1,95– 1,90)	50,0 – 2,0 (2,05 –2,00)
I/σI	13,30 (2,02)	44,26 (9,35)	27,07 (2,03)	30,72 (2,08)	19,85 (0,86)	24,59 (0,63)
I/σI anomalous signal	- -	1,80 (0,84)	- -	- -	- -	- -
CC ½ [%]	99,9 (88,9)	100,0 (98,4)	100,0 (80,6)	100,0 (84,2)	100,0 (26,0)	100,0 (30,5)
R_{meas} [%]	22,4 (804,6)	7,9 (52,3)	9,3 (236,9)	6,4 (200,2)	13,8 (453,1)	9,4 (341,2)
R_{pim} [%]	4,1 (149,6)	1,2 (8,4)	1,5 (38,0)	1,1 (33,3)	2,3 (87,8)	2,2 (110,0)
Completeness [%]	99,9 (99,6)	99,9 (100,0)	99,9 (99,9)	100,0 (100,0)	100,0 (100,0)	94,4 (99,5)
Measured reflections	704884 (117173)	741478 (53140)	916020 (322357)	1044585 (364931)	726316 (39214)	305382 (12253)
Unique reflections	23701 (4051)	18399 (1362)	24133 (8273)	28942 (10119)	20471 (1472)	16718 (1273)
Redundancy	29,74 (28,92)	40,30 (39,02)	37,96 (38,96)	36,09 (36,06)	35,48 (26,64)	18,27 (9,63)
Wilson B [Å^2]	44,25	45,23	47,13	46,75	46,98	49,53
Refinement						
Resolution [Å]	40,45 – 1,80	-	49,62 – 1,80	49,86 - 1,70	49,43 - 1,90	40,62 - 1,99
R_{work} / R_{free} [%]	21,14 / 24,55	-	21,49 / 25,37	22,78 / 26,40	21,96 / 26,48	22,46 / 27,24
No. of atoms						
Protein	1683	-	1652	1702	1698	1686
H_2O	80	-	67	54	45	80
Ligands:						
MurNAc / -1P	-	-	41	24**	41	20
UxxP / UDP	-	-	14	29**	29	72
SO_4^{2-}	20	-	5	5	10	10
Mg^{2+} / Na^+ / NH_4^+	-	-	-	2	5	-
Glycerol	42	-	28	28	32	28
B-factors [Å^2]						
Protein	46,68	-	51,84	55,95	52,88	56,85
H_2O	49,96	-	56,86	53,13	51,24	56,83
Ligands:						
MurNAc / -1P	-	-	48,71	68,05	86,90	49,08
UxxP / UDP	-	-	86,05	106,52	108,32	72,69
Mg^{2+} / Na^+ / NH_4^+	-	-	-	68,95	65,54	
SO_4^{2-}	88,38	-	94,51	105,33	141,71	112,02
Glycerol	48,38	-	57,29	71,61	48,66	84,64
r.m.s.** deviations						
Bond length [Å]	0,015	-	0,007	0,008	0,005	0,009
Bond angles [°]	1,526	-	1,074	1,192	1,002	1,348
Ramachandran plot						
Favored regions [%]	94,17	-	98,17	93,04	94,3	94,71
Allowed regions [%]	4,93	-	1,83	6,09	4,82	3,96
Outliers [%]	0,90	-	0,00	0,87	0,88	1,32

* values in parantheses refer to the highest resolution shell

** ligands refined without hydrogens

Table 1: Data collection and refinement statistics for MurU data sets

For complex and data set identifiers please refer to Table 2. For those structures that have been deposited in the PDB the respective PDBid is given.

4.3.3 Model building of MurU

To further investigate how the MurU binds and interacts with its substrates and to elucidate its reaction mechanism the structures of several enzyme-substrate / substrate-analog complexes were also determined. For this purpose diffraction data were collected from crystals soaked with various substrate (-analogs) combinations at different concentrations and for various soaking times. Table 2 lists all combinations that yielded a complete data set and assigns identifiers to the respective complex structures, which will be used throughout this work. Of all enzyme-substrate complex structures, those of MurU in complex with (i) MurNAc-α1P and the β-γ non-hydrolysable UTP analog UppNHp (complex #1 in Table 1), and with (ii) MurNAc-α1P, the α-β non-hydrolysable UTP analog UpNHpp, and 20 mM $MgCl_2$ (complex #2 in Table 1) turned out to be the most valuable for substrate binding analysis, since these two data sets yielded the most clearly defined unbiased difference electron density (mF_o-DF_c) for the substrates.

The final structural models contain all residues with the exception of some poorly ordered segments between G134 and H136 and around A154 that are missing in the structures of the native enzyme and complex #1 (Table 1) as well as the C-terminal His_6-tag that is missing in all structures. Complex #1 was used as the reference structure for structural comparison and 177 atoms of the native structure as well as 175 atoms of the complex #2 structure were aligned with rms deviations of 0.179 Å and 0.144 Å, respectively. For more experimental details on data collection and model quality see Table 1 and the pdb-vaild reports (section 11.1) of the data sets that have been deposited in the PDB.

The data used for structure determination were processed applying a high-resolution cutoff according to what has been established by Diederichs & Karplus in 2012 (Karplus & Diederichs, 2012). In this work the authors introduce *CC½* as a robust and statistically informative quantity useful for defining the high-resolution cutoff in crystallography.

Diederichs & Karplus state, that there is no valid basis for applying the formerly applied criterion that data are not useful beyond a resolution where R_{meas} (or R_{merge} or R_{pim}) rises above ~0.6 (60%) and that the signal-to-noise ratio (the mean $I/\sigma I$) is not a good cutoff criterion as well since there is the mean σI values can be mismatched. Referring to this, the significance of CC½ that, according to Diderichs & Karplus, is given at values of 10.0 (0.1%) was used as high-resolution cutoff criterion for our data sets. This however causes comparably high values for R_{meas} (or R_{merge} or R_{pim}) in high-resolution shells. Nevertheless these resolution bins still contain significant diffraction signals. Thus the values R_{meas} and R_{pim} in Table 1 are merely listed for the sake of completeness but had no influence on our decisions in data processing. Nevertheless, to validate the high-resolution cutoff applied, data were reprocessed using (i) $R_{meas} \leq 0.6$ (60%) and (ii) $I/\sigma I \geq 2.0$ as a high-resolution cutoff (Table 3). Refinement of the respective data set against the MurU molecular model revealed, that the difference between the Refinement R-factors R_{work} and R_{free} was significantly reduced when applying the significance of *CC½* as

cutoff (bottom line highlighted in green in Table 3). This in turn confirms, that that including the data from higher resolution shells actually improves the model, since the difference between R_{work} and R_{free} correlates with model quality in that the closer the two values are together the less overfit the model is (Karplus & Diederichs, 2012).

loop no.	x-tal identifier	crystallization condition	cryo-protectant	soaking component(s)	soaking time
2560	**native**	**0,1 M MES pH 6,5 + 1,3 M AS**	**3.06M AS + 10%Glycerol**	**-**	**-**
67	**NH_4I-soak**	**0,1 M MES pH 6,5 + 1,3 M AS**	**3.3M AS, 5% Glycreol; 0.1M MES pH 6.5**	**250mM NH_4I (for I-SAD phasing)**	**10 min**
3928	**complex # 1**	**0,1 M MES pH 6,0, 1,5 M AS**	**0.1 M MES pH 6, 15% Glycerol, 2.3 mM M1P + 3.3 mM UppNHp**	**2.3 mM MurNAc-1P + 3.3 mM UppNHp**	**75 min**
1633	**complex # 2**	**0,1M HEPES pH 6,4 + 1,1 M AS**	**0,8 M AS, 80 mM MES pH 6,0, 16 mM MgCl2, 20% Glycerol**	**2,8 mM UpNHpp; 2,8 mM M1P; 20 mM MgCl2 in 1M AS; 0,1M MES pH 6,0**	**2 min**
3296		**0,1 M HEPES pH 7.0, 1,4 M AS**	**1,0 2nd solution + 0,25 Glycerol**	**1st: 5,25 mM UpNHpp (1,0 M AS) 2nd: 2,52 mM M1P; 0,84 mM UpNHpp (1,5 M AS) final: 1,25 mM M1P + 3,67 mM UpNHpp**	**1st: 2 h 2nd: 22 h**
188		**0,1 M Tris pH 8,5 + 1,1 M AS**	**1,5 M AS + 15% Glycerol**	**100 mM MurNAc + 100 mM UDP**	**100'**
3077		0,1 M Citrate pH 5.0 + 1,1 M AS	0.5 UpNHpp 0.5 MurNAc in condition + 0.25 Glycerol (75%)	1st 5,25 mM UpNHpp + 2nd 5 mM MurNAc	1. 35' + 2. 90'
3291		0,1 M HEPES pH 6.4 1,0 M AS	1,0 2nd solution + 0,25 Glycerol	1st: 5,25 mM UpNHpp (1,5 M AS) 2nd: 2,52 mM M1P; 0,84 mM UpNHpp (1,5 M AS) final: 1,26 mM M1P + 3,04 mM UpNHpp	1st: 2:09 h 2nd: 22:20 h
64		0,1 M Tris pH 8,0 + 1,1 M AS	1,5 M AS + 15% Glycerol	100 mM MurNAc + 22,5 mM UppNHp	100 min
3297		0,1 M HEPES pH 7,5 + 1,0 M AS	1,5 M AS + 15% Glycerol	ca. 1,3 mM MurNAc-1P + 16,25 mM UppNHp	60 min
2560		0,1 M MES pH 6,0 + 1,5 M AS	0.1 M MES pH 6, 15% Glycerol, 2.3 mM M1P + 3.3 mM UppNHp	2.3 mM MurNAc-1P + 3.3 mM UppNHp	75 min
3055		0,1 HEPES pH 6,4 + 1,1 M AS	0.5 UpNHpp in condition + 0.25 Glycerol (75%)	5,25 mM UpNHpp	2:10 h
3036		0,1 HEPES pH 6,4 + 1,1 M AS	0.5 UpNHpp in condition + 0.25 Glycerol (75%)	50 mM UpNHpp	120 min
3058		0,1 M HEPES pH7,5 + 1,0 M AS	0.15 M HEPES pH 7, 1.5 M $(NH_4)_2SO_4$, 15% Glycerol	20 mM UMP in Cryo	o/n
3098		0,1 M HEPES pH7,5 + 1,0 M AS	0.15 M HEPES pH 7, 1.5 M $(NH_4)_2SO_4$, 15% Glycerol	100 mM UMP in Cryo	o/n

Table 2: Overview of all crystals of MurU that yielded a complete data set and could be processed.
Listed are the respective crystallization condition along with the cryo-protecting solution as well as the compounds used for soaking with their respective (effective) soaking concentrations and soaking times. For crystals that have been soaked sequentially the order in which the different solutions were added is indicated along with the respective soaking time. The data sets listed can be unambiguously identified by their 'loop no.' Those data sets that will be explicitly referred to in this work have additionally been assigned an 'x-tal identifier'.

Parameter	native dataset			complex #1: MurNAc-α1P + UppNHp			complex #2: Mg^{2+} + MurNAc-α1P + UpNHpp		
high-resolution cutoff	significance of CC½	I/σI ≥ 2.0	R_{meas} ≤ 60 %	significance of CC½	I/σI ≥ 2.0	R_{meas} ≤ 60 %	significance of CC½	I/σI ≥ 2.0	R_{meas} ≤ 60 %
Data collection									
Beamline	ID14-4, ESRF	ID14-4, ESRF	ID14-4, ESRF	X06DA, SLS PXIII	X06DA, SLS PXIII	X06DA, SLS PXIII	X06DA, SLS PXIII	X06DA, SLS PXIII	X06DA, SLS PXIII
Wavelength [Å]	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0
Spacegroup	P 6_122	P 6_122	P 6_122	P 6_122	P 6_122	P 6_122	P 6_122	P 6_122	P 6_122
Cell dimensions									
a; b; c [Å]	72,61; 72,61; 158,47	72,61; 72,61; 158,47	72,61; 72,61; 158,47	72,35, 72,35, 162,48	72,35, 72,35, 162,48	72,35, 72,35, 162,48	72,75; 72,75; 163,13	72,75; 72,75; 163,13	72,75; 72,75; 163,13
α; β; γ [°]	90,0; 90,0; 120,0	90,0; 90,0; 120,0	90,0; 90,0; 120,0	90,0; 90,0; 120,0	90,0; 90,0; 120,0	90,0; 90,0; 120,0	90,0; 90,0; 120,0	90,0; 90,0; 120,0	90,0; 90,0; 120,0
Resolution (Å)	50,00 – 1,80	50,0 - 1,97	50,0 - 2,55	50,0 – 1,8	50,0 - 2,06	50,0 - 2,25	50,0 - 1,7	50,0 - 1,93	50,0 - 2,00
	(1.92 - 1.80)	(2,02 - 2,97)	(2,62 - 2,55)	**(2,08 - 1,8)**	(2,11 - 2,06)	(2,31 - 2,25)	**(1,97 - 1,70)**	(1,98 - 1,93)	(2,05 - 2,00)
I/σI	13,30	16,49	26,00	27,07	39,45	49,45	30,72	43,44	47,63
	(2,02)	(4,11)	(13,34)	(2,03)	(4,22)	(7,83)	(2,08)	(4,26)	(5,85)
CC ½ [%]	99,90	99,90	99,90	100,00	100,00	100,00	100,00	100,00	100,00
	(88,90)	(97,60)	(98,10)	(80,60)	(88,90)	(97,60)	(84,20)	(93,10)	(96,80)
R_{mease} [%]	22,40	18,70	12,5	9,30	7,10	6,00	6,40	5,40	5,20
	(804,60)	(371,14)	(65,80)	(236,90)	(120,30)	(64,90)	(200,20)	(100,70)	(74,70)
R_{pim} [%]	4,11	3,42	2,33	1,51	1,16	0,98	1,07	0,90	0,87
	149,61	67,21	12,00	37,95	19,50	10,35	33,34	16,88	12,31
Completeness [%]	99,9	100	100	99,9	99,9	99,9	100,00	100,00	100,00
	(99,60)	(100,00)	(100,30)	(99,90)	(100,00)	(99,99)	(100,00)	(100,00)	(100,00)
Measured reflections	704884	544346	248886	916020	610272	471888	1044585	721340	651521
	(117173)	(40190)	(19014)	(322357)	(44154)	(35470)	(364931)	(51018)	(47749)
Unique reflections	23701	18243	8647	24133	16297	12621	28942	20004	18043
	(4051)	(1318)	(632)	(8273)	(1160)	(902)	(10119)	(1434)	(1297)
Redundancy	29,74	29,84	28,78	37,96	37,45	37,39	36,09	36,06	36,11
	(28,92)	(30,49)	(30,09)	(38,96)	(38,06)	(39,32)	(36,06)	(35,58)	(36,81)
Wilson B [$Å^2$]	44,25	46,59	50,68	47,13	48,46	50,06	46,75	46,94	47,38
Refinement									
Resolution [Å]	40,45 – 1,80	40,45 - 1,97	40,45 - 2,55	49,62 – 1,80	49,62 - 2,06	49,62 - 2,25	49,86 - 1,70	49,86 - 1,93	49,86 - 2,00
R_{work} [%]	21,14	22,5	20,74	21,49	20,25	19,43	22,78	22,07	22,72
R_{free} [%]	24,55	26,95	26,21	25,37	24,59	24,21	26,40	26,25	26,51
$\Delta R_{work} - R_{free}$ [%]	3,41	4,45	5,47	3,88	4,34	4,78	3,62	4,18	3,79

˙ values in parantheses refer to the highest resolution shell

Table 3 Data processing and refinement statistics for MurU data sets truncated using different high-resolution cutoffs.
The bottom line highlighted in green gives the differences between the refinement R-factors R_{work} and R_{free} with the smallest difference for each data set shaded in a darker green. Note that the smallest $\Delta R_{work} - R_{free}$ is achieved when applying CC as high-resolution cut-off criterion.

4.4 Overall structure or MurU

4.4.1 Domain organization

As expected, MurU features the general structural pattern of sugar nucleotidyltransferases, which basically consists of an elaborate (N-terminal) core domain that has a Rossmann-like fold and contains the catalytically active center and the dinucleotide binding site as well as a C-terminal auxiliary domain (Singh *et al.*, 2012). According to the conserved domain database (CDD) of the BLAST® web server, MurU's core domain can be classified as 'NTP-transferase-like-1' (CDD-accession number: cd06422). It is composed of a α/β/α-sandwich with a central, twisted, mixed β-sheet made up of 7 β-strands that is flanked on both sides by a total of 8 α-helices. A 9th α-helix is c-terminally protruding from the core domain. This 9th helix basically makes up the whole C-terminal domain, which, in comparison to structurally related enzymes, is only rudimentary present in MurU (Fig. 24) and which could not be assigned any auxiliary function. (Gisin *et al.*, 2013)

With the help of the enzyme-substrate complex structures, the catalytically active site of MurU could be located at a solvent exposed cleft, the bottom of which is formed by one edge of the central β-sheet and which is lined by the α-helices H5 and H8 on one side as well as the N-terminal amino acids 1-13 on the other (Fig. 24). This N-terminal section, which also comprises strand β1, is widely conserved among SNTs and carries the canonical signature motif GxGxR (G9-R13 in MurU, see Fig. 24) that has been referred to previously (Singh *et al.*, 2012). The third side of the substrate-binding cleft is flanked by a subdomain, which is inserted between strands β5 and β6 of the central β-sheet. This subdomain consists of a 3-stranded antiparallel β-sheet (strands β8-β10) that is connected to β5 by a highly flexible loop that was not fully visible in any of our crystal-structures (Fig. 24). Functionally, this subdomain is involved in the coordination of MurNAc-α1P - or generally the sugar moiety in structurally related enzymes - and is thus often referred to as the 'sugar-binding domain' (see also (Blankenfeldt *et al.*, 2000)).

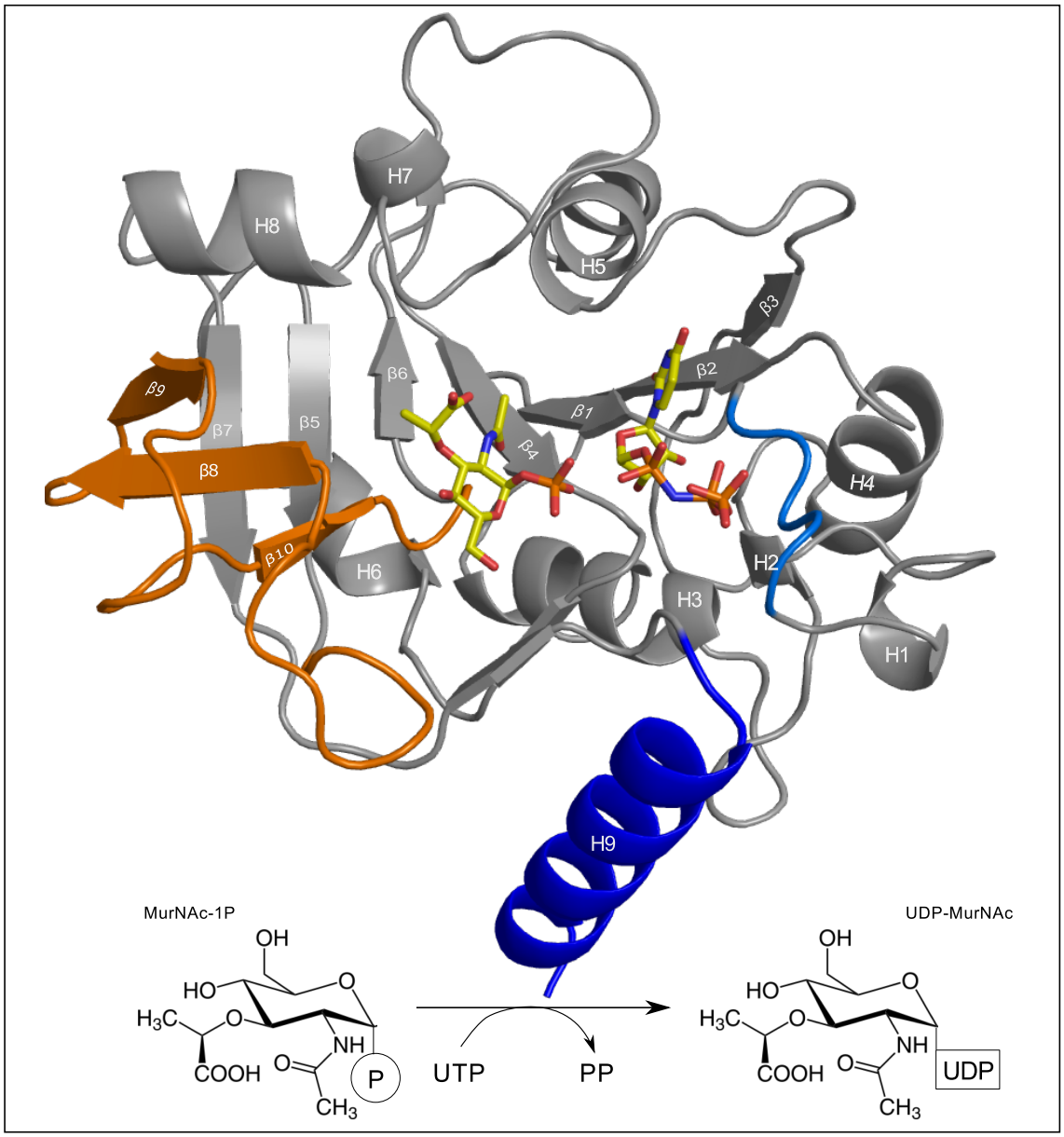


Fig. 24: Overall structure of a ternary complex of MurU, its substrate MurNAc-α1P, and the substrate analog UppNHp.
Substrates are depicted as sticks and colored according to the atom type (oxygens in red, nitrogens in blue, carbons in yellow and phosphorus in orange). The protein is shown in cartoon representation, with the colors highlighting the subdomains referred to in the text (blue: C-terminal Helix, orange: sugar-binding domain, marine-blue: SNT signature motif). Secondary structural elements are numbered from N- to C- terminus with H indicating α-helices and β indicating β-strands.
Bottom: Schematic representation of the reaction catalyzed by MurU.
Note, that ligand positions are based on diffraction data from two different crystals. The coordinates for the protein chain as well as for MurNAc-α1P are derived from the data set of complex #1, whereas those for UpNHpp originate from complex #2 (see Table 1 and Table 2).

4.5 The MurNAc-α1P binding site of MurU

The data sets for complex #1 and complex #2 (see Table 1 and Table 2) revealed very clear, unbiased difference electron density (mF_{obs}-DF_{calc}) for MurNAc-α1P, and thus this ligand could be modeled with confidence (Fig. 25). The MurNAc-α1P sugar ring assumes a $^4C_1$ chair conformation in all our MurU structures, which is the usual stable sugar ring configuration. MurNAc-α1P forms hydrogen bonds to the side chain carboxylates of D205 and D140, as well as polar contacts to the side chain of N105 (Fig. 25A). Furthermore, there seems to be a polar interaction to the backbone carbonyl of F160 (for the sake of clarity not explicitly shown in Fig. 25). The side chains of F141 and F160 provide a hydrophobic patch that accommodates the hydrophobic portion of the MurNAc-α1P lactyl moiety and the sugar ring. The side chains of L181 and L185, which are facing towards the lactyl-CH_3, also contribute to this interaction (Fig. 25A+B). The carboxyl group of the lactyl moiety is accommodated in the vicinity of K180 in a solvent-filled groove displaying a predominantly positive surface potential (Fig. 25C). The structures of the two complexes also contain a bound sulfate ion that was presumably incorporated from the crystallization solution. This sulfate binds in the vicinity of the MurNAc-α1P lactyl group (Fig. 25) and has elevated temperature factors (B-factors) in all structures, indicating low occupancy and/or high mobility.

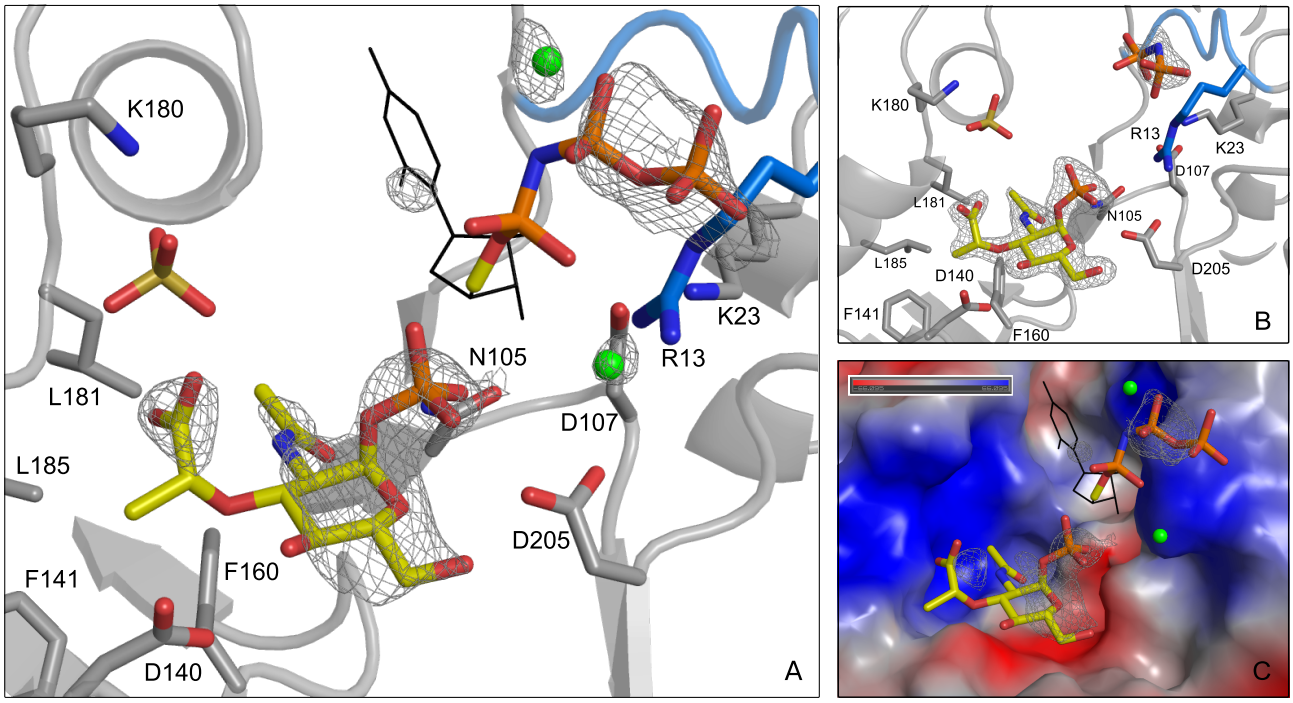


Fig. 25: (A,B) Crystal structures of MurU soaked with MurNAc-α1P + UppNHp and $MgCl_2$ (panel A, complex #2) and MurNAc-α1P + UppNHp (panel B, complex #1) with omit difference maps (mF_{obs}-DF_{calc}) for the ligands contoured at 2.5 σ.
Parts of the substrates for which omit difference electron density could be obtained are depicted as sticks, Mg^{2+}-ions as green spheres, and those parts of UpNHpp which no or only $2mF_{obs}$-DF_{calc}- density could be obtained for are represented as black lines. The SNT signature motif is highlighted in marine-blue. Side chains that are involved in ligand coordination are represented as sticks. Ligands and side chains are colored by atom type (oxygens in red, nitrogens in blue, phosphorus in orange, sulfur in gold, magnesium in green, and carbons in the colors used for the domains in figure 1.
(C) Surface representation of the substrate-binding site (complex #2) colored according to the surface potential (gradient from blue = positive to red = negative).
Note, that the parts of UpNHpp for which no reliable density could be obtained were modeled in the structurally most plausible conformation by relying on geometrical restraints and with the clear electron density for the pyrophosphate part as an anchor point

4.6 Coordination of PP / pNHp / P – the reaction's "leaving-group"

In contrast to MurNAc-α1P, the densities obtained for the uridyl-substrates were much weaker and did not allow for a confident placement of the uridyl ring. Nevertheless the pNHp portion of UppNHp (Fig. 25) or the pyrophosphate (pp) of UpNHpp (Fig. 25B), respectively, could be clearly localized. They are being coordinated by backbone nitrogens of residues 10-12 of the SNT signature motif as well as the side chain of K23 (Fig. 25A+B). The positive surface potential of this motif is significantly enhanced by the side chain of R13 and helps to compensate the negative charge of the pyrophosphate (Fig. 25C). The slightly different orientation that is adopted by the pNHp in comparison to the α-phosphate of UpNHpp can be attributed to the different geometrical environment of nitrogen compared to oxygen in the respective ligands.

In the non-liganded structure of MurU, a SO_4^{2-} is coordinated instead of the pyrophosphate at the respective site. However, compared to the native data set, the densities in the complexes are larger, and their elongated shape can not be explained by a spherical sulfate ion, thus clearly arguing for the presence of the pyrophosphate moiety of the non-hydrolysable substrate UpNHpp, or the pyrophosphate mimic pNHp respectively, being retained at the active site. This indicates that the positive surface potential of residues G9-R13 provides a preformed binding site for negatively charged moieties, which might be of catalytical relevance (Fig. 25).

As the uridyl moieties were not visible in either structure, their approximate location and orientation was modeled based on the visible portion of the nucleotides and geometrical restraints (black tracings in Fig. 25B and C). Independently performed molecular dynamics (MD) simulations (Fig. 26A) arrived at a conformation of the uridyl moiety that almost coincided with the orientations modeled, suggesting that the depicted conformation might be a good approximation of the structure that UTP actually adopts when bound to MurU. However it would not be valid to deduce any specific interactions with the protein from these models.

4.7 Coordination of catalytically relevant Mg^{2+} at the active site

An EDTA-induced enzymatic inactivation indicated that MurU requires Mg^{2+}, or equivalent cations, as a cofactor to achieve its full enzymatic activity (Gisin *et al.*, 2013). Therefore, 20 mM $MgCl_2$ were included in the soaking solution when solving the structure of complex #2. An unbiased difference electron density map (mF_o-DF_c) revealed two spherical features near the α-phosphate of MurNAc-α1P and the pyrophosphate moiety of UpNHpp that could either result from water molecules or from an ion of similar size (Fig. 25A+B). However, in MurU, MurNAc-α1P binding was obviously independent of the presence or absence of Mg^{2+} in the soaking solution, indicating that the Mg^{2+} is not required for substrate binding per se. To justify the modeling of Mg^{2+} in our structure, we analyzed the coordination environment at the putative cation positions (which will be termed position A / $_A Mg^{2+}$ and position B / $_B Mg^{2+}$ hereafter). At position A, the side chains of D205 and D107 (Fig. 25B) as well as the backbone nitrogen of G207 (not explicitly shown in Fig. 25B) would stabilize a putative $_A Mg^{2+}$. A Mg^{2+} in position B, in contrast, would not be coordinated by any amino acids of MurU but by the phosphates of the uridyl substrate. A similar coordination of divalent cations in a SNT has been described previously (Jagtap *et al.*, 2013). Moreover, MD simulations also predicted the presence of Mg^{2+} at the MurU active site. With the crystal structure as a starting model, the positions and radii of putative metal-ions could be calculated and optimized by keeping only the protein atoms close to their crystallographic positions through the application of a high force field constant on the protein

coordinates (Fig. 26B). Albeit not being suitable to distinguish between Mg^{2+}, NH_4^+ or Na^+ the ion positions predicted in this MD-approach were quite robust in different simulation scenarios and coincided nicely with the ion-positions deduced from the X-ray data. Taken together, it seems thus reasonable and plausible to model Mg^{2+}-ions at the two catalytically relevant positions A and B.

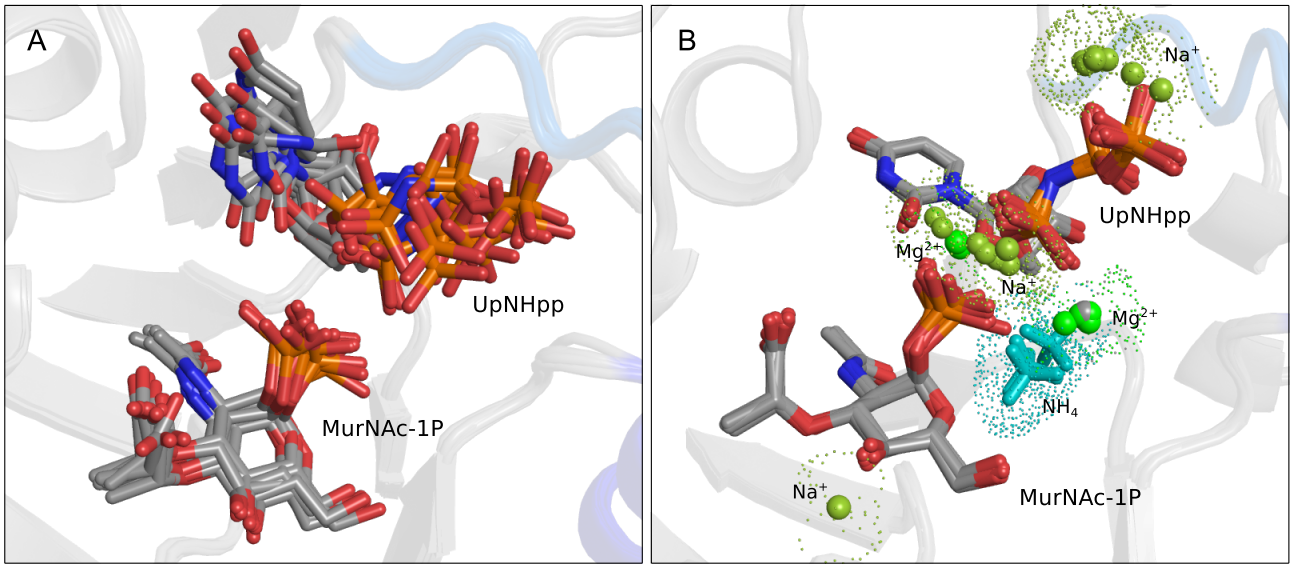


Fig. 26: Molecular dynamics simulations predicting ligand positions and conformations.
A) MD-simulation of the positions of UTP and MurNAc-α1P by applying a high force field constant to only the protein atoms. Metal ions were not included in this simulation scenario. B) MD-simulations of the positions of possible cations by restraining both protein and substrate atoms to their crystallographic positions. In this scenario Mg^{2+}, Na^+ and NH_4^+ were included., since these ions cannot be distinguished by their atomic radii.

4.8 Structural changes upon ligand binding

The comparison of the complex structures #1 and #2 with the non-liganded enzyme reveals no large domain rearrangements upon ligand binding. In particular, there is no significant movement of the C-terminal helix, and no reorientation of the β-sheet insertion between β5 and β6 (the 'sugar binding domain') relative to the core domain, which is in contrast to what has been observed for other SNTs with larger C-terminal domains (Pelissier *et al.*, 2010). However, there are some distinct side-chain reorientations of residues involved in MurNAc-α1P coordination upon substrate binding (Fig. 27).

The side chain D140 is rotated by almost 90° and reoriented towards the sugar substrate, thereby facilitating hydrogen bonding to C_4-OH of MurNAc-α1P (Fig. 27). Interestingly, this side chain has rather low B-factors in both its substrate-bound and its non-liganded state, indicating that the described movement is a switch between the two distinct states. The resulting hydrogen bond between D140 and the C_4-OH of MurNAc-α1P might thus be significant for substrate affinity.

Another, more subtle, reorientation takes place at the conserved R13 of the signature motif. Here, a slight reorientation of the R13-guanidinium group towards the pp-moiety of UpNHpp (or UppNHp, respectively) can be observed (Fig. 27). The B-factors observed for that residue's side chain were slightly higher than those of the neighboring residues in all structures, with the complex structures showing significantly higher B-factors for the guanidino-group and a steeper „B-factor-gradient" along the side chain than the non-liganded structure. This reflects a certain flexibility of R13, which argues in favor of a putative role in catalysis. A reorientation of this conserved arginine has also been reported for TmGMPase and seems to have a catalytical relevance (Pelissier *et al.*, 2010). Together with K23, the residues from the signature motive, and the Mg^{2+} in position B, R13 compensates the negative charge of UTP and thus contributes to the above-mentioned coordination of the pyrophosphate-leaving group. Furthermore R13 might also play a role in the stabilization of the reaction's transition state.

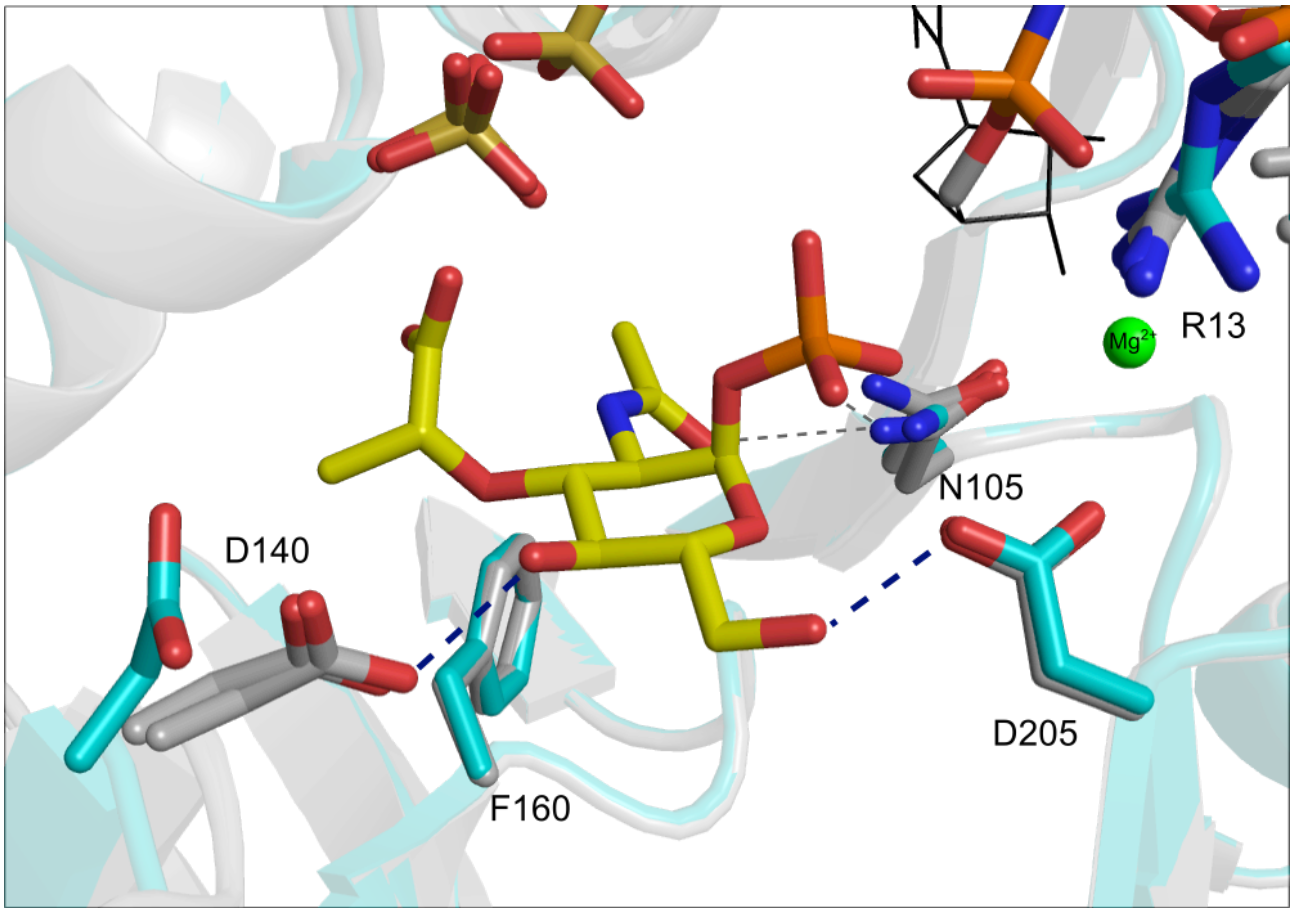


Fig. 27: Superposition of complex structures #1 and #2 (gray) with the native structure (cyan). The MurNAc-α1P, UppNHp and Mg^{2+} -ions are shown, together with a sulfate ion near K180 that is seen in all three structures. Substrates for which omit difference electron density could be obtained are depicted as sticks, Mg^{2+}-ions as green spheres, and those parts of UpNHpp which no or only $2mF_o$-DF_c- density could be obtained for are represented as black lines. Ligands and side chains involved in ligand coordination are represented as sticks and colored by atom type (oxygens in red, nitrogens in blue, phosphorous in orange, sulfur in gold, magnesium in green, and carbons in the colors used for the respective structure).
Note that D140 is rotated by about 90° towards the C_4-hydroxyl group of MurNAc-α1P in both ligand-bound structures.

4.9 Substrate binding in MurU and comparison to GlmU

Other residues involved in substrate binding, such as D205, F106, and N105 remain essentially unchanged upon substrate binding (Fig. 27). It appears that MurU provides a preformed binding pocket for MurNAc-α1P that requires no large domain rearrangements for substrate binding.

This is in accordance with the fact that, despite substrate binding is not affecting the domain arrangement of MurU, the reaction educts are bound to the enzyme similarly as it has been described for other SNTs. For example, the comparison of MurU with the N-acetylglucosamine-1-phosphate uridylyltransferase GlmU from *Mycobacterium tuberculosis* (GlmU^{Mtb},(Jagtap *et al.*, 2013)) reveals that the two enzymes bind their substrates in a similar orientation and position. In both cases, the respective sugar substrate is bound in a similar orientation and the P_β and P_γ of the uridylsubstrate are pointing away from the sugar-binding site to interact with the backbone nitrogen atoms of the signature motif (Fig. 25B+C, Fig. 28). Fig. 28 nicely shows that the sugar moiety of the reaction product UDP-GlcNAc (Fig. 28A) binds to GlmU in a very similar orientation as the educt Glc-1P (Fig. 28B) ((Jagtap *et al.*, 2013, Vithani *et al.*, 2014) and that this orientation closely resembles that of MurNAc-α1P bound to MurU. It has been proposed for GlmU^{Mtb} that the interactions made by the sugar phosphate moiety determine the enzyme's substrate binding mode (Zhang *et al.*, 2009) and judging from comparison drawn here it seems plausible that this is also true for MurU.

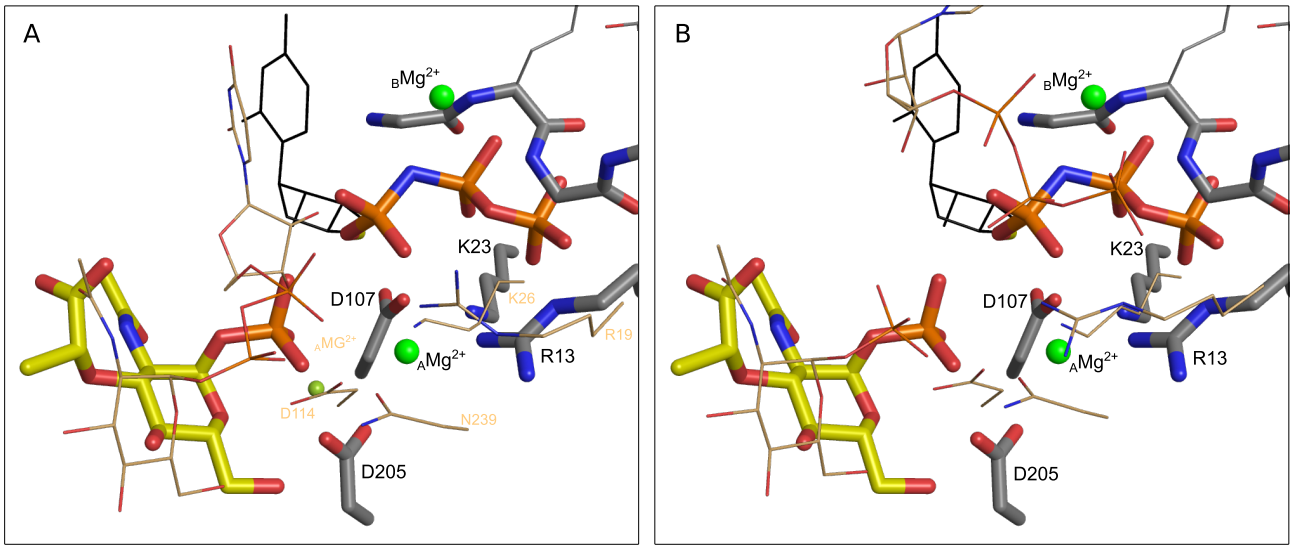


Fig. 28: Comparison of substrate binding in MurU and GlmU^{Mtb}
Overlay of substrate coordinates from MurNAc-α1P and UpNHpp bound to MurU (complex #2) with GlmU^{Mtb} in complex with A) UDP-GlcNAc (PDBid 3DJ4) and B) GlcNAc-1P + ATP (PDBid: 4K6R).
Note that in the structure shown in panel B ATP is bound to the enzyme instead of its physiological substrate UTP.

4.10 MurU orthologs – features and conservation

There are MurU orthologs in the human- or plantpathogenic γ-proteobacteria *Acinetobacter baumanii*, *Legionella pneumophila*, *Aeromonas hydrophila*, *Pseudomonas syringae*, *Xanthomonas oryzae*, in the pathogenic β-proteobacteria *Neisseria meningitides*, *Stenotrophomonas maltophila*, *Bordetella pertussis*, *Xylella fastidiosa*, and *Ralstonia solanacerum* as well as in the α-proteobacteria *Agrobacterium tumifaciens*, *Mesorhizobium loti*, and *Caulobacter crescentus* as well as in some *Bdellovibrio* and *Mycococcus* species.

4.10.1 Structural similarities to related sugar-nucleotidyltransferases

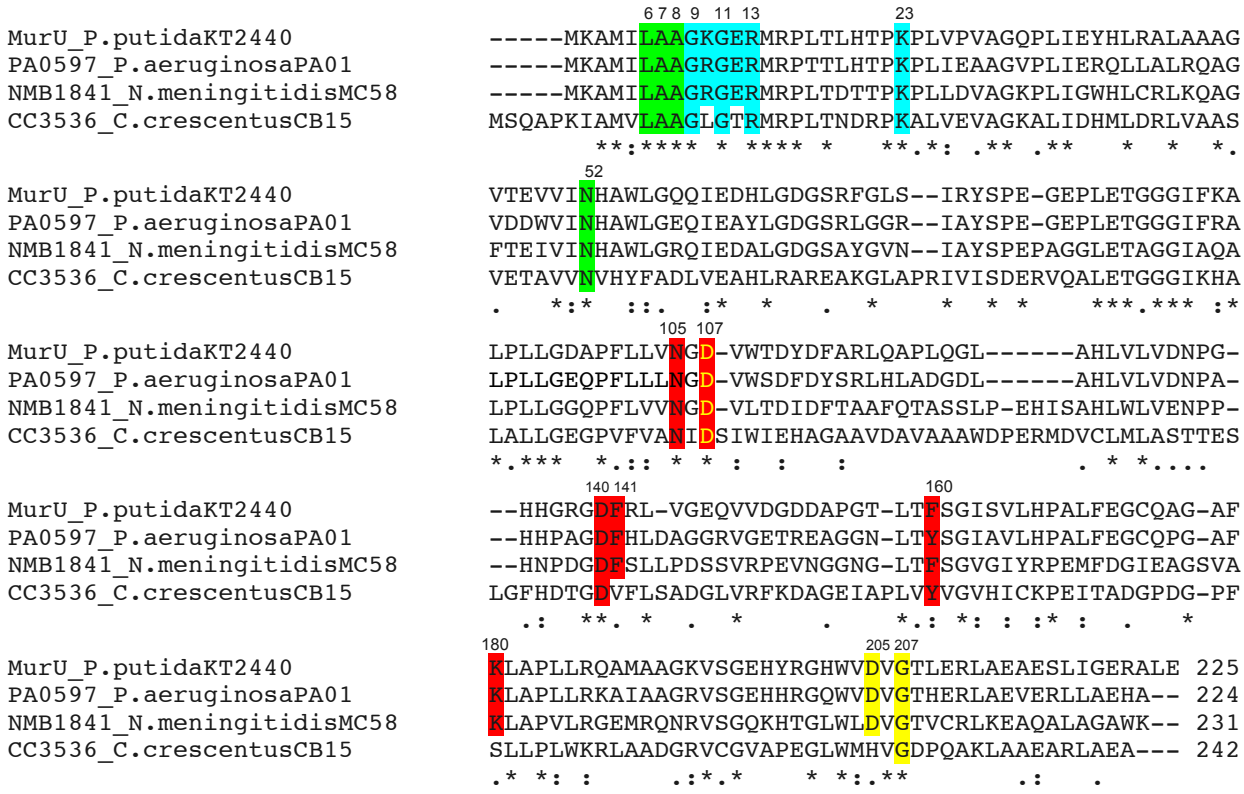


Fig. 29: Multiple sequence alignment of MurU with orthologs from different organisms: From top to bottom: *P.putida KT2440*, MurU; *P.aeruginosa PAO1*, putative nucleotidyl transferase (70.8% identity, 80.5% similarity), *N.meningitidis MC58*, mannose-1-phosphate guanylyltransferase (56.2% identity, 69.1% similarity); *C.crescentus CB15*, nucleotidyltransferase family protein (34.3% identity, 46.1% similarity); The alignment was performed with clustalw2 (https://www.ebi.ac.uk).
Residues predicted to be involved in substrate- or metal-binding by the CDD-search tool (provided under http://www.ncbi.nlm.nih.gov) are highlighted as follows: green: coordination of the NTP-substrate red: coordination of the sugar-1P substrate, yellow: coordination of metal ions; blue: SNT signature motif
Note, that D107 is predicted to be involved in both, the coordination of the sugar-1P substrate and metal ions.

Gisin et. al could show that MurU from *P.putida* as well as its orthologs from *P.aeruginosa PA01*, *N.meningitidis MC58* and *C.crescentus CB15* were all able to restore fosfomycin phenotypes in *murU*

mutants of *P.putidaKT2440* (Gisin *et al.*, 2013). These enzymes display quite high sequence identities to MurU ranging from 70,8 % (80,5 % similarity) for *P.aeruginosa PA01* over 45,0 % (56,2 % similarity) for *N.meningitidis MC58* to 34,3 % (46,1 % similarity) for *C.crescentus CB15*. A multiple sequence alignment and a BLAST®-search against the conserved domain database (CDD) revealed some highly conserved residues. Among these L6-R13 N52, N105 and D107 were assigned to contribute to ligand binding, whereas D107, D205 and G207 (all MurU-numbering) were classified as a metal binding feature by the CDD-search tool implemented in the search algorithm (See Fig. 29 and appendix section 11.4).

Since for none of the MurU orthologs an experimental 3D-structure was available, a direct structural comparison was not possible. Nevertheless, the BLAST®-results agree quite well with the structural data for MurU. From the structures presented above it can be stated, that N105 and D107 contribute to the coordination of the non-specific parts of the sugar substrate, whereas G9-R13 constitute the Nucleotidyltransferase signature motif and seem to be involved in the coordination of the UTP phosphate-part (Fig. 25). Additionally, the highly conserved K23 seems to contribute to the coordination of the phosphate-moiety as well. Residues L6-A8 and N52 on the other hand are potential candidates for coordinating uridyl-moiety of UTP (not explicitly shown). However, the experimental data for that part of the uridylsubstrate are insufficient. Thus this aspect cannot be satisfactorily clarified here.

4.10.2 MurU defines a 'minimal domain' for sugar-nucleotidyltransferase activity.

To better understand the functional relevance of some of MurU's structural features it is helpful to compare it to structurally and functionally related nucleotidyltransferases.

A structural similarity search using the DALI-server (Holm & Rosenstrom, 2010) revealed a significant consensus between MurU and several other SNTs. Of those, the glucose-1-phosphate thymidylyltransferase RmlA from *P.aeruginosa* (PDBid: 1G1L, (Blankenfeldt *et al.*, 2000)), the mannose-1-phosphate guanylyltransferase GMPase of *T.maritima* (PDBid: 2X5Z, (Pelissier *et al.*, 2010)) as well as the N-terminal SNT-domain of bifunctional GlcNAc-1P uridylyltransferase GlmU from *E.coli* (PDBid: 1FWY, (Brown *et al.*, 1999)) were chosen for structural comparison to MurU. This selection was predominantly made to cover a range of substrates that, taken together, feature the specificity requirements of MurU as closely as possible, and thus provide a helpful guideline to integrate the structural analysis of MurU into a broader context of SNTs. The comparison of the overall structures of *P. putida* MurU ($MurU^{P.p.}$), *P. aeruginosa* RmlA ($RmlA^{P.a.}$), *E .coli* GlmU ($GlmU^{E.c.}$) and *T. maritima* GMPase ($GMPase^{T.m.}$) clearly visualizes the common domain organization of SNTs (Fig. 30). All four enzyme-folds can be subdivided into an N-terminal and a C-terminal domain, with the N-

terminal domain housing the sugar-nucleotidyltransferase activity. The enzymes' N-terminal domains are similar in structure as they all feature nucleotide binding motifs showing a Rossmann-fold as well as a central seven-stranded β-sheet, with an additional 'sugar-binding' domain inserted between strands β5 and β6. The C-terminal part of MurU, however, is strikingly different from that of RmlA$^{P.a.}$, GlmU$^{E.c.}$ and GMPase$^{T.m}$. While in the latter three enzymes the C-terminal domain is rather large and projects quite prominently from the N-terminal nucleotidyltransferase domain, the corresponding region of MurU is small, essentially consisting of a single α-helix (Fig. 30).

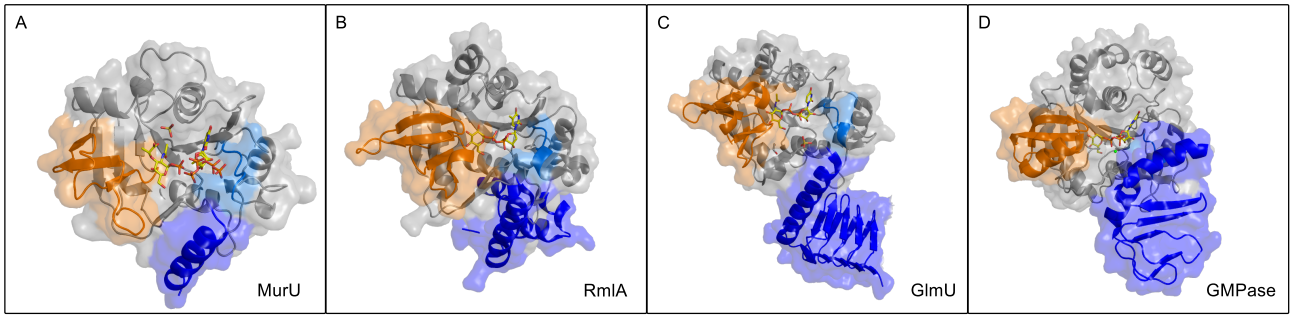


Fig. 30: Combined Cartoon representations of the MurU structure (same orientation as in Fig. 24) compared to those of the structurally related Glc-1P thymidylyltransferase RmlA, (*P. aeruginosa*, PDB-id: 1G1L), GlcNAc uridylyltransferase GlmU (*E.coli*, PDB-id: 1FWY) and Man-1P guanylyltransferase GMPase (*T.maritima*, PDB-id: 2X5Z) structures. All four proteins are shown with a semitransparent surface. Of the multimeric enzymes RmlA$^{P.a.}$, GlmU$^{E.c.}$ and GMPase$^{T.m}$ only one monomer is depicted. The coloring highlights similarities in domain organization. Blue: C-terminal domain, orange: sugar-binding domain, marine-blue: SNT signature motif. Note that, despite the obvious structural similarities, the sequence identities of these enzymes to MurU are quite low (RmlA: 22,6 %; GlmU: 16,6 %; GMPase: 18,7 %)

For several SNTs it is known, that the C-terminal domain is involved in the regulation of enzymatic activity by mediating oligomerization and/or by binding allosteric inhibitors (Singh *et al.*, 2012), and in some cases it may even influence substrate specificity (Pelissier *et al.*, 2010).

For example, GMPase$^{T.m}$, a monofunctional dimer, shows quite high selectivity for Man-1P, whereas the bifunctional PMI/GMPase of *P.furiosus* is rather promiscuous regarding its sugar-substrates. A C-terminally truncated version of the latter enzyme, however, shows a significantly higher specificity for its 'true' substrate Man-1P. (Pelissier *et al.*, 2010). The other two enzymes referenced are also multimers (GlmU$^{E.c.}$ is a trimer and GMPase$^{T.m}$ a tetramer) with their C-terminal domain mediating multimerization. By contrast, the monofunctional MurU, according to gel filtration profiles and enzyme activity tests, is monomeric in its physiologically active form (Fig. 18B), which means that a multimerization mediating (C-terminal) domain is not essential for this enzyme's functionality.

The role of the truncated C-terminal domain in MurU is unclear. No significant reorientation of the C-terminal helix α9 can be observed upon substrate binding, and no residues from α9 are involved in substrate binding or the coordination of the catalytically relevant Mg^{2+}-ions (Fig. 25, Fig. 27). It can thus be stated that, even though the C-terminal domain has some regulatory functions in related enzymes, MurU does not seem to require an 'elaborate' C-terminus for enzymatic activity. This defines

the N-terminal core domain as the 'minimal functional domain' required for MurNAc-α1P uridylyltransferase activity.

4.10.2.1 Binding cleft of MurU is more solvent exposed than in related SNTs

Another striking structural difference when comparing MurU's overall structure to related ones manifests at its the substrate binding cleft, which is much more solvent exposed than it is the case for example for RmlA$^{P.a.}$, GlmU$^{E.c.}$ and GMPase$^{T.m}$ (Fig. 30).

The structure of MurU reveals several flexible and poorly ordered regions. Two loop regions of the sugar binding domain (P133-R138 and D153-G156) as well as two additional loops that line the active site pocket (E79-P82 and G173-F179) either show elevated B-factors or cannot be modeled completely in some data sets (G134-H136 and around A154) (Fig. 31).

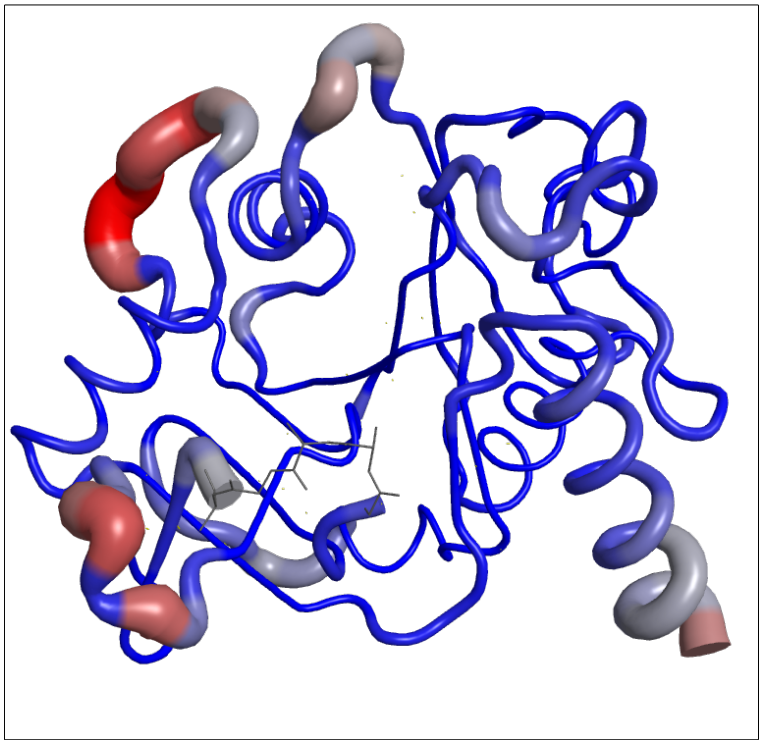


Fig. 31: Graphical representation of the crystallographic B-factors of the native structural model of MurU.
The blue-gray-red color spectrum ranges from a arbitrary value of 20 (deepest blue) to 100 (deepest red) and represents the B-factors of the amino acid backbone which is depicted in a B-Factor putty representation that also features the B-Factors by the width of the backbone trace (thin trace = low B-factors, thick trace = high B-factors). For the loop region between G134 and H136 no electron density could be observed in the native data set, thus these amino acids are depicted as gray lines.
Note that the minimum- and maximum values for the B-factor spectrum do not correspond to the minimal and maximal B-factors of the protein model, which are actually above respectively below these limits.

However, none of these high-B-factor regions seems to make a distinct subdomain-movement upon substrate binding (Fig. 27) as it has for example been described for the C-terminal domain of related enzymes (Pelissier *et al.*, 2010). Nevertheless it is interesting to note that especially the loops lining the active site pockets of RmlA$^{P.a.}$ (especially aa 187-197), GMPase$^{T.m.}$ and GlmU$^{E.c.}$ (especially aa 188-198) are obviously shielding the substrates much more from solvent than it is the case in MurU (Fig. 30).

5. Discussion MurU

5.1 Possible catalytic model

Extensive studies of GlmU-orthologs of various organisms have shown that the phosphoryl transfer reaction involves nucleophilic activation of the phosphoryl oxygen of the sugar substrate, which in turn attacks the α-phosphate of the NTP-substrate resulting in the formation of the NDP-sugar product and a concomitant release of pyrophosphate. This has been proposed to occur in a S_N2 reaction for GlmU, with the reaction proceeding through a pentavalent phosphorane involving two catalytically relevant Mg^{2+}-ions with distinct functions (Jagtap *et al.*, 2013). In this scenario, one Mg^{2+} ($_AMg^{2+}$) enables nucleophile activation and substrate coordination and is coordinated by distinct active site residues, whereas a second Mg^{2+} ($_BMg^{2+}$) contributes to the stabilization of the reaction's transition state. Due to its specific requirements for coordination geometry, $_BMg^{2+}$ promotes product formation by changing its coordination state and thus destabilizing the scissile bond (see Fig. 28 and (Jagtap *et al.*, 2013)). According to the structural data and the likely locations of the magnesium ions presented in this work, a similar mechanism seems also plausible for MurU (Fig. 32).

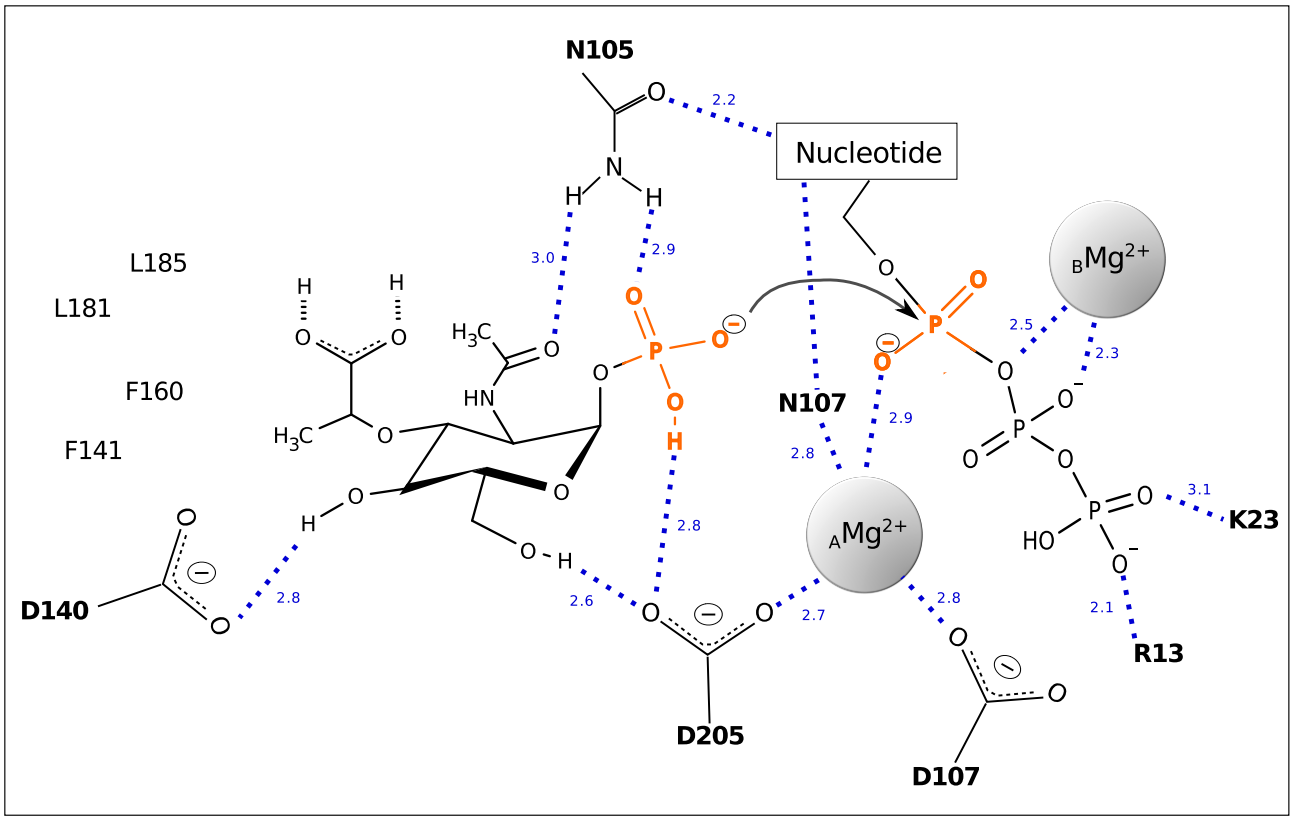


Fig. 32: Possible catalytic model for MurU
The reacting substrates and putative catalytic as well as coordinating residues are depicted together with the catalytically relevant Mg^{2+} ions $_AMg^{2+}$ and $_BMg^{2+}$. The dashed lines indicate polar contacts with the numbers indicating the respective interatomic distances in Å. The two reacting phosphates are highlighted in orange.

Thereby the extensive coordination of the leaving group of the UMP-transfer would result in a stabilization of the reaction product, which in turn destabilizes the scissile bond between P_α and P_β of UTP and thus renders it susceptible to a nucleophilic attack by the phosphate group of MurNAc-α1P. Such an attack might well be facilitated by the $_A Mg^{2+}$ ion in position A (Fig. 32), in analogy to what has been described for GlmU^{Mtb}. However, referring to our structure, it is possible that residue D205 not only coordinates the Mg^{2+}-ion, but that this interaction moreover enables the D205 carboxyl group to catalytically abstract a proton of the MurNAc-α1P phosphate group, which in turn initiates the nucleophilic attack (Fig. 32). However, at this point this remains mere speculation and further experiments will be needed to clarify the role of D205 in the UMP-transfer reaction.

All in all, it can be stated that the findings concerning a putative reaction mechanism of MurU are in good accordance to what has been proposed for other SNTs. The suggested mechanism follows the basic principle of a compensation of electrostatic repulsion between the two substrates, which facilitates their orientation along the reaction trajectory. This leads to a destabilization of the scissile bond and, concomitantly, to a stabilization of the reaction's transition state and products.

A destabilization of the scissile bond by distinct coordination of the leaving group of the phosphoryl transfer reaction has also been described for GlmU^{Mtb} (Jagtap *et al.*, 2013). In their structure of GlmU^{Mtb} Jagtab et al. could localize both reaction products bound to the active site. They argue that this represents a state of the enzyme immediately following the transfer reaction with particularly the pyrophosphate leaving-group being prevented from diffusing out of the active site by crystal packing. Jagtab et al. furthermore generalize and categorize the mechanistic principles of all SNTs that share a similar catalytic domain and catalyze a similar reaction. They subdivide SNTs into two groups that either employ two Mg^{2+} ions (group I) in catalysis or replace for Mg^{2+} by a Lys-residue (group II) (Jagtap *et al.*, 2013)). Based on a structural comparison of the active sites of MurU and GlmU, both enzymes (Fig. 28) can accordingly be considered to belong to group I and are thus likely to share a common mechanistic principle.

5.2 Sequence of substrate binding

Regardless, of MurU's rather solvent exposed substrate binding cleft, good electron density could be observed for MurNAc-α1P even at comparably low concentrations (see Table 2). In contrast, for the uridyl-moiety of the UTP-analogs the density obtained was generally less good and a structure of MurU with only UTP (or an analog) bound to the active site could not be obtained at all. Furthermore it has been shown previously for the related GlmU^{Mtb} that UTP alone does not bind to the enzyme unless the sugar-1P substrate is supplied along (Jagtap *et al.*, 2013).

It thus seems plausible that MurNAc-α1P is bound first to the enzyme's active center and UTP comes in second, together with the Mg^{2+} required to compensate for the negative charges, and that both Mg^{2+} and the pyrophosphate are released immediately after the uridyl transfer. This is further supported by similar findings in soaking experiments with GlmU^{Mtb} (Jagtap *et al.*, 2013, Zhang *et al.*, 2009).

5.3 MurU as minimal functional domain

MurU's strict specificity for MurNac-α1P (Gisin *et al.*, 2013) together with its rudimentary C-terminal domain leaves room for hypotheses about the influence of the C-terminus on substrate specificity. One could imagine the C-terminal domain to function as an auxiliary binding domain that holds in place even such sugars that are less well coordinated in the active site than the genuine sugar-substrate. Yet another scenario has been proposed by Pelissier et al. (Pelissier *et al.*, 2010). They suggest that the C-terminal domain might add flexibility to the binding pocket, thus allowing for a broader range of sugars to bind and orient along the reaction trajectory. Either scenario would provide an explanation for the exquisite selectivity of MurU with respect to its sugar substrate.

One could further imagine that the relatively easily accessible substrate-binding site might in some way compensate for MurU's rudimentary C-terminus. Since the solvent exposed binding pocket allows for the crucial Mg^{2+}-ions to freely access the enzyme's active site, an auxiliary function of the C-terminus in metal ion coordination, as it has been described for example for GMPase$^{T.m.}$ (Pelissier *et al.*, 2010) is not needed. However this cannot be directly concluded from the MurU structures presented here. Actually, due to the fact that the binding mode of the uridylsubstrate remains somewhat speculative, it cannot be ruled out, that a certain movement of MurU's C-terminal helix resembling the GMPase$^{T.m.}$ domain rearrangement might take place. Nevertheless, based on the structural data presented in this work, it can be stated, that MurU does not require an 'elaborate' C-terminal domain for the coordination of substrates and Mg^{2+}. Moreover, no significant reorientation of the C-terminal α-helix could be observed upon substrate binding (Fig. 27). Consequently, the N-terminal core domain of MurU can be considered as kind of the 'minimal functional domain' required for MurNAc-1P

uridylyltransferase activity. Nevertheless the question of whether or to what extend the C-terminal helix has to be included into this uridylyltransferase domain to confer functionality still has to be clarified.

With respect to the structural similarity among the active centers of SNTs the concept of the 'minimal functional domain' might be even generalizable for these enzymes. However, since the C-terminal domain has some regulatory functions - and in some cases influences the substrate promiscuity - in related enzymes, this hypothesis has to be treated with caution and needs to be validated.

5.4 Mg-soaks and enzyme activity in the crystal

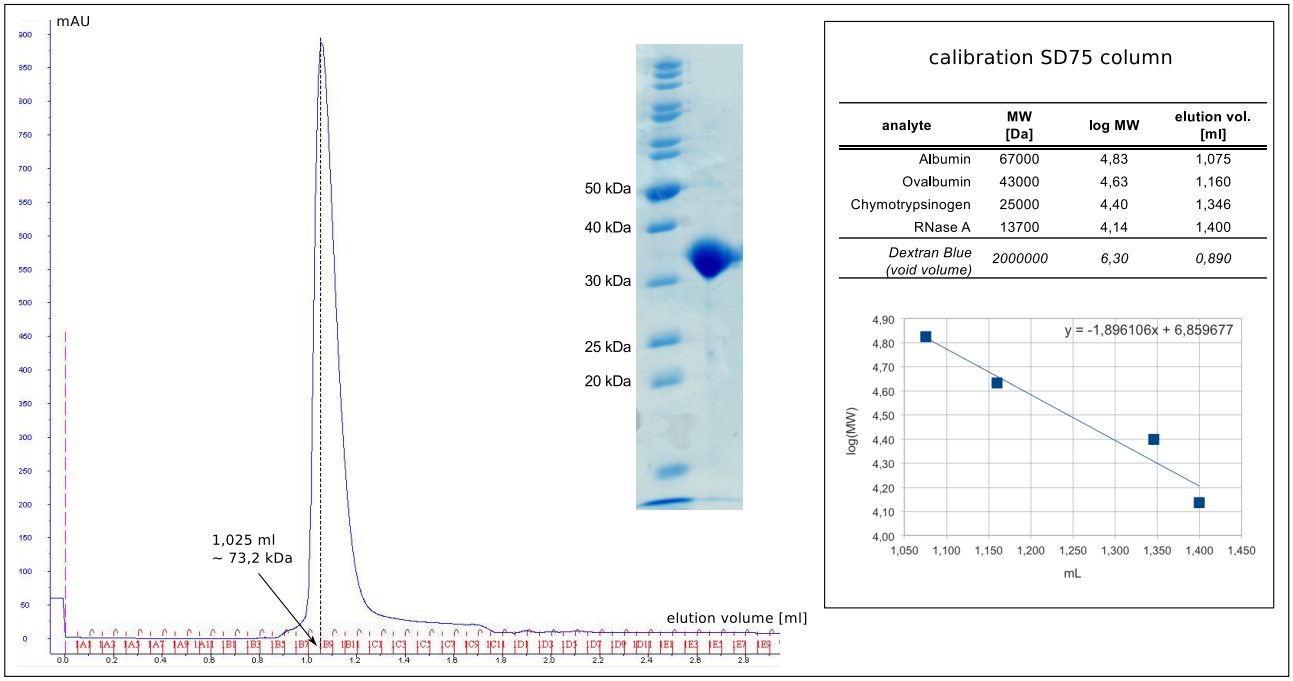
Obtaining a complex structure of MurU and its substrates with bound Mg^{2+} cofactor required a careful optimization of crystal soaking conditions. Crystals of the apoenzyme tended to break upon addition of Mg^{2+}-ions or their diffraction power was heavily impaired. In data sets from co-crystallization approaches no ligand density could be detected.

These difficulties can partially be explained by the fact that in the presence of Mg^{2+} the enzyme has its full activity, meaning that the reaction products are readily released and cannot be trapped in the crystal. This is especially true for the uridyl-component, which, judging from the quality of the observed electron density, seems to be less tightly coordinated. Furthermore the in- and out- diffusion of substrates, and possibly also a dynamic movement of the C-terminal helix or some flexible parts within the sugar domain, might cause a disruption in crystal packing resulting in cracks and/or a loss of diffraction. Nevertheless, the fragility of MurU crystals in the presence of Mg^{2+} is a clear hint that the enzyme is still active in the crystal, adding physiological significance to conclusions based on its crystal structure.

6. Results MurQ (YbbI)

6.1 Crystallization of MurQ

MurQ was crystallized as uncleaved His_6-fusion construct. Cloning, expression, and purification of MurQ was performed by Amanda Duckworth (AG Mayer, IMIT Uni Tübingen). Prior to crystallization each sample was analyzed on an analytical SD75 gel filtration column and via SDS-PAGE to confirm the protein to be homogeneous, pure and approximately of the expected molecular weight. The exemplary analytical SD75 gel filtration profile depicted in Fig. 33 shows that MurQ eluted approximately at the molecular weight of the dimer (2x 34,2 kDa) and SDS-PAGE analysis proved the corresponding crystallization sample to be pure and suitable for crystallization.



analyte	MW [Da]	log MW	elution vol. [ml]
Albumin	67000	4,83	1,075
Ovalbumin	43000	4,63	1,160
Chymotrypsinogen	25000	4,40	1,346
RNase A	13700	4,14	1,400
Dextran Blue (void volume)	*2000000*	*6,30*	*0,890*

Fig. 33: Elution profile of His_6-MurQ in an analytical gel filtration run using a SD75 column and SDS-PAGE analysis of the corresponding crystallization sample. Insert: Calibration-curve for the SD75 column used. For His_6-MurQ an elution volume of 1,025 ml was detected. This corresponds to a molecular weight of approximately 73,2 kDa, which is in good accordance to the theoretical molecular weight of the dimer(2x 34,2 kDa). SDS-PAGE analysis proved the corresponding sample to be reasonably pure for crystallization. Note that on an SDS-gel the protein is denatured and thus runs as a monomer approximately at its theoretical molecular weight of 34,2 kDa.
The table on the right lists the analytes used to calibrate the SD75 column with their molecular masses and the corresponding elution volumes. The linear fit below was used to calculate the molecular mass of His_6-MurQ from its elution volume.

An initial screen with a protein concentration of 8,6 mg/ml yielded crystals of MurQ in several conditions (see Fig. 34). Of those, the crystals grown in 0,1 M citrate pH 4 and 9,7% PEG 3350 diffracted best at the home source. Therefore fine screens to grow diffraction quality crystals were designed around this condition.

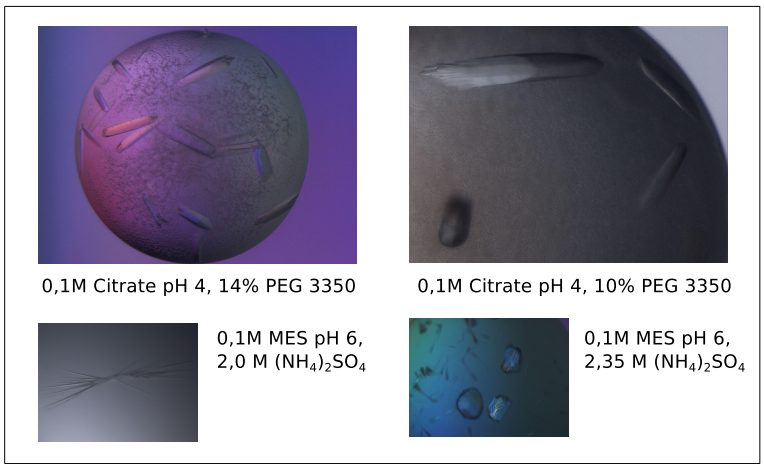


Fig. 34: Crystallization conditions for MurQ
Depicted are some typical crystal forms that appeared in an initial screen, which was set up with MurQ at a concentration of 8,6 mg/ml in Tris buffer.

All in all, data sets of about a dozen MurQ crystals were recorded, but the final structure determination was performed with the diffraction data from only one singe crystal that was grown in 0,1 M citrate pH 5, 9,7% PEG 3350 with a protein concentration of 8,6 mg/ml within 10 day at 20°C (Fig. 35).

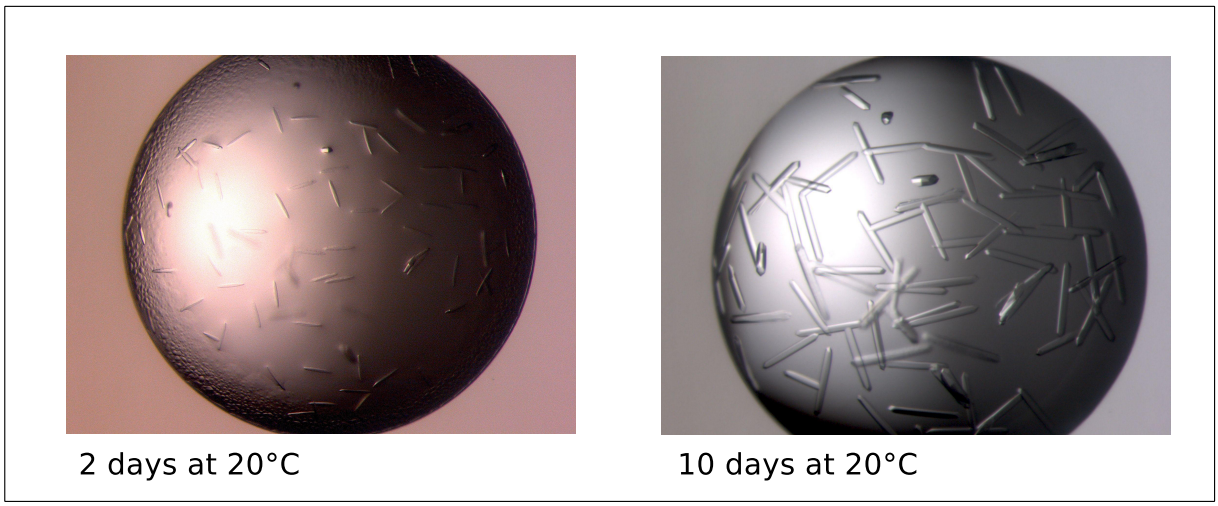


Fig. 35: Crystallization of MurQ
Crystals used for the structure determination of MurQ were grown over 10 days using 8,6 mg/ml MurQ in Tris buffer mixed 1:1 with a crystallization buffer containing 0,1M citrate pH 5,0 and 9,7% PEG 3350.

6.2 Structure determination of MurQ

6.2.1 Phase determination of MurQ

For the structure determination of MurQ diffraction data down to a nominal resolution of 2.27 Å were used. In this context an R_{meas} of ≤ 60 % has been applied as resolution cutoff criterion, since at that time in early 2012, CC½ has not yet been established.

The phases for the data set of wild type (wt) MurQ were derived via molecular replacement using the structure of the structurally related putative phosphosugar isomerase HI0754 from *H.influenzae* (PDBid: 1NRI) as a search model.

6.2.2 Model building and Refinement of MurQ

In the refined molecular model for MurQ all residues could be built according to the experimental electron density. Detailed statistical information on data collection and model quality are listed in Table 4 and the pdb-valid report (see section 11.1.4).

Data collection	
Parameter	**native dataset**
Data collection	
Beamline	SLS PXI
Wavelength [Å]	1.0
Spacegroup	P 3_221
Cell dimensions	
a; b; c [Å]	167,44 167,44 65,67
α; β; γ [°]	90,0 90,0 120,0
Resolution (Å)	50,0 - 2,27 (2,33 - 2,27)
I/σI	16,47 (4,03)
CC ½ [%]	99,8 (91,5)
R_{mease} [%]	8,0 (47,3)
Completeness [%]	99,2 (99,8)
Measured reflections	244753
Unique reflections	48658 (3576)
Redundancy	5,1 (5,1)
Wilson B [$Å^2$]	40,16

* values in parantheses refer to the highest resolution shell

Refinement	
Parameter	**native dataset**
Refinement	
Resolution [Å]	48,68 - 2,27
R_{work} / R_{free} [%]	14,69 / 18,02
No. of atoms	
Protein	6657
H_2O	460
Ligands:	
Citrate	13
total	7130
B-factors [$Å^2$]	
Protein	39,1
H_2O	105,4
Ligands:	
Citrate	65,2
r.m.s.** deviations	
Bond length [Å]	0,009
Bond angles [°]	1,165
Ramachandran plot	
Favored regions [%]	98,7
Allowed regions [%]	1,4
Outliers [%]	0

** root mean square

Table 4: Data collection and Refinement statistics for the native structure of MurQ.

6.2.3 Crystal packing of MurQ and spacegroup determination

All data from MurQ crystals were indexed and processed in spacegroup $P3_221$ (Space group no. 154) as assigned with the help of XDS and POINTLESS. MurQ crystallized with 3 protomers in the asymmetric unit (asu) (Fig. 36).

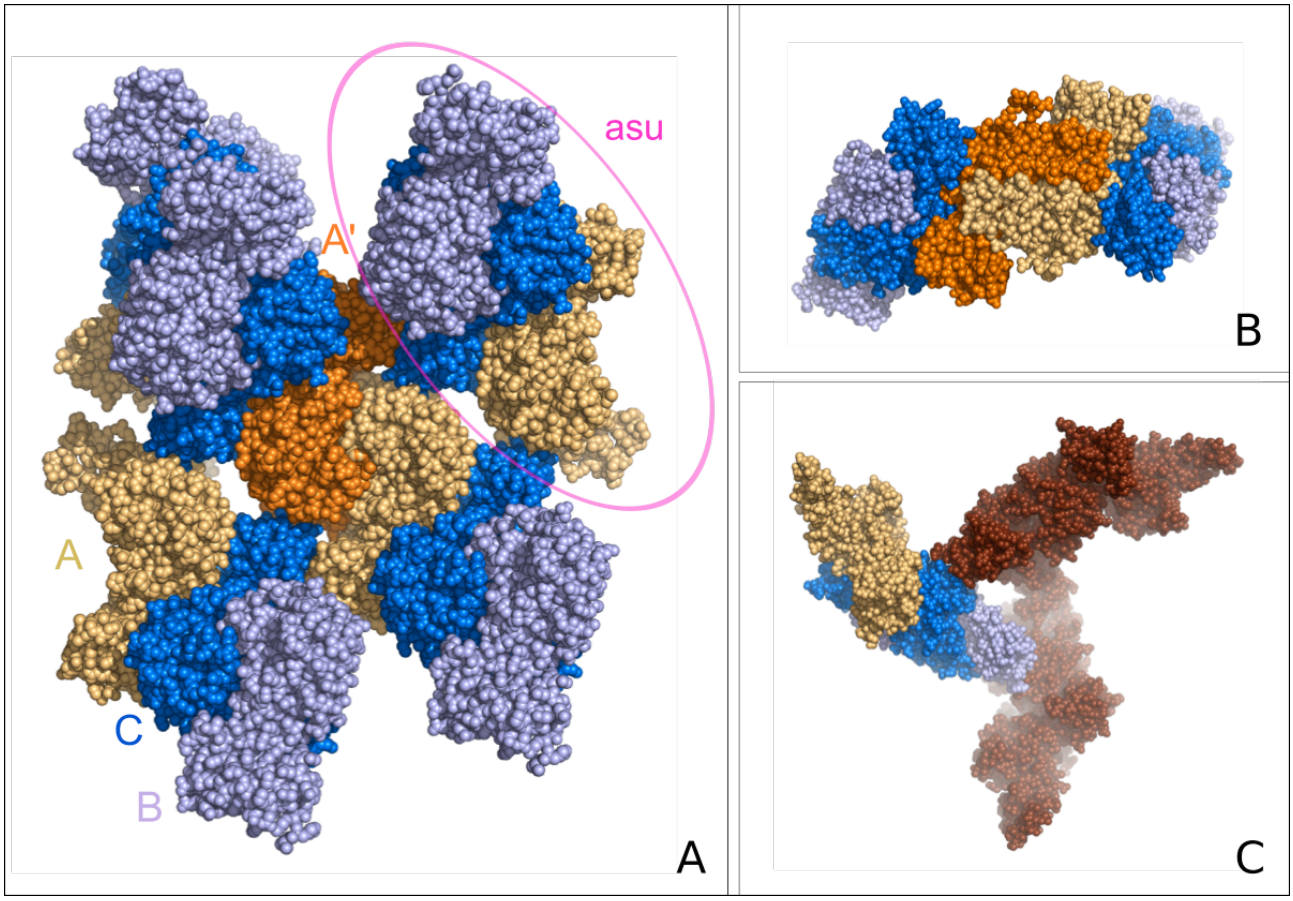
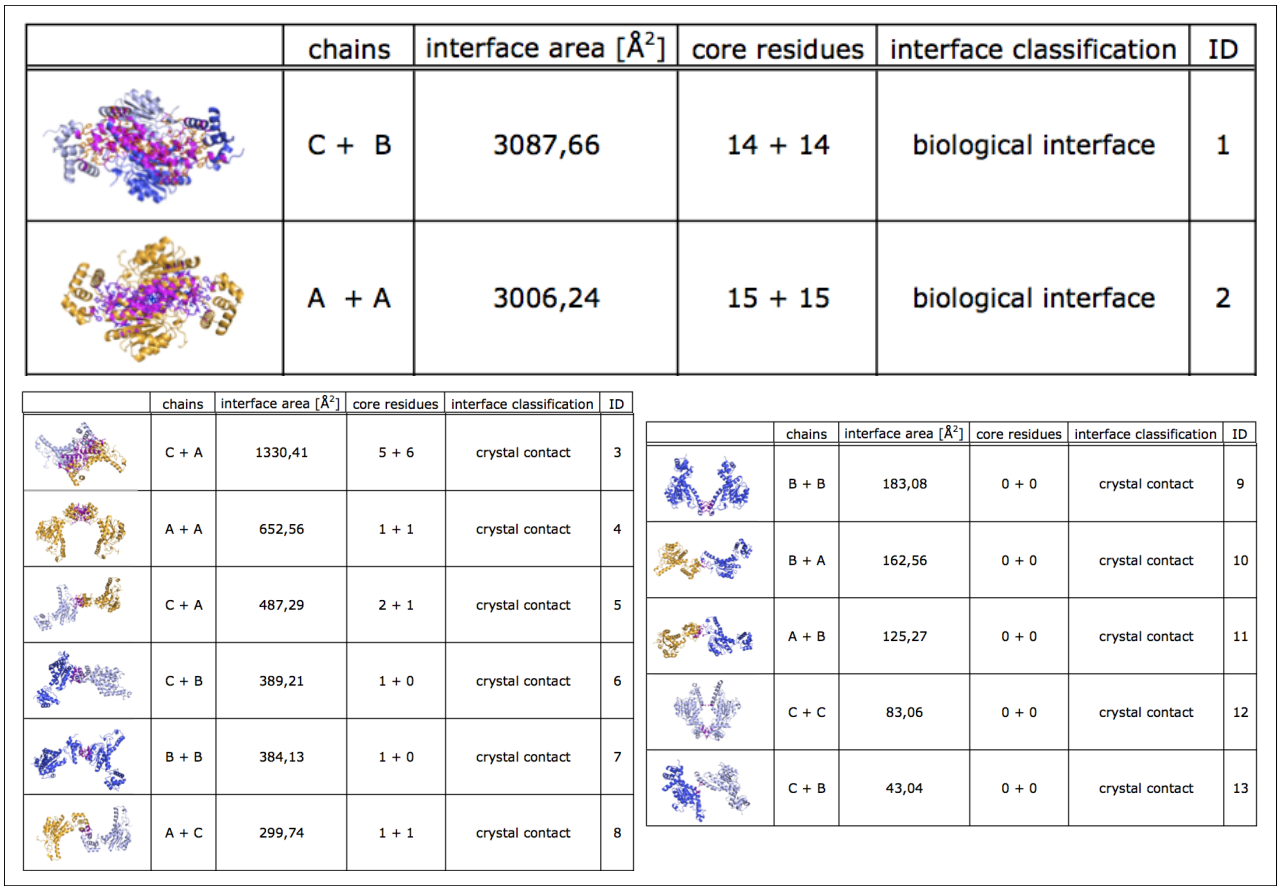


Fig. 36: Crystal packing of a typical MurQ crystal
A: Crystal packing of MurQ with 3 monomers per asu. The letters indicate the protein chain IDs used in the molecular model. Note, that the physiological dimers are AA and BC respectively. To highlight the AA-dimer one single chain A (A') is depicted in a darker orange. B: View along the 2-fold (crystal-) symmetry axis. The coloring is the same as in panel A. C: View along the 3-fold (crystal-) symmetry axis. One asu is colored as in A, the symmetry mates are colored in brown.

From a structural model based on a related structure (e.g. PDBid 1NRI) and its gel filtration profiles, MurQ's physiologically active form is known to be a homodimer, which means that the 3 protomers represent 1,5 functional units per asu. To further confirm the homodimeric nature of active MurQ, a surface contact analysis was performed using the EPPIC web server ((Duarte *et al.*, 2012) see Fig. 37). This analysis of all inter molecular contacts in the crystal revealed 13 contacting interfaces out of which only those between chain B and chain C as well as between two chains A were classified as biological interfaces. Moreover, the AA'-dimer and the BC-dimer could be aligned with an rms-

deviation of 0,265 Å (0,197 Å when only the Cα-atoms were aligned). This confirms, that the two dimers are functionally redundant and represent the biologically active unit and moreover implies that the dimer-interfaces (B-C and A-A') are the only biological interfaces in the crystal. Consequently, due to the crystal packing with 3 protomers in the asu, the two interfaces B-C and A-A' are redundant with chains B and C being linked by pseudo 2-fold non-crystallographic symmetry and chains A and A' from the neighboring asu by the 2-fold crystal symmetry axis (see Fig. 36B).



	chains	interface area [Å²]	core residues	interface classification	ID
	C + B	3087,66	14 + 14	biological interface	1
	A + A	3006,24	15 + 15	biological interface	2

	chains	interface area [Å²]	core residues	interface classification	ID
	C + A	1330,41	5 + 6	crystal contact	3
	A + A	652,56	1 + 1	crystal contact	4
	C + A	487,29	2 + 1	crystal contact	5
	C + B	389,21	1 + 0	crystal contact	6
	B + B	384,13	1 + 0	crystal contact	7
	A + C	299,74	1 + 1	crystal contact	8

	chains	interface area [Å²]	core residues	interface classification	ID
	B + B	183,08	0 + 0	crystal contact	9
	B + A	162,56	0 + 0	crystal contact	10
	A + B	125,27	0 + 0	crystal contact	11
	C + C	83,06	0 + 0	crystal contact	12
	C + B	43,04	0 + 0	crystal contact	13

Fig. 37: Crystal interface analysis using the EPPIC web server (Duarte *et al.*, 2012).
Depicted are all 13 contacting interfaces present in the MurQ crystal together with their respective surface area, the number of core residues involved in the inter chain contact and the classification according to the EPPIC-algorithm as biological interface or crystal contact. Note that, in this context, the term 'core residue' refers to a classification protocol applied through the EPPIC-algorithm for a detailed description of contacting interfaces

The Matthews probability of 0,71 for that crystal packing, calculated with Matthews Probability Calculator online tool, was convincing. Moreover, when processed in $P3_221$ the data yielded good phases and well interpretable electron density maps that could be readily refined confirming that the spacegroup assignment and data processing was performed correctly.

Overall structure of MurQ

The MurQ monomer consists of two distinct domains, a larger N-terminal core domain and a smaller C-terminal domain that is linked to the core by a loop region (Fig. 38A). Considering the topology of the N-terminal core domain of the MurQ monomer depicted in Fig. 38A, the enzyme can be clearly assigned to the family of mono-SIS domain proteins. It features the characteristic α-β-α sandwich structure containing a five-stranded parallel β-sheet that is flanked on both sides by a total 8 α-helices (see also Fig. 8).

Furthermore, in the structure that served as molecular replacement search model a significant portion of the C-terminal domain has been missing, whereas in this work the full-length structure of MurQ is presented. The analyzed crystals contain MurQ in its physiologically active form as a dimer as shown in Fig. 38C (and also in Fig. 36 and Fig. 37).

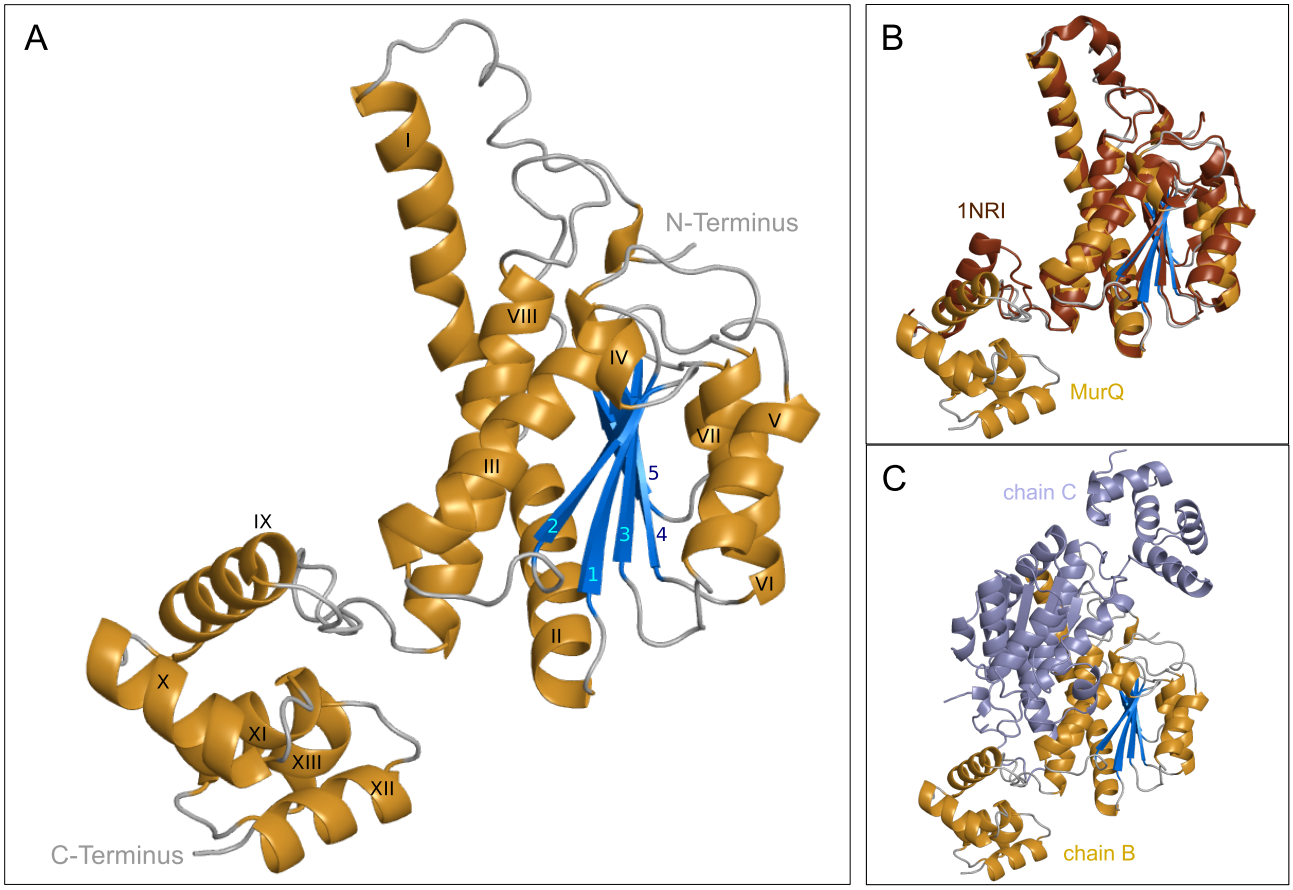


Fig. 38: Structural model of *B.subtilis* MurQ
A: Schematic representation of a MurQ$^{B.s.}$ monomer colored according to secondary structural elements.
The coloring and numbering of secondary structural elements is the same as in the schematic model for the SIS domain topology depicted in Fig. 8: α-helices: orange with roman numbers; β-strands: blue with arabic numbers; loops: gray.
B: Superposition of MurQ$^{B.s.}$ with the x-ray crystal structure of the putative phosphosugar isomerase from *H.influenzae* (PDBid: 1NRI) that has been used as a search model for molecular replacement.
C: The physiologically active dimer, represented by chains B (coloring as in panel A) and C (light blue).

6.3 Catalytically relevant residues and active sites of MurQ

Mutational studies have established, that residues E87 and E118 are essential for MurQ activity (A. Duckworth, AG Mayer; personal communication). Mutation of E87 (E87A) significantly reduces and a mutation of E118 (E118A) completely abolishes enzyme activity. These two residues are located at the physiological dimer interface in a way that brings E87 of one chain and E118 of the other in spatial proximity, suggesting that the surface cavity at this site functions as MurQ's substrate binding site. Due to the inherent twofold rotational symmetry of the MurQ homodimer (in Fig. 39 illustrated for chains B and C) the enzyme has two 'reciprocal' active sites at the respective symmetry related locations of the dimer interface (insets in Fig. 39). Further analysis of these active sites reveals two additional residues, D119 next to E118 and K237 of the same chain as E87, which are positioned in a way that suggests their potential participation in substrate binding or catalysis. However, so far, the influence of these residues on catalytic activity or substrate binding has not been investigated.

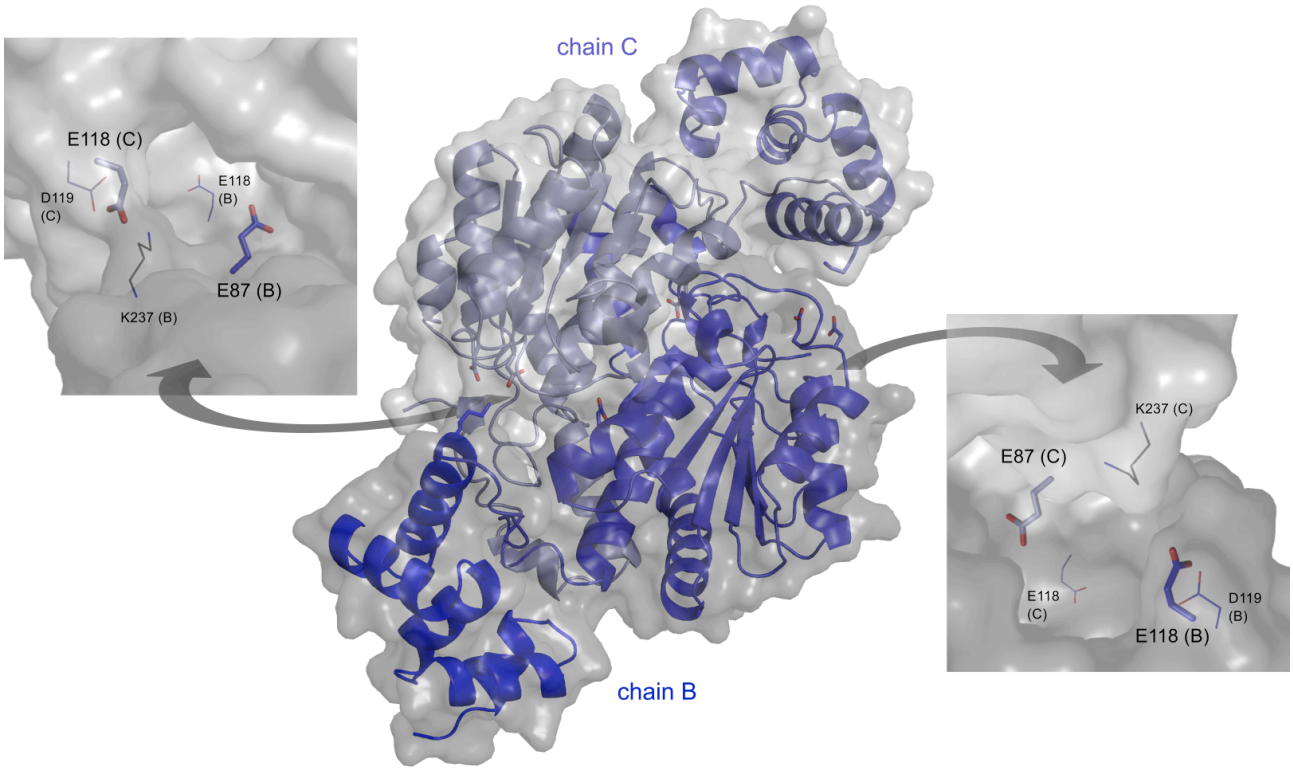


Fig. 39: MurQ's two active sites at the dimer Interface
Combined cartoon- and surface representation of the BC-dimer (chain B (blue on dark gray surface; chain C: light blue on light gray surface). The two insets show a close-up view of the two reciprocal active sites at the dimer interface from a perspective looking into the active site pocket, either from the C-terminal domain onto the core domain (left inset) or vice versa (right inset). The glutamates relevant for enzyme activity (E87 and E118) are represented as sticks; all other residues discussed in the text are depicted as lines. Colors are according to atom type (red: oxygen, blue: nitrogen) or reflect the chain identity (carbon).

The loop region linking the C-terminal domain to the core displays elevated B factors compared to the rest of the protein. This implicates mobility of the C-terminal domain relative to the N-terminal core that contains the SIS-motif and thus the enzyme's substrate binding site. Considering this, a certain rearrangement of the two domains upon substrate binding seems possible, which might contribute to a mutual allosteric relation between the two active sites regulating enzyme activity.

7. Discussion MurQ

7.1 Comparison with MurQ from *H.influenzae*

In November 2013, a new structure of MurQ from *H.influenzae* ($MurQ^{H.i.}$) was determined (Hadi *et al.*, 2013). This enzyme has been known previously as the uncharacterized putative etherase YfeU, the structure of which (PDBid: 1NRI) has served as a molecular replacement search model for the data sets of $MurQ^{B.s.}$ presented in this work. The structure of $MurQ^{H.i.}$ has been published in complex with a ring-opened substrate analog and satisfactorily explains the reaction mechanism of MurQ-like etherases. In contrast to what has been proposed in 2008 (Hadi *et al.*, 2008), Hadi et al. now assume that MurQ-like etherases apply a one base mechanism with a Glu residue (E89, *H.influenzae* numbering) functioning to both deprotonate at the C2 position and assist the departure of the lactyl ether at C3 position (see Fig. 40 and (Hadi *et al.*, 2013)).

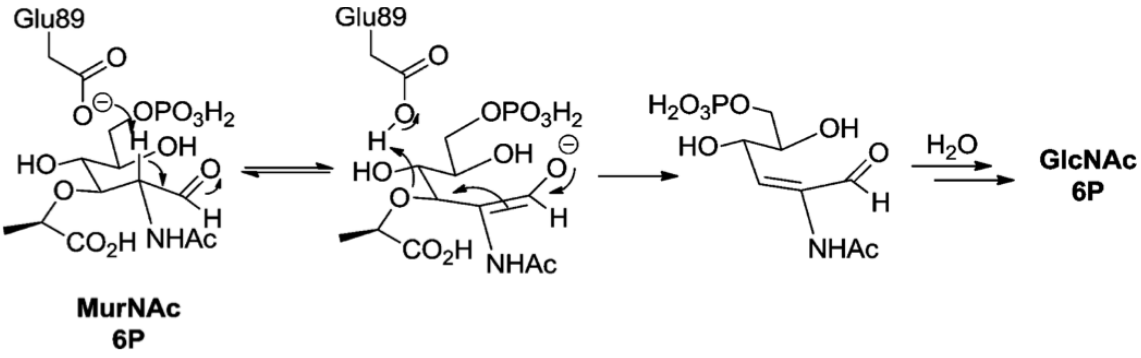


Fig. 40: Simplified scheme of the reaction mechanism of MurQ-like etherases as proposed by Hadi et al. Adapted from (Hadi *et al.*, 2013).

A structural comparison of MurQ from Gram-negative *H.influenzae* and from Gram-positive *B.subtilis* revealed that there are no significant differences between the overall structures of the two enzymes (Fig. 41) as well as between the topologies of their active sites (Fig. 42).

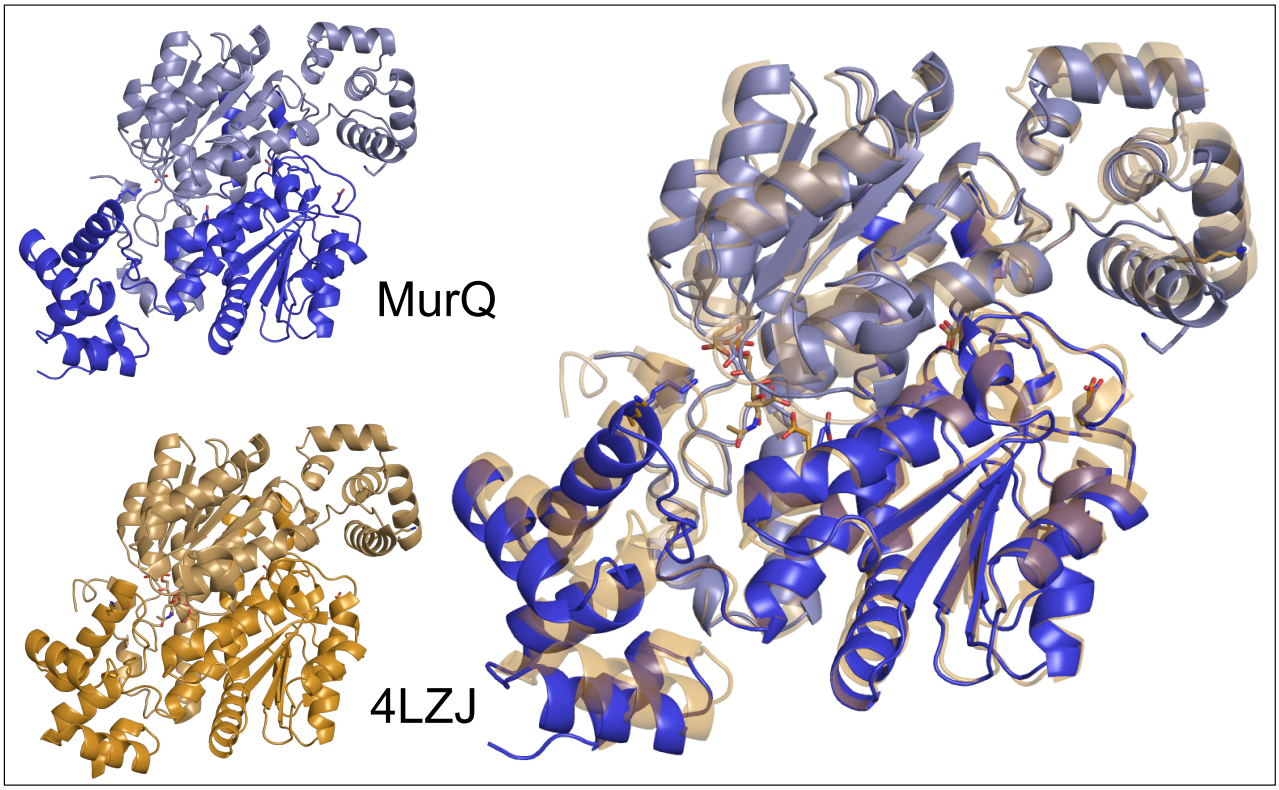


Fig. 41: Overall structural comparison of MurQ$^{H.i.}$ and MurQ$^{B.s.}$
MurQ$^{H.i.}$ (gold, PDBid: 4LZJ) is superposed onto MurQ$^{B.s.}$ (MurQ; blue: chain B, light blue: chain C). For MurQ$^{H.i.}$ the ring-opened substrate analog is shown in line-representation and for both enzymes the residues discussed in the text are represented as sticks. Colors are according to atom type (red: oxygen, blue: nitrogen) or reflect the chain/molecule identity (carbon).

In particular, the structure of MurQ$^{H.i.}$, which was determined using the same molecular replacement model as used for MurQ$^{B.s.}$, also comprised the whole C-terminal domain, which had been missing in the initial structure. This decreased the informative value of the C-terminal domain topology in the MurQ$^{B.s.}$ structure and the functional comparison between MurQ-like etherases from Gram-positive and Gram-negative organisms. Consequently, the answers to the questions that were initially intended to address with the structural analysis of MurQ$^{B.s.}$ had become largely obsolete with the published structure of MurQ$^{H.i.}$.

However, having a closer look to substrate binding sites of MurQ$^{B.s.}$ and MurQ$^{H.i.}$ revealed some slight, but potentially significant, differences in the side chain orientations of some active site residues. This brought some structural considerations concerning the reaction mechanism into focus that have not yet been accounted for (Fig. 42). In the complex structure of MurQ$^{H.i.}$, only one binding site is occupied by the ring-opened substrate analog (Fig. 42B), whereas the structure of MurQ$^{B.s.}$ represents the completely unoccupied from of the enzyme. From the structural comparison depicted in Fig. 42 it gets obvious that the orientations of the core- and C-terminal domains relative to each other remain unchanged upon substrate binding. Only the active site glutamates that are crucial for enzymatic activity (E87 and E118 in MurQ$^{B.s}$ / E89 and E120 in MurQ$^{H.i.}$) display a distinct reorientation when

substrate is bound. However, the orientations of these residues at the empty binding site of MurQ$^{H.i.}$ do not seem to be influenced by substrate coordination at the 'reciprocal' binding site, as their orientation is the same as in the unoccupied MurQ$^{B.s.}$ structure (Fig. 42C). This argues against an allosteric communication between the two active sites of MurQ. Nevertheless, helix IX of the C-terminal domain is rotated about 15° between MurQ$^{B.s}$ and MurQ$^{H.i.}$ and a lysine residue sitting at its proximal end (K237 in MurQ$^{B.s}$) is oriented differently in the two enzymes. In MurQ$^{H.i.}$, this Lys contributes to substrate coordination, whereas in the MurQ$^{B.s}$-structure it is too far away from the substrate binding site to do so in the state the *B.subtilis* enzyme has been crystallized in. Consequently, assuming that K237 of MurQ$^{B.s}$ also contributes to substrate binding in a similar manner as its homolog (K239 of MurQ$^{H.i.}$), a rearrangement of the C-terminal domain relative to the core upon substrate binding still seems possible for MurQ$^{B.s}$. This is one aspect of the structural analysis of MurQ, which still needs to be experimentally clarified for the specific case of *B.subtilis* and which might moreover add new significance to the comparison of MurQ between Gram-positive and Gram-negative organisms.

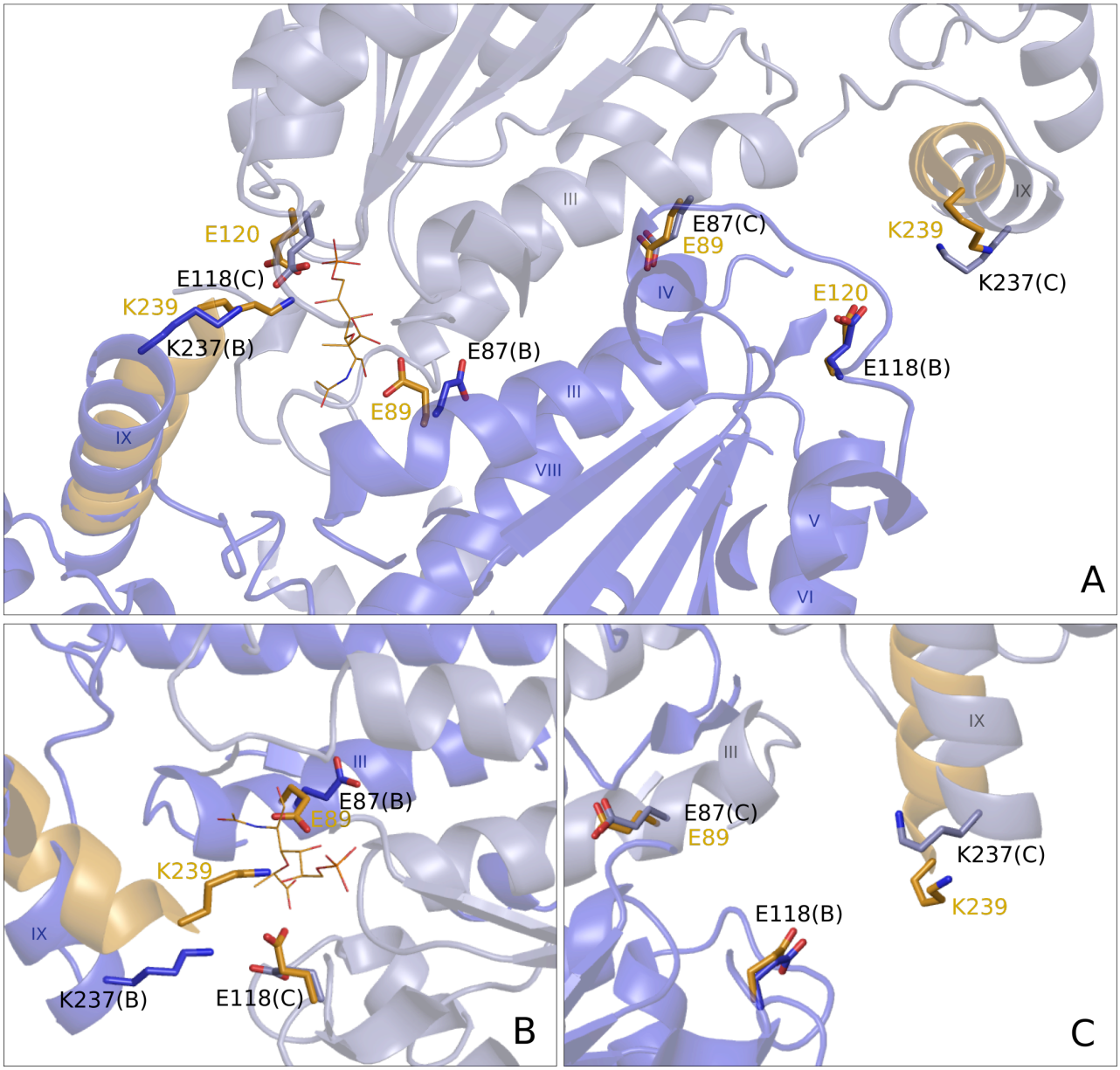


Fig. 42: Structural comparison of the active sites of MurQ$^{B.s.}$ and MurQ$^{H.i.}$
Panel A: Overview of both active sites at the physiological dimer interface. MurQ$^{H.i.}$ (gold, PDBid: 4LZJ) is superposed onto MurQ$^{B.s.}$ (MurQ; blue: chain B, light blue: chain C). For clarity, of MurQ$^{H.i.}$ only relevant secondary structural elements are shown. For MurQ$^{H.i.}$ the ring-opened substrate analog is shown in line-representation and for both enzymes the residues discussed in the text are represented as sticks. Colors are according to atom type (red: oxygen, blue: nitrogen) or reflect the chain/molecule identity (carbon). The small insets show a side-to-side comparison (upper left) and an overlay of the overall structures of both enzymes (lower right)
Panel B: Close-up view of the substrate binding site with the ring-opened substrate analog bound to MurQ$^{H.i.}$.
Panel C: Close-up view of the substrate binding site that is not occupied in MurQ$^{H.i}$

8. Summary

In this work, the high resolution structures of MurU of *P.putida* in native and two ligand-bound states as well as the structure of native MurQ from *B.subtilis* were solved.

Taken together, these structures contribute to the general understanding of bacterial cell wall recycling on a molecular level. Both structures provide insight into the catalytic mechanisms of the respective enzyme and might stimulate further scientific projects in this field. Apart from that, the complex structures of MurU may provide the structural basis for structure guided drug design to overcome fosfomycin resistance in antibiotic combination therapy.

8.1 Summary MurU

The structural analysis of MurU allows for the identification of critical catalytic residues and cofactors and reveals structural differences to other, related, SNTs. Moreover, the results presented in this work provide a structural basis for suggesting a catalytic mechanism and may furthermore be a starting point for a structural explanation of the remarkably narrow specificity of MurU for MurNAc-α1P.

MurU binds its substrate MurNAc-α1P through numerous polar and hydrophobic interactions, and specificity for the MurNAc group over related compounds such as GlcNAc or GalNAc is in part achieved through contacts that specifically target the unique lactyl group of the sugar. Although the location and orientation of the uridyl ring of the second substrate remains somewhat unclear as no reliable electron density could be observed for it, the pyrophosphate portions of UpNHpp and the pNHp of UppNHp are clearly visible and allow placement of at least this part of the nucleotide with confidence. The pyrophosphate mimics one of the products of the uridyl-transfer reaction and is tightly coordinated by residues of the SNT signature motif. Magnesium ions, or other divalent cations, have been found to be essential for or to enhance the enzymatic activities of other SNTs (Jagtap *et al.*, 2013, Gisin *et al.*, 2013, Singh *et al.*, 2012, Maruyama *et al.*, 2007), and it is likely that MurU likewise relies on bound magnesium ions to perform the nucleotidyl transfer reaction.

In comparison to the functionally related SNTs RmlA$^{P.a.}$, GlmU$^{E.c.}$ and GMPase$^{T.m}$, the substrate binding cleft of MurU is significantly more solvent exposed. Regarding this fact and considering the distinct binding mode for MurNAc-α1P, it seems likely that MurNAc-α1P binds first to the enzyme, in a preformed binding pocket, followed by Mg^{2+}-coordinated UTP. After uridyl-transfer the remaining pyrophosphate together, with the Mg^{2+}, would then be released immediately after the reaction, followed by the products UDP-MurNAc and pyrophosphate.

The recent rediscovery of fosfomycin in antibiotic combination therapy against multidrug-resistant

bacteria has identified intrinsic fosfomycin resistance as a remarkable clinical problem (Michalopoulos *et al.*, 2011). MurU significantly contributes to this resistance making a specific inhibitor for this enzyme a desirable objective. Since MurU is highly specific for its sugar substrate (Gisin *et al.*, 2013), the MurNAc-α1P binding pocket might be a promising target for structure guided inhibitor design. Although this pocket is shallow and solvent accessible, it still contains features that could be exploited, such as the hydrophobic pocket that accommodates the methyl group of the lactyl moiety of MurNAc-α1P or the charged surface that faces the carboxylate group of MurNAc-α1P. The design of an initial inhibitor should therefore mimic the lactyl pattern of MurNAc-α1P, and perhaps involve optimization of contacts that favor its binding over the actual substrate.

8.2 Summary MurQ

The structure of MurQ nicely shows how the active sites of this dimeric enzyme are made up at the dimer interface and confirms the glutamate residues E118 and E87, which have been found to be essential for enzymatic activity, to be located at these active sites. Due to its homodimeric nature, MurQ has two structurally redundant active sites, with one of the catalytically relevant glutamates being contributed from one chain and the second from the other. The same is true for lysine K237 and aspartate D119, which might also be involved in catalysis. The elevated B-factors of the loop region linking the C-terminal domain of MurQ to its active site-containing core implicate mobility of the C-terminal domain. This in turn suggests a mutual allosteric relation between the two active sites with the C-terminal domain presumably functioning as kind of a lid for the substrate binding site, which might exert a regulatory influence on enzyme activity.

The structural comparison between MurQ$^{B.s.}$ from Gram-positive *B.subtilis* and MurQ$^{H.i.}$ from Gram-negative *H.influenzae* clearly shows the distinct structural homology of the two enzymes. Furthermore it reveals a rearrangement of the catalytically relevant E87 (E89 in MurQ$^{H.i.}$) and a slight movement of the C-terminal domain, and thus K237 (K239 in MurQ$^{H.i.}$), upon substrate binding. This suggests that the coordination of substrate to the active site might be the trigger for a potential allosteric communication between the two active sites. However, to my knowledge, no studies explicitly addressing this aspect have been published so far. Moreover, these rearrangements have not yet been shown for a MurQ ortholog of a Gram-positive organism and mutational studies investigating the roles of K237 and D119 (*B.subtilis* numbering) are currently also not available. Despite their distinct structural homology, these aspects add new significance to the comparison of MurQ between Gram-positive and Gram-negative organisms, and raise new questions to be answered.

9. Zusammenfassung

In dieser Arbeit werden die per Röntgenkristallstrukturanalyse in submolekularer Auflösung bestimmten Strukturen zweier Enzyme, MurU aus *P.putida* und MurQ aus *B.subtilis*, präsentiert. Beide Enzyme sind an der Zellwandsynthese, genauer an der metabolischen Wiederverwertung von Zellwandkataboliten, im jeweiligen Organismus beteiligt.

Von der Nucleotidyltransferase MurU wurden, neben der nativen, außerdem zwei Strukturen mit im aktiven Zentrum koordinierten Substrat (-analoga) veröffentlicht. Die Analyse dieser Strukturen erlaubt die Identifikation katalytisch relevanter Aminosäuren und Cofaktoren und deckt strukturelle Unterschiede zu anderen, verwandten Nucleotidyltransferasen auf. Darüberhinaus eignen sich die in dieser Arbeit präsentierten Resultate als strukturelle Basis für die hypothetische Formulierung eines katalytischen Mechanismus, und könnten außerdem einen Ausgangspunkt für eine strukturelle Erklärung der strikten Substratspezifität von MurU für MurNAc-α1P bieten.

Durch die aktuelle Wiederentdeckung des Antibiotikums Fosfomycin für die Antibiotika Kombinationstherapie im Kampf gegen multiresistente Bakterien, offenbarte sich die intrinsische Fosfomycin Resistenz vieler Bakterien als erhebliches klinisches Problem (Michalopoulos *et al.*, 2011). MurU ist an einem Stoffwechselweg beteiligt, welcher die metabolische Wiederverwertung von Zellwandkataboliten unabhängig macht von dem durch Fosfomycin gehemmten Enzym und trägt daher entscheidend zu einer intrinsischen Fosfomycin Resistenz bei. Dieser Zusammenhang macht die Entwicklung eines spezifischen Inhibitors für MurU zu einem sehr erstrebenswerten Ziel. In diesem Zusammenhang könnte, aufgrund der strikten Spezifität von MurU für sein Zucker-Substrat, dessen Bindungstasche für MurNAc-α1P ein vielversprechendes Zielobjekt für die Struktur basierte Entwicklung eines MurU-spezifischen Inhibitors darstellen.

Die Struktur der Etherase MurQ zeigt sehr anschaulich, wie ihre beiden aktiven Zentren an der Dimer-Grenzfläche unter Beteiligung jeweils beider Monomere gebildet werden. Durch die Analyse der aktiven Zentren kann außerdem sehr anschaulich bestätigt werden, dass die beiden Glutamate E87 und E118 für die Enzymaktivität relevant sein müssen, wie bereits in Mutations-Studien gezeigt werden konnte. Aufgrund eines strukturellen Vergleichs der in dieser Arbeit präsentierten Struktur von MurQ aus dem Gram-positiven Organismus *B.subtilis* mit der im gleichen Zeitraum publizierten Struktur von MurQ aus Gram-negativen Organismus *H.influenzae* ((Hadi *et al.*, 2013)), erscheint außerdem eine gegenseitige allosterische Beeinflussung der beiden aktiven Zentren denkbar. Dies würde gleichzeitig einen Ausgangpunkt für die Formulierung einer Theorie zur Regulation von MurQ liefern. In diesem Zusammenhang ist die Datenlage zum momentanen Zeitpunkt allerdings bei weitem nicht ausreichend.

Zusammenfassend betrachtet tragen die in dieser Arbeit analysierten Strukturen von MurU und MurQ zu einem generellen Verständnis der bakteriellen Wiederverwertung von Zellwandkataboliten auf molekularer Ebene bei. Dabei ergeben sich aus beiden Strukturen Einblicke in den katalytischen Mechanismus des jeweiligen Enzyms. Außerdem bringen beide Projekte neue Aspekte in den Focus des Interesses, welche gegebenenfalls weitere Forschungen auf dem Gebiet der bakteriellen Wiederverwertung von Zellwandkataboliten stimulieren. Insbesondere die Strukturen der Enzym-Substrat Komplexe von MurU könnten sich in diesen Zusammenhang als strukturelle Basis in der pharmazeutischen Forschung erweisen.

10. References

Abendroth, J., Gardberg, A. S., Robinson, J. I., Christensen, J. S., Staker, B. L., Myler, P. J., Stewart, L. J. & Edwards, T. E. (2011). *J. Struct. Funct. Genomics* **12**, 83-95.

Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C. & Zwart, P. H. (2010). *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213-221.

Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. (2001). *Proc. Natl. Acad. Sci. U. S. A.* **98**, 10037-10041.

Bateman, A. (1999). *Trands in Biochmical Sciences* **24**, 94-95.

Baykov A.A., Evtushenko O.A. & Avaeva S.M. (1988). *Anal. Biochem.* **171**, 266-270.

Blankenfeldt, W., Asuncion, M., Lam, J. S. & Naismith, J. H. (2000). *EMBO J.* **19**, 6652-6663.

Borisova, M., Gisin, J. & Mayer, C. (2014). *Microb. Drug Resist.* **20**.

Brown, E. D., Vivas, E. I. & Walsh, C. T. (1995). *J. Bacteriol.* **177**, 4194–4197.

Brown, K., Pompeo, F. r., Dixon, S., Mengin-Lecreulx, D., Cambillau, C. & Bourne, Y. (1999). *EMBO J.* **18**, 4096–4107.

Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12-21.

Dahl, U., Jaeger, T., Nguyen, B. T., Sattler, J. M. & Mayer, C. (2004). *J. Bacteriol.* **186**, 2385-2392.

Dauter, M. & Dauter, Z. (2007). *Methods Mol. Biol.*, 2 ed., pp. 149-158: Humana Press Inc., Totowa, N.

DeLano (2012). *The PyMOL Molecular Graphics Software.* Version 1.5.0.4.

Diederichs, K. & Karplus, P. A. (1997). *Nat. Struct. Biol.* **4**, 269 - 275.

Duarte, J. M., Srebniak, A., Schärer, M. A. & Capitani, G. (2012). *BMC Bioinformatics* **13**.

Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486-501.

Evans, P. (2006). *Acta Crystallogr. D Biol. Crystallogr.* **62**, 72-82.

F.M. Kahan, J.S. Kahan, PJ Cassidy & Kropp, H. (1974). *Ann N Y Acad Sci.* **10**, 364-386.

G. F. White, N. J. Russell & Tidswell, E. C. (1996). *Microbiol. Rev.* **60**, 216-232.

Garman, E. F. (2014). *Science* **343**, 1102-1108.

Gasteiger E., H. C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A.; ProtParam: Protein Identification and Analysis Tools on the ExPASy Server;, Vol. John M. Walker (ed): The Proteomics Protocols Handbook, Humana Press (2005). pp. 571-601.

Gisin, J., Schneider, A., Nagele, B., Borisova, M. & Mayer, C. (2013). *Nat. Chem. Biol.* **9**, 491-493.

Hadi, T., Dahl, U., Mayer, C. & Tanner, M. E. (2008). *Biochemistry* **47**, 11547-11558.

Hadi, T., Hazra, S., Tanner, M. E. & Blanchard, J. S. (2013). *Biochemistry* **52**, 9358-9366.

Holm, L. & Rosenstrom, P. (2010). *Nucleic Acids Res.* **38**, 545-549.

iycr2014 (2014). *International Year of Crystalogrphy 2014,* http://www.iycr2014.org.

Jaeger, T., Arsic, M. & Mayer, C. (2005). *J. Biol. Chem.* **280**, 30100-30106.

Jaeger, T. & Mayer, C. (2008). *Cell. Mol. Life Sci.* **65**, 928-939.

Jagtap, P. K., Verma, S. K., Vithani, N., Bais, V. S. & Prakash, B. (2013). *J. Mol. Biol.* **425**, 1745-1759.

Johnson, J. W., Fisher, J. F. & Mobashery, S. (2013). *Ann. N. Y. Acad. Sci.* **1277**, 54-75.

Jollès, P. (1996). *Lysozmes; Model Enzymes in Biochemistry and Biology*. Birkhäuser Verlag, Basel.

Kabsch, W. (2010). *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125-132.

Kantardjieff, K. A. & Rupp, B. (2003). *Protein Sci.* **12**, 1865-1871.

Karplus, P. A. & Diederichs, K. (2012). *Science* **336**, 1030-1033.

Litzinger, S., Duckworth, A., Nitzsche, K., Risinger, C., Wittmann, V. & Mayer, C. (2010). *J. Bacteriol.* **192**, 3132-3143.

Litzinger, S. & Mayer, C. (2010). *Prokaryotic Cell Wall Compounds*, edited by H. König, H. Claus & A. Varma, pp. 3-52: Springer Verlag Berlin Heidelberg.

Lovering, A. L., Safadi, S. S. & Strynadka, N. C. (2012). *Annu. Rev. Biochem.* **81**, 451-478.

Malito, E. Stamp Collecting Project, Stamp Collecting Project.

Maruyama, D., Nishitani, Y., Nonaka, T., Kita, A., Fukami, T. A., Mio, T., Yamada-Okabe, H., Yamada-Okabe, T. & Miki, K. (2007). *J. Biol. Chem.* **282**, 17221-17230.

Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491-497.

Mayer, C. (2012). Bacterial Cell Wall Recycling, eLS, pp. 3-52. John Wiley & Sons, Ltd.

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Crystallogr.* **40**, 658-674.

Michalopoulos, A. S., Livaditis, I. G. & Gougoutas, V. (2011). *Int. J. Infect. Dis.* **15**, 732-739.

NATURE, s. i. (2014). *Nature* **505**, 585-716.

Painter, J. & Merritt, E. A. (2006). *J. Appl. Cryst.* **39**, 109-111.

Park, J. T. & Uehara, T. (2008). *Microbiol. Mol. Biol. Rev.* **72**, 211-227.

Pelissier, M. C., Lesley, S. A., Kuhn, P. & Bourne, Y. (2010). *J. Biol. Chem.* **285**, 27468-27476.

Ramachandran, R. a. S. (1963). *J. Mol. Biol.* **7**, 95-99.

Rupp, B. (2010). *Biomolecular Crystallography*. Garland Science Taylor & Frncis Group, LLC.

Schleifer, K. H. & O., K. (1972). *Bacteriol. Rev.* **36**, 407-477.

Schneck, M. (2008). Diploma thesis thesis, Eberhard-Karls-Universität Tübingen, Tübingen.

SCIENCE, s. i. (2014). *Science* **343**, 1049-1168.

Sheldrick, G. (2010). *Acta Crystallogr. D Biol. Crystallogr.* **66**, 479-485.

Shockman, G. D., Daneo-Moore, L., Karyama, R. & Massidda, O. (1996). *Microb. Drug Resist.* **2**, 95-98.

Silhavy, T. J., Kahne, D. & Walker, S. (2010). *Cold Spring Harb. Perspect. Biol.* **2**, a000414.

Singh, S., Phillips, G. N., Jr. & Thorson, J. S. (2012). *Nat. Prod. Rep.* **29**, 1201-1237.

Sumner, T. (2014). *Science* **343**.

Teplyakov, A., Obmolova, G., Badet-Denisot, M.-A. & Badet, B. (1999). *Protein Sci.* **8**, 596-602.

Trempe, J. F., Shenker, S., Kozlov, G. & Gehring, K. (2011). *Protein Sci.* **20**, 745-752.

Vithani, N., Bais, V. & Prakash, B. (2014). *Acta Crystallogr F Struct Biol Commun* **70**, 703-708.

Vollmer, W., Blanot, D. & de Pedro, M. A. (2008). *FEMS Microbiol. Rev.* **32**, 149-167.

Vonrhein, C., Blanc, E., Roversi, P. & Bricogne, G. (2007). *Methods Mol. Biol.* **364**, 215-230.

Weiss, M. S. (2001). *J. Appl. Cryst.* **34**, 130 - 135.

Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A. & Wilson, K. S. (2011). *Acta Crystallogr. D Biol. Crystallogr.* **67**, 235-242.

http://www.pdb.org PDB: RCSB.

Zhang, Z., Bulloch, E. M., Bunker, R. D., Baker, E. N. & Squire, C. J. (2009). *Acta Crystallogr. D Biol. Crystallogr.* **65**, 275-283.

11. Appendix

11.1 PDB valid reports

11.1.1 4Y7T – native MurU

11.1.2 4Y7V – complex #1

11.1.3 4Y7U complex #2

11.1.4 native MurQ

11.2 Complex structure: MurU with UDP + MurNAc (loop no. 0188)

11.3 *In silico* characterization of the Protein contructs

11.3.1 MurU-His_6

11.3.2 His_6-MurQ

11.4 BLAST® search results MurU

11.5 Crystallization Screens

11. Appendix

11.1 PDB valid reports

11.1.1 4Y7T – native MurU

11.1.2 4Y7V – complex #1

11.1.3 4Y7U complex #2

Full wwPDB X-ray Structure Validation Report (i)

Feb 16, 2015 – 05:08 AM EST

PDB ID	:	4Y7T
Title	:	Structural analysis of MurU
Authors	:	Renner-Schneck, M.G.; Stehle, T.
Deposited on	:	2015-02-16
Resolution	:	1.80 Å(reported)

DISCLAIMER
This is a preliminary version of the new style of wwPDB validation report.
We welcome your comments at validation@mail.wwpdb.org
A user guide is available at http://wwpdb.org/ValidationPDFNotes.html

CONFIDENTIAL VALIDATION REPORT

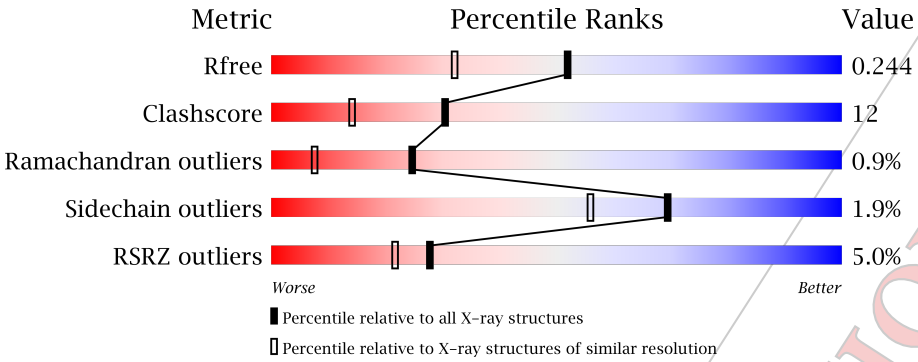
The following versions of software and data (see references) were used in the production of this report:

MolProbity	:	4.02b-467
Mogul	:	1.16 November 2013
Xtriage (Phenix)	:	dev-1439
EDS	:	stable24195
Percentile statistics	:	21963
Refmac	:	5.8.0049
CCP4	:	6.3.0 (Settle)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP)	:	stable24195

1 Overall quality at a glance

The reported resolution of this entry is 1.80 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	66092	3513 (1.80-1.80)
Clashscore	79885	4461 (1.80-1.80)
Ramachandran outliers	78287	4404 (1.80-1.80)
Sidechain outliers	78261	4403 (1.80-1.80)
RSRZ outliers	66119	3515 (1.80-1.80)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Mol	Chain	Length	Quality of chain
1	A	231	

The following table lists non-polymeric compounds that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Geometry	Electron density
3	SO4	A	306	-	X

CONFIDENTIAL VALIDATION REPORT

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 1825 atoms, of which 24 are hydrogens and 0 are deuterium.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

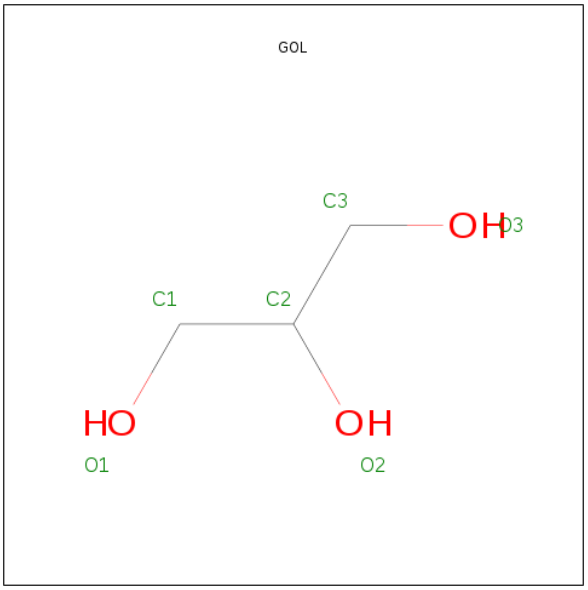
- Molecule 1 is a protein called Nucleotidyl transferase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	218	1683	1090	299	289	5	0	11	0

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	224	LEU	-	expression tag	UNP E4RE40
A	225	GLU	-	expression tag	UNP E4RE40
A	226	HIS	-	expression tag	UNP E4RE40
A	227	HIS	-	expression tag	UNP E4RE40
A	228	HIS	-	expression tag	UNP E4RE40
A	229	HIS	-	expression tag	UNP E4RE40
A	230	HIS	-	expression tag	UNP E4RE40
A	231	HIS	-	expression tag	UNP E4RE40

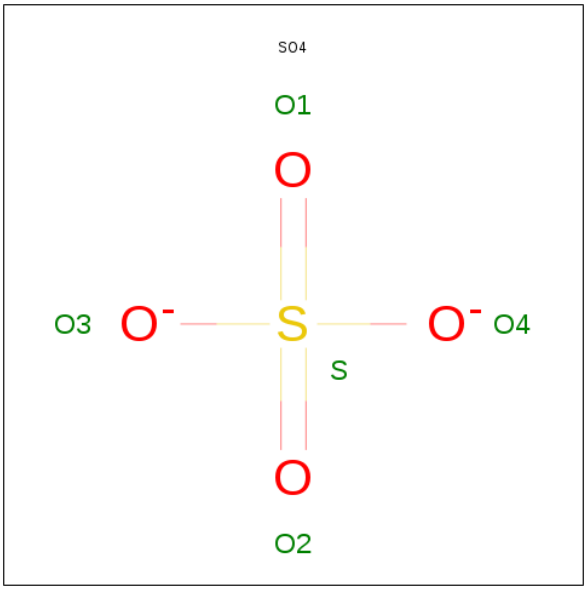
- Molecule 2 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).



VALIDATION REPORT

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	1	Total C H O 28 6 16 6	0	1
2	A	1	Total C H O 14 3 8 3	0	0

- Molecule 3 is SULFATE ION (three-letter code: SO4) (formula: O_4S).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total O S 5 4 1	0	0
3	A	1	Total O S 5 4 1	0	0
3	A	1	Total O S 5 4 1	0	0
3	A	1	Total O S 5 4 1	0	0

- Molecule 4 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	80	Total O 80 80	0	0

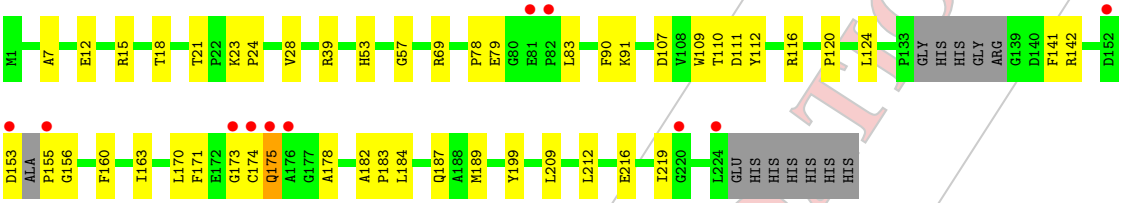
CONFIDENTIAL VALIDATION REPORT

3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density (RSRZ > 2). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

- Molecule 1: Nucleotidyl transferase

Chain A:



CONFIDENTIAL VALIDATION REPORT

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 61 2 2	Depositor
Cell constants a, b, c, α, β, γ	72.61Å 72.61Å 158.47Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	40.45 – 1.80 40.45 – 1.80	Depositor EDS
% Data completeness (in resolution range)	99.9 (40.45-1.80) 99.9 (40.45-1.80)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.36 (at 1.79Å)	Xtriage
Refinement program	PHENIX	Depositor
R, R_{free}	0.211 , 0.245 0.224 , 0.244	Depositor DCC
R_{free} test set	1185 reflections (5.26%)	DCC
Wilson B-factor (Å^2)	35.4	Xtriage
Anisotropy	0.528	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 42.8	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning	$< \|L\| > = 0.50, < L^2 > = 0.33$	Xtriage
Outliers	1 of 23701 reflections (0.004%)	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	1825	wwPDB-VP
Average B, all atoms (Å^2)	47.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.80% of the height of the origin peak. No significant pseudotranslation is detected.*

[1] Intensities estimated from amplitudes.

CONFIDENTIAL VALIDATION REPORT

5 Model quality

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: GOL, SO4

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	#$\|Z\|>5$	RMSZ	#$\|Z\|>5$
1	A	0.67	0/1755	0.73	0/2377

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

5.2 Close contacts

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogens added by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, and the number in parentheses is this value normalized per 1000 atoms of the molecule in the chain. The Symm-Clashes column gives symmetry related clashes, in the same way as for the Clashes column.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	1683	0	1727	40	0
2	A	18	24	24	1	0
3	A	20	0	0	0	0
4	A	80	0	0	2	4
All	All	1801	24	1751	41	4

Clashscore is defined as the number of clashes calculated for the entry per 1000 atoms (including hydrogens) of the entry. The overall clashscore for this entry is 12.

All (41) close contacts within the same asymmetric unit are listed below.

CONFIDENTIAL VALIDATION REPORT

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:A:83:LEU:O	1:A:178:ALA:HA	1.83	0.78
1:A:112:TYR:HE1	1:A:116[B]:ARG:HH21	1.38	0.71
1:A:171:PHE:O	1:A:174:CYS:HB2	1.93	0.68
1:A:28:VAL:HG22	1:A:212[A]:LEU:CD1	2.23	0.68
1:A:173:GLY:N	1:A:174:CYS:HA	2.08	0.67
1:A:174:CYS:O	1:A:175:GLN:CB	2.45	0.65
1:A:28:VAL:HG22	1:A:212[A]:LEU:HD11	1.78	0.65
1:A:171:PHE:HA	1:A:174:CYS:SG	2.37	0.65
1:A:110:THR:HG21	1:A:163[B]:ILE:HD11	1.79	0.64
1:A:7:ALA:HB1	1:A:24:PRO:HG3	1.80	0.64
1:A:153:ASP:O	1:A:155:PRO:HD3	1.98	0.63
1:A:141:PHE:HE2	1:A:189:MET:HE1	1.67	0.60
1:A:174:CYS:O	1:A:175:GLN:HB2	2.01	0.59
1:A:153:ASP:C	1:A:155:PRO:HD3	2.21	0.59
1:A:171:PHE:C	1:A:174:CYS:HB2	2.24	0.58
1:A:182:ALA:HB3	1:A:183:PRO:HD3	1.89	0.54
1:A:142[B]:ARG:HH22	1:A:156:GLY:N	2.06	0.54
1:A:142[B]:ARG:HH22	1:A:155:PRO:C	2.12	0.53
1:A:175:GLN:NE2	4:A:403:HOH:O	2.42	0.52
1:A:141:PHE:HE2	1:A:189:MET:CE	2.23	0.52
1:A:12:GLU:HA	1:A:15:ARG:NH1	2.26	0.50
1:A:170:LEU:O	1:A:184[B]:LEU:HD23	2.13	0.49
1:A:109:TRP:CZ2	1:A:219:ILE:HG12	2.49	0.47
1:A:141:PHE:CE1	1:A:160:PHE:HB2	2.49	0.47
1:A:79:GLU:HG2	1:A:83:LEU:HD12	1.96	0.46
1:A:111:ASP:OD1	1:A:199:TYR:OH	2.24	0.46
1:A:18:THR:HA	1:A:21:THR:O	2.17	0.45
1:A:23:LYS:HD2	1:A:107:ASP:OD1	2.17	0.44
1:A:187:GLN:HG3	2:A:302:GOL:H11	1.99	0.44
1:A:141:PHE:CE2	1:A:189:MET:CE	3.01	0.44
1:A:12:GLU:HA	1:A:15:ARG:HH11	1.81	0.43
1:A:23:LYS:N	1:A:24:PRO:HD2	2.33	0.43
1:A:83:LEU:HA	1:A:83:LEU:HD23	1.83	0.43
1:A:142[A]:ARG:NH2	1:A:155:PRO:O	2.51	0.43
1:A:124:LEU:HB3	1:A:170:LEU:HD13	2.02	0.42
1:A:155:PRO:HA	1:A:156:GLY:HA2	1.77	0.41
1:A:90:PHE:CE1	1:A:174:CYS:HB3	2.56	0.41
1:A:53:HIS:ND1	1:A:57:GLY:HA3	2.36	0.41
1:A:78:PRO:O	1:A:91:LYS:HE2	2.21	0.40

All (4) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

CONFIDENTIAL VALIDATION REPORT

Atom-1	Atom-2	Distance(Å)	Clash(Å)
4:A:402:HOH:O	4:A:402:HOH:O[7_555]	2.05	0.15
4:A:402:HOH:O	4:A:423:HOH:O[7_555]	2.13	0.07
4:A:424:HOH:O	4:A:424:HOH:O[12_555]	2.16	0.04
4:A:423:HOH:O	4:A:423:HOH:O[7_555]	2.19	0.01

5.3 Torsion angles

5.3.1 Protein backbone

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
1	A	223/231 (96%)	210 (94%)	11 (5%)	2 (1%)	25 7

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	175	GLN
1	A	120	PRO

5.3.2 Protein sidechains

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution. The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
1	A	169/177 (96%)	164 (97%)	5 (3%)	53 34

All (5) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	39[A]	ARG
1	A	39[B]	ARG
1	A	69[A]	ARG
1	A	69[C]	ARG

Continued on next page...

CONFIDENTIAL VALIDATION REPORT

Continued from previous page...

Mol	Chain	Res	Type
1	A	209	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

5.3.3 RNA (i)

There are no RNA chains in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains (i)

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates (i)

There are no carbohydrates in this entry.

5.6 Ligand geometry (i)

7 ligands are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
2	GOL	A	301[A]	-	5,5,5	0.22	0	5,5,5	0.68	0
2	GOL	A	301[B]	-	5,5,5	0.33	0	5,5,5	0.42	0
2	GOL	A	302	-	5,5,5	0.45	0	5,5,5	0.59	0
3	SO4	A	303	-	4,4,4	0.29	0	6,6,6	0.30	0
3	SO4	A	304	-	4,4,4	0.24	0	6,6,6	0.32	0
3	SO4	A	305	-	4,4,4	0.22	0	6,6,6	0.20	0
3	SO4	A	306	-	4,4,4	0.07	0	6,6,6	0.16	0

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral

CONFIDENTIAL VALIDATION REPORT

centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	GOL	A	301[A]	-	-	0/4/4/4	0/0/0/0
2	GOL	A	301[B]	-	-	0/4/4/4	0/0/0/0
2	GOL	A	302	-	-	0/4/4/4	0/0/0/0
3	SO4	A	303	-	-	0/0/0/0	0/0/0/0
3	SO4	A	304	-	-	0/0/0/0	0/0/0/0
3	SO4	A	305	-	-	0/0/0/0	0/0/0/0
3	SO4	A	306	-	-	0/0/0/0	0/0/0/0

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

5.7 Other polymers

There are no such residues in this entry.

5.8 Polymer linkage issues

There are no chain breaks in this entry.

CONFIDENTIAL VALIDATION REPORT

6 Fit of model and data

6.1 Protein, DNA and RNA chains

In the following table, the column labelled '#RSRZ> 2' contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95^{th} percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled 'Q< 0.9' lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2	OWAB(Å^2)	Q<0.9
1	A	218/231 (94%)	0.01	11 (5%) 28 22	28, 43, 83, 109	0

All (11) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	173	GLY	6.1
1	A	153	ASP	4.2
1	A	174	CYS	3.5
1	A	224	LEU	3.1
1	A	155	PRO	2.8
1	A	175	GLN	2.7
1	A	220	GLY	2.6
1	A	176	ALA	2.6
1	A	82	PRO	2.4
1	A	152	ASP	2.2
1	A	81	GLU	2.2

6.2 Non-standard residues in protein, DNA, RNA chains

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates

There are no carbohydrates in this entry.

6.4 Ligands

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains.

CONFIDENTIAL VALIDATION REPORT

The B-factors column lists the minimum, median, 95^{th} percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSR	LLDF	B-factors(Å^2)	Q<0.9
3	SO4	A	306	5/5	0.18	2.09	114,115,116,116	0
2	GOL	A	301[A]	6/6	0.17	1.70	40,48,53,54	14
2	GOL	A	301[B]	6/6	0.17	1.70	40,48,53,54	14
3	SO4	A	304	5/5	0.20	0.75	88,91,91,93	5
3	SO4	A	305	5/5	0.13	0.59	88,89,90,90	5
2	GOL	A	302	6/6	0.14	-0.05	39,47,56,57	14
3	SO4	A	303	5/5	0.05	-1.57	52,55,59,61	0

6.5 Other polymers (i)

There are no such residues in this entry.

CONFIDENTIAL VALIDATION REPORT

Full wwPDB X-ray Structure Validation Report (i)

Feb 16, 2015 – 06:15 AM EST

PDB ID	:	4Y7V
Title	:	Structural analysis of MurU
Authors	:	Renner-Schneck, M.G.; Stehle, T.
Deposited on	:	2015-02-16
Resolution	:	1.80 Å(reported)

DISCLAIMER
This is a preliminary version of the new style of wwPDB validation report.
We welcome your comments at validation@mail.wwpdb.org
A user guide is available at http://wwpdb.org/ValidationPDFNotes.html

CONFIDENTIAL VALIDATION REPORT

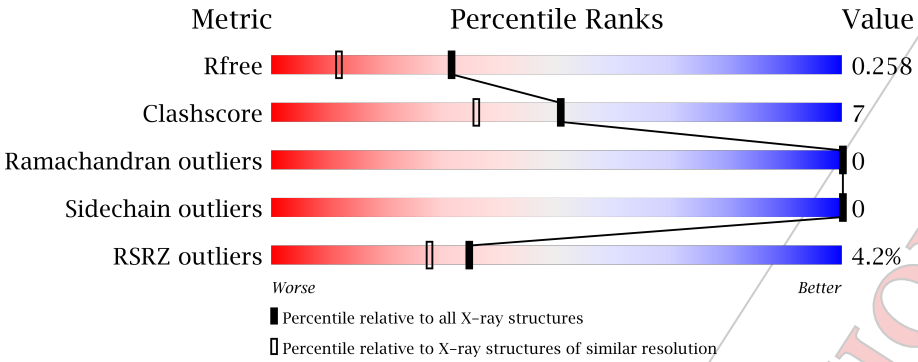
The following versions of software and data (see references) were used in the production of this report:

MolProbity	:	4.02b-467
Mogul	:	1.16 November 2013
Xtriage (Phenix)	:	dev-1439
EDS	:	stable24195
Percentile statistics	:	21963
Refmac	:	5.8.0049
CCP4	:	6.3.0 (Settle)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP)	:	stable24195

1 Overall quality at a glance

The reported resolution of this entry is 1.80 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	66092	3513 (1.80-1.80)
Clashscore	79885	4461 (1.80-1.80)
Ramachandran outliers	78287	4404 (1.80-1.80)
Sidechain outliers	78261	4403 (1.80-1.80)
RSRZ outliers	66119	3515 (1.80-1.80)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Mol	Chain	Length	Quality of chain
1	A	231	

CONFIDENTIAL VALIDATION REPORT

2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 1807 atoms, of which 38 are hydrogens and 0 are deuterium.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

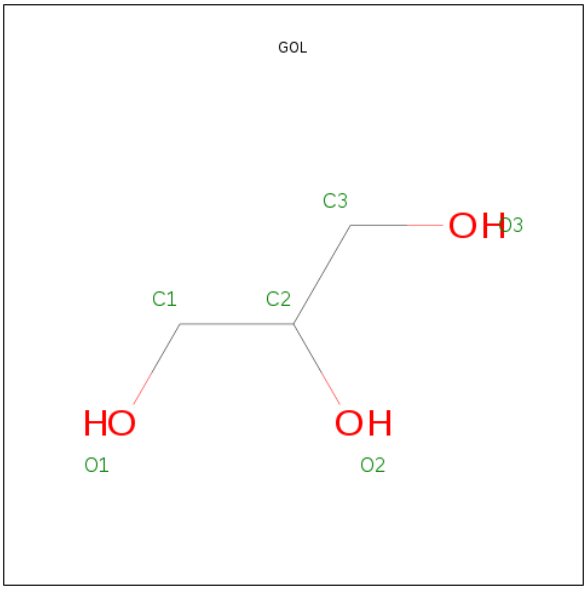
- Molecule 1 is a protein called Nucleotidyl transferase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	216	1652	1068	292	287	5	0	9	0

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	224	LEU	-	expression tag	UNP E4RE40
A	225	GLU	-	expression tag	UNP E4RE40
A	226	HIS	-	expression tag	UNP E4RE40
A	227	HIS	-	expression tag	UNP E4RE40
A	228	HIS	-	expression tag	UNP E4RE40
A	229	HIS	-	expression tag	UNP E4RE40
A	230	HIS	-	expression tag	UNP E4RE40
A	231	HIS	-	expression tag	UNP E4RE40

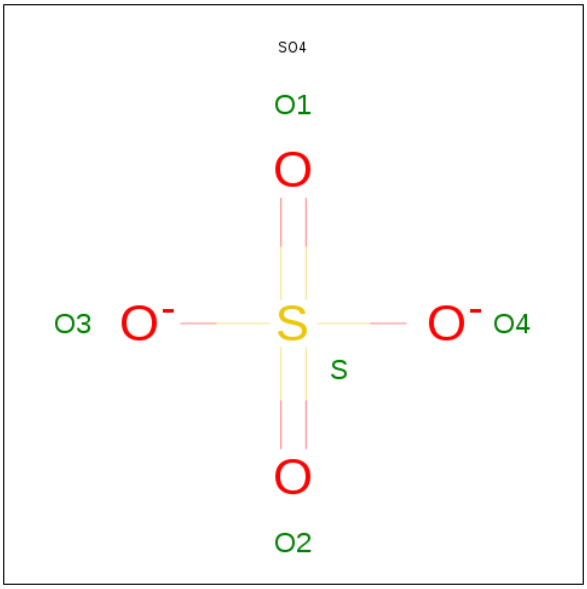
- Molecule 2 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).



VALIDATION REPORT

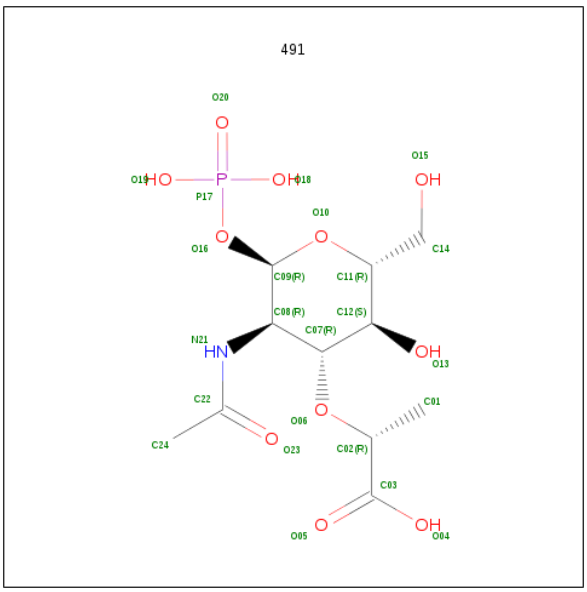
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	1	Total C H O 28 6 16 6	0	1

- Molecule 3 is SULFATE ION (three-letter code: SO4) (formula: O_4S).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total O S 5 4 1	0	0

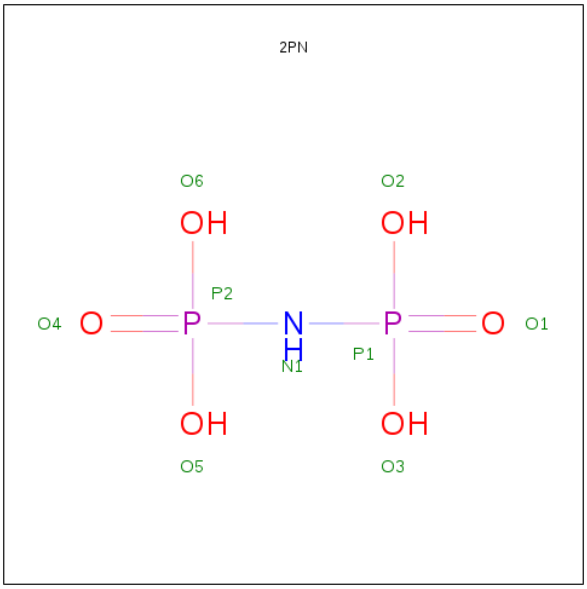
- Molecule 4 is 2-(acetylamino)-3-O-[(1R)-1-carboxyethyl]-2-deoxy-1-O-phosphono-alpha-D-glucopyranose (three-letter code: 491) (formula: $C_{11}H_{20}NO_{11}P$).



VALIDATION REPORT

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf
4	A	1	Total	C	H	N	O	P	0	0
			41	11	17	1	11	1		

- Molecule 5 is IMIDODIPHOSPHORIC ACID (three-letter code: 2PN) (formula: $H_5NO_6P_2$).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
5	A	1	Total	H	N	O	P	0	0
			14	5	1	6	2		

- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	67	Total	O	0	0
			67	67		

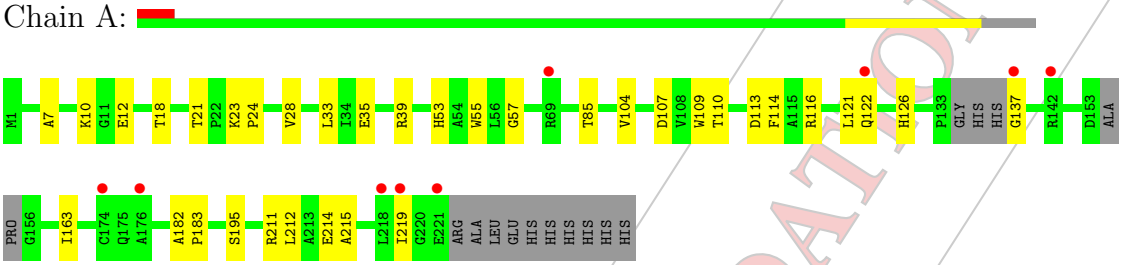
CONFIDENTIAL VALIDATION REPORT

3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density (RSRZ > 2). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

- Molecule 1: Nucleotidyl transferase

Chain A:



CONFIDENTIAL VALIDATION REPORT

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 61 2 2	Depositor
Cell constants a, b, c, α, β, γ	72.35Å 72.35Å 162.48Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	49.62 – 1.80 49.61 – 1.80	Depositor EDS
% Data completeness (in resolution range)	100.0 (49.62-1.80) 100.0 (49.61-1.80)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.06 (at 1.79Å)	Xtriage
Refinement program	PHENIX	Depositor
R, R_{free}	0.215 , 0.254 0.231 , 0.258	Depositor DCC
R_{free} test set	1207 reflections (5.26%)	DCC
Wilson B-factor (Å^2)	40.0	Xtriage
Anisotropy	0.178	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.37 , 49.3	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Outliers	0 of 24133 reflections	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	1807	wwPDB-VP
Average B, all atoms (Å^2)	52.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.89% of the height of the origin peak. No significant pseudotranslation is detected.*

[1] Intensities estimated from amplitudes.

CONFIDENTIAL VALIDATION REPORT

5 Model quality

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: GOL, 2PN, 491, SO4

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	$\#\|Z\| > 5$	RMSZ	$\#\|Z\| > 5$
1	A	0.37	0/1715	0.53	0/2325

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

5.2 Close contacts

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogens added by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, and the number in parentheses is this value normalized per 1000 atoms of the molecule in the chain. The Symm-Clashes column gives symmetry related clashes, in the same way as for the Clashes column.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	1652	0	1681	25	0
2	A	12	16	16	0	0
3	A	5	0	0	1	0
4	A	24	17	0	0	0
5	A	9	5	1	1	0
6	A	67	0	0	1	0
All	All	1769	38	1698	25	0

Clashscore is defined as the number of clashes calculated for the entry per 1000 atoms (including hydrogens) of the entry. The overall clashscore for this entry is 7.

All (25) close contacts within the same asymmetric unit are listed below.

CONFIDENTIAL VALIDATION REPORT

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:A:12:GLU:H	5:A:304:2PN:HN1	1.45	0.64
1:A:113:ASP:HB3	1:A:116[B]:ARG:HD2	1.81	0.63
1:A:109:TRP:CZ2	1:A:219:ILE:HG12	2.34	0.62
1:A:182:ALA:HB3	1:A:183:PRO:HD3	1.83	0.61
1:A:85:THR:HB	3:A:302:SO4:O1	2.03	0.58
1:A:28:VAL:HG22	1:A:212[A]:LEU:CD1	2.35	0.56
1:A:28:VAL:HG22	1:A:212[B]:LEU:CD1	2.38	0.52
1:A:110:THR:HG21	1:A:163[B]:ILE:HD11	1.92	0.50
1:A:18:THR:HA	1:A:21:THR:O	2.11	0.49
1:A:7:ALA:HB1	1:A:24:PRO:HG2	1.94	0.49
1:A:114:PHE:HE1	1:A:163[A]:ILE:HD12	1.77	0.48
1:A:10:LYS:HG3	1:A:55:TRP:CD2	2.51	0.46
1:A:211:ARG:HE	1:A:211:ARG:HA	1.80	0.46
1:A:35:GLU:O	1:A:39[A]:ARG:HG3	2.16	0.45
1:A:28:VAL:HG21	1:A:33:LEU:HD13	1.98	0.45
1:A:137:GLY:HA3	6:A:446:HOH:O	2.17	0.45
1:A:104:VAL:HG13	1:A:163[A]:ILE:CD1	2.47	0.44
1:A:215:ALA:O	1:A:219:ILE:HG13	2.16	0.44
1:A:211:ARG:NH2	1:A:214[A]:GLU:OE2	2.51	0.44
1:A:53:HIS:ND1	1:A:57:GLY:HA3	2.32	0.44
1:A:23:LYS:HD2	1:A:107:ASP:OD1	2.18	0.43
1:A:211:ARG:NH2	1:A:214[B]:GLU:OE2	2.52	0.43
1:A:28:VAL:HG22	1:A:212[B]:LEU:HD11	2.01	0.43
1:A:122:GLN:CB	1:A:195:SER:HB3	2.49	0.42
1:A:121:LEU:HD22	1:A:126:HIS:CG	2.56	0.41

There are no symmetry-related clashes.

5.3 Torsion angles

5.3.1 Protein backbone

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	A	219/231 (95%)	215 (98%)	4 (2%)	0	100	100

There are no Ramachandran outliers to report.

CONFIDENTIAL VALIDATION REPORT

5.3.2 Protein sidechains

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution. The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
1	A	165/177 (93%)	165 (100%)	0	100 100

There are no protein residues with a non-rotameric sidechain to report.

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

5.3.3 RNA

There are no RNA chains in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates

There are no carbohydrates in this entry.

5.6 Ligand geometry

5 ligands are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
2	GOL	A	301[A]	-	5,5,5	0.41	0	5,5,5	0.18	0
2	GOL	A	301[B]	-	5,5,5	0.40	0	5,5,5	0.31	0

CONFIDENTIAL VALIDATION REPORT

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\|>2$	Counts	RMSZ	$\#\|Z\|>2$
3	SO4	A	302	-	4,4,4	0.22	0	6,6,6	0.12	0
4	491	A	303	-	24,24,24	1.65	6 (25%)	35,35,35	1.14	2 (5%)
5	2PN	A	304	-	8,8,8	2.08	3 (37%)	9,13,13	3.90	1 (11%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	GOL	A	301[A]	-	-	0/4/4/4	0/0/0/0
2	GOL	A	301[B]	-	-	0/4/4/4	0/0/0/0
3	SO4	A	302	-	-	0/0/0/0	0/0/0/0
4	491	A	303	-	-	0/19/39/39	0/1/1/1
5	2PN	A	304	-	-	0/2/6/6	0/0/0/0

All (9) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	A	303	491	C07-C08	-3.76	1.46	1.53
5	A	304	2PN	P2-O4	3.70	1.50	1.46
5	A	304	2PN	P1-O1	3.38	1.50	1.46
4	A	303	491	O16-C09	-3.28	1.35	1.42
4	A	303	491	O10-C09	3.06	1.49	1.41
4	A	303	491	C22-N21	2.70	1.44	1.34
5	A	304	2PN	P2-N1	2.56	1.66	1.64
4	A	303	491	C24-C22	2.51	1.55	1.50
4	A	303	491	C02-C03	-2.41	1.48	1.52

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
5	A	304	2PN	P2-N1-P1	-11.05	111.49	130.07
4	A	303	491	O04-C03-C02	3.65	121.67	113.63
4	A	303	491	O06-C07-C12	2.14	112.59	107.17

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

CONFIDENTIAL VALIDATION REPORT

5.7 Other polymers

There are no such residues in this entry.

5.8 Polymer linkage issues

There are no chain breaks in this entry.

CONFIDENTIAL VALIDATION REPORT

6 Fit of model and data

6.1 Protein, DNA and RNA chains

In the following table, the column labelled '#RSRZ> 2' contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95^{th} percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled 'Q< 0.9' lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2	OWAB(Å^2)	Q<0.9
1	A	216/231 (93%)	0.14	9 (4%) 35 28	35, 50, 76, 103	0

All (9) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	137	GLY	6.0
1	A	221	GLU	3.4
1	A	122	GLN	2.9
1	A	176	ALA	2.9
1	A	219	ILE	2.8
1	A	69[A]	ARG	2.4
1	A	174	CYS	2.3
1	A	142	ARG	2.2
1	A	218	LEU	2.1

6.2 Non-standard residues in protein, DNA, RNA chains

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates

There are no carbohydrates in this entry.

6.4 Ligands

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95^{th} percentile and maximum values of B factors

CONFIDENTIAL VALIDATION REPORT

of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSR	LLDF	B-factors(Å^2)	Q<0.9
2	GOL	A	301[B]	6/6	0.20	0.60	50,60,63,63	14
5	2PN	A	304	9/9	0.15	0.59	52,87,117,121	14
3	SO4	A	302	5/5	0.13	0.43	94,94,94,94	0
4	491	A	303	24/24	0.13	0.41	39,49,57,58	41
2	GOL	A	301[A]	6/6	0.20	0.41	50,60,62,63	14

6.5 Other polymers (i)

There are no such residues in this entry.

CONFIDENTIAL VALIDATION REPORT

Full wwPDB X-ray Structure Validation Report (i)

Feb 16, 2015 – 06:16 AM EST

PDB ID	:	4Y7U
Title	:	Structural analysis of MurU
Authors	:	Renner-Schneck, M.G.; Stehle, T.
Deposited on	:	2015-02-16
Resolution	:	1.70 Å(reported)

DISCLAIMER
This is a preliminary version of the new style of wwPDB validation report.
We welcome your comments at validation@mail.wwpdb.org
A user guide is available at http://wwpdb.org/ValidationPDFNotes.html

CONFIDENTIAL VALIDATION REPORT

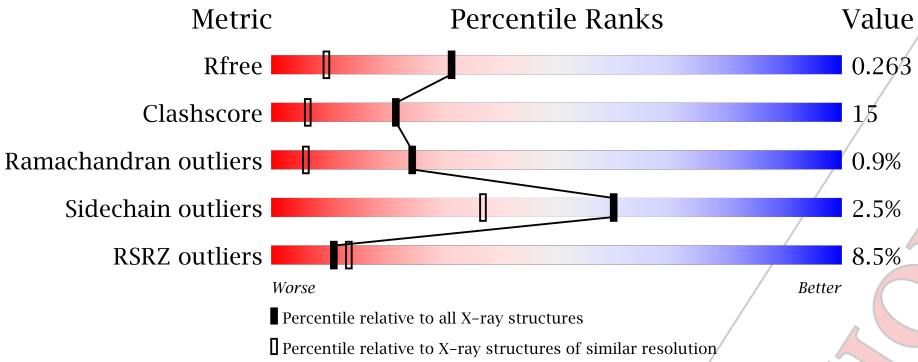
The following versions of software and data (see references) were used in the production of this report:

MolProbity	:	4.02b-467
Mogul	:	1.16 November 2013
Xtriage (Phenix)	:	dev-1439
EDS	:	stable24195
Percentile statistics	:	21963
Refmac	:	5.8.0049
CCP4	:	6.3.0 (Settle)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP)	:	stable24195

1 Overall quality at a glance

The reported resolution of this entry is 1.70 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	66092	2456 (1.70-1.70)
Clashscore	79885	2929 (1.70-1.70)
Ramachandran outliers	78287	2878 (1.70-1.70)
Sidechain outliers	78261	2878 (1.70-1.70)
RSRZ outliers	66119	2456 (1.70-1.70)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Mol	Chain	Length	Quality of chain
1	A	231	

The following table lists non-polymeric compounds that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Geometry	Electron density
4	MG	A	305	-	X
5	2KH	A	306	-	X
6	491	A	307	-	X

CONFIDENTIAL VALIDATION REPORT

2 Entry composition

There are 7 unique types of molecules in this entry. The entry contains 1844 atoms, of which 16 are hydrogens and 0 are deuterium.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

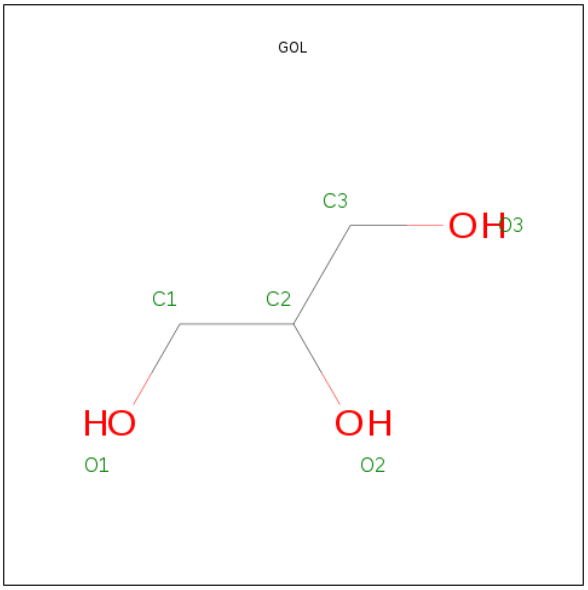
- Molecule 1 is a protein called Nucleotidyl transferase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	224	1702	1099	304	294	5	0	8	0

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	224	LEU	-	expression tag	UNP E4RE40
A	225	GLU	-	expression tag	UNP E4RE40
A	226	HIS	-	expression tag	UNP E4RE40
A	227	HIS	-	expression tag	UNP E4RE40
A	228	HIS	-	expression tag	UNP E4RE40
A	229	HIS	-	expression tag	UNP E4RE40
A	230	HIS	-	expression tag	UNP E4RE40
A	231	HIS	-	expression tag	UNP E4RE40

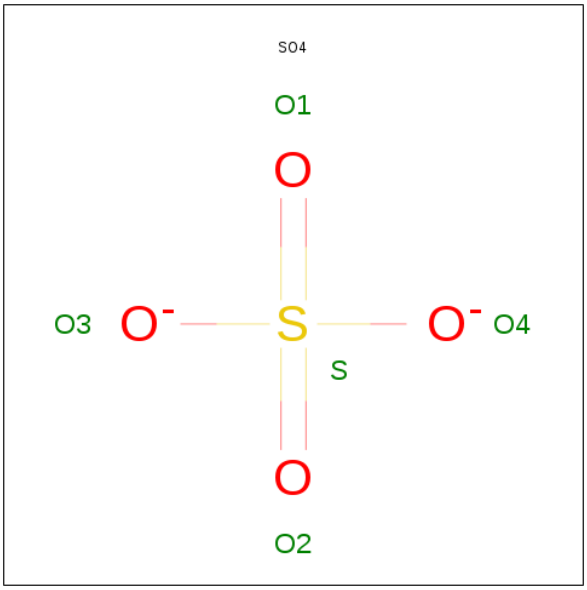
- Molecule 2 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).



VALIDATION REPORT

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	1	Total C H O 14 3 8 3	0	0
2	A	1	Total C H O 14 3 8 3	0	0

- Molecule 3 is SULFATE ION (three-letter code: SO4) (formula: O_4S).



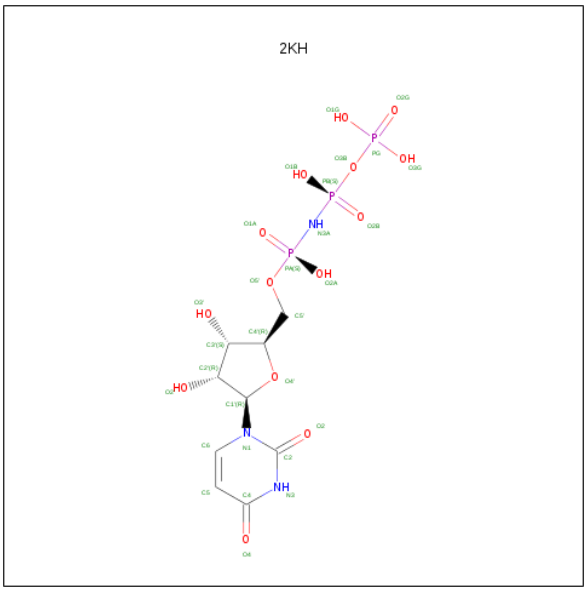
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total O S 5 4 1	0	0

- Molecule 4 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	2	Total Mg 2 2	0	0

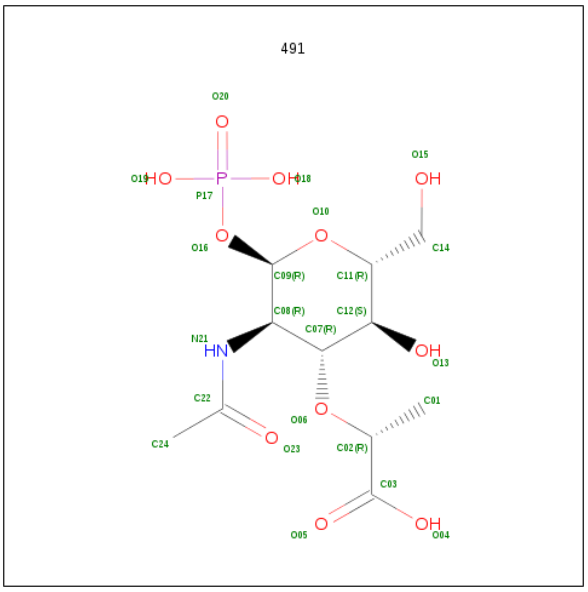
- Molecule 5 is 5'-O-[(S)-hydroxy{[(S)-hydroxy(phosphonooxy)phosphoryl]amino}phosphoryl]uridine (three-letter code: 2KH) (formula: $C_9H_{16}N_3O_{14}P_3$).

CONFIDENTIAL VALIDATION REPORT



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
			Total	C	N	O	P		
5	A	1	29	9	3	14	3	0	0

- Molecule 6 is 2-(acetylamino)-3-O-[(1R)-1-carboxyethyl]-2-deoxy-1-O-phosphono-alpha-D-glucopyranose (three-letter code: 491) (formula: $C_{11}H_{20}NO_{11}P$).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
			Total	C	N	O	P		
6	A	1	24	11	1	11	1	0	0

- Molecule 7 is water.

CONFIDENTIAL VALIDATION REPORT

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
7	A	54	Total O 54 54	0	0

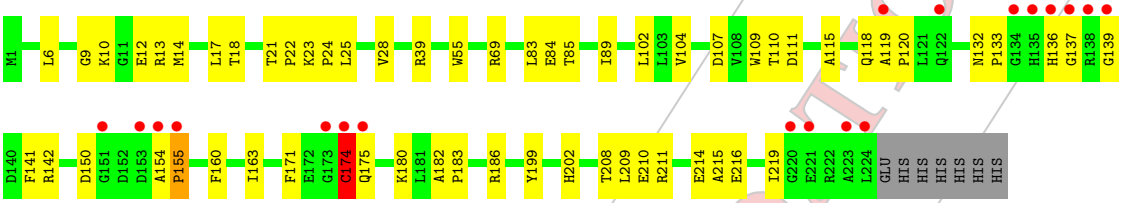
CONFIDENTIAL VALIDATION REPORT

3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density (RSRZ > 2). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

- Molecule 1: Nucleotidyl transferase

Chain A:



CONFIDENTIAL VALIDATION REPORT

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 61 2 2	Depositor
Cell constants a, b, c, α, β, γ	72.75Å 72.75Å 163.13Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	49.86 – 1.70 49.86 – 1.70	Depositor EDS
% Data completeness (in resolution range)	100.0 (49.86-1.70) 100.0 (49.86-1.70)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.18 (at 1.70Å)	Xtriage
Refinement program	PHENIX	Depositor
R, R_{free}	0.228 , 0.264 0.228 , 0.263	Depositor DCC
R_{free} test set	1446 reflections (5.00%)	DCC
Wilson B-factor (Å^2)	39.2	Xtriage
Anisotropy	0.211	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.36 , 47.6	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning	$< \|L\| > = 0.44$, $< L^2 > = 0.27$	Xtriage
Outliers	0 of 28942 reflections	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	1844	wwPDB-VP
Average B, all atoms (Å^2)	57.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.83% of the height of the origin peak. No significant pseudotranslation is detected.*

[1] Intensities estimated from amplitudes.

CONFIDENTIAL VALIDATION REPORT

5 Model quality

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: GOL, MG, 2KH, SO4, 491

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	$\#\|Z\| > 5$	RMSZ	$\#\|Z\| > 5$
1	A	0.37	0/1768	0.53	0/2400

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

5.2 Close contacts

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogens added by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, and the number in parentheses is this value normalized per 1000 atoms of the molecule in the chain. The Symm-Clashes column gives symmetry related clashes, in the same way as for the Clashes column.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	1702	0	1728	51	0
2	A	12	16	16	0	0
3	A	5	0	0	1	0
4	A	2	0	0	0	0
5	A	29	0	16	12	0
6	A	24	0	0	0	0
7	A	54	0	0	2	0
All	All	1828	16	1760	53	0

Clashscore is defined as the number of clashes calculated for the entry per 1000 atoms (including hydrogens) of the entry. The overall clashscore for this entry is 15.

CONFIDENTIAL VALIDATION REPORT

All (53) close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:A:107:ASP:OD2	5:A:306:2KH:H6	1.71	0.89
1:A:12:GLU:N	5:A:306:2KH:O1G	2.10	0.85
1:A:9:GLY:HA3	5:A:306:2KH:H12	1.64	0.80
1:A:182:ALA:HB3	1:A:183:PRO:HD3	1.70	0.74
1:A:13:ARG:NE	5:A:306:2KH:O3G	2.22	0.73
1:A:142[A]:ARG:NH2	1:A:155:PRO:HG2	2.05	0.72
1:A:13:ARG:HG3	5:A:306:2KH:O3G	1.95	0.67
1:A:137:GLY:HA3	1:A:154:ALA:HB1	1.78	0.66
1:A:13:ARG:CG	5:A:306:2KH:O3G	2.43	0.66
1:A:39[A]:ARG:HD2	7:A:448:HOH:O	1.98	0.63
1:A:115:ALA:O	1:A:118:GLN:HG2	2.01	0.61
1:A:174:CYS:HA	7:A:414:HOH:O	2.01	0.61
1:A:136:HIS:HE1	1:A:139:GLY:H	1.50	0.59
1:A:23:LYS:HB3	1:A:24:PRO:HD3	1.85	0.59
1:A:150:ASP:OD1	1:A:186:ARG:NH2	2.39	0.56
1:A:10:LYS:HG3	1:A:55:TRP:CD2	2.41	0.55
1:A:142[B]:ARG:HG2	1:A:142[B]:ARG:HH11	1.72	0.55
1:A:171:PHE:O	1:A:174:CYS:HB2	2.07	0.54
1:A:23:LYS:NZ	1:A:107:ASP:OD2	2.39	0.53
1:A:136:HIS:CE1	1:A:139:GLY:H	2.25	0.53
1:A:209:LEU:HD22	1:A:209:LEU:N	2.24	0.52
1:A:102:LEU:CD2	1:A:104:VAL:HG23	2.40	0.52
1:A:28:VAL:O	1:A:216:GLU:HG2	2.11	0.50
1:A:9:GLY:CA	5:A:306:2KH:H12	2.39	0.50
1:A:14:MET:HE3	1:A:17:LEU:HD12	1.93	0.50
1:A:211:ARG:NH2	1:A:214:GLU:OE2	2.46	0.49
1:A:83:LEU:O	1:A:84:GLU:HB2	2.12	0.48
1:A:22:PRO:HG2	1:A:25:LEU:HG	1.96	0.47
1:A:215:ALA:O	1:A:219:ILE:HG13	2.14	0.47
1:A:6:LEU:O	5:A:306:2KH:O2'	2.33	0.47
1:A:23:LYS:N	1:A:24:PRO:CD	2.78	0.47
5:A:306:2KH:O1B	5:A:306:2KH:O1G	2.27	0.47
1:A:13:ARG:HB2	5:A:306:2KH:O3G	2.15	0.46
1:A:208:THR:OG1	1:A:211:ARG:HG2	2.15	0.46
1:A:111:ASP:OD2	1:A:202:HIS:HB2	2.16	0.45
1:A:110:THR:HG21	1:A:163[A]:ILE:HD11	1.97	0.45
1:A:109:TRP:CZ2	1:A:219:ILE:HG12	2.52	0.44
1:A:142[A]:ARG:HH22	1:A:155:PRO:HG2	1.80	0.44
1:A:13:ARG:CB	5:A:306:2KH:O3G	2.65	0.43
5:A:306:2KH:H4	5:A:306:2KH:H3	1.65	0.42
1:A:154:ALA:N	1:A:155:PRO:CD	2.83	0.42
1:A:110:THR:HG21	1:A:163[B]:ILE:HD11	2.00	0.41

Continued on next page...

CONFIDENTIAL VALIDATION REPORT

Continued from previous page...

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:A:132:ASN:HA	1:A:133:PRO:HD3	1.87	0.41
1:A:141:PHE:HE2	1:A:160:PHE:HB2	1.86	0.41
1:A:210:GLU:OE2	1:A:211:ARG:HD2	2.20	0.41
1:A:142[B]:ARG:HG2	1:A:142[B]:ARG:NH1	2.36	0.41
1:A:154:ALA:N	1:A:155:PRO:HD3	2.35	0.41
1:A:10:LYS:HE3	1:A:10:LYS:HB2	1.95	0.41
1:A:85:THR:HB	3:A:303:SO4:O2	2.21	0.41
1:A:119:ALA:HB1	1:A:120:PRO:HD2	2.03	0.41
1:A:18:THR:HA	1:A:21:THR:O	2.21	0.41
1:A:111:ASP:OD1	1:A:199:TYR:OH	2.28	0.40
1:A:89:ILE:HG21	1:A:171:PHE:CZ	2.56	0.40

There are no symmetry-related clashes.

5.3 Torsion angles

5.3.1 Protein backbone

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
1	A	230/231 (100%)	214 (93%)	14 (6%)	2 (1%)	25 6

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	174	CYS
1	A	155	PRO

5.3.2 Protein sidechains

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution. The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

CONFIDENTIAL VALIDATION REPORT

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
1	A	168/177 (95%)	163 (97%)	5 (3%)	53 29

All (5) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	69[A]	ARG
1	A	69[C]	ARG
1	A	174	CYS
1	A	175	GLN
1	A	180	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

5.3.3 RNA (i)

There are no RNA chains in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains (i)

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates (i)

There are no carbohydrates in this entry.

5.6 Ligand geometry (i)

Of 7 ligands modelled in this entry, 2 are monoatomic - leaving 5 for Mogul analysis.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
2	GOL	A	301	-	5,5,5	0.38	0	5,5,5	0.36	0

CONFIDENTIAL VALIDATION REPORT

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\|>2$	Counts	RMSZ	$\#\|Z\|>2$
2	GOL	A	302	-	5,5,5	0.36	0	5,5,5	0.18	0
3	SO4	A	303	-	4,4,4	0.21	0	6,6,6	0.09	0
5	2KH	A	306	4	30,30,30	3.63	14 (46%)	40,47,47	2.65	12 (30%)
6	491	A	307	-	24,24,24	1.81	6 (25%)	35,35,35	1.03	2 (5%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	GOL	A	301	-	-	0/4/4/4	0/0/0/0
2	GOL	A	302	-	-	0/4/4/4	0/0/0/0
3	SO4	A	303	-	-	0/0/0/0	0/0/0/0
5	2KH	A	306	4	-	0/17/38/38	0/2/2/2
6	491	A	307	-	-	0/19/39/39	0/1/1/1

All (20) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	A	306	2KH	C5-C4	9.04	1.47	1.37
5	A	306	2KH	C2-N1	6.90	1.45	1.38
5	A	306	2KH	PB-O3B	-6.49	1.51	1.59
5	A	306	2KH	C6-N1	5.71	1.44	1.35
5	A	306	2KH	PG-O3B	-5.25	1.51	1.60
5	A	306	2KH	C4-N3	4.77	1.43	1.36
5	A	306	2KH	PA-N3A	4.41	1.68	1.64
5	A	306	2KH	C3'-C2'	-4.41	1.41	1.53
5	A	306	2KH	C6-C5	4.21	1.47	1.38
5	A	306	2KH	C2'-C1'	-4.11	1.47	1.53
5	A	306	2KH	PB-N3A	4.10	1.67	1.64
6	A	307	491	C07-C08	-3.87	1.46	1.53
5	A	306	2KH	PB-O2B	3.47	1.50	1.46
6	A	307	491	O16-C09	-3.35	1.35	1.42
6	A	307	491	O10-C09	3.35	1.50	1.41
5	A	306	2KH	C2-N3	3.05	1.46	1.36
5	A	306	2KH	PA-O1A	3.02	1.49	1.46
6	A	307	491	C22-N21	2.98	1.46	1.34
6	A	307	491	C02-C03	-2.92	1.47	1.52
6	A	307	491	C24-C22	2.70	1.56	1.50

All (14) bond angle outliers are listed below:

CONFIDENTIAL VALIDATION REPORT

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
5	A	306	2KH	PA-N3A-PB	-8.16	117.35	130.03
5	A	306	2KH	PG-O3B-PB	-6.83	108.60	132.05
5	A	306	2KH	C2-N1-C1'	-6.27	114.27	118.21
5	A	306	2KH	C4'-O4'-C1'	-4.63	104.63	109.72
5	A	306	2KH	O5'-C5'-C4'	4.41	125.16	108.96
5	A	306	2KH	O4'-C1'-N1	3.49	115.73	108.08
5	A	306	2KH	O3B-PB-N3A	3.28	115.69	106.59
5	A	306	2KH	C6-N1-C2	3.14	123.89	119.51
5	A	306	2KH	O2A-PA-O1A	-2.84	103.91	109.90
6	A	307	491	O04-C03-C02	2.78	119.76	113.63
5	A	306	2KH	O1A-PA-N3A	-2.63	107.86	111.83
5	A	306	2KH	O5'-PA-N3A	2.54	114.77	107.16
5	A	306	2KH	O3'-C3'-C4'	2.11	117.27	111.07
6	A	307	491	O10-C11-C12	2.06	113.56	109.73

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

5.7 Other polymers

There are no such residues in this entry.

5.8 Polymer linkage issues

There are no chain breaks in this entry.

CONFIDENTIAL VALIDATION REPORT

6 Fit of model and data

6.1 Protein, DNA and RNA chains

In the following table, the column labelled '#RSRZ> 2' contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95^{th} percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled 'Q< 0.9' lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2			OWAB(Å^2)	Q<0.9
1	A	224/231 (96%)	0.44	19 (8%)	11	13	35, 51, 104, 145	0

All (19) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	137	GLY	13.5
1	A	154	ALA	13.3
1	A	136	HIS	5.8
1	A	224	LEU	5.6
1	A	175	GLN	5.3
1	A	134	GLY	4.8
1	A	135	HIS	4.6
1	A	155	PRO	4.5
1	A	220	GLY	4.4
1	A	153	ASP	3.7
1	A	223	ALA	3.4
1	A	139	GLY	3.2
1	A	138	ARG	2.9
1	A	151	GLY	2.8
1	A	221	GLU	2.5
1	A	174	CYS	2.3
1	A	119	ALA	2.3
1	A	173	GLY	2.2
1	A	122	GLN	2.0

6.2 Non-standard residues in protein, DNA, RNA chains

There are no non-standard protein/DNA/RNA residues in this entry.

CONFIDENTIAL VALIDATION REPORT

6.3 Carbohydrates (i)

There are no carbohydrates in this entry.

6.4 Ligands (i)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95^{th} percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSR	LLDF	B-factors(Å^2)	Q<0.9
5	2KH	A	306	29/29	0.34	9.70	79,91,159,196	29
6	491	A	307	24/24	0.18	3.27	40,64,88,90	24
4	MG	A	305	1/1	0.15	2.61	77,77,77,77	0
2	GOL	A	302	6/6	0.27	1.22	66,79,84,84	14
2	GOL	A	301	6/6	0.20	1.22	53,67,75,80	0
3	SO4	A	303	5/5	0.11	0.24	104,105,105,105	5
4	MG	A	304	1/1	0.07	-1.54	60,60,60,60	1

6.5 Other polymers (i)

There are no such residues in this entry.

CONFIDENTIAL VALIDATION REPORT

11.2 Complex structure: MurU with UDP + MurNAc (loop no. 0188)

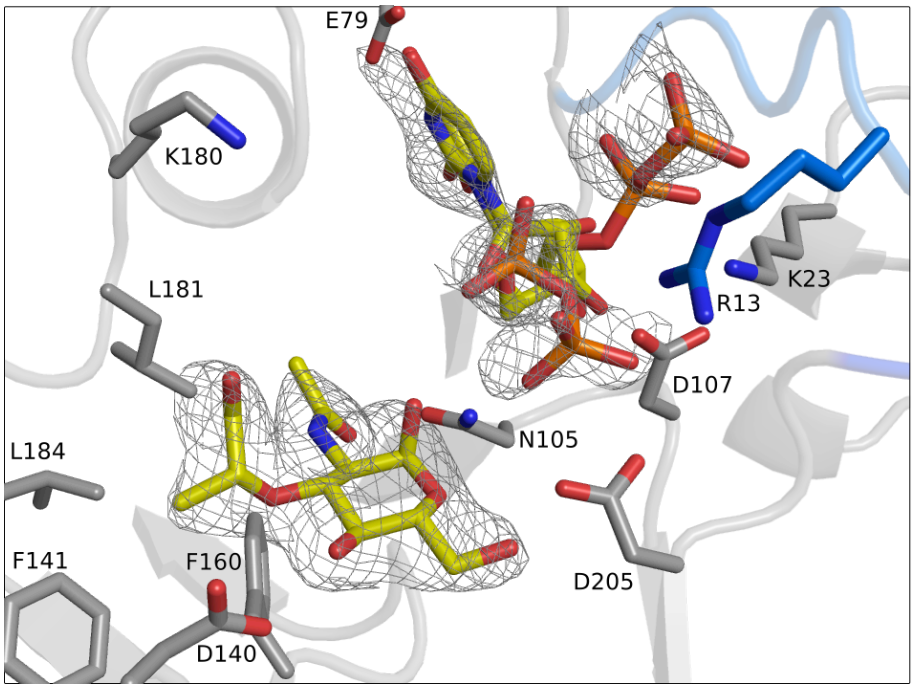


Fig.-S 1: Crystal structure of MurU soaked with MurNAc and UDP with omit difference maps (*mFo-DFc*) for the ligands contoured at 3.0 σ. Ligands and side-chains that contribute to ligand coordination are depicted in stick representation and colored by atom type.

Interestingly UDP adopts two conformations. One with the α -phosphate pointing away from MurNAc allowing the α -phosphate to be coordinated by the signature motif and the other with the α -phosphate oriented in a way that would be suitable for the formation of UDP-MurNAc in the active enzyme.

11.3 *In silico* characterization of the Protein contructs

11.3.1 MurU-His$_6$

```
         10         20         30         40         50         60
MKAMILAAGK GERMRPLTLH TPKPLVPVAG QPLIEYHLRA LAAAGVTEVV INHAWLGQQI

         70         80         90        100        110        120
EDHLGDGSRF GLSIRYSPEG EPLETGGGIF KALPLLGDAP FLLVNGDVWT DYDFARLQAP

        130        140        150        160        170        180
LQGLAHLVLV DNPGHHGRGD FRLVGEQVVD GDDAPGTLTF SGISVLHPAL FEGCQAGAFK

        190        200        210        220        230
LAPLLRQAMA AGKVSGEHYR GHWVDVGTLE RLAEAESLIG ERALEHHHHH H
```

Number of amino acids: 231
Molecular weight: 24879.3
Theoretical pI: 5.93

Amino acid composition:

Ala (A)	25	10.8%	Phe (F)	8	3.5%
Arg (R)	12	5.2%	Pro (P)	14	6.1%
Asn (N)	3	1.3%	Ser (S)	7	3.0%
Asp (D)	12	5.2%	Thr (T)	8	3.5%
Cys (C)	1	0.4%	Trp (W)	3	1.3%
Gln (Q)	8	3.5%	Tyr (Y)	4	1.7%
Glu (E)	15	6.5%	Val (V)	16	6.9%
Gly (G)	29	12.6%	Pyl (O)	0	0.0%
His (H)	16	6.9%	Sec (U)	0	0.0%
Ile (I)	8	3.5%			
Leu (L)	32	13.9%	(B)	0	0.0%
Lys (K)	6	2.6%	(Z)	0	0.0%
Met (M)	4	1.7%	(X)	0	0.0%

Total number of negatively charged residues (Asp + Glu): 27
Total number of positively charged residues (Arg + Lys): 18

Atomic composition:

Carbon	C	1119	Sulfur	S	5
Hydrogen	H	1741			
Nitrogen	N	319	**Formula:**	$C_{1119}H_{1741}N_{319}O_{316}S_5$	
Oxygen	O	316	**Total number of atoms:**	3500	

Extinction coefficients:

Extinction coefficients are in units of M^{-1} cm^{-1}, at 280 nm measured in water.

Ext. coefficient	22460
Abs 0.1% (=1 g/l)	0.903, assuming all pairs of Cys residues form cystines
Ext. coefficient	22460
Abs 0.1% (=1 g/l)	0.903, assuming all Cys residues are reduced

Apx. 1: *In silico* characterization of the MurU-His$_6$ expression construct using the ProtParam online tool of the ExPASy server (Gasteiger E.).

11.3.2 His_6-MurQ

```
        10         20         30         40         50         60
METRGSHHHH HHGSSEPLNL HRLTTESRNS QTVEIHKANT LGILKMINNE DMKVAAAVQE

        70         80         90        100        110        120
VLPDIKTAVD CAYESFQNGG RLIYTGAGTS GRLGVMDAVE CPPTYSVSPD QVIGIMAGGP

       130        140        150        160        170        180
EAFLQAAEGI EDSEEAGAED LRNIQLTSND TVIAIAASGR TPYAAGALRY ARKVGAHTIA

       190        200        210        220        230        240
LTCNENSAIS KDADHSIEVV VGPEAITGST RMKAATAHKM ILNMISTAVM VKIGKVYENL

       250        260        270        280        290        300
MVDVNVSNKK LKERAISIIQ SLTNASYDTA RYTLEQADHH VKTAIVMLKT STDQKQAQTL

       310
LDEANGFIDK AIEHYHP
```

Number of amino acids: 317
Molecular weight: 34197.6
Theoretical pI: 5.91

Amino acid composition:

Ala (A)	38	12.0%	Ile (I)	24	7.6%	Tyr (Y)	9	2.8%
Arg (R)	12	3.8%	Leu (L)	21	6.6%	Val (V)	22	6.9%
Asn (N)	16	5.0%	Lys (K)	17	5.4%	Pyl (O)	0	0.0%
Asp (D)	16	5.0%	Met (M)	11	3.5%	Sec (U)	0	0.0%
Cys (C)	3	0.9%	Phe (F)	3	0.9%			
Gln (Q)	11	3.5%	Pro (P)	9	2.8%	(B)	0	0.0%
Glu (E)	22	6.9%	Ser (S)	22	6.9%	(Z)	0	0.0%
Gly (G)	21	6.6%	Thr (T)	25	7.9%	(X)	0	0.0%
His (H)	15	4.7%	Trp (W)	0	0.0%			

Total number of negatively charged residues (Asp + Glu): 38
Total number of positively charged residues (Arg + Lys): 29

Atomic composition:

Carbon	C	1476
Hydrogen	H	2389
Nitrogen	N	427
Oxygen	O	477
Sulfur	S	14

Formula: $C_{1476}H_{2389}N_{427}O_{477}S_{14}$
Total number of atoms: 4783

Extinction coefficients:

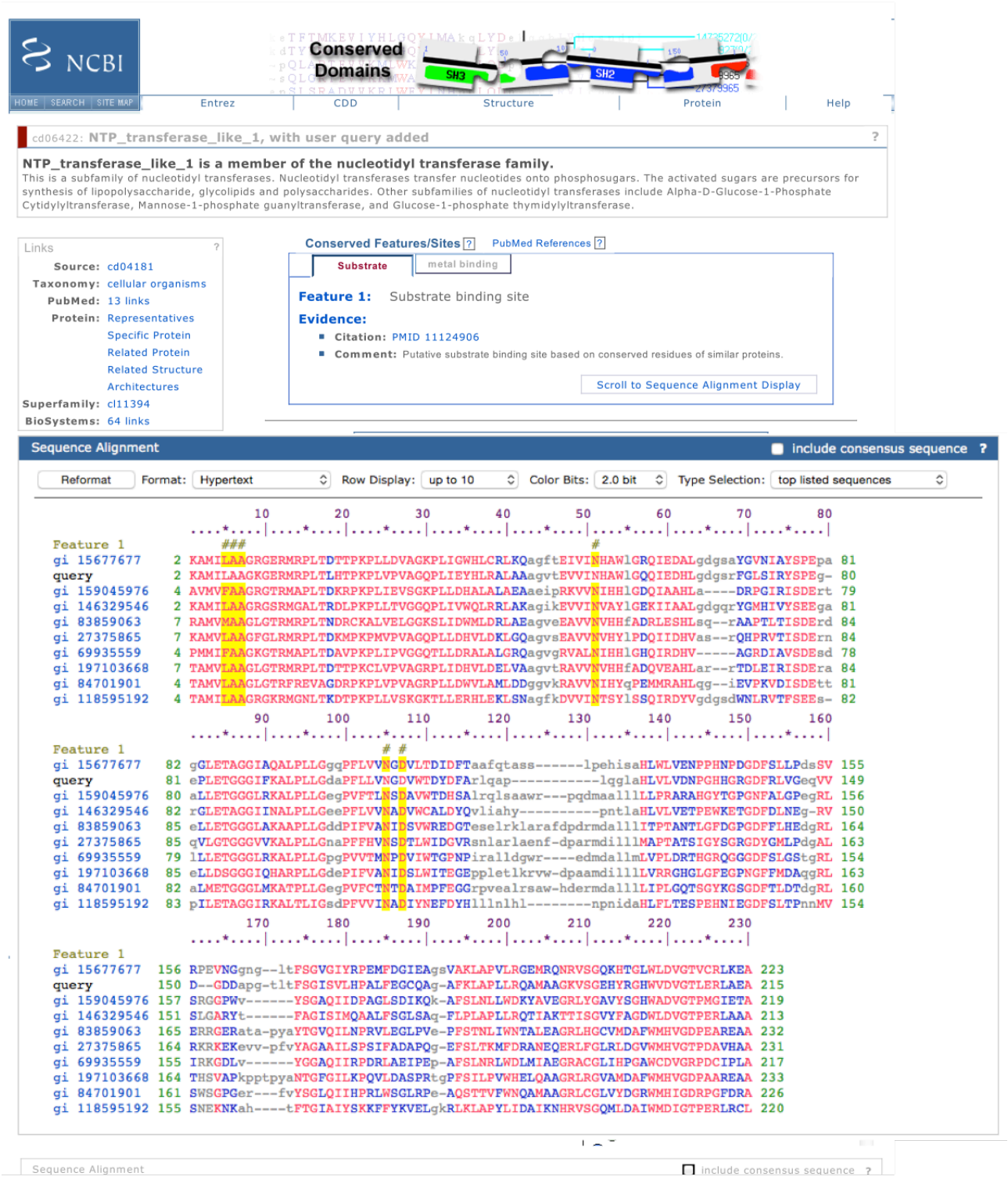
This protein does not contain any Trp residues. Experience shows that this could result in more than 10% error in the computed extinction coefficient.
Extinction coefficients are in units of M^{-1} cm^{-1}, at 280 nm measured in water.

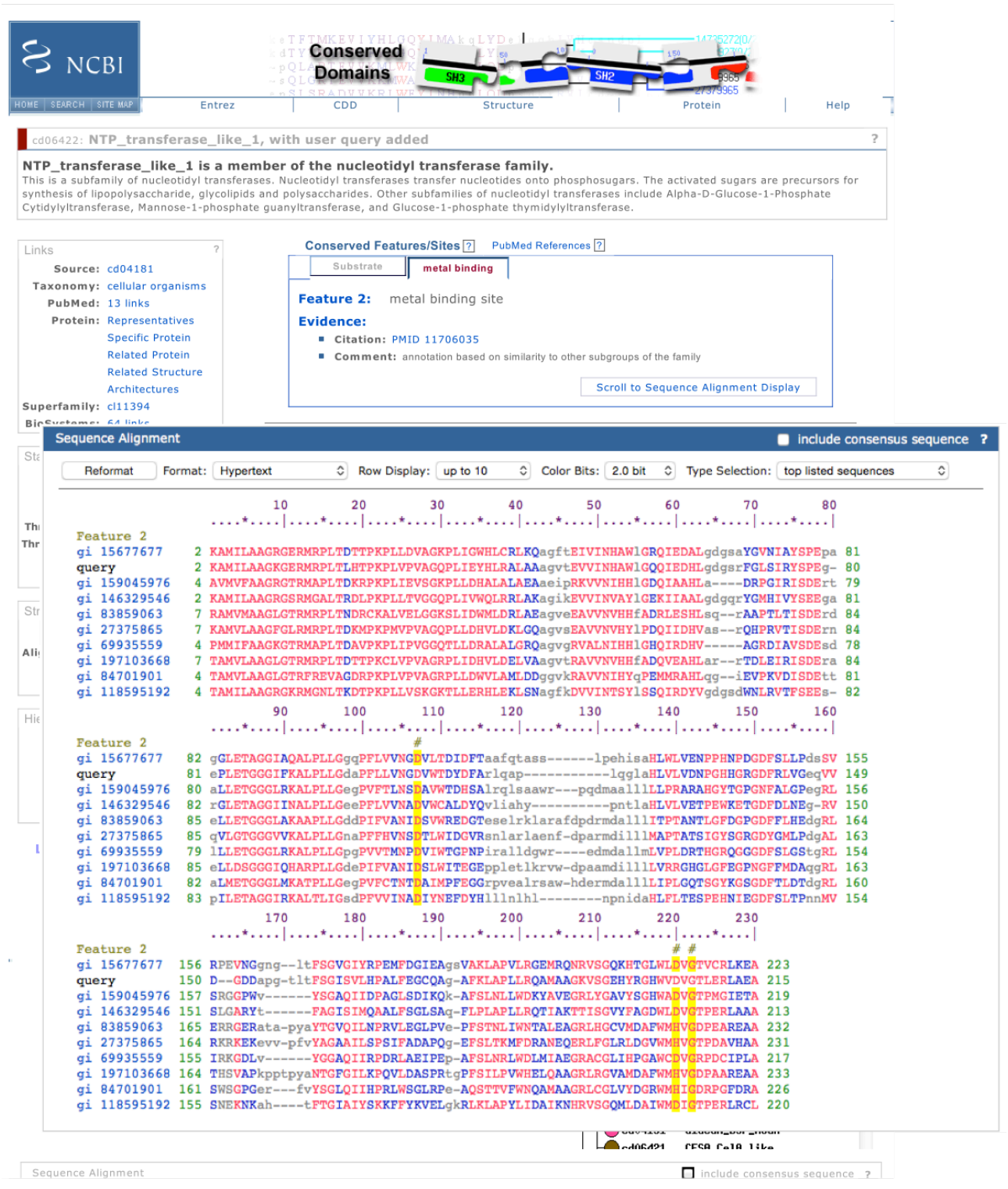
Ext. coefficient 13535
Abs 0.1% (=1 g/l) 0.396, assuming all pairs of Cys residues form cystines

Ext. coefficient 13410
Abs 0.1% (=1 g/l) 0.392, assuming all Cys residues are reduced

Apx. 2: *In silico* characterization of the His_6-MurQ expression construct using the ProtParam online tool of the ExPASy server (Gasteiger E.).

11.4 BLAST® search results MurU





Apx. 3: Results of a BLAST®-query of MurU's amino acid sequence against the conserved domain database (CDD) using the CD-Search Tool provided under http://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi.

F

11.5 Crystallization Screens

Screen	GridScreenAmSulf_PEG3350									
Date										
Proteinconc.	7 mg/ml									
Wellvolume	0,55 ml									
A1-H6										
Buffer 1	Citrate	**pH**	4	**Unit**	M	**conc**	0,95	**to-be con**	0,1	
Buffer 2	Citrate	**pH**	5	**Unit**	M	**conc**	1	**to-be con**	0,1	
Buffer 3	MES	**pH**	6	**Unit**	M	**conc**	1	**to-be con**	0,1	
Buffer 4	HEPES	**pH**	7	**Unit**	M	**conc**	1	**to-be con**	0,1	
Buffer 5	Tris	**pH**	8	**Unit**	M	**conc**	1	**to-be con**	0,1	
Buffer 6	BICINE	**pH**	9	**Unit**	M	**conc**	1	**to-be con**	0,1	
Salt				**Unit**		**conc**	20	**to-be con**		
Precipitant 1				**Unit**		**conc**	4	**to-be con**		
Precipitant 2	Ammonium Sulfate			**Unit**	M	**conc**	4	**start con**	**0,1**	**end conc** **1,45**
A7-H12										
Buffer 1	Citrate	**pH**	4	**Unit**	M	**conc**	0,95	**to-be con**	0,1	
Buffer 2	Citrate	**pH**	5	**Unit**	M	**conc**	1	**to-be con**	0,1	
Buffer 3	MES	**pH**	6	**Unit**	M	**conc**	1	**to-be con**	0,1	
Buffer 4	HEPES	**pH**	7	**Unit**	M	**conc**	1	**to-be con**	0,1	
Buffer 5	Tris	**pH**	8	**Unit**	M	**conc**	1	**to-be con**	0,1	
Buffer 6	BICINE	**pH**	9	**Unit**	M	**conc**	1	**to-be con**	0,1	
Salt				**Unit**		**conc**	20	**to-be con**		
Precipitant 1				**Unit**		**conc**	1	**to-be con**		
Precipitant 2	PEG 3350			**Unit**	%w/v	**conc**	50	**start con**	10	**end conc** 22
D1-H6										
Buffer 1	Citrate	pH	4	**Unit**	M	**conc**	0,95	**to-be con**	0,1	
Buffer 2	Citrate	pH	5	**Unit**	M	**conc**	1	**to-be con**	0,1	
Buffer 3	MES	pH	6	**Unit**	M	**conc**	1	**to-be con**	0,1	
Buffer 4	HEPES	pH	7	**Unit**	M	**conc**	1	**to-be con**	0,1	
Buffer 5	Tris	pH	8	**Unit**	M	**conc**	1	**to-be con**	0,1	
Buffer 6	BICINE	pH	9	**Unit**	M	**conc**	1	**to-be con**	0,1	
Salt				**Unit**		**conc**	20	**to-be con**		
Precipitant 1				**Unit**		**conc**	1	**to-be con**		
Precipitant 2	Ammonium Sulfate			**Unit**	M	**conc**	4	**start con**	1,9	**end conc** 3,25
D7-H12										
Buffer 1	Citrate	pH	4	**Unit**	M	**conc**	0,95	**to-be con**	0,1	
Buffer 2	Citrate	pH	5	**Unit**	M	**conc**	1	**to-be con**	0,1	
Buffer 3	MES	pH	6	**Unit**	M	**conc**	1	**to-be con**	0,1	
Buffer 4	HEPES	pH	7	**Unit**	M	**conc**	1	**to-be con**	0,1	
Buffer 5	Tris	pH	8	**Unit**	M	**conc**	1	**to-be con**	0,1	
Buffer 6	BICINE	pH	9	**Unit**	M	**conc**	1	**to-be con**	0,1	
Salt				**Unit**		**conc**	20	**to-be con**		
Precipitant 1				**Unit**		**conc**	100	**to-be con**		
Precipitant 2	PEG 3350			**Unit**	%w/v	**conc**	50	**start con**	**26**	**end conc** **38**

AmSulf/PEG fine

8,6 mg/ml
0,50 ml

Citrate	**pH**	5	**Unit**	M	**conc**	1	**to-be conc**	0,1		
Citrate	**pH**	5,5	**Unit**	M	**conc**	1	**to-be conc**	0,1		
MES	**pH**	6	**Unit**	M	**conc**	1	**to-be conc**	0,1		
HEPES	**pH**	6,4	**Unit**	M	**conc**	1	**to-be conc**	0,1		
HEPES	**pH**	7	**Unit**	M	**conc**	1	**to-be conc**	0,1		
Bicine	**pH**	9	**Unit**	M	**conc**	1	**to-be conc**	0,1		
			Unit		**conc**	2	**to-be conc**			
			Unit		**conc**	2	**to-be conc**			
(NH4)2SO4			**Unit**	M	**conc**	4	**start conc**	**1,5**	**end conc**	**3**
Citrate	**pH**	3	**Unit**	M	**conc**	1	**to-be conc**	0,1		
Citrate	**pH**	3,5	**Unit**	M	**conc**	1	**to-be conc**	0,1		
Citrate	**pH**	4	**Unit**	M	**conc**	0,95	**to-be conc**	0,1		
Citrate	**pH**	4,5	**Unit**	M	**conc**	1	**to-be conc**	0,1		
Citrate	**pH**	5	**Unit**	M	**conc**	1	**to-be conc**	0,1		
Citrate	**pH**	5,5	**Unit**	M	**conc**	1	**to-be conc**	0,1		
			Unit	M	**conc**	2	**to-be conc**	0,1		
			Unit		**conc**	100	**to-be conc**			
PEG 3350			**Unit**	%v/v	**conc**	50	**start conc**	2	**end conc**	20

Screen Buffer-AS_fine
Date
Proteinconc. 8,0 mg/ml
Wellvolume 0,50 ml

A1-H6

Buffer 1	Citrate	**pH**	5	**Unit**	M	**conc**	1	**to-be conc**	0,1	
Buffer 2	Citrate	**pH**	5,5	**Unit**	M	**conc**	1	**to-be conc**	0,11	
Buffer 3	Citrate	**pH**	6	**Unit**	M	**conc**	0,95	**to-be conc**	0,12	
Buffer 4	MES	**pH**	6	**Unit**	M	**conc**	1	**to-be conc**	0,13	
Buffer 5	MES	**pH**	6,5	**Unit**	M	**conc**	1	**to-be conc**	0,14	
Buffer 6	HEPES	**pH**	6,4	**Unit**	M	**conc**	1	**to-be conc**	0,15	
Salt				**Unit**		**conc**	2	**to-be conc**		
Precipitant 1				**Unit**		**conc**	2	**to-be conc**		
Precipitant 2	AmSulf			**Unit**	M	**conc**	3,95	**start conc**	**2**	**end conc** **1**

A7-H12

Buffer 1	HEPES	**pH**	7	**Unit**	M	**conc**	1	**to-be conc**	0,05	
Buffer 2	HEPES	**pH**	7,5	**Unit**	M	**conc**	1,2	**to-be conc**	0,06	
Buffer 3	Tris	**pH**	8	**Unit**	M	**conc**	0,8	**to-be conc**	0,07	
Buffer 4	Tris	**pH**	8,5	**Unit**	M	**conc**	2	**to-be conc**	0,08	
Buffer 5	BICINE	**pH**	9	**Unit**	M	**conc**	1	**to-be conc**	0,09	
Buffer 6	BICINE	**pH**	9,5	**Unit**	M	**conc**	1	**to-be conc**	0,1	
Salt				**Unit**		**conc**	2	**to-be conc**		
Precipitant 1				**Unit**		**conc**	100	**to-be conc**		
Precipitant 2	AmSulf			**Unit**	M	**conc**	3,95	**start conc**	2	**end conc** 1

Screen	AS_fine									
Date										
Proteinconc.	8,0 mg/ml									
Wellvolume	1,00 ml									
A1-H6										
Buffer 1	Citrate	**pH**	5	**Unit**	M	**conc**	1	**to-be conc**	0,1	
Buffer 2	Citrate	**pH**	5,5	**Unit**	M	**conc**	1	**to-be conc**	0,1	
Buffer 3	MES	**pH**	6,5	**Unit**	M	**conc**	0,95	**to-be conc**	0,1	
Buffer 4	MES	**pH**	6	**Unit**	M	**conc**	1	**to-be conc**	0,1	
Buffer 5	HEPES	**pH**	6,4	**Unit**	M	**conc**	1	**to-be conc**	0,1	
Buffer 6	HEPES	**pH**	7	**Unit**	M	**conc**	1	**to-be conc**	0,1	
Salt				**Unit**		**conc**	10	**to-be conc**		
Precipitant 1				**Unit**		**conc**	10	**to-be conc**		
Precipitant 2	AS			**Unit**	M	**conc**	4	**start conc**	**1,5**	**end conc** **1,12**
A7-H12										
Buffer 1	Citrate	**pH**	5	**Unit**	M	**conc**	1	**to-be conc**	0,1	
Buffer 2	Citrate	**pH**	5,5	**Unit**	M	**conc**	1,2	**to-be conc**	0,1	
Buffer 3	MES	**pH**	6,5	**Unit**	M	**conc**	0,8	**to-be conc**	0,1	
Buffer 4	MES	**pH**	6	**Unit**	M	**conc**	2	**to-be conc**	0,1	
Buffer 5	HEPES	**pH**	6,4	**Unit**	M	**conc**	1	**to-be conc**	0,1	
Buffer 6	HEPES	**pH**	7	**Unit**	M	**conc**	1	**to-be conc**	0,1	
Salt				**Unit**		**conc**	2	**to-be conc**		
Precipitant 1				**Unit**		**conc**	100	**to-be conc**		
Precipitant 2	AS			**Unit**	M	**conc**	4	**start conc**	1,13	**end conc** 0,75

				Screen	MRS_Cit-PEG_fine					
Date										
Proteinconc.	10,0 mg/ml									
Wellvolume	0,75 ml									
Buffer	Citrate	**pH**	5	**Unit**	M	**conc**	1	**to-be conc**	0,1	
Salt				**Unit**		**conc**	1	**to-be conc**		
Salt 2				**Unit**		**conc**	1	**to-be conc**		
Precipitant 1	PEG 1000			**Unit**	%w/v	**conc**	50	**Startconc**	5	**Endconc** 10
Precipitant 2	PEG 1000			**Unit**	%w/v	**conc**	50	**Startconc**	10,5	**Endconc** 15
Precipitant 3	PEG 1000			**Unit**	%w/v	**conc**	50	**Startconc**	15,5	**Endconc** 20
Precipitant 4	PEG 3350			**Unit**	%w/v	**conc**	50	**Startconc**	5	**Endconc** 10
Precipitant 5	PEG 3350			**Unit**	%w/v	**conc**	50	**Startconc**	10,5	**Endconc** 15
Precipitant 6	PEG 3350			**Unit**	%w/v	**conc**	50	**Startconc**	15,5	**Endconc** 20
Precipitant 7	PEG 4000			**Unit**	%w/v	**conc**	50	**Startconc**	5	**Endconc** 10
Precipitant 8	PEG 4000			**Unit**	%w/v	**conc**	50	**Startconc**	10,5	**Endconc** 15
Precipitant 9	PEG 4000			**Unit**	%w/v	**conc**	50	**Startconc**	15,1	**Endconc** 20
Precipitant 10	PEG 6000			**Unit**	%w/v	**conc**	50	**Startconc**	5	**Endconc** 10
Precipitant 11	PEG 6000			**Unit**	%w/v	**conc**	50	**Startconc**	10,5	**Endconc** 15
Precipitant 12	PEG 6000			**Unit**	%w/v	**conc**	50	**Startconc**	15,5	**Endconc** 20

MRS_PEG-fineE87Q

8,0 mg/ml
0,80 ml

Citrate	**pH**	3,5	**Unit**	M	**conc**	1	**to-be conc**	0,1		
Citrate	**pH**	4,5	**Unit**	M	**conc**	1	**to-be conc**	0,1		
Citrate	**pH**	5	**Unit**	M	**conc**	1	**to-be conc**	0,1		
Citrate	**pH**	5,5	**Unit**	M	**conc**	1	**to-be conc**	0,1		
MES	**pH**	6	**Unit**	M	**conc**	1	**to-be conc**	0,1		
MES	**pH**	6,4	**Unit**	M	**conc**	1	**to-be conc**	0,1		
			Unit		**conc**	2	**to-be conc**			
			Unit		**conc**	2	**to-be conc**			
PEG 3350			**Unit**	%w/v	**conc**	50	**start conc**	**12**	**end conc**	**15,8**

Citrate	**pH**	3,5	**Unit**	M	**conc**	1	**to-be conc**	0,1		
Citrate	**pH**	4,5	**Unit**	M	**conc**	1	**to-be conc**	0,1		
Citrate	**pH**	5	**Unit**	M	**conc**	1	**to-be conc**	0,1		
Citrate	**pH**	5,5	**Unit**	M	**conc**	1	**to-be conc**	0,1		
MES	**pH**	6	**Unit**	M	**conc**	1	**to-be conc**	0,1		
MES	**pH**	6,4	**Unit**	M	**conc**	1	**to-be conc**	0,1		
			Unit		**conc**	2	**to-be conc**			
			Unit		**conc**	2	**to-be conc**			
PEG3350			**Unit**	%w/v	**conc**	50	**start conc**	16	**end conc**	20

Screen CoCrys_PEGuAS
Date
Proteinconc. 7,0 mg/ml
Wellvolume 0,50 ml

A1-H6

Buffer 1	Citrate	**pH**	4,5	**Unit**	M	**conc**	1	**to-be con**	0,1	
Buffer 2	Citrate	**pH**	5	**Unit**	M	**conc**	1	**to-be con**	0,1	
Buffer 3	Citrate	**pH**	5,5	**Unit**	M	**conc**	1	**to-be con**	0,1	
Buffer 4	MES	**pH**	6	**Unit**	M	**conc**	1	**to-be con**	0,1	
Salt		0		**Unit**		0 **conc**	2	**to-be con**	0	
Precipitant 1		0		**Unit**		0 **conc**	2	**to-be con**	0	
Precipitant 2	PEG 3350			**Unit**	%	**conc**	50	**start con**	**2**	**end conc** **20**

E1-H6

Buffer 1	Citrate	**pH**	5,5	**Unit**	M	**conc**	1	**to-be con**	0,1	
Buffer 2	MES	**pH**	6	**Unit**	M	**conc**	1	**to-be con**	0,1	
Buffer 3	HEPES	**pH**	6,4	**Unit**	M	**conc**	1	**to-be con**	0,1	
Buffer 4	HEPES	**pH**	7	**Unit**	M	**conc**	1	**to-be con**	0,1	
Salt		0		**Unit**		0 **conc**	2	**to-be con**	0,2	
Precipitant 1		0		**Unit**		0 **conc**	1	**to-be con**	0,1	
Precipitant 2	AS			**Unit**	M	**conc**	4	**start con**	1,7	**end conc** 2,5

Screen	MRS_PEG3350-Cit_P								
Date									
Proteinconc.	8,0 mg/ml								
Wellvolume	0,70 ml								
A1-H6									
Buffer 1	Citrate	**pH**	4,5 **Unit**	M	**conc**	1 **to-be conc**	0,1		
Buffer 2	Citrate	**pH**	4,5 **Unit**	M	**conc**	1 **to-be conc**	0,1		
Buffer 3	Citrate	**pH**	4,5 **Unit**	M	**conc**	1 **to-be conc**	0,1		
Buffer 4	Citrate	**pH**	5 **Unit**	M	**conc**	1 **to-be conc**	0,1		
Buffer 5	Citrate	**pH**	5 **Unit**	M	**conc**	1 **to-be conc**	0,1		
Buffer 6	Citrate	**pH**	5 **Unit**	M	**conc**	1 **to-be conc**	0,1		
Salt	Na2HPO4		**Unit**	M	**conc**	1 **to-be conc**	0,01		
Precipitant 1			**Unit**		**conc**	1 **to-be conc**			
Precipitant 2	PEG 3350		**Unit**	%v/v	**conc**	50 **start conc**	**12 end conc**		**20**
A7-H12									
Buffer 1	Citrate	**pH**	4,5 **Unit**	M	**conc**	1 **to-be conc**	0,1		
Buffer 2	Citrate	**pH**	4,5 **Unit**	M	**conc**	1 **to-be conc**	0,1		
Buffer 3	Citrate	**pH**	4,5 **Unit**	M	**conc**	1 **to-be conc**	0,1		
Buffer 4	Citrate	**pH**	5 **Unit**	M	**conc**	1 **to-be conc**	0,1		
Buffer 5	Citrate	**pH**	5 **Unit**	M	**conc**	1 **to-be conc**	0,1		
Buffer 6	Citrate	**pH**	5 **Unit**	M	**conc**	1 **to-be conc**	0,1		
Salt			**Unit**		**conc**	2 **to-be conc**			
Precipitant 1			**Unit**		**conc**	100 **to-be conc**			
Precipitant 2	PEG 3350		**Unit**	%v/v	**conc**	50 **start conc**	12 **end conc**		20