Structural and mutational characterisation of *Shigella* Pathogenicity Factors

Dissertation
zur Erlangung des Doktorgrades
der Naturwissenschaften
(Dr. rer. nat.)

dem Fachbereich Pharmazie
der Philipps-Universität Marburg
vorgelegt von

Naomi Tidten
aus Bonn

Marburg/Lahn 2007

Die Untersuchungen zur vorliegenden Arbeit wurden auf Anregung von Herrn PD Dr. Klaus Reuter und Herrn Prof. Dr. Gerhard Klebe am Institut für Pharmazeutische Chemie des Fachbereichs Pharmazie der Philipps-Universität Marburg in der Zeit von Mai 2004 bis Dezember 2007 durchgeführt.

Vom Fachbereich Pharmazie der Philipps-Universität Marburg als Dissertation am 15.02.2008 angenommen.

Erstgutachter: Prof. Dr. Gerhard Klebe
Zweitgutachter: PD Dr. Klaus Reuter

Tag der mündlichen Prüfung: 15.02.2008

In Erinnerung an meinen Großvater

Es gibt mehr Leute, die kapitulieren, als solche, die scheitern.

Henry Ford (1863–1947)

Table of contents

Figures and Tables

1 Introduction

1.1 Bacillary dysentery – Shigellosis

Bacillary dysentery, an acute inflammatory bowel disease, is caused by enteroinvasive bacteria of the genus *Shigella*. After oral uptake the bacteria reach the colon, where they have the capacity to invade the colonic and rectal epithelium in humans, thereby causing the acute mucosal inflammation that characterises the disease. This destructive recto-colitis is responsible for the dysenteric symptoms [1].

Shigellosis is spread all-over the world, even though only 1.5 million of the 165.7 million annual cases occur in industrialised countries. The vast majority takes place in developing countries, often epidemically, with a death rate of 1.1 million *p.a.*. Shigellosis is mainly a disease of impoverished people which in about 69 % (112.6 million) of all cases and 61 % (660.000) of all fatalities affects children between the ages of one and five years [2].

Shigellae classification

Shigellae are GRAM-negative, nonsporulating, facultative anaerobic bacilli, belonging to the family of *Enterobacteriaceae*. Genetic analysis revealed that *Shigella* belongs to the core of *Escherichia coli* strains [3]. There are four 'species' and a varying number of serotypes of *Shigella* [1,2]:

- *Shigella dysenteriae* (16 serotypes), only rarely observed in industrialised countries (1 %). *S. dysenteriae* subtype 1 ("Shiga bacillus") accounts for brisk and deadly epidemics in the poorest populations, especially in South Asia and sub-Saharan Africa. This subtype produces a potent cytotoxin (Shiga toxin), that is capable to

induce severe inflammations, including kidney, brain, and red blood cell damages [4].

- *Shigella flexneri* (6 serotypes), the most abundant species worldwide, dominating in developing countries with 55 – 85 % (industrialised countries: 16 %).

- *Shigella sonnei* (1 serotype), the most abundant species in industrialised countries (77 %).

- *Shigella boydii* (8 serotypes), less common (6 %) and almost equally distributed all over the world.

1.2 Pathogenesis by *Shigellae*

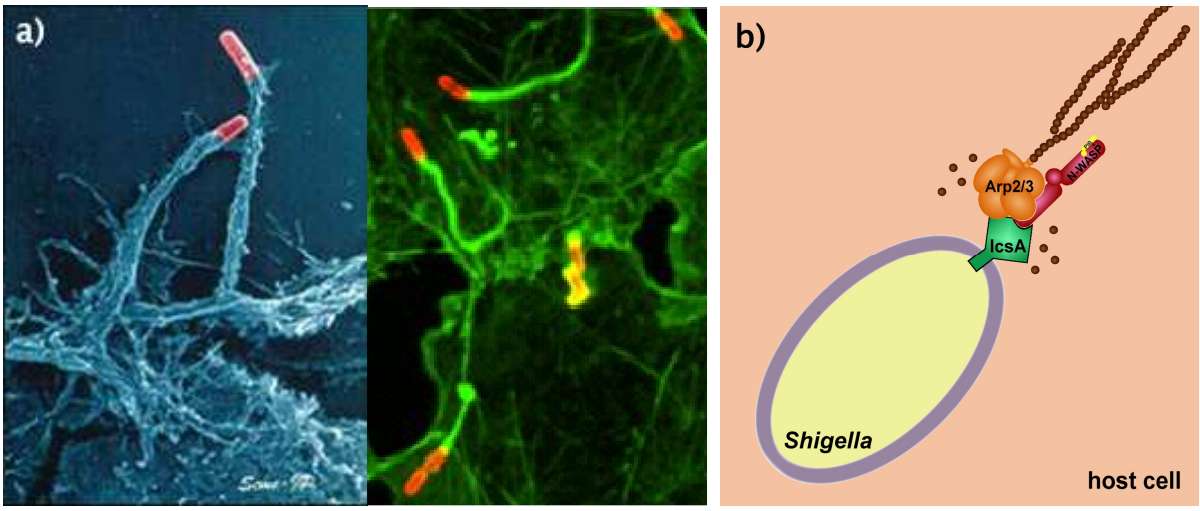
Shigella is a highly contagious microorganism. As few as 10 – 100 bacteria can cause the disease in adults. Bacterial dysentery remains essentially limited to the intestinal mucosa, where *Shigellae* are capable to invade and colonise the colon epithelium [1].

The invasion into epithelial cells does not occur through the apical side of the cells, but – after a passage of the colon epithelium's tight junctions *e.g. via* Membranous epithelial cells (M cells), (Figure 2.3, p. 4) [5] – from the basal side (Figure 2.5, p. 4) [6,7].

M cells, specialised to transport antigens (bacteria, *etc*.) through the colon epithelium, present them to associated macrophages. *Shigellae* have the unusual capacity to enter macrophages without damage. After phagocytic uptake, they escape from the phagosome and induce apoptosis of the macrophage. Thus, intact *Shigellae* are released into the mucosa (Figure 2.4, p. 4).

Invasive *Shigellae* induce massive reorganisation of the subcortical cytoskeletal network causing the formation of a macropinocytic vacuole, followed by the entry into the epithelial cells and quick lysis of the endocytic vacuole [8] (Figure 2.6). Escape into the host cytoplasm is followed by an actin-dependent motility process [9]. Once inside the epithelial cells, *Shigellae* propagate and spread to adjacent cells (Figure 1a; Figure 2.6, p. 4) – well protected from immune cells. Infected epithelial cells attract macrophages and polymorphonuclear leucocytes (PMN) from subepithelial tissues. The latter disrupt the integrity of the epithelial barrier and facilitate further *Shigella* invasion from the colon (Figure 2.2 and .7, p. 4).

Figure 1: Intra- and intercellular spread.
a) *Shigellae* with actin tails are coloured red[1]. **b**) IcsA / N-WASP / ARP2/3 mediated intracellular mobility of *Shigella*. G-actin molecules are indicated by brown spheres, F-actin by brown filaments.

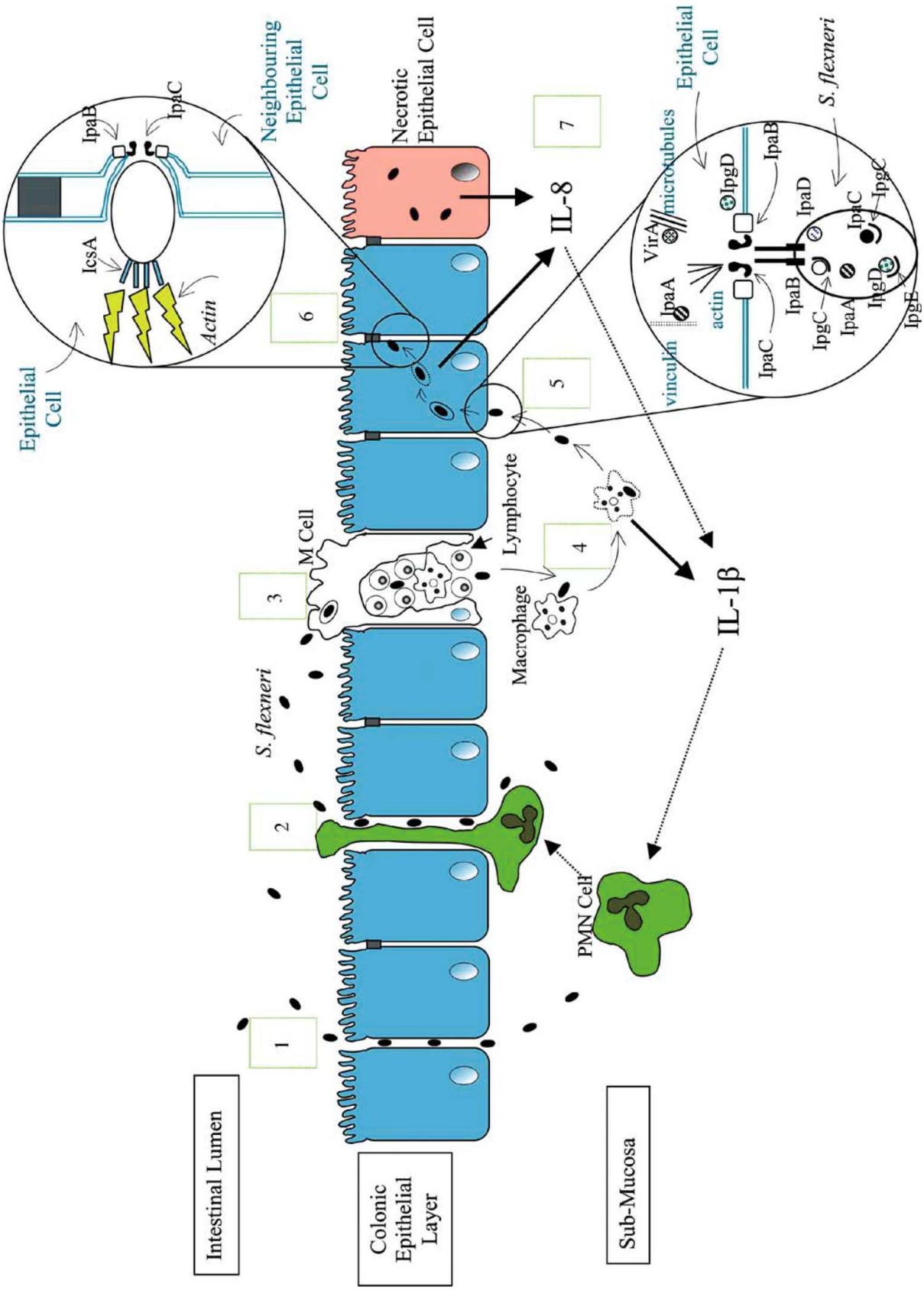


In summary, the invasive phenotype of *Shigellae* results in rupture, invasion and destruction of its various cellular targets and finally the mucosal barrier [8,10]. Nevertheless, in contrast to macrophages, PMN are able to kill *Shigellae* inside their phagosomes and eventually cope with the infection [7].

---

1) http://www.shigellablog.com/shigella_1.jpg &
http://www.med.monash.edu.au/assets/images/microbiology/Shigella.jpg

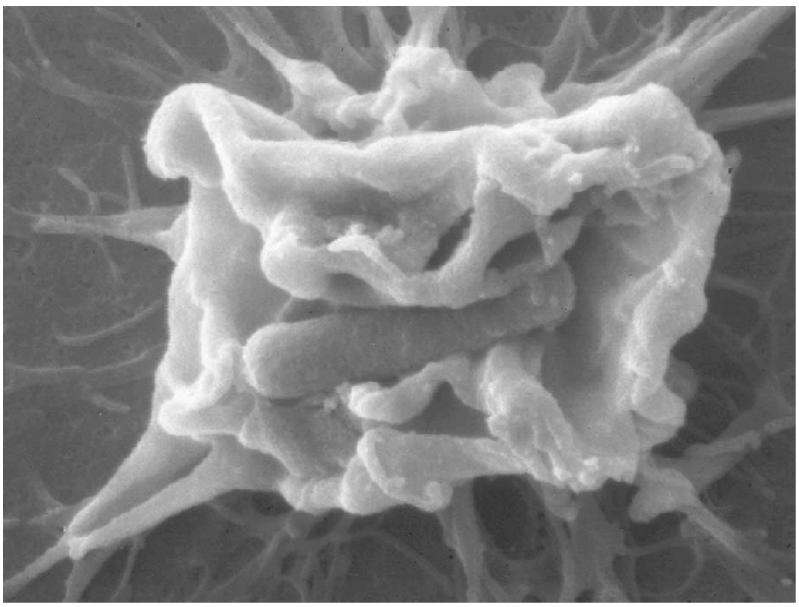
Figure 2: Pathogenesis of *S. flexneri* [7].
Lumenal bacteria invade the colonic epithelial layer by three known mechanisms.
1. *S. flexneri* can manipulate the tight junction proteins expressed by epithelial cells, allowing paracellular movement of bacteria into the sub-mucosa.
2. Polymorphonuclear leucocytes (PMN) cells recruited by IL-8 and IL-1β produced in response to *S. flexneri* invasion create gaps between epithelial cells, through which *S. flexneri* can transmigrate into the sub-mucosa.
3. Endocytic M cells transcytose bacteria, releasing them into an intraepithelial pocket filled with B and T lymphocytes and macrophages.
4. Macrophages phagocytose the bacteria. *S. flexneri* escapes the phagosome and induces the macrophage to undergo apoptosis. The apoptotic macrophage releases IL-1β.
5. Submucosal *S. flexneri* contact the basolateral membrane of epithelial cells, activating secretion of proteins through their type III secretion system. Proteins chaperoned in the cytosol of *S. flexneri* are secreted into the epithelial cell's cytoplasm through a pore formed by IpaB and IpaC. IpaC polymerises actin, IpgD dissociates the plasma membrane from the actin cytoskeleton, VirA destabilises microtubules and IpaA forms a complex with vinculin, depolymerising actin. This creates cell surface extensions which form around the bacterium, driving the epithelial cell to take up *S. flexneri* into a vacuole.
6. IpaB and IpaC lyse the vacuole, releasing *S. flexneri* into the epithelial cell's cytoplasm. The *S. flexneri* protein, IcsA is displayed on only one pole of the bacterium, creating a polymerised actin tail behind the bacterium. This propels *S. flexneri* through the cytoplasm until it contacts the plasma membrane. The force of the contact creates a protrusion into the neighbouring epithelial cell. Both membranes are lysed by IpaB and IpaC, releasing *S. flexneri* into the neighbouring epithelial cell.
7. Intracellular *S. flexneri* induces the epithelial cell to release IL-8. IL-8 and the IL-1β released from apopotic macrophages are chemotactic to PMN cells (represented by dotted arrows), attracting and inducing them to migrate through the epithelial layer to the lumen. This epithelial disruption amplifies *S. flexneri* invasion of the epithelial layer.


Intestinal Lumen
Colonic Epithelial Layer
Sub-Mucosa
1
2
3
4
5
6
7
S. flexneri
M Cell
Lymphocyte
Macrophage
PMN Cell
IL-8
IL-1β
Necrotic Epithelial Cell
Neighbouring Epithelial Cell
Epithelial Cell
IpaB
IpaC
IcsA
Actin
Epithelial Cell
microtubules
VirA
IpaA
vinculin
actin
IpaB
IpaC
IpgD
IpaD
IpgC
IpaA
IpgD
IpaC
IpgC
IpgE
S. flexneri

1.3 Molecular basis of the *Shigella* invasion mechanism

In *Shigella*, a 220 kb virulence plasmid encodes most of the genes required to express the key steps of the invasive phenotype, enabling cell entry, intracellular mobility, cell spreading and induction of apoptosis [11,12]. The *ipa/mxi-spa* locus, a 30 kb region on the virulence plasmid, is necessary and sufficient to cause entry into epithelial cells *via* macropinocytosis [8]. The virulence factors encoded by this locus facilitate reprogramming of the cellular machinery of epithelial as well as immune cells by activating innate transport and signalling pathways [13,14].

Figure 3: Entrance structure during *Shigella* invasion.
Electron microscope image[2].



Upon contact with the basolateral side of an epithelial cell, the type III secretion apparatus (TTS), a macromolecular structure containing a needle protruding from the bacterial surface [15], delivers *Shigella* effector proteins straight from the bacterial cytoplasm into the cytoplasm of the host cell target. For this purpose, a pore is formed within the cytoplasmic membrane

2) http://biop.ox.ac.uk/www/lea/website/images/shigella_invade.jpg

by the virulence factors IpaB and IpaC [11] (Figure 2.5, p. 4). In addition, IpaC activates an Src tyrosine kinase and thus a signalling cascade within the host cell, causing actin dependent filopodia and lamellipodia formation, and leading to a macropinocytotic uptake of the bacterium [16] (Figure 3). These events result from a multilateral cross-talk between *Shigella* and host cell signaling pathways which regulate the cytoskeleton, essentially by causing actin nucleation and polymerisation at the eukaryotic cell membrane [14].

Once internalised, the macropinocytotic vacuole is quickly lysed by the bacterium. Thereby it is released into the host cell cytoplasm. The bacterium assembles an F-actin comet at one of its poles enabling mobility within the host cell [17] (Figure 2.6, p. 4), and causing a very efficient process of intra-/ intercellular colonisation. This intra-/intercellular spread is a crucial step in the virulence of *Shigella* and is driven by the outer membrane protein IcsA (VirG). This protein is accumulated at the old pole of the bacterial surface and binds to the host cell proteins N-WASP and ARP2/3 leading to a polymerisation of actin, and pushing the bacterium forward inside the host cell [18,19] (Figure 1b, p. 3; Figure 2.6, p. 4).

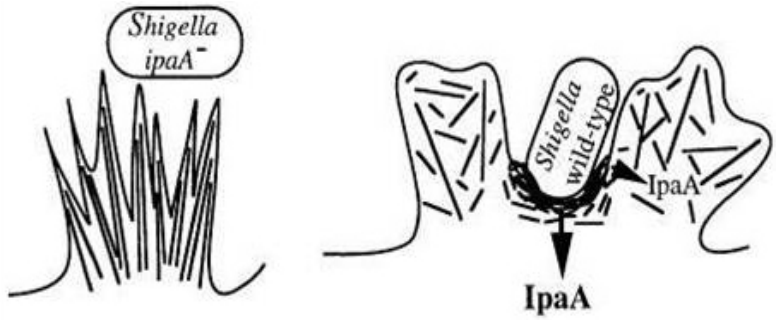
The described virulence factors are only the most prominent ones among a vast set of further virulence factors involved in the generation of pathogenicity. Below, several pathogenicity factors were selected for a more detailed description of functions.

1.3.1 Pathogenicity factor IpaA, its chaperone Spa15 and the host cell protein Vinculin

Prerequisite for invasion is the TTS mediated translocation of several *Shigella* effector proteins, particularly Ipa ("invasion plasmid encoded antigen") proteins, to the host cytoplasm [11] (Figure 2.5, p. 4). Among them, IpaA is involved in the maturation of the entry focus [9]. Deletion of its gene does not eliminate the entry capacity, but induces significant attenuation [1].

Expression of the *ipaA* gene occurs independently of the activity of the TTS. Within the bacteria, IpaA is stored in association with the chaperone Spa15 to maintain its secretion-competent state. In the absence of Spa15, IpaA is thought to form premature soluble complexes in the bacterial cytoplasm, preventing IpaA secretion [20]. The crystal structure of dimeric Spa15 has been solved at a resolution of 1.8 Å [21].

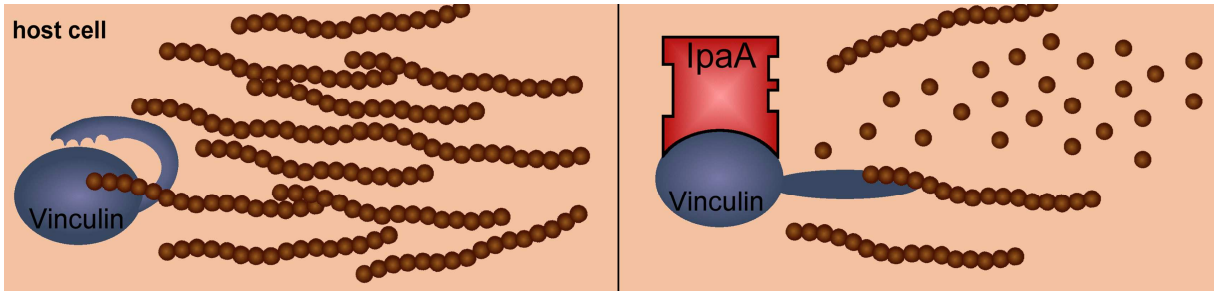
Figure 4: Effect of IpaA on the cytoskeletal rearrangements of the enty focus [9].
In *ipaA*⁻ strains of *Shigella*, uncontrolled polymerisation of actin within the host cell leads to the formation of protrusions, pushing the bacteria away from the cellular surface and thus, hampering inavsion. In contrast, in wild-type *Shigella*, cell extensions are not formed in the contact area between bacterium and host cell, promoting an efficient uptake of the bacterium by the host cell.



Cytoskeletal structures are the targets for *Shigella* entry associated proteins like IpaA. To promote a fully functional entry focus, besides the induction of the massive cytoskeletal extensions described above, a control by IpaA is required. If uncontrolled, the formation of these gigantic cell extensions would tend to push the bacterial body away (Figure 4). Actin polymerisation in the contact region of the bacterium and the host cell is avoided by a binding of IpaA to the *N*-terminal head domain of Vinculin [22]. This cytoskeleton-associated protein orchestrates the formation of cell adherence plaques within the host cell, leading to a depolymerisation of available F-actin from adjacent cytoskeletal elements. Thus, a pseudo-adherence plaque is formed [9]. In combination with the cytoskeletal extensions surrounding the bacterial body, a structure is formed that is productive to penetrate. Mutants that do not produce IpaA enter epithelial cells with much lower efficiency than their wild-type counterpart, whereas they induce entry

foci that are larger and much longer lasting than those induced by the wild-type strain (Figure 4) [22]. Entry foci induced by *ipaA* mutants show massive and persistent actin polymerisation, but fail to form the pseudo-adherence plaque that is required for the generation of an endocytic vacuole. Binding of IpaA to the *N*-terminal head domain of Vinculin activates the dormant form of Vinculin and induces its opening, thus revealing a functional actin binding site that participates in the bundling process in its *C*-terminal portion. In addition, this interaction also results in actin depolymerisation (Figure 5).

Figure 5: IpaA Vinculin interaction.
Binding of IpaA to the head domain of the host cell protein Vinculin induces opening of Vinculin and F-actin depolymerisation. G-actin molecules are indicated by brown spheres, F-actin by brown filaments.



1.3.2 Pathogenicity factor IpgD and its chaperone IpgE

Another pathogenicity factor involved in the maturation of the entry focus is IpgD [23]. Deletion of its gene also does not eliminate the entry capacity of the mutants, but significant reduction of invasion events has been observed [1].

Like IpaA, IpgD is produced independently of TTS activity. Within the bacterial cytoplasm, IpgD is stored in association with its specific chaperone IpgE [23]. IpgE is necessary and sufficient for the stability of IpgD within the cytoplasm. When used to infect tissue culture cells, *ipgD* as well as *ipgE* mutants elicit entry structures that have a different morphology compared with those produced by the wild-type strain [23]. In particular, actin rearrangements induced upon contact of bacteria with epithelial cells are

altered. The mutants provoke fewer actin rearrangements and less membrane ruffling on the cell surface [23].

Phosphoinositides play a central role in the control of several cellular events including actin cytoskeleton organisation. IpgD is translocated into the host cell cytoplasm and acts as a potent inositol-4-phosphatase that specifically dephosphorylates phosphatidylinositol-4,5-bisphosphate [PtdIns-(4,5)P_2] leading to the accumulation of phosphatidylinositol-5-monophosphate [PtdIns(5)P]. This transformation of PtdIns(4,5)P_2 into PtdIns(5)P by IpgD is responsible for dramatic morphological changes of the host cell, leading to a decrease in membrane tether force associated with membrane blebbing and actin filament remodelling, promoting membrane ruffling at the entry site (Figure 2.5, p. 4). Although IpgD is not absolutely required for entry of bacteria into cultured cells, it is implicated in entry focus formation [23-25]. After secretion, IpgD is thought to form a complex with IpaA within the host cell [23].

1.3.3 Pathogenicity factor OspD1

Transcription of several effector genes is regulated by TTS activity which is mediated by the transcriptional activator MxiE as well as by its co-activator IpgC. OspD1 was identified to negatively control expression of these genes regulated by secretion activity. OspD1 associates with the chaperone Spa15 and the activator MxiE and acts as an anti-activator until it is secreted. The TTS activity is induced upon contact of bacteria with host cells. After secretion of OspD1, it cannot longer act as an antiactivator within the bacterial cell which ultimately leads to the transcription of MxiE dependent effector genes [26,27].

1.3.4 Pathogenicity factor IpgB2

Because of their ubiquitous role in cell biology, Ras-like small G proteins are a common target for virulence factors secreted by bacterial pathogens [30]. There are well over 100 eukaryotic small G proteins that regulate a myriad of cellular processes including growth, differentiation, membrane trafficking, cytoskeletal organisation, and nuclear import. Several key features of small G proteins allow them to perform these essential functions. First, they act as molecular switches cycling between GTP bound (active) and GDP bound (inactive) conformations. Second, most small G proteins are post-translationally modified. The addition of isoprenoid lipid moieties allows them to associate with specific subcellular compartments. Third, they transmit signaling events in a GTP-dependent manner by activating and/or recruiting downstream target proteins to their sites of action.

IpgB2 subverts host cell signal transduction cascades by mimicking the signaling properties of Ras-like GTPases. It stimulates cellular responses analogous to GTP-active RhoA, a Ras-related GTP binding protein that induces actin stress fiber assembly. IpgB2 neither binds guanine nucleotides nor does it show any sequence similarity corresponding to the conserved GTPase domain, suggesting that it is a functional but not a structural mimic. However, the activity of IpgB2 is dependent on an invariant WxxxE motif found in numerous effectors as targeting sequence. Although the bacterial protein lacks many of the biochemical attributes of GTPases, it induces actin cytoskeletal dynamics in the host cell by stimulating common signaling pathways [31].

1.3.5 Pathogenicity factor VirF

The expression of virulence genes is organised hierarchically [32]. Expression of a plasmid encoded key transcription activator for virulence gene expression, namely *virF*, directly activates the transcription of virulence factor

genes like *icsA* and of *virB*. The *virB* gene product then activates the transcription of the *mxi/spa*- and *ipa* genes. Thus, VirF is in the centre of pathogenicity regulation. The amount of VirF depends on various environmental factors, such as pH, osmolarity, temperature and nutrition factors. Below 30° C *Shigella* is not virulent. The *virF* gene transcription is fully activated at 37° C, pH 7.4, physiological osmolarity and in the presence of free amino acids (arginine / methionine). Transcription factors encoded by the bacterial chromosome with expression sensitive to these environmental factors control *virF* expression [32-34]. Apart from this 'classical' regulation, the VirF level obviously also depends on an accurate translational machinery [34]. Efficient translation of *virF* mRNA at the ribosome requires the presence of modified tRNA molecules. Modification of tRNA bases is very common in nature [35]. Two modifications were demonstrated to exhibit major influence on the *virF* mRNA translational speed. In specific tRNA molecules the hyper-modified bases queuine in position 34, and 2-methylthio-*N* 6-isopentenyladenine in position 37 have to be present. *Shigella* mutants lacking one of these modifications show significantly reduced virulence, approved by mutational studies where gene knock-out in the tRNA modification pathway resulted in the absence of these modifications [36].

1.4 tRNA guanine transglycosylase, the gene product of *vacC* is a target for structure-based drug design

The *Shigella vacC* gene product, a tRNA guanine transglycosylase (TGT), catalyses the incorporation of a queuine precursor into tRNA at position 34, the anticodon 'wobble position'. Mutation of the *vacC* gene reduces both, the VirF level as well as the haemolytic activity to 50 – 60 % of the wild-type [34,37]. Thus, tRNA modifying enzymes represent promising targets for the development of antibiotics. Inhibition of such specific tRNA modification steps should result in significantly reduced virulence of *Shigella*. A crystal

structure of the structurally very similar *Zymomonas mobilis* TGT is available [38]. This TGT is employed for structure-based drug design [39-44].

tRNA maturation involves a series of post-transcriptional processing steps resulting in fully functional tRNA molecules. Among these maturation steps nucleoside modification is the most remarkable one. Modifications within or around the anticodon are proposed to fine-tune translational speed and fidelity or to influence the occurence of frame-shifting events. Nevertheless, the present understanding is still rudimentary [45,46]. *E.g.* only very recently, a putative second function of the *vacC* gene product has been identified. *E. coli* TGT was found to be able to recognise the *virF* mRNA and perform the site-specific modification of a single base. Thus, TGT has been discussed to modulate the VirF translation by a second mechanism, involving a putative riboswitch consisting of the VirF encoding mRNA. A UGU sequence that is specifically recognised by TGTs was identified, exhibiting adjacent sequences (nucleotides 410 – 433 of the *virF* mRNA) that enable the formation of a hairpin structure, similar to the anticodon stemloop of tRNAs [47].

1.5 Aims of the project

1.5.1 Pathogenicity factors IpaA, IpgE, IpgB2 and OspD1

The roles of many *S. flexneri* virulence genes have been studied in a variety of cell culture experiments such as invasion assays and plaque assays or through the use of *in vivo* animal models [7]. For the present study, the following pathogenicity factors were selected for further analyses, especially concerning the determination of their three dimensional structures: IpaA, IpgE, IpgB2, and OspD1. Therefore, several preceding procedures have to be established, such as cloning and overexpression of the selected genes, protein purification protocols, as well as high throughput screening of crystallisation conditions. Hereafter, the determination of crystal structures of single, or complexed pathogenicity factors should be possible.

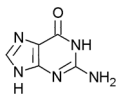
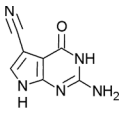
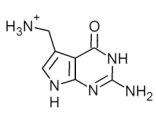
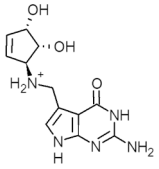
Research on *S. flexneri* virulence factors, including structure analyses, will yield further understanding of the molecular basis of *S. flexneri* mediated invasion and destruction of the intestinal mucosa, as well as the role of the host's subsequent innate responses. A comprehensive understanding of *S. flexneri*'s disease-causing mechanisms at the molecular level is essential to develop innovative antipathogenicity molecules against this disease.

1.5.2 Substrate selectivity and specificity study of tRNA guanine transglycosylases of the three domains of life, based on a *Zymomonas mobilis* TGT model system

Z. mobilis TGT is an established target for structure-based drug design [39-44]. Nevertheless, TGTs are evolutionary ancient enzymes, present in all three domains of life. The catalytic mechanism is thought to be highly conserved, even if the enzymes exhibit different *in vivo* substrates in a domain-specific manner. *In vitro*, however, 'foreign' substrates are as well

accepted by some of the TGTs. While archeal TGTs are highly selective and accept only guanine and the actual substrate $preQ_0$, bacterial TGTs are able to incorporate besides the natural substrate $preQ_1$ also guanine and $preQ_0$. Eukaryotic TGTs are considered as the most unselective enzymes in this context (Table 1). Thus, to be able to specifically inhibit bacterial TGT, it is of high importance to study the selectivity and specificity determining features of these enzymes. *Z. mobilis* TGT was selected to serve as a model system to analyse these features in a mutational study.

Table 1: *In vivo* and *in vitro* substrates of TGTs of the three domains [42].
X: *In vivo* substrate; (X): Accepted as a substrate *in vitro*, –: Not accepted as a substrate.

TGT	Exchanged base	 guanine	 $preQ_0$	 $preQ_1$	 queuine
Archaeal	G^{15}	(X)	X	–	–
Bacterial	G^{34}	(X)	(X)	X	–
Eukaryotic	G^{34}	(X)	(X)	(X)	X

1.6 References

[1] Sansonetti PJ. Rupture, invasion and inflammatory destruction of the intestinal barrier by Shigella, making sense of prokaryote-eukaryote crosstalks. FEMS Microbiology Reviews 2001;25:3-14.

[2] Kotloff K, Winickoff J, Ivanoff B, Clemens J, Swerdlow D, Sansonetti P, Adak G, Levine M. Global burden of Shigella infections: implications for vaccine development and implementation of control strategies. WHO Bull 1999;77:651-66.

[3] Escobar-Páramo P, Clermont O, Blanc-Potard AB, Bui H, Le Bouguénec C, Denamur E. A Specific Genetic Background Is Required for Acquisition and Expression of Virulence Factors in Escherichia coli. Molecular Biology and Evolution 2004;21:1085-94.

[4] O'Loughlin EV, Robins-Browne RM. Effect of Shiga toxin and Shiga-like toxins on eukaryotic cells. Microbes and Infection 2001;3:493-507.

[5] Sansonetti PJ, Phalipon A. M cells as ports of entry for enteroinvasive pathogens: mechanisms of interaction, consequences for the disease process. Semin. Immunol. 1999;11:193-203.

[6] Sansonetti PJ. Genetic and molecular basis of epithelial cell invasion by Shigella species. Rev. Infect. Dis. 1991;13:S285-S92.

[7] Jennison AV, Naresh KV. Shigella flexneri infection: pathogenesis and vaccine development. FEMS Microbiol. Rev. 2004;28:43-58.

[8] Sansonetti PJ. Microbes and Microbial Toxins: Paradigms for Microbial-Mucosal Interactions III. Shigellosis: from symptoms to molecular pathogenesis. Am. J. Physiol. Gastrointest. Liver Physiol. 2001;280:G319-G23.

[9] Bourdet-Sicard R, Rudiger M, Sansonetti PJ, Tran Van Nhieu G. Vinculin is unfolded by the Shigella protein IpaA, and the complex promotes F-actin depolymerization. EMBO J 1999;18:5853-62.

[10] Zychlinsky A, Sansonetti PJ. Apoptosis as a pro-inflammatory event: what can we learn from bacteria-induced cell death? Trends Microbiol. 1997;5 (201-204).

[11] Buchrieser C, Glaser P, Rusniok C, Nedjari H, D'Hauteville H, Kunst F, Sansonetti P, Parsot C. The virulence plasmid pWR100 and the repertoire of proteins secreted by the type III secretion apparatus of Shigella flexneri. Mol Microbiol 2000;38 (4):760-71.

[12] Sansonetti PJ, Hale TL, Dammin GJ, Kapfer C, Colins H, Formal SB. Alterations in the pathogenicity of Escherichia coli K12 after transfer of plasmid and chromosomal genes from Shigella flexneri. Infect. Immun. 1983;39:1392-402.

[13] Fernandez MI, Sansonetti PJ. Shigella interaction with intestinal epithelial cells determines the innate immune response in shigellosis. Int. J. Med. Microbiol. 2003;293:55-67.

[14] Van Nhieu GT, Bourdet-Sicard R, Duménil G, Blocker A, Sansonetti PJ. Bacterial signals and cell responses during Shigella entry into epithelial cells. Cell. Microbiol. 2000;2:187-93.

[15] Blocker A, Gounon P, Larquet E, Niebuhr K, Cabiaux V, Parsot C, Sansonetti P. The tripartite type III secreton of Shigella flexneri inserts IpaB and IpaC into host membranes. J Cell Biol 1999;147 (3):683-93.

[16] Adam T, Arpin M, Prevost MC, Gounon P, Sansonetti PJ. Cytoskeletal rearrangements and the functional role of T-plastin during entry of Shigella flexneri into HeLa cells. J Cell Biol 1995;129 (2):367-81.

[17] Bernardini ML, Mounier J, d'Hauteville H, Coquis-Rondon M, Sansonetti PJ. Identification of icsA, a plasmid locus of Shigella flexneri which governs bacterial intra- and intercellular spread through interaction with F-actin. Proc Natl Acad Sci USA 1989;86:3867-71.

[18] Lett MC, Sasakawa C, Okada N, Sakai T, Makino S, Yamada M, Komatsu K, Yoshikawa M. VirG, a plasmidcoded virulence gene of Shigella flexneri: identification of the VirG protein and determination of the complete coding sequence. J Bacteriol. 1988;171:353-9.

[19] Makino S, Sasakawa C, Kamata K, Kurata T, Yoshikawa M. A genetic determinant required for continuous reinfection of adjacent cells on large plasmid in Shigella flexneri 2a. Cell 1986;46:551-5.

[20] Page AL, Sansonetti PJ, Parsot C. Spa15 of Shigella flexneri, a third type of chaperone in the type III secretion pathway. Mol Microbiol 2002;43 (6):1533-42.

[21] van Eerde A, Hamiaux C, Pérez J, Parsot C, Dijkstra BW. Structure of Spa15, a type III secretion chaperone from Shigella flexneri with broad specificity. Embo Reports 2004;5:477-83.

[22] Tran Van Nhieu G, Ben Zeev A, Sansonetti PJ. Modulation of bacterial entry in epithelial cells by association between vinculin and the Shigella IpaA invasin. Embo J 1997;16:2717-29.

[23] Niebuhr K, Jouihri N, Allaoui A, Gounon P, Sansonetti PJ, Parsot C. IpgD, a protein secreted by the type III secretion machinery of Shigella flexneri, is

chaperoned by IpgE and implicated in entry focus formation. Mol. Microbiol. 2000;38:8-19.

[24] Niebuhr K, Giuriato S, Pedron T, Philpott DJ, Gaits F, Sable J, Sheetz MP, Parsot C, Sansonetti PJ, Payrastre B. Conversion of PtdIns(4,5)P2 into PtdIns(5)P by the *S. flexneri* effector IpgD reorganizes host cell morphology. Embo J 2002;21:5069-78.

[25] Allaoui A, Menard R, Sansonetti PJ, Parsot C. Characterization of the Shigella flexneri ipgD and ipgF genes, which are located in the proximal part of the mxi locus. Infect Immun 1993;61 (5):1707-14.

[26] Menard R, Sansonetti P, Parsot C. The secretion of the Shigella flexneri Ipa invasins is activated by epithelial cells and controlled by IpaB and IpaD. Embo J 1994;13 (22):5293-302.

[27] Pettersson J, Nordfelth R, Dubinina E, Bergman T, Gustafsson M, Magnusson KE, Wolf-Watz H. Modulation of virulence factor expression by pathogen target cell contact. Science 1996;273:1231-3.

[28] Miller VL. Connections between transcriptional regulation and type III secretion? Curr Opin Microbiol 2002;5:211-5.

[29] Parsot C, Ageron E, Penno C, Mavris M, Jamoussi K, d'Hauteville H, Sansonetti PJ, Demers B. A secreted anti-activator, OspD1, and its chaperone, Spa15, are involved in the control of transcription by the type III secretion apparatus activity in Shigella flexneri. Mol. Microbiol. 2005;56:1627-35.

[30] Boquet P. Small GTP binding proteins and bacterial virulence. Microbes and Infection 2000;2:837-43.

[31] Alto NM, Shao F, Lazar CS, Brost RL, Chua G, Mattoo S, McMahon SA, Ghosh P, Hughes TR, Boone C, Dixon JE. Identification of a bacterial type III effector family with G protein mimicry function. Cell 2006;124:133-45.

[32] Dorman CJ, Porter ME. The Shigella virulence gene regulatory cascade: a paradigm of bacterial gene control mechanisms. Mol. Microbiol. 1998;29:677-84.

[33] Durand JM, Björk GR. Putrescine or a combination of methionine and arginine restores virulence gene expression in a tRNA modificationdeficient mutant of Shigella flexneri: a possible role in adaptation of virulence. Mol Microbiol 2003;47:519-27.

[34] Durand JM, Dagberg B, Uhlin BE, Björk GR. Transfer RNA modification, temperature and DNA superhelicity have a common target in the regulatory network of the virulence of Shigella flexneri: the expression of the virF gene. Mol Microbiol 2000;35:924-35.

[35] Björk G. Stable RNA modification. Cellular and Molecular Biology 1996;2:861-86.

[36] Durand JM, Björk GR, Kuwae A, Yoshikawa M, Sasakawa C. The modified nucleoside 2-methylthio-N6-isopentenyladenosine in tRNA of Shigella flexneri is required for expression of virulence genes. J Bacteriol 1997;179:5777-82.

[37] Durand JM, Okada N, Tobe T, Watarai M, Fukuda I, Suzuki T, Nakata N, Komatsu K, Yoshikawa M, Sasakawa C. vacC, a virulence-associated chromosomal locus of Shigella flexneri, is homologous to tgt, a gene encoding tRNA-guanine transglycosylase (Tgt) of Escherichia coli K-12. J Bacteriol 1994;176 (15):4627-34.

[38] Romier C, Reuter K, Suck D, Ficner R. Crystal structure of tRNA-guanine transglycosylase: RNA modification by base exchange. Embo J 1996;15 (11):2850-7.

[39] Gradler U, Gerber HD, Goodenough-Lashua DM, Garcia GA, Ficner R, Reuter K, Stubbs MT, Klebe G. A new target for shigellosis: rational design and crystallographic studies of inhibitors of tRNA-guanine transglycosylase. J Mol Biol 2001;306 (3):455-67.

[40] Meyer EA, Brenk R, Castellano RK, Furler M, Klebe G, Diederich F. De novo design, synthesis, and in vitro evaluation of inhibitors for prokaryotic tRNA-guanine transglycosylase: a dramatic sulfur effect on binding affinity. Chembiochem 2002;3 (2-3):250-3.

[41] Brenk R, Naerum L, Gradler U, Gerber HD, Garcia GA, Reuter K, Stubbs MT, Klebe G. Virtual screening for submicromolar leads of tRNA-guanine transglycosylase based on a new unexpected binding mode detected by crystal structure analysis. J Med Chem 2003;46 (7):1133-43.

[42] Brenk R, Stubbs MT, Heine A, Reuter K, Klebe G. Flexible adaptations in the structure of the tRNA-modifying enzyme tRNA-guanine transglycosylase and their implications for substrate selectivity, reaction mechanism and structure-based drug design. Chembiochem 2003;4 (10):1066-77.

[43] Brenk R, Meyer EA, Reuter K, Stubbs MT, Garcia GA, Diederich F, Klebe G. Crystallographic study of inhibitors of tRNA-guanine transglycosylase suggests a new structure-based pharmacophore for virtual screening. J Mol Biol 2004;338 (1):55-75.

[44] Stengl B, Meyer EA, Heine A, Brenk R, Diederich F, Klebe G. Crystal Structures of tRNA-guanine Transglycosylase (TGT) in Complex with Novel and Potent Inhibitors Unravel Pronounced Induced-fit Adaptations and Suggest Dimer Formation Upon Substrate Binding. J Mol Biol 2007;370 (3):492-511.

[45] Bjork GR, Durand JM, Hagervall TG, Leipuviene R, Lundgren HK, Nilsson K, Chen P, Qian Q, Urbonavicius J. Transfer RNA modification: influence on translational frameshifting and metabolism. FEBS Lett 1999;452 (1-2):47-51.

[46] Iwata-Reuyl D. Biosynthesis of the 7-deazaguanosine hypermodified nucleosides of transfer RNA. Bioorg Chem 2003;31 (1):24-43.

[47] Hurt JK, Olgen S, Garcia GA. Site-specific modification of Shigella flexneri virF mRNA by tRNA-guanine transglycosylase in vitro. Nucleic Acids Res 2007;35 (14):4905-13.

2 Selected *Shigella* Pathogenicity Factors

In cooperation with Johannes Schulze Wischeler [1] and Christian Hasewinkel [2] the pathogenicity factor genes *ipaA, spa15, ipgB2, ospD1,* and *ipgE* were cloned into different plasmid vectors. For some of the pathogenicity factors it is reported that the first 20 *N-terminal* amino acids serve as a signal peptide sequence for TTS secretion or could have a disadvantageous influence on protein solubility due to high hydrophobicity [3-5]. Therefore, for some of the selected genes, the first 60 bp were deleted, leading to shortened versions. In order to find suitable expression systems, several different variants were created. Plasmid constructs leading to the production of tagless, GST-, His6- or Strep-tagged proteins as well as full length and shortened versions were cloned. In addition, it is known that the presence of specific chaperones takes influence on the solubility and stability of these proteins [6-8]. Thus, also co-expression vectors were used. The chaperone Spa15 has been identified to stabilise IpaA [7], and OspD1 [8], while IpgE is known to specifically stabilise IpgD [6]. Additionally, effects of Spa15 on IpgB2 solubility were also analysed. In case of IpaA it was assumed, that co-expression with other interaction partners could avoid aggregation of this pathogenicity factor. Vinculin, the host interaction partner of IpaA, mainly binds with its head domain to this bacterial effector [9,10]. Hence, also the part of the human *vinculin* gene corresponding to the head domain (vinc_hd) was cloned into the co-expression vector with *ipaA.*

2.1 Cloning and Expression of ipaA, spa15, vinc_hd, ipgB2, ospD1, ipgE

The following pathogenicity factor genes were amplified *via* PCR from the *Shigella* virulence plasmid pCP301 [11] using the primers listed in Table 4 (p. 33): *ipaA, spa15, ipgB2, ospD1, ipgE* [1,2]. The plasmid vector pCMV-SPORT6 containing the human *vinculin* gene was purchased from the Deutsche Ressourcenzentrum für Genomforschung GmbH. The first 2,550 bps of the *vinculin* gene encoding the head domain were as well amplified *via* PCR using the primers listed in Table 4 (p. 33). Subsequently, all PCR products were digested with the corresponding restriction enzymes listed in Table 2 (p. 23), purified *via* minispin colums (peQlab) and cloned into the plasmid vectors pPR-IBA2, pET-Duet-1, pGEx-6p-mod [12]. The resulting constructs are listed in Table 2 (p. 23).

In case of *ipgE* first expression experiments had already shown that the expression level is very low, possibly due to rare codon usage. Therefore, three rare codons were changed to codons frequently used in *E. coli via* site-directed mutagenesis (primer no. 19-22, Table 4, p. 33) to improve the expression level. Codons of Arg4 (CGA), Leu10 (CTA), Ile11 (ATA) were exchanged by the frequently found *E. coli* codons CGC (Arg4), CTG (Leu10), ATT (Ile11), respectively.

The complete sequences of all final constructs were afterwards checked by MWG Biotech AG, Ebersberg. Subsequently, the conctructs were transformed into the following CaCl2-competent expression hosts based on BL21 (DE3) and expression levels were checked at 37 °C, 17 °C, and 20 °C, respectively: BL21 (DE3) pLysS, RosettaTM 2, CodonPlus®-RIL, GOLD.

Table 2: Constructs containing the diffenent genes or combinations of genes, expression hosts and levels [1,2].

#	name (1)	vector	gene(s)	restriction sites	Primer no.(2)	expressed protein	expression strains BL21 (DE3)	expression(3) level / remarks	
1	N8	pPR-IBA2	*spa15, ipaA*	5'-Sac II 3'-Esp 3I; 5'-Esp 3I 3'-Bam HI	1-4	Strep-Spa15, IpaA	pLysS	*spa15(3b):* high, soluble fraction	*ipaA(3b):* not detected
							RosettaTM 2		
							CodonPlus® RIL		low, insoluble fraction
							Gold		
2	pCH3	pET-Duet-1	*spa15, ipaA-20*	5'-Bam HI 3'-Hin dIII; 5'-Nde I 3'-Xho I	5-8	His6-Spa15, IpaA-20	pLysS	*spa15:* high, soluble fraction	*ipaA:* not detected
							RosettaTM 2		
							CodonPlus® RIL		low, insoluble fraction
							Gold		
3	pCH2	pET-Duet-1	*vinc_hd, ipaA-20*	5'-Asc I 3'-Not I; 5'-Nde I 3'-Xho I	9, 10	His6-Vinc_hd, IpaA-20	pLysS	*vinc_hd:* not detected	*ipaA:* moderate, insoluble fraction
							Gold		
							RosettaTM 2		
							CodonPlus® RIL		
4	331	pET-Duet-1	*ipgB2(4)*	5'-Nde I 3'-Xho I	11, 12	IpgB2	pLysS	not detected	
							Gold		
							RosettaTM 2	moderate, insoluble fraction	
							CodonPlus® RIL	moderate, partially soluble fraction	
5	507	pET-Duet-1	*spa15, ipgB2*	5'-Bam HI 3'-Hin dIII; 5'-Nde I 3'-Xho I	--	His6-Spa15, IpgB	pLysS	*spa15:* high, soluble fraction	*ipgB2:* low, insoluble fraction
							RosettaTM 2		
							Gold		not detected
							CodonPlus® RIL		

#	name (1)	vector	gene(s)	restriction sites	Primer no. (2)	expressed protein	expression strains BL21 (DE3)	expression(3) level / remarks	
6	502	pET-Duet-1	*ipgB2-20(4)*	5'-Nde I 3'-Xho I	13, 14	IpgB2-20	pLysS	not detected	
							Gold	moderate, insoluble fraction	
							RosettaTM 2		
							CodonPlus® RIL		
7	pCH4	pGEX-6P-mod	*ipgB2*	5'-Nde I 3'-Xho I	--	GST-IpgB2	pLysS	moderate, insoluble fraction(3a)	
							Gold		
							RosettaTM 2		
							CodonPlus® RIL	high, insoluble fraction(3a)	
8	529	pGEX-6P-mod	*ipgB2-20(4)*	5'-Nde I 3'-Xho I	--	GST-IpgB2-20	pLysS	moderate, insoluble fraction(3a)	
							Gold		
							RosettaTM 2		
							CodonPlus® RIL	high, insoluble fraction(3a)	
9	341	pET-Duet-1	*ospD1*	5'-Nde I 3'-Xho I	15, 16	OspD1	pLysS	not detected	
							Gold		
							RosettaTM 2	moderate, insoluble fraction	
							CodonPlus® RIL	moderate, partially soluble fraction	
10	pCH6	pET-Duet-1	*spa15, ospD1*	5'-Bam HI 3'-Hin dIII; 5'-Nde I 3'-Xho I	--	His6-Spa15, OspD1	pLysS	*spa15:* high, soluble fraction	*ospD1:* high, insoluble fraction
							Gold		
							RosettaTM 2		
							CodonPlus® RIL		

#	name (1)	vector	gene(s)	restriction sites	Primer no. (2)	expressed protein	expression strains BL21 (DE3)	expression(3) level / remarks
11	527	pGEX-6P-mod	*ospD1*	5'-Nde I 3'-Xho I	--	GST-OspD1	pLysS	moderate, insoluble fraction(3a)
							Gold	
							RosettaTM 2	
							CodonPlus® RIL	high, partially soluble fraction(3a)
12	pCH7	pET-Duet-1	*ipgE-mod(5)*	5'-Nde I 3'-Xho I	17, 18 (19-22)	IpgE	pLysS	moderate, insoluble fraction
							Gold	
							RosettaTM 2	moderate, partially soluble fraction
							CodonPlus® RIL	
13	pCH8	pGEX-6P-mod	*ipgE-mod(5)*	5'-Nde I 3'-Xho I	-- (19-22)	GST-IpgE	pLysS	moderate, soluble fraction(3c)
							Gold	
							RosettaTM 2	high, soluble fraction(3c)
							CodonPlus® RIL	

1) According to diploma theses of Johannes Schulze Wischeler and Christian Hasewinkel [1,2].

2) According to Table 4, p. 11.

3) All expression experiments were performed under the following conditions: 1 mM IPTG, 3 h, 37 °C. Additionally tested conditions are annotated explicitly as follows: **a**) 1 mM IPTG, 6 h, 20 °C; **b**) 1 mM IPTG, 10 h, 17 °C; **c**) 1 mM IPTG, 4 h, 30 °C.

4) In multi cloning site II (MCS II)

5) Codons of Arg^{4} (CGA), Leu^{10} (CTA), Ile^{11} (ATA) were mutated to the frequently found *E. coli* codons CGC (Arg^{4}), CTG (Leu^{10}), ATT (Ile^{11}), respectively (primer numbers: 19-22).

Results of all expression tests are listed in Table 2 (p. 23). In most cases successful overexpression could be obtained but frequently, the resulting protein was present in form of inclusion bodies. Co-expression of *ipaA* and the sequence encoding the Vinculin head domain failed, while co-expression of the chaperone Spa15 had no significant influence on the solubility of IpgB2, OspD1, and IpaA. Similarly, deletion of the first 20 *N-terminal* amino acids seems not to take dramatic influence on protein solubilities. GST-tagging resulted only in case of IpgE in a partially better solubility of the overproduced gene product. In most cases, expression in BL21 (DE3) CodonPlus®-RIL performed best. This expression host provides additional tRNA molecules decoding 'rare codons' thus featuring a higher expression rate for foreign genes with 'unusual codons'.

2.2 Purification of Pathogenicity Factors IpgB2, OspD1 and IpgE

Purification of proteins from inclusion bodies is usually performed under denaturing conditions (6 M guaninidinium chloride/ 8 M urea) yielding unfolded protein and therefore requires a subsequent refolding step. Thus, in the first attempts only the soluble fractions of expression experiments were further examined. Only in case of IpgB2, OspD1 and IpgE at least partially soluble overexpression products were detected. Hence, these three pathogenicity factors were purified under native conditions.

2.2.1 IpgB2

After cell lysis GST-tagged IpgB2 and IpgB2-20 (from production in BL21 (DE3) CodonPlus®-RIL containing plasmid no. 7 (pCH4, Table 2, p. 23), at 1 mM IPTG, 3 h, 37 °C) were purified from the cell lysate by addition of glutathione sepharose, equilibrated in PBS (Phosphate Buffer Saline) and 3 h incubation at 20 °C. After several PBS washing steps, PreScissionTM protease

(GE Healthcare) were added to release free IpgB2, and IpgB2-20, respectively, during over night incubation. However, the small amount of protein gained from first expression experiments was not sufficient to successfully observe cleaved, GST-free IpgB2(-20). In addition untagged IpgB2(-20) may be less soluble, compared to the GST-tagged version, putatively leading to its precipitation [2].

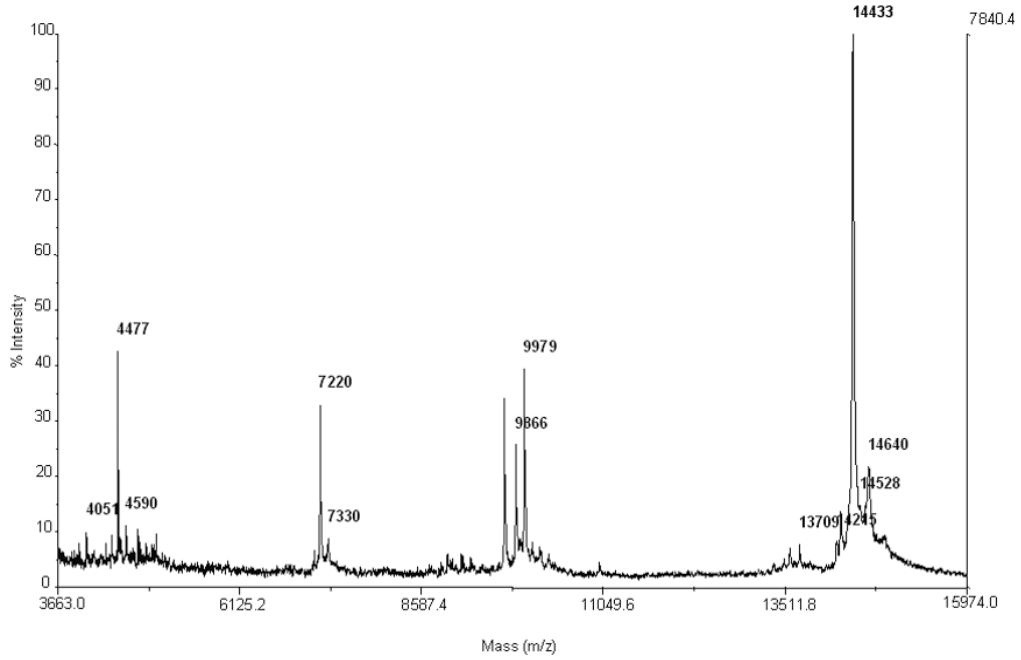
2.2.2 OspD1

GST-tagged OspD1 (from production in BL21 (DE3) CodonPlus®-RIL containing plasmid no. 11 (527, Table 2, p. 23), at 1 mM IPTG, 3 h, 37 °C) was purified *via* glutathione sepharose, as described for IpgB2. After several PBS washing steps, elution of GST-OspD1 was achieved upon addition of 20 mM glutathione. After a subsequent gelfiltration, the sample was concentrated to 6 mg/ml utilising Vivaspin columns (Sartorius), yielding 0.4 mg protein [2].

2.2.3 IpgE

GST-tagged IpgE (from production in BL21 (DE3) CodonPlus®-RIL containing plasmid no. 12 (pCH7, Table 2, p. 23), at 1 mM IPTG, 4 h, 30 °C, 2 L culture volume) was purified *via* glutathione sepharose and PreScissionTM protease cleavage of the GST-tag, as described for IpgB2. Size exclusion chromatography and concentration of the sample yielded 4.5 mg pure protein at a final concentration of 8 mg/ml. The protein was characterised by MALDI-TOF mass spectrometry (determined mass: 14,433 Da, Figure 6, which is in good agreement with the expected value of 14,455 Da) [2].

Figure 6: MALDI-TOF mass spectrum of IpgE [2].



2.3 Crystallisation of Pathogenicity Factors OspD1 and IpgE

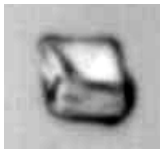
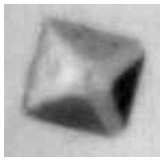
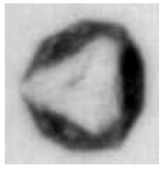
Crystallisation was performed by the vapour diffusion technique at 18 °C. Initial high-throughput screens were set up at the 200 nl scale using a Cartesian pipetting robot and 96-well Greiner sitting drop plates. The following commercial screens were used: Sigma Basic, Sigma Extension, Nextal Classic, Wizard I and II, PEG-Ions.

In case of GST-OspD1, no crystals or crystalline patterns appeared inside the droplets over a period of about 2 month. Nevertheless, in about 50 % of the droplets protein precipitation was observed. Thus, it can be assumed, that the applied protein concentration (6 mg/ml) is basically sufficient for a further screening.

In case of IpgE the conditions listed in Table 3, led to the formation of small crystals [2]. One crystal with a diameter of about 20-50 µm was mounted on a rotating anode X-ray source. Upon radiation, no reflections could be detected. Salt crystals of similar size would have shown few but strong

reflections. Thus, it is likely that these crystals grown under the identified conditions, consist of protein and not salt. However, the crystal size has to be optimised.

Table 3: IpgE crystallisation conditions identified by highthroughput screening [2].

Crystal Image	Protein Concentration	Crystallisation Condition	Screen, Well
	4 mg/ml	20 % 1,4-Butandiol 0.2 M Li2SO4 0.1 M MES pH 6.0	Wizard I&II C2
	6 mg/ml	15 % Jeffamin M-600 25 mM CsCl 50 mM MES pH 6.5	Sigma Extension H9
	4 mg/ml	15 % Jeffamin M-600 25 mM CsCl 50 mM MES pH 6.5	Sigma Extension H9

2.4 Homology Modelling of Pathogenicity Factors IpaA and IpgE

Since no experimental phases have been determined yet for the above described pathogenicity factors, prospective *de novo* structure determination *via* X-ray methods requires either the production of non-native protein crystals such as seleno methionine or heavy metal derivatives. On the other hand, also a homology model could yield appropriate phases by performing molecular replacement. In that case, structure solution could be performed without the need of synchrotron radiation at variable wavelength.

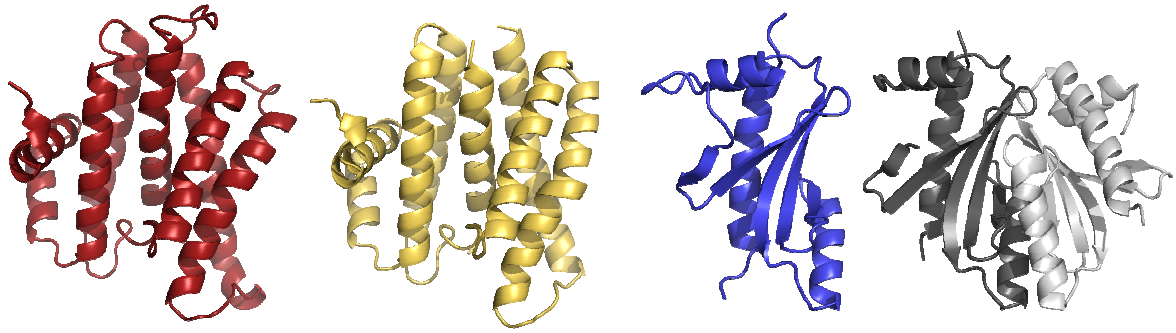
In case of IpaA and IpgE, homologues with already known crystal structures could be identified in the NCBI GenBank® sequence database [13] *via* BLAST query. The respective Clustal W 1.83 [14] sequence alignments are shown in chapter 5.2.3 (Appendix). Sequence identities of IpaA with a *Salmonella*

typhimurium invasion protein (SipA) fragment, and of IpgE with a TTS secretion chaperone of *Salmonella enterica* (SigE) were scored to 29 % (IpaA : SipA), and 24 % (IpgE : SigE). Conserved substitutions were observed in 27 % (IpaA : SipA), and 29 % (IpgE : SigE), leading to a homology of 56 % (IpaA : SipA), and 53 % (IpgE : SigE). Further 15 % (IpaA : SipA), and 14 % (IpgE : SigE) were identified as semi-conserved substitutions, respectively. Based on these sequence alignments, and the crystal structures of the SipA fragment (amino acids 48 – 264, PDB codes 2fm8 – SipA fragment, 2fm9 – SipA fragment in complex with InvB, a TTS chaperone) [15] in case of IpaA, and SigE (PDB code 1k3s – SigE dimer) [16], in case of IpgE respectively, homology models were generated using the program MODELLER [17,18].

The IpaA homology model (fragment of amino acids 48 – 261) shows an α–helix bundle, consisting of eight helices arranged in a compact fold (Figure 7a). The IpgE homology model exhibits the SigE fold with α–β–β–β–α–β–β–α topology [16] (Figure 7b).

Figure 7: homology models of IpaA and IpgE.
a) IpaA homology model for amino acids 48 – 261 (red) and template structure of *S. typhimurium* SipA, 2fm9, chain A (yellow). **b)** IpgE homology model (blue) and template structure of *S. enterica* SigE, 1k3s, chain A (dark grey) and B (light grey).

a) **b)**



2.5 Summary and Outlook

The pathogenicity factor genes *ipaA, ipgB2, ospD1, spa15,* and *ipgE* were cloned into several plasmid vectors. In case of OspD1 and IpgE purification protocols could be successfully established. Purified IpgE was identified *via* mass spectrometry and crystallisation conditions were detected.

Future attempts should include the following suggestions:

- **IpaA, Spa15, Vinculin head domain**: Codon optimisation, similar to the case of IpgE as reported above and separate protein production should be considered. Crystallisation screening should be performed with separated samples as well as with mixed samples to possibly gain crystals of complexes.

- **IpgB2 and OspD1**: Larger expression culture volumes, and purification of the GST-tagged variants should be tried. Crystallisation screening could be carried out with cleaved (tagless), and as well with GST-tagged proteins. GST-tagging obviously does not take sufficient influence on IpgB2 or OspD1 solubility. Thus, additionally tagless expression and purification of both IpgB2 variants, as well as OspD1 should be carried out directly by refolding from inclusion bodies. This procedure would profit from inclusion body purification, where the overproduced protein is available with rather low amounts of cellular protein impurities. Subsequently, a refolding protocol would have to be established. Several methods such as fast or slow, stepwise or continuous dilution, as well as additive screening (e.g. redox reagents like GSH/GSSG) should be performed.

- **IpgE**: Larger crystallisation droplet volumes have to be tested, hopefully leading to protein crystals large enough and suitable for X-ray structure analysis. Micro- and macroseeding techniques should be attempted, as well as the screening around the identified conditions (e.g. variation of the precipitant's concentration, or the pH of the reservoir solution). If crystals could be gained producing diffraction patterns of sufficient quality upon X-ray radiation, molecular replacement with the homology model, or a poly-alanine model retrieved from the homology model could be carried out. As well, seleno methionine or heavy metal derivatives of the IpgE crystals should be produced in order to determine experimental phases.

2.6 Materials and Methods

2.6.1 Primers used for PCR amplification of pathogenicity factor genes from the virulence plasmid pCP301

Cloning techniques, expression, protein purification and crystallisation protocols are described in the diploma theses [1,2].

Table 4: Primers used for amplification of genes *via* PCR attaching restriction sites to the 3'- and 5'-ends.

No.	Name	Sequence
1	IpaA forward	5'-CGG CAT ATG CAT AAT GTA AAT AAT ACT CAA GCG CC-3'
2	IpaA backward	5'-CGG CTC GAG TTA ATC CTT ATT GAT ATT CTT TAA TAC-3'
3	Iba2-Spa15-strep-f	5'-CAT ATG GCT AGC GGC GCC GAG AC-3'
4	Iba2-Spa15-strep-b	5'-GTC TCG GCG CCG CTA GCC ATA TG-3'
5	Spa15-DUETforward	5'-GGC GGA TCC CAT GAG TAA CAT TAA TTT AGT TCA ATT AGT TAG-3'
6	Spa15-DUETbackward	5'-GGC AAG CTT ATA AGA CCC CAT TTA AGA TTT CCA TCC-3'
7	pPR-Iba2-IpaA-20-f	5'-GGC GCC GAG ACC GCG GGT ATG TCA ATT ATT ACA GAT TTA-3'
8	pPR-Iba2-IpaA-20-b	5'-CTA ATT CTG TAA TAA TTG ACA TAC CCG CGG TCT CGG CGC C-3'
9	Vin_head_f	5'-GGC GGC GCG CCA TGC CAG TGT TTC ATA CGC GC-3'
10	Vin_head_b	5'-GGC GCG GCC GCT TCA AGG TCT GGT GGA GGC G-3'
11	ipgB2-Nde I-f	5'-CCG CAT ATG ATT ATA ATG CTT GGA ACA TCT TTT AAT AAT TTT GGA ATC-3'
12	ipgB2-Xho I-b	5'-GGC CTC GAG TCA GAA AGG CGA TTC TAA ATT TGT AAT ATA GTC-3'
13	ipgB2-20-Nde I-f	5'-GGC CAT ATG TAT TTT TCA GGA AAG GTG GAT GAG ATT ATC CG-3'
14	ipgB2-20-Xho I-b	5'-GGC CTC GAG TCA GAA AGG CGA TTC TAA ATT TGT AAT ATA GTC-3'
15	ospD1-Nde I-f	5'-CCG CAT ATG TCA ATA AAT AAC TAT GGA TTA CAT CCA GC-3'

No.	Name	Sequence
16	ospD1-Xho I-b	5'-GGC CTC GAG TTA AAA CAA CTC TTG TAT TTT GTC AGA AAT AAT ATT AAA CAT GTC-3'
17	ipgE-Nde I-f	5'-GGC CAT ATG GAA GAT TTA GCA GAT GTT ATT TGC CG-3'
18	ipgE-Xho I-b	5'-GCC GGA TCC TTA ATA CCC CTT CAT TCT TCG CGC-3'
19	IpgE-mut10-f	5'-GCA GAT GTT ATT TGC CGC GCC TTG GGT ATC C-3'
20	IpgE-mut10-b	5'-GGA TAC CCA AGG CGC GGC AAA TAA CAT CTG C-3'
21	IpgE-mut31-f	5'-GCT TGA TGA TGA TGT GCT GAT TTA TAT TGA AAA AGA AGG AGA TTC-3'
22	IpgE-mut31-b	5'-GAA TCT CCT TCT TTT TCA ATA TAA ATC AGC ACA TCA TCA TCA AGC-3'

2.6.2 Buffers, Enzymes

PreScissionTM protease — 2 u/µl, GE Healthcare

gelfiltration column — HiLoadTM 26/60 SuperdexTM 200
320 mL bed volume

PBS (Phosphate Buffer Saline): 140 mM NaCl
10 mM Na2HPO4
2.8 mM KCl
1.7 mM KH2PO4
pH 7.3

GST cleavage buffer: 50 mM Tris/HCl
pH 7.0
150 mM NaCl
1 mM EDTA
1 mM DTT
0.2 % (v/v) Triton X-100

2.6.3 Cleavage of the GST-tag

After cell lysis GST-tagged proteins were purified from 40 ml cell lysate by addition of 2 ml glutathione sepharose, equilibrated in PBS (Phosphate Buffer Saline) and 3 h incubation at 20 °C. After several PBS washing steps,

200 units of PreScissionTM protease (GE Healthcare) were added with GST cleavage buffer to release free tagless proteins, during over night incubation.

2.6.4 Homolgy modelling

Amino acid sequences were extracted from the deposited crystal structures with PDB codes: 2fm9 (chain A), 2fm8 (chain C), in case of IpaA homolgy modeling, and 1k3s (chains A and B), in case of IpgE homolgy modeling, respectively. These sequences were aligned with the *S. flexneri* IpaA, and IpgE sequences (from SWISS-PROT [19]), respectively, utilising Clustal W 1.83 [14]. The sequences aligned with 29 %, and 24 % identities, respectively. Based on the alignment, ten homology models for IpaA and IpgE were calculated with MODELLER [18]. The best models were selected by picking the model with the lowest value of the MODELLER objective function, which is reported in the second line of the model PDB file.

2.7 References

[1] Schulze Wischeler J. Erstellung und Optimierung verschiedener Expressionssysteme für die Shigellen-Invasine IpaA, IpgB2 und OspD1 mit dem Chaperon Spa15. Pharmazeutische Chemie. Diplomarbeit: Martin-Luther-Universität Halle-Wittenberg / Philipps-Universität Marburg, (2007).

[2] Hasewinkel C. Klonierung und Expression der Shigella-Pathogenitätsgene ipgE, ipgB2, ospD1 und ipaA mit Vinculin oder spa15. Pharmazeutische Chemie. Diplomarbeit: Philipps-Universität Marburg, 2006.

[3] Picking WL, Nishioka H, Hearn PD, Baxter MA, Harrington AT, Blocker A, Picking WD. IpaD of Shigella flexneri is independently required for regulation of Ipa protein secretion and effi-cient insertion of IpaB and IpaC into host membranes. Infect Immun 2005;73:1432-40.

[4] Allaoui A, Sansonetti PJ, Parsot C. MxiJ, a lipoprotein involved in secretion of Shigella Ipa invasins, is homologous to YscJ, a secretion factor of the Yersinia Yop proteins. J Bacteriol 1992;174 (23):7661-9.

[5] Andrews GP, Maurelli AT. mxiA of Shigella flexneri 2a, which facilitates export of invasion plasmid antigens, encodes a homolog of the low-calcium-response protein, LcrD, of Yersinia pestis. Infect Immun 1992;60 (8):3287-95.

[6] Niebuhr K, Jouihri N, Allaoui A, Gounon P, Sansonetti PJ, Parsot C. IpgD, a protein secreted by the type III secretion machinery of Shigella flexneri, is chaperoned by IpgE and implicated in entry focus formation. Mol. Microbiol. 2000;38:8-19.

[7] Page AL, Sansonetti PJ, Parsot C. Spa15 of Shigella flexneri, a third type of chaperone in the type III secretion pathway. Mol Microbiol 2002;43 (6):1533-42.

[8] Parsot C, Ageron E, Penno C, Mavris M, Jamoussi K, d'Hauteville H, Sansonetti PJ, Demers B. A secreted anti-activator, OspD1, and its chaperone, Spa15, are involved in the control of transcription by the type III secretion apparatus activity in Shigella flexneri. Mol. Microbiol. 2005;56:1627-35.

[9] Bourdet-Sicard R, Rudiger M, Sansonetti PJ, Tran Van Nhieu G. Vinculin is unfolded by the Shigella protein IpaA, and the complex promotes F-actin depolymerization. EMBO J 1999;18:5853-62.

[10] Tran Van Nhieu G, Ben Zeev A, Sansonetti PJ. Modulation of bacterial entry in epithelial cells by association between vinculin and the Shigella IpaA invasin. Embo J 1997;16:2717-29.

[11] Jin Q, Yuan Z, Xu J, Wang Y, Shen Y, Lu W, Wang J, Liu H, Yang J, Yang F, Zhang X, Zhang J, Yang G, Wu H, Qu D, Dong J, Sun L, Xue Y, Zhao A, Gao Y, Zhu J, Kan B, Ding K, Chen S, Cheng H, Yao Z, He B, Chen R, Ma D, Qiang B, Wen Y, Hou Y, Yu J. Genome sequence of Shigella flexneri 2a: insights into pathogenicity through comparison with genomes of Escherichia coli K12 and O157. Nucleic Acids Res 2002;30 (20):4432-41.

[12] Grimm C. Röntgenkristallographische Studien an Nukleotidkofaktor-bindenden Enzymen. Fachbereich Pharmazie. Dissertation: Philipps-Universität Marburg, 2001.

[13] Benson D, Karsch-Mizrachi I, Lipman D, Ostell J, Wheeler D. GenBank. Nucleic Acids Res 2007;35:D21-D5.

[14] Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, Thompson JD. Multiple sequence alignment with the Clustal series of programs. Nucleic Acids Res 2003;31 (13):3497-500.

[15] Lilic M, Vujanac M, Stebbins CE. A common structural motif in the binding of virulence factors to bacterial secretion chaperones. Mol Cell 2006;21 (5):653-64.

[16] Luo Y, Bertero MG, Frey EA, Pfuetzner RA, Wenk MR, Creagh L, Marcus SL, Lim D, Sicheri F, Kay C, Haynes C, Finlay BB, Strynadka NC. Structural and biochemical characterization of the type III secretion chaperones CesT and SigE. Nat Struct Biol 2001;8 (12):1031-6.

[17] Marti-Renom MA, Stuart A, Fiser A, Sánchez R, Melo F, Sali A. Comparative protein structure modeling of genes and genomes. Annu. Rev. Biophys. Biomol. Struct. 2000;29:291-325.

[18] Eswar N, Marti-Renom MA, Webb B, Madhusudhan MS, Eramian D, Shen M, Pieper U, Sali A. Comparative Protein Structure Modeling With MODELLER. Current Protocols in Bioinformatics 2000;Supplement 15:5.6.1-5.6.30.

[19] Boeckmann B, Bairoch A, Apweiler R, Blatter MC, Estreicher A, Gasteiger E, Martin MJ, Michoud K, O'Donovan C, Phan I, Pilbout S, Schneider M. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. Nucleic Acids Res 2003;31 (1):365-70.

3 Glu^{235} is critical for substrate selectivity in TGT

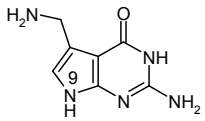
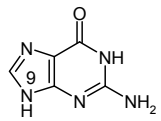
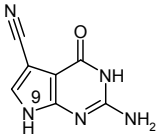
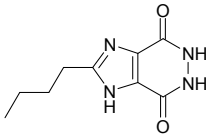
Glutamate versus glutamine exchange swaps substrate selectivity in tRNA-guanine transglycosylase: Insight into the regulation of substrate selectivity by kinetic and crystallographic studies[1]

Transfer RNA-guanine transglycosylases (TGT, E.C. 2.4.2.29) constitute a family of tRNA-modifying enzymes present in all three domains of life. They catalyse the base exchange of specific guanine residues in tRNAs by 7-substituted-7-deazaguanines in a domain specific manner [2,3]. Archaeal TGT inserts preQ$_0$ (for chemical formula see Table 5, p. 40) at position 15 of the majority of known archaeal tRNAs where it is further converted to archaeosine [4]. Bacterial TGT, in contrast, is involved in a pathway introducing the hypermodified base queuine (Q) into the anticodon position 34 (the "wobble position") of just four different tRNAs [5] (Figure 8, p. 40). These are specific for His, Tyr, Asp and Asn, respectively, and share a U^{33}G^{34}U^{35} sequence in common [6,7]. The enzyme performs the first tRNA dependent step in queuine biosynthesis replacing guanine^{34} (G^{34}) by preQ$_1$ [5] (for chemical formula see Table 5, p. 40). This premodified base is produced from GTP [8] by means of the *queC*, *queD*, *queE* and *queF* gene products [9,10]. While the exact functions of QueC, QueD and QueE still have to be figured out, QueF has been identified as an NADPH-dependent oxidoreductase catalysing the reduction of preQ$_0$ to preQ$_1$ [11]. After incorporation into tRNA, preQ$_1$ is converted to the functional Q base in two subsequent steps performed by S-adenosylmethionine:tRNA ribosyltransferase-isomerase (QueA) [12-14] and an unknown coenzyme B$_{12}$-dependent enzyme [15].

---

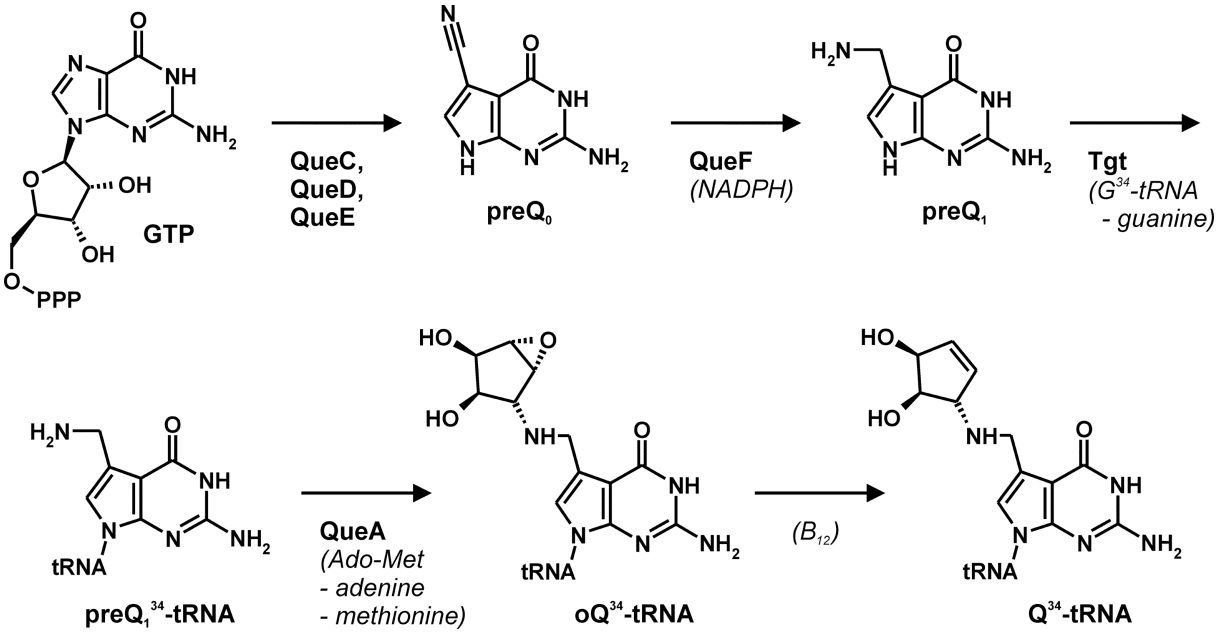
[1]) This chapter is extracted from: Tidten N, Stengl B, Heine A, Garcia GA, Klebe G, Reuter K. J Mol. Biol. 2007; 374 (3): 764-76. Mutagenesis was performed by both first authors; enzyme purification, kinetic characterisation and determination of crystal structures 2z1v, 2pwv were performed by Bernhard Stengl; assay modification and determination of crystal structures 2pwu, 2oko, 2pot, 2z1x, 2z1w were performed by Naomi Tidten.

Table 5: Structures of natural substrates of bacterial TGT and the inhibitor 2-butyl-1*H*-imidazole-4,5-dicarboxylic acid hydrazide (BIH).
In the line "conformation" the functional group of the Leu[231]/Ala[232] peptide bond exposed to the binding pocket is given.

conformation	carbonyl Leu[231] *H*-bonded Glu[235] (deprotonated)	amide Ala[232] *H*-bonded Glu[235] (protonated)	amide Ala[232] stacking Glu[235] (protonated)
natural substrates			
	preQ$_1$	**guanine**	**preQ$_0$**
	K_M: 0.7 µM	K_M: 1.2 µM	K_M: 0.9 µM
inhibitor			
		BIH	
		K_{ic}: 62 µM	

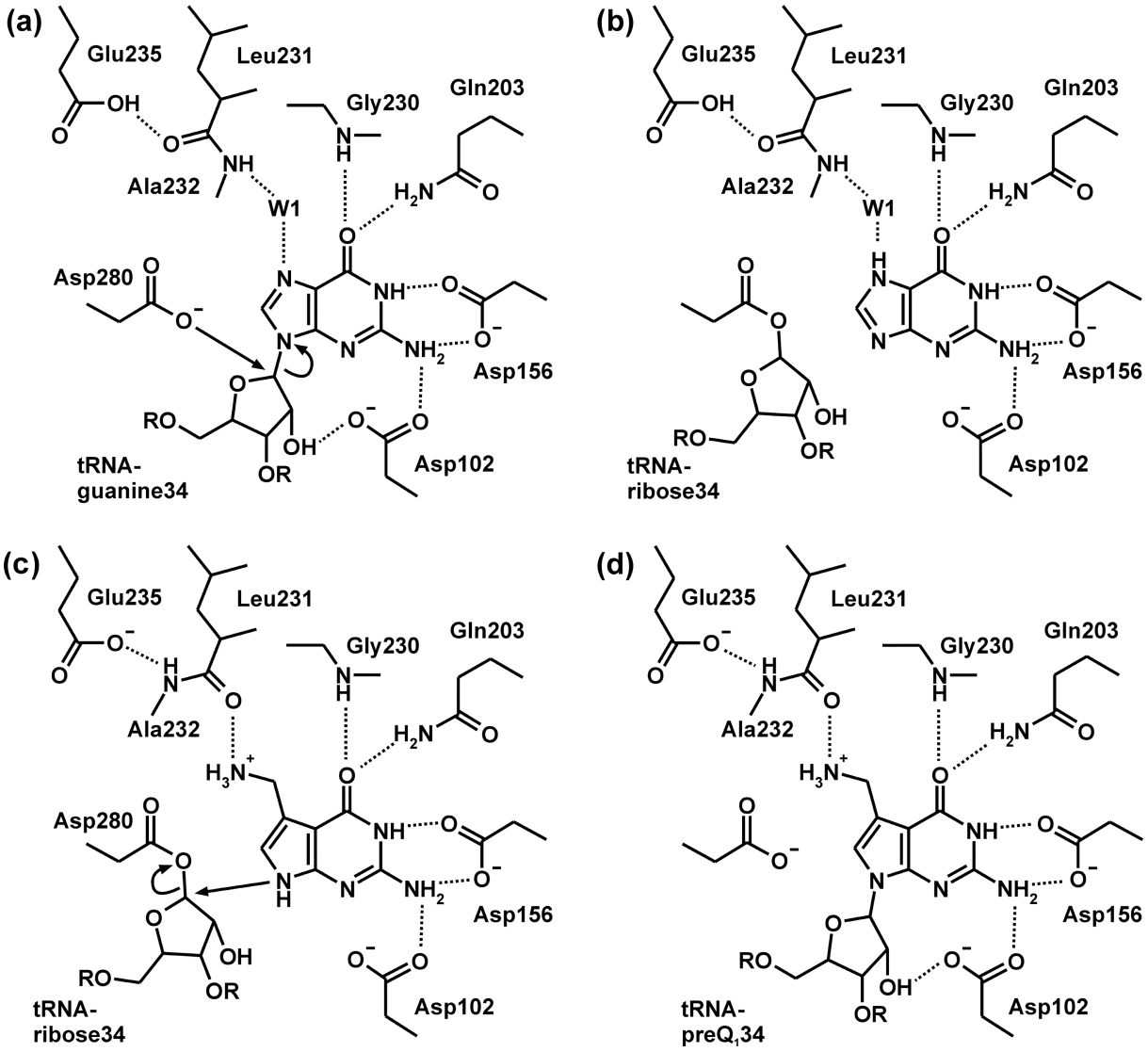
The base exchange catalysed by bacterial TGT follows a ping-pong reaction mechanism [16] which is well documented by crystal structure analyses of TGT from *Zymomonas mobilis* [17-19] (Figure 9, p. 41). In a first step, tRNA binds to TGT and the glycosidic bond of G[34] is cleaved via a nucleophilic attack of Asp[280] producing a covalent TGT·tRNA complex intermediate. Subsequently, preQ$_1$ replaces guanine within the binding pocket and is incorporated into tRNA in a reverse reaction step. During the base exchange step, replacement of guanine by preQ$_1$ requires an adjustment of the binding pocket geometry (Figure 9b, c). Both guanine and preQ$_1$ are recognised by functional groups of Asp[102], Asp[156], Gln[203], and Gly[230]. However, to distinguish between the two bases, a peptide bond formed by Leu[231] and Ala[232] has to flip. G[34]-tRNA binding results in a water mediated contact of the G[34]-*N7* to the main chain amide of Ala[232]. Although the moderate resolution of the G[34]-tRNA bound TGT crystal structure does not allow to unambiguously extract this structural detail [18], the structure of TGT bound to the guanine analogous inhibitor 2-butyl-1*H*-imidazole-4,5-

Figure 8: Queuine modification pathway.
AdoMet: *S*-adenosylmethionine; B_{12}: coenzyme B_{12}; oQ: epoxyqueuine.



dicarboxylic acid hydrazide (BIH) (Table 5, p. 40), clearly suggests this binding mode [20]. In order to permit efficient $preQ_1$ binding the composition of the binding pocket has to be altered. Proper recognition of the $preQ_1$ amino methyl group requires the *CO* acceptor functionality of the Leu^{231}/Ala^{232} peptide bond to be exposed instead of the *NH* donor (Figure 9c, p. 41). Both peptide bond geometries are stabilised from the protein interior using the side chain carboxyl of Glu^{235}, a residue strictly conserved in all bacterial TGTs. Obviously depending on its protonation state, Glu^{235} either donates an *H*-bond to the backwards oriented carbonyl oxygen or accepts an *H*-bond from the amide of the flipped peptide bond (Figure 9b, c, p. 41). Consistently, the peptide switch is also observed in crystal structures of the ligand free enzyme in a pH dependent manner. TGT crystals obtained at pH 5.5 show the amide of the Leu^{231}/Ala^{232} peptide bond exposed to the binding pocket [19], while the opposite orientation is observed in crystals obtained at pH 8.5 (Figure 10d, p. 44).

Figure 9: Base exchange mechanism in bacterial TGT.



In addition to $preQ_1$, bacterial TGT *in vitro* also accepts the biosynthetic precursor of $preQ_1$, namely $preQ_0$, as a substrate [19,21]. This substrate promiscuity is a consequence of the above-mentioned peptide switch as binding of $preQ_0$ is facilitated by the Leu^{231}/Ala^{232} peptide bond conformation required for the accommodation of guanine. Since in bacterial cells both $preQ_0$ and $preQ_1$ are present (even at unknown concentrations), bacterial TGT must be able to discriminate between these substrates and preferentially incorporate $preQ_1$ into tRNA. For *Escherichia coli* TGT there exists evidence that this is achieved through a lower affinity of the enzyme to $preQ_0$, since K_M determination revealed a six-fold higher value for $preQ_0$

(2.4 µM) compared to $preQ_1$ (0.4 µM) [21] (for *Z. mobilis* TGT no kinetic data regarding $preQ_0$ and $preQ_1$ have been reported yet). Although the available complex structures of $preQ_0$ and $preQ_1$ bound to *Z. mobilis* TGT do not provide an ostensible explanation for this selectivity, they reveal some significant differences between the binding modes of the two substrates [19].

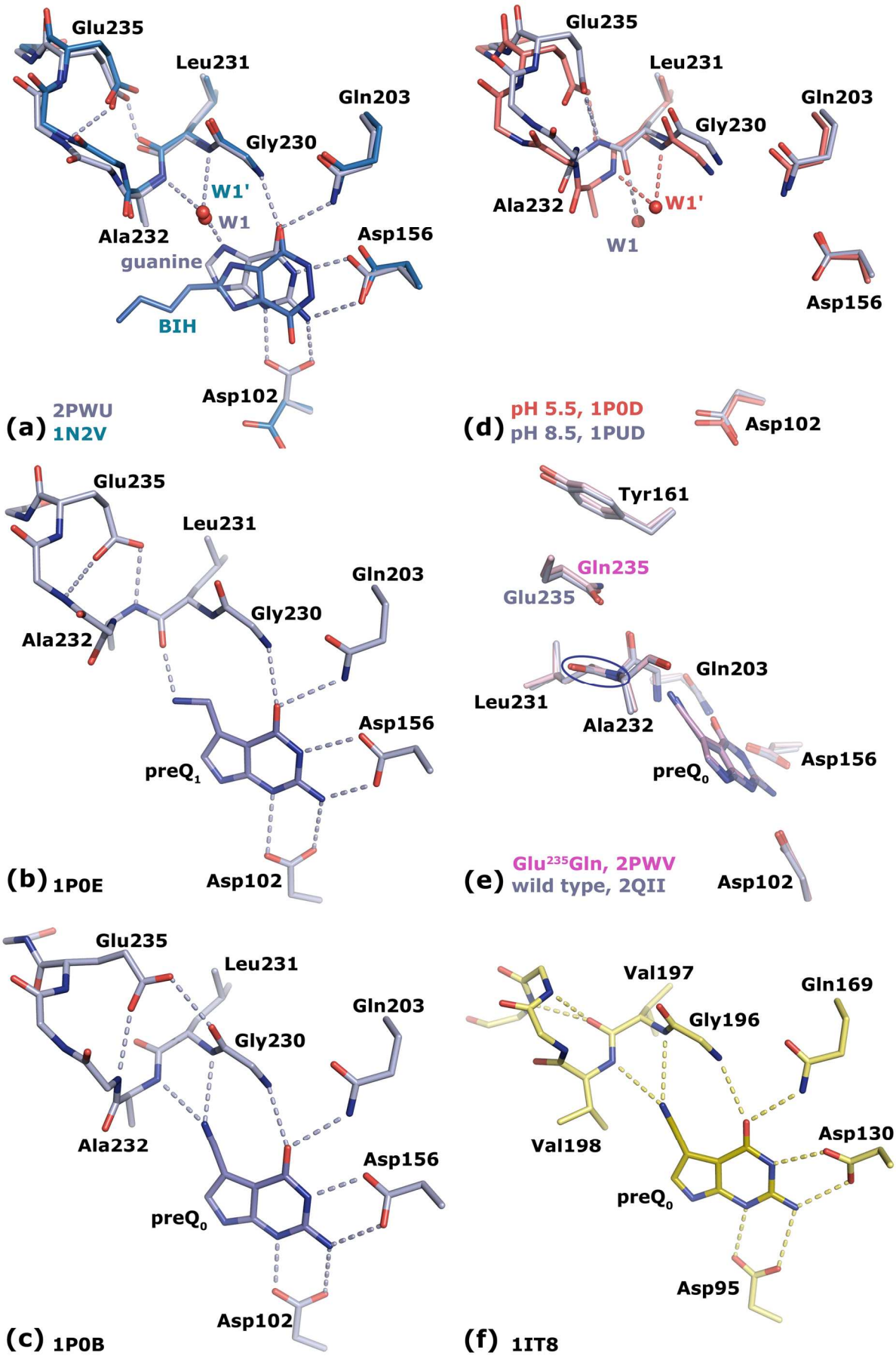
In the TGT·$preQ_1$ complex structure the Glu^{235} carboxyl function is clearly present in a deprotonated state, promoting the formation of an *H*–bond (2.8 Å) to the main chain amide of Ala^{232}. This results in the exposure of the Leu^{231} main chain carbonyl into the binding pocket enabling it to interact with the presumably protonated and thus positively charged amino methyl group of $preQ_1$ (Figure 10b, p. 44). It should be noted that in this scenario the charges of the Glu^{235} carboxylate and the protonated $preQ_1$ amino methyl group compensate for each other.

In the TGT·$preQ_0$ complex the Leu^{231}/Ala^{232} peptide bond is flipped compared to the $preQ_1$ bound situation, *i.e.* the Ala^{232} main chain amide is facing the binding pocket, where it *H*–bonds the ligand's acceptor nitrile group. As mentioned above, this conformation of the Leu^{231}/Ala^{232} peptide bond exactly corresponds to the one observed upon binding of the guanine analogue inhibitor BIH. However, the way this conformation is stabilised by the Glu^{235} side chain carboxyl, is significantly different. Indeed, the Glu^{235} carboxyl function does not *H*–bond in its protonated state the backwards oriented Leu^{231} carbonyl. It rather stacks with its hydrophobic π face between the Leu^{231}/Ala^{232} peptide bond and the aromatic ring system of the Tyr^{161} side chain. This stacking interaction is stabilised by *H*–bonds formed by the Glu^{235} carboxyl group to the main chain amide of Val^{233} (3.1 Å) and to the carbonyl oxygen of Gly^{230} (2.8 Å). The latter *H*–bond clearly indicates that also here the Glu^{235} carboxyl function is present in a protonated state. A shift of the Glu^{235} side chain by *ca.* 1.5 Å towards the binding pocket is prerequisite for the stacking interaction. This shift is rendered possible by a

Glu^{235}/Gly^{236} peptide bond flip which so far had not been observed in any other TGT crystal structure [19] (Figure 10c, e, p. 44).

The present work was aimed to confirm and further analyse the role of Glu^{235} in the Leu^{231}/Ala^{232} peptide switch and in substrate promiscuity of bacterial TGT. In order to mimic a permanently protonated state of the Glu^{235} carboxyl function we introduced a conservative Glu^{235}Gln mutation into the *Z. mobilis* TGT enzyme. This mutation was expected to firmly fix the peptide switch in its *NH* exposing form independent of the applied pH or any bound substrate. This chapter describes the consequences of this mutation on substrate specificity of *Z. mobilis* TGT as studied by means of enzyme kinetics and crystal structure analyses.

Figure 10: Crystal structures of bacterial and archaeal TGT.
(**a**) The superposition of bacterial TGT in complex with the TGT inhibitor BIH (*C* atoms in cyan) and guanine (*C* atoms in grey) shows in both cases an identical conformation for the Leu^{231}/Ala^{232} peptide bond stabilised by an *H*-bond of the Leu^{231} carbonyl to the protonated Glu^{235} side chain carboxyl.
(**b**) Bacterial TGT in complex with preQ$_1$. The Leu^{231}/Ala^{232} peptide bond is switched compared to the guanine/BIH bound situation stabilised by an *H*-bond of the Ala^{232} amide to the deprotonated Glu^{235} carboxyl.
(**c**) Bacterial TGT in complex with preQ$_0$. The Leu^{231}/Ala^{232} peptide bond conformation corresponds to that observed for guanine/BIH bound TGT, but stabilisation surprisingly occurs through a stacking interaction with the protonated Glu^{235} carboxyl concomitant with a Glu^{235}/Gly^{236} peptide bond flip.
(**d**) Superposition of ligand-free bacterial TGT crystallised at pH 5.5 (*C* atoms in red) and pH 8.5 (*C* atoms in grey). Protonation of Glu^{235} at pH 5.5 leads to a binding pocket geometry observed upon binding of guanine/BIH. Deprotonation of Glu^{235} at pH 8.5 results in a geometry as induced upon preQ$_1$ binding.
(**e**) Superposition of bacterial TGT and TGT(Glu^{235}Gln) both in complex with preQ$_0$ reveals identical binding pocket geometries. The Glu^{235} side chain is stacking between the aromatic ring system of Tyr^{161} and the Leu^{231}/Ala^{232} peptide bond (encircled).
(**f**) Archaeal TGT with bound preQ$_0$. The amide group of the Val^{197}/Val^{198} peptide bond (corresponding to Leu^{231}/Ala^{232} in bacterial TGT) is unable to flip. It permanently exposes the Val^{198} *NH*, since the opposing Val^{197} *CO* is involved in an invariant backbone *H*-bond network.


Glu235
Leu231
Gln203
Gly230
W1'
W1
Ala232
guanine
Asp156
BIH
Asp102
(a) 2PWU
1N2V
Glu235
Leu231
Gln203
Gly230
Ala232
Asp156
preQ1
Asp102
(b) 1P0E
Glu235
Leu231
Gln203
Gly230
Ala232
Asp156
preQ0
Asp102
(c) 1P0B
Glu235
Leu231
Gly230
Gln203
W1'
W1
Ala232
Asp156
Asp102
(d) pH 5.5, 1P0D
pH 8.5, 1PUD
Tyr161
Gln235
Glu235
Gln203
Leu231
Ala232
preQ0
Asp156
Asp102
(e) Glu235Gln, 2PWV
wild type, 2QII
Val197
Gln169
Gly196
Val198
Asp130
preQ0
Asp95
(f) 1IT8

3.1 Results

3.1.1 Crystal structure of wild type *Z. mobilis* TGT in complex with guanine

The crystal structure of *Z. mobilis* TGT in complex with the inhibitor 2-butyl-1*H*-imidazole-4,5-dicarboxylic acid hydrazide (BIH) had shown, for the first time, a conformation of the Leu^{231}/Ala^{232} peptide bond, in which the main chain amide of Ala^{232} was exposed to the binding pocket [20]. This had been in contrast to all other crystal structures of *Z. mobilis* TGT determined previously, where this peptide bond was oriented such that the main chain carbonyl of Leu^{231} was facing the binding pocket. Since guanine just like BIH lacks the exocyclic aminomethyl group of preQ$_1$, it had immediately been postulated that binding of guanine to TGT resulted in the same conformation of the Leu^{231}/Ala^{232} peptide bond as binding of BIH. Due to the lack of a high resolution crystal structure of guanine-bound TGT, this assumption had not yet been proven so far. Thus, in order to figure out the exact geometry of the TGT binding pocket with guanine bound, we crystallised *Z. mobilis* TGT in the presence of guanine hydrochloride and determined the complex structure at a resolution of 1.77 Å (Table 7, p. 62). In the electron density map the bound ligand was clearly defined. As observed for the TGT·BIH complex, the Leu^{231}/Ala^{232} peptide bond was present in the *NH* exposing conformation stabilised by a strong *H*-bond (2.4 Å) which was formed between the backwards oriented Leu^{231} carbonyl oxygen and the protonated carboxyl group of Glu^{235}. Also in absolute accordance with the situation in the TGT·BIH complex an interstitial water was bound *H*-bonding both the Ala^{232} main chain amide and the guanine *N7* (Figure 10a, p. 44). Thus, the presented structure confirmed that binding of guanine by TGT leads to the same Leu^{231}/Ala^{232} peptide bond orientation and Glu^{235} conformation as binding of the BIH inhibitor.

3.1.2 Enzymatic characterisation of wild type *Z. mobilis* TGT

Since no kinetic data were available for the *Z. mobilis* TGT enzyme regarding $preQ_0$ and $preQ_1$ we determined and accordingly redetermined k_{cat} and K_M concerning guanine, $preQ_0$, $preQ_1$ and tRNA. Redetermination of the Michaelis/Menten parameters for guanine and tRNA was done by means of a well established assay monitoring the incorporation of radio-labelled guanine into tRNA [22]. Since no radio-labelled $preQ_0$ and $preQ_1$ is commercially available, we used for these substrates a guanine washout assay which in essence was first described by Hoops *et al.* [21] Here, the decrease of radioactivity in tRNA labelled in position 34 with [8-^3H]-guanine due to the incorporation of the respective substrate is monitored. The kinetic parameters for all four substrates are summarised in Table 6 (p. 47). With respect to $preQ_0$ and $preQ_1$ the six-fold difference in K_M reported for the *E. coli* enzyme was not observed for *Z. mobilis* TGT. Within the experimental errors the K_M-values of $preQ_0$ and $preQ_1$ were virtually identical and did not differ from that measured for guanine. Instead, the k_{cat}-values of $preQ_0$ and $preQ_1$ differed significantly by a factor of almost 10. While, compared to guanine, $preQ_1$ was incorporated into tRNA with a higher rate, the incorporation rate of $preQ_0$ was substantially lower than that of guanine. Obviously, in contrast to *E. coli* TGT, in *Z. mobilis* TGT the preferred insertion of $preQ_1$ into tRNA is not achieved by a higher affinity of $preQ_1$ to the enzyme compared to $preQ_0$, but by a clearly higher turnover number.

Table 6: Kinetic parameters for wild type TGT and TGT(Glu^{235}Gln).

wild type Tgt	$tRNA^{Tyr}$ *	[^3H]-guanine	$preQ_1$	$preQ_0$
K_M [µM]	0.9 ± 0.2	1.2 ± 0.2	0.7 ± 0.2	0.9 ± 0.2
k_{cat} [s^{-1}] **	$5.4 \cdot 10^{-2}$	$5.6 \cdot 10^{-2}$	$10.2 \cdot 10^{-2}$	$1.2 \cdot 10^{-2}$
k_{cat}/K_M [$\mu M^{-1}s^{-1}$]	$6.0 \cdot 10^{-2}$	$4.6 \cdot 10^{-2}$	$14.6 \cdot 10^{-2}$	$1.3 \cdot 10^{-2}$

Tgt(Glu^{235}Gln)	$tRNA^{Tyr}$ *	[^3H]-guanine	$preQ_1$	$preQ_0$
K_M [µM]	1.0 ± 0.1	3.3 ± 0.3	32 ± 7	< 0.5
k_{cat} [s^{-1}] **	$7.0 \cdot 10^{-2}$	$7.6 \cdot 10^{-2}$	$10.0 \cdot 10^{-2}$	$0.6 \cdot 10^{-2}$
k_{cat}/K_M [$\mu M^{-1}s^{-1}$]	$7.0 \cdot 10^{-2}$	$2.4 \cdot 10^{-2}$	$0.3 \cdot 10^{-2}$	$> 1.2 \cdot 10^{-2}$

3.1.3 Construction and enzymatic characterisation of *Z. mobilis* TGT(Glu^{235}Gln)

To further investigate the influence of the Glu^{235} carboxyl on the Leu^{231}/Ala^{232} peptide switch and thus on substrate selectivity we introduced a Glu^{235}Gln mutation into *Z. mobilis* TGT by site directed mutagenesis. The carboxamide of the mutated residue was intended to mimic a permanently protonated and electrically neutral Glu^{235} side chain carboxyl. Subsequently, we determined Michaelis/Menten parameters for the mutated enzyme concerning guanine, preQ$_0$, preQ$_1$ and tRNA. The results are summarised in Table 6. As expected, the mutation had no significant influence on the affinity of the tRNA substrate reflected by a virtually unchanged K_M compared to wild type TGT. A slight increase of k_{cat} ($7.0 \cdot 10^{-2}$ s^{-1} vs. $5.4 \cdot 10^{-2}$ s^{-1} measured for the wild type enzyme) using guanine as second substrate was not considered significant with respect to the error range of the assay. Alike, the Glu^{235}Gln mutation obviously had no influence on the turnover numbers for any of the three substrate bases. The k_{cat} values regarding guanine, preQ$_0$ and preQ$_1$ determined for the mutated enzyme agreed well with those determined for the wild type enzyme. In contrast, however, the mutation caused a measurable change in K_M with respect to the substrate bases. While K_M(guanine) was only marginally changed, the most drastical impact of the mutation was observed for K_M(preQ$_1$). Although this base was still accepted as a substrate, K_M was increased by a factor of almost 50 suggesting a noticeable decrease in affinity. Unlike guanine and preQ$_1$, preQ$_0$ seemed to bind with an increased affinity to the mutated enzyme as K_M for this substrate was, compared to wild type TGT, lowered from 0.9 µM to less than 0.5 µM. Due to the slow turnover rate achieved by TGT, in particular with respect to this substrate, an exact determination of K_M was not possible. To assure an excess of substrate even at low concentrations the amount of enzyme used in the assay would have had to be kept too low to obtain a reasonable signal to noise ratio.

In summary, while the Glu^{235}Gln mutation had no significant influence on the turnover number for any of the three substrate bases, K_M was drastically increased for preQ$_1$, marginally elevated for guanine and obviously lowered (even though to an unknown extent) for preQ$_0$. Thus, regarding k_{cat}/K_M as a measure of catalytic efficiency, the Glu^{235}Gln mutation resulted in a clearing or even inversion of substrate preference from preQ$_1$ to preQ$_0$ (see Table 6, p. 47). This agrees well with the hypothesis, that the Leu^{231}/Ala^{232} peptide conformation necessary for guanine and preQ$_0$ accommodation is in wild type TGT stabilised by the protonated state of Glu^{235}, which was mimicked by the Glu^{235}Gln mutation.

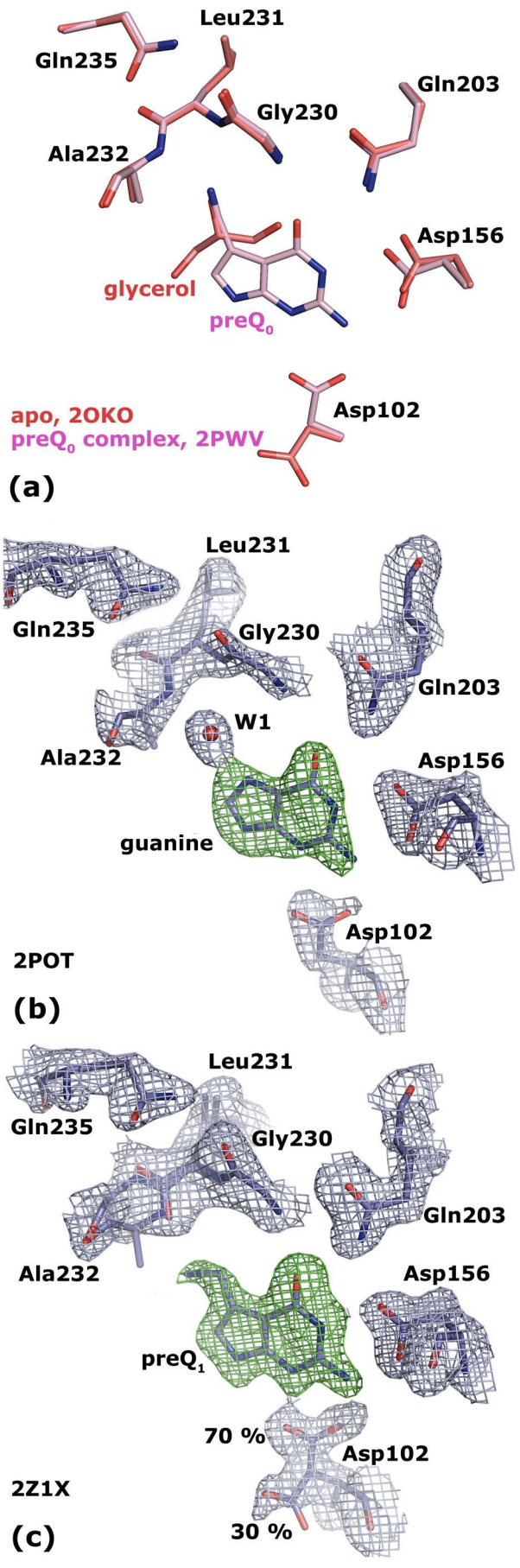
3.1.4 Crystal structure analyses of TGT(Glu^{235}Gln)

To allow the interpretation of the measured enzyme kinetic data at a structural level, we crystallised the mutated enzyme in its apo form as well as in complex with its substrates and the guanine analogous inhibitor BIH (Table 7, p. 62).

Two structures of ligand-free TGT(Glu^{235}Gln) derived from crystals grown at pH 5.5 and pH 8.5, respectively, were determined at nominal resolutions of better than 1.6 Å. In the electron densities of both structures the residues of and near the active site were excellently defined. As expected, in contrast to wild type TGT no pH-dependent conformational change was observed for the Leu^{231}/Ala^{232} peptide bond. Independently whether crystallisation had been performed at pH 5.5 or 8.5, the carbonyl of this peptide bond was oriented backwards while the amide was exposed into the binding pocket. Surprisingly, however, the way this conformation was stabilised by the Gln^{235} carboxamide did not correspond to the way it is stabilised by the protonated Glu^{235} carboxyl in wild type apo TGT crystallised at pH 5.5. Rather, the geometry of the binding pocket was virtually identical to the one observed in the structure of wild type TGT with preQ$_0$ bound. No *H*-bond was formed

Figure 11: Crystal structures of TGT(Glu^{235}Gln).

(**a**) Superposition of TGT(Glu^{235}Gln) in its apo form (crystallised at pH 5.5; the structure obtained at pH 8.5 looks identical) and in complex with preQ$_0$ reveals equal binding pocket geometries. (**b**) $2|F_o| - |F_c|$ electron density map of TGT(Glu^{235}Gln) in complex with guanine contoured at 1.5 σ (blue). The green density represents an $|F_o| - |F_c|$ map with the ligand omitted from the calculation contoured at 3.0 σ. (**c**) $2|F_o| - |F_c|$ electron density map of TGT(Glu^{235}Gln) in complex with preQ$_1$ contoured at 1.5 σ (blue). The green density represents an $|F_o| - |F_c|$ map with the ligand omitted from the calculation contoured at 3.0 σ. The electron density reveals a split conformation for the Asp^{102} side chain indicating an uncomplete occupancy of the binding pocket with the substrate base. The Leu^{231}/Ala^{232} peptide bond is present in two different conformations. The occupancy of the ligand and of the predominant Leu^{231}/ Ala^{232} and Asp^{102} conformers refine to 0.7.



between the Gln^{235} carboxamide and the Leu^{231} carbonyl oxygen. Instead, the carboxamide stacked between the Leu^{231}/Ala^{232} peptide bond and the side chain aromatic ring system of Tyr^{161}. As a consequence, the Gln^{235} side chain was shifted towards the binding pocket concomitant with a Gln^{235}/Gly^{236} peptide flip so far only

observed in $preQ_0$ bound wild type TGT (with Gln^{235} being replaced by Glu^{235}). To investigate the influence of a bound substrate on the active-site geometry of TGT($Glu^{235}Gln$), we determined the crystal structures of the mutated enzyme in complex with $preQ_0$, guanine, the BIH inhibitor, and $preQ_1$, respectively.

The TGT($Glu^{235}Gln$)·$preQ_0$ complex structure was determined at a resolution of 1.70 Å. The electron density for the $preQ_0$ ligand was clearly defined. Superposition of the TGT($Glu^{235}Gln$)·$preQ_0$ complex structure and the structures of the ligand-free mutated enzyme revealed that the binding pocket geometry was virtually identical in all three structures (Figure 11a, p. 50) and exactly corresponded to the one observed in $preQ_0$ bound wild type TGT (Figure 10e, p. 44). The nitrile nitrogen of $preQ_0$ was present at a position occupied by a glycerol oxygen picked up from the cryo buffer in the structures of uncomplexed mutated TGT (Figure 11a, p. 50).

The crystal structures of TGT($Glu^{235}Gln$) in complex with guanine and with the guanine analogous inhibitor BIH were determined at a resolution of 1.80 Å and 1.63 Å, respectively. In both structures the electron densities for the ligands as well as for the residues constituting the binding pockets were excellently defined (for the guanine complex see Figure 11b (p. 50). It was clearly visible that also binding of guanine or BIH did, compared to the apo-form, not result in any conformational change of the Leu^{231}/Ala^{232} peptide bond or of Gln^{235}. As in guanine or BIH-bound wild type TGT, the exposed Ala^{232} main chain amide formed an *H*-bond to an interstitial water molecule, which was further *H*-bonded to *N* 7 of guanine or the corresponding nitrogen of BIH, respectively. In contrast to wild type TGT, however, this conformation was not stabilised by an *H*-bond formed between the Leu^{231} carbonyl and the carboxamide of Gln^{235} but by a stacking interaction between the Leu^{231}/Ala^{232} peptide bond and the Gln^{235} side chain.

In order to understand the way $preQ_1$ was bound by the mutated TGT we determined the crystal structure of the TGT(Glu^{235}Gln)·$preQ_1$ complex at a resolution of 1.63 Å. Also in this structure the electron density attributable to the ligand was very well defined. A split conformation of the Asp^{102} side chain, however, indicated an incomplete occupancy of the binding pocket with $preQ_1$ (Figure 11c, p. 50). In apo structures of TGT the Asp^{102} side chain carboxylate is oriented towards the outside of the binding pocket *H*-bonding the side chain amide of Asn^{70} as well as the main chain amide of Thr^{71}. Once a substrate base has bound, the Asp^{102} side chain rotates into the binding pocket towards the ligand in order to *H*-bond its exocyclic amine at position 3 [23]. In the TGT(Glu^{235}Gln)·$preQ_1$ complex structure *ca.* 30 % of Asp^{102} exhibited a side chain conformation pointing towards the outside of the binding pocket suggesting a portion of the binding pocket being unoccupied by the ligand. This was in line with the fact that $K_M(preQ_1)$ was drastically increased for TGT(Glu^{235}Gln). As in all other structures of TGT(Glu^{235}Gln) the Gln^{235} side chain stacked on top of the Leu^{231}/Ala^{232} peptide bond. Accordingly, the conformation of the Gln^{235}/Gly^{236} peptide bond remained also here unchanged compared to all other TGT(Glu^{235}Gln) structures as well. Interestingly, however, the electron density attributable to the Leu^{231}/Ala^{232} peptide bond could only been interpreted such that it was present in two different conformations. One conformer corresponded exactly to the one observed in the remaining TGT(Glu^{235}Gln) structures with the main chain amide of Ala^{232} being exposed to the binding pocket. Indeed, this conformation was hardly compatible with a bound $preQ_1$ molecule due to unfavourable interactions between the Ala^{232} main chain amide and the amino methyl group of $preQ_1$. The fact that this conformation could be refined to an occupancy of *ca.* 30 % agreed well with the assumption that a portion of the binding pocket was devoid of ligand. In the second conformer the Leu^{231}/Ala^{232} peptide bond was flipped enabling the exposed Leu^{231} carbonyl oxygen to *H*-bond the amino methyl group of $preQ_1$. This geometry combining the exposed Leu^{231} carbonyl with the stacking conformation of

Gln^{235} (or Glu^{235}) had never been observed in any TGT crystal structure before. As a consequence, exposure of the Leu^{231} carbonyl into the binding pocket does not necessarily require the deprotonated carboxyl of Glu^{235} *H*-bonding the backwards oriented Ala^{232} main chain amide. It is also compatible with a stacking interaction of the Gln^{235} side chain, although this mode of stabilisation is obviously energetically less favourable as reflected by the drastically increased K_M(preQ$_1$) measured for TGT(Glu^{235}Gln).

3.2 Discussion

In this study we have confirmed and further analysed the role of Glu^{235} as a general acid/base not directly involved in catalysis but facilitating the binding of various substrates with different *H*-donor and -acceptor properties. The high resolution crystal structure of *Z. mobilis* TGT in complex with guanine presented in this work gives the final proof that binding of each of the three substrate bases (guanine, preQ$_0$ and preQ$_1$) triggers a different geometry of the TGT binding pocket. Prerequisite of the substrate promiscuity observed in TGT is the ability of the Leu^{231}/Ala^{232} peptide bond to flip and thus to change its orientation depending on the bound substrate. Remarkably, the bound substrate not only influences the orientation of this peptide bond but, moreover, the way it is stabilised by the strictly conserved Glu^{235}. Obviously, the highly flexible interplay of the Leu^{231}/Ala^{232} peptide bond and the general acid/base Glu^{235} facilitates an optimal adaptation of the TGT binding pocket to each of its substrate bases. This is reflected by the virtually identical K_M values of *Z. mobilis* TGT for guanine, preQ$_0$ and preQ$_1$ as determined in this study. Moreover, our results show that the preferred incorporation of the "correct" substrate preQ$_1$ into tRNA over its biosynthetic precursor preQ$_0$ is achieved by a *ca.* 10 fold faster turnover rate. It should be noted, however, that to the best of our knowledge it has not been investigated so far, if QueF, the nitrile reductase reducing preQ$_0$ to preQ$_1$, is also acting on preQ$_0$ incorporated into tRNA. In this case preQ$_0$ could represent an *in vivo*

substrate of TGT as well. The observed differences in k_{cat} for the three substrate bases may hardly be explained by their binding modes observed in the corresponding crystal structures, but possibly may be deduced from their chemical properties. In each case nitrogen *N*9 of the respective base nucleophilically attacks *C*1 of the covalently bound tRNA ribose[34] during catalysis (Figure 9c, p. 41 and Table 5, p. 40). In $preQ_1$ this nitrogen is the most potent nucleophile, as it is in no electronic conjugation with the exocylic amino methyl group. In guanine, *N9* is less nucleophilic due to an electron withdrawing effect of *N7* being in conjugation with *N*9. The $preQ_0$ substrate is the least reactive base in this series. Here, the exocyclic nitrile function is conjugated to the endocyclic nitrogen, resulting in an even more pronounced electron withdrawing effect. This is fully consistent with the observed decrease in turnover in the order $preQ_1$ > guanine > $preQ_0$. The virtually identical K_M-values for the three substrate bases indicate that the conformation of the Leu[231]/Ala[232] peptide bond as well as the way it is stabilised by the Glu[235] side chain carboxyl may have little influence on the energetic state of the substrate binding pocket.

With intent to create a bacterial TGT with altered substrate specificity Glu[235] was mutated to Gln in an attempt to mimic a permanently protonated Glu[235] side chain carboxyl. Since the Leu[231]/Ala[232] peptide bond conformation required for the accommodation of $preQ_1$ involves stabilisation by a deprotonated Glu[235] the mutation was expected to strongly disfavour or even disable $preQ_1$ binding while leaving unaffected or even improving the binding of guanine and $preQ_0$. Kinetic characterisation of TGT(Glu[235]Gln) showed that the change of K_M(guanine) caused by the mutation was at the border of significance, while $K_M(preQ_0)$ was (to an unknown extent) decreased compared to wild type TGT. Although $K_M(preQ_1)$ of the mutated enzyme was increased by a factor of almost 50, the remaining affinity to this substrate was still unexpectedly high. The crystal structures of TGT(Glu[235]Gln) in its apo-form as well as in complex with various ligands provided an explanation to the observed kinetic data. The same geometry of the binding pocket was

observed, no matter if the mutated enzyme was crystallised in its apo-form (independent of the applied pH) or in complex with guanine, BIH, or $preQ_0$. In each case, the Ala^{232} main chain amide was exposed to the binding pocket, stabilised by a stacking interaction with the Gln^{235} side chain carboxamide. This conformation was analogous to that observed in wild type TGT bound to $preQ_0$. Seemingly, a surprisingly strong *H*-bond formed between the Leu^{231} carbonyl and the protonated Glu^{235} side chain carboxyl (2.4 Å) plays an important role in stabilising the binding pocket geometry observed in wild type TGT crystallised at pH 5.5 or in complex with guanine/BIH. Due to its less electronegative nitrogen and the presence of two hydrogens over which the positive partial charge is distributed the carboxamide of Gln^{235} is obviously not able to form as strong an *H*-bond to the Leu^{231} carbonyl rendering the stacking interaction energetically more favourable. Hence, the binding pocket of TGT(Glu^{235}Gln) seems predestined for the accommodation of $preQ_0$, since independent of any pH no structural rearrangement is required for its binding. *A priori* the binding pocket is present in a geometry identical to the one induced upon $preQ_0$ binding in wild-type TGT. This finding is supported by the (even though to an unknown extent) reduced $K_M(preQ_0)$ determined for TGT(Glu^{235}Gln). As a surprise, the crystal structure of TGT(Glu^{235}Gln) in complex with $preQ_1$ revealed a binding pocket geometry featuring an exposed Leu^{231} carbonyl stabilised by a Gln^{235} stacking interaction. This unexpected geometry showed that the stacking interaction still allows the Leu^{231}/Ala^{232} peptide flip. As a result, the substrate specificity achieved by the Glu^{235}Gln mutation was not as strict as that observed for archaeal TGT which is absolutely unable to bind $preQ_1$. It permanently exposes the amide group of the Val^{197}/Val^{198} peptide bond (corresponding to Leu^{231}/Ala^{232} in bacterial TGT) allowing both the binding of guanine and of $preQ_0$ but not of $preQ_1$. The peptide flip necessary to accommodate this base is rendered impossible, since the opposing carbonyl of this peptide bond is involved in an invariant backbone *H*-bond network which does not imply the carboxyl group of an acidic amino acid [2,24] (Figure 10f, p. 44).

In summary, the mutated TGT(Glu^{235}Gln) investigated in this study behaved in several details different from what was expected, although a more conservative exchange than glutamic acid / glutamine can hardly be performed. The present work clearly shows the indispensability of crystal structure analyses for the correct interpretation of kinetic data obtained from mutated proteins.

Furthermore, this study provides an example that mutation of a residue which does not even directly interact with any ligand enables the inversion of substrate selectivity of an enzyme. This may generally underline the importance to consider residues in the "second sphere" of binding pockets during attempts of creating enzymes with new binding properties.

The successful modulation of selectivity is an important step towards the understanding of selectivity and specificity determining features in TGTs. In *E. coli* TGT replacement of Asp^{156} (*Z. mobilis* numbering) by Asn altered the specificity towards xanthine even though at the expense of a reduced catalytic activity [25]. In this context it might be interesting to study mutational exchanges in the close neighbourhood of the attacking nucleophile Asp^{280}. In bacterial TGT this residue is tightly kept in position by *H*–bonds formed with Gly^{261} and Tyr^{258} [18]. In archaeal TGT the tyrosine is replaced by a strictly conserved histidine [2]. Likely, this exchange will have pronounced influence on the catalytic properties.

These considerations provide a perspective towards an ambitious goal: the modification of substrate specificity towards bases other than preQ$_0$ or preQ$_1$ and their efficient incorporation into the wobble position of tRNAs. This would allow to study the translational process in more detail via well designed modulation of the accuracy of the wobble base pairing.

3.3 Summary

Bacterial tRNA-guanine transglycosylase (TGT) catalyses the exchange of guanine in the wobble position of particular tRNAs by the modified base $preQ_1$. *In vitro*, however, the enzyme is also able to insert the immediate biosynthetic precursor, $preQ_0$, into those tRNAs. This substrate promiscuity is based on a peptide switch in the active site, gated by the general acid/base Glu^{235}. The switch alters the properties of the binding pocket to allow either the accommodation of guanine or $preQ_1$. The peptide conformer recognising guanine, however, is also able to bind $preQ_0$. To investigate selectivity regulation, kinetic data for *Z. mobilis* TGT were recorded. They show that selectivity in favour of the actual substrate $preQ_1$ over $preQ_0$ is not achieved by a difference in affinity but via a higher turn-over rate. Moreover, a TGT(Glu^{235}Gln) variant was constructed. The mutation was intended to stabilise the peptide switch in the conformation favouring guanine and $preQ_0$ binding. Kinetic characterisation of the mutated enzyme revealed that the Glu^{235}Gln exchange has, with respect to all substrate bases, no influence on k_{cat}. In contrast, $K_M(preQ_1)$ is drastically increased while $K_M(preQ_0)$ seemes to be decreased. Hence, regarding k_{cat}/K_M as an indicator for catalytic efficiency, selectivity of TGT in favour of $preQ_1$ is abolished or even inverted in favour of $preQ_0$ for TGT(Glu^{235}Gln). Crystal structures of the mutated enzyme confirm that the mutation strongly favours the binding pocket conformation required for the accommodation of guanine and $preQ_0$. The way this is achieved, however, significantly differs from what was predicted based on crystal structures of wild type TGT.

3.4 Materials and Methods

3.4.1 Cloning and TGT preparation

The QuikChangeTM site-directed mutagenesis kit (Stratagene) was used to introduce a Glu^{235}Gln mutation into the wild type *tgt* expression plasmid pET9d-ZM4 [26] following the vendor's protocol. The primers used for mutagenesis were E235Q-s (5'-GGG GGA TTG GCT GTG GGT CAA GGA CAG GAT GAA ATG-3') and E235Q-a (5'-CAT TTC ATC CTG TCC TTG ACC CAC AGC CAA TCC CCC-3') (mutation underlined). Sequencing of the entire *tgt* gene (MWG Biotech, Ebersberg) confirmed the presence of the desired mutation as well as the absence of any further unwanted mutation. Subsequently, the mutated plasmid pET9d-ZM4-E235Q was transformed into *E. coli* BL21 (DE3) pLysS cells. These cells were used for the preparation of the TGT enzyme following the method described by Romier *et al.* [27].

3.4.2 Preparation of tRNA^{Tyr}

Preparation of *E. coli* tRNA^{Tyr} (ECY2) *via* in vitro transcription was done following the method described by Curnow *et al.* [28].

3.4.3 Kinetic parameters

Kinetic parameters for TGT and TGT(Glu^{235}Gln) have been determined for the different substrates essentially as described in Meyer *et al.* [22].

Michaelis-Menten parameters for tRNA and guanine were determined monitoring incorporation of radioactively labelled guanine into tRNA^{Tyr} in position 34. Kinetic data for both substrates were determined separately in triplicate and average values were calculated. Kinetic parameters for guanine were measured using 150 nM TGT, 15 µM unlabelled tRNA^{Tyr}, and variable

concentrations of [8-^{3}H]-guanine (0.9 Ci/mmol,; Hartmann Analytic) in the range of 0.5 – 20 µM in buffer solution (200 mM HEPES pH 7.3, 20 mM $MgCl_2$) and 2.95 µM (≡ 5 % of critical micellar concentration) Tween 20 (Roth). Kinetic parameters for tRNA were measured using 150 nM TGT, 20 µM [8-^{3}H]-guanine, and variable concentrations of $tRNA^{Tyr}$ (0.25 – 15 µM). Initial velocities of the base exchange reaction in counts per minute were converted to [µM/min] using a calibration constant derived from liquid scintillation counting of [8-^{3}H]-guanine solutions with variable concentrations. Kinetic parameters were determined via double-reciprocal linearisation using the method of Eadie-Hofstee and linear regression using GraFit [29].

Michaelis-Menten parameters for $preQ_0$ and $preQ_1$ were calculated *via* monitoring the loss of [8-^{3}H]-guanine from $tRNA^{Tyr}$ radio-labelled in position 34. To produce radioactively labelled tRNA 50 µM unmodified $tRNA^{Tyr}$ was incubated with 0.5 µM TGT and 10 µM [8-^{3}H]-guanine (11.8 Ci/mmol, Hartmann Analytic) in TGT assay buffer for 2 h. TGT and free guanine/ [8-^{3}H]-guanine were extracted from the reaction mixture by the addition of equal volumes of Roti-Phenol (Roth) and chloroform : isoamylalcohol (24:1). The aqueous supernatant was once more treated with an equal volume of chloroform : isoamylalcohol (24:1). From the aqueous supernatant containing the radioactively labelled tRNA samples were retained to determine the specific activity of the radio-labelled tRNA. An additional purification step was performed via gel filtration using NAP-columns (GE Healthcare, Life Science) and TGT assay buffer. From the eluted solution again samples were retained to derive the final concentration of the radio-labelled tRNA from its specific activity. For each preparation it amounted to *ca.* 20 µM. Kinetic parameters for $preQ_0$ and $preQ_1$ were measured using 150 nM wild type TGT or TGT($Glu^{235}Gln$) and 8 µM radio-labelled $tRNA^{Tyr}$. The concentration of $preQ_0$ was varied in a range of 0.5 – 15 µM. For $preQ_1$ variable concentrations in a range of 0.5 – 15 µM for wild type TGT and 2 – 80 µM for TGT($Glu^{235}Gln$) were applied. Initial velocities in counts per minute were calculated from the decreasing tritium labelling level of tRNA due to the

incorporation of the respective substrate bases. Initial velocities in counts per minute were converted to [µM/min] using a calibration constant derived from liquid scintillation counting of guanine/ [8-^{3}H]-guanine solutions with variable concentrations. Kinetic parameters were determined *via* double-reciprocal linearisation using the method of Eadie-Hofstee and linear regression using GraFit [29].

3.4.4 Crystal structure analyses

TGT(Glu^{235}Gln) crystals suitable for ligand soaking were produced in a two step procedure. Droplets were prepared by mixing 2 µL of concentrated protein solution (14 mg/mL) with 2 µL reservoir solution [100 mM MES, pH 5.5, 1 mM DTT, 13 % (w/v) PEG 8,000, 10 % (v/v) DMSO]. Micro-crystals were grown at 291 K using the hanging-drop vapour diffusion method in the presence of 1.0 mL of reservoir solution of the respective seeding buffer. Micro crystals grew within two weeks.

Subsequently, macro-seeding was performed under similar conditions to obtain crystals of the uncomplexed protein. Again droplets were prepared by mixing 2 µL of concentrated protein solution with 2 µL macro-seeding reservoir solution [100 mM MES, pH 5.5, 1 mM DTT, 8 % (w/v) PEG 8,000, 10 % (v/v) DMSO]. One micro-crystal was transferred into this solution. Single crystals with a size of approximately 0.7 x 0.7 x 0.2 mm^3 grew within two to four weeks.

To obtain crystals of apo-TGT(Glu^{235}Gln) at pH 8.5, buffers were used identical to those described above but with 100 mM Tris/HCl pH 8.5 instead of 100 mM MES pH 5.5 as buffer system.

To allow co-crystallisation of the mutated TGT in complex with preQ$_0$ the compound was dissolved in DMSO and added to the macro-seeding droplet to a final concentration of 20 mM. Crystals of mutated TGT complexed to

$preQ_1$, BIH and guanine as well as crystals of wild type TGT bound to guanine were produced by a soaking procedure. The compounds were dissolved in DMSO and added to 2 µL of a crystallisation droplet to a final concentration of 10 to 20 mM. Finally, a single TGT crystal was transferred into the droplet and soaked for 30 min.

For data collection, crystals were cryoprotected using glycerol. The glycerol and PEG 8,000 concentrations of the macro-seeding buffers were increased stepwise by transferring the crystals to six different 2 µL cryo-droplets each with 5 to 30 min incubation times [glycerol concentrations (v/v): 5 % → 10 % → 15 % → 20 % → 25 % → 30 %; and PEG 8,000 concentrations (w/v): 5.0 % → 6.3 % → 7.5 % → 8.0 % → 8.8 % → 9.8 %, respectively]. These droplets also contained the ligands at equivalent concentrations compared to soaking and co-crystallisation conditions. The cryo-soaked crystals were flash-frozen in liquid N_2. Data sets were collected at cryo conditions (100 K) with CuKα radiation (λ = 1.5418 Å) using a Rigaku RU-300 rotating-anode generator at 50 kV and 90 mA equipped with Xenocs focussing optics and an R-AXIS IV detector. All crystals tested exhibited monoclinic symmetry in space group C2 containing one monomer per asymmetric unit with Matthews coefficients of 2.3 - 2.4. Data processing and scaling was performed using the HKL2000 package [30]. For all refined structures unit cell dimensions for the crystals, data collection and processing statistics are given in Table 7.

For crystals grown at pH 5.5 coordinates of the apo TGTt crystal structure grown at pH 5.5 (PDB-code: 1P0D) were directly applied for initial rigid-body refinement of the protein molecule followed by repeated cycles of conjugate gradient energy minimisation, simulated annealing and B-factor refinement using the CNS program package [31]. For the crystal grown at pH 8.5 the coordinates of the apo TGT crystal structure grown at pH 8.5 (PDB-code: 1PUD) were applied. Refinement at the later stages for all other structures was performed with SHELXL [32]. Here, up to 50 cycles of conjugate gradient minimisation were performed with default restraints on bonding geometry

and B-values: 5 % of all data were used for R_{free} calculation. Amino acid side chains were fit to σA-weighted 2 $|F_o|$ – $|F_c|$ and $|F_o|$ – $|F_c|$ electron density maps using Coot [33]. Water, glycerol molecules, and ligands were located in the difference electron density and added to the model for further refinement cycles. Anisotropic conjugate gradient refinement resulted in a significant improvement of the models built for apo TGT(Glu^{235}Gln) (pH 5.5 and 8.5), TGT(Glu^{235}Gln)·BIH, and TGT(Glu^{235}Gln)·preQ$_1$. During the last refinement cycles, riding *H*-atoms were introduced for the protein residues (not for the ligand) without using additional parameters. All final models were validated using PROCHECK [34]. Data refinement statistics are given in Table 7.

3.4.5 Protein Data Bank Accession Codes

The following Protein Data Bank (PDB) accession codes were allocated to the crystal structures determined during this work (also given in Table 7):

TGT·preQ$_0$: **2QII**
TGT·guanine: **2PWU**
TGT(Glu^{235}Gln) pH 5.5: **2OKO**
TGT(Glu^{235}Gln) pH 8.5: **2Z1V**
TGT(Glu^{235}Gln)·guanine **2POT**
TGT(Glu^{235}Gln)·BIH **2Z1W**
TGT(Glu^{235}Gln)·preQ$_0$ **2PWV**
TGT(Glu^{235}Gln)·preQ$_1$: **2Z1X**

3.4.6 Crystallographic Tables

Table 7: Crystallographic data collection and refinement statistics.

a) number in parentheses is for highest resolution shell

b) $R(I)_{sym} = \Sigma\ |I - \langle I\rangle|\ /\ \Sigma I$ with I representing the observed intensities and ⟨I⟩ representing the mean observed intensity

c) $R_{work} = \Sigma_{hkl}\ |F_o - F_c|\ /\ \Sigma_{hkl}\ |F_o|$.

d) R_{free} was calculated as for R_{work} but on 5 % of the data excluded from the refinement.

Crystal data	Wild type · guanine (2pwu)	Glu235Gln, pH 5.5 (2oko)	Glu235Gln, pH 8.5 (2z1v)	Glu235Gln · guanine (2pot)
A. Data processing				
Space group	*C2*	*C2*	*C2*	*C2*
a, b, c (Å)	90.2, 64.5, 70.8	90.4, 65.2, 70.5	90.8, 65.2, 70.5	89.1, 63.7, 71.1
β (°)	92.8	96.4	96.1	92.9
Resolution range[a] (Å)	50–1.77 (1.80–1.77)	50–1.50 (1.53–1.50)	20–1.55 (1.58–1.55)	50–1.80 (1.83–1.80)
Total no. of reflections	105,401	210,241	132,929	90,611
No. of unique reflections	36,424	59,104	56,093	33,979
Completeness[a] (%)	94.8 (83.9)	91.5 (85.1)	95.6 (78.5)	97.6 (83.4)
Redundancy	2.8	3.5	2.3	2.5
$R(I)_{sym}$[a,b] (%)	4.4 (38.9)	3.2 (15.9)	5.8 (23.9)	4.1 (46.7)
$I/\sigma(I)$[a]	22 (2.3)	33 (8.0)	16 (3.8)	21 (1.5)
B. Refinement				
R_{work}[c] / R_{free}[d] (%)	20.0/27.1	12.5/17.4	14.3/20.3	21.9/28.6
No. of atoms/residues (molecules)				
Protein	2702/358	2824/365	2860/369	2688/355
Water	101	346	305	86
Glycerol (cryo buffer)	18/3	96/16	24/4	42/7
Ligand	14/1	-	-	11/1
Mean *B*–factors (Å^2)				
Protein	32.8	21.8	21.6	39.0
Water	34.0	34.8	31.3	40.4
Glycerol (cryo buffer)	52.6	48.1	39.4	62.6
Ligand	60.4	-	-	53.5
rmsd angle (°)	2.0	2.2	2.2	2.0
rmsd bond (Å)	0.007	0.011	0.010	0.006
(An)isotropic refinement	Isotropic	Anisotropic	Anisotropic	Isotropic

Crystal data	Glu235Gln · BIH (2z1w)	Glu235Gln · preQ0 (2pwv)	Glu235Gln · preQ1 (2z1x)
A. Data processing			
Space group	*C2*	*C2*	*C2*
a, b, c (Å)	90.9, 65.3, 70.6	90.5, 65.2, 70.4	89.0, 64.2, 70.6
β (°)	96.7	96.3	93.1
Resolution range[a] (Å)	50–1.63 (1.66–1.63)	20–1.70 (1.73–1.70)	50–1.63 (1.66–1.63)
Total no. of reflections	170,346	124,483	103,473
No. of unique reflections	47,068	42,679	42,180
Completeness[a] (%)	92.5 (83.9)	98.3 (96.9)	89.8 (93.1)
Redundancy	3.6	2.8	2.3
$R(I)_{sym}$[a,b] (%)	3.2 (17.2)	5.9 (49.1)	7.8 (31.6)
$I/\sigma(I)$[a]	33 (8.3)	12 (2.2)	24 (2.0)
B. Refinement			
R_{work}[c] / R_{free}[d] (%)	13.2/19.4	20.1/24.5	15.3/21.6
No. of atoms/residues (molecules)			
Protein	2801/362	2799/361	2790/364
Water	298	228	172
Glycerol (cryo buffer)	60/10	36/6	48/8
Ligand	15/1	13/1	13/1
Mean *B*–factors (Å^2)			
Protein	23.5	22.8	25.7
Water	34.0	28.6	32.9
Glycerol (cryo buffer)	54.8	48.5	53.0
Ligand	22.2	24.9	25.2
rmsd angle (°)	2.1	2.1	2.1
rmsd bond (Å)	0.010	0.009	0.008
(An)isotropic refinement	Anisotropic	Isotropic	Anisotropic

3.4.7 Acknowledgements

This research was supported by the NIH (Grant GM065489) and by the Deutsche Forschungsgemeinschaft (Grants KL1204/1-3, KL1204/9-1 and the Graduiertenkolleg "Protein Function at the Atomic Level"). We thank Christian Sohn for his support during X-ray data collection, Hans-Dieter Gerber for synthesising $preQ_1$ and Andreas Blum for helpful discussions.

3.5 References

[1] Tidten N, Stengl B, Heine A, Garcia GA, Klebe G, Reuter K. Glutamate versus glutamine exchange swaps substrate selectivity in tRNA-guanine transglycosylase: Insight into the regulation of substrate selectivity by kinetic and crystallographic studies. J Mol. Biol. 2007; 374 (3): 764-76.

[2] Stengl B, Reuter K, Klebe G. Mechanism and substrate specificity of tRNA-guanine transglycosylases (TGTs): tRNA-modifying enzymes from the three different kingdoms of life share a common catalytic mechanism. Chembiochem 2005;6 (11):1926-39.

[3] Iwata-Reuyl D. Biosynthesis of the 7-deazaguanosine hypermodified nucleosides of transfer RNA. Bioorg Chem 2003;31 (1):24-43.

[4] Watanabe M, Matsuo M, Tanaka S, Akimoto H, Asahi S, Nishimura S, Katze JR, Hashizume T, Crain PF, McCloskey JA, Okada N. Biosynthesis of archaeosine, a novel derivative of 7-deazaguanosine specific to archaeal tRNA, proceeds via a pathway involving base replacement on the tRNA polynucleotide chain. J Biol Chem 1997;272 (32):20146-51.

[5] Okada N, Nishimura S. Isolation and characterization of a guanine insertion enzyme, a specific tRNA transglycosylase, from Escherichia coli. J Biol Chem 1979;254 (8):3061-6.

[6] Nakanishi S, Ueda T, Hori H, Yamazaki N, Okada N, Watanabe K. A UGU sequence in the anticodon loop is a minimum requirement for recognition by Escherichia coli tRNA-guanine transglycosylase. J Biol Chem 1994;269 (51):32221-5.

[7] Curnow AW, Garcia GA. tRNA-guanine transglycosylase from Escherichia coli. Minimal tRNA structure and sequence requirements for recognition. J Biol Chem 1995;270 (29):17264-7.

[8] Kuchino Y, Kasai H, Nihei K, Nishimura S. Biosynthesis of the modified nucleoside Q in transfer RNA. Nucleic Acids Res 1976;3 (2):393-8.

[9] Reader JS, Metzgar D, Schimmel P, de Crecy-Lagard V. Identification of four genes necessary for biosynthesis of the modified nucleoside queuosine. J Biol Chem 2004;279 (8):6280-5.

[10] Gaur R, Varshney U. Genetic analysis identifies a function for the queC (ybaX) gene product at an initial step in the queuosine biosynthetic pathway in Escherichia coli. J Bacteriol 2005;187 (20):6893-901.

[11] Van Lanen SG, Reader JS, Swairjo MA, de Crecy-Lagard V, Lee B, Iwata-Reuyl D. From cyclohydrolase to oxidoreductase: discovery of nitrile reductase activity in a common fold. Proc Natl Acad Sci U S A 2005;102 (12):4264-9.

[12] Van Lanen SG, Kinzie SD, Matthieu S, Link T, Culp J, Iwata-Reuyl D. tRNA modification by S-adenosylmethionine:tRNA ribosyltransferase-isomerase. Assay development and characterization of the recombinant enzyme. J Biol Chem 2003;278 (12):10491-9.

[13] Mathews I, Schwarzenbacher R, McMullan D, Abdubek P, Ambing E, Axelrod H, Biorac T, Canaves JM, Chiu HJ, Deacon AM, DiDonato M, Elsliger MA, Godzik A, Grittini C, Grzechnik SK, Hale J, Hampton E, Han GW, Haugen J, Hornsby M, Jaroszewski L, Klock HE, Koesema E, Kreusch A, Kuhn P, Lesley SA, Levin I, Miller MD, Moy K, Nigoghossian E, Ouyang J, Paulsen J, Quijano K, Reyes R, Spraggon G, Stevens RC, van den Bedem H, Velasquez J, Vincent J, White A, Wolf G, Xu Q, Hodgson KO, Wooley J, Wilson IA. Crystal structure of S-adenosylmethionine:tRNA ribosyltransferase-isomerase (QueA) from Thermotoga maritima at 2.0 A resolution reveals a new fold. Proteins 2005;59 (4):869-74.

[14] Grimm C, Ficner R, Sgraja T, Haebel P, Klebe G, Reuter K. Crystal structure of Bacillus subtilis S-adenosylmethionine:tRNA ribosyltransferase-isomerase. Biochem Biophys Res Commun 2006;351 (3):695-701.

[15] Frey B, McCloskey J, Kersten W, Kersten H. New function of vitamin B12: cobamide-dependent reduction of epoxyqueuosine to queuosine in tRNAs of Escherichia coli and Salmonella typhimurium. J Bacteriol 1988;170 (5):2078-82.

[16] Goodenough-Lashua DM, Garcia GA. tRNA-guanine transglycosylase from E. coli: a ping-pong kinetic mechanism is consistent with nucleophilic catalysis. Bioorg Chem 2003;31 (4):331-44.

[17] Romier C, Reuter K, Suck D, Ficner R. Crystal structure of tRNA-guanine transglycosylase: RNA modification by base exchange. Embo J 1996;15 (11):2850-7.

[18] Xie W, Liu X, Huang RH. Chemical trapping and crystal structure of a catalytic tRNA guanine transglycosylase covalent intermediate. Nat Struct Biol 2003;10 (10):781-8.

[19] Brenk R, Stubbs MT, Heine A, Reuter K, Klebe G. Flexible adaptations in the structure of the tRNA-modifying enzyme tRNA-guanine transglycosylase and their implications for substrate selectivity, reaction mechanism and structure-based drug design. Chembiochem 2003;4 (10):1066-77.

[20] Brenk R, Naerum L, Grädler U, Gerber HD, Garcia GA, Reuter K, Stubbs MT, Klebe G. Virtual screening for submicromolar leads of tRNA-guanine transglycosylase based on a new unexpected binding mode detected by crystal structure analysis. J Med Chem 2003;46 (7):1133-43.

[21] Hoops GC, Townsend LB, Garcia GA. tRNA-guanine transglycosylase from Escherichia coli: structure-activity studies investigating the role of the aminomethyl substituent of the heterocyclic substrate PreQ1. Biochemistry 1995;34 (46):15381-7.

[22] Meyer EA, Donati N, Guillot M, Schweizer WB, Diederich F, Stengl B, Brenk R, Reuter K, Klebe G. Synthesis, biological evaluation, and crystallographic studies of extended guanine based (lin-Benzoguanine) inhibitors for tRNA-guanine transglycosylase (TGT). Helv. Chim. Acta 2006;89 (4):573-97.

[23] Brenk R, Meyer EA, Reuter K, Stubbs MT, Garcia GA, Diederich F, Klebe G. Crystallographic study of inhibitors of tRNA-guanine transglycosylase suggests a new structure-based pharmacophore for virtual screening. J Mol Biol 2004;338 (1):55-75.

[24] Ishitani R, Nureki O, Fukai S, Kijimoto T, Nameki N, Watanabe M, Kondo H, Sekine M, Okada N, Nishimura S, Yokoyama S. Crystal structure of archaeosine tRNA-guanine transglycosylase. J Mol Biol 2002;318 (3):665-77.

[25] Todorov KA, Garcia GA. Role of aspartate 143 in Escherichia coli tRNA-guanine transglycosylase: alteration of heterocyclic substrate specificity. Biochemistry 2006;45 (2):617-25.

[26] Reuter K, Ficner R. Sequence analysis and overexpression of the Zymomonas mobilis tgt gene encoding tRNA-guanine transglycosylase: purification and biochemical characterization of the enzyme. J Bacteriol 1995;177 (18):5284-8.

[27] Romier C, Ficner R, Reuter K, Suck D. Purification, crystallization, and preliminary x-ray diffraction studies of tRNA-guanine transglycosylase from Zymomonas mobilis. Proteins 1996;24 (4):516-9.

[28] Curnow AW, Kung FL, Koch KA, Garcia GA. tRNA-guanine transglycosylase from Escherichia coli: gross tRNA structural requirements for recognition. Biochemistry 1993;32 (19):5239-46.

[29] Leatherbarrow R. GraFit 4.09 edit. Erithacus Software Limited, USA 1999.

[30] Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. 1997;276:307-26.

[31] Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL. Crystallography and NMR system: A new software suite for macromolecular structure determination. Acta Crystallogr. sect. D 1998;54:905-21.

[32] Sheldrick GM, Schneider TR. SHELXL: high-resolution refinement. Methods Enzymol. 1997;277b:319-43.

[33] Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallogr. sect. D 2004;60:2126-32.

[34] Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Crystallogr. 1993;26:283-91.

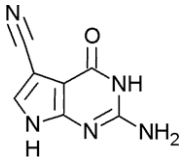
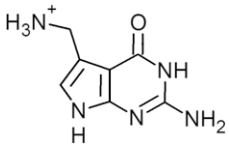
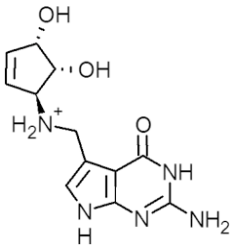
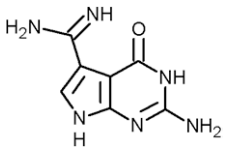
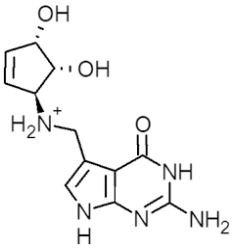
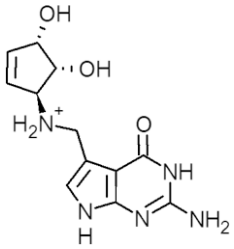
4 Characterisation of a model system for the human TGT binding pocket

4.1 Introduction

Transfer RNA-guanine transglycosylases (TGTs, EC 2.4.2.29), evolutionary ancient enzymes, are involved in the hypermodification of tRNAs [1]. TGTs catalyse a base exchange reaction, excorporating a certain guanine base within specific tRNAs and introducing 7-substituted-7-deazaguanines [1,2]. The enzymes are found in all three domains of life. Although distinct bases are incorporated into tRNA in a domain-specific manner, the catalytic subunits of TGTs are structurally well conserved [2].

The molecular differences of archaeal and bacterial TGTs have been discussed in chapter 3. As described, archaeal TGTs exchange a guanine base at position 15 within the D-arm of tRNAs and are involved in a hypermodification pathway, finally leading to archaeosine-tRNAs [3] (Table 8). In contrast, in bacteria and eukaryotes, guanine^{34} of the tRNA anticodon loop is exchanged and the final hypermodification product within the tRNAs is the base queuine (Table 8). Nevertheless, bacterial and eukaryotic TGTs still differ in their *in vivo* substrate bases (Table 8). Bacterial TGTs incorporate a relatively small modified base, namely preQ$_1$, which is afterwards further modified to the functional queuine within the tRNA by S-adenosylmethionine:tRNA ribosyltransferase-isomerase (QueA) [4-6] and an unknown coenzyme B$_{12}$-dependent enzyme [7]. In contrast, eukaryotic TGTs directly incorporate queuine into their tRNA substrates [8]. In bacteria and eukaryotes, most likely, the translational properties of the tRNA substrates are influenced by these modifications [9,10].

Table 8: Three domains of life: Final tRNA modifications and substrate bases of TGTs.

TGT	Archaeal	Bacterial	Eukaryotic
Exchanged base	G^{15}	G^{34}	G^{34}
Substrate base			
	$preQ_0$	$preQ_1$	queuine
Final tRNA modification			
	archaeosine	queuine	queuine

An evolutionary highly conserved mechanism is assumed to originate from an ancient ancestor that existed even before the separation of the three domains [2]. Bacterial TGTs have been well characterised by extensive kinetic studies and crystal structure analyses of *Zymomonas mobilis* TGT [11–16]. Four specific tRNA substrates are known: $tRNA^{Tyr,\ Asp,\ Asn,\ His}$, sharing a common $U^{33}G^{34}U^{35}$ sequence [17,18]. The TGT catalysed reaction follows a ping-pong mechanism [19]. First, the tRNA substrate binds to the enzyme and Asp^{280} (*Z. mobilis* numbering) performs a nucleophilic attack, cleaving the *N*-glycosidic bond of $guanine^{34}$ at the wobble position of the anticodon loop and producing a covalent TGT · tRNA intermediate complex [20]. Subsequently, the modified base $preQ_1$ [21] (Table 8) replaces guanine within the binding pocket. In a reverse reaction step, it is incorporated into tRNA (Figure 9, chapter 3).

Bacterial TGTs are homodimers of approximately 43 kDa per monomer [22]. In contrast, mammalian TGTs appear to be composed of a heterodimer [23]

consisting of one approximately 60 kDa subunit, and one smaller, presumably the catalytic subunit. In *Homo sapiens*, this 44 kDa subunit shows high homology to bacterial TGTs (43 % sequence identity between *H. sapiens* and *Z. mobilis* [24]). In addition, a homology model based on *Z. mobilis* TGT and the *Caenorhabditis elegans* sequence (38.4 % sequence identity) indicate a very high structural similarity of the active sites of bacterial and eukaryotic TGTs (five of seven active site residues are strictly conserved) [25].

Mutation of the *tgt* gene in *Shigella flexneri* results in a significant loss of pathogenicity of the bacterium [26]. *Shigellae* are the causative agents of bacterial dysentery and effect some 600,000 infant deaths per year, mainly in developing countries with sub-standard hygiene and water supplies. Extensive studies of *S. flexneri* have resulted in a detailed understanding of the pathogenesis, involving bacterial invasion into colonic epithelial cells followed by intracellular multiplication and spreading into adjacent cells [27]. The virulence was shown to be multifactorial and is under control of various genes encoded by the chromosome and the large virulence plasmid [28]. Mutation of the *tgt* gene causes a reduced synthesis of the VirF protein, a key regulatory protein for the full expression of virulence. This effect is a consequence of the less efficient translation of the *virF* mRNA, presumably due to the absence of modified tRNA molecules [26] or a putative riboswitch-like function of the *virF* mRNA itself [29]. TGT has been suggested to be able to modify the *virF* mRNA as well, thereby influencing its translation competent conformation [29]. In any case, TGT is essentially required by the bacterium to produce the virulence factor VirF and its pathogenic phenotype. Therefore, TGT has been established as a target for structure-based drug design [30].

Queuine modification is thought to affect the speed and accuracy of the translational process and was reported to exhibit pleiotropic effects on cellular metabolism [1]. Hence, to be able to specifically inhibit the bacterial

target while having unaffected the human counterpart, it is of utmost importance to study selectivity-determining features.

To date, no crystal structure of an eukaryotic, in particular of human TGT (hTGT) is yet available, although multiple sequences of eukaryotic TGTs resulting from several genome projects [24,31,32] have been deposited during the last years. So far, not even the heterologous production of an active eukaryotic TGT could be achieved. Rather, the human TGT catalytic subunit was reported to be toxic to *E. coli* rendering the isolation of an active catalytic subunit of mammalian TGT impossible [24].

Hitherto, the *C. elegans* TGT homology model is the only source for a comparison of eukaryotic and bacterial TGT active sites in structural terms [25]. Romier *et al.* discussed that most presumably a spatial expansion of the binding pocket in eukaryotic TGTs enables the binding of extended $preQ_1$-like substrates such as queuine. This enlargement of the binding pocket was discussed to be due to an eukaryote-specific $Val^{233}Gly$ (*Z. mobilis* numbering) replacement. A further $Cys^{158}Val$ replacement was suggested to facilitate additional van der Waals contacts to the cyclopentenyl moiety of queuine [25].

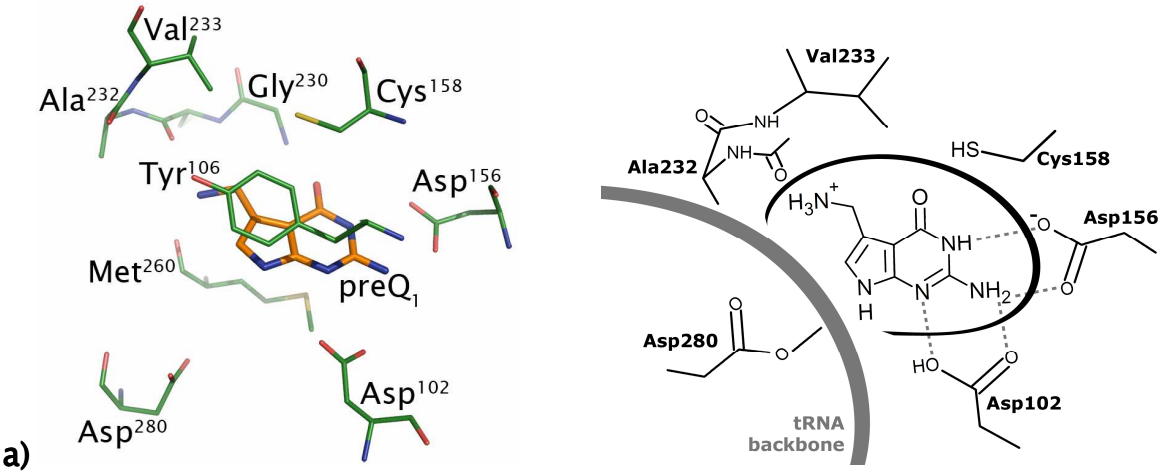
The aim of this project is to study the differences between bacterial TGT as a drug design target and human TGT as the host's counterpart. In this study, based on *Z. mobilis* TGT a genetically engineered model system for human TGT was designed, to probe the properties of such an extended binding pocket.

4.2 Results

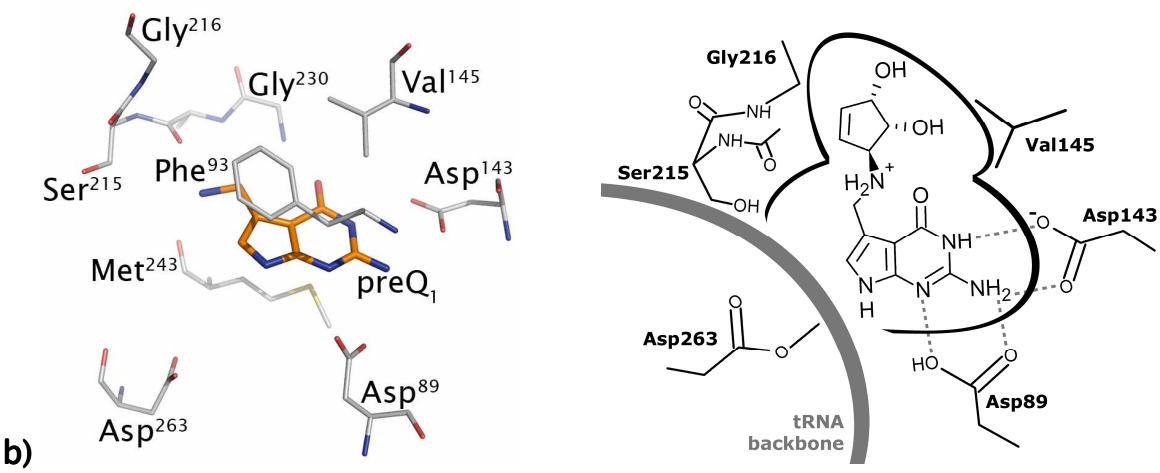
4.2.1 Sequence comparisons and homology modelling

Until now, structural information of eukaryotic TGTs was only gained from a homology model of the *C. elegans* enzyme [25]. However, in recent years, a large number of eukaryotic TGT sequences have been deposited [24,31,32]. Since the human TGT sequence is now available [24], a homology model was constructed for human TGT based on *Z. mobilis* TGT crystal structures. The sequences were aligned utilising Clustal W 1.83 [33] (chapter 5.2.1, Appendix). Based on this alignment, ten homology models for human TGT were generated with MODELLER 6a [34,35]. The best scored model is shown in Figure 12b. It can be clearly seen that the residues inside the binding pocket are in general highly conserved among bacterial and eukaryotic TGTs (Figure 12a). As in the *C. elegans* TGT homology model [25], also in the human TGT model, a Val^{233} (*Z. mobilis* numbering)/ Gly^{216} (human TGT numbering) exchange can be identified as the crucial difference between the human and bacterial TGT binding pockets. This leads to an enlarged binding pocket which is required for queuine binding. The replacement of Cys^{158}/Val^{145} is observed as well (Figure 12a, b). Thus, as expected, considerations of Romier *et al.* [25] for the *C. elegans* model are also found for the human TGT homology model. It has already been reported that the Val^{233}/Gly^{216} and Cys^{158}/Val^{145} replacements are highly conserved among the bacterial and eukaryotic domains [25], respectively. Taking into account the newly deposited sequences in a multiple sequence alignment of 100 bacterial, and 100 eukaryotic TGT sequences, Val^{233} is found to be strictly conserved among bacteria, while Gly^{216} is to 79 % present in eukaryotes. The conservation of Cys^{158} (99 %) in bacterial, and also of Val^{145} (59 %) in

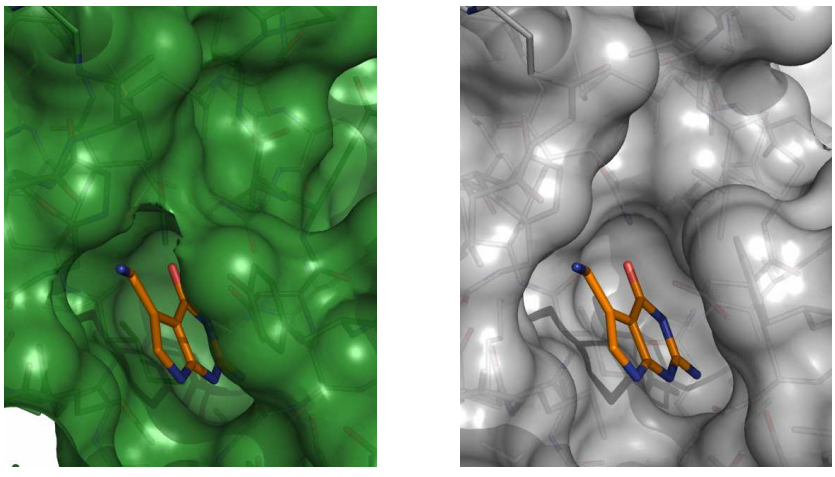
Figure 12: Homology modeling.
a) *Z. mobilis* TGT crystal structure 1ozq (green) in complex with the substrate $preQ_1$ (orange), showing amino acids of the active site, and a schematic representation of the active site. **b)** Homology model for human TGT (grey) superimposed with the substrate $preQ_1$ (orange) from 1ozq, showing amino acids of the active site, and a schematic representation of the active site. **c)** Comparison of the active sites' solvent accessible surfaces of *Z. mobilis* TGT (green) and homology model for human TGT (grey).



a)



b)



c)

eukaryotic TGTs is sufficiently high to consider these two sites as characteristic for the respective domains (chapter 5.2.2, Appendix). Both sites were therefore chosen for a mutational study, in order to generate a model for the human TGT binding pocket based on the well established *Z. mobilis* TGT system. The mutated bacterial TGT was expected to accommodate queuine and to give new insights into the substrate recognition properties of human TGT.

4.2.2 Construction and enzymatic characterisation of *Z. mobilis* TGT variants

Z. mobilis TGT, serving as the basis for this mutational study, exhibits one residue within the active site, that is unique among bacteria: Tyr^{106}. Usually, at the corresponding position a phenylalanine is found in nearly all bacterial TGTs. With 99 % it is very highly conserved. Brenk *et al.* had already shown, that a $Tyr^{106}Phe$ TGT variant exhibits no significant changes of the enzymatic properties [14]. In eukaryotic TGTs a phenylalanine at the corresponding position (Phe^{93} in human TGT) is as well conserved to 65 %. Thus, to generate a model system as authentic as possbile, the *Z. mobilis tgt* gene, already carrying the $Tyr^{106}Phe$ mutation, was chosen as template for site-directed mutagenesis. Since the $Tyr^{106}Phe$ replacement does not affect the enzyme's properties [14], in the following, this mutation as it is considered in all studies will not be mentioned explicitly.

To study the significance of residues Cys^{158} (Val^{145} in human TGT) and Val^{233} (Gly^{216} in human TGT) with respect to substrate specificity, a $Cys^{158}Val$ variant (named TGT($C^{158}V$) in the following), a $Val^{233}Gly$ variant (named TGT($V^{233}G$) in the following), and a $Cys^{158}Val/Val^{233}Gly$ variant (named TGT($C^{158}V/V^{233}G$) in the following) were created by site-directed mutagenesis of the *tgt*-Y106F template. All different variants could be produced and purified according to standard protocols [12,36].

Subsequently, to be able to compare the enzymatic properties of the variants with the wild-type enzyme, kinetic measurements were performed by means of a well established assay detecting the incorporation of [8-^3H]-guanine in exchange with the native guanine[34] from the tRNA substrate. All variants were active at the level of detection. While the K_M values for guanine and tRNA remained almost unchanged within the experimental error (Table 9, p. 77), k_{cat} was significantly decreased in all variants compared to the wild-type enzyme. In case of the TGT(C^{158}V) variant, the turnover number was decreased by factors of 3-5, while in case of the TGT(V^{233}G) variant it was decreased by a factor of about one order of magnitude, and by almost two orders of magnitude for TGT(C^{158}V/V^{233}G).

Surprisingly, all exchanged residues, which are fairly distant from the catalytic residue Asp^{280} (Figure 12, p. 74), seem to take a dramatic influence on the catalytic activity, even though affinity towards the substrates remains almost unchanged.

A further comparison of the human and bacterial TGT binding pockets (Figure 12, p. 74) reveals a third residue, which is as well quite proximate to the substrate and differs in the two TGTs: Ala^{232} (*Z. mobilis* numbering) is indeed not strictly conserved (63 %) among bacteria. In about 37 % of the remaining cases Ser^{232} is found which corresponds to Ser^{216} in the human TGT. In eukaryotes this residue is conserved to 51 %. Thus, to exmanine, whether this residue is required by an enlarged binding pocket for efficient enzyme activity, two additional variants were generated: TGT(A^{232}S/V^{233}G), and TGT(C^{158}V/A^{232}S/V^{233}G). Kinetic characterisation of the newly generated variants revealed, however, that the Ala^{232}Ser exchange has no significant influence on catalytic activity (Table 9).

Next, incorporation of preQ$_1$ and queuine into tRNA was analysed. Currently, no radio-labelled preQ$_1$ and queuine are commercially available. Thus, a 'reverse assay strategy' as already described previously [16,37] was

performed. There, the decrease of radioactivity in [8-^3H]-guanine^{34}-tRNA is monitored in consequence of the replacement of the labeled guanine by the respective substrate.

Table 9: Kinetic charcteriation of the TGT variants.

wild-type TGT	tRNA^{Tyr} *	[8-^3H]-guanine	preQ$_1$
K_M [µM]	0.9 ± 0.20	1.2 ± 0.20	0.9 ± 0.10
k_{cat} [10^{-3} s^{-1}] **	53.6 ± 0.10	55.8 ± 0.04	106.0 ± 2.72
k_{cat}/K_M [10^{-3} μM^{-1} s^{-1}]	59.6	46.5	117.8

TGT(C^{158}V)	tRNA^{Tyr} *	[8-^3H]-guanine	preQ$_1$
K_M [µM]	2.1 ± 0.15	1.4 ± 0.15	16.0 ± 2.35
k_{cat} [10^{-3} s^{-1}] **	10.2 ± 0.24	17.6 ± 0.20	35.0 ± 0.82
k_{cat}/K_M [10^{-3} $\mu M^{-1}s^{-1}$]	4.9	12.6	2.2

TGT(V^{233}G)	tRNA^{Tyr} *	[8-^3H]-guanine	preQ$_1$
K_M [µM]	2.0 ± 0.32	1.0 ± 0.22	2.1 ± 0.33
k_{cat} [10^{-3} s^{-1}] **	5.4 ± 0.28	4.6 ± 0.24	32.6 ± 1.12
k_{cat}/K_M [10^{-3} $\mu M^{-1}s^{-1}$]	2.7	4.6	15.5

TGT(A^{232}S/V^{233}G)	tRNA^{Tyr} *	[8-^3H]-guanine	preQ$_1$
K_M [µM]	2.0 ± 0.24	1.7 ± 0.11	n.d.
k_{cat} [10^{-3} s^{-1}] **	2.2 ± 0.08	5.6 ± 0.10	n.d.
k_{cat}/K_M [10^{-3} $\mu M^{-1}s^{-1}$]	1.1	3.3	-

TGT(C^{158}V/V^{233}G)	tRNA^{Tyr} *	[8-^3H]-guanine	preQ$_1$
K_M [µM]	1.4 ± 0.44	2.7 ± 0.32	n.d.
k_{cat} [10^{-3} s^{-1}] **	0.8 ± 0.08	0.6 ± 0.01	n.d.
k_{cat}/K_M [10^{-3} $\mu M^{-1}s^{-1}$]	0.6	0.2	-

TGT(C^{158}V/A^{232}S/V^{233}G)	tRNA^{Tyr} *	[8-^3H]-guanine	preQ$_1$
K_M [µM]	0.3 ± 0.04	4.2 ± 0.76	n.d.
k_{cat} [10^{-3} s^{-1}] **	0.2 ± 0.01	0.4 ± 0.01	n.d.
k_{cat}/K_M [10^{-3} $\mu M^{-1}s^{-1}$]	0.7	0.1	-

Wild-type TGT and variants characterised with substrates tRNA^{Tyr}, [8-^3H]-guanine, and [8-^3H]-guanine^{34}-tRNA^{Tyr}, respectively. * 2nd substrate: [8-^3H]-guanine, ** taking into account the presence of TGT as a homodimer able to bind only one tRNA molecule at a time [38].

K_M values for preQ$_1$ could only be determined for the wild-type enzyme and the variants TGT(V^{233}G) and TGT(C^{158}V). Data from literature [16] are perfectly reproduced for the wild-type enzyme by finding a K_M(preQ$_1$) of

0.9 µM, which approximately corresponds to the K_M(guanine) value, and a k_{cat} twice as high compared to the [8-^3H]-guanine incorporation. In case of TGT(C^{158}V), the K_M value for preQ$_1$ increases by a factor of almost 20 with respect to the wild-type TGT, although the Cys^{158}Val replacement hardly influences the K_M(guanine) value. The k_{cat} value is slightly decreased by a factor of 3 in comparison to the wild-type enzyme, which is analogously observed in the guanine measurements. K_M(preQ$_1$) for TGT(V^{233}G) does not significantly increase. Hence, the Val^{233}Gly replacement seemed to have no dramatic influence on substrate affinities. Furthermore, the k_{cat}(preQ$_1$) value is only slightly decreased by a factor of 3 with respect to the wild-type enzyme.

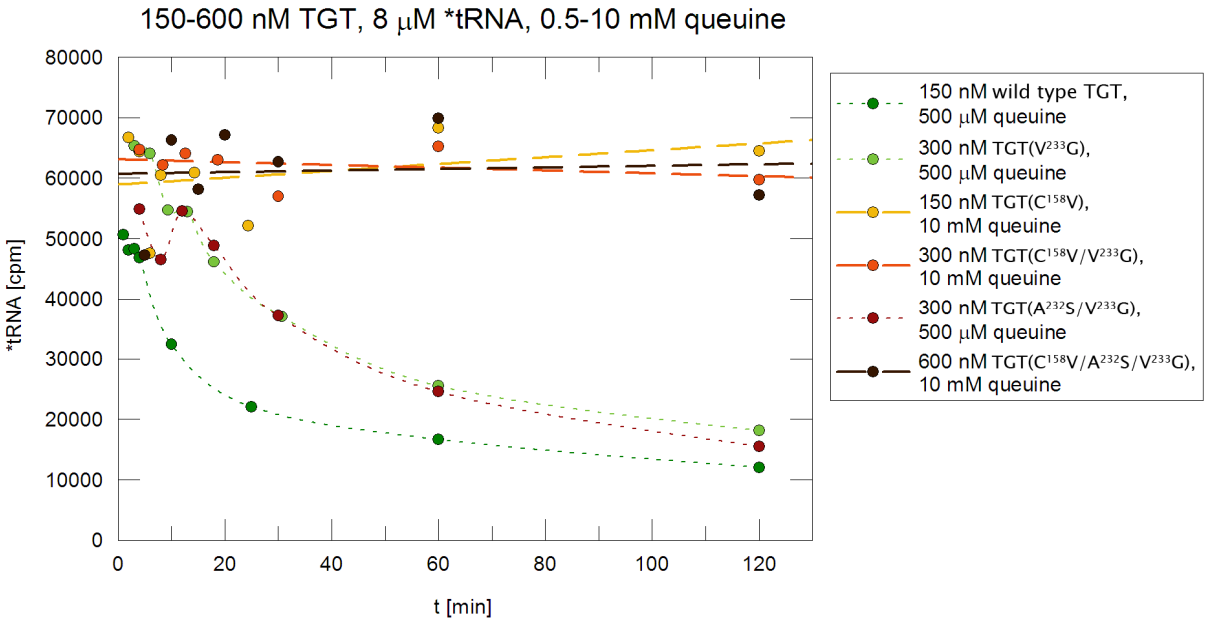
The reverse 'washout assay', where [8-^3H]-guanine is excorporated from tRNA, implicates a considerably lower signal to noise ratio compared to the basic assay, where the incorporation of [8-^3H]-guanine into tRNA is monitored. In combination with the dramatically reduced turnover number of the TGT(C^{158}V/(A^{232}S)/V^{233}G) variants, the decrease of radioactivity per time intervall is not sufficient for an accurate determination of Michaelis-Menten parameters. The amount of enzyme used in the assay cannot be further elevated to obtain a better signal, since an excess of substrate has to be assured.

In case of queuine, for the variants TGT(V^{233}G), and TGT(A^{232}S/V^{233}G), only very high queuine concentrations (30 - 500 µM) lead to a detectable decrease of the radioactive signal due to the excorporation of [8-^3H]-guanine. However, even at a concentration of 10 mM queuine, the remaining variants TGT(C^{158}V), TGT(C^{158}V/V^{233}G) and TGT(C^{158}V/A^{232}S/V^{233}G) are not able to excoporate [8-^3H]-guanine (Figure 13).

Suprisingly, in the control reaction, for the wild-type enzyme an even faster decrease of the radioactive signal is observed, indicating, that the wild-type enzyme is obviously able to accept queuine as a substrate. In conclusion, all

variants with the Cys[158]Val exchange seem to be unable to excorporate [8-^3H]-guanine in the presence of queuine as substrate.

Figure 13: Excorporation of [8-^3H]-guanine from labelled tRNA^{Tyr} in the presence of queuine.

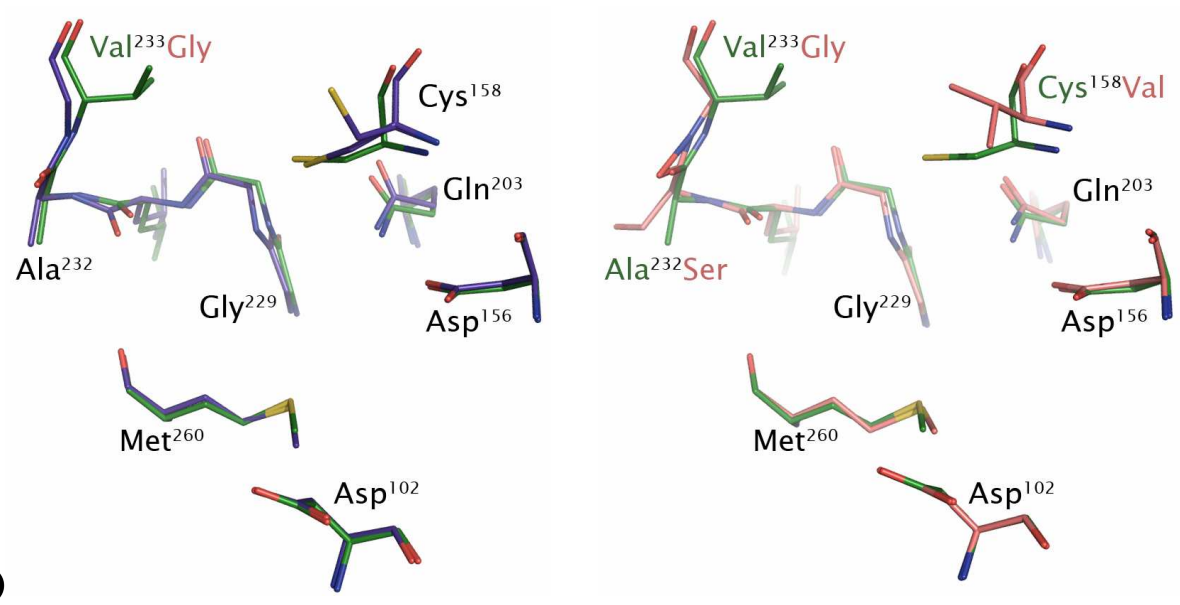


4.2.3 Crystal structures

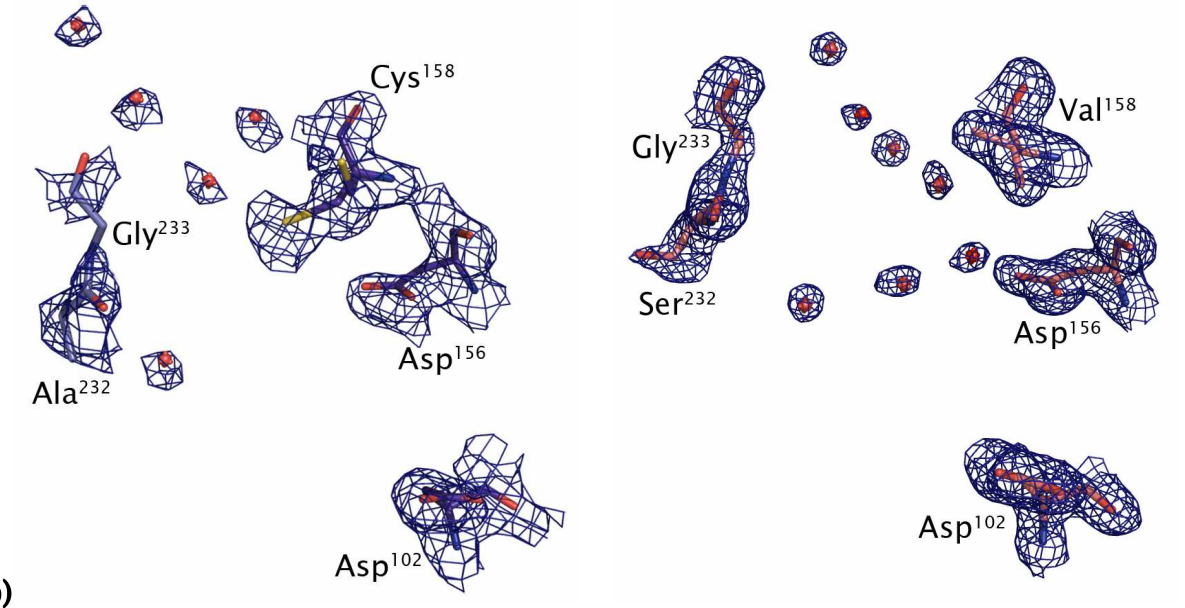
To study the constructed human TGT binding pocket model in structural terms, two variants were crystallised. The TGT(V[233]G) variant, representing the assumed minimal requirement to generate an enlarged binding pocket for queuine accommodation, as well as the TGT(C[158]V/A[232]S/V[233]G) variant, representing the binding pocket, probably most similar to the human TGT binding pocket, were selected for this purpose. Besides the uncomplexed proteins, also preQ$_1$- and queuine-complexed structures were studied for these variants and will be described in the following.

The superposition of the apo-structures of TGT(V[233]G) and TGT(C[158]V/A[232]S/V[233]G) with wild-type TGT (Figure 14a) shows, as expected, that the backbones adopt identical conformations. In addition, there is a dramatic

Figure 14: Uncomplexed (apo) TGT(V^{233}G) and TGT(C^{158}V/A^{232}S/V^{233}G) crystal structures.
a) TGT(V^{233}G) (blue, 2.25 Å resolution, left), and TGT(C^{158}V/A^{232}S/V^{233}G) (red, 1.60 Å resolution, right) apo structures, superimposed with wild-type TGT (green, PDB code 1pud).
b) $2|F_o|-|F_c|$ density contoured at a σ level of 1.5 is coloured blue. TGT(V^{233}G) (blue, left), and TGT(C^{158}V/A^{232}S/V^{233}G) (red, right) binding pockets, respectively.



a)



b)

gain of space inside the binding pocket due to the amino acid replacements, especially of Val^{233}Gly. The generated space is occupied by well-defined water molecules (Figure 14b).

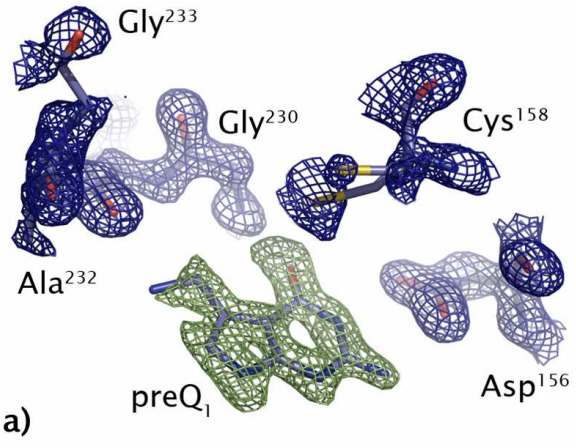
In case of the TGT(V^{233}G) variant, the exchanged residue Val^{233}Gly is only poorly defined. For the Cys^{158} side chain a double conformation is observed. In general, the *B*-factors in the region next to this residue are relatively high in all deposited TGT structures (~ 55 Å^2), compared to other active site residues. In addition, there are two other TGT variant structures, exhibiting a double conformation of the Cys^{158} side chain: 1pxg (a Asp^{280}Glu variant in complex with preQ$_1$, and 1ozq (a Tyr^{106}Phe variant in complex with preQ$_1$).

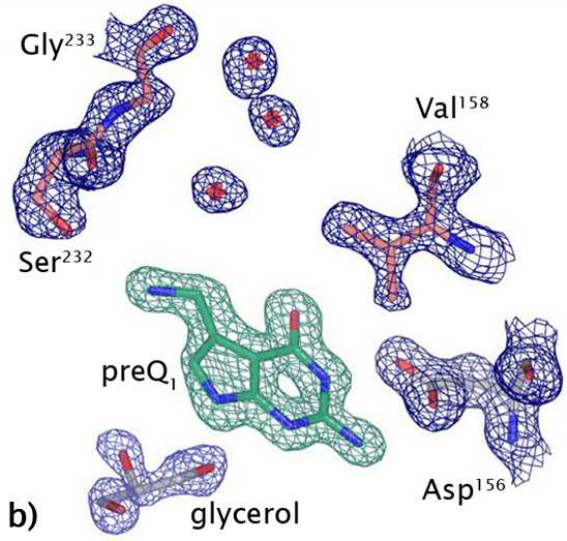
Figure 15: preQ$_1$ bound TGT(V^{233}G) and TGT(C^{158}V/A^{232}S/V^{233}G) crystal structures.

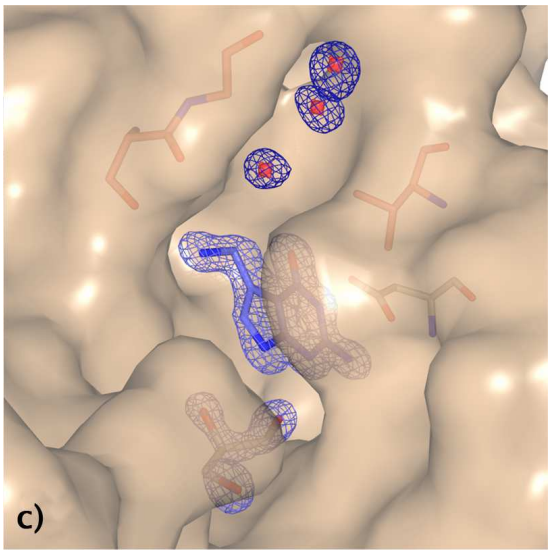
a) TGT(V^{233}G) is represented in blue, 2|F$_o$|-|F$_c$| density contoured at a σ level of 1.5 is coloured blue, |F$_o$|-|F$_c$| density contoured at σ a level of 3 is coloured green (1.19 Å resolution).

b) TGT(C^{158}V/A^{232}S/V^{233}G) is represented in red. 2|F$_o$|-|F$_c$| density contoured at a σ level of 1.5 is coloured blue, |F$_o$|-|F$_c$| density contoured at σ a level of 3 is coloured green (1.46 Å resolution).

c) Solvent accessible surface representation of the TGT(C^{158}V/A^{232}S/V^{233}G) binding pocket, showing electron densities of the ligand and water molecules.







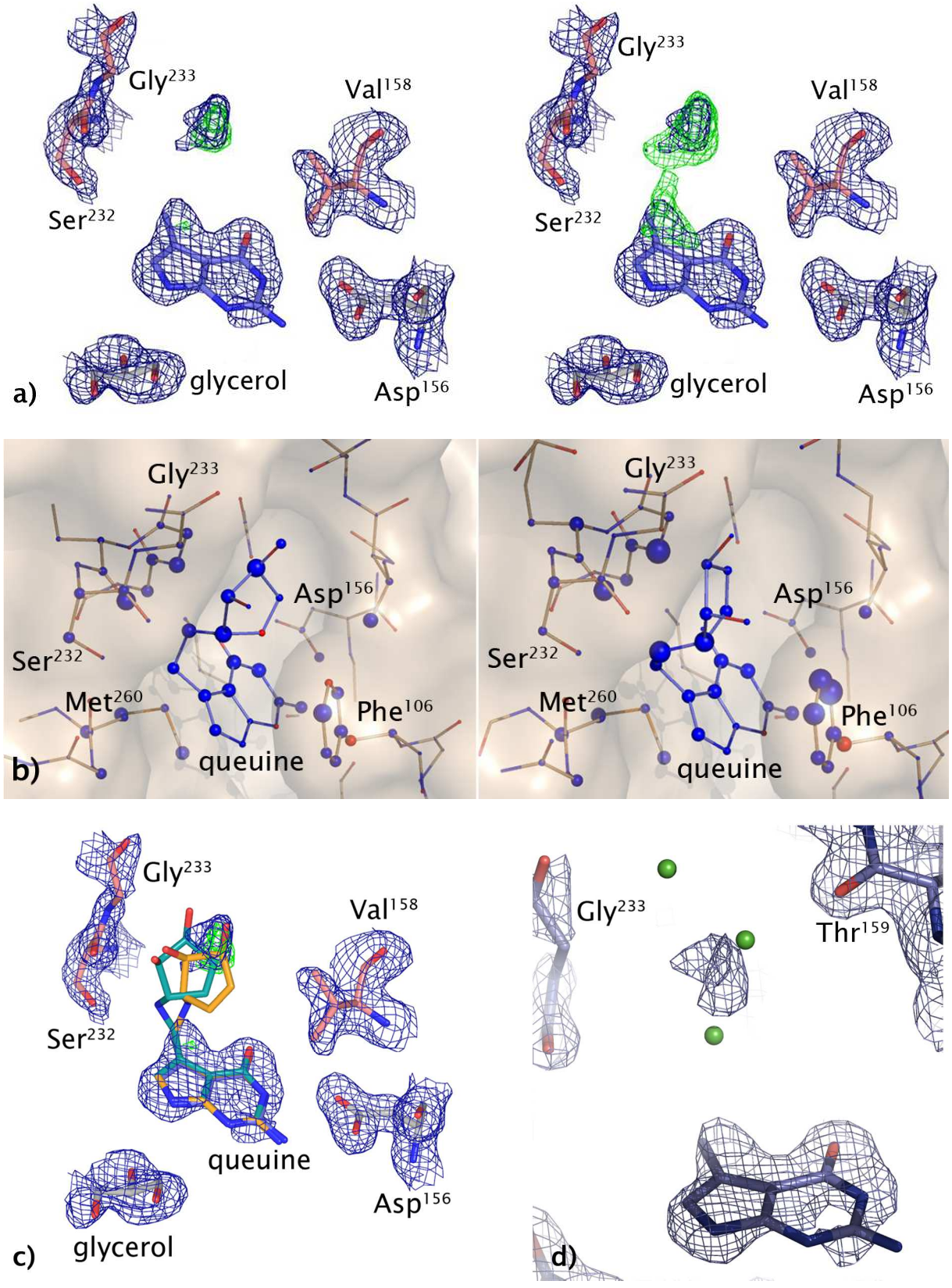
Nevertheless, in case of the TGT($C^{158}V/A^{232}S/V^{233}G$) variant, all exchanged residues show well-defined electron density (Figure 14b).

The $preQ_1$ complex structures of the two variants show, as expected, an analogous binding mode of the ligand (Figure 15). The protein backbone superimposes almost perfectly. However, the high *B*-factors of $preQ_1$ in the TGT($V^{233}G$) variant binding pocket indicate, that the ligand is probably not fully occupied. Again, the Gly^{233} density is relatively weak and the Cys^{158} side chain was refined with a double conformation, although the splitting is only weakly indicated. In case of the TGT($C^{158}V/A^{232}S/V^{233}G$) variant, all exchanged residues show well defined electron density (Figure 15b). The newly created subpocket is occupied by well-defined water molecules (Figure 15c), most of them at similar positions compared to the apo structures.

In case of queuine, co-crystallised with the TGT($V^{233}G$) variant, partial density corresponding to the purine heterocycle could be assigned. However, attempts to refine the position of the heterocycle did not reveal conclusive results. Obviously, the ligand shows only minor population or extended disorder in the binding pocket making any refinement of its posistion unreasonable.

Figure 16: Crystal structure of TGT($C^{158}V/A^{232}S/V^{233}G$) co-crystallised with queuine.

a) $2|F_o|-|F_c|$ density contoured at a σ level of 1.5 is coloured blue, $|F_o|-|F_c|$ density contoured at σ levels of 3 (left), and 2 (right), respectively, is coloured green (1.60 Å resolution).
b) Queuine, docked into the crystal structure with GOLD [39,40]. Two of the highest scored ($DrugScore^{CSD}$ [41]) docking poses are shown. Favourable per atom contributions to the overall score are indicated by blue spheres, while the size of the sphere corresponds to the respective contribution, disfavourable per atom contributions are indicated by red spheres.
c) Superposition of the two docking poses of queuine with the crystal structure of TGT($C^{158}V/A^{232}S/V^{233}G$) in complex with queuine.
d) Superposition of three active site bound water molecules from the $preQ_1$·TGT($C^{158}V/A^{232}S/V^{233}G$) crystal structure (represented as green spheres) with electron density of the queuine·TGT($C^{158}V/A^{232}S/V^{233}G$) crystal structure.


Gly233
Val158
Ser232
glycerol
Asp156
a)
Gly233
Val158
Ser232
glycerol
Asp156
Gly233
Asp156
Ser232
Met260
Phe106
queuine
b)
Gly233
Asp156
Ser232
Met260
Phe106
queuine
Gly233
Val158
Ser232
queuine
Asp156
glycerol
c)
Gly233
Thr159
d)

In case of the queuine·TGT(C^{158}V/A^{232}S/V^{233}G) variant complex structure, the electron density of the purine heterocycle is well defined. For the cyclopentenyl moiety only poorly defined electron density can be observed. Comparing the traces of the electron density and its spacial location with respect to the apo and preQ$_1$ structures, it appears obvious that the remaining rather large difference electron density peak does not originate from some bound water molecules, present in the apo or preQ$_1$ structures (Figure 16d). Apparently, the enlarged binding pocket does not offer appropriate interaction partners for the hydroxyl groups of the cyclopentenyl moiety in *H*-bonding distance. Therefore, it is assumed that the ring adopts multiple orientations resulting in a pronounced disorder. Docking queuine into this structure using GOLD [39,40] reveals several binding orientations of the cyclopentenyl ring. None of them establishes strong interactions to the protein (Figure 16b). The two best scored docking solutions overlap in an area that nicely coincides with the observed difference electron density peak. The latter peak could not be assigned to any of the previously detected water molecules in the uncomplexed or preQ$_1$ bound cases (Figure 16d).

4.2.4 Molecular dynamics simulation

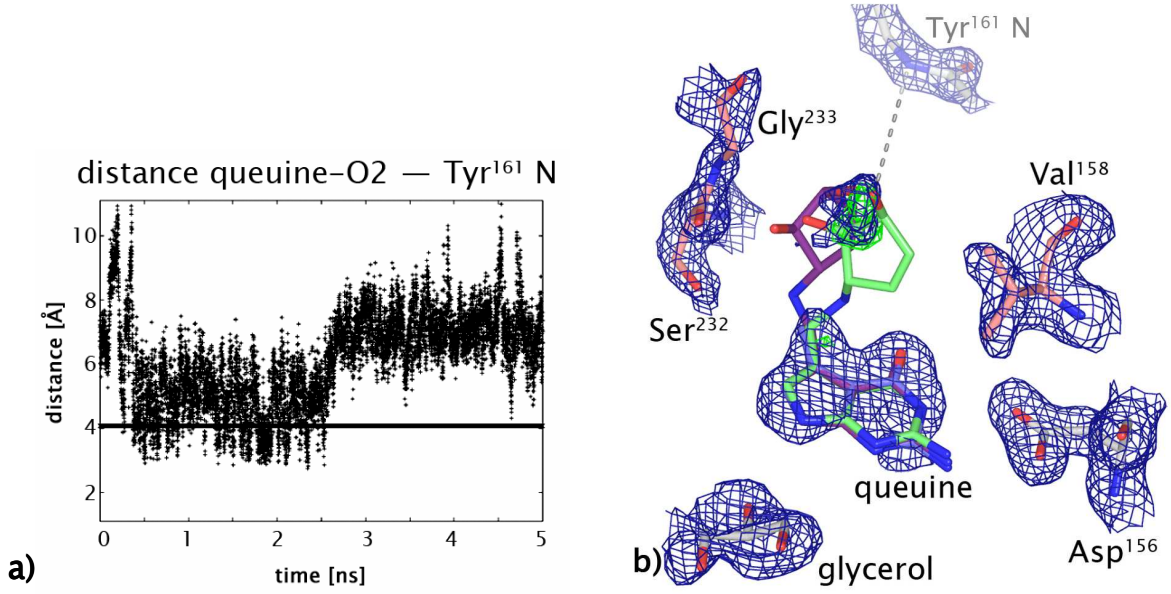
To probe a possible distribution of the cyclopentenyl moiety over several orientations in absence of appropriate interaction partners for directional interactions, molecular dynamics (MD) simulations of the complex were performed. As starting geometry, the TGT(C^{158}V/A^{232}S/V^{233}G) variant structure complexed by queuine was used. To generate a reasonable pose for queuine, the ligand was docked into the structure with GOLD [39,40] and minimised with the MAB force field in MOLOC [42]. Analysis of the MD trajectory revealed that, in fact, the cyclopentenyl moiety shows high residual mobility inside the binding pocket (Figure 17a). Two snapshots were extracted from the simulation, and superimposed based on a C$_\alpha$ fit of all

protein residues with the TGT(C^{158}V/A^{232}S/V^{233}G) crystal structure (Figure 17b).

Figure 17: MD simulation of TGT(C^{158}V/A^{232}S/V^{233}G) in complex with queuine.

a) Distance between an atom of the flexible cyclopentenyl moiety and a backbone amide nitrogen from a residue, deep inside the new subpocket, measured over the MD simulation time.

b) Two queuine conformers from MD simulation snapshots superimposed with the TGT(C^{158}V/A^{232}S/V^{233}G) crystal structure and the electron density.



Very similar to the previously described docking study, the region of the remaining difference electron density that could not be assigned to water molecules, fits nicely with the central part of the cyclopentenyl moiety considering two different orientations of queuine during the MD simulation. These results support the hypothesis that queuine is bound to the TGT(C^{158}V/A^{232}S/V^{233}G) variant. *Via* its central part, the obviously highly flexible cyclopentenyl moiety creates the difference electron density of the ligand in this region.

Additionally, also an MD simulation of the uncomplexed wild-type enzyme was performed. Interestingly, also the region between the crucial residues (Val^{233}, Cys^{158}) exhibits pronounced flexibility (Figure 18). It possibly creates a small subpocket, similar to the one opened in the TGT(C^{158}V/A^{232}S/V^{233}G) variant. Moreover, volumes of the different binding pockets were calculated

Figure 18: MD simulation of apo wild-type TGT.
Distance between Val^{233} Cα and Cys^{158} Sγ, measured over the MD simulation time. The green vertical lines indicate snapshots, that have been extracted for the calculations of binding pocket volumes, listed in Table 10.

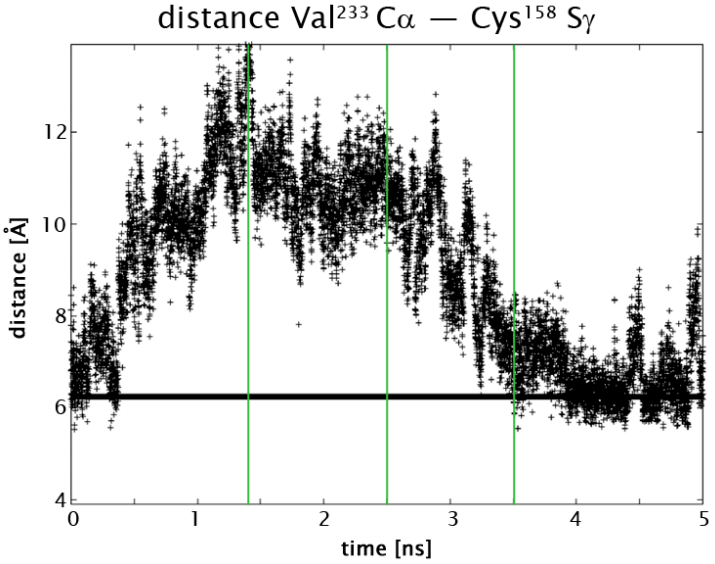


Table 10: Binding pocket volumes calculated with CASTp [43].

TGT	V [$10^2 \cdot Å^3$]
Wild-type TGT·preQ$_1$ (1p0e)	2.5
TGT(C^{158}V/A^{232}S/V^{233}G)·queuine	6.5
hTGT homology model	6.0
Wild-type TGT (apo), MD snapshot 1.4 ns	7.5
Wild-type TGT (apo), MD snapshot 2.5 ns	7.1
Wild-type TGT (apo), MD snapshot 3.5 ns	2.5

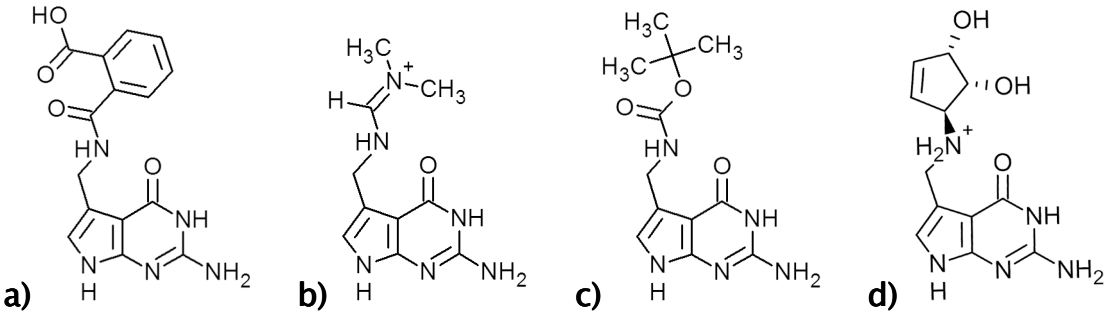
with CASTp [43] as listed in Table 10. The binding pockets of both, the TGT(C^{158}V/A^{232}S/V^{233}G) as well as the hTGT homology model, were found to possess more than twice as large binding pockets, compared to wild-type bacterial TGT. In light of these values it is remarkable to note that the binding pocket volume of the wild-type bacterial TGT can significantly open up as indicated by some frames during the MD simulation due to the mentioned flexibility in the region between the residues Val^{233} and Cys^{158}. The binding pocket then reaches volumes even larger than the eukaryotic

TGT binding pocket models, large enough for queuine accomodation (Table 10). After several nanoseconds, the original state, where the subpocket is closed again, is adopted, as indicated by the small volume of the binding pocket of the third MD snapshot (Figure 18, Table 10). Therefore, it even might be possible that also the wild-type enzyme is able to bind queuine. Interestingly in this context, wild-type TGT crystals are destroyed under queuine soaking conditions. In case of TGT, such findings often indicate that the soaked ligand is able to bind to the active site but requires structural adaptation of the enzyme destroying crystal contacts. In agreement with this, wild-type TGT was found to be able to excorporate labeled guanine from tRNA in the presence of queuine (Figure 13, p. 79).

4.2.5 Co-crystallisation of the TGT variants with preQ$_1$ derivatives

To further probe the new, enlarged binding pocket in structural terms, for several preQ$_1$ derivatives co-crystallisation trials were attempted with the two TGT variants. All selected derivatives, such as a phthalic acid derivative, a dimethylformimide derivative, or a t-butyloxycarbonyl (Boc)-protected preQ$_1$ (Figure 19), were easily available as intermediates or by-products of the queuine synthesis. These compounds feature spatial requirements which exceed beyond preQ$_1$. In contrast, they resemble, in steric terms, more the shape of queuine.

Figure 19: preQ$_1$ derivatives.
a) Phthalic acid derivative of preQ$_1$ **b)** Dimethylformimide derivative of preQ$_1$ **c)** Boc-preQ$_1$ **d)** Queuine.



Co-crystallisation and soaking attempts of the two TGT variants with the phthalic acid derivative failed. Rather the exposed microseeding or full size crystals were destroyed in the presence of this compound.

Co-crystallisation trials of the two TGT variants with the dimethyl formimide derivative did not yield any bound ligand in the structures.

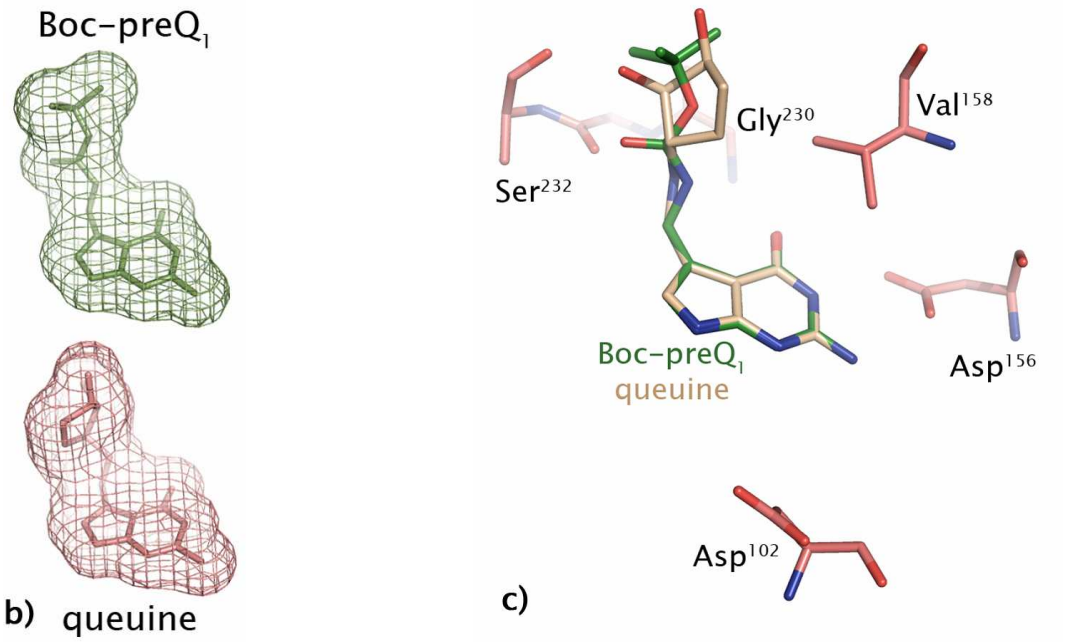


Figure 20: Crystal structure of Boc-preQ$_1$ bound to the TGT(C^{158}V/A^{232}S/V^{233}G).
a) 2|F$_o$|-|F$_c$| density contoured at a σ level of 1.5 is coloured blue, |F$_o$|-|F$_c$| density contoured at a σ level of 3 is coloured green (1.26 Å resolution). (1): usual orientation of carbonyl in all other deposited TGT structures, (2): usual position of the following residue 233. **b)** Steric comparison of Boc-preQ$_1$ with queuine. **c)** Queuine docking pose superposed with the crystal structure of TGT(C^{158}V/A^{232}S/V^{233}G) in complex with Boc-preQ$_1$.

Interestingly, a complex structure of the TGT($C^{158}V/A^{232}S/V^{233}G$) variant with Boc-protected $preQ_1$ could be determined (in the corresponding trial with the TGT($V^{233}G$) variant only an unoccupied binding pocket could be observed) (Figure 20a). In this complex, the side chain of Ser^{232} and the loop Gly^{233}-Gly^{234} are completely disordered, but the orientation of the Ser^{232} carbonyl group indicates already a different conformation of the loop 232-234, compared to all other TGT structures, leading to a probably even larger, and 'opened' binding pocket. Volume comparisons of Boc-$preQ_1$ with queuine reveals good steric similarity (Figure 20b). In addition, the observed Boc-$preQ_1$ binding conformation superimposes well with the assumed queuine binding modes from docking results (Figure 20c).

In the radioactive 'washout assay' the TGT($C^{158}V/A^{232}S/V^{233}G$) variant is able to excorporate labeled guanine from tRNA in the presence of 10 µM Boc-$preQ_1$ (chapter 5.3.2f, Appendix). Thus, the enzyme variant seems to be able to not only bind but also accept this $preQ_1$ derivative as substrate.

4.3 Discussion

4.3.1 The turnover numbers of the TGT variants decrease dramatically

In this study it was attempted to modify substrate specificity of bacterial TGT towards human TGT to investigate properties of an eukaryotic TGT binding pocket by means of a model system. On the basis of a multiple sequence alignment and homology modelling, three residues were selected for mutation – apart from the Tyr^{106}Phe exchange. The TGT variants, however, showed unexpected kinetic behaviour. While K_M values for guanine and tRNA remained almost unchanged within the experimental error for all variants, k_{cat} values were reduced by factors of one to two orders of magnitude in several cases.

These results indicate, that although the mutation sites are rather distant from the catalytic residue Asp^{280}, they play a crucial role for efficient enzyme turnover rates. As already suggested by Romier *et al.* [11], residues 233 and 158 perform van der Waals interactions to the purine heterocycle of the substrate. They might be important for stabilisation of the complex during catalysis. However, the Cys^{158} side chain was observed to be not completely rigid in the TGT(V^{233}G) variant crystal structures. A possible explanation could be that upon Gly^{233}Val replacement an appropriate bulky partner is missing to fix the orientation of the cystein side chain. This probably leads to the slight increase of the K_M value towards preQ$_1$.

4.3.2 The turnover numbers of TGT variants increase for the substrate base preQ$_1$

For preQ$_1$ k_{cat} values for the different TGT variants are significantly higher. The enhancement of k_{cat} for preQ$_1$ incorporation has already been

discussed [16,37] and possibly correlates with the nucleophilicity of nitrogen *N*9 in the respective substrates. In each case nitrogen *N*9 nucleophilically attacks *C*1 of the covalently bound tRNA ribose[34] during catalysis. In $preQ_1$, this nitrogen is the most potent nucleophile, as it is in no electronic conjugation with the exocylic amino methyl group. In guanine, *N*9 is less nucleophilic due to an electron withdrawing effect of *N*7 being in conjugation with *N*9. This is fully consistent with the observed decrease in turnover in the order $preQ_1$ > guanine.

4.3.3 Increased K_M for $preQ_1$

However, the K_M value for $preQ_1$ of the TGT($C^{158}V$) variant is drastically increased (factor 17.5, Table 9). This loss in binding affinity for $preQ_1$ could be due to different van der Waals contacts provided by Val^{158} replacing Cys^{158}. Already the K_M values for guanine show a slight tendency to lower values, upon $Cys^{158}Val$ exchange (TGT($C^{158}V/V^{233}G$) and TGT($C^{158}V/A^{232}S/V^{233}G$) variants, Table 9). Only a few eukaryotic TGTs have been kinetically characterised in literature: from rat liver [8], rabbit erythrocytes [44], and wheat germ [45]. For most of these TGTs no kinetic data concerning $preQ_1$ were published, except for wheat germ [45], where the authors stated that the affinity of TGT towards $preQ_1$ was not exactly determined but found to be at least 100-fold reduced in comparison to queuine. Taking into account that eukaryotic TGTs have no physiological need to incorporate $preQ_1$ into tRNA, as queuine is directly inserted to tRNAs ($preQ_1$ is not a relevant substrate in eukaryotes), the affinity loss of $preQ_1$ towards the TGT variants constructed in this study suggests another eukaryotic TGT specific property. Here, the affinity loss observed for the purine heterocycle moiety could possibly be compensated in the actual substrate queuine by additional hydrophobic interactions of Val^{158} with the cyclopentenyl moiety. This was already suggested by Romier *et al.* [11]. Furthermore, additional hydrogen bonds could be formed between the hydroxyl groups of the cyclopentenyl moiety

and the protein. This study did not reveal any favourable interaction partners for these hydroxyl groups, and the cyclopentenyl moiety remains disordered in the co-crystal structure. However, a reconsideration of the homology model and the multiple sequence alignment suggests another residue to be exchanged to better approximate human TGT: Tyr^{161} (*Z. mobilis* numering)/ Ser^{148} (*H. sapiens* numbering). This residue was not considered in the first attempt, since it is quite distant from the substrate binding pocket in the bacterial wild-type TGT binding pocket. It points towards the surface on the back side of the protein. In bacterial TGTs to 75 % tyrosine or phenylalanine is found at this position. In eukaryotic TGTs in 50 % threonine or serine (as in *H. sapiens* TGT) is present, potentially featuring a hydrogen bond towards one of the hydroxyl groups as shown in Figure 21. Queuine was docked into the human TGT homology model using GOLD [39,40] and minimised in the MAB force field of MOLOC [42], leaving only the ligand and the binding sites residues (Asp^{89}, Phe^{93}, Asp^{143}, Val^{145}, Ser^{148}) flexible. During minimisation, the Ser^{148} side chain rotates towards queuine by about 30 °. Also the MD simulation of the complex structure revealed a limited number of interaction sites for the cyclopentenyl moiety of queuine inside the chimeric model binding pocket. Creating an additional interaction site by the $Tyr^{161}Ser$ exchange could fix queuine in the binding pocket.

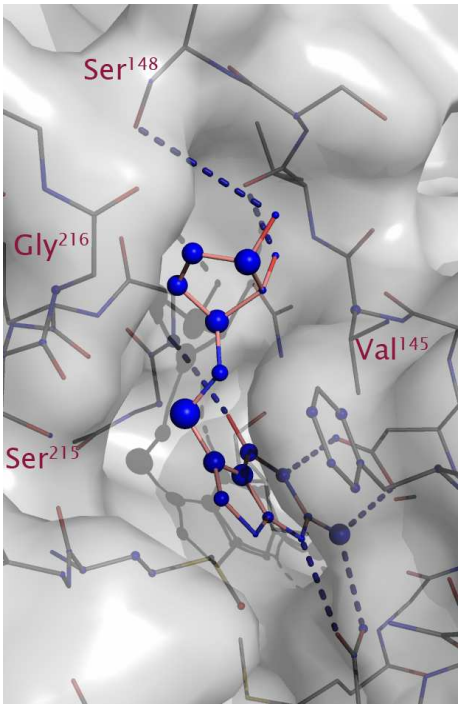


Figure 21: Queuine docked into the hTGT homology model after minimisation with the MAB force field of MOLOC [42].
The GOLD [39,40] docking pose was scored with the knowledge-based scoring function $DrugScore^{CSD}$[41]. Favourable per atom contributions are indicated by blue spheres, while the size of the sphere corresponds to the score, disfavourable per atom contributions are indicated by red spheres.

4.3.4 *Z. mobilis* wild-type TGT is eventually able to bind queuine

Surprisingly, in the radioactive 'washout assay', the wild-type enzyme showed an excorporation of [8-^{3}H]-guanine in the presence of queuine. Okada *et al.* [21] reported that *E. coli* TGT is unable to accept queuine as a substrate. In their assay 50 nM [^{14}C]-guanine^{34}-tRNA^{Tyr} were incubated with 217 nM *E. coli* TGT and 7.5 µM queuine. Already at this concentration, an excorporation of about 10 % of radioactivity is noticeable (Figure 22). In the present study, a concentration of 500 µM led to an even higher amount of radioactivity excorporation (Figure 13, p. 79).

Figure 22: Okada *et al.* [21], Figure 2.

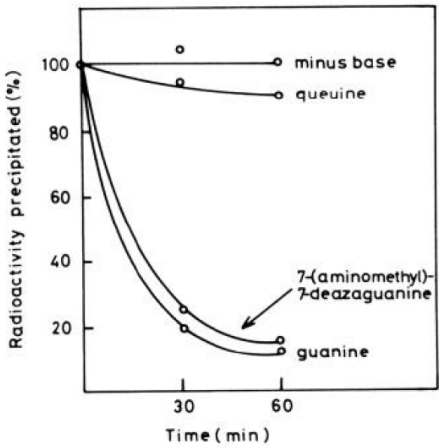


FIG. 2. Release of [^{14}C]guanine from [^{14}C]guanine-labeled *E. coli* G-tRNA by 7-(aminomethyl)-7-deazaguanine. The reaction mixture contained 0.1 nmol of [^{14}C]guanine-labeled *E. coli* G-tRNA^{Tyr} (5,000 cpm) and 20 µg of *E. coli* tRNA transglycosylase (Fraction V) with or without 15 nmol of the specified base in 2 ml of 70 mM Tris-HCl (pH 7.5), 20 mM $MgCl_2$. Samples of 0.6 ml taken after incubation for 0, 30, and 60 min were chilled and mixed with an equal volume of 30% trichloroacetic acid. The precipitate formed was collected on a glass fiber filter-paper disc, washed four times with cold 5% trichloroacetic acid and dried, and its radioactivity was counted in a liquid scintillation counter.

However, these indirect assays only show that the presence of queuine leads to the excorporation of labeled guanine from tRNA. Thus, it is only cautiously assumed but not proven that it is queuine that is incorporated into tRNA in exchange. In this study, also MALDI-TOF/MS analyses of a RNA oligo^{Asp} product of the TGT catalysed reaction were performed. The aim was to record

the base incorporated into tRNA by the enzyme. The results are listed in chapter 4.5 but will not be discussed here, since they did not allow a conclusive interpretation free of any contradictions.

Nevertheless, binding of queuine to wild-type *Z. mobilis* TGT seems not to be impossible: wild-type TGT crystals soaked with queuine immediately dissolved, possibly indicating a binding of queuine to the enzyme. In addition, MD simulations of the wild-type apo structure revealed a flexibility of the region around residues 158 and 233, possibly enabling a binding of queuine between these residues (Figure 18, Table 10, p. 86). Finally, the radioactive 'washout assay' clearly showed excorporation of [8-^{3}H]-guanine in the presence of queuine.

4.3.5 Outlook

In the future, a new crystallisation screen for the wild-type TGT in complex with queuine should be performed to determine the complex structure, or to reveal, what kind of compound is bound to the enzyme.

The variant crystals could be soaked with the cyclopentenyl moiety or other small molecule fragments in addition to $preQ_1$, to assess, whether they are able to bind to the generated subpocket. The variants are only models of a eukaryotic TGT binding pocket and possibly do not provide the cyclopentenyl moiety's interaction partners in appropriate distances. Eventually, in a free (not $preQ_1$-associated form), the cyclopentenyl moiety is able to adopt a better orientation leading to a better interaction pattern towards the binding pocket of the models. Finally, even though enzymatic characterisation of the variants has been proven difficult or even impossible, structural analysis of an additional $Tyr^{161}Ser$ variant could be performed, to further probe the binding pocket of the model system. However, it has to be considered, that amino acid residues further remote from the active site might affect indirectly active site residues as previously reported for TGT [16]. This might provide a slightly different but significantly better interaction pattern towards the queuine cyclopentenyl moiety actually given within an eukaryotic TGT binding pocket.

In conclusion, this study emphasises the importance to characterise the original protein structure to evaluate substrate specificity. Even for quite similar proteins, the composition of the active site, including remote residues taking indirect impact can be different enough to modulate enzymatic activity or specificity. There is still much to be learnt about TGTs originating from different domains of life that cannot be gleaned from homology models.

4.4 Summary

tRNA-guanine transglycosylases (TGT) - present in all three domains of life - catalyse a base exchange reaction in tRNAs. In bacteria, the mechanistic pathway of the guanine exchange towards the modified base $preQ_1$ is well characterised due to crystal structure analyses and elaborate kinetic studies of *Zymomonas mobilis* TGT [11-16]. As TGT plays a key role in pathogenicity of *Shigella flexneri* [26], the causative agent of bacterial dysentery, it has been established as a target for structure-based drug design [30]. However, eukaryotes also possess a TGT with high homology to the bacterial one and most residues inside the binding pocket are conserved [25]. In an analogous reaction the base queuine is incorporated into tRNA. Hence, it is of utmost importance to study selectivity-determining features.

Since no crystal structure of an eukaryotic TGT is yet available, in this study based on *Z. mobilis* TGT a genetically engineered human TGT binding pocket was designed as a model system. Three amino acids specific for the human TGT binding pocket were introduced by site-directed mutagenesis of the *Z. mobilis* TGT gene, ($Cys^{158}Val$, ($Ala^{232}Ser$), $Val^{233}Gly$). After purification, kinetic measurements and crystallisation were performed. Modified assays showed that - surprisingly - the *Z. mobilis* TGT variants have dramatically slowed down so that kinetic parameters could not be determined for incorporation of queuine into tRNA. Nevertheless, crystal structure analysis and molecular dynamics simulations indicate a possible binding of queuine to the active site of the variants. Additionally, it was observed that the wild-type enzyme is able to excorporate labeled guanine from tRNA in the presence of queuine. MD simulations indicate that the enzyme is able to open up a subpocket that potentially could accommodate queuine.

4.5 Supplement

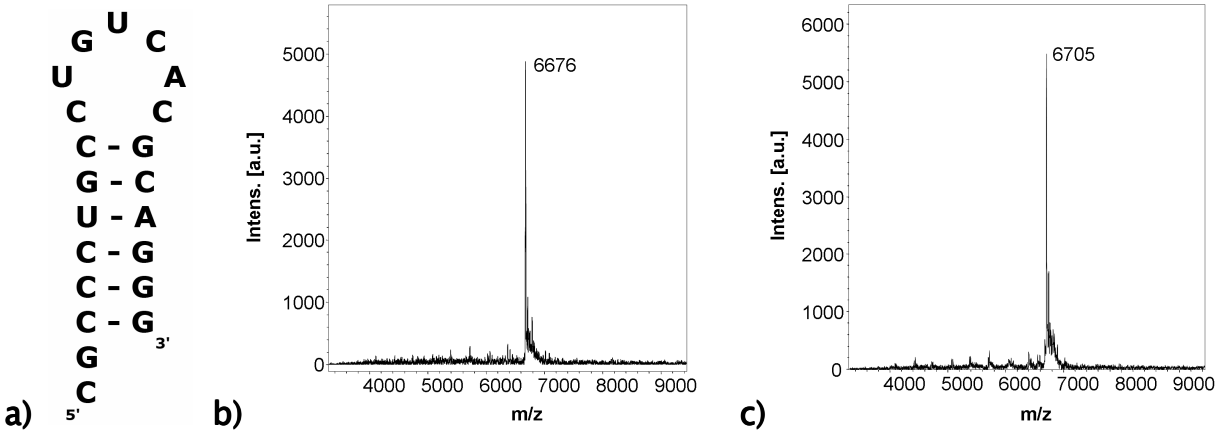
4.5.1 MALDI-TOF/MS analysis of the RNA product

To record the base incorporated into tRNA by the TGT variants and the wild-type enzyme, another assay was performed. In this assay the RNA product of the reaction was subsequently analysed *via* Matrix-assisted laser desorption ionisation time-of-flight mass spectrometry (MALDI-TOF/MS). However, highly charged full length tRNA is not applicable for an accurate MALDI-TOF/MS analysis, and secondly, the relative mass difference between native and modified tRNA would be too small (Table 11). It is reported that TGT also

Table 11: Relative mass differences of native and modified RNA substrates.

RNA	mass [Da]	Δmass [%]
native oligo^{Asp}	6678	
preQ$_1$-oligo^{Asp}	6707	0.43 %
queuine-oligo^{Asp}	6805	1.90 %
native tRNA^{Tyr}	27492	
preQ$_1$-tRNA^{Tyr}	27521	0.11 %
queuine-tRNA^{Tyr}	27619	0.46 %

Figure 23: MALDI-TOF/MS analysis of native and preQ$_1$-modified RNA oligo^{Asp}.
a) Sequence and assumed mini-helix structure of the tRNA^{Asp} anticodon stemloop RNA oligo, **b)** MALDI-TOF mass spectrum of unmodified RNA oligo^{Asp}, **c)** MALDI-TOF mass spectrum of preQ$_1$-modified RNA oligo^{Asp} (Δmass to (b): 29 Da).



accepts so-called 'mini-helices' [18] as RNA-substrate. Thus, a 21mer RNA oligo (corresponding to the $tRNA^{Asp}$ anticodon stemloop) was purchased (Figure 23a). First control experiments with the wild-type enzyme and 500 µM $preQ_1$ showed that, in fact, the RNA $oligo^{Asp}$ is accepted as substrate and $preQ_1$ is incorporated. This is indicated as native and $preQ_1$-modified RNA $oligo^{Asp}$ could be clearly distinguished in the MALDI-TOF mass spectra. The observed mass difference corresponds to the expected values (Figure 23b, c; Table 12). Also lower $preQ_1$ concentrations were tested. Already a concentration of 5 µM lead to a detectable, but only partial guanine to $preQ_1$ replacement, while at a concentration of 1 µM only unmodified RNA $oligo^{Asp}$ can be detected.

Table 12: MALDI-TOF/MS analysis of RNA $oligo^{Asp}$ reaction products.

TGT	substrate	Δmass[1] [Da]
wild-type TGT	$preQ_1$	29 ± 2
TGT($V^{233}G$)	$preQ_1$	28 ± 2
TGT($C^{158}V/V^{233}G$)	$preQ_1$	28 ± 2[(2)]
wild-type TGT	queuine	27 ± 2[(2)]
TGT($V^{233}G$)	queuine	27 ± 2[(2)]
TGT($C^{158}V$)	queuine	-
TGT($C^{158}V/V^{233}G$)	queuine	-
TGT($C^{158}V/A^{232}S/V^{233}G$)	queuine	-
wild-type TGT	Boc-$preQ_1$	38 ± 2[(2)]
TGT($V^{233}G$)	Boc-$preQ_1$	38 ± 2[(2)]
TGT($C^{158}V$)	Boc-$preQ_1$	38 ± 2[(2)]
TGT($C^{158}V/V^{233}G$)	Boc-$preQ_1$	38 ± 2[(2)]
TGT($C^{158}V/A^{232}S/V^{233}G$)	Boc-$preQ_1$	38 ± 2[(2)]

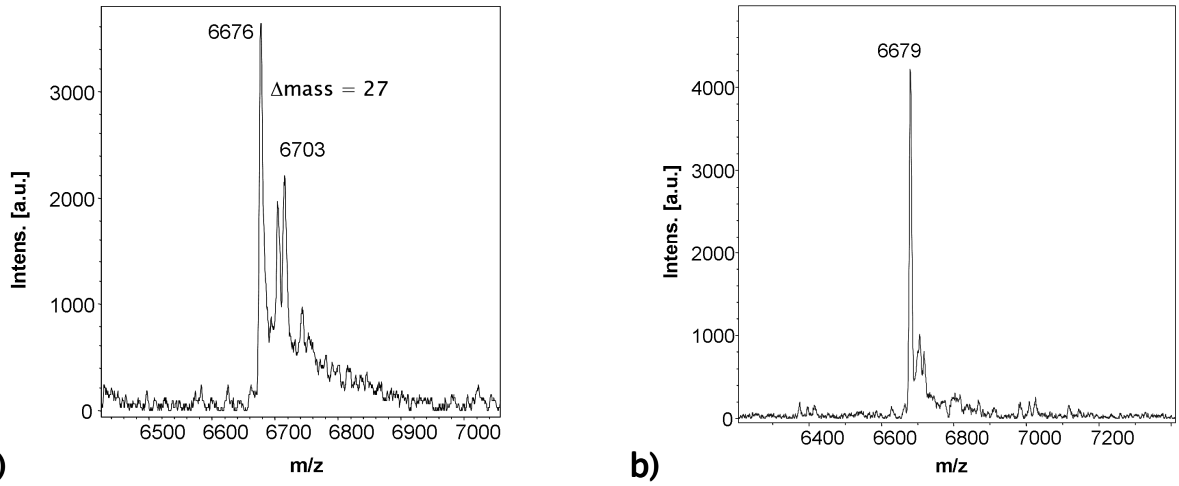
1) Mass differences to native oligo (6678 Da): 29 Da would correspond to a $preQ_1$ modification of the RNA $oligo^{Asp}$, 127 Da to a queuine modification, 128 Da to a Boc-$preQ_1$ modification, **2)** Only partial modification observed.

Also the TGT($V^{233}G$) variant (decreased k_{cat} by a factor of about 3 – 10 with respect to the wild-type TGT values, Table 9, p. 77) is able to incorporate $preQ_1$ efficiently (Table 12), once incubated with 500 µM $preQ_1$. It can be assumed, that the TGT($C^{158}V$) variant (decreased k_{cat} by a factors of about 3

with respect to the wild-type TGT values, Table 9, p. 77) is able to incorporate $preQ_1$ with similar efficiency, since the (irreversible) exchange reaction has been executed for more than 16 h. For such a long time span even the increased $K_M(preQ_1)$ value should not matter significantly. The TGT($C^{158}V/V^{233}G$) variant (decreased k_{cat} by a factor of about two orders of magnitude with respect to the wild-type TGT values, Table 9, p. 77) is found to be rather inefficient, thus only a reduced quantity of $preQ_1$ modification is detected besides unmodified RNA oligo^{Asp}. In light of this finding it can be assumed, that also the TGT($C^{158}V/A^{232}S/V^{233}G$) variant with a similarly reduced k_{cat} (Table 9, p. 77) is only capable to produce minor amounts of $preQ_1$-modified RNA oligo^{Asp}.

Figure 24: MALDI-TOF/MS analysis of RNA oligo^{Asp} reaction products upon incubation with queuine.

a) MALDI-TOF mass spectrum of partially $preQ_1$-modified RNA oligo^{Asp}, produced by wild-type TGT incubated with queuine, **b)** MALDI-TOF mass spectrum of native (unmodified) RNA oligo^{Asp}, produced by TGT($C^{158}V$) incubated with queuine.



Incubation of the RNA oligo^{Asp} with 500 µM pure queuine, results in no significant queuine modification. In case of TGT($C^{158}V$), TGT($C^{158}V/V^{233}G$), TGT($C^{158}V/A^{232}S/V^{233}G$) only unmodified RNA oligo^{Asp} could be isolated (Figure 24b). These results are in good agreement with the radioactive 'washout assay', where these three variants were found to be not sufficiently active to produce a measurable signal. However, in case of the remaining TGT($V^{233}G$) variant and the wild-type enzyme, the mass spectra showed

surprising results. A signal with a mass difference is observed, that suggests incorporation of $preQ_1$ (Figure 24a; Table 12, p. 98). Contamination of queuine by $preQ_1$ can be excluded as high pressure liquid chromatography coupled with electrospray ionisation mass spectrometry[1] (HPLC/ESI-MS) and ^1H-NMR spectrum[2] analyses did not provide any evidence. In addition, since the radioactive 'washout assay' suggested also Boc-$preQ_1$ as a substrate, RNA oligo^{Asp} products of the corresponding reaction were analysed as well (chapter 5.4.j-m, Appendix). The detected peaks do not correspond to any expected masses. Rather, it is awkward to interprete the calculated mass differences (Table 12, p. 98). Thus, MALDI-TOF/MS analysis of the RNA product seems to be not appropriate to determine the base that had been incorporated by TGT.

4.5.2 MALDI-TOF/MS assay drawbacks

MALDI-TOF/MS analysis of the RNA oligo^{Asp} products obtained after incubation with the TGT variants was only successful for those variants, that exhibited an only moderately decreased reaction velocity. Due to the temperature sensitivity of the RNA oligo^{Asp} substrate, the reaction mixture was incubated at 25 °C. To compensate for the probably lower enzymatic activity at this temperature, incubation time was elongated to 16 h (in comparison, the [8-^3H]-guanine assay which is performed at 37 °C was only elongated from 4 min to 30 min to receive a measurable signal of the variants). Yet, there are additional difficulties. Curnow *et al.* [18] reported a 6-fold increase of K_M for a 17-base tRNA^{Tyr} anticodon stemloop. For wild-type TGT only a rough estimation of the K_M value for the RNA oligo^{Asp} substrate could be performed *via* Lineweaver-Burk linear regression of [8-^3H]-guanine assay data (chapter 5.3.1f, Appendix). Already for the wild-

---

1) Performed by Anja Vogt, Institut für Pharmazeutische Chemie, Philipps-Universität Marburg.

2) Performed by Hans-Dieter Gerber, Institut für Pharmazeutische Chemie, Philipps-Universität Marburg.

type enzyme, the approximate K_M with 1.5 mM is quite high. For the even less active variants, assuming no further increase of the K_M, the catalytic efficiency would be by far too low to record detectable modification of the RNA oligo^{Asp}. Curnow *et al.* [18] reported that a 4-base elongation of their minihelix led to a stemloop that is stable at 37 °C. This could be a possibility to increase the performance of the less active variants, eventually leading to a detectable modification of the RNA oligo^{Asp}. Nevertheless, the conditions applied during MALDI-TOF/MS analysis of the RNA oligo^{Asp} have to be examined first. The results described above cannot be interpreted consistently but they might indicate that the stability of the base modification is not sufficient to survive the applied conditions in the MALDI-TOF/MS experiments.

In literature, analysis of tRNA^{Tyr} RNase T_1 hydrolysis by PEI-cellulose chromatography had revealed the position of modification in tRNA [21], and tRNA modification with CNBr and RPC-5 analysis could distinguish between native and queuine modified tRNAs [46-48]. These techniques could possibly support the analysis of full length tRNA after incubation with the TGT variants.

4.6 Materials and Methods

4.6.1 Homolgy modelling

Amino acid sequences were extracted from the deposited crystal structures with PDB codes: 1r5y (chain A), 1y5v (chain A), 1y5w (chain A), 1ozm (chain A), 1ozq (chain A), and 1q2r (chain A, and B, respectively). These sequences were aligned with the human TGT sequence (from SWISS-PROT Release 54.3 [49]) utilising Clustal W 1.83 [33]. The sequences aligned with 41 -42 % identity. Based on the alignment, ten homology models for human TGT were calculated with MODELLER 6a [35]. The best model (model no. 8) was selected by picking the model with the lowest value (11,429.19) of the MODELLER objective function, which is reported in the second line of the model PDB file.

4.6.2 Construction of the TGT mutants C158V, V233G, C158V-V233G, A232S-V233G, C158V-A232S-V233G

Site-directed mutagenesis was performed using the QuickChange™ kit (Stratagene) following the vendor's protocol. The oligonucleotides listed in Table 13 were used to introduce the mutations into the Y106F mutant of the *tgt* gene on the expression plasmid pET9d-ZM-Y106F [14]. Sequencing of the entire *tgt* gene (MWG Biotech, Ebersberg) confirmed the presence of the desired mutations as well as the absence of any further unwanted mutation. Subsequently, the mutated plasmids were transformed into *E. coli* BL21 (DE3) pLysS cells.

Table 13: Oligonucleotides used in mutagenesis.

mutation site/ oligo name	sequence (5' to 3')[1]
C158V-f	5'-GTA ATG GCC TTT GAC GAA **GT**C ACG CCT TAT CCA GC-3'
C158V-b	5'-GC TGG ATA AGG CGT G**AC** TTC GTC AAA GGC CAT TAC-3'
V233G-f	5'-GCT GTT GGG GGA TTG GCT G**G**G GGT GAA GGA CAG GAT G-3'
V233G-b	5'-C ATC CTG TCC TTC ACC C**C**C AGC CAA TCC CCC AAC AGC-3'
A232S-V233G-f	5'-GCT GTT GGG GGA TTG **T**CT G**G**G GGT GAA GGA CAG GAT G-3'
A232S-V233G-b	5'-C ATC CTG TCC TTC **A**CC C**C**C AGA CAA TCC CCC AAC AGC-3'

[1]) The mutated bases are printed ind bold letters and highlighted with grey background.

4.6.3 Expression, Purification

Z. mobilis TGT variants were prepared as described elsewhere [12,36]. *E. coli* $tRNA^{Tyr}$ (ECY2) was prepared as described by Curnow *et al.* [50]. Radio-labeling of $tRNA^{Tyr}$ was performed as already described [16].

4.6.4 Determination of kinetic parameters

Kinetic parameters for the TGT variants were determined for the different substrates essentially as described by Meyer *et al.* [51]. Michaelis-Menten parameters for tRNA and guanine were determined monitoring incorporation of radioactively labelled guanine into $tRNA^{Tyr}$ in position 34. Kinetic data for both substrates were determined separately in duplicate and average values were calculated. Kinetic parameters for guanine were measured using 150 nM (wild-type, TGT($V^{233}G$), TGT($C^{158}V$), TGT($A^{232}S/V^{233}G$) variants) - 300 nM (TGT($C^{158}V/V^{233}G$), TGT($C^{158}V/A^{232}S/V^{233}G$) variants) TGT, 15 µM unlabelled *E. coli* $tRNA^{Tyr}$, and variable concentrations of [8-^{3}H]-guanine (0.9 Ci/mmol, Hartmann Analytic) in the range of 0.5 – 20 µM in buffer solution (200 mM HEPES pH 7.3, 20 mM $MgCl_2$, 0.036 % (v/v) Tween 20 (Roth)). Kinetic parameters for tRNA were measured using 150 nM (wild-type,

TGT(V^{233}G), TGT(C^{158}V), TGT(A^{232}S/V^{233}G) variants) – 300 nM (TGT(C^{158}V/V^{233}G), TGT(C^{158}V/A^{232}S/V^{233}G) variants) TGT, 20 µM [8-^{3}H]-guanine, and variable concentrations of tRNA^{Tyr} (0.25 – 15 µM). Initial velocities of the base exchange reaction in counts per minute were converted to [µM/min] using a calibration constant derived from liquid scintillation counting of [8-^{3}H]-guanine solutions with variable concentrations. Kinetic parameters were determined *via* regression using GraFit [52].

Michaelis-Menten parameters for preQ$_1$ were calculated *via* monitoring the loss of [8-^{3}H]-guanine from tRNA^{Tyr} radio-labelled in position 34. Kinetic parameters for preQ$_1$ were measured using using 150 nM (wild-type, TGT(V^{233}G), TGT(C^{158}V), TGT(A^{232}S/V^{233}G) variants) – 300 nM (TGT(C^{158}V/V^{233}G), TGT(C^{158}V/A^{232}S/V^{233}G) variants) TGT and 8 µM radio-labelled tRNA^{Tyr}. The concentration of preQ$_1$ was varied in ranges of 0.5 – 15 µM, 1 – 50 µM, and 25 – 750 µM, respectively. Initial velocities in counts per minute were calculated from the decreasing tritium labelling level of tRNA due to the incorporation of the respective substrate bases. Initial velocities in counts per minute were converted to [µM/min] using a calibration constant derived from liquid scintillation counting of guanine/ [8-^{3}H]-guanine solutions with variable concentrations. Kinetic parameters were determined *via* regression using GraFit [52].

4.6.5 MALDI-TOF/MS analysis

342 pmol of the RNA oligo^{Asp} were incubated with 1.78 µM TGT, 1 – 500 µM preQ$_1$ or queuine in 200 mM HEPES, pH 7.3, 20 mM $MgCl_2$ at 25 °C for at least 16 h. TGT and free preQ$_1$ or queuine were extracted from the reaction mixture by the addition of equal volumes of Roti-Phenol (Roth) and chloroform : isoamylalcohol (24:1). The aqueous supernatant was once more treated with an equal volume of chloroform : isoamylalcohol (24:1). An additional purification step was performed *via* gel filtration using NAP-

columns (GE Healthcare, Life Science). The RNA oligo was eluted with pure nuclease-free water and vacuum dried. The samples were further purified utilising magnetic beads and analysed by MALDI-TOF/MS by Dr. Silke Seeber and Nico Vogel (Institut für Biochemie, Friedrich-Alexander-Universität Erlangen- Nürnberg).

4.6.6 Crystallisation, data collection, and processing

Recombinant *Z. mobilis* TGT was expressed in *E. coli* BL21 (DE3), purified and crystallised as described previously [12,36]. TGT variant crystals suitable for ligand soaking were produced in a two step procedure. Droplets were prepared by mixing 2 µl of concentrated protein solution (14 mg/ml) with 2 µl reservoir solution [100 mM Tris/HCl, pH 8.5, 1 mM DTT, 13 % (w/v) PEG 8,000, 10 % (v/v) DMSO]. Micro-crystals were grown at 291 K using the hanging-drop vapour diffusion method in the presence of 1.0 ml of reservoir solution of the respective seeding buffer. Micro crystals grew within two weeks.

Subsequently, macro-seeding was performed under similar conditions to obtain crystals of the uncomplexed protein. Again droplets were prepared by mixing 2 µl of concentrated protein solution with 2 µl macro-seeding reservoir solution [100 mM Tris/HCl, pH 8.5, 1 mM DTT, 5 % (w/v) PEG 8,000, 10 % (v/v) DMSO]. One micro-crystal was transferred into this solution. Single crystals with a size of approximately 0.7 x 0.7 x 0.2 mm^3 grew within two to four weeks.

To allow co-crystallisation of the TGT variants in complex with queuine the compound was dissolved in DMSO and added to the macro-seeding droplet to a final concentration of 1 to 20 mM. Crystals of the TGT variants complexed to $preQ_1$ were produced by a soaking procedure. The compound was dissolved in DMSO and added to 2 µl of a crystallisation droplet to a final

concentration of 10 to 20 mM. Finally, a single TGT crystal was transferred into the droplet and soaked for 30 min.

For data collection, crystals were cryoprotected using glycerol. The glycerol and PEG 8,000 concentrations of the macro-seeding buffers were increased stepwise by transferring the crystals to six different 2 µl cryo-droplets each with 5 to 30 min incubation times [glycerol concentrations (v/v): 5 % → 10 % → 15 % → 20 % → 25 % → 30 %; and PEG 8,000 concentrations (w/v): 5.0 % → 6.3 % → 7.5 % → 8.0 % → 8.8 % → 9.8 %, respectively]. These droplets also contained the ligands at equivalent concentrations compared to soaking and co-crystallisation conditions. The cryo-soaked crystals were flash-frozen in liquid N_2. Data sets were collected at cryo conditions (100 K) at the BESSY (Berliner Elektronenspeicherring-Gesellschaft für Synchrotronstrahlung, Berlin, Germany), PSF (Proteinstrukturfabrik) beamline BL-14.2 and with CuKα radiation (λ = 1.5418 Å) using a Rigaku RU-300 rotating-anode generator at 50 kV and 90 mA equipped with Xenocs focussing optics and an R-AXIS IV detector. All crystals tested exhibited monoclinic symmetry in space group C2 containing one monomer per asymmetric unit with Matthews coefficients of 2.3 - 2.4. Data processing and scaling was performed using the HKL2000 package [53]. For all refined structures unit cell dimensions for the crystals, data collection and processing statistics are given in Table 14.

Coordinates of the apo TGT crystal structure grown at pH 8.5 (PDB-code: 1pud) were slightly modified (deletion of coordinates at the mutation sites) before applying for initial rigid-body refinement of the protein molecule followed by repeated cycles of conjugate gradient energy minimisation, simulated annealing and B-factor refinement using the CNS program package [54]. Refinement at the later stages for all structures was performed with SHELXL [55]. Here, up to 50 cycles of conjugate gradient minimisation were performed with default restraints on bonding geometry and B-values: 5 % of all data were used for R_{free} calculation. Amino acid side chains were fit to σA-weighted $2|F_o|-|F_c|$ and $|F_o|-|F_c|$ electron density maps using

Coot [56]. Water, glycerol molecules, and ligands were located in the difference electron density and added to the model for further refinement cycles. Anisotropic conjugate gradient refinement resulted in a significant improvement of the models built for 2nso, 2nqz, 3bld, 3bll. During the last refinement cycles, riding H-atoms were introduced for the protein residues (not for the ligand) without using additional parameters. All final models were validated using PROCHECK [57]. Data refinement statistics are given in Table 14.

Table 14: Crystallographic table.

[a]) number in parentheses is for highest resolution shell

[b]) $R(I)_{sym} = \Sigma\ |I - \langle I\rangle|\ /\ \Sigma I$ with I representing the observed intensities and $\langle I\rangle$ representing the mean observed intensity

[c]) $R_{work} = \Sigma_{hkl}\ |F_o - F_c|\ /\ \Sigma_{hkl}\ |F_o|$.

[d]) R_{free} was calculated as for R_{work} but on 5 % of the data excluded from the refinement.

Crystal data	TGT($C^{158}V/A^{232}S/V^{233}G$) (2nso)	TGT($C^{158}V/A^{232}S/V^{233}G$)·$preQ_1$ (2nqz)	TGT($V^{233}G$) (3bl3)
A. Data processing			
Space group	*C2*	*C2*	*C2*
a, b, c (Å)	90.73, 64.83, 70.81	89.90, 64.93, 71.18	90.69, 64.77, 70.40
β (°)	96.1	96.36	95.94
Resolution range[a] (Å)	50–1.60 (1.63–1.60)	50–1.46 (1.49–1.46)	50–2.25 (2.30–2.25)
Total no. of reflections	153,286	223,937	281,752
No. of unique reflections	53,101	69,580	49,829
Completeness[a] (%)	98.6 (79.8)	98.6 (80.7)	93.4 (66.4)
Redundancy	2.9	3.2	2.8
$R(I)_{sym}$[a,b] (%)	7.1 (34.5)	6.3 (21.4)	11 (28.1)
$I/\sigma(I)$[a]	15 (2.0)	18 (3,9)	9 (2.4)
B. Refinement			
R_{work}[c]/R_{free}[d] (%)	16.2/22.1	15.00/19.84	18.67/22.73
No. of atoms/residues (molecules)			
Protein	2769/364	2737/360	2755/365
Water	185	258	216
Glycerol (cryo buffer)	12/2	38/6	6/1
Ligand	–	14	-
Mean *B*–factors (Å^2)			
Protein	24.2	20.1	19.9
Water	30.7	30.8	28.4
Glycerol (cryo buffer)	36.0	32.4	30.8
Ligand	–	22.6	-
rmsd angle (°)	2.0	2.1	1.3
rmsd bond (Å)	0.009	0.011	0.008
(An)isotropic refinement	Anisotropic	Anisotropic	Isotropic

Crystal data	TGT($V^{233}G$) · $preQ_1$ (3bld)	TGT($C^{158}V/A^{232}S/V^{233}G$) · queuine (3blo)	TGT($C^{158}V/A^{232}S/V^{233}G$) · Boc–$preQ_1$ (3bll)
A. Data processing			
Space group	*C2*	*C2*	*C2*
a, b, c (Å)	89.66, 64.88, 70.33	90.38, 64.78, 70.32	89.95, 64.85, 70.47
β (°)	95.55	95.91	95.56
Resolution range[a] (Å)	20–1.19 (1.22–1.19)	20–1.60 (1.63–1.60)	20–1.26 (1.29–1.26)
Total no. of reflections	428,522	122,910	328,720
No. of unique reflections	116,963	51,989	97,859
Completeness[a] (%)	91.0 (62.8)	97.5 (89.2)	90.4 (63.3)
Redundancy	3.7	2.4	3.4
$R(I)_{sym}$[a,b] (%)	4.1 (26.2)	5.2 (29.9)	3.7 (31.9)
$I/\sigma(I)$[a]	21 (2.7)	18 (2.1)	21 (2.2)
B. Refinement			
R_{work}[c]/R_{free}[d] (%)	15.08/18.54	18.02/21.28	14.92/18.21
No. of atoms/residues (molecules)			
Protein	2653/348	2664/349	2631/347
Water	250	177	227
Glycerol (cryo buffer)	12/2	6/1	18/3
Ligand	13/1	20/1	20(1)
Mean *B*-factors (Å^2)			
Protein	18.6	24.1	18.1
Water	28.4	30.8	28.9
Glycerol (cryo buffer)	32.8	35.0	33.1
Ligand	33.0	39.9	28.9
rmsd angle (°)	2.1	2.2	2.2
rmsd bond (Å)	0.013	0.010	0.012
(An)isotropic refinement	Anisotropic	Isotropic	Anisotropic

4.6.7 Molecular Dynamics Simulation

The MD simulation and all setup steps were performed with the Amber 8.0 suite of programs using the Amber 1999 force field [58]. The TGT($C^{158}V/A^{232}S/V^{233}G$) variant queuine complex structure was used as starting coordinates. The ligand and all crystallographic water molecules were removed. Queuine was docked into the structure using GOLD [39,40] and minimised in the MAB force field of MOLOC [42], defining only the queuine atoms as flexible, while protein atoms were left rigid. For the apo wild-type structure simulation, 1pud was used. Missing residues and side chains were added and minimised with the Amber 1999 force field. After estimating protonation states of all histidines by visual inspection of interactions, histidines were set to the appropriate form (double protonated: 'HIP'/ protonated at Nδ: 'HID' (HIS#: 73, 257, 319, 332, 333)/ protonated at Nε: 'HIE' (HIS#: 90, 127, 133, 145, 349)). Cysteins coordinating the Zn^{2+} ion were set to 'CYM' (CYS#: 319, 320, 323). Hydrogen atoms were added with PROTONATE. The protein was initially subjected to 200 steps of minimisation using a generalised Born solvation model. Subsequently, the systems were solvated in a box of 2563 (TGT($C^{158}V/A^{232}S/V^{233}G$)·queuine complex)/ 2350 (apo wild-type TGT) TIP3P water molecules [59], and two (TGT($C^{158}V/A^{232}S/V^{233}G$)·queuine complex)/ three (apo wild-type TGT) sodium ions were added to ensure neutrality. After 200 steps of minimisation of the solvated systems, the MD simulations were started by heating the solvent to 300 K over a period of 20 ps and cooling to 100 K over 5 ps, keeping the protein fixed. Then the entire system was brought to 300 K over a period of 25 ps and the simulation was carried on for 5 ns under NPT conditions, using a time step of 2 fs and PME [60] for evaluating the electrostatic interactions. Energy data were saved every 20 fs, protein coordinates every 0.5 ps. CARNAL was used for further analysis of the trajectory and VMD 1.8.2 [61] for visualisation.

4.6.8 Alignment and Figures

Alignment of structures with similar or identical sequences was performed with the alignment function implemented in Pymol (http://www.pymol.org). Figures were prepared using Isis Draw (MDL, San Leandro, USA) and Pymol.

4.6.9 Protein Data Bank Accession Codes

The following Protein Data Bank (PDB) accession codes were allocated to the crystal structures determined during this work: 2nso, 2nqz, 3bl3, 3bld, 3bll, 3blo (also given in Table 14).

4.7 Acknowledgements

This work was supported by the NIH (Grant GM065489) and by the Deutsche Forschungsgemeinschaft (Graduiertenkolleg "Protein Function at the Atomic Level"). We gratefully acknowledge the BESSY (Berliner Elektronenspeicher-ring-Gesellschaft für Synchrotronstrahlung) in Berlin, Germany, for providing beamtime at the PSF (Proteinstrukturfabrik) beamline BL-14.2. We thank Christian Sohn for his support during X-ray data collection, Claudia Sass for performing magnetic beads purification of the RNA oligo samples, Hans-Dieter Gerber for synthesising $preQ_1$ and queuine and Andreas Blum for helpful discussions.

4.8 References

[1] Iwata-Reuyl D. Biosynthesis of the 7-deazaguanosine hypermodified nucleosides of transfer RNA. Bioorg Chem 2003;31 (1):24-43.

[2] Stengl B, Reuter K, Klebe G. Mechanism and substrate specificity of tRNA-guanine transglycosylases (TGTs): tRNA-modifying enzymes from the three different kingdoms of life share a common catalytic mechanism. Chembiochem 2005;6 (11):1926-39.

[3] Watanabe M, Matsuo M, Tanaka S, Akimoto H, Asahi S, Nishimura S, Katze JR, Hashizume T, Crain PF, McCloskey JA, Okada N. Biosynthesis of archaeosine, a novel derivative of 7-deazaguanosine specific to archaeal tRNA, proceeds via a pathway involving base replacement on the tRNA polynucleotide chain. J Biol Chem 1997;272 (32):20146-51.

[4] Van Lanen SG, Kinzie SD, Matthieu S, Link T, Culp J, Iwata-Reuyl D. tRNA modification by S-adenosylmethionine:tRNA ribosyltransferase-isomerase. Assay development and characterization of the recombinant enzyme. J Biol Chem 2003;278 (12):10491-9.

[5] Mathews I, Schwarzenbacher R, McMullan D, Abdubek P, Ambing E, Axelrod H, Biorac T, Canaves JM, Chiu HJ, Deacon AM, DiDonato M, Elsliger MA, Godzik A, Grittini C, Grzechnik SK, Hale J, Hampton E, Han GW, Haugen J, Hornsby M, Jaroszewski L, Klock HE, Koesema E, Kreusch A, Kuhn P, Lesley SA, Levin I, Miller MD, Moy K, Nigoghossian E, Ouyang J, Paulsen J, Quijano K, Reyes R, Spraggon G, Stevens RC, van den Bedem H, Velasquez J, Vincent J, White A, Wolf G, Xu Q, Hodgson KO, Wooley J, Wilson IA. Crystal structure of S-adenosylmethionine:tRNA ribosyltransferase-isomerase (QueA) from Thermotoga maritima at 2.0 A resolution reveals a new fold. Proteins 2005;59 (4):869-74.

[6] Grimm C, Ficner R, Sgraja T, Haebel P, Klebe G, Reuter K. Crystal structure of Bacillus subtilis S-adenosylmethionine:tRNA ribosyltransferase-isomerase. Biochem Biophys Res Commun 2006;351 (3):695-701.

[7] Frey B, McCloskey J, Kersten W, Kersten H. New function of vitamin B12: cobamide-dependent reduction of epoxyqueuosine to queuosine in tRNAs of Escherichia coli and Salmonella typhimurium. J Bacteriol 1988;170 (5):2078-82.

[8] Shindo-Okada N, Okada N, Ohgi T, Goto T, Nishimura S. Transfer ribonucleic acid guanine transglycosylase isolated from rat liver. Biochemistry 1980;19 (2):395-400.

[9] Bjork GR. Genetic dissection of synthesis and function of modified nucleosides in bacterial transfer RNA. Prog Nucleic Acid Res Mol Biol 1995;50:263-338.

[10] Bjork GR, Durand JM, Hagervall TG, Leipuviene R, Lundgren HK, Nilsson K, Chen P, Qian Q, Urbonavicius J. Transfer RNA modification: influence on translational frameshifting and metabolism. FEBS Lett 1999;452 (1-2):47-51.

[11] Romier C, Reuter K, Suck D, Ficner R. Crystal structure of tRNA-guanine transglycosylase: RNA modification by base exchange. Embo J 1996;15 (11):2850-7.

[12] Reuter K, Ficner R. Sequence analysis and overexpression of the Zymomonas mobilis tgt gene encoding tRNA-guanine transglycosylase: purification and biochemical characterization of the enzyme. J Bacteriol 1995;177 (18):5284-8.

[13] Romier C, Reuter K, Suck D, Ficner R. Mutagenesis and crystallographic studies of Zymomonas mobilis tRNA-guanine transglycosylase reveal aspartate 102 as the active site nucleophile. Biochemistry 1996;35 (49):15734-9.

[14] Brenk R, Stubbs MT, Heine A, Reuter K, Klebe G. Flexible adaptations in the structure of the tRNA-modifying enzyme tRNA-guanine transglycosylase and their implications for substrate selectivity, reaction mechanism and structure-based drug design. Chembiochem 2003;4 (10):1066-77.

[15] Grädler U, Ficner R, Garcia GA, Stubbs MT, Klebe G, Reuter K. Mutagenesis and crystallographic studies of Zymomonas mobilis tRNA-guanine transglycosylase to elucidate the role of serine 103 for enzymatic activity. FEBS Lett 1999;454 (1-2):142-6.

[16] Tidten N, Stengl B, Heine A, Garcia GA, Klebe G, Reuter K. Glutamate versus glutamine exchange swaps substrate selectivity in tRNA-guanine transglycosylase: Insight into the regulation of substrate selectivity by kinetic and crystallographic studies. J Mol. Biol. 2007;374 (3):764-76.

[17] Nakanishi S, Ueda T, Hori H, Yamazaki N, Okada N, Watanabe K. A UGU sequence in the anticodon loop is a minimum requirement for recognition by Escherichia coli tRNA-guanine transglycosylase. J Biol Chem 1994;269 (51):32221-5.

[18] Curnow AW, Garcia GA. tRNA-guanine transglycosylase from Escherichia coli. Minimal tRNA structure and sequence requirements for recognition. J Biol Chem 1995;270 (29):17264-7.

[19] Goodenough-Lashua DM, Garcia GA. tRNA-guanine transglycosylase from E. coli: a ping-pong kinetic mechanism is consistent with nucleophilic catalysis. Bioorg Chem 2003;31 (4):331-44.

[20] Okada N, Noguchi S, Kasai H, Shindo-Okada N, Ohgi T, Goto T, Nishimura S. Novel mechanism of post-transcriptional modification of tRNA. Insertion of bases of Q precursors into tRNA by a specific tRNA transglycosylase reaction. J Biol Chem 1979;254 (8):3067-73.

[21] Okada N, Nishimura S. Isolation and characterization of a guanine insertion enzyme, a specific tRNA transglycosylase, from Escherichia coli. J Biol Chem 1979;254 (8):3061-6.

[22] Romier C, Ficner, R., Suck, D. Structural basis for base exchange by tRNA-guanine transglycosylases. Modification and Editing of RNA 1998:169-82.

[23] Winkler ME. Genetics and regulation of base modification in the tRNA and rRNA of prokaryotes and eukaryotes. Modification and Editing of RNA 1998:441-69.

[24] Deshpande KL, Katze JR. Characterization of cDNA encoding the human tRNA-guanine transglycosylase (TGT) catalytic subunit. Gene 2001;265 (1-2):205-12.

[25] Romier C, Meyer JE, Suck D. Slight sequence variations of a common fold explain the substrate specificities of tRNA-guanine transglycosylases from the three kingdoms. FEBS Lett 1997;416 (1):93-8.

[26] Durand JM, Okada N, Tobe T, Watarai M, Fukuda I, Suzuki T, Nakata N, Komatsu K, Yoshikawa M, Sasakawa C. vacC, a virulence-associated chromosomal locus of Shigella flexneri, is homologous to tgt, a gene encoding tRNA-guanine transglycosylase (Tgt) of Escherichia coli K-12. J Bacteriol 1994;176 (15):4627-34.

[27] Sansonetti PJ. Molecular and cellular bases of intestinal epithelial cell invasion by Shigella flexneri. C. R. Acad. Sci. ser. III 1997;320:729-34.

[28] Dorman CJ, Porter ME. The Shigella virulence gene regulatory cascade: a paradigm of bacterial gene control mechanisms. Mol. Microbiol. 1998;29:677-84.

[29] Hurt JK, Olgen S, Garcia GA. Site-specific modification of Shigella flexneri virF mRNA by tRNA-guanine transglycosylase in vitro. Nucleic Acids Res 2007;35 (14):4905-13.

[30] Grädler U, Gerber HD, Goodenough-Lashua DM, Garcia GA, Ficner R, Reuter K, Stubbs MT, Klebe G. A new target for shigellosis: rational design and crystallographic studies of inhibitors of tRNA-guanine transglycosylase. J Mol Biol 2001;306 (3):455-67.

[31] Misra S, Crosby MA, Mungall CJ, Matthews BB, Campbell KS, Hradecky P, Huang Y, Kaminker JS, Millburn GH, Prochnik SE, Smith CD, Tupy JL, Whitfied EJ, Bayraktaroglu L, Berman BP, Bettencourt BR, Celniker SE, de Grey AD, Drysdale RA, Harris NL, Richter J, Russo S, Schroeder AJ, Shu SQ, Stapleton M, Yamada C, Ashburner M, Gelbart WM, Rubin GM, SE. L. Annotation of the Drosophila melanogaster euchromatic genome: a systematic review. Genome Biol. 2002;3 (12):RESEARCH0083.

[32] Kawai J SA, Shibata K, Yoshino M, Itoh M, Ishii Y, Arakawa T, Hara A, Fukunishi Y, Konno H, Adachi J, Fukuda S, Aizawa K, Izawa M, Nishi K, Kiyosawa H, Kondo S, Yamanaka I, Saito T, Okazaki Y, Gojobori T, Bono H, Kasukawa T, Saito R, Kadota K, Matsuda H, Ashburner M, Batalov S, Casavant T, Fleischmann W, Gaasterland T, Gissi C, King B, Kochiwa H, Kuehl P, Lewis S, Matsuo Y, Nikaido I, Pesole G, Quackenbush J, Schriml LM, Staubli F, Suzuki R, Tomita M, Wagner L, Washio T, Sakai K, Okido T, Furuno M, Aono H, Baldarelli R, Barsh G, Blake J, Boffelli D, Bojunga N, Carninci P, de Bonaldo MF, Brownstein MJ, Bult C, Fletcher C, Fujita M, Gariboldi M, Gustincich S, Hill D, Hofmann M, Hume DA, Kamiya M, Lee NH, Lyons P, Marchionni L, Mashima J, Mazzarelli J, Mombaerts P, Nordone P, Ring B, Ringwald M, Rodriguez I, Sakamoto N, Sasaki H, Sato K, Schönbach C, Seya T, Shibata Y, Storch KF, Suzuki H, Toyo-oka K, Wang KH, Weitz C, Whittaker C, Wilming L, Wynshaw-Boris A, Yoshida K, Hasegawa Y, Kawaji H, Kohtsuki S, Hayashizaki Y; RIKEN Genome Exploration Research

Group Phase II Team and the FANTOM Consortium. Functional annotation of a full-length mouse cDNA collection. Nature 2001;409 (6821):685-90.

[33] Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, Thompson JD. Multiple sequence alignment with the Clustal series of programs. Nucleic Acids Res 2003;31 (13):3497-500.

[34] Marti-Renom MA, Stuart A, Fiser A, Sánchez R, Melo F, Sali A. Comparative protein structure modeling of genes and genomes. Annu. Rev. Biophys. Biomol. Struct. 2000;29:291-325.

[35] Eswar N, Marti-Renom MA, Webb B, Madhusudhan MS, Eramian D, Shen M, Pieper U, Sali A. Comparative Protein Structure Modeling With MODELLER. Current Protocols in Bioinformatics 2000;Supplement 15:5.6.1-5.6.30.

[36] Romier C, Ficner R, Reuter K, Suck D. Purification, crystallization, and preliminary x-ray diffraction studies of tRNA-guanine transglycosylase from Zymomonas mobilis. Proteins 1996;24 (4):516-9.

[37] Hoops GC, Townsend LB, Garcia GA. tRNA-guanine transglycosylase from Escherichia coli: structure-activity studies investigating the role of the aminomethyl substituent of the heterocyclic substrate PreQ1. Biochemistry 1995;34 (46):15381-7.

[38] Stengl B, Meyer EA, Heine A, Brenk R, Diederich F, Klebe G. Crystal Structures of tRNA-guanine Transglycosylase (TGT) in Complex with Novel and Potent Inhibitors Unravel Pronounced Induced-fit Adaptations and Suggest Dimer Formation Upon Substrate Binding. J Mol Biol 2007;370 (3):492-511.

[39] Verdonk ML, Cole JC, Hartshorn MJ, Murray CW, Taylor RD. Improved Protein-Ligand Docking Using GOLD. Proteins 2003;52:609-23.

[40] Cole JC, Nissink JWM, Taylor R. Protein-Ligand Docking and Virtual Screening with GOLD. Virtual Screening in Drug Discovery 2005.

[41] Velec HF, Gohlke H, Klebe G. DrugScore(CSD)-knowledge-based scoring function derived from small molecule crystal data with superior recognition rate of near-native ligand poses and better affinity prediction. J Med Chem 2005;48 (20):6296-303.

[42] Gerber PR, Müller K. MAB, a generally applicable molecular force field for structure modelling in medicinal chemistry. J Comput Aided Mol Des 1995;9 (3):251-68.

[43] J. Liang HE, P. Fu, P.V. Sudhakar S. Subramaniam. Analytical shape computing of macromolecules I: molecular area and volume through alpha shape. Proteins 1998;33:1-17.

[44] Howes NK, Farkas WR. Studies with a homogeneous enzyme from rabbit erythrocytes catalyzing the insertion of guanine into tRNA. J Biol Chem 1978;253 (24):9082-7.

[45] Walden TL, Jr., Howes N, Farkas WR. Purification and properties of guanine, queuine-tRNA transglycosylase from wheat germ. J Biol Chem 1982;257 (22):13218-22.

[46] Seno T, Kobayashi M, Nishimura S. Purification of Excherichia coli methionine tRNAF and methionine tRNAM and studies on their biophysical and biochemical properties. Biochim Biophys Acta 1968;169 (1):80-94.

[47] Pearson RL, Weiss JF, Kelmers AD. Improved separation of transfer RNA's on polychlorotrifuoroethylene-supported reversed-phase chromatography columns. Biochim Biophys Acta 1971;228 (3):770-4.

[48] Saneyoshi M, Nishimura S. Selective inactivation of amino acid acceptor and ribosome-binding activities of Escherichia coli tRNA by modification with cyanogen bromide. Biochim Biophys Acta 1971;246 (1):123-31.

[49] Boeckmann B, Bairoch A, Apweiler R, Blatter MC, Estreicher A, Gasteiger E, Martin MJ, Michoud K, O'Donovan C, Phan I, Pilbout S, Schneider M. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. Nucleic Acids Res 2003;31 (1):365-70.

[50] Curnow AW, Kung FL, Koch KA, Garcia GA. tRNA-guanine transglycosylase from Escherichia coli: gross tRNA structural requirements for recognition. Biochemistry 1993;32 (19):5239-46.

[51] Meyer EA, Brenk R, Castellano RK, Furler M, Klebe G, Diederich F. De novo design, synthesis, and in vitro evaluation of inhibitors for prokaryotic tRNA-guanine transglycosylase: a dramatic sulfur effect on binding affinity. Chembiochem 2002;3 (2-3):250-3.

[52] Leatherbarrow R. GraFit 4.09 edit. Erithacus Software Limited, USA 1999.

[53] Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. 1997;276:307-26.

[54] Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL. Crystallography and NMR system: A new software suite for macromolecular structure determination. Acta Crystallogr. sect. D 1998;54:905-21.

[55] Sheldrick GM, Schneider TR. SHELXL: high-resolution refinement. Methods Enzymol. 1997;277b:319-43.

[56] Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallogr. sect. D 2004;60:2126-32.

[57] Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Crystallogr. 1993;26:283-91.

[58] Case DA, Cheatham TE, Darden T, Gohlke H, Luo R, Merz KM, Onufriev A, Simmerling C, Wang B, Woods RJ. The Amber biomolecular simulation programs. Journal of Computational Chemistry 2005;26 (16):1668-88.

[59] Jorgensen WL, Chandrasekhar J, Madura JD, Impey WI, Klein ML. Comparison of simple potential functions for simulating liquid water. The Journal of Chemical Physics 1983;79 (2):926-35.

[60] Darden T, York D, Pederson L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. The Journal of Chemical Physics 1993;98 (12):10089-92.

[61] Humphrey W, Dalke A, Schulten K. VMD: Visual molecular dynamics. Journal of Molecular Graphics 1996;14 (1):33-&.

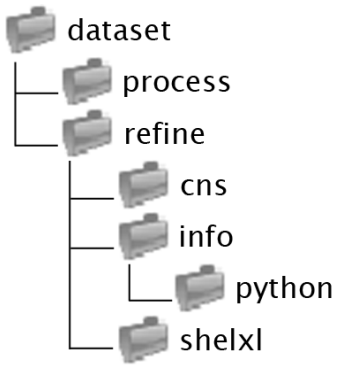
5 Appendix

5.1 X-ray refinement Toolkit - Manual

This toolkit is a python [1] based script set, written to assist the setup of X-ray refinement cycles. The respective scripts extract all necessary values from log files, subsequently edit input files for the following refinement cycle and start the corresponding refinement programs. Afterwards, the new log files are checked and several values are extracted and written to a summary file (**info.txt**, File 2, p. 122). The script is adjusted to the refinement strategy commonly applied in our group: Data processing should be performed with HKL2000 [2], rigid body refinement and simulated annealing procedures with cns [3], and refinement with either cns or shelxl [4]. Generated electron density maps are converted by the scripts for a direct read into coot [5]. This toolkit is currently available under **/nfs/extern/pub/inhouse/(incoming/) tidten/**.

5.1.1 Arrangement of folders and subfolders

Figure 25: Folder arrangement required by the toolkit.



A short README file describes the usage of the toolkit. For proper performance of the toolkit, a certain arrangement of folders and subfolders has to be maintained (Figure 25). Images of the dataset can be stored in any directory. However, the HKL2000 processing log file and reflections file (**scaleui.log**, **output.sca**) have to be available under **dataset/process/**. After having processed the dataset, and thus having generated the folder **dataset/process/**, the script **make_dir.py** can be applied to generate the folder **dataset/refine/** with its subfolders.

Additionally, the python toolkit files are automatically copied to **dataset/refine/python/**.

dataset/refine/ contains three subfolders: **cns/** (for rigid body refinement and simulated annealing procedures performed with cns, and cns refinement cycles), **shelxl/** (for shelxl refinement cycles), **info/** (containing the toolkit files in the subfolder **python/** and the summary file (**info.txt**, File 2, p. 122), where important information is entered during the usage of the toolkit).

5.1.2 default.txt

Before the toolkit is used, several files have to be prepared. The HKL2000 files (**scaleui.log**, **output.sca**) have to be present in **dataset/process/** as described above. The file **default.txt** (File 1) provided by the toolkit, has to be edited and stored under **dataset/refine/info/**. In **default.txt** names of files, which are read while the script runs, can be changed if desired. The 'short name' of protein and ligand is used to generate the file names of each refinement cycle (*e.g.* **01-prot-lig.ins**).

File 1: **default.txt**

```
Please change default-file-names here:

CNS:
pdb-file, coordinates for MR:           start.pdb
sca-file, HKL2000 output:               output.sca
log-file, HKL2000:                      scaleui.log
short name of your protein:             prot
short name of your ligand:              lig

SHELXL:
dfx-file, inserted in shelxl-ins-file:    GOL.dfx
start- and end-aa in refinement-model:    11-383
seq-file, sequence of your whole protein: prot.seq
```

5.1.3 Ligand restraints (shelxl dfx file)

A shelxl restraint file for glycerol molecules is provided by the toolkit (**GOL.dfx**). For any other ligand molecules, restraints have to be manually generated utilising the shelxpro option **[J]** and may then be simply appended to the **GOL.dfx** file in **dataset/refine/info/python/**.

5.1.4 Protein sequence file

A file containing the (entire) one letter amino acid sequence of the protein has to be present in **dataset/refine/info/**. This file is used to generate additional output files after a shelxl run as described later. Storing the entire protein sequence instead of only the refined residues has the advantage that during PDB deposition of the final structure, the complete sequence is already available.

5.1.5 start.pdb

A pdb file containing coordinates for rigid body refinement and simulated annealing procedures performed by cns has to be provided under **dataset/ refine/cns/** and named as declared in **default.txt** (**start.pdb**).

5.1.6 First cns run

Subsequently, structure solution can be started from **dataset/refine/cns/** utilising the script **cns1.py**. This script first requests data collection information, such as date of data collection, detector type, exposure time, detector distance, oscillation range, radiation wavelength. These values have to be declared upon PDB deposition. Thus, correct completion is recommended but not necessary for proper function of the toolkit, except

File 2: info.txt

DATA:

date-name:			070316-vac-boc
detector	t	dist.	phi	lambda
marccd	5", 3"	70,150	160-290	0.97803

	all	highest resol.	lowest resol.
Rsym	3.7 %	23.7 %	3.3 %
I/sig	34	2.4	52
redundancy	3.6	--	--
completeness	93.7 %	92.2 %	87.9 %
mosaicity/resolution	0.911°	1.34 Å	20- 3.3 Å

	tot. no.	unique	rej./excl.
frames	500	--	0
reflexes	1542926	300004	0

cell/SpGr	a: 89.949	b: 64.849	c: 70.471
C2	alpha: 90.0	beta: 95.56	gamma: 90.0

##
##

REFINEMENT:

Name (cns)	Rfree	Rwork	ΔR
00-vac-boc	26.84 %	26.46 %	0.4 %

##

Name (shelxl)	ANIS?	HFIX?	WGHT
01-vac-boc	False	True	0.1

Cycles	aa	HOH	GOL	LIG
50	11-383	0	0	2

double conf: 9	ALA: 35	gaps: 39

shelxl	Rfree	R1
Fo > 4sig(Fo)	28.09 %	24.36 %
all, ΔR: 3.4 %	28.37 %	24.96 %
ratio/merge	0.9921	25.01 %

##

the radiation wavelength, which is essentially required and inserted into input files later on. Thereafter, processing data, such as R_{sym}, I/σ, resolution, number of reflections, unit cell, are extracted from HKL2000 files and written to the summary file (**info.txt**, File 2, p. 122) in **dataset/refine/info/**.

Before cns runs are started by the script, the corresponding input files (**make_cv.inp**, **generate.inp**, **rigid.inp**, **refine.inp**) are edited according to the previously extracted information. Additionally, the HKL2000 reflection file **output.sca** is copied to the folder **dataset/refine/cns/** and converted to the cns compatible xpl format. The log files written by the cns runs are read and the script is aborted if errors are found, so that manual adjustment of input files or others (such as **start.pdb**) is possible. The subsequent start of the script **cns1.py** then utilises the manually edited cns input file from **dataset/refine/cns/**. cns runs are started in the following sequence: **make_cv**, **generate**, **rigid**, **refine**. After successful cns runs, maps generated by the **refine** run are converted to the coot compatible ccp4 format, R_{work}/R_{free} values are extracted from the log file and written to the summary file **info.txt**.

In general, the toolkit script suggests default answers (given in brackets behind the question), which can be applied by pressing the '**ENTER**' key. If a descision has to be made, the answer given in capitals is defined as the default answer (*e.g.* '**(y/N)**' means, if only the '**ENTER**' key is pressed, the adopted answer is '**N**').

5.1.7 Second and following cns runs

All following cns runs should be started by the script **cns2ff.py**. **cns2ff.py** first requests the name of the latest pdb file of the previous coot refinement and splits this file into separate pdb files containing protein atoms, ligand atoms, ions, water molecules, and glycerol molecules (declared as 'prosthetic

group'), respectively, if present. The cns input files **generate.inp**, **minimize.inp**, **bindividual.inp**, **water_pick.inp**, **model_map.inp** are edited and cns runs are started in the following sequence: **generate**, **minimze**(1), **bindividual**, **water_pick**, **minimize**(2), **model_map** (fofc), **model_map** (3fo2fc). If a non-protein, non-water, non-glycerol residue is present, the user is asked to decide whether the respective residue should be considered as an ion or as a ligand. In the first case, the standard ion restraints are added to the input files, while in the second case, the user is asked to provide name and path of a restraint file. After successful cns runs, maps are converted to the coot compatible ccp4 format, R_{work}/R_{free} values are extracted from the log file and written to the summary file **info.txt**.

5.1.8 First shelxl run

For high resolution datasets, the **cns2ff.py** script can be skipped, and **shelxl1.py** can be started from **dataset/refine/shelxl/**. This script first copies the last cns refinement pdb file and the reflection file to **dataset/refine/shelxl/** and generates the shelxl input file (**.ins**) as well as the hkl file (reflections, marked for the calculation of R_{free}, which were not used during cns refinement are maintained) by utilising shelxpro and a batch file. Subsequently, in a separate shell, the generated ins file may be edited with the shelxpro option **[U]**, *e.g.* for the refinement of double conformation occupancies or the transformation of water molecule coordinates to the nearest symmetry equivalent. Additional command lines may be inserted manually as well. If required, also a new ins file can be generated manually with the shelxpro option **[I]**, but the file name has to be kept consistently. The script is resumed by typing in '**y**' ("**Have you finished? (y)**"). In the next step, the ins file is edited by the script. The number of refinement cycles is changed (**CGLS** line) according to the user input, and '**-1**' is added automatically to the line for R_{free} calculation, '**ANIS**' for anisotropic refinement can be inserted, **HFIX** restraints can be switched on, and the **GOL.dfx**

restraints file is inserted. The **HKLF** code is set to **3** (F) per default (since the hkl file is generated from the cns **.cv** file, maintaining the set of reflections used for R_{free} calculation) but can be changed to **4** (F^2), if required. **HOPE** can be switched on and off as desired and the weighting factor (**WGHT**) can be changed from **0.1** (default value) to **0.2** for final refinement cycles. A shelxl run is then started by the script, writing a log file automatically. After a successful shelxl run, R_{work}/R_{free} values, as well as the number of refined water, ligand and glycerol molecules are written to the summary file **info.txt**. Additionally, the following files are generated: **01-prot-lig.lig**, **01-prot-lig.ala**, **01-prot-lig.gap**, **01-prot-lig.dbl**. The **.lig** file contains all residue names, which are considered as 'ligand' (non-protein, non-water, non-glycerol. Thus, also ions are counted in the respective field of **info.txt** and listed in the **.lig** file). The **.ala** file contains all residue numbers and original names, which are present as alanine in the refinement pdb file, but have a different side chain according to the sequence file (**dataset/refine/info/prot.seq**). The **.gap** file contains all missing residues, according to the sequence file, and is used to automatically generate '**HFIX 33**' lines in the next refinement cycle for the *N*-terminal end of the gap, if required. The **.dbl** file contains all double conformations and their corresponding refined occupancies. The respective total numbers of residues in gaps, with double conformations, artificial alanines, or ligands are listed as well in the summary file **info.txt**.

For huge datasets, an analogous **shelxh1.py** script is available, starting shelxh instead of shelxl. Nevertheless, the folder has to be named 'shelxl' anyway. The default answers differ slightly in **shelxh1.py**: **HOPE** is switched off, **HFIX** restraints are switched on and '**ANIS**' is inserted per default. These settings can be changed if desired.

5.1.9 Second and following shelxl runs

All following shelxl runs should be started with the script **shelxl2ff.py**. This script skips the file copying from cns performed in **shelxl1.py**, and instead changes the file name of the already existing hkl file corresponding to the new refinement cycle. In addition, all steps of the script **cleanup_pdb.py**, which is also seperately available (chapter 5.1.11, p. 128), are performed. Chain IDs (generated by coot *e.g.* upon molecule merging or sorting) are deleted, ligands and glycerols are sorted before the water molecules (the coot 'merge molecules' function usually appends glycerols to the end of the pdb file and thus after already refined water molecules). Water molecules suggested by the previous shelxl run exhibit '**W**' as atom names, which are not considered by the subsequent shelxpro generated **.ins** file if not changed to '**O**'. All water molecules have to be checked during coot refinement and deleted, if required. The remaining '**W**'-water molecules will be changed to '**O**', and thus be considered and refined in the next cycle. The remaining steps are performed analogously to **shelxl1.py**.

It has to be noted, that the pdb file from which the shelxl input file for the subsequent refinement cycles is generated, is always chosen by the user (per default, a pdb file with the ending **-coot-0.pdb** of the previous refinement cycle is selected by the script). Thus, if *e.g.* several variants starting from the same pdb file are tested during refinement, also pdb files from a refinement cycle before the previous one can be selected for continuation. The numbering of the refinement cycles will not be affected, but it is possible to add comments concerning the refinement procedure to the summary file at the right end of the lines, without disturbing the architecture of the tables. Such comments will be kept but not considered by the toolkit.

A respective **shelxh2ff.py** script is available as well.

5.1.10 Utilising the toolkit after a manual refinement

In addition, there are also two scripts provided, allowing the extraction of data processing and shelx refinement information of a refinement that had been performed manually before, or is even finished (**read_all_infos_of_finished_refinements.py**). A subsequent use of the toolkit is facilitated by **shelxl2ff_after_manual_refinement.py**.

First, for the use of **read_all_infos_of_finished_refinements.py**, it has to be assured, that for each **.ins** file **.log** and **.lst** files with corresponding names are available. If not, these **.ins** files have to be renamed (such as **.ins_**). The script will not consider them. Besides, a protein sequence file has to be provided in **dataset/refine/info/** (see also chapter 5.1.4, p. 121). **read_all_infos_of_finished_refinements.py** writes all information to a summary file in **dataset/refine/shelxl/info.txt** as already described above (chapter 5.1.8, p. 124) and **.lig**, **.ala**, **.gap** and **.dbl** files are generated for the corresponding refinement cycles. The sequence of refinement cycles is not obtained from the **.ins** file names but from the date and time on which the files had been generated.

If, subsequently, refinement is continued utilising the toolkit, the summary file has to be copied to **dataset/refine/info/**. Afterwards, the first shelxl run has to be started with the script **shelxl2ff_after_manual_refinement.py**. This script is able to accept pdb files with arbitrary file names. Furthermore, the user can specify the starting number of the subsequent refinement cycle (*e.g.* **06-prot-lig.ins**), to keep the overall numbering consistent.

All following shelxl runs should be started with the script **shelxl2ff.py** (chapter 5.1.9, p. 126).

5.1.11 Additional scripts

analize_pdb.py – Counts number of protein atoms (considering atom occupancies of double conformations), water molecules, glycerol atoms and ligand atoms.
– Calculates mean B-values of protein atoms, water molecules, glycerol atoms and ligand atoms.
– Values are directly displayed within the shell and can be used for crystallographic tables in publications.

cleanup_folder.py – Deletes all disc space consuming map files except the newest (fcf files are never deleted).
– Deletes all intermediate coot-pdb-files (such as **01-prot-log-coot-5.pdb**) except the newest.
– Deletes all session files and folders generated by coot.
– gzips all files within the present folder.

cleanup_pdb.py – Deletes chain IDs within a pdb file generated by coot *e.g.* upon molecule merging or sorting.
– Sorts ligands and glycerols before water molecules (the coot 'merge molecules' function usually appends glycerols to the end of the pdb file and thus after already refined water molecules).
– Renames water molecule atom names from '**W**' (suggested water molecules by shelxl) to '**O**'.

make_dir.py – Generates the **refine** folder and its subfolders which are required by the toolkit.
– Copies the toolkit files to **dataset/refine/info/python/**.

5.1.12 Template files provided by the toolkit

bindividual.inp cns input file.

default.txt see chapter 5.1.2, p. 120.

generate.inp cns input file.

GOL.dfx shelx restraint file for glycerol molecules.

info-leer.txt empty **info.txt** file, may be edited in the first three lines and stored as **dataset/refine/info/info.txt** before the first run of the toolkit. In this case inserted values are read in automatically by the script and the user is not asked during the usage of **cns1.py** (chapter 5.1.6, p. 121).

make_cv.inp cns input file.

minimize.inp cns input file.

model_map.inp cns input file.

refine.inp cns input file.

rigid.inp cns input file.

tgt.seq excample file, containing the (entire) one letter amino acid sequence of tRNA guanine transglycosylase (TGT).

water_pick.inp cns input file.

5.2 Multiple Sequence alignments

Amino acid residues are coloured by their properties, according to the Clustal W colour code: red – hydrophobic, blue – acidic, magenta – basic, green – hydrophilic. consensus symbols: * – residue identical in all sequences in the alignment, : – conserved substitutions, . – semi-conserved substitutions.

5.2.1 Multiple sequence alignment of human TGT with selected bacterial TGT structures

ClustalW 1.83 [6] multiple sequence alignment of human TGT (hTGT) with selected bacterial TGT structures as template for homology modelling. Three wild-type *Zymomonas mobilis* TGT structures (PDB codes: 1r5y, 1y5v, 1y5w – resolutions: 1.20 Å and 1.58 Å, respectively), two Y106F variant structures (PDB codes: 1ozm, 1ozq – resolutions: 1.95 Å and 1.90 Å, respectively), and one structure of a covalent TGT-tRNA intermediate (PDB code: 1q2r (chain A and B) – resolution: 2.90 Å) were chosen.

```
1R5Y.pdb_A      RPRFSFSIAAREGKARTGTIEMKRGVIRTPAFMPVGTAATVKALKPETVRATGADIILGN 60
1Y5W.pdb_A      RPRFSFSIAAREGKARTGTIEMKRGVIRTPAFMPVGTAATVKALKPETVRATGADIILGN 60
1Q2R.pdb_A      RPRFSFSIAAREGKARTGTIEMKRGVIRTPAFMPVGTAATVKALKPETVRATGADIILGN 60
1Y5V.pdb_A      RPRFSFSIAAREGKARTGTIEMKRGVIRTPAFMPVGTAATVKALKPETVRATGADIILGN 60
1Q2R.pdb_B      RPRFSFSIAAREGKARTGTIEMKRGVIRTPAFMPVGTAATVKALKPETVRATGADIILGN 60
1OZQ.pdb_A      RPRFSFSIAAREGKARTGTIEMKRGVIRTPAFMPVGTAATVKALKPETVRATGADIILGN 60
1OZM.pdb_A      RPRFSFSIAAREGKARTGTIEMKRGVIRTPAFMPVGTAATVKALKPETVRATGADIILGN 60
hTGT            ---MRLVAECSRSRARAGELWLPHGTVATPVFMPVGTQATMKGITTEQLDALGCRICLGN 57
                      :     .:** * : : :*.: **.****** *::* :... :   *. * *.*

1R5Y.pdb_A      TYHLMLRPGAERIAKLGGLHSFMGWDRPILTDSGGYQVM--LSLTKQSEEGVTF------ 112
1Y5W.pdb_A      TYHLMLRPGAERIAKLGGLHSFMGWDRPILTDSGGYQV------TKQSEEGVTF------ 108
1Q2R.pdb_A      TYHLMLRPGAERIAKLGGLHSFMGWDRPILTDSGGYQVMSLSSLTKQSEEGVTFKSHLDG 120
1Y5V.pdb_A      TYHLMLRPGAERIAKLGGLHSFMGWDRPILTDSGGYQVM-----TKQSEEGVTFK----- 110
1Q2R.pdb_B      TYHLMLRPGAERIAKLGGLHSFMGWDRPILTDSGGYQVMSLSSLTKQSEEGVTFKSHLDG 120
1OZQ.pdb_A      TYHLMLRPGAERIAKLGGLHSFMGWDRPILTDSGGFQVMSLSSLTKQSEEGVTFKSHLDG 120
1OZM.pdb_A      TYHLMLRPGAERIAKLGGLHSFMGWDRPILTDSGGFQVMSLSSLTKQSEEGVTFKSHLDG 120
hTGT            TYHLGLRPGPELIQKANGLHGFMNWPHNLLTDSGGFQMVSLVSLSEVTEEGVRFRSPYDG 117
                *:** *:** * *   ***.**.* : :******:*:     .  :*** *

1R5Y.pdb_A      ---MLSPERSIEIQHLLGSDIVMAFDECTPYPATPSRAASSMERSMRWAKRSRDAFDSRK 169
1Y5W.pdb_A      ---MLSPERSIEIQHLLGSDIVMAFDECTPYPATPSRAASSMERSMRWAKRSRDAFDSRK 165
1Q2R.pdb_A      SRHMLSPERSIEIQHLLGSDIVMAFDECTPYPATPSRAASSMERSMRWAKRSRDAFDSRK 180
1Y5V.pdb_A      --HMLSPERSIEIQHLLGSDIVMAFDECTPYPATPSRAASSMERSMRWAKRSRDAFDSRK 168
1Q2R.pdb_B      SRHMLSPERSIEIQHLLGSDIVMAFDECTPYPATPSRAASSMERSMRWAKRSRDAFDSRK 180
1OZQ.pdb_A      SRHMLSPERSIEIQHLLGSDIVMAFDECTPYPATPSRAASSMERSMRWAKRSRDAFDSRK 180
1OZM.pdb_A      SRHMLSPERSIEIQHLLGSDIVMAFDECTPYPATPSRAASSMERSMRWAKRSRDAFDSRK 180
hTGT            NETLLSPEKSVQIQNALGSDIIMQLDDVVSSTVTGPRVEEAMYRSIRWLDRCIAAHQR-- 175
                   :*.** *:::*  ***** * :*.    .  .. .:  *: **  *.  *..
```

```
1R5Y.pdb_A      EQAENAALFGIQQGSVFENLRQQSADALAEIGFDGYAVGGLAVGEGQDEMFRVLDFSVPM 229
1Y5W.pdb_A      EQAENAALFGIQQGSVFENLRQQSADALAEIGFDGYAVGGLAVGEGQDEMFRVLDFSVPM 225
1Q2R.pdb_A      EQAENAALFGIQQGSVFENLRQQSADALAEIGFDGYAVGGLAVGEGQDEMFRVLDFSVPM 240
1Y5V.pdb_A      EQAENAALFGIQQGSVFENLRQQSADALAEIGFDGYAVGGLAVGEGQDEMFRVLDFSVPM 228
1Q2R.pdb_B      EQAENAALFGIQQGSVFENLRQQSADALAEIGFDGYAVGGLAVGEGQDEMFRVLDFSVPM 240
1OZQ.pdb_A      EQAENAALFGIQQGSVFENLRQQSADALAEIGFDGYAVGGLAVGEGQDEMFRVLDFSVPM 240
1OZM.pdb_A      EQAENAALFGIQQGSVFENLRQQSADALAEIGFDGYAVGGLAVGEGQDEMFRVLDFSVPM 240
hTGT            --PDKQNLFAIIQGGLDADLRATCLEEMTKRDVPGFAIGGLSGGESKSQFWRMVALSTSR 233
                    ::  **.* **.:  :**  .   ::. .. *:*:***: ** :.  .:  .:.

1R5Y.pdb_A      LPDDKPHYLMGVGKPDDIVGAVERGIDMFDCVLPTRSGRNGQAFTWDGPINIRNARFSED 289
1Y5W.pdb_A      LPDDKPHYLMGVGKPDDIVGAVERGIDMFDCVLPTRSGRNGQAFTWDGPINIRNARFSED 285
1Q2R.pdb_A      LPDDKPHYLMGVGKPDDIVGAVERGIDMFDCVLPTRSGRNGQAFTWDGPINIRNARFSED 300
1Y5V.pdb_A      LPDDKPHYLMGVGKPDDIVGAVERGIDMFDCVLPTRSGRNGQAFTWDGPINIRNARFSED 288
1Q2R.pdb_B      LPDDKPHYLMGVGKPDDIVGAVERGIDMFDCVLPTRSGRNGQAFTWDGPINIRNARFSED 300
1OZQ.pdb_A      LPDDKPHYLMGVGKPDDIVGAVERGIDMFDCVLPTRSGRNGQAFTWDGPINIRNARFSED 300
1OZM.pdb_A      LPDDKPHYLMGVGKPDDIVGAVERGIDMFDCVLPTRSGRNGQAFTWDGPINIRNARFSED 300
hTGT            LPKDKPRYLMGVGYATDLVVCVALGCDMFDCVFPTRTARFGSALVPTGNLQLRKKVFEKD 293
                 **.***:*:** * .  ::  *  * ******:*** .* * *:.  * ::::   ...

1R5Y.pdb_A      LKPLDSECHCAVCQKWSRAYIHHLIRAGEILGAMLMTEHNIAFYQQLMQKIRDSISEGRF 349
1Y5W.pdb_A      LKPLDSECHCAVCQKWSRAYIHHLIRAGEILGAMLMTEHNIAFYQQLMQKIRDSISEGRF 345
1Q2R.pdb_A      LKPLDSECHCAVCQKWSRAYIHHLIRAGEILGAMLMTEHNIAFYQQLMQKIRDSISEGRF 360
1Y5V.pdb_A      LKPLDSECHCAVCQKWSRAYIHHLIRAGEILGAMLMTEHNIAFYQQLMQKIRDSISEGRF 348
1Q2R.pdb_B      LKPLDSECHCAVCQKWSRAYIHHLIRAGEILGAMLMTEHNIAFYQQLMQKIRDSISEGRF 360
1OZQ.pdb_A      LKPLDSECHCAVCQKWSRAYIHHLIRAGEILGAMLMTEHNIAFYQQLMQKIRDSISEGRF 360
1OZM.pdb_A      LKPLDSECHCAVCQKWSRAYIHHLIRAGEILGAMLMTEHNIAFYQQLMQKIRDSISEGRF 360
hTGT            FGPIDPECTCPTCQKHSRAFLHALLHSDNTAALHHLTVHNIAYQLQLMSAVRTSIVEKRF 353
                 : *:* .* * .*:: :*:::* *:  .:  .   :* *** :  .** . :* ** .  *

1R5Y.pdb_A      SQFAQDFRARYF---------------------- 361
1Y5W.pdb_A      SQFAQDFRARYF---------------------- 357
1Q2R.pdb_A      SQFAQDFRARYFARNS------------------ 376
1Y5V.pdb_A      SQFAQDFRARYF---------------------- 360
1Q2R.pdb_B      SQFAQDFRARYF---------------------- 372
1OZQ.pdb_A      SQFAQDFRARYF---------------------- 372
1OZM.pdb_A      SQFAQDFRARYF---------------------- 372
hTGT            PDFVRDFMGAMYGDPTLCPTWATDALASVGITLG 387
                 ::  **
```

5.2.2 Multiple sequence alignment of eukaryotic and bacterial TGTs

Residues in proximity of the $preQ_1$ substrate are listed. Frequencies of the corresponding residues found in a multiple sequence alignment of 100 eukaryotic, and bacterial TGT sequences (from SWISS-PROT [7]), respectively, are given in percent.

Z. mobilis / *H. sapiens* #	Eukaryotic TGTs														Bacterial TGTs								
D102 / D89	D	C	--	L	A	F	G	I	Y						D								
	69	10	7	5	4	2	1	1	1						100								
S103 / S90	S	--	L	A	G	R	V	F	Q	T					S								
	78	7	3	3	2	2	2	1	1	1					100								
Y106 / F93	F	D	--	Y	A	Q	T	S	R						F	Y							
	65	18	6	5	2	1	1	1	1						99	1							
D156 / D143	D	A	S	Y	--	C	E	N	V	F					D								
	70	9	7	4	4	2	1	1	1	1					100								
C158 / V145	V	G	C	E	L	--	A	Q	W	Y	H	S			C	V							
	59	11	7	6	5	4	2	2	1	1	1	1			99	1							
Y161 / S148	T	S	-	A	V	D	W	N	G	M	I	P	L	K	Y	F	M	L	G	N	W	R	H
	28	22	19	11	6	3	2	2	2	1	1	1	1	1	56	19	8	7	4	3	1	1	1
I201 / I184	I	V	A	P	R	C	T	N	S	K	--				I								
	69	15	5	4	1	1	1	1	1	1	1				100								
Q203 / Q186	Q	E	V	H	T	S	L	R	--						Q	H							
	67	11	9	5	3	2	1	1	1						98	2							
G229 / G212	G	D	N	--	S	H	L	Y							G								
	79	10	3	3	2	1	1	1							100								
G230 / G213	G	--	L	N	F	Q	D								G								
	91	3	2	1	1	1	1								100								
L231 / L214	L	F	S	C	P	V	G	Y	T						L	V							
	69	15	6	5	1	1	1	1	1						99	1							
A232 / S215	S	L	A	Q	C	G	H	--	R	T	F	Y			A	S							
	51	11	10	7	5	5	4	3	1	1	1	1			63	37							
V233 / G216	G	--	S	A	V	Q	N	T							V								
	79	9	3	3	2	2	1	1							100								
M260 / M243	M	L	--	F	V	I	Y	H	G	E	C	P			M								
	62	18	4	3	3	3	2	1	1	1	1	1			100								
G261 / G244	G	I	V	L	--	D	R	N	M	C					G								
	70	9	5	5	5	2	1	1	1	1					100								

5.2.3 Multiple sequence alignment of IpaA and IpgE homologues

a) IpaA sequence (line 3) aligned with fragments of *Salmonella typhimurium* SipA (amino acids 48 – 264) from PDB codes 2fm9, chain A (line 1) and 2fm8, chain C (line 2).

```
2FM9_A          ------------------------------------------------PQLEDFPALIKQ 12
2FM8_C          ----------------------QATNLAANLSAVRESATATLSGEIKGPQLEDFPALIKQ 38
IpaA            MHNVNNTQAPTFLYKATSPSSTEYSELKSKISDIHSSQTS-LKTPASVSEKENFATSFNQ 59
                                                        .: *:*.: ::*

2FM9_A          ASLDALFKCGKDAEALKEVFTNSNNVAGKKAIMEFAGLFRSALNATSDSPEAKTLLMKVG 72
2FM8_C          ASLDALFKCGKDAEALKEVFTNSNNVAGKKAIMEFAGLFRSALNATSDSPEAKTLLMKVG 98
IpaA            KCLDFLFSSSGKEDVLRSIYSNSMNAYAKSEILEFSNVLYSLVHQNGLNFENEKGLQKIV 119
                 .** **... . :.*:.:::** *. .*. *:**:.:: * :: .. . * :. * *:

2FM9_A          AEYTAQIIKDGLKEKSAFGPWLPETKKAEAKLENLEKQLLDIIKNNTGGELSKLSTNLVM 132
2FM8_C          AEYTAQIIKDGLKEKSAFGPWLPETKKAEAKLENLEKQLLDIIKNNTGGELSKLSTNLVM 158
IpaA            AQYSELIIKDKLSQDSAFGPWSAKNKKLHQLRQNIEHRLALLAQQHTSGEALSLGQKLLN 179
               *:*:  **** *.:.****** .:.** .   :*:*::*  : :::*.**  .*. :*:

2FM9_A          QEVMPYIASCIEHNFGCTLDPLTRSNLTHLVDKAAAKAVEALDMCHQKLTQEQGTSVGRE 192
2FM8_C          QEVMPYIASCIEHNFGCTLDPLTRSNLTHLVDKAAAKAVEALDMCHQKLTQEQGTSVGRE 218
IpaA            TEVSSFIKNNILAELKLSNETVSSLKLDDLVDAQAKLAFDSLRNQRKNTIDSKGFGIGKL 239
                ** .:* . *  ::  : :.::  :* .***  *  *.::*   :::  :.:* .:*:

2FM9_A          ARHLEMQTLIPLLLRNVFAQIPA------------------------------------- 215
2FM8_C          ARHLEMQTLIPLLLRNVFAQIP-------------------------------------- 240
IpaA            SRDLNTVAVFPELLRKVLNDILEDIKDSHPIQDGLPTPPEDMPDGGPTPGANEKTSQPVI 299
               :*.*:  :::* ***:*: :*
```

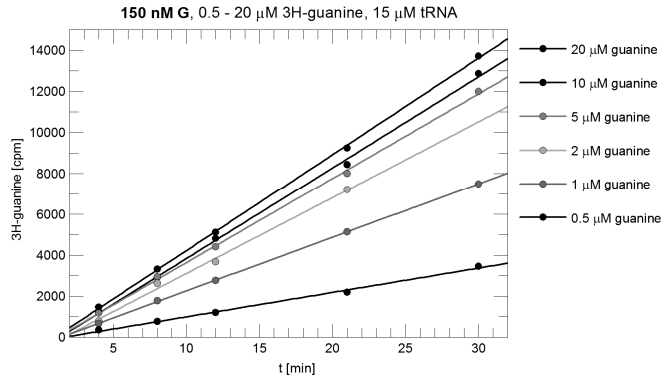
b) IpgE sequence (line 3) aligned with the *Salmonella enterica* TTS chaperone SigE from PDB codes 1k3s, chain A (line 1) and chain B (line 2).

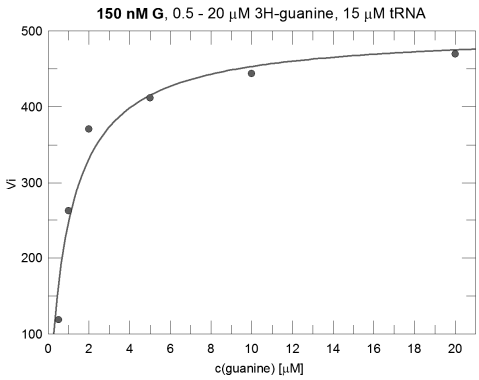
```
1K3S_A          MESLLNRLYDALG-----LDEPLLIIDDGIQVYFNESDHTLEMCCPFMPLPDDILTLQHF 55
1K3S_B          -ESLLNRLYDALG-----LDAPLLIIDDGIQVYFNESDHTLEMCCPFMPLPDDILTLQHF 54
IpgE            MEDLADVICRALGIPLIDIDDQAIMLDDDVLIYIEKEGDSINLLCPFCALPENINDLIYA 60
                *.* : :  ***     :*   :::**.: :*:::...:::: *** .**::*  * :

1K3S_A          LRLNYTSAVTIGADADNTALVALYRLPQTSTEEEALTGFELFISNVKQLKEHYA------ 109
1K3S_B          LRLNYTSAVTIGADADNTALVALYRLPQTSTEEEALTGFELFISNVKQLKEH-------- 106
IpgE            LSLNYSEKICLATDDEGGNLIARLDLTGINEFEDVYVNTEYYISRVRWLKDEFARRMKGY 120
               * ***:. : :.:* :.  *:*   *.  .  *:. .. * :**.*: **:.
```

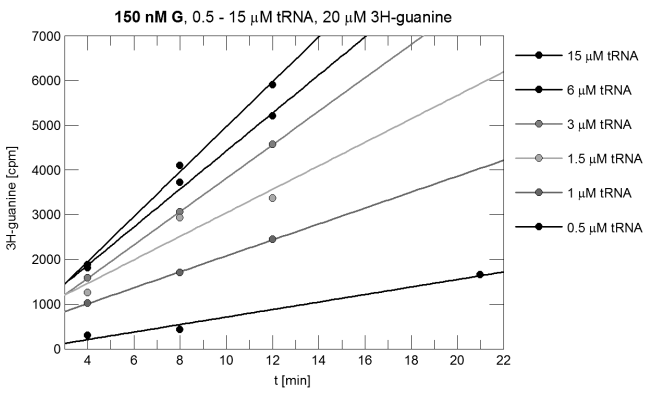
5.3 Kinetic measurements

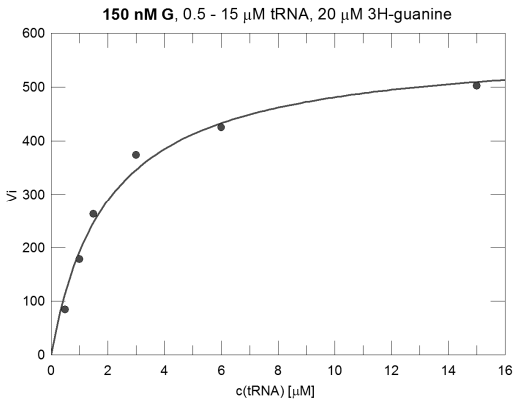
5.3.1 [8-^3H]-guanine TGT assay

a) Initial velocities plotted against time, Michaelis Menten curve of Tyr^{106}Phe / Val^{233}Gly TGT variant for [8-^3H]-guanine and tRNA^{Tyr}.

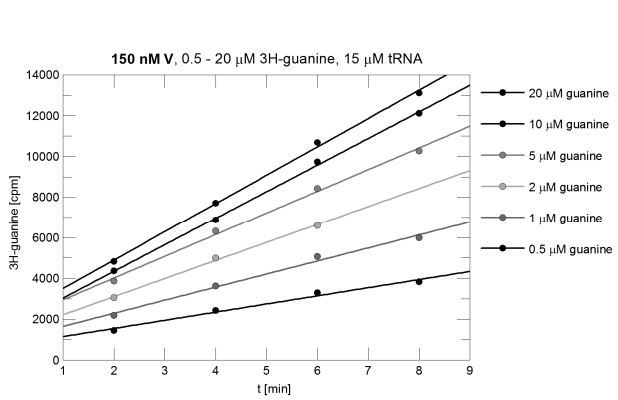


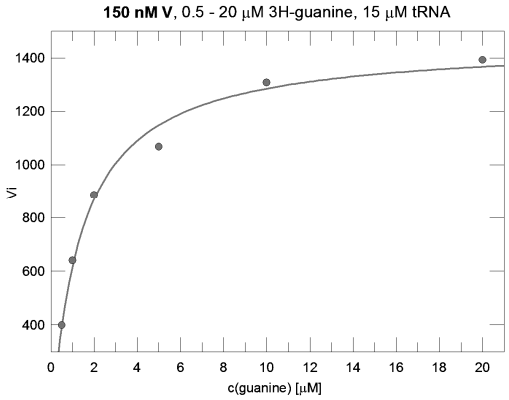


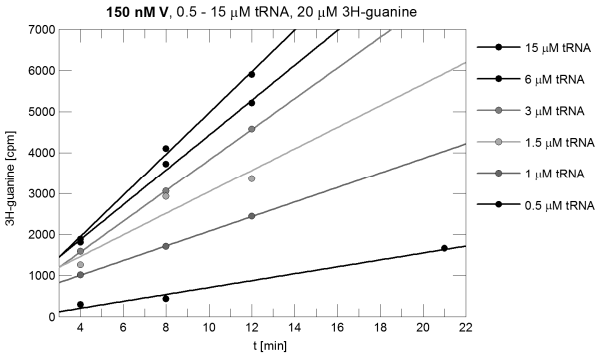


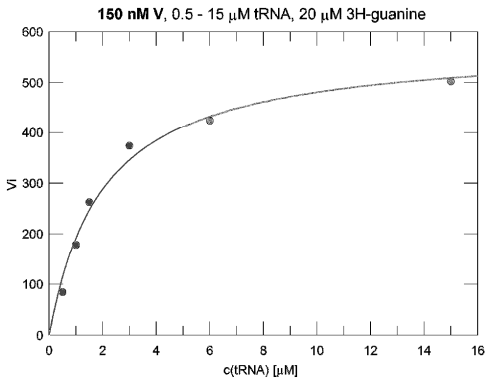


b) Initial velocities plotted against time, Michaelis Menten curve of Tyr[106]Phe / Cys[158]Val TGT variant for [8-^3H]-guanine and tRNA^{Tyr}.

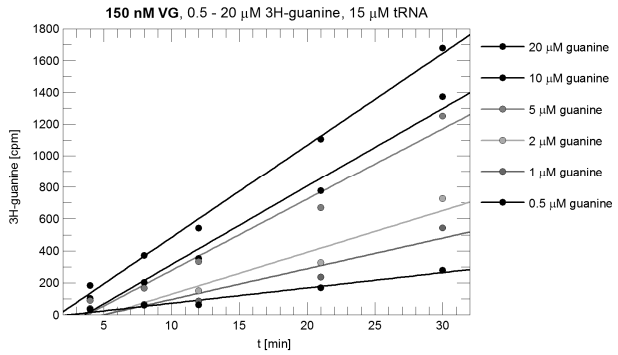


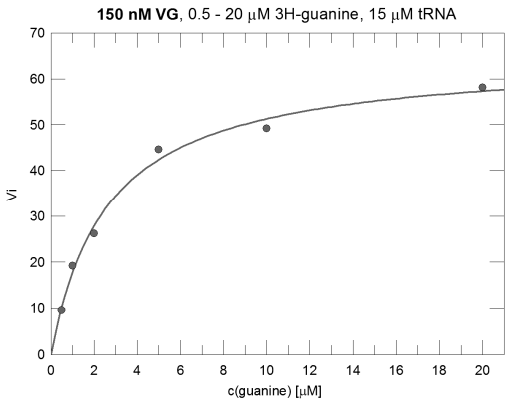


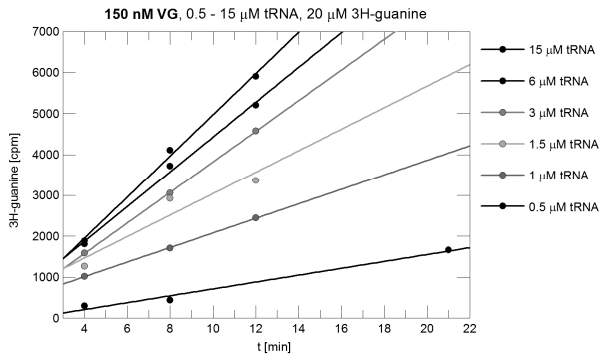


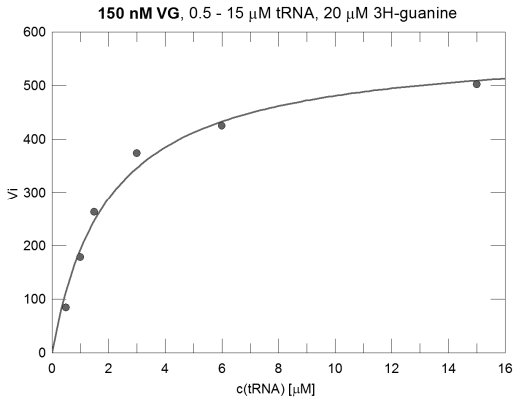


c) Initial velocities plotted against time, Michaelis Menten curve of Tyr[106]Phe / Cys[158]Val /Val[233]Gly TGT variant for [8-^3H]-guanine and tRNA^{Tyr}.

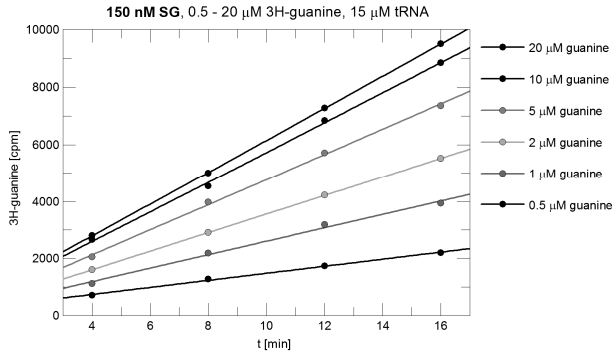


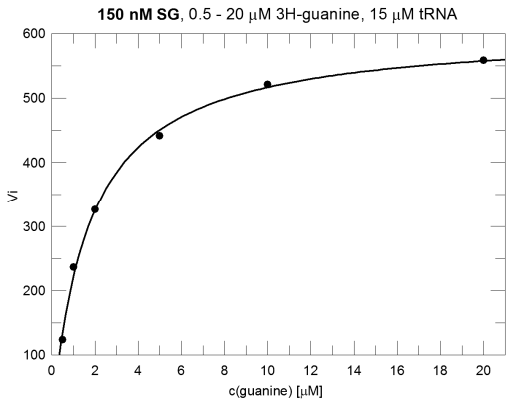


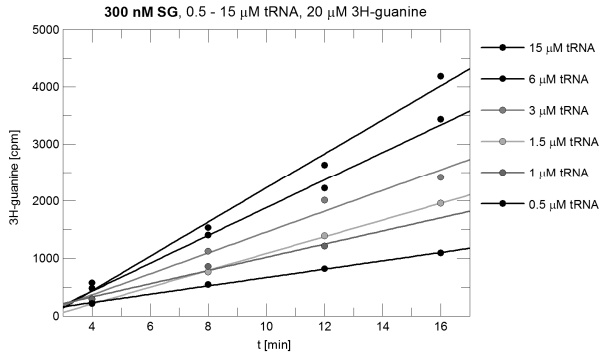


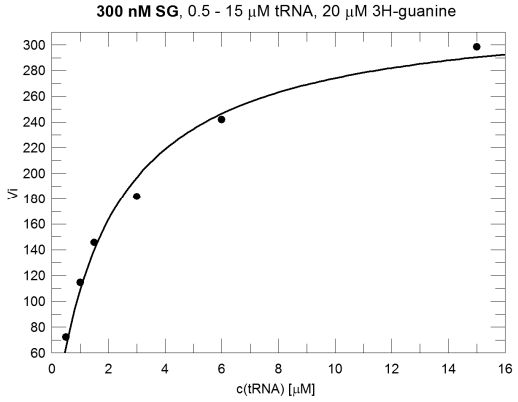


d) Initial velocities plotted against time, Michaelis Menten curve of Tyr^{106}Phe / Ala^{232}Ser /Val^{233}Gly TGT variant for [8-^{3}H]-guanine and tRNA^{Tyr}.

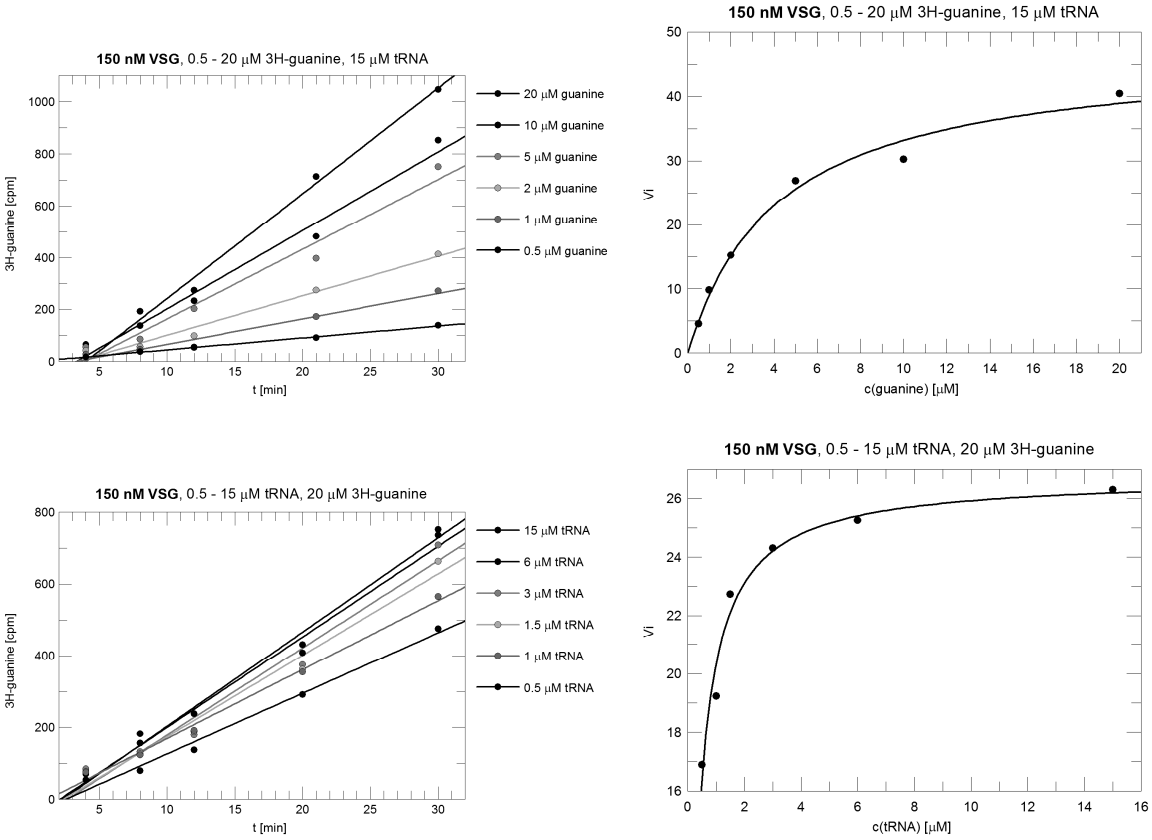




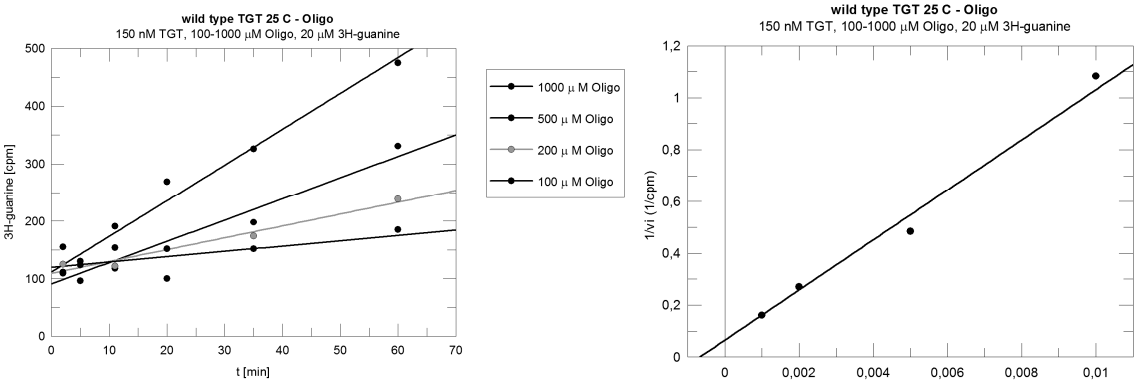




e) Initial velocities plotted against time, Michaelis Menten curve of Tyr[106]Phe / Cys[158]Val /Ala[232]Ser /Val[233]Gly TGT variant for [8-^3H]-guanine and tRNA^{Tyr}.

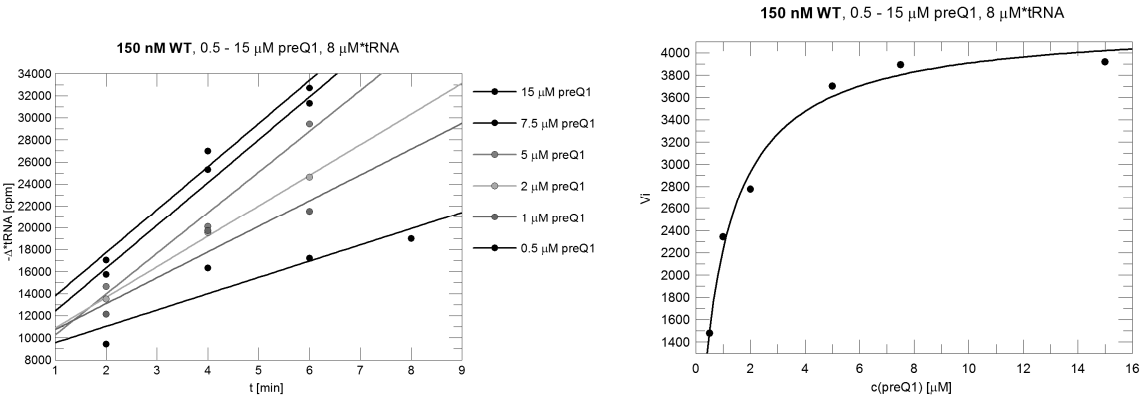


f) Initial velocities plotted against time, Lineweaver/Burk diagramm of wild-type TGT for [8-^3H]-guanine and RNA oligo^{Asp} (tRNA^{Asp} anticodon stemloop).

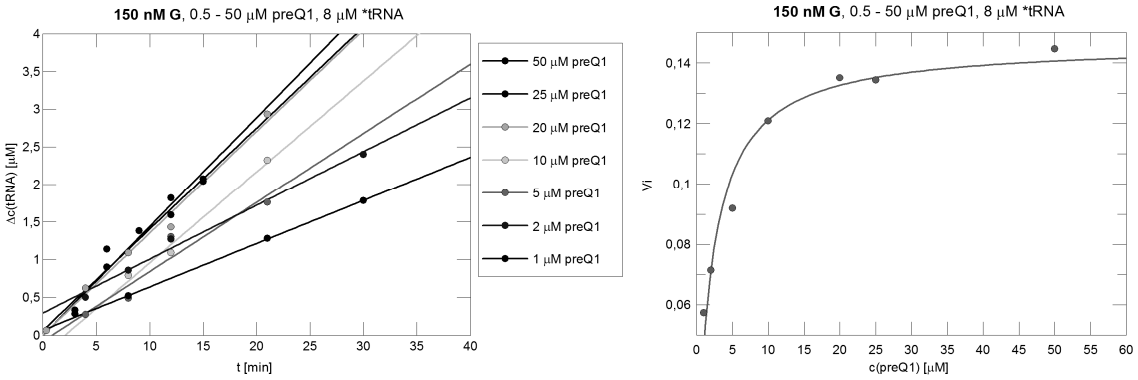


5.3.2 [8-^{3}H]-guanine-tRNA^{Tyr} TGT 'washout assay'

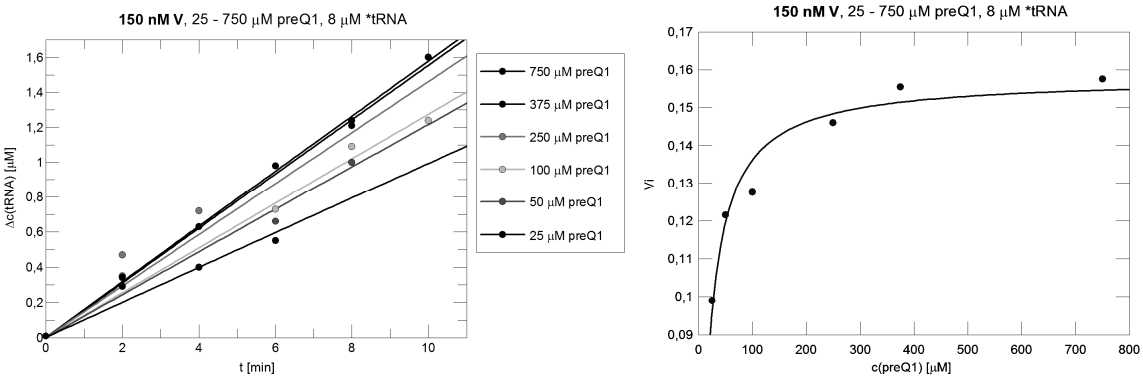
a) Initial velocities plotted against time, Michaelis Menten curve of wild-type TGT and preQ$_1$.



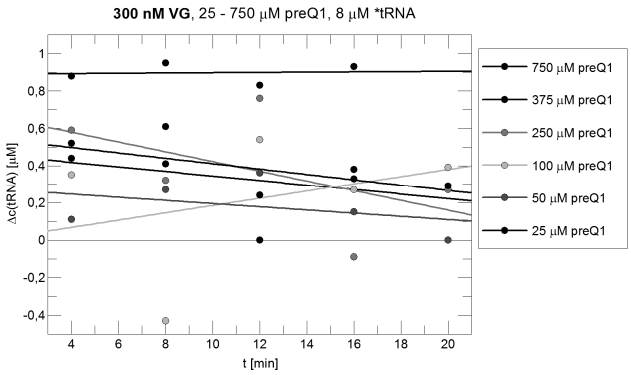
b) Initial velocities plotted against time, Michaelis Menten curve of Tyr^{106}Phe / Val^{233}Gly TGT variant and preQ$_1$.



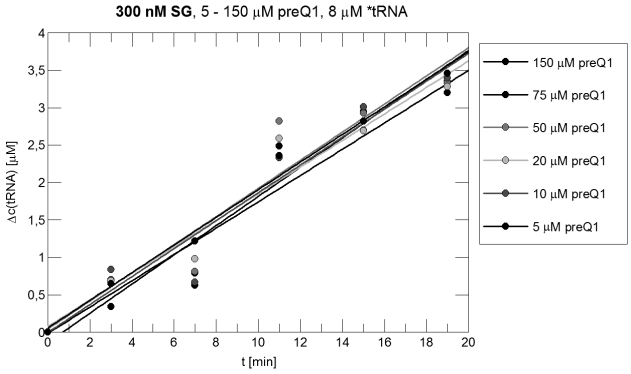
c) Initial velocities plotted against time, Michaelis Menten curve of Tyr^{106}Phe / Cys^{158}Val TGT variant and preQ$_1$.



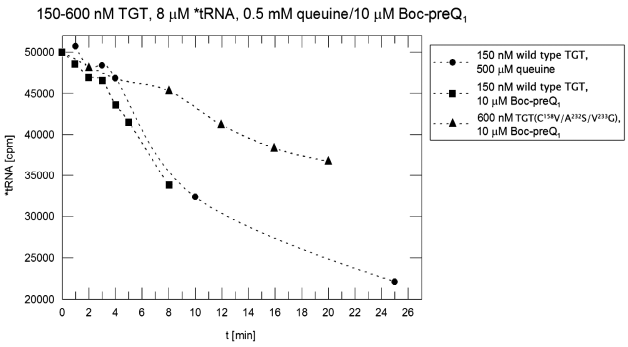
d) Initial velocities of Tyr[106]Phe / Cys[158]Val /Val[233]Gly TGT variant and preQ$_1$ plotted against time.



e) Initial velocities plotted against time, Michaelis Menten curve of Tyr[106]Phe / Ala[232]Ser /Val[233]Gly TGT variant and preQ$_1$.

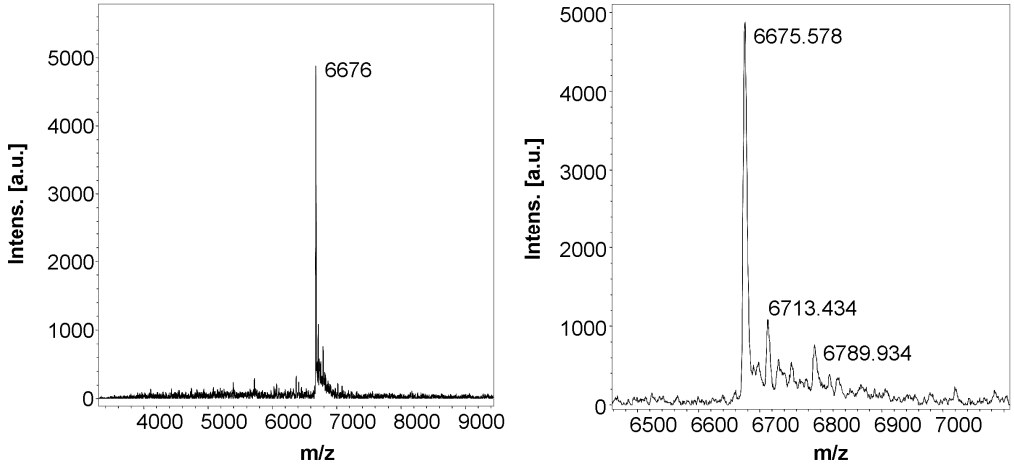


f) Excorporation of [8-^3H]-guanine from labelled tRNA^{Tyr} in the presence of queuine and Boc-preQ$_1$

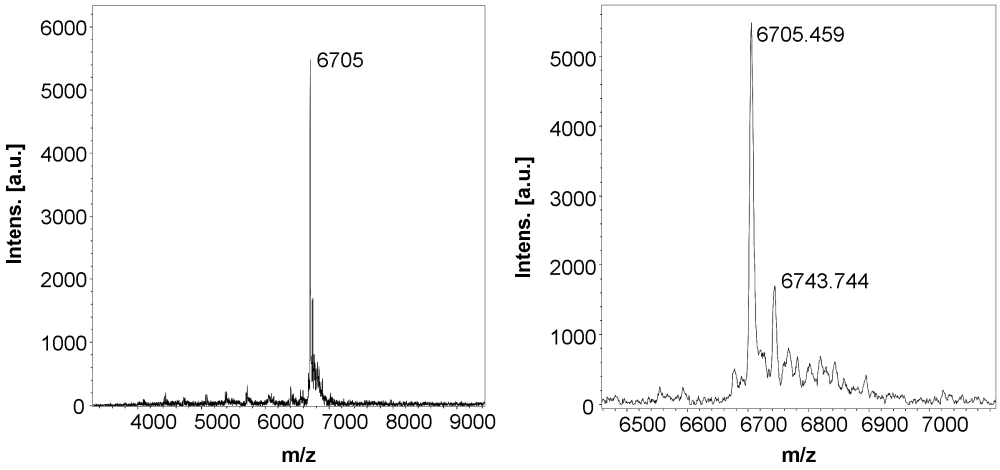


5.4 MALDI-TOF/MS analysis of RNA oligo^{Asp} reaction products

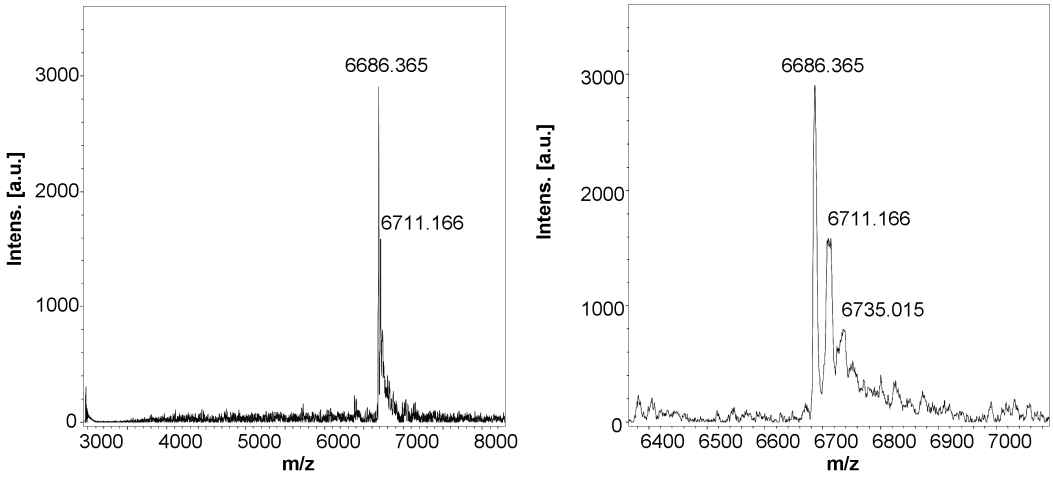
a) Native 21mer RNA oligo^{Asp}.



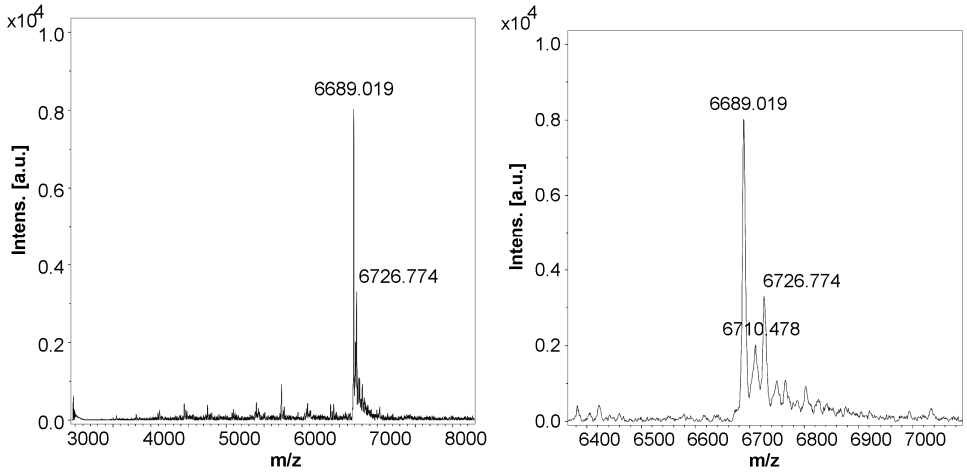
b) preQ$_1$-modified RNA oligo^{Asp} (produced by wild-type TGT, and 500 µM preQ$_1$).



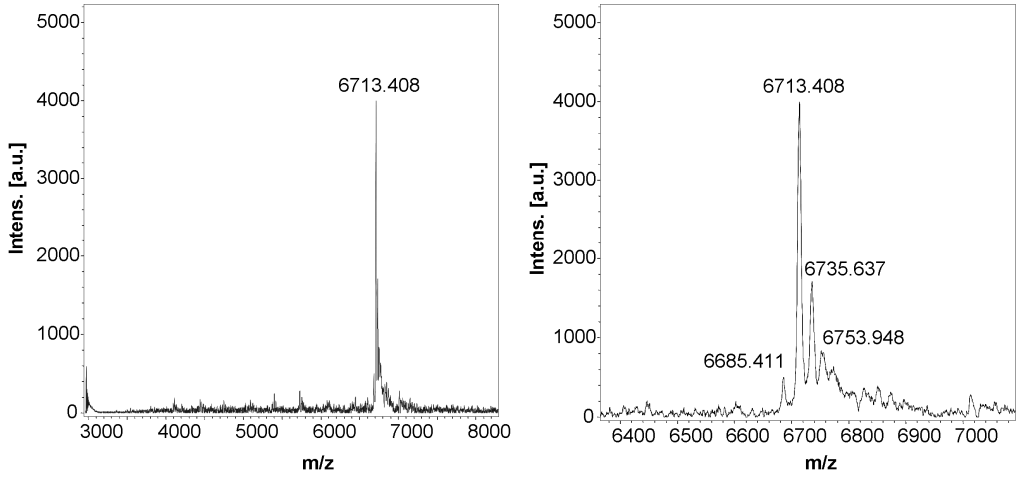
c) preQ$_1$-modified RNA oligo^{Asp} (produced by wild-type TGT, and 5 µM preQ$_1$).



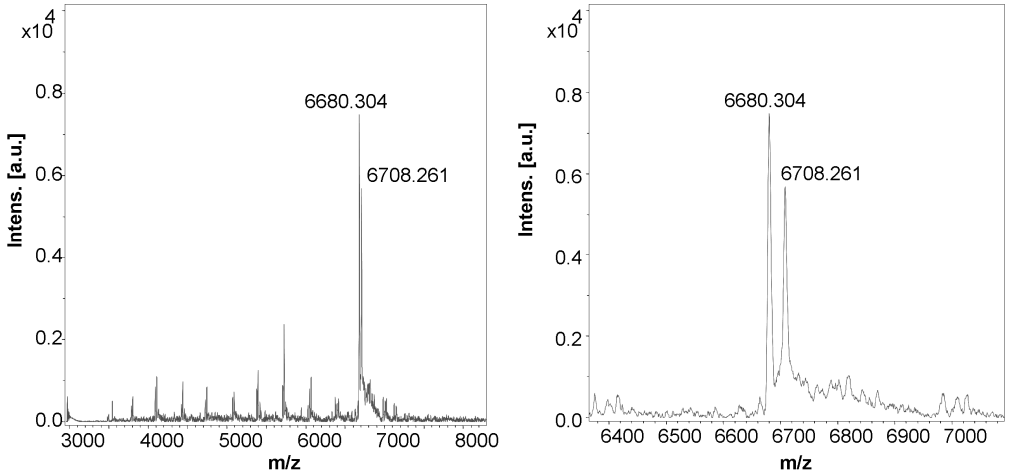
d) preQ$_1$-modified RNA oligo^{Asp} (produced by wild-type TGT, and 1 µM preQ$_1$).



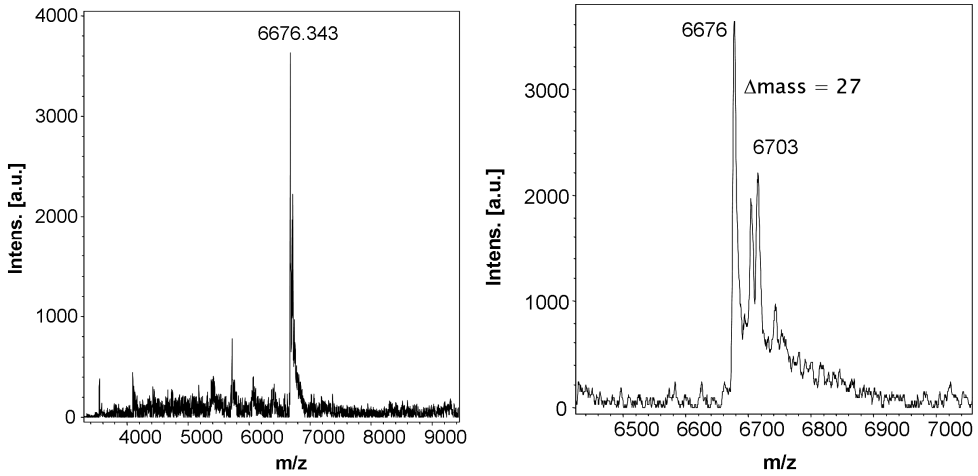
e) preQ$_1$-modified RNA oligo^{Asp} (produced by Tyr^{106}Phe /Val^{233}Gly TGT variant).



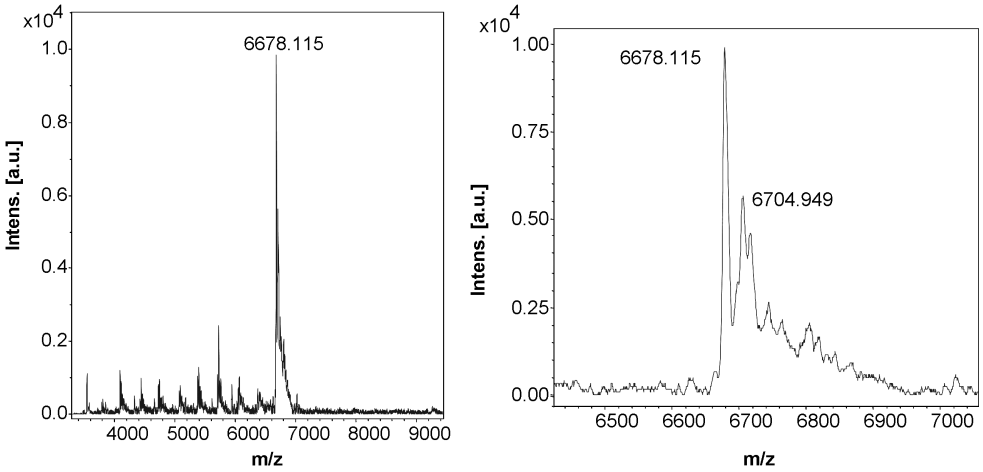
f) preQ$_1$-modified RNA oligo^{Asp} (produced by Tyr^{106}Phe /Cys^{158}Val /Val^{233}Gly TGT variant).



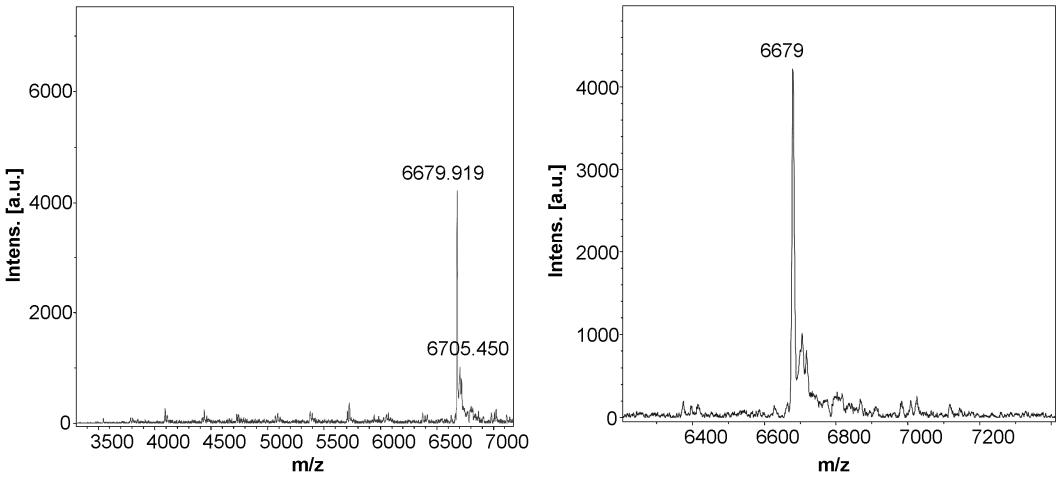
g) Partially preQ$_1$-modified RNA oligo^{Asp} (produced by wild-type TGT and queuine).



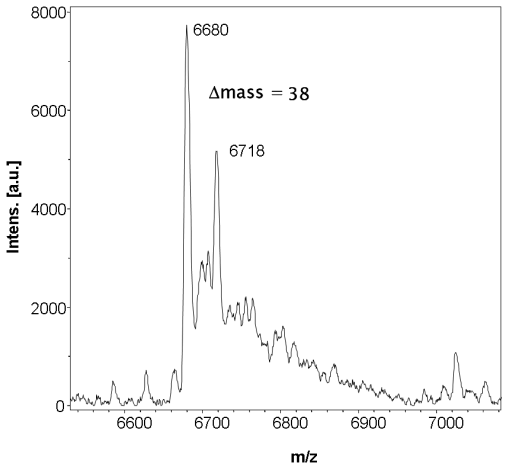
h) Partially preQ$_1$-modified RNA oligo^{Asp} (produced by Tyr^{106}Phe /Val^{233}Gly TGT variant and queuine).



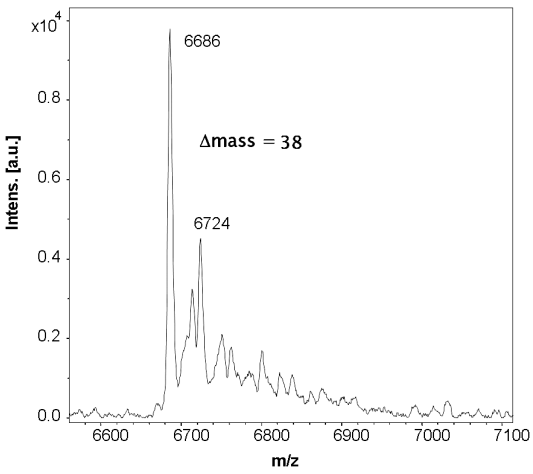
i) Unmodified RNA oligo^{Asp} (produced by Tyr^{106}Phe /Cys^{158}Val TGT variant and queuine).



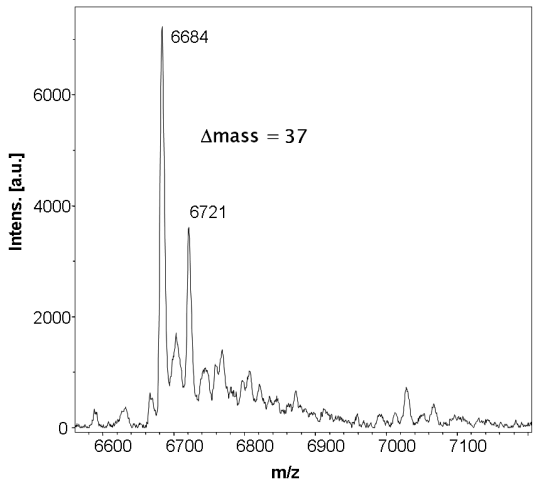
j) RNA oligo^{Asp} (produced by wild-type TGT and Boc-preQ$_1$).



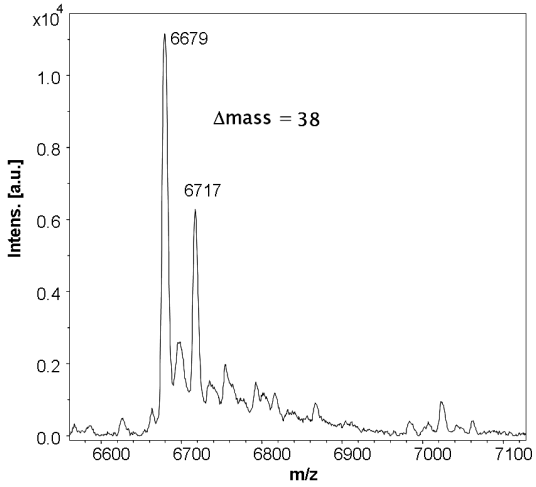
k) RNA oligo^{Asp} (produced by Tyr^{106}Phe / Cys^{158}Val TGT variant and Boc-preQ$_1$).



l) RNA oligo^{Asp} (produced by Tyr^{106}Phe / Cys^{158}Val /Val^{233}Gly TGT and Boc-preQ$_1$).



m) RNA oligo^{Asp} (produced by Tyr^{106}Phe / Cys^{158}Val /Ala^{232}Ser /Val^{233}Gly TGT variant and Boc-preQ$_1$).



5.5 References

[1] www.python.org.

[2] Otwinowski Z, Minor W. Processing of X-ray Diffraction Data Collected in Oscillation Mode. Methods in Enzymology, Macromolecular Crystallography 1997;276 (part A):307-26.

[3] Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL. Crystallography and NMR system: A new software suite for macromolecular structure determination. Acta Crystallogr. sect. D 1998;54:905-21.

[4] Sheldrick GM, Schneider TR. SHELXL: high-resolution refinement. Methods Enzymol. 1997;277b:319-43.

[5] Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallogr. sect. D 2004;60:2126-32.

[6] Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, Thompson JD. Multiple sequence alignment with the Clustal series of programs. Nucleic Acids Res 2003;31 (13):3497-500.

[7] Boeckmann B, Bairoch A, Apweiler R, Blatter MC, Estreicher A, Gasteiger E, Martin MJ, Michoud K, O'Donovan C, Phan I, Pilbout S, Schneider M. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. Nucleic Acids Res 2003;31 (1):365-70.

5.6 Abbreviations

archaeosine	7-formamidino-7-deazaguanosine
BIH	2-butyl-1*H*-imidazole-4,5-dicarboxylic acid hydrazide
Boc	t-butyloxycarbonyl
Da	Dalton
ESI	electrospray ionisation
GSH	glutathione (reduced form)
GSSG	glutathione disulphide (oxidised form)
GST	glutathione S transferase
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
His_6	hexa histidine
HPLC	high pressure liquid chromatography
hTGT	human TGT
ics	intra/inter cellular spread
ipa	invasion plasmid encoded antigen
ipg	invasion plasmid gene
IPTG	Isopropyl-β-D-1-thiogalactopyranoside
M cell	membranous epithelial cell
MALDI	Matrix-assisted laser desorption/ionisation
mRNA	messenger RNA
MS	mass spectrometry
mxi	membrane expression of invasion plasmid antigens
NMR	nuclear magnectic resonance
osp	outer *Shigella* protein
PBS	phosphate buffer saline
PEG	polyethylene glycole
PMN	polymorphonuclear leukocytes
$preQ_0$	7-cyano-7-deaza-guanine
$preQ_1$	7-(aminomethyl)-7-deazaguanine
queuine	7-(((4,5-cis-dihydroxy-2-cyclopenten-1-yl)amino)methyl)-7-deazaguanine
rho	ras homolog gene family
sig	*Salmonella* invasion gene
sip	*Salmonella* invasion protein
spa	surface presentation of invasion plasmid antigens

TGT	tRNA guanine transgylcosylase
TOF	time of flight
Tris	trishydroxymethylaminomethane
tRNA	transfer RNA
TTS	type III secreton
vinc_hd	Vinculin head domain
vir	virulence

Zusammenfassung

Strukturelle Charakteriserung und Mutationsstudie von *Shigella* Pathogenitätsfaktoren

An der Bakterienruhr, verursacht durch *Shigella*-Bakterien, erkranken jährlich 163 Mio. Menschen. Die Erreger dringen in die Colon-Epithelzellen des Wirtes ein, vermehren sich dort und infizieren weitere Epithelzellen. Durch diesen Prozess wird ein Großteil des Epithels zerstört. Die Fähigkeit der *Shigellen* zur Invasion ist auf einem Virulenzplasmid kodiert. Die Expression der sog. Invasin-Gene wird durch den Virulenzfaktor VirF kontrolliert. Die Translation der *virF*-mRNA ist von der Aktivität eines tRNA-modifizierenden Enzyms, der tRNA-Guanin-Transglycosylase (TGT), abhängig. TGT-Defektmutanten von *Shigellen* sind kaum noch in der Lage, Wirtszellen zu penetrieren.

Klonierung, Expression, Aufreinigung und Kristallisation von Invasinen

Im ersten Teil der Arbeit wurden die Invasingene *ipaA*, *spa15*, *ipgB2*, *ospD1*, *ipgE*, die für den *Shigella*-Invasionsmechanismus an unterschiedlicher Stelle von Bedeutung sind, kloniert und durch *E. coli* in großen Mengen produziert, lagen jedoch im Falle von IpaA, OspD1, IpgB2 unlöslich in Form von Einschlusskörpern („Inclusion Bodies“) vor. In anschließenden Arbeiten sollen die Pathogenitätsfaktoren aus den Einschlusskörpern in eine lösliche Form überführt, weiter aufgereinigt und kristallisiert werden. IpgE wurde als Glutathion-S-Transferase-Fusionsprotein erfolgreich in löslicher Form in *E. coli* überproduziert und mittels Affinitätschromatographie, sowie Größenausschlusschromatographie gereinigt. Die Identität des Proteins wurde mittels Massenspektrometrie verifiziert. Durch ein breit angelegtes Kristallisationsscreening wurden Bedingungen identifiziert, die zur IpgE-Kristallisation geeignet sind. Diese sollen in nachfolgenden Arbeiten für eine röntgenkristallographische Strukturaufklärung von IpgE weiter optimiert werden.

Mutagenesestudie zur Substrat-Spezifität und -Selektivität der TGT

Im zweiten Teil der Arbeit wurden Mutagenesestudien an einem Target der Wirkstoffentwicklung gegen *Shigella* - der TGT - durchgeführt, um neue Erkenntnisse über Substrat-Spezifität und -Selektivität gewinnen zu können.

Die TGTs, evolutionär sehr alte Enzyme, sind in allen drei Domänen des Lebens vertreten, unterscheiden sich allerdings in ihrer Substrat-Spezifität und -Selektivität. So katalysieren zwar alle drei Varianten eine Austauschreaktion, bei der eine Guaninbase in tRNAs durch eine modifizierte Base ersetzt wird. Die eingefügte modifizierte Base ist jedoch in den drei Domänen verschieden (Archae: $preQ_0$; Bakterien: $preQ_1$; Eukaryoten: Queuin). Interessanterweise können eukaryotische TGTs *in vitro* auch die beiden Substrate ihrer verwandten TGTs erkennen und in tRNAs inserieren, während ihre archaeaelle Variante außer dem eigenen Substrat nur Guanin in der Bindetasche aufnehmen kann. Die bakterielle TGT erkennt zwar ebenfalls das archaeaelle Substrat, kann aber das eukaryotische Substrat nicht umsetzen.

Die unterschiedlichen Substratspezifitäten der archaeaellen und bakteriellen TGTs beruhen auf einem Peptidbindungsflip im aktiven Zentrum, der in Bakterien durch den Protonierungszustand von Glutamat^{235} kontrolliert wird. Der Peptidflip ändert die Eigenschaften der Bindetasche so, dass entweder Guanin oder $preQ_1$ gebunden werden kann. Das Konformer, das Guanin erkennt, ist ebenfalls in der Lage, $preQ_0$ zu binden. In Archaea kann die entsprechende Peptidbindung ihre Orientierung nicht ändern, da sie durch eine invariante Rückgratwechselwirkung stabilisiert wird. Für eine TGT(Glu^{235} Gln)-Variante, die, ähnlich dem archaeaellen Enzym, eine bevorzugte Stabilisierung des Peptipflips in derjenigen Orientierung gewährleisten sollte, die das Binden von Guanin und $preQ_0$ bevorzugt, konnte gezeigt werden, dass die Selektivität zugunsten von $preQ_0$ umgekehrt wird. Zahlreiche Kristallstrukturen der mutierten TGT in Komplex mit den verschiedenen Substraten konnten belegen, dass die Bindetasche im mutierten Enzym tatsächlich hauptsächlich in der Konformation vorliegt, die das Binden von Guanin und $preQ_0$ ermöglicht.

Die Substrat-Promiskuität des eukaryotischen Enzyms basiert auf der Tatsache, dass die Substratbindetasche sehr groß sein muss, um Queuin aufnehmen zu können. Die entscheidenden Interaktionen werden jedoch zu dem heterozyklischen, guaninähnlichen Teil des Moleküls ausgebildet, der in allen drei Substraten analog ist. So können auch die wesentlich kleineren, aber in diesem Bereich sehr ähnlichen Substrate $preQ_0$ und $preQ_1$ durch die eukaryotische TGT erkannt und umgesetzt werden. Im Vergleich eines Homologiemodells der humanen TGT mit einer Kristallstruktur der *Zymomonas mobilis* TGT wurden Aminosäuren in der Bindetasche identifiziert, die für eine Vergrößerung der Bindetasche in eukaryotischen TGTs verantwortlich sein könnten. Um die Unterschiede der bakteriellen und humanen TGT-Bindetaschen näher zu beleuchten, wurden *Z. mobilis* TGT-Varianten konstruiert, bei denen diese Reste ausgetauscht wurden, um eine solch vergrößerte Bindetasche zu schaffen, und biochemisch sowie strukturell charakterisiert. Die erzeugten TGT-Varianten zeigten eine signifikante Erniedrigung ihrer Wechselzahlen. Es muss daher angenommen werden, dass die ausgetauschten Reste, trotz ihrer relativ großen Entfernung zum katalytischen Rest Asp^{280}, einen entscheidenden Einfluss auf die katalytische Aktivität des Enzyms ausüben. In kinetischen Messungen wurde beobachtet, dass lediglich eine der TGT-Varianten TGT($Val^{233}Gly$), sowie unerwarteter Weise bereits das *Z. mobilis* Wildtyp-Enzym bei sehr hohen Queuin-Konzentrationen in der Lage sind, radioaktiv markiertes Guanin aus tRNA in Gegenwart von Queuin auszubauen. Analyse von Komplex-Kristallstrukturen und Molekular-Dynamik-Simulationen unterstützen diese Ergebnisse.

Danksagungen

Bei Herrn Prof. Dr. *GERHARD KLEBE* möchte ich mich für das interessante Thema, seine stete Diskussionsbereitschaft, seine großzügige Förderung zahlreicher Tagungsreisen, die große Freiheit bei der Arbeit, die zu der sehr produktiven Arbeitsatmosphäre innerhalb seiner Gruppe führt, seine hilfreiche Beratung bei Bewerbungen, die finanzielle Unterstützung und nicht zuletzt für die Zusammenstellung und Führung einer trotz ihrer Größe so harmonischen und menschlich angenehmen Arbeitsgruppe ganz herzlich bedanken.

Herrn PD Dr. *KLAUS REUTER* danke ich herzlich für die Betreuung meiner Arbeit, die vielen Gespräche bei Problemen, seine stete Hilfsbereitschaft, sowie die große Gastfreundschaft, die ich als „Untermieter“ in seinem Büro erfahren habe.

Dr. *ANDREAS HEINE* danke ich für die Einführung in die kristallographischen Methoden, seine immer interessierte Bereitschaft schwierige Kristallstrukturen zu begutachten und die sehr netten Synchrotron-Reisen, bei denen ich viel von ihm gelernt habe.

Prof. Dr. *CHRISTOPH SOTRIFFER* danke ich ganz besonders für die Einführung in das Aufsetzen und Analysieren von MD Simulationen, die Diskussionen und das Korrekturlesen vieler Posterentwürfe. Während der Betreuung des Erstsemesterpraktikums habe ich seine immer sehr geduldige, höfliche, äußerst kompetente und stets interessierte Art mit Menschen umzugehen sehr zu schätzen gelernt. Für die Planung der vielen schönen Wanderungen in Hirschegg und die damit verbundenen Erinnerungen möchte ich mich ebenfalls ganz herzlich bedanken.

Dr. *BERNHARD STENGL* danke ich ganz besonders für die äußerst produktive und nette Kooperation nicht nur bezüglich der E235Q-Mutagenesestudie.

Allen Mitgliedern und Ehemaligen der Arbeitsgruppe danke ich für die nette Arbeitsatmosphäre und den guten Gruppenzusammenhalt. Den Computer-Administratoren Dr. *NILS WESKAMP* und Dr. *MATTHIAS ZENTGRAF* danke ich für ihre ständige Bereitschaft zur schnellen und freundlichen Hilfe. *GERD NEUDERT* und *SVEN SIEBLER* danke ich für die Einführung in die Programmiersprache Python.

Bei *JOHANNES SCHULZE WISCHELER* und *CHRISTIAN HASEWINKEL* möchte ich mich für die nette Zusammenarbeit während ihrer Diplomarbeiten bedanken.

ANDREAS RATJE danke ich für die Etablierung und freundliche Übergabe des MALDI-TOF/MS assays.

Dr. *SILKE SEEBER* und *NICO VOGEL* (Universität Erlangen-Nürnberg) danke ich für die Durchführung zahlreicher MALDI-TOF/MS Analysen. *CLAUDIA SASS* danke ich für die „Magnetic beads“-Aufreinigung der Proben.

HANS-DIETER GERBER danke ich für die Synthesen von Queuin, $preQ_1$ und -Derivaten.

Bei *ANJA VOGT* möchte ich mich für die unkomplizierte und freundliche Zusammenarbeit bzgl. der LC/MS Analyse von Queuin-Proben bedanken.

Ich danke allen meinen Praktikanten, insbesondere *LARA-MARIA DREWES*, *JOHANNES SCHULZE WISCHELER*, *NINA ZIMMER*, *NADJA HERKERT* und *DORIS DOBRIN*, für die Unterstützung bei der Klonierung und Expression von *Shigella* Invasinen.

Bei *STEFFI DÖRR* bedanke ich mich für ihre tatkräftige Hilfe bei der TGT-Aufreinigung.

PD Dr. *THOMAS ADAM* (Charité Berlin) danke ich für Bereitstellung und Präparation des *Shigella* Virulenzplasmids pCP301.

JÖRG KAHNT (MPI Marburg) danke ich für die MALDI-TOF/MS Analyse von IpgE.

Dem DFG Graduiertenkolleg „Proteinfunktion auf atomarer Ebene“ danke ich für die finanzielle Unterstützung.

MEINER FAMILIE, insbesondere *MEINEN ELTERN* möchte ich für ihre immerwährende Unterstützung, ihr stetes Interesse und den seelisch-moralischen Beistand in allen Lebenslagen meinen zutiefst empfundenen Dank aussprechen.

Ganz besonders danken möchte ich *INKEN PAYSEN*, *TINA HUNGER* und *ALWIN SCHÄFER* für die enge Freundschaft, die stets aufmunternde Worte, viel Verständnis und geduldiges Anhören von Problemen.

SIMONE, *GERO* und *ROMAN SCHERMUTZKI* möchte ich für die außerordentliche menschliche Bereicherung neben der Promotion ganz herzlich danken. Meinem Patenkind *ROMAN* dabei zusehen zu dürfen, wie er die Welt um sich herum entdeckt, empfinde ich als ein ganz großes Geschenk.

Dr. *PAUL CZODROWSKI*, Dr. *ANDREAS BLUM*, *JARK BÖTTCHER*, Dr. *VIDA ZOHRABI*, Dr. *MATTHIAS ZENTGRAF*, Dr. *KATRIN SILBER*, *JOHANNES SCHULZE WISCHELER*, *LISA ENGLERT* und *BERNHARD BAUM* möchte ich ganz besonders für die vielen netten Stunden, Gespräche und die schöne Zeit in Marburg während und neben der Promotion danken.

TORSTEN LUKSCH verdanke ich eine ausgelassen schöne und erfüllte Zeit in Marburg, in der er durch sein einfühlsames Verständnis und seine Geduld, seine tiefe Freundschaft und Verbundenheit und nicht zuletzt durch seine Liebe mich und mein Leben unendlich bereichert hat.

Erklärung

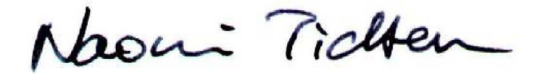
Ich versichere, dass ich meine Dissertation

Structural and mutational characterisation of *Shigella* Pathogenicity Factors

selbständig, ohne unerlaubte Hilfe angefertigt und mich dabei keiner anderen als der von mir ausdrücklich bezeichneten Quellen bedient habe.

Die Dissertation wurde in der jetzigen oder einer ähnlichen Form noch bei keiner anderen Hochschule eingereicht und hat noch keinen sonstigen Prüfungszwecken gedient.

Marburg, den 12.12.2007



(Naomi Tidten)

Curriculum Vitae

Naomi Tidten (Dipl.-Biochem.)

Date of Birth: Dec, 8th 1978 in Bonn, Germany.

School Education

1985 – 1992	Deutsche Schule Tokyo, Japan (Accredited German School Abroad).
06/1998	High School Diploma, Gymnasium Icking, Germany.

University Studies

1998 – 2003	Study of biochemistry (Universität Bayreuth, Germany).
08/2002 – 02/2003	Diploma thesis, Topic: "*Cloning and characterisation of miniGC-C, an extracellular domain fragment of the Guanylyl-cyclase-C"* (Lehrstuhl für Struktur und Chemie der Bio-polymere, research group of Prof. Dr. P. Rösch, Universität Bayreuth, Germany).

Internships

07/1998 – 09/1998	Bayerische Landesanstalt für Bodenkultur und Pflanzenbau – Bavarian Institute of Crop and Plant Cultivation, Freising, Germany.
09/1998 – 10/1998	Prof. Dr. Robert Huber, Max Planck Institut für Biochemie, Martinsried, Germany.
02/2004 – 05/2004	Aventis Pharma, Department of Structural Biology, Frankfurt/ Main, Germany.

Scientific Work

05/2003 – 09/2003 Research stay (Kwansei Gakuin University, Sanda, Prof. Dr. H. Yamaguchi, and Osaka University, Prof. Dr. Y. Hidaka, Japan), supported by a German Academic Exchange Service (DAAD) grant.

10/2003 – 01/2004 Research stay (Lehrstuhl für Struktur und Chemie der Biopolymere, research group of PD Dr. U. C. Marx, Universität Bayreuth, Germany) supported by a grant of the Universität Bayreuth (Gesetz zur Förderung des wissenschaftlichen und künstlerischen Nachwuchses in Bayern – BayNwFG).

since 05/2004 PhD project: "*Structural and mutational characterisation of* Shigella *Pathogenicity Factors"* (Institut für Pharmazeutische Chemie, research group of Prof. Dr. G. Klebe/ PD Dr. K. Reuter, Philipps-Universität Marburg, Germany), supported by a DFG Graduiertenkolleg grant "Protein function at the atomic level".

Publications

Schulz A., Marx U. C., Tidten N., Lauber T., Hidaka Y., Adermann K.:
Side chain contributions to the interconversion of the topological isomers of guanylin-like peptides, J Pept. Sci. **2005** *Jun;* 11*(6):* 319-30.

Tidten N., Stengl B., Heine A., Garcia G. A., Klebe G., Reuter K.:
Glutamate versus Glutamine exchange waps substrate selectivity in tRNA-guanine transglycosylase: Isight into the regulation of substrate selectivity by kinetic and crystallographic studies, J Mol. Biol. **2007** *Nov;* 374*(3):* 764-76.

Oral Presentations

09/2005 Anglo-German-Japanese Biochemistry Meeting, Kyoto, Japan: *Substrate specificities of eukaryotic and eubacterial tRNA-guanine transglycosylases.*

09/2006 Graduiertenkolleg 541 *„Proteinfunktion auf atomarer Ebene"* Symposium, Marburg, Germany: *Mutational study on substrate specificity of tRNA-guanine transglycosylases.*

Poster Presentations

01/2005 First FECS school on protein chemistry, Brixen, Italy:
Unravelling substrate specificity & flexibility in binding pockets of the TGT system
(Tidten N., Stengl B., Meyer E., Heine A., Brenk R., Diederich F., Reuter K.).

03/2005 56. Mosbacher Kolloquium der Gesellschaft für Biochemie und Molekularbiologie, Mosbach (Baden), Germany:
Studies of Peptide Hormone/Receptor Interactions by NMR Spectroscopy
(Vitzithum K., Link N., Schuster D., Tidten N., Lauber T., Rösch P., Marx U.C.).

03/2006 Gesellschaft Deutscher Chemiker (GDCh) Annual Meeting "Frontiers in Medicinal Chemistry", Frankfurt/Main, Germany:
Mutational studies on substrate specificity in tRNA guanine transglycosylase binding pockets
(Tidten N., Sotriffer C., Gerber H.-D., Stengl B., Reuter K., Klebe G.).

05/2006 8th International School on the Crystallography of Biological Macromolecules, Como, Italy:
Flexibility & specificity of TGT elucidated by MD and crystal structure analysis
(Tidten N,. Sotriffer C., Gerber H.-D., Stengl B., Reuter K., Klebe G.).

08/2006 1st European Chemistry Congress, Budapest, Hungary:
Unravelling substrate specificity & flexibility in binding pockets of the TGT system
(Tidten, N., Sotriffer, C., Gerber, H.-D., Stengl, B., Reuter, K., Klebe, G.).

03/2007 Gesellschaft Deutscher Chemiker (GDCh) Annual Meeting "Frontiers in Medicinal Chemistry", Berlin, Germany:
Glu vs. Gln exchange swaps substrate specificity in TGT - Insight into the regulation of substrate selectivity
(Tidten N., Stengl B., Heine A., Klebe G., Reuter K.).