Structural and functional study of the SaPIbov1 Staphylococcus aureus pathogenicity island regulator protein

Ph. D. Thesis

Kinga Nyíri
Chemical Engineering MSc

Supervisor
Prof. Beáta G. Vértesy
Ph. D., D.Sc.

Budapest
2018
My individual role in the preparation of the Ph. D thesis

The data obtained in my Ph.D. thesis is the result of the combined effort of our group and collaborators, therefore ‘first person plural’ is used throughout the text.

The thesis is based on the first-author and co-author articles listed in Chapter 8.1 and 8.2. In case of the first-author papers I planned, performed and evaluated most of the experiments and wrote the majority of the text of the manuscripts. Contributions of major importance of the co-authors to these first-author articles are listed in Table 1.

Table 1. Contributions of major importance of the co-authors to my first-author publications

<table>
<thead>
<tr>
<th>First-author paper</th>
<th>Contributions of major importance of the co-authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nyíri et al 2015 PLoS One</td>
<td>Measurement and evaluation of synchrotron radiation dichroism spectrum of Stl were done by Dr. József Kardos and Dr. András Micsonai.</td>
</tr>
<tr>
<td>Nyíri et al 2015 Structural Chemistry</td>
<td>Mycobacterium tuberculosis dUTPase protein was provided by Dr. Judit Eszter Szabó.</td>
</tr>
<tr>
<td>Nyíri et al 2018 Scientific Reports</td>
<td>Measurement and evaluation of small angle X-ray scattering (SAXS) spectra of Stl, human dUTPase and their complexes were performed by Dr. Haydyn D. T. Mertens and Dr. Dmitri I. Svergun.</td>
</tr>
<tr>
<td></td>
<td>Measurement and evaluation of hydrogen-deuterium exchange mass spectra of Stl, human dUTPase and their complexes were performed by Matthew J. Harris and Dr. Antoni Borysik.</td>
</tr>
<tr>
<td></td>
<td>Measurement and evaluation native exchange mass spectra of Stl, human dUTPase and their complexes were performed by Dr. Olivér Ozohanics.</td>
</tr>
<tr>
<td></td>
<td>Isothermal titration calorimetry measurements and data evaluation were performed together with Dr. Gergely Nagy.</td>
</tr>
<tr>
<td></td>
<td>In vitro inhibition measurement of human dUTPase activity in the presence and absence of Stl and the electrophoretic mobility shift assay experiment were performed by Dr. Borbála Tihanyi.</td>
</tr>
</tbody>
</table>

Results included in detail in the thesis from the articles in which I was a co-author are the results of experiments planned and performed by myself (cf. Table 2). A single exception is the native mass spectrometry of Φ11 phage dUTPase and Stl, in that case, I prepared the material for the measurements, which have been performed and evaluated by Dr. Olivér Ozohanics.
<table>
<thead>
<tr>
<th>Co-author paper</th>
<th>My contribution included in the thesis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sample preparation for native mass spectrometry.</td>
</tr>
<tr>
<td></td>
<td>Preparation of Stl protein for the other <em>in vitro</em> tests.</td>
</tr>
<tr>
<td>Hirmondó et al 2015 DNA Repair</td>
<td>Planning, performing and evaluation of native gel electrophoresis experiments of mycobacterial dUTPase and Stl.</td>
</tr>
</tbody>
</table>
# Table of contents

Acknowledgement.................................................................................................................. 6

Abbreviations .......................................................................................................................... 7

1 Introduction .......................................................................................................................... 9

1.1 *Staphylococcus aureus* pathogenicity island 'life cycle' and the role of Stl protein................................................................................................................................. 9

1.2 Physiological role of dUTPases ......................................................................................... 11

1.3 General structure of dUTPases ...................................................................................... 15

1.3.1 General structure of trimeric dUTPases .................................................................... 15

1.3.2 General structure of dimeric dUTPases ..................................................................... 18

1.4 Species-specific dUTPase segments and their role ......................................................... 21

1.5 dUTPase as a potential drug target .................................................................................. 25

2 Aims ..................................................................................................................................... 31

3 Materials and Methods ....................................................................................................... 33

3.1 Cloning and mutagenesis .............................................................................................. 33

3.1.1 Cloning ....................................................................................................................... 33

3.1.2 Mutagenesis ............................................................................................................. 33

3.2 Protein expression and purification ................................................................................. 34

3.2.1 Protein expression .................................................................................................... 34

3.2.2 Protein purification ................................................................................................. 34

3.3 Protein characterization .................................................................................................. 36

3.3.1 Homology modeling and *in silico* predictions ....................................................... 36

3.3.2 Synchrotron radiation circular dichroism (SRCD) measurements and CD spectrum analysis .......................................................................................................................... 37

3.3.3 Chemical crosslinking ............................................................................................... 37

3.3.4 Electrophoretic mobility shift assay (EMSA) ............................................................ 38

3.3.5 Native polyacrylamide gel electrophoresis (native PAGE) ....................................... 38
3.3.6 Steady-state kinetics experiments ................................................................. 38
3.3.7 Isothermal titration calorimetry (ITC) .......................................................... 39
3.3.8 Electrospray ionization mass spectrometry .................................................... 39
3.3.9 Hydrogen-deuterium exchange mass spectrometry (HDX-MS) ..................... 40
3.3.10 Small angle X-ray scattering ....................................................................... 41

4 Results and Discussion ..................................................................................... 44

4.1 Structural and functional characterization of Stl ................................................ 44
  4.1.1 Secondary structure of Stl ............................................................................. 44
  4.1.2 Domains of Stl and their functions ............................................................. 46
  4.1.3 Interaction surface of Stl with φ11 phage dUTPase .................................... 53
  4.1.4 Molecular shape and structure of Stl in solution ........................................ 54

4.2 Role of the phage specific loop of φ11 phage dUTPase .................................... 56
  4.2.1 Role of the phage specific loop of φ11 phage dUTPase in Stl binding ......... 56
  4.2.2 Role of the phage specific loop of φ11 phage dUTPase in perturbation of
      Stl-DNA interaction ....................................................................................... 61

4.3 Interaction of Stl with the human dUTPase ...................................................... 66

4.4 Interaction of Stl with a dimeric phage dUTPase ............................................. 78

5 Summary ......................................................................................................... 85

6 Összefoglalás ................................................................................................... 88

7 References ....................................................................................................... 91

8 Publication list ................................................................................................ 108
  8.1 List of my first-author publications related to the thesis ................................ 108
  8.2 List of my co-author publications related to the thesis .................................. 108
  8.3 List of other publications .............................................................................. 109
  8.4 List of conference presentations related to the thesis .................................... 110
    8.4.1 Oral presentations at conferences ............................................................. 110
    8.4.2 Poster presentations at conferences ....................................................... 110
Appendix ......................................................................................................................... 112
Appendix Figures ............................................................................................................. 112
Appendix Tables ............................................................................................................... 121
Acknowledgement

First and foremost, I would like to express my heartfelt appreciation to my supervisor, Prof. Beáta G. Vértessey, for her continuous support, motivation, enthusiasm, and immense knowledge. I feel honored that I could work as her colleague in various projects.

I would like to offer my special thanks to all my collaborators: I want to thank Dr. József Kardos and Dr. András Micsonai for their contribution to this work by performing the synchrotron radiation circular dichroism spectroscopy measurements. I am deeply grateful to Matthew J. Harris and Dr. Antoni Borysik for their collaboration in the hydrogen-deuterium exchange mass spectrometry experiments and to Dr. Olivér Ozohanics and Dr. Károly Vékey for the native mass spectrometry experiments and their innovative ideas supporting my work. I owe so much for Dr. Haydyn D. T. Mertens and Dr. Dmitri I. Svergun for their contribution to small angle X-ray scattering analysis. I am deeply grateful to Dr. Gergely Nagy not just for his kind assistance with the isothermal titration calorimetry measurements, but for his constructive comments and warm encouragement. I want to thank Dr. Borbála Tihanyi for her contribution in performing some *in vitro* assays. I am indebted to Dr. Judit Eszter Szabó and to Dr. Rita Hirmondó for their cooperation and insightful suggestions. I owe a lot to all my colleagues for their friendly advice and for creating supportive environment in the laboratory.

I also wish to thank Bianka Köhegyi and Judit Matejka, former undergraduate students, who contributed remarkably to these projects and their constant enthusiasm and diligent work was truly motivating.

I am grateful to Prof. László Nyulászi, the Head of the George Oláh PhD School for supporting my Ph.D. studies.

I would like to thank Prof. László Buday director of the Institute of Enzymology, Prof. György Keserü and Prof. György Pokol former and current directors of the MTA Research Centre for Natural Sciences for allowing me to accomplish my Ph.D. research at the Institute.

I wish to gratefully acknowledge the Ph.D. fellowship grant and funding for short-term fellowships provided by Richter Gedeon Ltd.; travel grants from the Hungarian Chemical Society and ISBC; the EMBO Travel grant; the FEBS Youth Travel Fund, the New National Excellence Program (ÚNKP) Pre-Doctoral fellowship of the Ministry of Human Capacities and the Pungor Ernő Pre-Doctoral fellowship award for supporting my Ph.D. studies and research.

Finally, I would like to express my deepest gratitude to my parents Julianna Losonczi and Sándor Nyíri for their devoted support of my studies and I offer my special thanks to the love of my life, Bence Quittner for his constant encouragement and love.
Abbreviations

APE  apurinic/apyrimidinic endonuclease
BER  base excision repair
CD   circular dichroism spectroscopy
C. jejuni  Campylobacter jejuni
CjDUT  Campylobacter jejuni dUTPase
C-terminus  carboxy-terminal end of a protein
CTD  carboxy-terminal domain/segment
D. mel  Drosophila melanogaster
DCD  dCTP deaminase
DCD-DUT  bifunctional dCTP deaminase-dUTPase
dCMP  deoxycytidine monophosphate
dCTP  deoxycytidine triphosphate
dNTP  deoxyribonucleoside triphosphate
dTMP  thymidine monophosphate
dTTP  thymidine triphosphate
dUDP  deoxyuridine diphosphate
dUMP  deoxyuridine monophosphate
dUPNPP  α,β-imido-dUTP
dsDNA  double-stranded DNA
DUT  dUTPase, deoxyuridine 5’-triphosphate nucleotidohydrolase
      (EC 3.6.1.23)
dUTP  deoxyuridine triphosphate
E. coli  Escherichia coli
EDTA  ethylenediaminetetraacetic acid
EMSA  Electrophoretic Mobility Shift Assay
φNM1DUT  S. aureus φNM1 phage dUTPase
GST  glutathione-S-transferase
hDUT  human dUTPase
HDX-MS  Hydrogen-Deuterium Exchange Mass Spectrometry
HEPES  4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
IPTG  isopropyl-β-D-thiogalactoside
ITC  Isothermal Titration Calorimetry
L. major  Leishmania major
LmDUT  Leishmania major dUTPase
MS  Mass Spectrometry
M. tuberculosis  Mycobacterium tuberculosis
MdDUT  Mycobacterium tuberculosis dUTPase
NCBI CDD  National Center for Biotechnology Information Conserved Domain
      Database
NMR  nuclear magnetic resonance
N-terminus  amino-terminal end of a protein
NTD  amino-terminal domain/segment
P. falciparum  Plasmodium falciparum
Pfam  The Pfam protein families database (EMBL-EBI)  
(https://pfam.xfam.org/)
PfDUT  *Plasmodium falciparum* dUTPase
PCR  polymerase chain reaction
PDB  Protein Data Bank (www.rcsb.org)
P-loop  phosphate-binding loop
PPI  pyrophosphate
*S. aureus*  *Staphylococcus aureus*
SaPI  *Staphylococcus aureus* Pathogenicity Island
SASBDB  Small Angle Scattering Biological Data Bank  
(www.sasbdb.org)
SaUGI  *S. aureus* UDG inhibitor
SAXS  Small angle X-ray scattering
SD  standard deviation
SDS-PAGE  sodium dodecyl sulfate polyacrylamide gel electrophoresis
Sn2  bimolecular nucleophilic substitution
SRCD  synchrotron radiation circular dichroism spectroscopy
Stl  master repressor of SaPIs, the term Stl refers to the repressor of  
*SaPI*~bov1~ throughout the text unless otherwise stated
Stl-CTD  carboxy-terminal domain/segment of Stl (residues 85–267)
Stl-NTD  amino-terminal domain/segment of Stl (residues 1–84)
TBE  TRIS-Borate-EDTA
*T. brucei*  *Trypanosoma brucei*
TbDUT  *Trypanosoma brucei* dUTPase
*T. cruzi*  *Trypanosoma cruzi*
TcDUT  *Trypanosoma cruzi* dUTPase
TCEP  Tris(2-carboxyethyl)phosphine
thymidine  thymine deoxyriboside (1)
TK  thymidine kinase
TRIS  2-amino-2-hydroxymethyl-propane-1,3-diol
TS  thymidylate synthase
U  uracil
UDG  Uracil DNA glycosylase
UDP  uridine diphosphate
UGI  UNG inhibitor
UniProt  Universal Protein Resource (www.uniprot.org)
UNG  Uracil-N-glycosylase
wild type, WT  In here used for a heterologously expressed protein of an organism  
with or without small peptide tags used as a laboratory reference
WSSV  white spot syndrome virus
Φ11DUT  *S. aureus* Φ11 phage dUTPase
1 Introduction

1.1 *Staphylococcus aureus* pathogenicity island ‘life cycle’ and the role of Stl protein

*Staphylococcus aureus (S. aureus)* is an opportunistic pathogen abundant in the human skin microbiome, which is responsible for several nosocomial infections of significant health hazard (2). Virulence in *S. aureus* is often associated with the horizontal transfer of mobile genetic elements, such as bacteriophages, plasmids and pathogenicity islands within the bacterial population. *S. aureus* is infamous for rapidly acquiring resistance against anti-staphylococcal drugs, developing multidrug resistant strains due to spread of SCCmec genomic island (3, 4). Meanwhile, genes responsible for superantigen-related human diseases like staphylococcal toxic shock syndrome are often encoded by 14–27 kb genetic elements so-called *S. aureus* pathogenicity islands (SaPIs) (5). In addition to spreading of SaPI-encoded virulence genes, mobilization of these pathogenicity islands also initiate the transfer of unlinked chromosomal segments containing virulence genes, which largely contribute to the dynamic evolution of virulent and antibiotic-resistant bacteria (6).

The horizontal gene transfer of SaPIs is induced and mediated by so-called SaPI specific helper bacteriophages or rarely by activation of helper prophages by SOS response (7, 8). In the absence of helper phages a SaPI-encoded master repressor, namely Stl, retains the SaPIs reside in the bacterial genome. The unique role of Stl protein in the regulation of SaPI mobilization was verified by deletion of the encoding gene (*stl*), which led to autonomous SaPI replication (9). The expression of a specific helper phage derepressor protein could counter the Stl-mediated repression of SaPI genes responsible for excision of the mobile genetic element. The excised SaPI genome is then circularized, replicated and packed to phage capsids, which facilitate transduction and preservation of the integrity of the mobile genetic element. SaPIs also interfere with phage propagation by hijacking the phage capsids, so that SaPI propagation can be considered as an alternative, Staphylococcus-specific defense against phage invasion replacing CRISPR, which are however abundant in other bacteria rarely found in Staphylococci (10).

The genomic architecture SaPIs somewhat resembles to that of temperate colipages (Figure 1). Similarly to the organization of *cl* and *cro* genes, which encode regulatory proteins of lambda phages, *stl* is located beside an oppositely oriented gene, namely *str*. Hence, the name of *stl* originates from the abbreviation of SaPI transcription leftward, which refers to the
position of that relative to *str* according to the conventional drawing of SaPI genome (7). Concordantly, the function of Stl as a master repressor of SaPI is similar to that of the gene product of *cl* which maintains phage lysogeny (11), since Stl controls the expression of Xis and Int proteins, which are responsible for SaPI excision (9, 12). It has also been shown, that like in the case of CI, production of Stl exerts a negative feedback on its own production (12, 13), rather than acting as an autoinducer as suggested formerly (9).

Figure 1. Similarities between the genomic architectures of SaPIbov1 and *E. coli* lambda phage.

Genes are represented by arrows (length approximately to scale), gene names are indicated above.

The coloring is based on function: recombination genes *int* and *xis* (excisionase) are yellow; transcription regulators are blue; replication genes are green; genes encoding hypothetical proteins are grey. Constructed based on Novick et al (7).

Despite the similarities of prophages and SaPIs, in contrast to the transcription activator function of Cro, the regulatory function of Str is questionable, since the in-frame deletion of *str* (ORF 19) in case of in SaPIbov1 had no apparent effect on the SaPI replication cycle (9). Besides, there is a marked difference between prophage and SaPI activation, since the repression of SaPIs is not directly relieved by SOS response (8).

Derepression of SaPIs is initiated by so-called helper phages. In all investigated cases a single phage encoded protein was responsible for the release of the repression of SaPI genes, so these antirepressor-repressor interactions are SaPI- and phage-specific (8, 14, 15). Deletion of the specific derepressor protein from the helper phage did not perturb phage viability, but resulted in drastically reduced SaPI production and transfer through classical generalized transduction (8, 14, 15). It has also been revealed by *in vitro* experiments that the derepression is due to the direct disruption of Stl-DNA complex by the derepressor protein (14, 16, 17). Since SaPI transduction interferes with phage propagation, thus phages rapidly introduce mutations in the non-essential derepressor proteins to reduce their affinity for the Stl repressor (13, 18).
Interestingly, several of the identified antirepressors also function as dUTPase enzymes (14, 16, 17), which alludes to the existence of a potential link between nucleotide metabolism and horizontal gene transfer in Staphylococcus.

The information available the SaPI proteins is currently very limited, although the function of some of those e.g. Stl, Xis and Int has been assigned, none of those were characterized in detail. In order to gain insight on molecular mechanism governing the SaPI life cycle properties of these proteins ought to be thoroughly studied.

1.2 Physiological role of dUTPases

The dUTPase (deoxyuridine 5'-triphosphate nucleotidohydrolase, EC 3.6.1.23) enzyme is responsible for regulation of the cellular dUTP level by breaking dUTP down into dUMP and inorganic pyrophosphate (19). At first glance, it is not evident, why deoxyuridine species are present in the nucleotide pool at all, since normally uracil is supposed to be incorporated only into the RNA, which calls for the mere existence of uridine derivatives. The presence of dUTP is explained by the peculiar mechanism of dTTP synthesis, which unlike to other dNTPs that are produced from the corresponding NTPs by reduction and phosphorylation catalyzed by ribonucleotide reductase (RNR) and nucleoside diphosphate kinase (NDPK) respectively, proceeds through dUMP intermediate (Figure 2).

dTTP is not synthesized directly from the 5-methyluridine triphosphate, but via methylation of dUMP formed by deamination of cytosine nucleotides (dCTP or dCMP) or in some specific cases by consecutive reduction and dephosphorylation of UDP (20–23) (Figure 2).

It is noteworthy that methylation reaction of dUMP to dTMP catalyzed by thymidylate synthase (TS) is the only de novo source of dTTP (24). Consequently, thymidylate synthase is a well-known target of antiparasitic and cancer therapies as the rapidly dividing cells are highly sensitive for depletion of thymine (24–28). The distinct synthesis path of thymine nucleotides could support the hypothesis that the evolutional transition from RNA to DNA might have proceeded through a uracil-DNA intermediate (29).

---

1 As thymidine denotes a deoxyribonucleoside by convention, the correct term to be used for ribonucleoside triphosphate form of thymine (in some cases also denoted as rTTP) is 5-methyluridine triphosphate (m5UTP).
Figure 2. Pyrimidine biosynthesis pathways (27). The enzymatic reactions that are not universal are represented by dashed lines. Nucleotides and co-factors are shown in white and black boxes, respectively. Abbreviations of enzymes are all in italics. The background is colored blue for traditional antiparasitic drug target enzymes, transparent background and blue lettering are applied for recently verified drug target enzymes, and other enzymes are shown with black lettering on grey background. Abbreviations are as follows: CDA: cytidine deaminase, CH2THF: 5,10-methylene tetrahydrofolate, DCD: dCTP deaminase, DCD-DUT: bifunctional dCTP deaminase – dUTPase, DCDT: dCMP deaminase, dCTPP1: dCTP pyrophosphatase 1, DUT: dUTPase, DHF: dihydrofolate, DHFR: dihydrofolate reductase, NDPK: nucleoside-diphosphate kinase, NK: nucleoside kinase, NMPK: nucleoside monophosphate kinase, RNR: ribonucleotide reductase, SHMT: serine hydroxymethyltransferase, THF: tetrahydrofolate, TK: thymidine kinase, TS, ThyX: classical and flavin dependent thymidylate synthase respectively, TMPK: dTMP kinase.

Although without dUTPase activity the dTTP synthesis is blocked due to lack of dUMP intermediate, based on the lack of cooperativity in dUTPases it has been proposed that the supply of dUMP for dTTP synthesis is likely to be regulated through the dCTP deamination. While dUTPase has an imminent role in the preventive repair due to the sanitization of the dNTP pool (30). Maintenance of ample ratio of dTTP and dUTP is crucial to avoid uracil incorporation into DNA because most of the DNA polymerases are unable to discriminate between the two nucleotides (31–35). At first glance, the presence of uracil in DNA seems harmless, due to its high similarity to thymine. Concordantly, it has been shown that by replacement of thymine to uracil, DNA maintained basically the same global structure (36), albeit the methyl group of thymine may establish some additional van der Waals interaction.
within the double-stranded DNA. Although A:U pairs are not mutagenic, those are still disruptive, since the methyl group of thymine is known to mediate the binding affinity of transcription factors (37, 38). Accumulation of unrepaired uracils in the DNA of nonreplicating cells upon aging have been associated to neurodegeneration caused by the perturbation of gene expression control (38).

However, the major reason why uracil is detected as an error in DNA is that the spontaneous deamination of cytosine bases results in uracil:guanine mismatches (36), which lead to the formation of G→A mutations upon replication, based on U:A base pairing. Cytosine deamination takes place with considerably high frequency, similar to that of the formation of likewise mutagenic 8-oxoguanine (100-500 per/day) and the spontaneous loss of pyrimidine bases (ca. 500/day) (39, 40).

To preserve the genetic information, base excision repair (BER) system removes the potentially mutagenic uracil from DNA. In the first step, uracil-DNA-glycosylase eliminates the uracil base, then the AP endonuclease cleaves phosphodiester backbone of DNA generating a free 3′-OH adjacent to the abasic site. This is followed by the excision of the abasic sugar, the extension of the 3′-OH by a DNA polymerase and finally, the ligation of the nick after gap filling (19, 41).

The most effective and most abundant glycosylase (UNG) acts both on U:G and U:A pairs (42–44). Thus if the number of thymine-replacing uracils is not limited by the effective dUTPase activity, repetitive futile cycles of the BER to remove genomic uracil generates numerous nicks, which could lead to DNA fragmentation (Figure 3) (19, 41). Single strand breaks at replication forks more readily lead to deleterious chromosome fragmentation, which explains the increased affinity of UNG to ssU-DNA (45). The aforementioned thymidylate synthase inhibitors, besides causing thymine depletion also induce dUTP misincorporation into the DNA and indeed exploit the overwhelming of BER mechanism to induce lethal DNA fragmentation in rapidly replicating parasite or cancer cells (28). Interestingly, class switch recombination, which is a normal physiological process in antibody diversification, is also based on the controlled fragmentation caused by uracil repair (46).
Figure 3. Schematic representation of uracil excision repair. Uracil base is excised by a uracil DNA glycosylase enzyme (UDG) and DNA is cleaved at the abasic site by the apurinic/apyrimidinic endonuclease (APE). Then DNA polymerase fills the gap, which is followed by ligation of the nick. Repair is efficient if the dUTP level is low due to normal dUTPase activity. Upon elimination or perturbation of dUTPase activity successive dUTP incorporation due to the high level of cellular dUTP overwhelms the repair system and leads to genome fragmentation as a consequence of excessive nick formation.

All in all, dUTPase activity is a key player in the preventive repair to avoid uracylation and consequent perturbation transcription or fragmentation of DNA. This is also demonstrated by numerous studies on the effect of inactivation, silencing or knock out of dUTPases in various species. It has been shown that lack of dUTPase activity leads to lethality in Escherichia coli (47), Mycobacterium smegmatis (23, 48), Trypanosoma brucei (49), Caenorhabditis elegans (50), Drosophila melanogaster (51) and Arabidopsis thaliana (52, 53). Based on these studies dUTPase was identified as a target for antiparasitic drug design (cf. Chapter 1.5). Although the RNAi silencing was not found to be lethal in human cell lines, most of those showed increased dUTP level and formation DNA strand breaks (54–56). It has also been proven that reduction of dUTPase activity supports thymidylate synthase inhibition and enhances cancer therapy treatment outcome (28, 57–62).

Surprisingly, numerous bacteria lack all of the previously described dUTPase genes (63). The viability of these bacteria can be explained by i) simultaneous low activity, lack or inhibition of UNG, ii) a promiscuous nucleotide pyrophosphatase with dUTPase activity iii) supply of dUTPase from bacteriophages (63). It is especially interesting that S. aureus strains also do not encode an endogenous dUTPase gene and even if integrated prophages contain the dUTPase gene, the expression of the protein is most likely being repressed (17).
In some of the *S. aureus* strains a specific inhibitor protein of the uracil DNA glycosylase, namely SaUGI presumably moderates uracil excision, while the survival of other strains is yet unexplained (63–65). It has also been demonstrated that the prophage-free *S. aureus* RN450 strain, which does not contain the dUTPase gene possesses elevated genomic uracil content compared to *dut+ ung*+ bacteria. It seems likely that mobile genetic elements may encode either dUTPase or SaUGI to escape the damaging uracil-DNA repair, which can impair their horizontal gene transfer (63). In addition to this, uracil content of a mobile genetic element might also prevent its integration into the genome of the new host, as it was recently demonstrated in the case of human immunodeficiency virus (66, 67). It is also tempting to hypothesize that some SaPIs recognize phages through dUTPase-StI interaction, in order to ensure their uracil-free replication (17). This could have a dual advantage: i) replicated SaPI genomes are not fragmented by the host repair mechanism, ii) mutation rate of phages is not elevated by the damaging BER, which hinder their ability to escape the SaPI interference.

1.3 General structure of dUTPases

Most dUTPases belong either to the all-β trimeric or to all-α dimeric structurally distinct enzyme families (27, 68). A few organisms encode monomeric dUTPases, which share a common fold with trimeric dUTPases (69).

1.3.1 General structure of trimeric dUTPases

The tertiary structure of trimeric dUTPases resembles a distorted β-barrel of jelly-roll-like fold (70). In the quaternary structure the subunits are entwined in a homotrimer due to the swapping of the C-terminal β-strands of the neighboring subunits (Figure 4) (71, 72). The intersubunit interactions resemble embracing arms, so that of the overlapping C-terminal segment is often termed as the arm of dUTPases (73, 74). A particular exception to this domain swapping is the dUTPase of *Plasmodium falciparum* in which the C-terminal β-strand due to a sharp turn folds alongside the first β-strand in the core of the protein (75). In the case of the shrimp white spot syndrome virus dUTPase a β-hairpin formed by a unique non-canonical sequence insert perturbs domain swapping (76).

The three active sites of the trimeric dUTPases are constituted from residues of the five conserved motifs from all the three subunits in a specific pattern: one subunit donates conserved Motifs 1, 2 and 4 and another contributes Motif 3 to form the substrate binding cavity between two subunits, which is closed by the flexible Motif 5 of the third protomer upon substrate
binding (Figures 4 and 5). The closure of the active site creates the catalytically competent active site architecture by providing conserved residues of key role in enzyme catalysis (77–79).

**Figure 4. The overall structure of trimeric dUTPases.** A) Cartoon representation of the trimeric human dUTPase (PDB ID: 3EHW). Chains colored by yellow, cyan and salmon, substrate analogue dUPNPP shown as sticks with atomic coloring (carbon: black, oxygen: red, nitrogen: blue, phosphorus: orange). B) Close up representing the unique architecture of the dUTPase active site. The substrate binding pocket is constituted at the interface of two subunits by conserved Motifs 1, 2 and 4 of one subunit and Motif 3 of the other subunit. Upon dUTP hydrolysis the pocket is closed by Motif 5 of the third subunit.

The β-hairpin formed by Motif 3 constitutes a tight substrate binding pocket, excluding the binding of purine deoxynucleotides and posing steric hindrance for thymine binding thus ensuring strict substrate specificity of dUTPases (80) (Figure 4). Specificity for uracil over cytosine is mainly the consequence of the greater number of H-bonds uracil can form with nucleobase binding cleft (80, 81). Tyrosine residue of Motif 3 serves as a steric gate to exclude the binding of ribonucleotides, demonstrated by the drastic increase of UTP hydrolysis upon tyrosine to alanine mutation of this specific residue in *Saccharomyces cerevisiae* dUTPase (71, 82) (Tyr84, Figure 5). Motifs 1, 2 and 4 and the Mg$^{2+}$ cofactor accommodate the phosphate chain of the substrate in the catalytically competent gauche conformation (Figure 5). The P-loop-like C-terminal arm of dUTPases has a crucial role in diphosphate/triphosphate ligand discrimination by promoting only the hydrolysis of dUTP via the coordination of the γ-phosphate (78) (Figure 4). It has also been suggested that an arginine finger-like bidentate interaction of the conserved arginine residue in Motif 5 with the γ-phosphate is essential for full
enzymatic activity (Arg159, Figure 5)(79), however other residues of Motif 5 also play a crucial role in catalysis (78). In addition to the positioning of the phosphate chain the so-called phenylalanine/tyrosine lid of Motif 5 establishes a π-π stacking interaction between the uracil ring of the substrate and the conserved aromatic residue (77) (Phe164, Figure 5). Due to alteration of Motif 5 in case of dUTPases from Bacillus subtilis and its prophage (YncF and YosS), this aromatic stacking interaction is provided by a phenylalanine residue close to Motif 3 (83).

Figure 5. Substrate binding of Φ11 phage dUTPase. A) Structure of substrate binding pocket. Cartoon representation of the trimeric Φ11 phage dUTPase (PDB ID: 4GV8) (84), chains colored by yellow, cyan and salmon, substrate analogue dUPNPP shown as sticks with atomic coloring (carbon: black, oxygen: red, nitrogen: blue, phosphorus: orange), residues important for substrate binding and catalysis are shown as sticks with atomic coloring (carbon: colored by chain, oxygen: red, nitrogen: blue), water and magnesium ion are represented by red and green spheres respectively. Black dashed lines represent hydrogen bonds, the surface of residues in van der Waals interaction with the substrate is shown in grey. A solid magenta arrow represents the attack of water on the α-phosphate B) Schematic representation of protein substrate interactions. The coloring is according to the legend shown on the bottom of the panel. C) Annotated sequence of Φ11 phage dUTPase. Conserved motifs boxed, residues shown on Panel A are denoted with cyan, ocher and salmon respective to their coloring on Panel A. Phage specific insert is underlined with dashed line. Secondary structure elements are annotated by ESPript based on structure (α – alpha helix, β – beta sheet, TT – strict β-turn, elements are numbered gradually from the amino- to the carboxy-terminal) (85).
Structural, solution dynamics and molecular modeling studies revealed that the C-terminal segment undergoes relatively large conformational changes during the enzymatic cycle compared to protein core in which only subtle conformation changes were detected (86–91). However, the degree of flexibility of the C-terminal segment and thus the amplitude of its movement varies between different dUTPases perhaps due to the environment provided by non-conserved residues. The flexibility of the C-terminal arm frequently prohibits its localization in the electron density map determined by X-ray diffraction method. For example, the arm of *E. coli* and Φ11 phage dUTPases is disordered even in the presence of substrate analogue dUPNPP (84, 92).

It has been shown that the kinetic mechanism of dUTPase enzymatic action comprises four major steps: i) a rapid substrate binding, ii) a relatively slow substrate-induced structural change of the enzyme-substrate complex into the catalytically competent conformation, iii) a rate-limiting hydrolysis step, and iv) a fast release of the products in a random fashion (89).

The hydrolysis is initiated by a nucleophilic attack of a water molecule on the α-phosphate along the line of the scissile bond (in-line attack), the reaction proceeds through S_n2 mechanism involving pentavalent phosphorane-like intermediate (Figure 6) (93). The catalytic water is positioned by a strictly conserved aspartate in Motif 3 (Figure 5), perturbation of the coordination of the catalytic water by mutation of this residue to the highly similar asparagine abolishes the enzymatic activity while retains its substrate binding affinity (92, 94).

![Figure 6. Mechanism of phosphate hydrolysis of trimeric dUTPases.](image)

Catalytic water attacks on the α-phosphate, dUTP is hydrolyzed to dUMP and pyrophosphate. (Figure is from Ref. 68.)

1.3.2 General structure of dimeric dUTPases

It has been shown that the sequence of the dUTPase gene of the protozoan *Leishmania major* differs significantly from the trimeric dUTPases, as it encodes a longer protein, in which the characteristic motifs of the trimeric dUTPases did not present (95). Later similar dUTPases have been identified in Trypanosoma species (49, 96), *Campylobacter jejuni* (97), *Deinococcus radiodurans* (98) and Staphylococcal bacteriophages (14, 15).
Structural studies revealed that this new family of dUTPases form dimers in solution (99), and monomers have a predominantly helical structure (100) (Figure 7). Rigid domains of the monomers form the dimer interface, where the two active sites of the dimers are accommodated. Upon substrate binding, similarly to that of trimeric dUTPases, a flexible loop of the dimeric dUTPases closes the active site (100).

Figure 7. Representation of dimeric dUTPase structure. A) Cartoon representation of the dimeric *Leishmania major* dUTPase (PDB ID:2YAY) (101), chains colored by yellow and cyan, substrate analogue dUPNPP shown as sticks with atomic coloring (carbon: black, oxygen: red, nitrogen: blue, phosphorus: orange) (27). B) Active center of LmDUT (PDB ID:2YAY) (101) colored as in Panel B, residues important for substrate binding and catalysis are shown as sticks with atomic coloring (carbon: colored by chain, oxygen: red, nitrogen: blue), water and catalytic calcium ions are represented by red and green spheres, respectively. Substrate analogue dUPNPP shown as sticks with atomic coloring according to Panel A. Black dashed lines represent hydrogen bonds (27). C) Annotated sequence of LmDUT. Conserved motifs boxed, residues shown on Panel B are denoted with cyan and ocher respective to their coloring on Panel B. Secondary structure elements are annotated by ESPript based on structure (α – alpha helix, β – beta sheet, η – 3_10-helix, TT – strict β-turn, elements are numbered gradually from the amino- to the carboxy-terminal) (85).

Dimeric dUTPases share six conserved motifs, although the presence of the C-terminal Motif 6 is not ubiquitous. Both active sites are built up by Motifs 1, 2, 4, 5 of one of the monomers and Motif 3 of the second monomer. Motifs 1-4 of the rigid domain form the ligand-binding pocket and coordinate the catalytic water and divalent metal ions. Motif 5 is located in the mobile domain and mostly responsible for phosphate binding. Glutamate and aspartate residues of Motif 2 and 4 have a crucial role in binding of the divalent metal cofactors,
mutation of these residues to alanine abolish the enzymatic activity (Glu48, Glu51, Glu76 and Asp79, Figure 7) (102, 103). This is a marked difference compared to trimer dUTPases, which can function without cofactor binding although less effectively (104). It has also been revealed by mutational analysis that lysine residues in Motif 3 are responsible for substrate recognition and/or catalysis (Lys59 and Lys62, Figure 7)(102, 103).

Despite the fact that the dimeric and trimeric dUTPases both hydrolyze dUTP the mechanism of the enzymatic catalysis completely differs from each other. In case of dimer dUTPases the reaction proceeds via an in-line nucleophilic attack of the catalytic water molecule on the β-phosphate (105) (Figure 8). The enzyme positions α- and β-phosphates in the proper orientation for catalysis by direct contacts, while the γ-phosphate has a minor or no role in substrate binding and catalysis (105). This mechanism explains why dimeric dUTPases are able to hydrolyze dUDP, while that functions an inhibitor of trimeric dUTPases.

![Figure 8. Mechanism of phosphate hydrolysis of dimeric dUTPases.](image)

The difference in the catalytic mechanism is also reflected by the different orientation of the phosphate tail of dUTP in the two types of dUTPases. In case of trimeric dUTPases the gauche (eclipsed) conformation of the phosphate chain is the catalytically competent arrangement, while the phosphate tail should adopt trans (staggered) conformation to permit catalysis in case of dimeric dUTPases (Figure 9).
Figure 9. **The conformation of substrate analogue in dimeric and trimeric dUTPases.** Overlay of dUPNPP substrate analogue molecules from the dimeric *Leishmania major* dUTPase (sticks with atomic coloring: carbon white, oxygen red, nitrogen blue, phosphorus orange, PDB ID:2YAY) (101) and that of the trimeric Φ11 phage dUTPase (black sticks, PDB ID: 4GV8) (84). The substrate analogue adopts a gauche conformation in case of the trimeric dUTPase while it is in trans conformation in case of the dimeric dUTPase. The position of the catalytic waters represented by spheres differs significantly in the two cases. The black and blue arrows symbolize the in-line attack of the catalytic water molecules.

The role of metal ions in the catalytic mechanism also differs between the two dUTPase families. In case of dimeric dUTPases, two divalent metal ions are involved in the substrate binding and the activation of the catalytic water by those is essential for the hydrolysis. On the contrary, trimeric dUTPases can hydrolyze dUTP in the absence of divalent metals although at a lower rate, due to the lack of positioning of the phosphate tail by the cofactor.

1.4 Species-specific dUTPase segments and their role

Several members of the trimeric dUTPase family possess species-specific additional segments besides the five conserved sequence motifs (cf. Chapter 1.3.1) (Figure 10). The role of these segments has been revealed in some cases, while their function is yet unidentified for some other dUTPases (Figure 10).

The amino-terminal extension of eukaryotic dUTPases serves either as a nuclear or as a mitochondrial localization signal (106–111). In case of rat dUTPase it was also proposed that it plays an important role in the interaction of the protein with the transcription factor PPARα (112).

In addition to the amino-terminal extension, a unique 28-residue-long segment has been identified at the C-terminus of the *Drosophila melanogaster* dUTPase. It has been shown by truncated mutants that this segment has no regulatory role *in vitro*, although it could modulate enzyme function *in vivo* (107, 113, 114). A retroviral nucleocapsid protein of the size of ca. 10 kDa is fused to the amino-terminal of the β-retroviral dUTPases. It has been hypothesized...
that this segment aids the localization of dUTPase to the DNA/RNA in order to enhance the fidelity of replication by the local lowering of the dUTP level at the site of reverse transcription (72, 115, 116).

**Figure 10. The species specific dUTPase segments** (117). Schematic representation of the species-specific segments of dUTPases from different sources and their function. Putative functions are designated by question marks.

Besides the extra segments attached to the amino- and carboxy-terminal of the proteins, various insertions could also be found between the conserved motifs. In mycobacterial dUTPases a short insert of five residues is situated between the conserved Motifs 4 and 5 (Figure 11). Deletion of this small segment did not affect the in vitro enzymatic activity although that was essential for the survival of bacteria due to yet undiscovered cellular activity (48). A longer, so-called pre-V segment was recently discovered in white spot syndrome virus dUTPase, which alters significantly the fold of Motif 5 (76). This ca. 20-residue-long insert hampers domain swapping by turning back the C-terminal arm protruding towards the neighboring subunit to the core of the protein via a β-hairpin (76). Interestingly this difference in the fold was not reflected in the enzymatic activity of the protein, which was found to be similar to that of other dUTPases (76).
In case of the *Plasmodium falciparum* dUTPase (PfDUT) a 25-residue-long insert consisting mostly of hydrophilic amino acid residues can be found between conserved Motifs 2 and 3 (75) (Figure 11). This can be classified as a so-called low complexity region typical of plasmodial proteins. The putative function of these disordered segments is to serve as a variable epitope, which may circumvent immune response of the host (118). The PfDUT protein did neither exhibit domain swapping, although this is not caused directly by the steric hindrance of the insert, the intersubunit interactions provided by this segment might facilitate the formation of the altered fold by compensating for the lack cohesive interactions of the C-terminal arm.

![Figure 11](image_url)

**Figure 11. Sequence alignment of trimeric dUTPases with non-canonical segments.**

Sequence of the human, *Drosophila melanogaster* (D. mel), white spot syndrome virus (WSSV), *Mycobacterium tuberculosis* (M. tub), *Plasmodium falciparum* (P. fal), *S. aureus* (P. 11 phage (P11) dUTPases (with the respective UniProt IDs: P33316-2, Q9V311, Q77378, P9WNS5, Q8II92, Q8SDV3) were aligned by ClustalW algorithm built in the NPS@ server. A active site motifs of dUTPases are bold and framed, Non-canonical segments are bold and red. Identical residues denoted with stars (*) strongly similar residues denoted with colons (:), weakly similar residues denoted with dots (.).

An interesting moonlighting function of a recently discovered phage dUTPase has been attributed to the phage-specific insert of the protein, which is situated between the third and fourth conserved motifs (16) (Figure 11). It has been shown that the dUTPase enzyme from the *Staphylococcus aureus* phage P11 (P11DUT) is involved in the regulation of pathogenicity island mobilization (16). P11DUT acts as a derepressor since it perturbs the binding of the master repressor Stl protein of *Staphylococcus aureus* pathogenicity island (SaPIbov1) to its
cognate DNA site(s), which leads to replication and transfer of the genetic element. We have demonstrated that Stl-dUTPase interaction is possible only if the dUTP is eliminated from the cellular nucleotide pool (17). The dUTP level in *S. aureus* is elevated, since the bacterium does not encode a dUTPase in its genome and staphylococcal prophage dUTPases are under repression in the lysogenic phase (17, 63). The dUTPase-induced activation of SaPIblov1 might link the genomic stability and the mobilization of this genetic element.

The main differences between SaPI inducing and noninducing *Staphylococcus* phage dUTPases have been observed within this ca. 30-40-residue-long phage-specific insert (16). This led to the hypothesis that this segment is essential for SaPI derepression. A role in dUTPase activity has also been assigned to the insert of Φ11DUT based on the inactivity of a truncated mutant enzyme (Φ11DUTΔ96A–134L), from which a 39-residue-long, so-called variable region containing the insert has been removed, based on sequence alignment of staphylococcal phage dUTPases without consideration of the position of catalytic motifs (16). Nevertheless, since the integrity of the fourth active site building motif (residues 133–136 in Φ11DUT, Figure 11) was also perturbed by this truncation the role of the insert in enzymatic activity was questionable. The importance of these residues in active site architecture has been proven based on the 3D crystal structure of Φ11DUT obtained by our group (84). It has also been revealed by the structure that the insert forms a β-pleated mini-domain, which did not perturb the fold of the enzyme core and has no contact with the active center. To investigate the potential role of the insert in dUTPase enzymatic activity we replaced a phenylalanine residue within the phage specific insert by a tryptophan fluorophore (Φ11DUTF108W) (84). The binding of either nucleotide did not induce any observable fluorescence spectral changes in case of this protein, which was confirmed to function as the wild type Φ11DUT (84).

Based on the Φ11DUT structure we have also created a mutant (Φ11DUTA101G–122Q) form, in which 20 residues out of the 26-residue-long phage specific insert were removed (84). This mutant verified that conclusions being drawn from the studies of the Φ11DUTΔ96A–134L mutant about the role of the insert in enzymatic activity were mistaken, as Φ11DUTA101G–122Q exhibited the wild type properties (84). The same observations have later been reported for the insert-less 80α phage dUTPase (119). Although it has been well established that the phage-specific insert has no effect on dUTPase activity, the role of that in Stl-binding, and derepression activity has not yet been characterized in detail.
1.5 dUTPase as a potential drug target

The paramount importance of dUTPase in preventive DNA repair (cf. Chapter 1.2) designate the enzyme as a potential antiparasitic drug target. As the human dUTPase belongs to the trimeric family, inhibitors against the dimeric enzyme family would presumably exhibit fewer side effects due to the fundamental differences between the host and pathogen enzymes. Following this strategy, drug design projects have been initiated targeting dUTPases from *Campylobacter, Leishmania* and *Trypanosoma* species (Table 3). These studies could exploit the high-resolution 3D crystal structures during the development of selective inhibitors, although to find effective molecules, flexibility of the protein should also be taken into account (97, 100, 101, 120).

In spite of the similarities between the members of the trimeric dUTPase family, this enzyme is still in the focus of relevant research efforts in case of *Mycobacterium tuberculosis* and *Plasmodium falciparum* (Table 3). These are two human pathogens of significant importance, the causative agents of tuberculosis and malaria, respectively. The lack of salvage pathways of thymidylate biosynthesis in these organisms greatly enhances the effectivity of those for dUTPase inhibitors (19, 121). Structure based drug design against these enzymes was made possible by the 3D structures of the mycobacterial and the plasmodial dUTPases reported in 2004 and 2005, respectively (75, 122).

An experimental assay applicable for high-throughput screening of candidates with effects on *M. tuberculosis* dUTPase activity has been developed (123). Virtual screening probing more than 2 million drug-like molecules targeting an essential species-specific segment in around the active site of the mycobacterial dUTPase have been carried out (48, 124, 125). This study provided a concise list of promising candidates, which would potentially exert lower off-target effects. The effectivity of these compounds has been demonstrated by in vitro assays and on infected guinea pig model (124).

The first high throughput biochemical assay identified no effective hits against the *Plasmodium falciparum* dUTPase within 5655 compounds that were proven to perturb parasite growth (126). While a small-scale substrate analogue inhibitor design study revealed that modifications on the uracil ring abolish ligand binding to the active site, though substituents at the 3’ and 5’ positions can be varied without loss of effectivity (127). Within this project a drug candidate showing 200-fold selectivity for the pathogen enzyme over the human dUTPase has been developed (128), although the lipophilicity and cell penetration properties of these
compounds are still to be improved. Recently QSAR models were also introduced to facilitate the design of novel compounds higher antimalarial bioactivities (129).

In spite of the excessive efforts still much work is ahead to develop dUTPase inhibitors applicable in antimicrobial therapy.
Table 3. Recent results on the inhibitor design against parasitic dUTPases (27)

<table>
<thead>
<tr>
<th>Organism</th>
<th>Year</th>
<th>Summary</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leishmania major</td>
<td>1997</td>
<td>Identifying LmDUT. Lysates of E. coli showed significant dUTPase activity increase after orthologous expression LmDUT gene.</td>
<td>(95)</td>
</tr>
<tr>
<td></td>
<td>2000</td>
<td>Verification of dimeric form of LmDUT. dUDP and dUTP are both substrates of LmDUT (dUDP is an inhibitor of trimeric dUTPases).</td>
<td>(99)</td>
</tr>
<tr>
<td></td>
<td>2001</td>
<td>Kinetic parameters of LmDUT for dUTP hydrolysis are comparable to that of hDUT. Dissimilarities in the binding of dUDP and dUMP as compared to the human enzyme suggest differences in the structure of the active sites.</td>
<td>(130)</td>
</tr>
<tr>
<td></td>
<td>2011</td>
<td>Crystal structure of LmDUT in complex with substrate analogues, product dUMP, and a substrate fragment (dU). Very tight ligand binding pocket, modifications at the uracil or ribose rings might perturb binding of an inhibitor. The presence of a single phosphate or similarly charged group is important to induce ligand binding conformation of LmDUT.</td>
<td>(101)</td>
</tr>
<tr>
<td>Trypanosoma cruzi</td>
<td>2004</td>
<td>Crystal structure of TcDUT. Major differences between the substrate-binding pocket of dimeric and trimeric dUTPases provides potential for selective inhibitor design. It was observed for the first time that ligand binding induces a large conformational change in case of dimeric dUTPases.</td>
<td>(100)</td>
</tr>
<tr>
<td></td>
<td>2006</td>
<td>Inhibitor design based on in silico docking. No in vivo and in vitro effects of the compounds. Protein flexibility has to be taken into account.</td>
<td>(120)</td>
</tr>
<tr>
<td>Trypanosoma brucei</td>
<td>2008</td>
<td>The dimeric TbDUT is a nuclear enzyme and down-regulation of its activity by RNAi proved that TbDUT is indispensable for efficient cell cycle progression and DNA replication in T. brucei.</td>
<td>(49)</td>
</tr>
<tr>
<td></td>
<td>2013</td>
<td>Conditional dUTPase knockout without adding thymidine caused impaired proliferation and lethality in T. brucei. Adding uracil, uridine or deoxyuridine could not rescue this phenotype. dUTPase has a major role in the provision of pyrimidine nucleotides in kinetoplastids.</td>
<td>(131)</td>
</tr>
<tr>
<td>Campylobacter jejuni</td>
<td>2004</td>
<td>The crystal structure of CjDUT. Mg²⁺ is important for enzymatic action. It was shown for the first time, that nucleophilic attack occurs on the β-phosphate in contrast to the trimeric enzymes, where it happens at the α-phosphate. Ligand binding causes large conformational change so as in the case of TcDUT (100).</td>
<td>(97)</td>
</tr>
<tr>
<td></td>
<td>2009</td>
<td>The difference in inhibition constants as compared to that of the trimeric dUTPase enzymes permits the design of specific inhibitors of CjDUT.</td>
<td>(132)</td>
</tr>
<tr>
<td></td>
<td>2011</td>
<td>Crystal structure of CjDUT with dUPNPP substrate analogue contains only two metal ions at the active site, while in the dUPNP complex three of those were identified. Glu49 is flipped away from the active site in the presence of the triphosphate, and no longer coordinates any of the metal ions.</td>
<td>(101)</td>
</tr>
</tbody>
</table>
## Table 3. Recent results on the inhibitor design against parasitic dUTPases (27)

<table>
<thead>
<tr>
<th>Organism</th>
<th>Year</th>
<th>Summary</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Campylobacter jejuni</em></td>
<td>2013</td>
<td>Crystallographic and NMR studies revealed that similarly to CjDUT in case of TbDUT nucleophilic attack also occurs on the $\beta$-phosphate of the substrate. Unlike in the trimer enzymes in case of TbDUT one of the divalent metal ions plays a direct role in catalysis.</td>
<td>(105)</td>
</tr>
<tr>
<td><em>Mycobacterium tuberculosis</em></td>
<td>2004</td>
<td>Crystal structure of MtDUT reveals that its binding pocket is similar to that of hDUT, hampering the design of MtDUT specific inhibitors.</td>
<td>(122)</td>
</tr>
</tbody>
</table>
|                               | 2008 | Identification of a bifunctional dCTP deaminase dUTPase in *M. tuberculosis*  
The lower similarity to hDUT and broader substrate specificity than MtDUT marks DCD-DUT enzyme a possible target for chemotherapy.  | (133)  |
|                               | 2008 | Introducing a highly sensitive fluorescent label to follow the MtDUT enzymatic reaction.  
Structure and activity of H145W MtDUT is not altered.                                                                                         | (123)  |
|                               | 2011 | Molecular modeling of MtDUT nucleotide binding, based on activation energies from QM-MM modeling hydrolysis is slower than the product release.                                                        | (134)  |
|                               | 2012 | The dUTPase enzyme is essential in *Mycobacterium smegmatis*.  
Mycobacteria-specific loop has no major effect on MtDUT activity *in vitro*, but a loop-specific function seems to be essential within the *in vivo* model *M. smegmatis*. | (48)   |
|                               | 2015 | Virtual screening of several million small molecules against the species-specific surface loop of MtDUT was performed.  
An optimized hit was conjugated to a phagocytosis stimulating tuftsin peptide derivative and encapsulated into PLGA nanoparticles.  
*In vivo* efficacy of this formulation was verified in guinea pig model. | (124)  |
|                               | 2016 | The prevention of DNA uracylation and the regulation of dNTP balance are decoupled in Mycobacteria and separately achieved by DUT and DCD-DUT enzyme functions, respectively.  | (23)   |
|                               | 2016 | RAMD simulations suggest that substrate binding in MtDUT happens through small amplitude conformational changes of the C-terminal arm, in which His21 plays an important role. | (91)   |
| *Plasmodium falciparum*       | 2005 | Development of selective inhibitor candidates against PfDUT with antiparasitic activity.  
Crystal structure of inhibitor-bound PfDUT.                                                                                                     | (75)   |
|                               | 2005 | Selective, nontoxic, drug-like inhibitor lead design against PfDUT.  
Analogues of dUMP with the variety of substituents at the 5' and 3' positions.  
Effectivity is not sufficient against *Leishmania* and *Trypanosoma* parasites.                                                            | (135)  |
|                               | 2006 | Acyclic uracil derivatives with similar or better antiplasmodial properties than those in Ref (135) especially with regards to selectivity.  
$K_i$ of the best active compound was 0.2 μM.                                                                                                   | (136)  |
|                               | 2007 | Study of PfDUT ligand binding.  
No significant conformational changes upon binding are inferred based on ITC measurements.                                                               | (137)  |
<table>
<thead>
<tr>
<th>Organism</th>
<th>Year</th>
<th>Summary</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Plasmodium falciparum</em></td>
<td>2009</td>
<td>Tritylated uracil acetamide derivatives containing amide bond between the β-C and N-1 of uracil ring were found to be weak inhibitors of PfDUT.</td>
<td>(138)</td>
</tr>
<tr>
<td></td>
<td>2010</td>
<td>Study on PfDUT and hDUT kinetics. Specific product inhibition of the plasmodial dUTPase compared to that of the human enzyme was caused by the substituent at the C-5 position of the uracil ring.</td>
<td>(139)</td>
</tr>
<tr>
<td></td>
<td>2011</td>
<td><em>In vitro</em> high-throughput screening for PfDUT inhibitors in a compound library of commercially available non-proprietary compounds did not identify any hits out of 3086 molecules.</td>
<td>(126)</td>
</tr>
<tr>
<td></td>
<td>2011</td>
<td>Modification on the uracil ring of the substrate analogue inhibitors impaired the effect of those on PfDUT. While there is room for variation of the 5’-trityl group and the 3’-substituent.</td>
<td>(127)</td>
</tr>
<tr>
<td></td>
<td>2011</td>
<td>Testing β-branched acyclic uridine analogues as PfDUT inhibitors. Kᵢ of the best inhibitor was 0.5 μM. This showed more than 200-fold selectivity compared to hDUT and EC₅₀= 0.61 μM growth inhibition of <em>P. falciparum</em>.</td>
<td>(128)</td>
</tr>
<tr>
<td></td>
<td>2011</td>
<td>Mutational analysis of residues important in the binding of uracil based inhibitors containing trityl groups at the 5’-position. F46A mutation of PfDUT leads to an increase in Kᵢ values while K96A mutation has an opposite effect.</td>
<td>(140)</td>
</tr>
<tr>
<td></td>
<td>2011</td>
<td>Testing lower lipophilicity and molecular weight diphenyl substituted inhibitors of PfDUT. Slightly decreased activity against both dUTPase and parasite than the corresponding trityl derivatives was observed.</td>
<td>(141)</td>
</tr>
<tr>
<td></td>
<td>2013</td>
<td>Characterization of novel conformationally restrained amide derivatives to overcome entropic disadvantages of former PfDUT inhibitors. Inhibitors showed similar or greater potency but lower selectivity (&lt;40x) in cellular assays, than the previous drug candidates against PfDUT.</td>
<td>(142)</td>
</tr>
<tr>
<td></td>
<td>2015</td>
<td>2D- and 3D-QSAR model using the LQTA-QSAR methodology on a series of PfDUT inhibitors with high predictive power facilitates the design of new compounds with higher antimalarial bioactivities.</td>
<td>(129)</td>
</tr>
</tbody>
</table>
It has been well-established that the modulation of dUTPase activity interferes with thymidylate synthase inhibition in cancer therapeutics outcome (28, 57–62). The first attempts to design human dUTPase inhibitors, which significantly differ from substrate analogue inhibitors, $\alpha,\beta$-imido-dUTP and $\alpha,\beta$-methylene-dUTP, resulted in only inactive compounds. Acyclic uracil derivatives were found to be effective ($K_i < 10 \mu M$) against both human and plasmodial dUTPases (136, 141). Combined efforts of scientists from academia and Taiho Pharmaceutical Co. Ltd. lead to highly effective dUTPase inhibitors with favorable pharmacokinetic characteristics (143–146). One of the compounds designed by the same group (TAS-114\textsuperscript{2}), successfully passed the Phase I trials (147), so that is a promising candidate to be used in cancer therapy in combination with thymidylate synthase inhibitors (57).

**Table 4.** Recent results on the inhibitor design against the human dUTPase

<table>
<thead>
<tr>
<th>Year</th>
<th>Summary</th>
<th>Ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>2001</td>
<td>Substitution at the 5’ position of dU resulted in derivatives with no inhibitory effect on cellular extracts from the human lung carcinoma cell line, which demonstrates high dUTPase activity.</td>
<td>(148)</td>
</tr>
<tr>
<td>2005</td>
<td>5’-trityl substituted uridine derivatives showed $K_i= 18 \mu M$ and 46 $\mu M$ against hDUT although those are more effective against PfDUT.</td>
<td>(75, 135)</td>
</tr>
<tr>
<td>2006</td>
<td>Several acyclic uracil derivatives showed lower than 10 $\mu M$ $K_i$ against hDUT.</td>
<td>(136)</td>
</tr>
<tr>
<td>2006</td>
<td>Some molecules from a library of triskelion uracil derivatives showed $3.3 \pm 1.1 \mu M$ and $5.5 \pm 1.1 \mu M$ IC\textsubscript{50} values.</td>
<td>(149)</td>
</tr>
<tr>
<td>2011</td>
<td>Certain acyclic uracil derivatives exhibited $1.4$ and $5.7 \mu M$ $K_i$ against hDUT, however those were even stronger inhibitors of PfDUT.</td>
<td>(141)</td>
</tr>
<tr>
<td>2012</td>
<td>Various trials of hDUT inhibitor development resulted in uracil derivatives of submicromolar IC\textsubscript{50} and perturbation of in vivo tumor growth in combination with 5-FU. The most potent compound showed IC\textsubscript{50} = 39 nM, EC\textsubscript{50} = 66 nM and favorable pharmacokinetic profile.</td>
<td>(143–146)</td>
</tr>
<tr>
<td>2014</td>
<td>Successful Phase I. trials of the human dUTPase inhibitor candidate (TAS-114) for treatment of colorectal cancer in combination with thymidylate synthase inhibitors.</td>
<td>(147)</td>
</tr>
<tr>
<td>2017</td>
<td>Reduction of dUTPase activity by RNAi and inhibitors increased the toxicity of 5-FU in case of various human cell lines. One of the uracil derivatives introduced in this study (Compound 2) demonstrated IC\textsubscript{50} of 25 nM.</td>
<td>(57)</td>
</tr>
<tr>
<td>2018</td>
<td>Crystal structure of the human dUTPase with Compound 2 of Ref 57</td>
<td>PDB:5H4J\textsuperscript{3}</td>
</tr>
</tbody>
</table>

\textsuperscript{2} Structure of TAS-114 has not yet been published.

\textsuperscript{3} http://www.rcsb.org/structure/5H4J
2 Aims

The dUTPase enzyme as a part of the preventive DNA repair, prohibits uracil misincorporation into the genome by hydrolysis of dUTP. As a consequence of this important role, dUTPase is essential for most organisms, and therefore it is a well-established target for antiparasitic drugs and cancer therapy (27). It has been shown that Stl, a Staphylococcus aureus pathogenicity island repressor protein interacts with trimeric dUTPase of the Φ11 phage (16). We have revealed that Stl also acts as an effective proteinaceous inhibitor of this dUTPase (17), which might constitute a potential for design of a specific inhibitor for in vivo studies on the effect of enzyme inhibition or later can serve as a starting point for the development of proteinaceous drugs. Such strategy was successfully applied in the case of the uracil-DNA glycosylase inhibitor UGI from bacteriophage PBS2, which provided potent inhibitory effect against UNGs from various sources, including the human enzyme (150–154). In addition to this, the better understanding of the process of horizontal gene transfer might contribute to the success of the fight against antibiotic-resistant S. aureus strains.

Towards this end, we focused our research to answer the following questions:

1. What structural characteristics does the Stl protein possess? Can we define domains within the Stl repressor, which are responsible for DNA binding, dimerization or protein-protein interaction as it has been demonstrated for other well-known bacteriophage repressors?

2. What is the role of the phage-specific insert of Φ11 phage dUTPase in Stl-binding, and derepression activity? Does Stl interact with dUTPases lacking the phage-specific insert? Based on the structural similarity of the core region of dUTPase isoenzymes of different species is it possible that the Stl protein does inhibit trimeric dUTPases from other organisms than Staphylococcal phages?

3. What is stoichiometric ratio of Stl and dUTPase in their complex(es)? What is the interaction surface of the dUTPase-Stl complexes? What is the structural background of the mutual inhibition of the Stl and dUTPase biological function in the complex(es) of the two proteins?

4. How does the Stl repressor interact with trimeric and dimeric dUTPase derepressor proteins? Is the interaction localized to distinct regions of Stl because of the large structural differences between trimeric and dimeric dUTPases or does Stl have a universal dUTPase-inhibitor segment?
Following an integrated structural biology approach we studied wild-type and specific mutants of Stl and dUTPase and their complexes by native gel electrophoresis, electrophoretic mobility shift assay, chemical crosslinking, steady state activity measurements, isothermal titration calorimetry, electrospray ionization native mass spectrometry, synchrotron radiation circular dichroism and most prominently hydrogen-deuterium exchange mass spectrometry (HDX-MS) and size-exclusion chromatography in line with small-angle X-ray scattering (SEC-SAXS) methods to provide answer for these questions.
3 Materials and Methods

3.1 Cloning and mutagenesis

Expression vectors of Stl (17), human dUTPase (155), Φ11 phage dUTPase (156) and its mutants (84) have been created previously. Expression vectors of the C-terminal segment of Stl (Stl-CTD), and φNM1 phage dUTPase were created by molecular cloning, while those of the N-terminal segment of Stl (Stl-NTD), Q40A,N41A mutant of Stl (Stl-AA) and His-tag free Stl (Stl-dHis) were created by mutagenesis.

3.1.1 Cloning

The gene encoding residues 85–263 of Stl was amplified by PCR from pGEX-4T-1 vector encoding GST-Stl protein by primers Stl-CTD-F and Stl-CTD-R (cf. Appendix Table A1.). The resulting insert was cloned in frame with the amino-terminal GST tag and the thrombin cleavage site to an empty pGEX-4T-1 vector by using EcoRI and XhoI restriction sites to create an expression vector of Stl-CTD (157).

The gene encoding φNM1 phage dUTPase was PCR amplified from a pET21a vector (14), provided by the courtesy of Prof. Terje Dokland, with primers NM1-GST-F and NM1-GST-R (cf. Appendix Table A1.). Then the insert was cloned into to an empty pGEX-4T-1 vector in frame with the amino-terminal GST tag and the thrombin cleavage site between BamHI and XhoI restriction sites.

The sequences of all constructs were verified by DNA sequencing performed by Eurofins MWG Operon and Microsynth AG. The sequences of the expressed proteins are summarized in Table A3 in the Appendix.

3.1.2 Mutagenesis

The expression vectors encoding the N-terminal 84 residues of Stl (Stl-NTD), Q40A,N41A mutant of Stl (Stl-AA) and His-tag free Stl (Stl-dHis) were created by QuikChange site-directed mutagenesis (Stratagene) using suitable mutagenic primers (cf. Appendix Table A2) and the original pGEX-4T-1 vector encoding Stl in frame with the amino-terminal GST tag and the thrombin cleavage site (157, 158). The resulting constructs were validated by DNA sequencing at Eurofins MWG Operon and Microsynth AG. The sequences of the expressed proteins are shown in Table A3 in the Appendix.
3.2 Protein expression and purification

Protein samples used throughout this study were produced and purified by myself according the following protocols, with the single exception of MtDUT, which was expressed and purified by Dr. Judit Eszter Szabó and Paula Dobrotka (159).

3.2.1 Protein expression

The sequences of the expressed proteins are listed in Table A3 in the Appendix. Proteins were expressed in *E. coli* BL21 Rosetta (pLysS) cells (Novagen) as described previously (17, 84, 155–158). Briefly transformed bacterial cells were propagated at 37 °C in 500 ml Luria broth medium till exponential growth phase (OD<sub>600</sub> ~ 0.6) and then protein expression was induced by addition of 0.5 mM isopropyl-β-D-thiogalactoside (IPTG). In case of the expression of GST-tagged proteins (all Stl derivatives and φNM1 phage dUTPase) from pGEX-4T-1 vectors cells were grown for 4 h at 30 °C. dUTPases were expressed in cultures grown at 37 °C for 4 h. Finally, the cells were harvested by centrifugation for 30 min at 3000 g and stored at -80 °C.

3.2.2 Protein purification

3.2.2.1 Cell lysis

Cell pellets were solubilized using Potter-Elvehjem homogenizer in 20 ml lysis buffer (composition defined case by case at the following sections) supplemented with 2 mM dithiothreitol (DTT), 1% Triton X-100, ca. 2 µg/ml RNase and DNase and one tablet of Complete ULTRA Tablets, Mini, EDTA-free protease inhibitor. Cell suspensions were sonicated (4 x 60 s), and centrifuged (16000 g for 30 min).

3.2.2.2 Purification of GST-tagged proteins

For purification of GST-tagged proteins (Stl, Stl-CTD, Stl-AA, φNM1DUT) the supernatant resulting from cell lysis procedure was loaded on a benchtop glutathione-agarose affinity-chromatography column (GE Healthcare) pre-equilibrated in lysis buffer (50 mM HEPES, pH=7.5, 200 mM NaCl). The column was washed with 50 mL lysis buffer and subsequently 80 Cleavage Units thrombin (GE Healthcare) was added in 4 mL lysis buffer to perform on-column cleavage for the removal of GST-tag. After overnight cleavage purified proteins were obtained in the flow-through. The column was regenerated by elution of the cleaved GST-tag applying 20 ml solution containing 100 mM reduced glutathione, 50 mM TRIS, pH=8.0 (157).
3.2.2.3 **Purification of His-tagged human dUTPase**

Purification of hDUT was performed based on the protocol developed previously in our group (111). The supernatant resulting from centrifugation after cell lysis was applied onto an 5 ml Ni-NTA column (Novagen) pre-equilibrated with lysis buffer (50 mM TRIS·HCl, pH=8.0, 300 mM NaCl) supplemented 15 mM imidazole. First, the contaminants were removed by washing the column with ten bed volumes of low salt and high salt buffers (50 mM HEPES pH=7.5, supplemented with 30 mM KCl or 300 mM KCl, respectively), then hDUT was eluted with 500 mM imidazole dissolved in low salt buffer. Afterwards the eluted hDUT was dialyzed overnight against 50 mM HEPES, pH=7.5, 300 mM NaCl, 5 mM MgCl₂ buffer. This was followed by gel-filtrated in the dialysis buffer on a Superdex 200 10/300 GL column in an AKTA Purifier (GE Healthcare) system. Tag-free human dUTPase was used (hDUTONTR) in the native mass spectrometry and the standard SAXS measurements. The His-tag was removed by adding 40 unit thrombin to 1 ml of 18 mg/ml dialyzed hDUT and incubated overnight at 20 °C prior to gel filtration (158).

3.2.2.4 **Purification of tag-free Φ11 phage dUTPases**

Purification of Φ11DUT and Φ11DUTΔ101G–122Q was performed as described previously (17, 156), briefly, supernatant resulting from centrifugation of cell lysate in 20 mM HEPES, pH=7.5, 100 mM NaCl, 5 mM MgCl₂, 10 mM β-mercaptoethanol was loaded on a Q-Sepharose column (5 ml) equilibrated in lysis buffer and the protein were subsequently eluted by applying 30 ml of a linear gradient up to 1 M NaCl. The dUTPases dissociated from the stationary phase at 0.3–0.5 M NaCl concentration. A second purification step was performed on a size exclusion column (Superdex 200 10/300 GL, GE Healthcare) with a molecular weight separation range between 10-600 kDa in lysis buffer containing 300 mM NaCl. AKTA Purifier (GE Healthcare) system, with Unicorn software (GE Healthcare) was used for both chromatographic steps, protein fractions were collected based on monitoring absorbance at 280 and 260 nm.

3.2.2.5 **Quality control and protein quantification**

Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) was utilized to check the quality of purified proteins. Samples were only used for further experiments if the resulting gel suggested at least 90% purity without detectable degradation of the protein (i.e. if the protein appeared as single band). Protein concentration was calculated based on the absorbance of the protein samples at 280 nm measured using a NanoDrop 2000 UV-Vis spectrophotometer and extinction coefficients based on the protein sequence estimated with the
Protparam on-line tool (http://web.expasy.org/protparam). The extinction coefficients of the expressed proteins are shown in Table A3 in the Appendix. The DNA contamination of the samples was checked based on UV absorbance determined at 260 nm, samples were only used for subsequent experiments if the ratio of absorbance values measured at 260 nm and 280 nm was lower than 0.8. All protein preparations matching quality criteria were either used freshly or were flash-frozen in small aliquots in liquid nitrogen, and stored at -80 °C. If required, protein samples were concentrated on Millipore centrifugal filters (10 kDa cutoff).

3.3 Protein characterization

3.3.1 Homology modeling and in silico predictions

The 3D homology model of Stl was constructed using the Phyre2 Server in intensive mode (157). Seven templates were selected by the program to model Stl protein based on heuristics to maximize confidence, percentage identity and alignment coverage (cf. Appendix Table A4) (160). Five out of the seven templates covered >90% of the Stl sequence, while the other two templates provided only partial coverage of the sequence, but with greater local similarity. In the resulting model, 97% of residues were modeled at >90% confidence. Homology prediction was performed by HHpred (161), and 3D structural models were generated subsequently with Modeller in two different ways: i) applying automatic template selection and ii) setting the Phyre2 hits manually selected as templates from the HHpred list (162). Functional domain search was accomplished within Pfam and NCBI conserved domain databases (163, 164). The possible position of the helix-turn-helix (HTH) DNA binding motif in the sequence was predicted by NPS@ server (165). Flexibility prediction during construct design was performed by using GeneSilico MetaDisorder service (166). To compare the homology model to the structural information obtained based on the SRCD spectroscopy experiments, the secondary structure composition of the Phyre2 model, including the helix content, was assigned using the DSSP algorithm (167) and the BeStSel and CONTIN secondary structure definitions (168, 169).

Since reasonable template structures were available in the case of the Φ11DUT\textsubscript{Δ101G-122Q} and φNM1DUT homology models were generated for these proteins by the SWISS-MODEL webserver (170–172) based on manually selected templates. The structure of the full-length Φ11DUT (PDB ID:4GV8) (84) served as template for the structure prediction of the truncated Φ11DUT\textsubscript{Δ101G-122Q} protein. The 3D homology model of φNM1DUT was constructed using the structure of the S. aureus phage φDI dUTPase (PDB ID: 5MYD) (15) as a template.
Reliability of the resulting model is established by the 75% sequence identity of the two proteins.

3.3.2 Synchrotron radiation circular dichroism (SRCD) measurements and CD spectrum analysis

SRCD spectrum of Stl was recorded at the DISCO beamline of SOLEIL French Synchrotron Facility (157). The concentration of Stl was 2.1 mg/ml in a buffer consisting of 50 mM HEPES, 200 mM NaCl, pH=7.5. A CaF$_2$ cell with a path length of 6.13 μm was used. 38 scans were accumulated in the 180-270 nm wavelength range at 1 nm steps with a lock-in time constant of 300 ms and integration time of 1200 ms. In this wavelength range and path length, the photomultiplier voltage did not exceed the 700 V limit. After baseline subtraction, the spectrum was corrected with the CSA calibration (173).

To estimate the secondary structure content, the resulting CD spectrum was analyzed by the BeStSel (168) and CONTIN methods (169). These algorithms distinguish two types of spectrally different helical components, helix1 and helix2. Helix1 is the regular, middle part of the helix where all the hydrogen bonds within the main chain amido groups are formed and helix2, a so-called “distorted helix”, consisting of the two-two residues at the ends of the helical segment with unsatisfied H-bonding. Taking into account the helix2 content and the helix2/helix1 ratio the number and average length of helices in the protein could be predicted (174). Note that BeStSel defines helix as α-helix while the CONTIN definition includes α-helix and 3_{10}-helix. As in general the 3_{10}-helix content of proteins is low we expected similar results for the two algorithms.

3.3.3 Chemical crosslinking

Samples of full-length Stl and Stl-CTD of 40 μM concentration were incubated with 0.1-0.3 mM disuccinimidyl suberate (DSS) at 20 °C for 60 min (158). Samples Stl and Φ11DUT or φNM1DUT of 20 μM concentration and the dUTPase-Stl mixtures of 1:1 molar ratio (20 μM total protein concentration) were incubated with 20 mM DSS at 20°C for one hour (175). The crosslinking reactions were quenched with the addition of 5 μL 100 mM (pH=7.5) TRIS buffer to 40 μL of samples and then those were analyzed by SDS-PAGE after electrolysis for 45 min at 200 V using a two-phase polyacrylamide gel; in which acrylamide concentration was 4% in the stacking gel and 12% in the resolving gel. Page Ruler prestained protein ladder (Thermo Fisher) was used as a molecular weight marker. Page Blue protein staining solution containing Coomassie Brilliant Blue dye was used to stain the gel.
3.3.4 Electrophoretic mobility shift assay (EMSA)

EMSA experiments were performed to test the DNA-binding ability of Stl-CTD and Stl-AA proteins (157), using a 57-mer dsDNA oligonucleotide encoding the sequence of the intergenic region between the str and the xis genes (later termed as str-xis in (176)). Proteins were mixed with 100 ng dsDNA in 20 μl total volume, in which final concentration of NaCl was set to 100 mM. After 15 min incubation at 4 °C, samples were loaded onto a polyacrylamide gel (8%), which was equilibrated with the TRIS-Borate-EDTA (TBE) assay buffer by a 1 h pre-electrophoresis at 100 V. Electrophoresis was performed for 70 min at room temperature at 150 V. Bands were detected with an Uvi-Tec gel documentation system (Cleaver Scientific Ltd., Ruby, UK) after staining with GelRed (Bioticum).

The same experimental setup was used to test the effect of addition of different dUTPases on the interaction of Stl to its cognate DNA binding site. In case of hDUT a 43-bp double-stranded oligonucleotide was used (later named as interR in (176)), which contains the same binding motif as the 57-mer dsDNA oligonucleotide utilized in the other experiments (158). The two strands of the both dsDNA oligonucleotides were custom synthesized by Eurofins MWG Operon and hybridized by controlled gradual cooling after 5 min incubation on 95 °C.

3.3.5 Native polyacrylamide gel electrophoresis (native PAGE)

Native gel electrophoresis experiments were performed in a two-phase polyacrylamide gel; acrylamide concentration was 4% in the stacking gel and 8% in the resolving gel (17, 158, 159). The gel was subjected to pre-electrophoresis with constant voltages of 100 V in assay buffer (TRIS-HCl, pH=8.5 for hDUT and pH=8.7 otherwise) for 30 min. Then 25 μl of the premixed samples was applied on the gel and electrophoresis was performed at 150 V for 1.5 h (phage dUTPases), 2.5 h (MtDUT) and 5 h (hDUT) depending on the pI difference between the specific dUTPases and Stl. The apparatus was placed on ice during electrophoresis in order to avoid heat-induced denaturation. The gel was stained with Comassie-Brilliant Blue dye.

3.3.6 Steady-state kinetics experiments

Steady state enzymatic activity of the dUTPases in the presence and absence of Stl was measured by following a standard protocol established within our research group (17, 158, 159, 177). During the dUTPase enzymatic action dUTP is hydrolyzed to dUMP and PP, and the reaction also results in proton release to the solution, which was followed continuously at
559 nm at 20 °C using a Specord 200 spectrophotometer (Analytic Jena, Germany) and 10 mm path length thermostatted plastic cuvettes. Reaction mixtures contained 50 nM dUTPase enzyme and varying concentrations of Stl in a buffer consisting of 1 mM HEPES (pH=7.5), 5 mM MgCl₂, 150 mM KCl, 40 μM Phenol Red indicator. After 5 min pre-incubation of the two proteins in the assay buffer the reaction was started by the addition of 30 μM dUTP. Initial velocity was determined from the slope of the first 10% of the progress curve. Stl inhibition data were fitted to the quadratic binding equation using Origin 8.1 software (OriginLab Corp., Northampton, MA, USA) describing 1:1 stoichiometry for the dissociation equilibrium with no cooperativity.

3.3.7 Isothermal titration calorimetry (ITC)

The ITC experiments were carried out at 20 °C on a Microcal ITC₂₀₀ instrument (Malvern Instruments, Malvern, UK). Prior to the measurements hDUTONTR and Stl samples were dialyzed overnight at 4 °C in a buffer of 20 mM HEPES (pH=7.5), 300 mM NaCl, 5 mM MgCl₂, 1 mM TCEP. In the experimental setup, the cell of the instrument was filled with Stl (15 μM) and the syringe with hDUTONTR (200 μM). The titrations were performed with the injection syringe rotating at 750 r.p.m. and included a series of 20 injections spaced 3 min apart from each other with injection volumes of 0.5 μl for the first titration and 2 μl for the subsequent 19 titrations. The apparent dilution heat observed at the end of the titration was subtracted from the integrated heat data. The dissociation constant (K_D), stoichiometry (N), enthalpy (ΔH) and entropy (ΔS) were determined from the analysis of the corrected data by using MicroCal ORIGIN 7.5 software, following the instructions of the manufacturer. The best fit to the data was obtained applying the one set of independent sites binding model of the software. This model is appropriate to describe binding to any number of sites (N) assuming that the sites have the same K_D and ΔH. Average and standard deviation (SD) of the fitted parameters were calculated from three parallel measurements (158).

3.3.8 Electrospray ionization mass spectrometry

A commercial Waters QTOF Premier mass spectrometer (Waters, Milford, MA, USA) equipped with electrospray ionization source (Waters, Milford, MA, USA) operated in positive ion mode was used to study the protein complexes. Proteins mixed in ca. 20 μM final concentration were subjected to buffer exchange to 200 mM NH₄HCO₃ buffer applying Vivaspin® 500 Polyethersulfone centrifugal concentrators of 10 kDa weight cutoff. Mass spectra were measured under native conditions: namely, the ions were generated from aqueous
5 mM NH$_4$HCO$_3$ buffer solution (pH=7.5) containing the human dUTPase (hDUT$^{\text{ONTR}}$), Stl or both protein constructs at ca. 0.5 µM monomer concentration. These conditions favor the transfer of the protein complexes from solution into the gas phase. The capillary voltage was set to 2600-2800 V, the sampling cone voltage was 125 V and the temperature of the source was kept at 80 °C, collision cell pressure was 3.43×10$^{-3}$ mbar and ion guide gas flow was 35 ml/min. Mass spectra were recorded applying the software MassLynx 4.1 (Waters, Milford, MA, USA) in the 1000-8000 m/z mass range (17, 158, 175).

3.3.9 Hydrogen-deuterium exchange mass spectrometry (HDX-MS)

HDX-MS measurements were performed on a Synapt G2Si HDMS coupled to an Acquity UPLC M-Class system with HDX and automation (Waters Corporation, Manchester, UK). The deuterium uptake of the Φ11DUT, Φ11DUT$^{\text{A101G-122Q}}$, hDUT, φNM1DUT and Stl proteins was determined using a continuous workflow with labelling taking place at 20 °C. Each protein was solubilized in 20 mM HEPES, 300 mM NaCl, 5 mM MgCl$_2$, pH=7.5 to a working concentration of 10-20 µM. Initiation of deuterium labelling was performed by diluting 5 µl of each protein sample into 95 µl of labeling buffer (20 mM HEPES, 300 mM NaCl, 5 mM MgCl$_2$ in D$_2$O, pD 7.1). After various incubation times (1, 10, 100 min), samples were quenched in 2.4% formic acid at 1 °C to prohibit further deuteration or back-exchange and were then digested on-line with a Waters Enzymate BEH pepsin column at 20 °C. Peptides were trapped on a Waters BEH C18 VanGuard pre-column for 3 minutes at a flow rate of 200 µl/min in 0.1% formic acid (pH=2.5) before being applied to a Waters BEH C-18 analytical column. Peptides were eluted by developing a linear gradient with 0.1% formic acid in acetonitrile, pH=2.5 at a flow rate of 40 µl/min. All trapping and chromatography stages of the trapping experiment were performed at 0.5 °C to minimize back-exchange. Determination of the bound HDX profile of each protein was carried out by pre-mixing the proteins at approximately equimolar concentrations (in terms of monomers). MS data were acquired using an MS$^E$ workflow in HD mode with extended range allowing to reduce detector saturation and maintain peak shapes. Non-deuterated reference acquisitions were obtained in sextuplicate for each protein along with labelling acquisitions, which were obtained in triplicate. The mass spectrometer was calibrated using NaI and MS data were obtained with lock mass correction using Leu-enkephalin.

Peptides were assigned with the ProteinLynx Global Server (PLGS) (Waters Corporation, Manchester, UK) software package, with the deuterium uptake of each assigned peptide being determined with DynamX v3.0 (Waters Corporation, Manchester, UK). Evaluation of the data fitting as well as determining the error of each dataset were performed as previously described.
Difference plots ($\Delta$mass) were prepared for each protein by subtraction of the bound from the unbound HDX-MS patterns of each protein. The average $\Delta$mass across all peptides at each residue was then calculated to obtain confidence bands. Residues with values exceeding the 95% confidence bands are noted and defined as part of the interaction surface of Stl and dUTPases. In all cases sequence coverage was above 90% and redundancy was above 3. Summarizing the signals obtained for overlapping peptides provided sub-peptide level resolution data and enabled us to gain information about specific regions involved in complex formation. We assumed that all subunits were involved in the protein-protein interactions such that the interacting surfaces were projected equally across all subunits.

3.3.10 Small angle X-ray scattering

Synchrotron radiation small angle X-ray scattering data were collected on the EMBL P12 beamline of the storage ring PETRA III (DESY, Hamburg) (Appendix Tables A5-A6), using a PILATUS 2M pixel detector (DECTRIS, Switzerland) (158). The hDUT and Stl (Stl-dHis) proteins were dialyzed overnight into a buffer consisting of 50 mM HEPES, pH=7.5, 300 mM NaCl, 5 mM MgCl$_2$. SAXS data for individual proteins were measured in standard mode while the samples were flowing through a temperature controlled capillary (1.2 mm I.D.) at 20 °C and 20 frames of 0.05 s exposure time were collected. The hDUT-Stl mixture (100 µL of ca. 8 mg/ml mixture with 1:1 hDUT:Stl molar ratio) was injected onto a GE Healthcare S200 Increase 10/300 (24 ml) column equilibrated in the dialysis buffer and in-line SEC-SAXS performed at a flow rate of 0.5 ml/min. A total of 3600 x 1 second SAXS data frames were recorded during elution.

The sample-to-detector distance was 3.1 m, covering a range of momentum transfer $0.01$ Å$^{-1} \leq s \leq 0.46$ Å$^{-1}$ ($s= 4\pi\sin\theta/\lambda$, where $2\theta$ is the scattering angle, and $\lambda= 1.24$ Å is the X-ray wavelength). No radiation damage was detected based on comparison of successive frames. Data from the detector were normalized to the transmitted beam intensity, averaged, placed on absolute scale relative to water and the scattering of buffer solutions subtracted. All data manipulations were performed using PRIMUSqt and the ATSAS software package (179). SEC-SAXS data were analyzed using CHROMIXS (180).

Data analysis and modelling were performed by Dr. Haydyn. D. T. Mertens following the procedure described below (158). The forward scattering ($I_0$) and radius of gyration ($R_g$) were determined from Guinier analysis (181) assuming that at very small angles ($s \leq 1.3/R_g$) the intensity is represented as $I_s = I_0 \exp(-(sR_g)^2/3))$. These parameters were also estimated from the full scattering curves applying the indirect Fourier transform method implemented in the
program GNOM (182), along with the distance distribution function \( p(r) \) and the maximum particle dimensions \( D_{\text{max}} \). Molecular masses of solutes were estimated from SAXS data by comparing the extrapolated forward scattering with that of a reference solution of bovine serum albumin (Merck KGaA, Darmstadt, Germany), the hydrated-particle/Porod volume \( V_p \), where molecular mass is estimated as 0.588 times \( V_p \), and from the excluded solvent volumes, \( V_{\text{ex}} \) obtained from \textit{ab initio} modeling in the program DAMMIF (183). Molecular mass estimation from in-line SEC-SAXS data was performed by DATMOW (184) routine implemented in ATSAS. The program CRYSOL was used for computation of theoretical scattering intensities (185).

Low resolution \textit{ab initio} shapes were reconstructed from SAXS data using the programs DAMMIF (183), which represents the macromolecule as a densely packed interconnected configuration of beads or chain-like ensemble of dummy residues, respectively, that best fits the experimental data \( I_{\text{exp}}(s) \) by minimizing the discrepancy according to the equation below, where \( N \) is the number of experimental points, \( c \) is a scaling factor and \( I_{\text{calc}}(s_j) \) and \( \sigma(s_j) \) are the calculated intensity and the experimental error at the momentum transfer \( s_j \), respectively.

\[
\chi^2 = \frac{1}{N-1} \sum_j \left[ \frac{I_{\text{exp}}(s_j) - cI_{\text{calc}}(s_j)}{\sigma(s_j)} \right]^2
\]

For both 3:3 and 3:2 hDUT:Stl complex data models were generated initially without symmetry (P1) and then enforcing P3 (3:3) or P2 (3:2) symmetries. The symmetry constrained models were checked to be consistent with the P1 reconstructions through superposition in SUPALM (186). As a quantification of the reliability of models, we considered effective resolution determined by the recently published Fourier shell correlation (FSC) approach (187) and normalized spatial discrepancy (NSD) values (188), which represent the measure of the real-space variation of the models. Ideal superposition of two compact and rigid structures with the same low resolution shapes gives NSD around 1. For the models containing flexible regions of significant length, e.g. flexible N/C termini like in the reported case, can still be considered similar in the overall shape if NSD is below 3.

Human dUTPase crystal structures (PDB IDs are 1Q5U (71) and 3EHW) were used as templates for modeling of the trimeric core of hDUT:Stl complexes. Rigid bodies of Stl were generated applying the protein structure prediction server Phyre2 (189) and dimer models from that were created by M-ZDOCK server (190) using C-terminal region (T87–N267) of the Stl model as the dimer interface based on our chemical crosslinking results (Figure 16). The flexibility of the Stl dimeric assembly was determined using EOM (186), where a genetic
algorithm is used to select an ensemble of best fitting configurations from a randomly generated pool. Multi-step rigid body refinement of hDUT and hDUT-Stl complexes in both 3:3 and 3:2 stoichiometries was performed by the program CORAL (179), where a simulated annealing (SA) based search of subunit arrangements and orientations, stoichiometry and conformations of missing terminal loops was conducted to fit the experimental SAXS data. In the first step, using the SEC-SAXS region 1 data, the position of the hDUT trimer was fixed and ambiguous distance restraints (10 Å) based on the potential contacts identified by hydrogen-deuterium exchange mass spectrometry (HDX-MS) defined between each hDUT (residues A37–R44 and L88–H92) and Stl monomer (residues Y98–Y113). For hDUT the interfacial residues identified by HDX-MS were filtered by surface accessibility (cut-off >10 Å²) based on the crystal structure (1Q5U) from the regions with negative HDX signals (H34–L50 and L88–G110). Refinement was conducted in CORAL applying P3 symmetry to generate a symmetric complex with 3:3 hDUT:Stl stoichiometry, to add missing terminal residues and to allow a flexible linker between the amino and carboxy-terminal domains of Stl. To generate a 3:2 hDUT:Stl complex, subsequent refinement was performed in CORAL against the major SEC-SAXS peak data (region 2) using the rigid body model of the 3:3 complex as input with one Stl monomer removed and the hDUT:Stl interface fixed. Multiple runs of modeling based on SAXS data were performed to verify the stability of the solution, and to establish the most typical 3D reconstructions using DAMAVER (191). From the top 20 resulting modelling based on SAXS and HDX-MS data the best few models were selected based on other experimental observations. SAXS data and the selected final models have been deposited at the SASBDB (www.sasbdb.org) with accession codes SASDC57, SASDC67, SASDC77, SASDC87.
4 Results and Discussion

4.1 Structural and functional characterization of Stl

4.1.1 Secondary structure of Stl

As there was no structural information available about the Stl repressor, first we have studied the secondary structural elements of the protein experimentally by synchrotron radiation circular dichroism (SRCD) spectroscopy (157), which is a well-established method for assessing several characteristics of protein conformation. The far-UV wavelength range (180-240 nm) of the CD spectra reflects the secondary structure composition (169, 192), while the near-UV wavelength range (250-380 nm) changes in tertiary structure can be detected (193). Compared to conventional light sources, synchrotron radiation offers extended wavelength range and increased signal/noise ratio in case of CD spectroscopy. This is especially beneficial for measurement of proteins, like Stl, which are sensitive to low salt conditions, since NaCl has high far-UV absorption, thus signals are superimposed on the higher background.

The observed spectral shape was characteristic of α-helical proteins (Figure 12, Panel A). BeStSel (168) and CONTIN (169) algorithms were applied for quantitative analysis of the secondary structure composition. Based on the CD spectrum the α-helix content of Stl was estimated to be 67.6% and 63.3%, respectively (Table 5). According to the prediction of both algorithms, the β-sheet content is low, close to zero within the accuracy of the methods.

Table 5. Secondary structure estimation of the Stl protein from the synchrotron radiation CD spectrum and in silico homology modelling (157).

<table>
<thead>
<tr>
<th></th>
<th>BeStSel CD analysis</th>
<th>Model</th>
<th>CONTIN CD analysis</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Helix1</td>
<td>44.6</td>
<td>40.0</td>
<td>44.5</td>
<td>42.5</td>
</tr>
<tr>
<td>Helix2</td>
<td>23.0</td>
<td>21.4</td>
<td>18.8</td>
<td>21.4</td>
</tr>
<tr>
<td>β-sheet</td>
<td>3.6</td>
<td>0.0</td>
<td>6.3</td>
<td>0.0</td>
</tr>
<tr>
<td>Turn</td>
<td>8.4</td>
<td>7.5</td>
<td>9.0</td>
<td>10.4</td>
</tr>
<tr>
<td>Others</td>
<td>20.5</td>
<td>31.1</td>
<td>21.4</td>
<td>25.7</td>
</tr>
<tr>
<td>Number of helices</td>
<td>16</td>
<td>15</td>
<td>13</td>
<td>15</td>
</tr>
<tr>
<td>Average helix length</td>
<td>11.8</td>
<td>11.5</td>
<td>13.5</td>
<td>11.9</td>
</tr>
</tbody>
</table>

* BeStSel and CONTIN algorithms were applied to estimate the secondary structure composition of Stl protein from the CD spectrum. The overall helix, β-sheet and turn+others contents are comparable although the two algorithms use different attributes to identify secondary structure components. The secondary structure content of the Phyre2 homology model was also calculated based on the BeStSel and CONTIN definitions (168, 169) by using the DSSP algorithm (167).
Figure 12. Experimental and in silico structure prediction of Stl (157). A) Normalized synchrotron radiation circular dichroism (SRCD) spectrum of Stl. Stl spectrum is with black, curves fitted by BeStSel and CONTIN algorithms are shown with red and blue dashed lines, respectively.

B) Homology model of the Staphylococcus aureus pathogenicity island repressor Stl. Cartoon representation of the homology model of the Stl protein created by Phyre2 Server (160). Based on the homology model Stl consists of 15 α-helices (> 60 %) connected with flexible loops, and seems to be divided into two segments: the amino-terminal segment (cyan and dark blue) and the carboxy-terminal segment (hotpink). A helix-turn-helix (HTH) DNA binding motif was predicted to reside within the amino-terminal segment by Pfam and NCBI CDD (163, 164). The position of the HTH according to NPS@ server is colored to dark blue (165).

In order to assess information about the 3D structure of Stl, in silico homology models were generated by Phyre2 Server (160) and Modeller (162). The predicted homology model provided by the Phyre2 Server (160) applying intensive mode is presented in Panel B of Figure 12, while the templates used for modeling are listed in Table A4 of the Appendix (157). Apart from a very short part of the N-terminus (residues 1–7) modeled by ab initio methods the 3D structural model of the Stl protein was created from templates with known structure exhibiting sequence similarity to the target. The high confidence of the alignment (> 95%) suggests that the core of the protein and the overall fold are modeled reliably, while the orientation of the surface loops is somewhat ambiguous.

To generate alternative structural models the Stl sequence was subjected to analysis by the HHpred server (161), which provided the list of the closest homologs of Stl. As all the templates of the Phyre2 model with the only exception of the human transcription factor Oct-1 were on the list, first we performed a 3D modeling by Modeller by choosing those as templates (162). This procedure provided a structural model for the most part of the protein, although the structure of the 33 residues at the C-terminal was not well-defined. According to both the Modeller and Phyre2 predictions the Stl is mostly α-helical and contains a helix-turn-helix (HTH) motif close to N-terminal of the protein, which is a typical DNA binding motif of
repressors, consisting of a helix responsible for specific recognition of the cognate DNA site linked to an other helix, which stabilizes the protein DNA interaction, oriented by a sharp turn (194–196). The overall structure and the position of the HTH within the models were similar even if the orientation of the helices at the C-terminal region of the protein differed between the two models (Appendix Figure A1). Since the predicted “all-α” fold type is characteristic for several protein superfamilies performing various functions and most of the templates showed only local sequence similarity to specific Stl segments, we did not analyze those in detail. In a second round we performed the prediction by Modeller with automatic template selection, which resulted in a well-defined structural model only for the N-terminal residues (1–75), which was indeed highly similar to the other two models.

We compared the 3D model of the entire protein created by Phyre2 server with the results of the SRCD spectra (Table 5). The composition of secondary structural elements and the estimated number of helices of Phyre-predicted 3D model agreed well with the data derived from the CD analysis, which justifies the reliability of the model. Thus we applied this 3D model as a starting point for experiment design. Based on the model it seems that the Stl protein consists of two individual segments (Figure 12, Panel B): the amino-terminal segment (cyan and dark blue) and the carboxy-terminal segment (hotpink), which can be responsible for different functions.

4.1.2 Domains of Stl and their functions

The function of Stl protein of *Staphylococcus aureus* pathogenicity islands (SaPIs) is similar to that of the master repressors of temperate phages as it controls the expression of Str, Xis and Int proteins, which are responsible for SaPI excision and replication (12). It has been demonstrated that Stl has a binding site between *stl* and *str* genes (16). We have also shown that two specific binding motifs of Stl reside within the *stl-str* intergenic region and one such motif also presents between the *str* and *xis* genes, which coincides with the repression of adjacent genes (176). Similarly to that of the 17 base pairs long palindromic binding sites of lambda phage CI repressor, Stl binds to a 23 base pairs long motif constituting two inverted repeat segment of 6 base pairs linked by a random spacer sequence of 5 base pairs, surrounded by overhangs of at least 3-bp on both ends (176, 197).

The functional similarities of the gene regulation mechanism of Stl and the CI repressor of temperate phages may establish the latter to serve as an adequate model for the main lifecycle regulator of SaPIs. The CI repressor proteins are constituted by two domains of separate functions, namely protein and DNA binding (198, 199). The protein binding domain generally
functions as an interaction surface for derepressor proteins and has a role in oligomerization of the repressor (199–202). In case of lambdoid phages the derepression is achieved through a RecA binding, which induces irreversible autoproteolytic inactivation of the maintenance repressor (203), while in various P2 related phages the repressor function is suspended reversibly via a noncovalent complex formation of the repressor with a derepressor protein (204–207). Besides serving as an interaction surface for effectors, the protein binding domain of the repressors is usually involved in oligomerization, which provides the opportunity for more complex and sensitive regulation mechanisms (208, 209). The DNA binding domain although performing the same function show different structural characteristics amongst specific repressors. In case of lambda phage CI repressor protein a helix-turn-helix (HTH) motif is responsible for DNA binding, while a winged-helix turn helix (wHTH) motif and the antiparallel β-strands of the ribbon-helix-helix (RHH) motif serve the same purpose in the MuR repressor protein of Mu phage and in Arc repressor of P22 phage, respectively (210, 211).

The 3D homology model of Stl suggests that the protein – similarly to the repressors discussed above – is seemingly divided into two segments, which may fold independently (Figure 12, Panel B). To identify the potential domains within the Stl repressor we estimated the flexibility of the Stl protein by MetaDisorder server (166) (Figure 13). The patterns provided by the different flexibility predictors differ significantly if specific residues are considered, still most of those agreed that residues between 90–260 form a somewhat less flexible segment.

Figure 13. Disorder prediction for Stl and designed protein constructs (157). Representative curves of flexibility prediction for Stl provided by MetaDisorder server (166). The mid-panel indicates the construct designed as horizontal bars. Bottom-panel shows the arrangement of α-helices along the sequence according to the Phyre2 model by purple rectangles except for the helix-turn-helix motif which is colored blue.
In parallel with the structural predictions, Pfam and NCBI CDD protein domain annotation engines (163, 164) located a helix-turn-helix DNA binding motif (HTH) at the amino-terminal part of the protein (between residues 15–68; with expectancy value of $7.97 \cdot 10^{-11}$). The position of the HTH has been narrowed down to residues 27–48 by NPS@ server (165), which matches perfectly with the position of the HTH in the 3D model (Figures 12 and 13). The flexibility of the Stl protein at the HTH position is somewhat more pronounced than in other regions, which might contribute to the recognition of cognate DNA binding sites (212). Similar pattern has been observed for HTH of the lambda phage CI repressor (Appendix Figure A2).

Based on these in silico predictions we hypothesized that Stl, like other well-known repressors, is built up by two segments, which may be responsible for distinct functions. According to the HTH predictions it is likely that the amino-terminal segment has a DNA binding function, while the less flexible carboxy-terminal segment may be involved in protein-protein interactions. It has been shown that Stl forms complex with phage dUTPases (16, 17), thus we speculated that the C-terminal segment is responsible for dUTPase binding. We have previously shown by native mass spectrometry that Stl forms dimers in solution (17), which evoked the idea that the C-terminal segment may also serve as a dimerization surface.

To experimentally verify these suggestions, we created constructs encoding the N-terminal and C-terminal segment of the protein and tested if those still have either DNA or dUTPase binding ability. The constructs were created based on the three-dimensional structural model provided by Phyre2 and the flexibility and HTH predictions. In case of the design of the amino-terminal construct the length of the DNA-bound bacteriophage repressor constructs deposited in the PDB was also taken into account (213–218).

The vector encoding the GST-tagged C-terminal (residues 85–263) segment was made by PCR based cloning. The resulting Stl-CTD protein was stable in solution after the cleavage of the fusion tag, which suggests that this segment might constitute an independently folded C-terminal domain. The construct encoding the GST-tagged N-terminal segment (residues 1–84) was derived by the insertion of a stop codon into the full-length sequence (157). The expression of this N-terminal segment did not result in a soluble protein despite containing the soluble fusion tag, which may indicate that it does not fold independently. Since in this way the N-terminal segment could not be analyzed on its own we decided to modify key residues in the helix-turn-helix motif and examine the effect of the mutations within the full-length context.

The mutations within the HTH were planned based on experimentally determined 3D structures of DNA-bound bacteriophage repressors, which contain HTH motif. We found seven
such proteins in the PDB (213–218), the best resolution structures of those, with the PDB IDs: 1LMB; 6CRO, 2OR1, 3CRO, 2R1J, 3ZHM, 3QWS were examined in detail. In addition to the structural similarity of these phage repressor HTHs, high level of sequence similarity was observed in case of five of those (Figure 14) (165). It has been revealed by these structures and various additional studies that the first helix stabilizes the protein-DNA interaction, while the second helix is responsible for specific recognition of the cognate DNA binding motif of the repressor (194–196).

In most of the cases, with the exception of Cro repressor, the first two polar residues of the second helix make direct H-bond interaction with DNA nucleobases, which suggests their essential role in binding and recognition (Figure 14) (157). If the structural model of the Stl HTH is superimposed with these repressor structures two similar polar residues, Q40 and N41, can be found at the same position. In addition to this sequence alignment also indicates the potential importance of these two residues in DNA binding of Stl (Figure 14). We created a Q40A, N41A double mutant Stl protein (Stl-AA) to eliminate the H-bonding ability at this positions, while preserving the helical secondary structure of the motif. Similar double alanine mutation of the respective residues in case of the TP901 phage CI repressor resulted in the loss of in vitro and in vivo function of the protein (218).

**Figure 14. DNA binding domain of bacteriophage repressors** (157). A) Sequence alignment of the HTH motifs of bacteriophage repressors and Stl. The number before each segment is the position of the first residue of the segment in the respective protein sequence. Helices are with grey background, similar residues are in bold, box highlights residues interacting with DNA nucleobases.

B) Experimentally determined structure of the DNA-bound CI bacteriophage repressor (PDB ID: 1LMB). DNA cartoons orange, protein cartoon: dark blue for HTH, otherwise cyan. DNA bases and DNA interacting amino acid residues are stick representation with atomic coloring (protein carbon yellow, DNA carbon green, oxygen red, nitrogen blue, phosphorus orange), H-bond interactions are represented by black dashed lines.

To test our hypothesis about domain functions, the DNA binding ability of Stl-CTD was examined with electrophoretic mobility shift assay. Samples containing Stl-CTD in various
concentrations (increasing from 2 μM up to 30 μM) and 100 ng of the double-stranded DNA oligo involving the cognate binding-site of Stl \((str-xis, (176))\) were subjected to native polyacrylamide gel electrophoresis (Figure 15).

The positive control containing 2 μM Stl clearly showed the expected shift of the DNA band compared to the position of the free DNA band due to complex formation between Stl and DNA (cf. also respective Figures in Ref. 17 and 176), while no band shift was observed in any of the samples containing Stl-CTD (Figure 15, Panel A). These data revealed that the C-terminal segment of Stl lacks the potential for DNA-binding, possibly because of the removal of the HTH segment. According to our expectation, the double alanine mutant \((Q40A, N41A)\) Stl construct \((Stl-AA)\) was found to be highly defective in DNA binding based on EMSA experiments, compared to the wild type protein (Figure 15, Panels B and C, cf. also Figure A3 in the Appendix) \((157)\). These \textit{in vitro} results about the DNA-binding ability of Stl-CTD and Stl-AA have later been verified also \textit{in vivo} by our Stl-based reporter system in \textit{Mycobacterium smegmatis} \((219)\).

![Figure 15. Testing the DNA binding ability of Stl-CTD and Stl-AA (157). A) Electrophoretic mobility shift assay of samples consisting of the dsDNA oligo containing the cognate Stl binding site mixed with Stl-CTD or Stl was performed to investigate the DNA-binding ability of the proteins. Species and concentrations of monomers are indicated on the figure. The band of the dsDNA was only shifted upwards if the wild type Stl was added, but there was no observable shift even at high concentrations Stl-CTD. B-C) Electrophoretic mobility shift assay was performed to compare the DNA binding ability of Stl (Panel B) and Stl-AA (Panel C). Species and concentrations of monomers are indicated on the figure. Wild type Stl caused shift of the dsDNA oligo at 1 μM concentration. The band of dsDNA oligo was only partially shifted upwards even at relatively high concentrations of Stl-AA (3 μM) (cf. also Figure A3 in the Appendix).](image)

The dimerization ability of Stl-CTD was checked by chemical crosslinking. Samples of Stl and Stl-CTD were incubated with excess covalent crosslinking agent \((\text{disuccinimidyl suberate, DSS})\) and the resulting samples were analyzed on denaturing PAGE (Figure 16). Bands showed up at the positions corresponding the Stl dimer and Stl-CTD dimer, confirming the dimerization ability of the C-terminal construct. This supported our hypothesis that the carboxy terminal segment of the Stl protein is responsible for dimerization, which was also
reinforced by the findings that the Q150STOP and K176STOP mutants of Stl were not functional in repression perhaps due to their impaired dimerization (18, 220).

<table>
<thead>
<tr>
<th>Protein (40 μM)</th>
<th>Stl</th>
<th>Stl-CTD</th>
</tr>
</thead>
<tbody>
<tr>
<td>DSS (mM)</td>
<td>0.2</td>
<td>0.1</td>
</tr>
</tbody>
</table>

Figure 16. Chemical crosslinking of full-length Stl and Stl-CTD (residues 85–267) (158). Upon addition of the crosslinking agent, DSS to Stl (32.9 kDa) and Stl-CTD (22.4 kDa) bands corresponding to the size of Stl and Stl-CTD dimers appeared. Species and concentrations given in monomers are indicated on the figure.

To test if Stl-CTD is involved in the Stl-Φ11DUT complex formation, we performed native polyacrylamide gel electrophoresis experiment with the mixture of the two proteins (Figure 17, Panel A). As the concentration of Stl-CTD was increased in the samples the band at the position of the free Φ11DUT gradually disappeared, while the band corresponding to the complex of two proteins showed up at a distinct position as compared to that of the individual proteins. These results are similar to that obtained with the full-length Stl protein (Figure 20 and Figure 2B in Ref 17), although the dUTPase band vanishes at 1:1 Stl to Φ11DUT ratio in case of the wild type Stl protein, while it is still present even in case of relatively high (3:1) Stl-CTD to Φ11DUT ratio. We also tested if this interaction between Stl-CTD and Φ11DUT perturbs the enzymatic activity of the latter (Figure 17, Panel C). The maximum inhibitory effect of Stl-CTD on dUTPase was found to be around 40% even at saturating inhibitor concentration, while the full-length Stl protein caused practically complete loss of dUTPase activity if added in high excess (17). The lower extent of maximal dUTPase inhibition exerted by Stl-CTD as compared to that of the wild type protein suggests that the amino-terminal segment of Stl may also contribute somehow to the interaction between the two proteins. Nevertheless, the apparent inhibitory constant of K<sub>i</sub> = 1.5 ± 0.5 nM obtained for Stl-CTD was comparable with that measured for Stl (K<sub>i</sub> = 1.2 ± 0.8 nM) (117).

The interaction of Stl-AA with dUTPase has been investigated, in order to check if the double mutation within the helix-turn-helix influences the dUTPase binding ability of Stl
(Figure 17, Panels B and D). The pattern of complex formation of Stl-AA with Φ11DUT on the native gel was very much alike to that of the wild type Stl protein (Figure 20 and Figure 2B in Ref 17). The inhibitory effect of Stl-AA on Φ11 phage dUTPase was also not perturbed by the mutations, as characteristics of inhibition by Stl-AA (K_i = 1.4 ± 0.9 nM and more than 90% maximal inhibition) were basically the same as that of Stl (17, 117).

**Figure 17. Testing the Φ11 phage dUTPase binding and inhibition of Stl-CTD and Stl-AA (157).**

**A-B)** Native gel electrophoresis experiments were performed to investigate the dUTPase binding ability of Stl-CTD and Stl-AA, respectively. Species and concentrations given in monomers are indicated. A band at distinct position comparing to the individual proteins shows up in case of the samples containing Φ11DUT and Stl-CTD, which clearly indicates the complex formation. The interaction of Stl-AA with Φ11 phage dUTPase is similar to that of the wild type Stl (Figure 20).

**C-D)** Activity of the Φ11 phage dUTPase was measured in the presence and absence of Stl-CTD and Stl-AA, respectively. Each measurement was repeated three times. The quadratic binding equation was fit to the data resulted in the apparent K_i of 1.5 ± 0.5 nM for Stl-CTD and 1.4 ± 0.9 nM for Stl-AA. The total change in amplitude of the activity was 40% in case of Stl-CTD, which is lower than that of the full-length Stl protein. The maximal inhibitory effect of Stl-AA was above 90%.

Comparing to the wild type Stl the inhibition ability of Stl-AA on Φ11 phage dUTPase is not perturbed (Figure 20).

In summary, these data show that the N-terminal segment of Stl governs the interaction of the repressor protein with its cognate DNA site, since the C-terminal segment of Stl lacks the potential for DNA-binding (157). Point mutations within the predicted helix-turn-helix motif drastically decrease DNA-binding ability of Stl, which supports the validity of our hypothesis about the structure and position of this motif. Although the C-terminal segment is
capable of folding on its own and is able to dimerize in solution, it still does not possess full functional capability for dUTPase binding and inhibition.

4.1.3 Interaction surface of Stl with Φ11 phage dUTPase

We have studied the interaction surface of Stl and Φ11DUT by hydrogen-deuterium exchange mass spectrometry (HDX-MS) (175), a pioneering technique that reports on changes in the local deuterium uptake of a protein in D₂O solvent (221). The isotope uptake rates are determined by the environment of exchangeable sites, comparison of the HDX-MS patterns recorded for the individual proteins and the complex can provide information about the interaction surface of a protein with its ligand or with other proteins (222–225). To gain regional information after quenching the exchange the protein is subjected to proteolytic digestion, and mass of the resulting peptides is analyzed by HPLC-MS workflow. The region in which decrease is observed is proposed to be directly involved in the protein-protein interaction (226, 227).

In case of Stl a difference plot containing data for 84 individual peptides has been obtained, covering 94% of the protein sequence (excluding tags) with redundancy of 3.8 (Figure 18 and Figure A4 in the Appendix). Based on our HDX-MS results we could identify regions of Stl which are responsible for complex formation. In the presence of DUTΦ11, Stl exhibits significant negative mass shifts across the entire length of the protein. This indicates that Stl undergoes a global decrease in dynamics upon binding to Φ11DUT which also implies the inherent flexibility of Stl protein.

**Figure 18. Experimental results and schematic models showing how Stl interacts with the Φ11 phage dUTPase (175).** A) HDX difference plots the mass change (Δmass) of each individual peptide of Stl upon mixing with Φ11DUT. The dashed line represents the 95% confidence bands evaluated over the whole dataset. B) cartoon and surface representation of the in silico 3D homology model of Stl colored according to HDX-MS following the color gradient on the bottom of the figure.
In addition to these overall changes, a specific tyrosine-rich region of the Stl protein (segment 98Y–113Y) exhibited a dramatically pronounced negative mass shift upon addition of dUTPase (Figure 18). This suggests that although the interaction is communicated as a global conformational tightening across the entire protein, binding is rather localized to a specific region. These findings are consistent with our previous results which showed dUTPase binding of Stl-CTD (residues 85–263) and Stl-AA. Interestingly we have not detected any significant positive HDX signal across the Stl sequence, which would be expected upon the dissociation of the Stl dimer. All in all, these HDX-MS data taken together with our previous results on Φ11DUT-Stl interaction argue for overlap between the dimerization surface of Stl and the Stl-dUTPase interface.

4.1.4 Molecular shape and structure of Stl in solution

Small angle X-ray scattering (SAXS) experiments were performed to assess the molecular shape and determine the solution structure of Stl at low-resolution (158). Molecular mass estimations for Stl based on SAXS data suggested that the protein predominantly forms dimers in solution (Table A5 in the Appendix), which is consistent with the dimer formation observed by native mass spectrometry and chemical crosslinking (cf. Figure 17) (17, 158). The ab initio shape envelope of the Stl dimer generated based on SAXS spectra is shown in Figure 19, Panel A. This ab initio model has an effective resolution of 4.9 ± 0.4 nm and average NSD of 1.5 ± 0.4. (Ideal superposition of two compact and rigid structures with the same low-resolution shapes gives NSD around 1 (188).)

The observed relatively low resolution of the ab initio models for the Stl dimer may indicate the internal flexibility of the protein. Therefore the flexibility of the Stl dimer was examined using an ensemble-based approach (186). In the first step, random configurations of the N-terminal region (M1-K83) were generated, while the C-terminal region (T87-N267) of the Stl model constituted a fixed dimer interface, based on our chemical crosslinking results (cf. Figure 16). The results of this procedure are represented in size-distributions calculated from the pool from which a genetic algorithm selected an ensemble of best-fitting configurations. The highly flexible domain arrangement of Stl protein is verified by the resulting similarly broad distributions of the random and the selected models (Figure 19, Panel B). This domain flexibility might correspond with the multiple functions of Stl and the dynamic change between DNA and dUTPase binding.
Figure 19. Models generated based on SAXS measurements for Stl (158). A) Overlay of the dimeric Stl model from the Phyre2 server and the *ab initio* shape envelope reconstructed from SAXS data (SASBDB ID.: SASDC67). Stl-NTD denotes the amino terminal segment while Stl-CTD designates the carboxy terminal segment of Stl according to Figure 13. B) Ensemble analysis of Stl SAXS data with EOM (186). R<sub>g</sub> size-distributions show the high level of flexibility of the selected best-fitting ensemble compared to that of the random pool. The presence of inter-domain flexibility was also reflected by the flexibility metric (R<sub>flex</sub> was found to be 88% for both the selected ensemble and the initial pool).

Rigid body models of Stl dimer were generated by M-ZDOCK server (190) applying the 3D structural model created by Phyre2 server (189) and defining the C-terminal region (T100–N280) of Stl as the dimer interface based on our previous results (cf. Chapters 4.1.2 and 4.1.3). The resulting structures were screened against the experimental SAXS data, the best candidate provided an excellent fit with the SAXS spectra (χ<sup>2</sup> = 1.0 between the model and experimental data, cf. Figure 29), which indicates that this is a valid model of the average structure in solution (Figure 19, Panel A).
4.2 Role of the phage specific loop of Φ11 phage dUTPase

Staphylococcal phage dUTPases possess a ca. 30 to 40-residue-long phage-specific segment inserted between the third and fourth conserved motif (cf. Chapter 1.4) (16). It has been shown that this segment has no role in the enzymatic activity of the protein (84, 119). As some phage dUTPases exhibited a moonlighting function, namely the mobilization of pathogenicity island through binding to the master repressor Stl protein (16), it was of our interest to explore the role of the phage specific insert in the dUTPase-Stl interaction and in the disruption of Stl-DNA complex (117).

4.2.1 Role of the phage specific loop of Φ11 phage dUTPase in Stl binding

We have quantitatively characterized the inhibitory effect of Stl on the wild type Φ11DUT enzyme (17), and proved that the truncated Φ11DUT<sup>Δ101G–122Q</sup> lacking the phage-specific insert exhibits the same activity as the full-length dUTPase (84). To test the role of the insert in Stl binding first we investigated the interaction of Stl with Φ11DUT<sup>Δ101G–122Q</sup> by enzyme activity measurements. We observed very similar inhibition characteristics for the two dUTPases, indicating that Stl is fully capable of binding to and inhibiting insert-less Φ11DUT mutant (Figure 20, Panel A). Both dUTPases were almost completely inhibited upon the addition of approximately equimolar amount of Stl, the apparent inhibitory constants did not differ in the case of the truncated mutant and full-length dUTPase (K<sub>i</sub> was found to be 1.2 ± 0.8 nM for Φ11DUT, while K<sub>i</sub> = 1.5 ± 0.6 nM was obtained for Φ11DUT<sup>Δ101G–122Q</sup>) (117).

Then the complex formation of the wild type and the truncated Φ11 phage dUTPase proteins with Stl were tested by native polyacrylamide gel electrophoresis. The resulting pattern of complexes was very similar for Φ11DUT<sup>Δ101G–122Q</sup> and Φ11DUT (Figure 20, Panel B) (117). A band at a distinct upper position as compared to Stl and dUTPase bands appear in case of samples of 1:1 dUTPase:Stl monomer ratio, corresponding to the formed complex (Complex A). The appearance of the excess dUTPase and a band of a second complex type (Complex B) was observed upon decreasing the amount of Stl in the protein mixtures. Despite the position of Complex B band is close to that of Stl, it certainly does not correspond to the shift of the free Stl band in protein mixtures since Φ11DUT was in excess in those samples, in which the band of Complex B is highly abundant (cf. also Figure 1B in Ref 17). The two complexes obviously represent different stoichiometries, based on the applied concentrations the Stl: Φ11DUT ratio is higher in Complex A than in Complex B.
Our previous native mass-spectrometry experiments revealed the existence of a complex with the molar mass of 125.500 (±0.080) kDa (17), which results from the interaction a Φ11DUT trimer of 55.1 kDa with an Stl dimer or with two Stl monomers of 65.8 kDa (Φ11DUT3Stl2) (Figure 20, Panel C). Peak series referring for the low amount of a species with the molar mass of 158.370 (±0.080) kDa has also been observed, which suggests the formation of a Φ11DUT3Stl3 complex (Figure 20, Panel C, blue peaks). The existence of Φ11DUT3Stl3 complex could explain the observed nearly complete inhibition of dUTPase enzyme activity by Stl. The N= 1.19 ± 0.12 DUT:Stl stoichiometry of the complex indicated by ITC also argues for the presence of the mixture of these two complexes in solution (17). Considering the composition of the samples and the relative intensity of the bands of the complexes and the free proteins on the native gel, we propose that Complex A reflects the Φ11DUT3Stl3 stoichiometry, and consequently, Complex B has the stoichiometry of Φ11DUT3Stl2 (Figure 20, Panel B).

Figure 20. Interaction of Stl with Φ11DUT and Φ11DUTΔ101G–122Q (17, 117). A) The enzymatic activity of 50 nM Φ11DUTΔ101G–122Q and Φ11DUT was measured in the presence and absence of Stl. Each measurement was repeated three times and quadratic binding equation was fit to the data. The table at the bottom indicates the characteristic data of the inhibition curves. B) Native gel electrophoresis experiment was performed to investigate the Stl binding ability of Φ11DUTΔ101G–122Q comparing to Φ11DUT. Species and concentrations given in monomers are indicated. Bands at distinct positions compared to that of the individual proteins in samples containing both Stl and dUTPases refer to the presence of two type of complexes, with different stoichiometry. Φ11DUTΔ101G–122Q shows very similar complexation pattern to Φ11DUT. C) Mass spectrum of Φ11DUT-Stl complex, peak series of 4826 (26+), 5019 (25+), 5228 (24+), 5456 (23+), 5704 (22+), 5976 (21+), 6276 (20+), 6609 (19+), 6975 (18+) referring for species with the molar mass of 125500 (±80) Da, which corresponds to a Φ11DUT3Stl2 complex. Although with very weak intensity peak series of 5280 (30+), 5657 (28+), 5867 (27+), 6093 (26+) is also present (highlighted with blue) reflecting the existence of a species with the molar mass of 158370 (±80) Da, arguing for the formation of a Φ11DUT3Stl3 complex.
The distribution of the two types of Stl-DUT complexes slightly differs between \( \Phi 11\text{DUT} \) and \( \Phi 11\text{DUT}^{\Delta 101G-122Q} \), which may reflect some difference between the Stl-binding affinity of the proteins, which is also suggested by BLI measurements \((K_D(\Phi 11\text{DUT}) = 1.84 \text{ nM}, K_D(\Phi 11\text{DUT}^{\Delta 101G-122Q}) = 11.4 \text{ nM})\)\(^4\) (119).

To gain more information about the interaction between Stl and \( \Phi 11 \) phage dUTPase we carried out hydrogen-deuterium exchange mass spectrometry experiments with complexes of \( \Phi 11\text{DUT} \) and \( \Phi 11\text{DUT}^{\Delta 101G-122Q} \) with Stl. According to the difference plots of Stl the two dUTPases interact with the same part of the protein (Figure 21).

![Figure 21. Difference plots of Stl upon complex formation with \( \Phi 11\text{DUT} \) and \( \Phi 11\text{DUT}^{\Delta 101G-122Q} \) (175). HDX difference plots the mass change (\( \Delta \text{mass} \)) of each individual peptide of Stl upon mixing with \( \Phi 11\text{DUT} \) (red) and \( \Phi 11\text{DUT}^{\Delta 101G-122Q} \) (blue). The dashed line represents the 95\% confidence bands evaluated over the whole dataset.](image)

The active site is involved in the Stl-binding in case of both dUTPases, although the active site of \( \Phi 11\text{DUT} \) is less well-hidden by Stl than that of \( \Phi 11\text{DUT}^{\Delta 101G-122Q} \) (Figure 22, Appendix Figures A5 and A6). It seems that the insert somewhat hinders the accessibility of the active site of \( \Phi 11\text{DUT} \), while also provides new sites for Stl-binding.

\(^4\) These values should be considered as approximations since in case of nanomolar \( K_D\)-s measuring the binding affinity by BLI with Ni-NTA biosensors is biased by the simultaneous dissociation of the immobilized His-tagged sensors (His-DUT) and the target (Stl). This is especially true if the dissociation is slow as in the case of Stl-DUT interactions. To obtain more reliable data more effective immobilization of the biosensor is needed (e.g. biotinylated protein and streptavidin sensors) while using the same bait (i.e. Stl) would also make a comparison of the values more straightforward.
Figure 22. HDX-MS difference data obtained for Φ11DUT and Φ11DUT$^{Δ101G-122Q}$ upon complex formation with Stl (175). A) Representation of the HDX-MS difference data on the surface of the Φ11DUT and Φ11DUT$^{Δ101G-122Q}$ (175). In case of the Φ11DUT the experimentally determined structure is shown (PDB ID: 4GV8), note that the position of the C-terminal 15 residues was not resolved in the crystal structure possibly due to flexibility. A three-dimensional structural model generated by SWISS-MODEL is displayed for the truncated mutant. The substrate analogue is shown as red sticks in order to ease visualization of the interference of Stl and substrate binding. The coloring is according to the scale at the bottom of the panel. (Views: side, top, bottom.) B) Difference plots of Φ11DUT and Φ11DUT$^{Δ101G-122Q}$ upon complex formation with Stl (175). HDX difference plots the mass change (Δmass) of each individual peptide of dUTPases upon mixing with Stl. The dashed line represents the 95% confidence bands evaluated over the whole dataset. Numbering of residues in Φ11DUT$^{Δ101G-122Q}$ follow that of Φ11DUT, yellow background highlights the deleted segment.

Based on these experimental data we propose that the phage specific insert of Φ11DUT is not essential for Stl-binding. This conclusion is further supported by our results obtained for the interaction between Stl and *Mycobacterium tuberculosis* dUTPase (MtDUT), which obviously does not contain the phage-specific insert (159) (cf. Figures 11 and 23). It was found
that Stl is able to reduce the activity of MtDUT by $81 \pm 10\%$, and the interaction can be described with a very low apparent inhibitory constant, $K_i = 6.8 \pm 4.4$ nM (159).

![Comparison of MtDUT and Φ11DUT structures](image)

**Figure 23. Comparison of MtDUT and Φ11DUT structures** (117). A) X-ray crystal structure of the *Mycobacterium tuberculosis* dUTPase (MtDUT, PDB ID: 2PY4), protein trimer is shown in cartoon representation, substrate analogues (dUPNPP) as sticks. The *Mycobacteria* specific insert colored dark blue other residues orange, ligands green. B) X-ray crystal structure of the Φ11 bacteriophage dUTPase (Φ11DUT, PDB ID: 4GV8), protein trimer is shown in cartoon representation, substrate analogues (dUPNPP) as sticks. The phage-specific insert is colored dark blue other residues cyan, ligands red. C) Superimposition of the MtDUT and Φ11DUT structures. Coloring and representation are according to Panels A and B, ligands are not shown. The phage-specific insert did not affect the core dUTPase fold.

Native mass spectrometry also confirmed the formation of MtDUT$_3$Stl$_2$ complex although the 3:3 complex was not observed, which can explain the ca. 20% residual activity of this dUTPase upon Stl binding (Figure 24, Panel A). Concordantly the complex formation between Stl and MtDUT was observed in native gel electrophoresis experiments, however the existence of two complex forms is also not evident from the resulting pattern (Figure 24, Panel B) (159). The inhibition of mycobacterial dUTPase activity by Stl has also been demonstrated *in vivo* in *Mycobacterium smegmatis* (159).
Figure 24. Complex formation of Stl with MtDUT. A) Native electrospray mass spectrum of the mixture of Stl and MtDUT. Peak series of 4569 (25+), 4761 (24+), 4968 (23+), 5194 (22+), 5442 (21+), 5713 (20+), 6017 (19+), 6364 (18+) argues for the existence of a complex with molar weight of 114.200 (±0.080) kDa, which corresponds to a complex formed by a trimer MtDUT of 54.0 kDa and an Stl dimer or two Stl monomers of 65.8 kDa (MtDUT3Stl2). B) Complex formation between MtDUT and Stl is apparent from native gel electrophoresis experiment, although the presence of two types of complexes is not evident. Species and concentrations given in monomers are indicated (159).

Our group has also shown that Stl binds to and inhibits the human (cf. Chapter 4.3) and the Drosophila melanogaster dUTPase (228), while our preliminary results suggest that Stl also binds to but does not inhibit dUTPases from E. coli and Plasmodium falciparum. Based on these results Stl shows high affinity for dUTPase binding in general, independent from the species specific segments.

4.2.2 Role of the phage specific loop of Φ11 phage dUTPase in perturbation of Stl-DNA interaction

Having established that the phage specific insert of Φ11DUT is dispensable for Stl-binding, we tested if this insert may have any role in perturbation of the Stl function. The regulatory role of Stl on the replication and excision of the SaPI mobile genetic elements is exerted through binding of the repressor to specific DNA segments and herewith preventing the expression of proteins required for SaPI mobilization (16, 176). It has been demonstrated by in vitro electrophoretic mobility shift assay experiments that Φ11DUT perturbs the DNA-binding of the Stl repressor (16, 17). It has also been shown that the phage dUTPase-Stl interaction results in derepression and SaPI replication in Staphylococcus aureus (16). Hence it was proven that the results from the in vitro experiments reflect well the in vivo derepression activity of Φ11DUT. Therefore, we tested the role of the insert in disturbing the DNA binding ability of Stl by EMSA experiments (Figure 25, Panel A). When Φ11DUT101G–122Q was added to the DNA-Stl mixture the DNA-band showed up at the position of the Stl-DNA complex, while it appeared at the same position as the free DNA in case of the positive control sample containing Φ11DUT, Stl and DNA in the same concentration. This result indicates that the insert is
necessary for the disruption of the Stl-DNA complex and so for the derepression activity of Φ11 phage dUTPase. Later the impaired SaPI induction ability of the Φ11DUT<sup>Δ101G-122Q</sup> protein has also been verified by in vivo experiments (119).

**Figure 25. Perturbation of Stl-DNA complex by Φ11DUT, Φ11DUT<sup>Δ101G-122Q</sup>, MtDUT and hDUT (117, 158, 159).** Electrophoretic mobility shift assay experiments were performed to investigate the ability of different dUTPases to disrupt Stl-DNA complex. Species and concentrations of monomers (in μM) are indicated on the figure. A) Φ11DUT<sup>Δ101G-122Q</sup> did not perturb the DNA-binding of Stl even at elevated concentration. B) Unless MtDUT lacks the phage specific insert it is able to disrupt Stl-DNA complex. C) DNA-binding ability of Stl is perturbed during complex formation with hDUT although this dUTPase was less effective in perturbation of Stl-DNA complex, than Φ11DUT and MtDUT. (EMSA experiment displayed on this panel was performed by Dr. Borbála Tihanyi.)

These data highlight the role of the insert in the derepression ability of Φ11DUT. On the other hand, derepression ability of both Φ11 and 80α phage dUTPases is highly sensitive to point mutations around the active site (119). In addition to this, a few, seemingly benign differences in the core region between the SaPI inducing Φ11 phage and Staphylococcus virus 71 dUTPases compared to that of ΦB2 phage, although leaving both the insert and the conserved motifs intact, lead to the loss of SaPI mobilization ability of ΦB2 phage dUTPase (18).

All in all, in the context of Φ11 phage dUTPase the insert seems a necessary but not sufficient condition for SaPI derepression. Although, it has been shown that under strong selection of 80α phages against SaPI<sub>bovi</sub> interference only mutations acquired in the dUTPase core region lead to resistance, so none of the effective hits included mutation in the insert (18). The fact that Stl does not target these potentially inactive dUTPases might also imply that the SaPI prefers phages possessing full-dUTPase activity to avoid uracil incorporation in their genome (cf. Chapter 1.2). This strategy could also provide the SaPIs an advantage by hindering the escape of phages because of the reduced mutation rate of those due to dUTPase activity.

As Stl seems to have universal binding capacity to trimeric dUTPases one may hypothesize that phages have two strategies to escape SaPI interference: i) evolve an insert between Motif 3 and 4, without the loss of dUTPase activity; ii) if the insert fails to serve this
purpose, mutations in the enzyme core can also be effective although with a potential fitness cost of reducing/losing the dUTPase.

In order to provide a more general view about the effect of dUTPase-Stl interaction on the DNA-Stl complex we tested the ability of Mycobacterium tuberculosis and human dUTPases, which obviously lack the phage specific insert (cf. Figures 10, 11, 23), to perturb the binding of Stl to DNA in vitro (Figure 25, Panels B and C). The electrophoretic mobility shift assay experiments revealed that MtDUT has similar capability to disrupt the Stl-DNA complex as the Φ11DUT, while hDUT is also able to disturb DNA-binding of Stl although less effectively than the former two dUTPases. In addition to this, our preliminary results suggest that Plasmodium falciparum dUTPase also has reduced, but still considerable ability to displace Stl from the protein-DNA complex. These results suggest that the insert is only necessary for perturbing Stl-DNA interaction in case of phage dUTPases. Our group revealed that Drosophila melanogaster dUTPase-Stl interaction does not affect Stl-DNA complexation (228), and according to our preliminary data E. coli dUTPase although binds to Stl also lacks the potential to block Stl-DNA interaction (cf. Figures 10 and 11). In conclusion, the available data suggest no obvious relationship between the ability of a specific dUTPase to disrupt the Stl-DNA complex and the inhibition of the enzyme by Stl (Table 6).

Table 6. Summary of Stl-dUTPase interactions.

<table>
<thead>
<tr>
<th></th>
<th>complex formation with Stl</th>
<th>inhibited by Stl</th>
<th>disruption of Stl-DNA complex</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Φ11DUT</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>(16, 17, 117)</td>
</tr>
<tr>
<td>Φ11DUTΔ101G–122Q</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>(117, 119)</td>
</tr>
<tr>
<td>80αDUT</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>(16)</td>
</tr>
<tr>
<td>MtDUT</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>(159)</td>
</tr>
<tr>
<td>hDUT*</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>(158)</td>
</tr>
<tr>
<td>D. mel dUTPase</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>(228)</td>
</tr>
<tr>
<td>PfDUT</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>#</td>
</tr>
<tr>
<td>E. coli dUTPase</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>#</td>
</tr>
</tbody>
</table>

1 José R. Penadés personal communication  
* cf. Chapter 4.3.  
# preliminary results

To demonstrate the complexity of this system, the potential complexation equilibria in the three-component mixture including the major steps are shown in Figure 26. We assume that

---

5 Although it has been shown that dUTPase is not essential for phage survival the effect of losing this enzymatic function on the viability of the phage has not been followed for multiple generations of phage progeny (16).
Stl, like other well-known prophage repressors, binds DNA as a dimer. Higher order complexes between Stl and DNA may also probably exist (12), which would provide the possibility of a more sophisticated regulatory scheme for the gene expression, however, this question has not yet been investigated. The existence of Stl dimer in solution was verified by chemical crosslinking, native mass-spectrometry and SAXS experiments (cf. Chapter 4.1.2 and 4.1.4) (17, 158). As there has been no indication for dissociation of Φ11DUT upon complexation with Stl (17), the monomer-trimer equilibrium was excluded from this model. This is also supported by the fact that trimeric dUTPases are extremely stable in this oligomer form in solution (176).

We found experimental indication for the existence of Φ11DUT/Stl₂ (Complex B) and Φ11DUT/Stl₃ (Complex A) complexes, although whether these include an Stl dimer bound to the dUTPase trimer or just monomers of Stl interact with Φ11DUT remains to be established. The complexity of this simplest model might justify the difficulties in finding a direct relationship between dUTPase-Stl interaction and perturbation of DNA-binding of Stl by that, which may be described with more even complex or non-equilibrium models.

![Figure 26. Schematic representation of the possible complex equilibria in the mixture of Stl (blue spheres), dUTPases (red triangles) and DNA (purple helices) (117). The dimerization of Stl was proven by our chemical crosslinking, electrospray ionization native mass-spectrometry (ESI-MS) and SAXS measurements (17, 158). We assume that, as the prophage regulator repressors in general, Stl also binds DNA as a dimer. Formation of Φ11DUT/Stl₂ (Complex A) and Φ11DUT/Stl₃ (Complex B) was suggested based on native ESI-MS and native gel electrophoresis experiments (17). It remains to be explored if Stl could bind dUTPase as a dimer or only as a monomer, so we included two possible arrangement of Stl in case of both complexes. We applied this basic model; although some more complex patterns may possibly exist within this three-component system.](image)
than a truncated mutant of Φ11 lacking conserved Motif 5 (K_D = 1.42 nM), despite the latter is ineffective in SaPI induction \textit{in vivo} (although it shows perturbation of Stl-DNA interaction \textit{in vitro}), while 80αDUT acts as a SaPI derepressor. When the dUTPase-Stl interaction and the derepression are investigated within the Staphylococcus cell one cannot exclude the interference of other bacterial or phage proteins that may perturb these processes directly or indirectly. To avoid these background effects, we created an Stl-based reporter system in \textit{Mycobacterium smegmatis}, which offers us the opportunity for high-throughput studies of Stl-DNA and Stl-dUTPase interactions (219). Although thorough \textit{in vitro} characterization of the resulting hits provided by the studies applying the latter system is also required, since the slight interference of mycobacterial dUTPase is overcome by overexpression of the proteins. Hopefully, this approach will reveal novel details about the connection between the Stl-binding and derepressor abilities of different dUTPases.
4.3 Interaction of Stl with the human dUTPase

It has been demonstrated that the staphylococcal repressor protein Stl binds to and inhibits the trimeric dUTPases from *M. tuberculosis* and *D. melanogaster in vitro*, even though these two proteins share only limited sequence similarity with the original phage dUTPase partners of Stl (17, 159, 228). Previous studies have also provided evidence that Stl is an effective inhibitor of the mycobacterial dUTPase *in vivo* (159). As the cellular activity of the human dUTPase has a significant effect on the success of the thymidylate synthase targeting chemotherapeutic agents in various tumor types and cancer cell lines, extensive efforts have been made to study the dUTPase inhibition *in cellulo* (28, 57–62). We set out to investigate if Stl could support these research projects as a tool, like the uracil-DNA glycosylase inhibitor (UGI) from bacteriophage PBS2 has been used to investigate inhibition of human UNG (154, 229).

First, we tested if Stl could bind to the human dUTPase by native polyacrylamide gel electrophoresis of samples constituting individual proteins and mixtures of the two proteins (Figure 27, Panel A). Due to the higher isoelectric point of hDUT (pI= 7.1) as compared to that of Stl (pI= 6.3), the band of dUTPase (white arrowhead) was found at an upper position relative to that of Stl (black arrowhead), which is a marked difference to the Φ11DUT-Stl and MtDUT-Stl native gels (cf. Figures 20 and 24). As the concentration of hDUT is increased in the samples containing fixed amount of Stl, a band at a distinct position became apparent (grey arrowhead), while the Stl band gradually disappeared (Figure 27).

**Figure 27. Identification and functional analysis of the interaction between Stl and hDUT (158).**

A-B) Formation of a hDUT-Stl complex is represented by the appearance of unique band (grey arrowhead) on native gel electrophoresis in samples containing the mixture of Stl (black arrowhead) and hDUT (white arrowhead). Panel B shows how complex formation is affected by the substrate analogue dUPNPP depending on the order of mixing. C) dUTPase enzymatic activity was measured in the presence and absence of Stl (representative measurement out of the three paralels). The apparent Kᵦ= 6.7 ± 2.4 nM was derived from the quadratic equation line fitted to the data. (Enzyme activity measurements shown here were performed by Dr. Borbála Tihanyi.)
The effect of the non-hydrolyzable substrate analogue dUPNPP on hDUT-Stl interaction was also tested and no dissociation of the pre-formed hDUT-Stl complex was observed upon addition of dUPNPP in 60-fold excess (Figure 27, Panel B). Nevertheless, if the same amount of this substrate analogue was pre-mixed with hDUT the complex formation of hDUT with Stl was not detectable. This result suggested that there is an interference between substrate and Stl-binding of hDUT, which was verified by enzyme activity assays revealing the potential of Stl to reduce the catalytic activity of human dUTPase by 70% (Figure 27, Panel C). Measuring the enzymatic activity at different concentrations of Stl provided an apparent inhibitory constant of 6.7 ± 2.4 nM, which indicates that Stl is a highly potent inhibitor of the human dUTPase. Comparing the data obtained for the inhibitory effect of Stl on hDUT with our former results acquired for MtDUT (Kᵢ = 6.8 ± 4.4 nM, residual activity of ca. 20% (159)) we can conclude that characteristics of inhibition are very similar in case of these two dUTPases. Based on this, we propose that Stl might have the potential for significantly reduce the cellular activity of the human dUTPase, as it was able to perturb dUTPase function in Mycobacterium smegmatis (159). Our former kinetic analysis revealed that Stl and dUTP compete for the active site of Φ11DUT, and the slow and tight binding character of the complex formation between Stl and the phage dUTPase (17). Similarly to our observations with Φ11DUT, in case of the human dUTPase inhibition was only detectable when the enzyme was pre-incubated with Stl before the addition of dUTP. This strong dependence of the inhibitory and binding ability of Stl to hDUT on the order of mixing of components in case of the kinetic studies and the native gel electrophoresis might indicate that Stl is a slow and tight binding competitive inhibitor of the human dUTPase.

Our isothermal titration calorimetry (ITC) experiments also verified that Stl and hDUT form a considerably strong complex (Figure 28, Panel A). The observed apparent affinity (Kₒ = 0.23 ± 0.02 μM) was close to the Kₒ of 0.10 ± 0.03 μM determined for the Φ11DUT-Stl interaction with ITC (17). Similarly to that of Φ11DUT-Stl complex formation, the binding of Stl to hDUT is featured by a dominant contribution of binding enthalpy further enhanced by a favorable entropic component of complexation. The best-fitting, one set of independent sites binding model resulted in N = 1.27 ± 0.01 for stoichiometry, which formally represents that 1.27 hDUT protomer interact with 1 Stl protein molecule (ie. a trimer hDUT forms complex with to 2.4 Stl protomers). Considering the stability of the dUTPase trimer (30, 90, 230) and taking into account that there is a dynamic equilibrium between the monomer-dimer forms of Stl (17), the determined 3 to 2.4 hDUT-Stl protomer ratio could refer to the existence of i) hDUT₃Stl₂ macromolecular complex formed by a dUTPase trimer and 2 Stl molecules
(1 dimer/2 monomers) or ii) hDUT$_3$Stl$_3$ assembly consisting of a hDUT trimer and 3 Stl molecules (1 dimer + 1 monomer/3 monomers) or the mixture of these complexes (158). These observations are consonant with our previous results about the stoichiometry of Φ11DUT-Stl and MtDUT-Stl complexes (cf. Chapter 4.2.1).

Figure 28. Experiments aiming to explore the stoichiometry of hDUT-Stl complex (158). A) Thermodynamic characteristics of equilibrium binding of hDUT to Stl based on isothermal titration calorimetry (ITC) measurements. A representative graph is shown out of the three parallel experiments. The table at the bottom displays the average and standard deviation of parameters determined by fitting a one set of sites model to the integrated titration data (N, stoichiometry, $K_D$, binding affinity, $\Delta H$, binding enthalpy, $-T\Delta S$, binding entropy, $\Delta G$, binding free enthalpy). Parameters from the ITC experiment performed with Φ11DUT and Stl are shown alongside, note that errors, in this case, indicate the reliability of the fitting of one set of sites model on the data (17). Data argues for the existence of hDUT$_3$Stl$_2$ and/or hDUT$_3$Stl$_3$ complexes. B) Part of the mass spectra (4300-5100 m/z region) obtained for the mixture of Stl and human dUTPase under native electrospray conditions. Peak series of 4294 (28$^+$), 4455 (27$^+$), 4622 (26$^+$), 4808 (25$^+$), 5007 (24$^+$), corresponds to the heteromer with the molar mass 120.0 ± 0.1 kDa, which refers to the presence of the hDUT$_3$Stl$_2$ complex, in which a human dUTPase trimer (54.0 kDa) binds to two Stl molecules (65.8 kDa). Peak with m/z value of 4700 corresponds to Stl monomer with charge +7 (cf. Supplementary Figure 1 of Ref 157). C) Chromatogram of SEC-SAXS experiment performed with the hDUT-Stl mixture. The integrated SAXS intensity is plotted against frame number. The inset displays an enlarged view of the major peak and the average radius of gyration ($R_g$) derived from the data applying a sliding window of 10 frames is represented by a solid black line. Based on correlation map analysis (231) of the SAXS data across the peak two regions could be identified within the peak.
Native electrospray ionization mass spectrometry (ESI-MS) experiments were conducted to gain additional experimental information about the molecular composition of the Stl-hDUT complex. In the native MS spectrum peaks of the hetero-oligomer appeared and displayed a Gaussian-like distribution, with the maximum at m/z 4622 (z= 26\(^+\)) (Figure 28, Panel B). The molecular mass of the complex was found to be 120.0 ± 0.1 kDa, which argues for the existence of complex with the molecular formula of hDUT\(_3\)Stl\(_2\) constituted by two Stl molecules (65.8 kDa) and a human dUTPase trimer (54.1 kDa).

To determine the probable stoichiometry and three-dimensional structures of hDUT-Stl complexes, in-line SEC-SAXS experiments were performed. The chromatogram showed a single peak with a minor leading edge (Figure 28, Panel C). SAXS parameters extracted from individual scattering frames across this peak were subjected to correlation map analysis (231), which identified the existence of two self-consistent regions including pools of similar frames (region 1 and region 2). The molecular weight parameters of region 1 and region 2 match with that of hDUT\(_3\)Stl\(_3\) and hDUT\(_3\)Stl\(_2\) complexes (Appendix Table A6), respectively, which is consistent with the obtained ITC and ESI-MS results. The significant difference in \(R_g\) and \(D_{\text{max}}\) of region 1 compared to that of region 2 (45 Å and 170 Å vs 39 Å and 140 Å, respectively) also suggests the presence of two type of complexes. The real-space distance distribution functions, \(P(r)\) for both SEC-SAXS regions are characteristic of extended structures, with the region 2 data clearly less extended than that of region 1 (Figure 29, Panel C) (158). \textit{Ab initio} shape reconstructions for the region 1 and region 2 peak data yielded envelopes in which molecular models of heterohexameric and heteropentameric hDUT:Stl assemblies fit well (Figure 31). The effective resolution of these \textit{ab initio} shape models was estimated to be 3.7 ± 0.3 nm and 5.3 ± 0.4 nm for the hDUT\(_3\)Stl\(_3\) and hDUT\(_3\)Stl\(_2\) models, respectively.

All in all, our native gel electrophoresis, native mass spectrometry, isothermal titration calorimetry and SEC-SAXS results presented herein argue altogether for the formation of complexes of 3:3 and 3:2 hDUT to Stl ratios. The DUT\(_3\)Stl\(_2\) stoichiometry was also suggested based on native MS for the complex of Φ11 phage, \textit{Mycobacterium tuberculosis} and \textit{D. melanogaster} dUTPases with Stl (17, 228); while the native gel electrophoresis and native MS results referred for the existence of Φ11DUT\(_3\)Stl\(_3\) complex (17, 117).
Figure 29. SAXS data obtained for Stl, hDUT and their complexes (158). A) SAXS data (circles) and model fits (rigid body models/RBM – solid lines, ab initio – dotted lines,) hDUT (pink), Stl (light blue) and hDUT-Stl complexes corresponding to SEC-SAXS peak region 1 (green) and peak region 2 (dark blue) (cf. Figure 28, Panel C). B) Guinier plots of data in Panel A representing linearity of data at low angles. C) Real-space pair distance distribution functions resulted from an indirect Fourier transformation of the SAXS data in Panel A displays the clear increase in maximum particle dimension in samples containing both Stl and hDUT that is caused by complex formation.

The interaction surface of the human dUTPase with Stl was studied by hydrogen-deuterium exchange mass spectrometry. In general, the obtained HDX-MS data provided further, convincing evidence of the hDUT-Stl complexation, and the results suggest that the interaction is localized to specific regions within the proteins. In the case of hDUT the most extensive changes were observed for peptides from the regions covering residues 34H–50L and while modest but persistent changes in case of peptides from the 89A–110G region were detected, which suggest the involvement of the first three conserved motifs of human dUTPase in Stl-binding (Figure 30, Panels A and C). It is worth mentioning that a region overlapping with the β- and γ-phosphate coordinating Motif 2 was also characterized with a
considerable negative HDX signal, which is consistent with the previous finding that only dUTP but not dUMP interferes with Stl binding of dUTPases (17, 119). In addition to these segments, significant mass shifts were also detected at the carboxy terminal region of the human dUTPase, which includes the fifth conserved motif. However, these ∆masses were declining by time, which refers to a weaker and transient interaction of this region with Stl (Figures A7 and A8 in the Appendix). Representation of the HDX-MS difference data on the surface of the human dUTPase clearly demonstrates that the active site is directly involved in Stl-binding (Figure 30, Panel C). As we have shown, Stl and substrate binding to hDUT are mutually exclusive (Figure 27, Panels B and C), similarly to that of Φ11DUT and MtDUT (17, 159), thus Stl interacts only with the apo form of dUTPase. In this substrate-free conformation, the C-terminal Motif 5 is flexible and the active site is accessible (71, 86, 89, 90). Taken together the enzyme kinetic and HDX-MS data, we propose a mechanistic model in which Stl docks into the substrate binding site of dUTPase unless the access to the cavity is blocked by either the closed conformation of the flexible Motif 5 or the substrate. The observed negative HDX signal at the C-terminal region of the human dUTPase upon complexation could indicate that the flexible Motif 5 may stabilize the hDUT-Stl complex, by forming further interactions with Stl. However, this potential additional effect ought to be transient according to the fluctuation of the H/D exchange rate observed in case of this segment, which is consistent with preceding results showing that Stl-binding of the Φ11DUT mutant lacking Motif 5 was not significantly perturbed (17, 119). Although mutational analysis indicated the more significant contribution of this segment to the complex formation between Stl and 80α phage dUTPase, the experimental data determined for the interaction of phage dUTPases with Stl are also in agreement with our model in general (119).

In the case of Stl, a relatively short Tyr-rich region of the protein (residues 98Y–113Y) showed significant H/D exchange rate decrease upon addition of dUTPase (Figure 30, Panels B and D). Thus this segment is supposed to be involved in the interaction with dUTPase as it has become less solvent accessible. Interestingly the HDX-MS difference plots obtained for Stl with hDUT and Φ11DUT are highly similar (cf. Figure 18 and Appendix Figures A7 and A9), suggesting that Stl has a universal trimeric dUTPase-binding motif. The lack of significant positive HDX signal upon addition of dUTPases to Stl could reinforce our previous postulation that the dimer interface of Stl overlaps with the Stl-dUTPase interaction surface (157) (cf. Chapter 4.1.3). The global decrease in the H/D exchange rate observed for the peptides
along the whole Stl sequence possibly refer to the decrease in the flexibility of the protein upon Stl-dUTPase complex formation (157).

Figure 30. Representation of the hydrogen-deuterium exchange mass spectrometry results referring to the hDUT-Stl interaction surface (158). A-B) The sequence of the human dUTPase and Stl proteins colored according to the obtained HDX signal applying the color-scheme displayed on Panel C. Numbering starts at the first residue of the UniProt sequences of the proteins (UniProt IDs: P33316-2 and Q9F0J8 respectively), the extension compared to UniProt sequence is in italics. The active site residues in case of hDUT and the DNA-binding motif of Stl are boxed. C) Representation of the HDX-MS difference data on the surface of apo state structure of the human dUTPase (PDB ID: 1Q5U) following the color-scheme displayed on this Panel. The last 13 residues from the C-terminal are omitted from the representation since the position of these residues was not resolved in the crystal structure presumably due to flexibility. Position of the substrate analogue is shown to ease visualization of the active site based on the structural alignment of the apo and ligand-bound structures (PDB ID: 3EHW). The substrate analogue is represented as spheres with elemental coloring (carbon white, nitrogen blue, oxygen red, phosphorus orange). The results suggest that Stl binds to substrate binding cavity of hDUT. D) Representation of the HDX-MS difference on the surface of the validated 3D homology model of Stl. Coloring is according to the scale at Panel C.

To assess the molecular shapes of hDUT and Stl and to determine the solution structure of the hDUT-Stl complex at low-resolution, SAXS experiments were performed (158). Data obtained from standard SAXS measurements of hDUT yielded shapes and structural parameters consistent with a single trimeric hDUT and were consistent with that of the previous SAXS and
X-Ray crystallographic studies (71, 90) (Appendix Table A5). SAXS data obtained by standard SAXS analysis of Stl samples suggested that the protein predominantly forms dimers in solution (Figures 19 and 29, Table A5 in the Appendix), matching with the results of the chemical crosslinking and native mass spectrometry experiments (Figure 16) (17). The best-fitting dimeric assembly generated from the Phyre2 model (Figures 19 and 29) was used for SAXS modeling procedures. The inherent flexibility of the protein was also verified and was taken into account for subsequent hybrid modelling procedures.

*Ab initio* shape reconstructions for peak region 1 and 2 data extracted from size-exclusion chromatography coupled SAXS experiments yielded envelopes with effective resolution of $3.7 \pm 0.3$ nm and $5.3 \pm 0.4$ nm, respectively (Figure 31). Both reconstructions could readily accommodate a trimeric hDUT core with the remaining volume occupied by three or two Stl monomers, respective to the size of the assembly. Interestingly, no reasonable solutions for the complexes could be acquired using a dimeric Stl model, suggesting that disruption of the Stl dimer is inevitable for the formation of this complex (158). This observation provides a plausible explanation for the dUTPase-induced perturbation of Stl-DNA complex, since Stl is expected to exert its repressor function as a dimer. We suggest that Stl binding to the trimeric dUTPases initiates dissociation of the Stl dimer which leads to impaired Stl-DNA interaction.

**Figure 31.** Models generated based on SEC-SAXS measurements for hDUT-Stl complex (158). Overlay of the best SAXS/HDX-MS hybrid models of the hDUT$_3$Stl$_3$ (Panel A) and hDUT$_3$Stl$_2$ (Panel B) complexes and the *ab initio* shape envelopes reconstructed from SAXS data (SASBDB ID.: SASDC77 and SASDC87, respectively). The hDUT monomers are displayed as red, yellow and blue cartoons. Stl monomers are shown as light-blue cartoons. Regions of missing sequence added during modeling as dummy residues are presented in wire format.
Rigid body structure calculations were performed applying the SAXS data and constraints based on HDX-MS measurements using the crystal structure of the trimeric hDUT core (PDB ID:1Q5U) and Phyre2 homology model of Stl (cf. Chapter 3.4.10). The top 20 models obtained for hDUT$_3$:Stl$_3$, which provided the best agreement with the SAXS data, were clustered based on the relative position of the N-terminal (residues 1–84, Stl-NTD) and C-terminal (residues 85–267, Stl-CTD) segments of Stl to the dUTPase substrate binding pocket (Table 7) (158).

Based on our previous experimental data we propose that in the complex Stl-CTD should block the active site, while Stl-NTD is most likely pointing towards the solvent (cf. HDX-MS results) (157, 158). The models in Cluster 1 are matching all these criteria, while in case of other group of models one or more of those is not fulfilled. It is important to note that the models showing the best $\chi^2$ statistics were also the ones satisfying these criteria (cf. models 11 and 17, in Cluster 1 in Table 7). A representative model from the top selection is shown on Figure 31 (model 11, $\chi^2 = 1.4$, fit is shown on Figure 29). The best model for the hDUT$_3$:Stl$_2$ complex corresponding to this quaternary structure (Figure 30) had also a good agreement with the SAXS data ($\chi^2 = 1.2$, Figure 29). Representation of the HDX-MS results on the surface of the selected models further supports the validity of those as most of the dUTPase and Stl surface which showed decrease in the H/D exchange rate, is buried in the complex (cf. Appendix Figures A10 and A11). Note that the HDX-MS data reflect the properties of proteolytic peptide fragments, thus it is possible that not every residue in a peptide contribute uniformly to the signal, and involved in protein-protein interaction. Recently developed strategies to simulate HDX-MS difference data for quantitative scoring of docking outputs, may stimulate further refinement studies beyond the scope of this work (232). Nevertheless, more detailed information about the role of individual residues could only be assessed by extensive mutagenesis study of this complex, which is the subject of a current project in our group. Our focused study on the structural requirements of human dUTPase-Stl interaction may serve as the starting point of future development of a species-specific dUTPase inhibitory peptide or protein.
Table 7. Clusters of the top 20 rigid-body models for hDUT, Stl obtained from SAXS and HDX-MS data and $\chi^2$ statistics of the acquired models. Clusters were made based on the relative position of the N-terminal (residues 1–84, NTD) and C-terminal (residues 85–267, CTD) segments of Stl to the dUTPase substrate binding pocket of hDUT. dUTPase (trimer) surface is coloured by grey, Stl models are shown as ribbons with different colors. Note that to ease visibility only one out of the 3 Stls from hDUT, Stl is presented. Since substrate and Stl can not simultaneously bind to the dUTPase, substrate analogues are shown as sticks with atomic coloring (carbon black, nitrogen blue, oxygen red, phosphorus orange) in order to indicate the position of the substrate binding pockets. Based on HDX-MS and other experimental evidence Stl-CTD directly interacts the dUTPase, while Stl-NTD has a limited contribution to Stl-dUTPase interaction (157, 158).

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Models</th>
<th>$\chi^2$</th>
<th>Relative orientation of Stl domains to hDUT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>11 (green) 12 (cyan) 17 (magenta)</td>
<td>1.4 2.3 1.4</td>
<td>CTD is close NTD points out 1 active site is blocked</td>
</tr>
<tr>
<td>2.</td>
<td>1 (green) 16 (cyan) 20 (magenta)</td>
<td>1.8 2.0 1.6</td>
<td>most of the CTD points out NTD contacts the dUTPase zero active site is blocked</td>
</tr>
<tr>
<td>Cluster</td>
<td>Models</td>
<td>$\chi^2$</td>
<td>Relative orientation of Stl domains to hDUT</td>
</tr>
<tr>
<td>---------</td>
<td>--------</td>
<td>--------</td>
<td>------------------------------------------</td>
</tr>
<tr>
<td>3.</td>
<td>2 (green) 10 (cyan) 14 (magenta)</td>
<td>1.7 1.5 1.5</td>
<td>CTD is close, NTD points out zero active site is blocked</td>
</tr>
<tr>
<td>4.</td>
<td>3 (green) 5 (cyan) 6 (magenta) 8 (yellow) 9 (salmon) 13 (purple) 18 (orange)</td>
<td>1.8 1.7 1.8 1.6 1.7 2.0 1.8</td>
<td>most of the CTD points out NTD contacts the dUTPase zero active site is blocked</td>
</tr>
<tr>
<td>5.</td>
<td>7 (green) 15 (cyan) 19 (magenta)</td>
<td>1.8 2.1 1.8</td>
<td>most of the CTD points out NTD contacts the dUTPase 1 active site is blocked</td>
</tr>
<tr>
<td>Cluster</td>
<td>Models</td>
<td>$\chi^2$</td>
<td>Relative orientation of StI domains to hDUT</td>
</tr>
<tr>
<td>---------</td>
<td>--------</td>
<td>---------</td>
<td>------------------------------------------</td>
</tr>
<tr>
<td>6.</td>
<td>4 (green)</td>
<td>2.2</td>
<td>most of the CTD points out NTD contacts DUT zero active site is blocked</td>
</tr>
</tbody>
</table>
4.4 Interaction of Stl with a dimeric phage dUTPase

It was discovered recently that not only the trimeric but the dimeric phage dUTPases are capable to interact with Stl of SaPl_{bov1} (14, 15, 103). This finding was unexpected since the structure of dimeric and trimeric dUTPases fundamentally differ from each other: the dimeric dUTPases are all-α helical proteins while the trimeric dUTPases have a β-pleated 3D fold (cf. Chapter 1.3). The two structurally distinct dUTPase families have only the dUTP binding ability in common, so it was compelling to speculate on the involvement of this region in Stl-binding. Although it has been shown that the enzymatic activity is not essential for SaPl mobilization capacity of these phage dUTPases (94, 103), which seemingly could argue against this hypothesis. Nevertheless, in case of Φ11DUT we have shown by HDX-MS that the Stl-dUTPase interaction surface includes the substrate binding pocket of the enzyme (cf. Chapter 4.2.1.).

To explore the Stl-binding characteristics of the dimeric phage dUTPase we examined the complex formation between Stl and φNM1DUT. It has been shown that Stl is able to reduce the enzymatic activity of the φNM1 phage dUTPase (14). We quantitatively analyzed the inhibition of φNM1DUT by Stl with performing enzyme activity measurements of φNM1 phage dUTPase in the presence of Stl of different concentrations (Figure 32, Panel A) (175). The maximal extent of inhibition exerted by Stl was found to be about 40%, thus more than half of the original enzymatic activity was maintained even at relatively high concentration of Stl. This is a marked difference compared to the complete loss of dUTPase enzymatic activity observed upon Φ11DUT-Stl complex formation within the same assay conditions (17, 117). Although this fact alone does not refer to weaker binding per se, as for example in case of competitive inhibition observed for the Φ11DUT Stl system (17), the extent of inhibitory effect on enzymatic activity is determined by the dissociation and association kinetics of both the inhibitor and the ligand. However, the apparent inhibitory constant obtained for φNM1DUT, (K_i= 34 ± 14 nM) was arguing for the slightly weaker binding of Stl to this dUTPase. The exact mechanism of inhibition can only be revealed by detailed kinetic and thermodynamic characterization of the dimerization and substrate binding of the φNM1DUT, which was beyond the scope of this study.

Next, we investigated the stoichiometry of the φNM1DUT-Stl complex by applying various biochemical and biophysical methods. The individual proteins and samples containing
the mixture of the two proteins in different relative concentrations were subjected to native gel electrophoresis (Figure 32, Panel B) (175). On the resulting gel the bands of φNM1DUT and Stl practically disappear in case of the samples containing the two proteins in ca. 1:1 molar ratio, while in parallel a band appeared at a third, distinct position. The band of the complex is located between the positions of the Stl and φNM1DUT dimer bands, which taken together with the DUT:Stl molar ratio of these samples suggests the formation of the heterodimeric φNM1DUT-Stl complex.

**Figure 32. Biophysical characterization of φNM1 phage dUTPase-Stl interaction (175).**

A) Enzyme activity of dimeric φNM1DUT in the presence and absence of Stl. The maximal extent of inhibition was about 40%, the binding is characterized by the apparent inhibitory constant of $K_i = 34 \pm 14 \text{nM}$. B) Native gel electrophoresis of Stl, φNM1DUT and the mixture of the two proteins of various molar ratios. The emerging band between that of the individual dimeric proteins in case of the samples containing the mixture of the two proteins suggest the formation of an Stl-φNM1DUT heterodimer. C) SDS-PAGE analysis of Stl, φNM1DUT and their mixture after chemical crosslinking, the mixture of the untreated proteins was also loaded on the gel as a control. φNM1DUT-Stl heterodimer is denoted with a red star. D) The native mass spectrum of the φNM1DUT-Stl mixture. Peak series with m/z values of 2946 (18+), 3119 (17+), 3315 (16+), 3535 (15+), 3788 (14+) indicates the presence of an assembly associated with the molar mass of 53020 ± 7 Da, which corresponds to a 1:1 complex of Stl (32.0 kDa) and φNM1DUT (21.0 kDa). Peaks corresponding to φNM1DUT dimer are also present in the spectrum.
The assignment of the bands on the native gels was further supported by the results obtained from our chemical crosslinking experiments (Figure 32, Panel C) (175). The SDS-PAGE analysis of the crosslinked samples containing the individual proteins resulted in the presence of two bands for each protein on the gel with molecular weights corresponding to the monomeric and dimeric forms of proteins (Figure 32, Panel C). The same treatment of the mixture of φNM1DUT and Stl in 1:1 molar ratio led to the appearance of a band between the position of Stl dimer and the φNM1DUT dimer bands, with a molecular weight of ca. 55 kDa. These results both refer to the formation of φNM1DUT:Stl heterodimer, which is consistent with previous observations applying shorter and less specific crosslinkers (14). In case of chemical crosslinking of Φ11DUT with Stl the resulting covalent complex barely entered the resolving gel possibly due to its extended size (cf. Figure A12 in the Appendix).

The 1:1 composition of the φNM1DUT-Stl complex was also confirmed by native mass spectrometry measurements (Figure 32, Panel D) (175). In addition to the peak series corresponding to φNM1DUT monomer and dimer peaks associated with the molar mass of 53.0 kDa were also observed in the spectrum which argues for the existence of a 1:1 complex of Stl (32.0 kDa) and φNM1DUT (21.0 kDa). We have not found indication for the presence of any higher order complex neither in the native mass spectrometry nor in the chemical crosslinking experiments. Therefore, we concluded that φNM1DUT and Stl form only one type of complex, which has 1:1 stoichiometry. Thus, the complex formation of dimeric and trimeric phage dUTPases with Stl markedly differ from each other. While Φ11DUT remains in the trimeric oligomeric state upon complex formation and interacts with 2 or 3 Stl molecules, the dimeric dUTPase from φNM1 bacteriophage dissociates to monomers, which then bind to an Stl monomer. These results suggest a plausible model for the mechanism of enzymatic inhibition of φNM1DUT by Stl, since the active site is located at the dimer interface of the protein, which is likely affected by Stl binding. Complex formation of Stl with φNM1DUT definitely necessitate the dissociation of the Stl dimer, that possibly prohibits the binding of the repressor to its cognate DNA site since Stl proposedly binds to DNA as a dimer. Based on the results obtained for hDUT and Stl a similar mechanism may govern the derepression action of the trimeric dUTPases.

We also wished to inquire whether Stl binds to dUTPases by the same segment or its promiscuity to bind and inhibit dUTPases of different folds is orchestrated by different binding regions involved in Stl-DUT interaction. In order to provide exclusive insight to the structural details of complex formation we performed hydrogen-deuterium exchange mass spectrometry
measurements (175). In case of the φNM1DUT-Stl complex no significant mass change was detected for peptides covering the region of the amino-terminal 200 residues, however peptides from the 60-residue-long segment form the very C-terminal part of the Stl sequence showed marked negative HDX signal (Figure 33 and Figure A13 in the Appendix). These results are consistent with the previous results indicating that binding of Stl-CTD to φNM1DUT is not perturbed despite this protein lacks the N-terminal segment (84 residues) of Stl (103, 220).

Comparing the difference plots of Stl in the presence of the dimeric and trimeric dUTPases reveal dramatically different Δmass profiles. In the presence of Φ11DUT, Stl exhibits significant negative mass shifts across the entire protein backbone although the binding is mostly localized to the protein region of residues 98Y–113Y (Figure 18). These results indicate that different segments of Stl interacts with the two phage dUTPases of different fold.

Figure 33. Experimental results and schematic models displaying how the Stl repressor protein interacts with the trimeric Φ11DUT and the dimeric φNM1DUT phage dUTPases (175). A) HDX difference plots the Δmass of each individual peptide of Stl upon mixing with of Φ11DUT (red) and φNM1DUT (green). The dashed line represents the 95% confidence bands evaluated over the whole dataset. B) The in silico 3D homology model of Stl colored according to HDX-MS data obtained upon binding to Φ11DUT (left) and φNM1DUT (right), respectively, following the color gradient on the bottom of the figure.

The HDX-MS results for φNM1DUT showed a significantly more complex binding pattern as compared to that of Φ11DUT (cf. Figure 34, Panels A and B, and Figure A14 in the Appendix) (175). Peptides spanning residues 15–36 and 155–171 of φNM1DUT showed marked negative HDX signal consistent with the binding of Stl and occlusion of these sites from isotope exchange. These regions contain active site residues of key-importance in enzymatic activity including Q17, D21 (residues of Motif 1 responsible for uracil binding), K159 and R166 (residues of Motif 5 responsible for phosphate binding) as determined by
sequence alignment against dimeric phage dUTPases of known structure and in case of K159 also by alanine mutagenesis (Figure 34, Panels B and C) (15, 103). This suggests that part of the active site is directly involved in the complex formation with Stl, as these residues of the φNM1DUT became less accessible to solvent upon Stl-binding. Note that the specific mutation of K159 to alanine did not abolish φNM1DUT-Stl complex formation \emph{in vitro}, thus this residue is not an essential factor in the protein-protein interaction (103).

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure34.png}
\caption{HDX-MS results for φNM1DUT upon mixing with Stl (175). A) HDX difference plot showing the Δmass of each individual peptide of φNM1DUT upon mixing with Stl. The dashed line represents the 95% confidence bands evaluated over the whole dataset. B) Sequence of φNM1DUT letter coloring is corresponding to HDX signal (color scheme shown in Panel C), active site building conserved motifs are boxed. C) The \textit{in silico} 3D homology model of φNM1DUT dimer colored according to HDX-MS data following the color gradient on the bottom, substrate analogue dUPNPP is shown as black sticks in order to ease visualization of the active sites (views: front, back, bottom).
}
\end{figure}

The φNM1 phage dUTPase also shows extended regions of positive Δmass in the presence of Stl including peptides covering residues 48–65, 89–97 and 116–129, indicating that these protein regions become more solvent accessible in the presence of Stl. To provide an explanation for this observation HDX-MS outputs were mapped directly onto a 3D homology model of φNM1DUT generated based on the crystal structure of φDI phage dUTPase (Figure 34, Panel C). According to the analysis of residue-residue interactions across the dimer interface of the \textit{in silico} 3D model performed with DIMPLOT (233) residues 45–58 and 111–124 are located on the dimer interface. Thus the increase in H/D exchange rate detected in
these regions could correspond to the increased solvent accessibility of these segments due to
the dissociation of \( \varphi \text{NM1DUT} \) dimer upon complex formation with Stl. All in all, the native
gel electrophoresis, chemical crosslinking, enzyme activity, native MS and HDX-MS
experiments all support the formation of a heterodimer complex constituting \( \varphi \text{NM1DUT} \) and
Stl.

Altogether, our experiments reveal previously unreported functional plasticity of Stl and
drastically different binding mechanisms of Stl for the two different phage dUTPases. The fact
that the SaPI repressor Stl divergently evolved to target phages encoding dUTPases may refer
to that the mobile genetic elements gain benefit from the uracil-free environment. One possible
advantage is that the low dUTP level, provided by the dUTPase enzymatic action, enhances the
fidelity of SaPI replication via diminution of the mutation rate (23), while reducing the potential
for the selective evolution of phages to escape SaPI interference in parallel.

Based on our results Figure 34 shows a schematic model, which describes interaction of
Stl with dimeric and trimeric dUTPases.

---

**Figure 35. Interaction of the Stl protein with dimeric and trimeric dUTPases.** Stl (orange ellipses)
forms dimers in solution and may bind to DNA (black) as dimers. Perturbation of Stl dimerization by
dUTPases leads to the dissociation of the Stl-DNA complex. Stl inhibits both the trimeric (light blue
triangles) and dimeric dUTPases (dark blue rectangles). The inhibition is based on competition
between Stl and the substrate, dUTP (black dots) in case of the trimeric dUTPases. Stl monomers and
dUTPase trimers form DUT\(_3\)Stl\(_2\) and DUT\(_3\)Stl\(_3\) complexes. The substrate binding site of dimeric
dUTPases, which is located at the dimerization surface of the enzyme, is impaired upon formation of a
heterodimeric complex of the dUTPase with Stl. This explains the reduction of the enzymatic activity
of dimeric dUTPases in the presence of Stl.
Stl protein dimerizes in solution and based on the similarity with other repressors and the symmetry of the specific binding site of the protein within the SaPI DNA it is assumed that Stl binds to DNA as dimers (Figure 34) (158, 176). Interaction of Stl monomers with dUTPases perturbs the dimerization of the repressor, hence it leads to the dissociation of the Stl-DNA complex. Trimeric dUTPases can form DUT$_3$Stl$_2$ and DUT$_3$Stl$_3$ complexes with Stl, while in the case of dimeric dUTPases the complex is a DUT-Stl heterodimer (Figure 34) (17, 158, 159, 228). Based on our results Stl binds directly to the active site of trimeric dUTPases and it acts as a competitive inhibitor of these enzymes (17, 175). Stl also reduces the activity of dimeric dUTPases, although via a different mechanism, than that observed for the trimeric enzymes. Inhibition of dimeric dUTPases by Stl is resulting from the of the abolition of the active site, which resides at the dimer interface of the enzyme, during complex formation between the two proteins (175).
Summary

This study aimed to explore the structure and the related functional characteristics of the SaPIbov1 Staphylococcus aureus pathogenicity island (SaPI) regulator Stl protein and its complexes with different dUTPases. This system functions as a molecular switch within S. aureus, which controls the horizontal gene transfer of the mobile genetic element via interaction of Stl with phage dUTPases leading to dissociation of Stl from its consensus DNA binding site and the subsequent expression of specific SaPI proteins responsible for mobilization of the pathogenicity island (16). In the meantime, Stl was also identified as a potent inhibitor of the phage dUTPase (16), which might constitute a basis for the development of specific dUTPase inhibitors for in vivo studies on the enzyme function.

First, we focused on to resolve how potential domains within the Stl repressor may be defined (157). Towards this end, we created an in silico structural model of the full-length Stl protein and validated the predicted fold by synchrotron radiation circular dichroism (SRCD) experiments. Based on this 3D model verified by CD results, we produced truncated and point mutants of the Stl protein and studied their function in DNA- and dUTPase-binding. We showed that the produced carboxy-terminal segment is an independently folded domain, which retains its dimerization ability and binding affinity to Φ11 phage dUTPase (Φ11DUT), but exhibits reduced inhibitory effect. The amino-terminal putatively DNA-binding segment of Stl was also studied by point mutations. Our experimental results convincingly support the predicted position of helix-turn-helix motif. These in vitro results were reinforced by our recent in vivo experiments aiming to investigate this molecular switch in Mycobacterium smegmatis (219).

The interaction surface of Stl with Φ11DUT was explored by hydrogen-deuterium exchange mass spectrometry (HDX-MS), which reports on changes in the local isotope exchange rate of the protein upon complex formation. In the presence of Φ11DUT, a specific region within the C-terminal domain of the Stl protein (residues 98Y–113Y) displayed a pronounced negative mass shift (175). This suggests that this region is principally responsible for the Stl-Φ11DUT interaction. In addition to this, significant negative mass shifts were observed across the entire length of Stl, referring to global conformational tightening of the protein upon complex formation. Taken together the experimental data we proposed that the dimerization and dUTPase-binding surface of Stl overlap with each other. Thus we propose that dUTPase initiates dissociation of the Stl dimer, which presumably perturbs the Stl-DNA interaction since Stl is expected to bind DNA as a dimer. Therefore, this postulation provides a potential mechanistic model of the derepression based on experimental data. The structure of
the Stl dimer was also modeled based on standard small-angle X-ray scattering measurements (SASBDB ID: SASDC67) (158). The resulting ab initio and rigid-body models referred for inherent flexibility of the protein, which hindered its crystallization.

The structure of the Φ11 phage dUTPase is highly similar to that of other trimeric dUTPases like the human and mycobacterial enzymes, although besides the well conserved protein structural core it contains phage-specific insert segment, which folds into a β-pleated mini-domain (84). We explored the role of the phage specific insert of Φ11DUT in the enzymatic activity, in the dUTPase-Stl interaction and in the disruption of Stl-DNA complex (84, 117). We have shown previously that the truncated enzyme mutant Φ11DUTΔ101G–122Q has similar catalytic properties as the full-length phage dUTPase (84). The insert was also found dispensable for complex formation with Stl based on native gel electrophoresis experiments and enzyme activity assays. Based on the observed inhibitory effect and negative HDX signals around the ligand binding cavity upon addition of Stl to Φ11DUT and Φ11DUTΔ101G–122Q we propose that Stl binds to the active site of both dUTPases. The HDX-MS results reflect that the insert partially hinders the accessibility of the active centrum, while also provides additional possibilities for Stl-binding. On the other hand, our electrophoretic mobility shift assay experiments indicated that Φ11DUTΔ101G–122Q does not effectively interfere with the interaction between Stl and DNA (117). These results suggest that within its own protein context the phage-specific insert is essential for the derepression function.

To provide a more general view on this subject we tested whether Mycobacterium tuberculosis and human dUTPases (MtDUT and hDUT) are able to bind to Stl and to disturb the Stl-DNA complex formation in vitro (117, 158, 159). We found that Stl is a potent inhibitor of both hDUT and MtDUT, and both dUTPases showed capability to disrupt the Stl-DNA interaction. Taking together these results with experimental data available for the Drosophila melanogaster dUTPase-Stl complex formation (228), and our preliminary results obtained for the interaction of Stl with Plasmodium falciparum and E. coli dUTPases we concluded that there is no obvious relationship between the inhibition of a specific enzyme by Stl and the ability of that dUTPase to disrupt the Stl-DNA complex.

Nevertheless, understanding the molecular mechanism of Stl action on dUTPase might constitute the first step towards the design of an efficient and selective proteinaceous inhibitor. Towards this end, we followed an integrated structural biology approach, based on the combination of isothermal titration calorimetry (ITC), native mass spectrometry, and most prominently size-exclusion chromatography in line with small-angle X-ray scattering and
hydrogen-deuterium exchange mass spectrometry to present detailed structural insights on the hDUT-Stl interaction (158).

The obtained experimental data allowed us to construct 3D structural models for heterohexameric (hDUT₃Stl₃) and heteropentameric (hDUT₃Stl₂) complexes for hDUT-Stl assemblies (SASBDB ID: SASDC77, SASDC87). Importantly, no rational solutions for the complexes could be realized by applying a dimeric model for Stl, suggesting that disruption of the Stl dimer could be a prerequisite for complex formation between Stl and trimeric dUTPases. The proposed perturbation of Stl dimerization by dUTPases could be the basis of the derepression ability of those since Stl is expected to exert its repressor function as a dimer.

It has been shown that representatives of both the β-pleated trimeric and all-α dimeric phage dUTPases interact with the Stl repressor protein within *Staphylococcus aureus* (14). We set out to characterize the structural background of the promiscuity of Stl for dUTPase binding. The obtained native gel electrophoresis, chemical crosslinking, enzyme activity, native MS and HDX-MS results all argue for the formation of a heterodimer complex constituting φNM1DUT and Stl, which suggest the disruption of the dUTPase dimer. Since both φNM1DUT and Stl function as a dimer, the formation of a heterodimer provides a plausible explanation for the mutual functional inhibition of the two proteins within the complex (175). In contrast to φNM1DUT, the Φ11 phage dUTPase remains in the trimeric oligomeric state upon complex formation with Stl (17). Our experiments revealed drastically different interaction surface of Stl with the dimeric φNM1DUT and trimeric Φ11DUT and also referred to distinct Stl-binding mechanisms of the two dUTPases (175). The presented adaptation of Stl to interact with the different dUTPases may serve as a basis to support the hypothesis, that the presence of dUTPases could have an advantage for the SaPI mobile genetic elements.
A **SaPI** 

**Staphylococcus aureus** patogenicitási sziget bioregulációjáért felelős fehérje szerkezeti és funkcionális vizsgálata

A bakteriális genom gyakran tartalmaz olyan mobilis genetikai elemeket, melyek a bakteriális kromoszómától többé-kevésbé független módon képesek replikálódni és ezt követően horizontális géntranszfer folyamatok során különböző törzsekbe átjutni. A **Staphylococcus aureus** (Sa) baktériumban egy ilyen ún. patogenicitási sziget (PI) terjedését egy fehérje-fehérje kölcsönhatás szabályozza, melyen belül a Φ11 bakteriofág dUTPáz fehérjéje a SaPI génaszakasz bioregulációjáért felelős Stl represszor fehérjével komplexet képez, és ezáltal lehetővé válik a replikáció és a génexpresszió (16). Munkám célja ezen molekuláris kapcsoló működési mechanizmusának felértékelése volt. A rendszer érdekességét adja továbbá, hogy a képződött fehérje-fehérje komplexben a dUTPáz enzimatikus aktivitása gátolt (16), ami lehetőséget nyújt új, fehérje-típusú dUTPáz inhibitorok fejlesztésére.

Először *in silico* módszerekkel prediktáltuk a DNS-kötő motívum pozícióját és létrehoztunk egy homológia modellt. Az *in silico* szerkezeti modell alapján a fehérje α-hélixekből és az azokat összekötő flexibilis szakaszokból áll; a másodlagos szerkezeti elemek hasonló eloszlását a fehérje szinkrotron sugárzási cirkuláris dikroizmus spektrumának elemzése is megerősítette. A háromdimenziós szerkezeti modell alapján az Stl fehérje két szegmensből áll: egy amino-terminális (N-terminális) részből, mely a valószínűsíthetően a DNS kötésért felelős hélix-turn-hélix szerkezeti motívumot tartalmazza és egy karboxi-terminális (C-terminális) szegmensből, mely feltehetően a fehérje-fehérje kölcsönhatásban játszik fontos szerepet. Kísérleteink igazolták, hogy a fehérje C-terminális szakasza egyaránt képes a dimerizációra és a dUTPáz inhibícióra, azonban a DNS-hez nem kötödik. A predikciós programok által meghatározott, hélix-turn-hélix motivumban két kiválasztott poláris oldallancú aminosavat alaninra cseréelve (Stl-AA) a kapott pontmutáns fehérje DNS-kötő képessége a vad típusú fehérjéhez viszonyítva jelentősen csökkent. Emellett egy *Mycobacterium smegmatis* modell rendszerben igazolható, hogy a ezek a mutáns Stl fehérjék *in vivo* sem képesek a DNS-hez kötödnek (219). Ezek az eredmények alátámasztják, hogy az *in silico* módszerek megfelelően jeleztek előre a hélix-turn-hélix motivum helyét a fehérje amino-terminális szakaszán (157).

Az Stl és a Φ11 bakteriofág dUTPáz (Φ11DUT) közötti kölcsönható felszínt hidrogén-deutérium cserés tömegspektrometriás (HDX-MS) módszerrel vizsgáltuk, amely az izotópcsere sebességének lokális változásairól ad információt (175). A Φ11DUT fehérje

Kisszögű röntgenszórás (SAXS) kísérletekkel meghatároztuk az Stl dimer oldatbeli alakját, és ez alapján alkottunk egy alacsony felbontású szerkezeti modellt (SASBDB ID: SASDC67) (158). A SAXS adatok emellett arra utaltak, hogy a fehérje flexibilis régiókat is tartalmaz ami megmagyarázza a fehérje kristályosítására tett kísérleteink sikertelenségét.


Ezen felül kimutattuk, hogy érdekes módon a Staphylococcus aeurus baktériumból származó Stl fehérje kölcsönhatásába lép a mikobakteriális és a humán dUTPáz fehérjével is, és jelentősen csökkenti azok katalitikus aktivitását (117, 158, 159). Emellett mindkét dUTPáz, ugyan különböző mértékben, de képes megbontani az Stl-DNS kölcsönhatást. Hasonló eredményeket kaptunk a Drosophila melanogaster dUTPáz esetében is (228). Ezen eredmények és további Plasmodium falciparum és az E. coli dUTPázok esetében rendelkezésünkre álló előzetes kísérleti tapasztalataink alapján nem állítható fel egyértelmű
kapcsolat a különböző trimer dUTPázok Stl-általi gátlása és azok Stl-DNS komplexképződést gátló hatása között.

Mindazonáltal a különböző dUTPázok és Stl közötti kölcsönhatófelszín és a komplexek szerkezetének ismerete lehetőséget teremt új, fehérje vagy peptid típusú dUTPáz inhibitor fejlesztésére. Ennek érdekében integrált szerkezeti biológiai megközelítést alkalmazva natív gél elektroforézis, izotermális titráló kalorimetria (ITC), natív tömegspektrometria, valamint gélszúréssel kapcsolt kisszögű röntgenszórás (SEC-SAXS) és HDX-MS módszerekkel kapott eredményeink kombinálásával kísérletet tettünk a humán dUTPáz (hDUT) Stl fehérjével alkotott komplexének szerkezeti modellezésére (158). A SAXS adatok alapján végzett modelllépítés során a dimer Stl modell felhasználásával nem jutottunk olyan modellhez, amely összeegyeztethető a más módszerekkel kapott kíséreti eredményekkel. Ezzel szemben a hDUT trimer és Stl monomerek kombinálásával kapott hDUT₃Stl₂ és hDUT₃Stl₁ szerkezeti modellek (SASBDB ID: SASDC77, SASDC87) jól megfelelnek az összes rendelkezésre álló kísérleti tapasztalatnak. Ez alapján az Stl dimer disszociációja feltehetően a dUTPázzal való kölcsönhatás előfeltétele. A dimerizációs egyensúly dUTPáz hatására bekövetkező eltolódása magyarázatot kínál a derepresszió mechanizmusára is, mivel az Stl fehérje valószínűleg dimerként kötődik a DNS-hez.

Érdekes módon az Stl fehérje a β-redős trimer dUTPázok mellett képes kölcsönhatásba lépni a teljesen eltérő szerkezetű φNM1 fágból származó α-helikális dimer dUTPázzal is (φNM1DUT) (14). A két különböző dUTPázzal való komplex-képzés szerkezeti háttérenek felderítésére natív gél elektroforézis, kovalens keresztkötési, natív tömegspektrometria és HDX-MS kísérleteket végeztünk (175). Ezek eredménye alapján a φNM1DUT és Stl fehérjék heterodimer komplexet alkotnak. Ez önmagában magyarázza a két fehérje interakciója során fellépő kölcsönös funkcióvesztést, ugyanis mindkét fehérje esetében csak a homodimer forma lehet aktiv: A dUTPáz aktív centrum a dimer kölcsönható felszínen helyezkedik el míg az Stl fehérje dimerként kötődik a DNS-hez. A HDX-MS mérések alapján az Stl fehérje különböző szegmenseivel kötődik a két dUTPázhoz. Az Stl fehérje különleges alkalmazkodása, amely lehetővé teszi a különböző szerkezeti dUTPáz izoenzimek felismerését, felveti annak a lehetőségét, hogy a fág dUTPázok valamilyen módon előnyt jelentenek a SaPi mobilis genetikai elemek számára.
7 References

16. Tormo-Más, M.Á., Mir-Sanchis, I., Shrestha, A., Tallent, S.M., Campoy, S., Lasa, Í., Barbé, J.,


76. Zang, K., Li, F. and Ma, Q. (2018) The dUTPase of white spot syndrome virus assembles its active


82.


221. Rey,M., Sarpe,V., Burns,K.M., Buse,J., Baker,C.A.H., van Dijk,M., Wordeman,L.,


8 Publication list

8.1 List of my first-author publications related to the thesis


3. **Nyíri K.**, Vértesy G. B. *Perturbation of genome integrity to fight pathogenic microorganisms (2017)* Biochimica et Biophysica Acta - General Subjects, 1861:(1) 3593–3612. review article


8.2 List of my co-author publications related to the thesis


8.3 List of other publications


8.4 List of conference presentations related to the thesis

8.4.1 Oral presentations at conferences


4. **Nyíri K., Matejka J., Harris M. J., Borysik A. J., Vértessy B. G.** *How drastically different families of dUTPases interact with the same inhibitor protein: structural insights (2017)* Proteins In Action Conference, Ceske Budejovice, Czech Republic (in English)

5. **Nyíri K., Mertens H., Tihanyi B., Nagy G. N., Kőhegyi B., et al.** *Towards development of peptide inhibitors of the human dUTPase (2017)* Molecular and cellular mechanisms of human diseases – 9th ÖGMBT Annual Meeting & 8th Life Science Meeting, Innsbruck, Austria (in English)

6. **Nyíri K., Mertens H., Tihanyi B., Nagy G. N., Kőhegyi B., et al.** *Exploring a proteinaceous inhibitor of the human dUTPase: pioneering structural model of the inhibitory complex (2018 July), 8th FEBS Young Scientists’ Forum (YSF) Prague, Czech Republic (in English)

8.4.2 Poster presentations at conferences


Appendix

Appendix Figures

Figure A1. Comparison of the 3D models generated by Phyre2 and Modeller (A) Ribbon representation of the homology model of the *Staphylococcus aureus* pathogenicity island repressor Stl produced by Phyre2 Server (160). Based on the homology model the protein is highly α-helical (74%), and seems to be divided into two segments: the amino terminal segment colored cyan and the carboxy-terminal segment colored hotpink. According to Pfam and NCBI CDD the protein is predicted to contain a helix-turn-helix DNA binding motif. The position of the HTH predicted by NPS@ server is colored to darkblue (162). (B) Ribbon representation of the homology model of Stl obtained by Modeller (162), predicted HTH colored yellow. (C) Superimposition of the two models. Both models agreed in that the protein is mostly α-helical and contains an N-terminal HTH motif.

Figure A2. IUPRED prediction for Stl és a lambda CI repressors (234). Coloring of the protein follows the figure legend. Position of HTH motifs within the proteins is highlighted.
Figure A3. Testing the DNA binding ability of Stl-AA (157). Electrophoretic mobility shift assay was performed to investigate the DNA binding ability of Stl-AA. Concentration of Stl monomer in the samples is indicated on the figure. The band of the dsDNA is only partially shifted upwards even if high concentrations of Stl-AA were applied. Wild type Stl shows shift of the same amount of DNA at concentration of 1 µM (cf. Figure 15 of the thesis).

Figure A4. Coverage map of HDX-MS difference plots for Stl upon complex formation with Φ11DUT (175). Coverage map describing the distribution of 84 individual peptides (horizontal bars) covering 94.0% of the Stl sequence with redundancy of 3.8 in the presence of Φ11DUT.
Figure A5. Coverage map of HDX-MS difference plots for Φ11DUT upon complex formation with Stl (175). Coverage map describing the distribution of 84 individual peptides (horizontal bars) covering 95.3% of the Φ11DUT sequence with redundancy of 6.0 in the presence of Stl.

Figure A6. Coverage map of HDX-MS difference plots for Φ11DUTΔ101G-122Q upon complex formation with Stl (175). Coverage map describing the distribution of 70 individual peptides (horizontal bars) covering 92% of the Φ11DUTΔ101G-122Q sequence with redundancy of 6.2 in the presence of Stl.
Figure A7. HDX-MS difference plots for hDUT (Panel A) and Stl (Panel B) displaying the change in isotope uptake upon complexation of the proteins (158). Labelling time points are indicated by different colors and the dashed lines represent the 95% confidence bands.

Figure A8. Coverage map of HDX-MS difference plots for hDUT upon complex formation with Stl (158). Coverage map of human dUTPase (hDUT) describing the distribution of 62 individual peptides (horizontal bars) covering 95.7% of the hDUT sequence with redundancy of 4.4 in the presence of Stl.
Figure A9. Coverage map of HDX-MS difference plots for StI upon complex formation with hDUT (158). Coverage map of StI describing the distribution of 63 individual peptides (horizontal bars) covering 94.0% of the StI sequence with redundancy of 3.0 in the presence of hDUT.
Figure A10. Surface representation of Stl-dUTPase complexes (158). Stl is colored orange, hDUT colored according to HDX-MS applying the color-scheme displayed on the bottom of the figure. Most of the dUTPase surface showing negative HDX-MS signal is buried by Stl in the complex.
Figure A11. Surface representation of Stl-dUTPase complexes (158). hDUT is colored orange, Stl colored according to HDX-MS applying the color-scheme displayed on the bottom of the figure. Most of the Stl surface which showed the most negative HDX-MS signal is buried by hDUT in the complex.
Figure A12. Chemical crosslinking experiment with Stl and Φ11DUT and the mixture of the two proteins (175). Upon addition of the crosslinking agent to the individual proteins bands corresponding to Stl monomer (32.9 kDa) and dimer (65.8 kDa), and to Φ11DUT monomer (18.4 kDa), dimer (36.8 kDa) and trimer (55.2 kDa) appeared on the gel resulted after denaturing electrophoresis. The crosslinked Φ11DUT-Stl complex barely entered the resolving gel perhaps due to its extended size. Species and concentrations given in monomers are indicated on the figure.

Figure A13. Coverage map of HDX-MS difference plots for Stl upon complex formation with φNM1DUT (175). Coverage map describing the distribution of 89 individual peptides (horizontal bars) covering 94.8% of the Stl sequence with redundancy of 3.8 in the presence of φNM1DUT.
Figure A14. Coverage map of HDX-MS difference plots for φNM1DUT upon complex formation with Stl (175). Coverage map describing the distribution of 70 individual peptides (horizontal bars) covering 97.3% of the φNM1DUT with redundancy of 3.5 in the presence of Stl.
Appendix Tables

**Table A1.** Primers used for PCR amplifications of genes encoding proteins of interest

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence (5’→3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NM1-GST-F</td>
<td>TATTGGATCCATGGCTAGCATAAACACATTAACA</td>
</tr>
<tr>
<td>NM1-GST-R</td>
<td>GGTCCTCGAGTTACAGTATCCTTTTCCTGCG</td>
</tr>
<tr>
<td>Stl-CTD-F</td>
<td>TATTGAATTCAGCCGACCCTGAACG</td>
</tr>
<tr>
<td>Stl-CTD-R</td>
<td>GGTCCTCGAGTTAGGATCTCTTTTCTCATATAATTTTCTCTGATG</td>
</tr>
</tbody>
</table>

**Table A2.** Primers used for mutagenesis

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence (5’→3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stl-AA-mut-F</td>
<td>CGCACCGGCTTTAGCGCAGGACCATTAGCAACCATGAAAACG</td>
</tr>
<tr>
<td>Stl-AA-mut-R</td>
<td>CGTTTCATGGTGGCTAAATGTCGCTGCTGCAAAAGCCGGTGC</td>
</tr>
<tr>
<td>Stl-dHis-mut-F</td>
<td>GTGGATCCCCCGAATTCAGCATGGAAGGCGCG</td>
</tr>
<tr>
<td>Stl-dHis-mut-R</td>
<td>CGCGCTTCCATGCTGAAATCCGGGGATCCAC</td>
</tr>
<tr>
<td>Stl-NT-mut-F</td>
<td>CGTTCAAGGCTGGCTCAGTTAGCTAAATAGCCTTTTCTTTAATTCTAGC</td>
</tr>
<tr>
<td>Stl-NT-mut-R</td>
<td>GCGATGAATTAAAGAAAGGCTATTAGCTAAGGGGCCGCAGGCTAG</td>
</tr>
</tbody>
</table>
Table A3. Sequence and properties of protein constructs (purification tags are in italics, residues deleted or modified by mutagenesis are underlined)

<table>
<thead>
<tr>
<th>Protein</th>
<th>Szekvencia</th>
<th>MW (kDa)</th>
<th>ε¹</th>
</tr>
</thead>
<tbody>
<tr>
<td>GST-tag</td>
<td>MSPILGYWKI KGLVQPTRTLLEYLEEKYEEHHLYERDEGDKWRNNKKF ELGLEFPNLPIYY1GDVLKQTSMAIRY1ADKHNNLGGCPKERAEL SMLEGAVLDIYRVSIAHSDSFTKLFVDLSKPLEMLKMFDSERLC HKTLYNLDHVTDPFNYLDALDVLYMPCMDAIFPPLCVCFKRIIE AIPOQDIYKLSSKYIAWPQLGQWATFPGDHPKSDLVPR</td>
<td>26.2</td>
<td>NA</td>
</tr>
<tr>
<td>Stl (GenBank ID: AAG29617.1)</td>
<td>GSPFMHXXXXHSGMAGGAQMAELPHTYGIITKTLRKMYLKLTQSKLS ERGFSQTISNHENGNNRIVGNEIEIYKGLGIFSYY1HLRISDEFEKEKGYSP TLNDPGFKDFDKMYSYNKAYNDDIYSSY1LYDETIKLLELLKESKINVNDIDY1VLKLYQ1LSTDEKTISINETYLANTRK SSDKKEVTVIIEIEGFEYKLKLLFTNLETHNDRKKALAIEEKEKLYGKLRTKNKLYLLENHYDAIKGKPMLYLYEPDRLERHQLKIIEE KDTN</td>
<td>32.9</td>
<td>1.087</td>
</tr>
<tr>
<td>Stl-dHis</td>
<td>GSPFMHXXXXHSGMAGGAQMAELPHTYGIITKTLRKMYLKLTQSKLS ERGFSQTISNHENGNNRIVGNEIEIYKGLGIFSYY1HLRISDEFEKEKGYSP TLNDPGFKDFDKMYSYNKAYNDDIYSSY1LYDETIKLLELLKESKINVNDIDY1VLKLYQ1LSTDEKTISINETYLANTRK SSDKKEVTVIIEIEGFEYKLKLLFTNLETHNDRKKALAIEEKEKLYGKLRTKNKLYLLENHYDAIKGKPMLYLYEPDRLERHQLKIIEE KDTN</td>
<td>32.0</td>
<td>1.117</td>
</tr>
<tr>
<td>Stl-NT segment</td>
<td>GSPFMHXXXXHSGMAGGAQMAELPHTYGIITKTLRKMYLKLTQSKLS ERGFSQTISNHENGNNRIVGNEIEIYKGLGIFSYY1HLRISDEFEKEKGYSP TLNDPGFKDFDKMYSYNKAYNDDIYSSY1LYDETIKLLELLKESKINVNDIDY1VLKLYQ1LSTDEKTISINETYLANTRK SSDKKEVTVIIEIEGFEYKLKLLFTNLETHNDRKKALAIEEKEKLYGKLRTKNKLYLLENHYDAIKGKPMLYLYEPDRLERHQLKIIEE KDTN</td>
<td>11.0</td>
<td>0.678</td>
</tr>
<tr>
<td>Stl-CTD</td>
<td>GSPFMHXXXXHSGMAGGAQMAELPHTYGIITKTLRKMYLKLTQSKLS ERGFSQTISNHENGNNRIVGNEIEIYKGLGIFSYY1HLRISDEFEKEKGYSP TLNDPGFKDFDKMYSYNKAYNDDIYSSY1LYDETIKLLELLKESKINVNDIDY1VLKLYQ1LSTDEKTISINETYLANTRK SSDKKEVTVIIEIEGFEYKLKLLFTNLETHNDRKKALAIEEKEKLYGKLRTKNKLYLLENHYDAIKGKPMLYLYEPDRLERHQLKIIEE KDTN</td>
<td>22.4</td>
<td>1.261</td>
</tr>
<tr>
<td>Stl-AA</td>
<td>GSPFMHXXXXHSGMAGGAQMAELPHTYGIITKTLRKMYLKLTQSKLS ERGFSQTISNHENGNNRIVGNEIEIYKGLGIFSYY1HLRISDEFEKEKGYSP TLNDPGFKDFDKMYSYNKAYNDDIYSSY1LYDETIKLLELLKESKINVNDIDY1VLKLYQ1LSTDEKTISINETYLANTRK SSDKKEVTVIIEIEGFEYKLKLLFTNLETHNDRKKALAIEEKEKLYGKLRTKNKLYLLENHYDAIKGKPMLYLYEPDRLERHQLKIIEE KDTN</td>
<td>32.8</td>
<td>1.090</td>
</tr>
<tr>
<td>Φ11DUT (UniProt ID: Q8SDV3)</td>
<td>MTNTLQVRLLSENARMPERNHKTDAGYDIFSIAETVVLPEFQEKAVIK TDVAVSIEPVEGVLTTSRSGVSSKTHLVIETGIDAGYHNGLNIG RKNDAISNGYITPGVFDIKEG1DSAIRQVGTYQINEGDKLAQLV TVPINTPELKVEEEFESERGEGFSSGVC</td>
<td>18.4</td>
<td>0.786</td>
</tr>
<tr>
<td>Φ11DUTΔ101G-122Q</td>
<td>MTNTLQVRLLSENARMPERNHKTDAGYDIFSIAETVVLPEFQEKAVIK TDVAVSIEPVEGVLTTSRSGVSSKTHLVIETGIDAGYHNGLNIG RKNDAISNGYITPGVFDIKEG1DSAIRQVGTYQINEGDKLAQLV TVPINTPELKVEEEFESERGEGFSSGVC</td>
<td>16.0</td>
<td>0.810</td>
</tr>
<tr>
<td>MtDUT (UniProt ID: P9WNS5)</td>
<td>MGSSHHHHHHHSGVLPRHGMSTTLAIVRLDFGLFLFSRAHDGDAG VDLYSAEVIAFLAPRRAAGKVRGTAIAVAVFPGMVLHFRSGLATRVG LSIIVNSPGTIDAGYRGEIKVALINLDFAPAAIPVHGRDIAQLV VRVIELVEVSSFDAGLASTRGGGHCXSQGS</td>
<td>18.0</td>
<td>0.166</td>
</tr>
<tr>
<td>hDUT (UniProt ID: P33316-2)</td>
<td>MGSSHHHHHHHSGVLPRHGMPCSETAPPSKRRAPPAEVGGMQRFARLSEHATAPTRGASAARAGVYLDASYDYTYPMKAVEKVKTDFIQI ALPSCUGVRAPSRGKALHHTDVGAVGVEDARYNGVNCVLNFGCK EKFEVKGDRIQAELICERIFEYIEEQVALDTERGSGFGGSTKN</td>
<td>19.9</td>
<td>0.524</td>
</tr>
</tbody>
</table>

¹ Extinction coefficient (g⁻¹ dm³ cm⁻¹)
**Table A3. Sequence and properties of protein constructs** (purification tags are in italics, residues deleted or modified by mutagenesis are underlined)

<table>
<thead>
<tr>
<th>Protein</th>
<th>Szekvencia</th>
<th>MW (kDa)</th>
<th>ε&lt;sup&gt;i&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>hDUT&lt;sup&gt;ONTR&lt;/sup&gt;</td>
<td>GSMPCSEETPAISFSKRARPAAEVGGMLRDFARLSEHATAPTRGSA RAAGYLYASAYDITYPPMEKAVKTQIALPSGCYGRVAPRSLGA AKHFDIYGAVIDEDYRGNGVVLNFNGKEKFEVKGDRIALCIE RIFYEIEQVQALDDTERSGGFGSTGKN</td>
<td>18.0</td>
<td>0.579</td>
</tr>
<tr>
<td>NM1DUT (UniProt ID: A0EWK2)</td>
<td>GSMASTNTLIDQLQELQIKEFDFDRIPTLNLDSQAYVVEFFE WFNTLETFXNNWKKKPKGKFLDQQLQELADMLAFLGSIAQGVSSEE IKEAIESSFKDETFHMNFKDKFEFQADAVVSTPQQIIIKEFIPQQQ AIVVIDIAYNLISDQLIDAYKKKRNHERQDGTAADAGKVY</td>
<td>21.0</td>
<td>0.951</td>
</tr>
</tbody>
</table>

<sup>i</sup> Extinction coefficient (g<sup>-1</sup> dm<sup>3</sup> cm<sup>-1</sup>)

---

**Table A4. Properties of the templates used by Phyre2 for the Stl homology model (157).**

<table>
<thead>
<tr>
<th>Template PDB code</th>
<th>Protein Name and Function</th>
<th>Aligned residues&lt;sup&gt;*&lt;/sup&gt;</th>
<th>Alignment Coverage&lt;sup&gt;*&lt;/sup&gt; (%)</th>
<th>Confidence&lt;sup&gt;#&lt;/sup&gt; (%)</th>
<th>Identity&lt;sup&gt;§&lt;/sup&gt; (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1E3O</td>
<td>Human transcription factor Oct-1, activator of promoters of genes for some small nuclear RNAs (snRNA), histone H2B and immunoglobulins</td>
<td>2-48</td>
<td>17&lt;sup&gt;§&lt;/sup&gt;</td>
<td>96.5</td>
<td>33&lt;sup&gt;§&lt;/sup&gt;</td>
</tr>
<tr>
<td>4YV9</td>
<td>Rgg protein, transcriptional regulator of <em>Streptococcus dysgalactiae</em></td>
<td>14-263</td>
<td>93</td>
<td>99.7</td>
<td>12</td>
</tr>
<tr>
<td>2GRM</td>
<td>Prgx, molecular switch controlling expression of conjugation and virulence genes encoded by plasmid pCF10 of <em>Enterococcus faecalis</em></td>
<td>14-261</td>
<td>92</td>
<td>99.8</td>
<td>12&lt;sup&gt;i&lt;/sup&gt;</td>
</tr>
<tr>
<td>4RYK</td>
<td>Lmo0325 protein, a putative transcriptional regulator from <em>Listeria monocytogenes</em> EGD-e</td>
<td>13-260</td>
<td>92</td>
<td>99.8</td>
<td>12</td>
</tr>
<tr>
<td>2QFC</td>
<td>PlcR, major virulence regulator of the <em>Bacillus cereus</em> group</td>
<td>12-261</td>
<td>93</td>
<td>99.7</td>
<td>13</td>
</tr>
<tr>
<td>2AXZ</td>
<td>Prgx, molecular switch controlling expression of conjugation and virulence genes encoded by plasmid pCF10 of <em>Enterococcus faecalis</em></td>
<td>14-261</td>
<td>92</td>
<td>99.7</td>
<td>13&lt;sup&gt;i&lt;/sup&gt;</td>
</tr>
<tr>
<td>2EBY</td>
<td>ybaQ, putative HTH-type transcriptional regulator from <em>E. coli</em></td>
<td>11-96</td>
<td>31</td>
<td>99.6</td>
<td>15</td>
</tr>
</tbody>
</table>

<sup>*</sup> Part of Stl sequence which is covered by alignment with the template sequence.

<sup>#</sup> Confidence represents the estimated precision, it is the probability that the match between the query sequence and a template is because of true homology (160).

<sup>§</sup> This template was included to the final model by the Phyre2 because of higher local identity.

<sup>i</sup> The difference in the sequence identity of Stl with the same type of protein occurs because the PDB entry 2GRM contains the structure of a longer protein construct and more residues are modeled in that structure, than those in 2AXZ, therefore the Phyre2 alignment differs somewhat in the two cases.
Table A5. SAXS Data collection and derived parameters for hDUT and Stl. Abbreviations: $M_r$: molecular mass; $R_g$: radius of gyration; $D_{max}$: maximal particle dimension; $V_p$: Porod volume; $V_{ex}$: Particle excluded volume (158).

<table>
<thead>
<tr>
<th>Data collection parameters</th>
<th>hDUT</th>
<th>Stl</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instrument</td>
<td>EMBL P12 beam line (PETRA-III, DESY, Hamburg)</td>
<td></td>
</tr>
<tr>
<td>Beam geometry</td>
<td>0.2 x 0.12 mm²</td>
<td></td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>1.24</td>
<td></td>
</tr>
<tr>
<td>$s$ range (Å⁻¹)</td>
<td>0.01-0.46</td>
<td></td>
</tr>
<tr>
<td>Exposure time (s)</td>
<td>1 (20×0.05 s)</td>
<td></td>
</tr>
<tr>
<td>Concentration range (mg/mL)</td>
<td>0.2-2.0</td>
<td>0.2-0.9</td>
</tr>
<tr>
<td>Temperature (K)</td>
<td>283</td>
<td>283</td>
</tr>
</tbody>
</table>

**Structural parameters**

- $I_0$ (cm⁻¹) [from $p(r)$]:
  - hDUT: 0.025 ± 0.001
  - Stl: 0.032 ± 0.001
- $R_g$ (Å):
  - hDUT: 29 ± 1
  - Stl: 32 ± 1
- $I_0$ (cm⁻¹) (from Guinier):
  - hDUT: 0.025 ± 0.001
  - Stl: 0.031 ± 0.001
- $R_g$ (Å) (from Guinier):
  - hDUT: 30 ± 1
  - Stl: 33 ± 1
- $D_{max}$ (Å):
  - hDUT: 100
  - Stl: 105
- Porod volume estimate (Å³):
  - hDUT: 117000 ± 10000
  - Stl: 100000 ± 20000
- Excluded volume estimate (Å³):
  - hDUT: 132000 ± 20000
  - Stl: 90000 ± 20000
- Dry volume calculated from:
  - sequence (Å³):
    - hDUT: 21815
    - Stl: 38738

**Molecular-mass determination**

- $I(0)$ (cm⁻¹) BSA (70,000 Da):
  - hDUT: 0.050 ± 0.001
- Molecular mass $M_r$ (Da): [from $I(0)$]
  - hDUT: 35000 ± 4000
  - Stl: 45000 ± 5000
- Molecular mass $M_r$ (Da): [from Porod volume ($V_p/1.6$)]
  - hDUT: 73000 ± 8000
  - Stl: 59000 ± 6000
- Molecular mass $M_r$ (Da): [from excluded volume ($V_{ex}/2$)]
  - hDUT: 66500 ± 10000
  - Stl: 45000 ± 5000
- Calculated monomeric $M_r$ from sequence (Da):
  - hDUT: 18029
  - Stl: 32015

**Software employed**

- Primary data reduction: RADAVER
- Data processing: PRIMUS/Qt
- Ab initio analysis: DAMMIF
- Validation and averaging: DAMAVER
- Hybrid modeling: CORAL, EOM
- Computation of model intensities: CRYSOL
- 3D graphics representations: PyMOL, UCSF Chimera

---

*a* Momentum transfer $s = 4\pi\sin(\theta)/\lambda$.

*b* Values reported for merged data sets (0.2 & 2.0, and 0.2 & 0.9 mg/mL⁻¹ for hDUT and Stl, respectively).

*c* Dry volume determined using the server:

http://www.basic.northwestern.edu/biotools/proteincalc.html
### Table A6. SAXS Data collection and derived parameters for hDUT:Stl SEC-SAXS.

Abbreviations: \( M_r \): molecular mass; \( R_g \): radius of gyration; \( D_{max} \): maximal particle dimension; \( V_p \): Porod volume; \( V_{ex} \): Particle excluded volume (158).

<table>
<thead>
<tr>
<th></th>
<th>hDUT:Stl (Region 1)</th>
<th>hDUT:Stl (Region 2)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data collection parameters</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instrument</td>
<td>EMBL P12 beam line (PETRA-III, DESY, Hamburg)</td>
<td></td>
</tr>
<tr>
<td>Beam geometry</td>
<td>0.2 x 0.12 mm²</td>
<td></td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>1.24</td>
<td></td>
</tr>
<tr>
<td>( s ) range (Å⁻¹) ( a )</td>
<td>0.01-0.46</td>
<td></td>
</tr>
<tr>
<td>Exposure time (s)</td>
<td>3600 (3600×1.0 s)</td>
<td></td>
</tr>
<tr>
<td>Concentration range (mg/mL)</td>
<td>~ 0.4</td>
<td>~ 0.4</td>
</tr>
<tr>
<td>Temperature (K)</td>
<td>283</td>
<td>283</td>
</tr>
<tr>
<td><strong>Structural parameters( b )</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( I_0 (cm⁻¹) ) [from ( p(r) )]</td>
<td>0.041 ± 0.001</td>
<td>0.047 ± 0.001</td>
</tr>
<tr>
<td>( R_g (Å) ) [from ( p(r) )]</td>
<td>45 ± 1</td>
<td>39 ± 1</td>
</tr>
<tr>
<td>( I_0 (cm⁻¹) ) (from Guinier)</td>
<td>0.041 ± 0.001</td>
<td>0.046 ± 0.001</td>
</tr>
<tr>
<td>( R_g (Å) ) (from Guinier)</td>
<td>44 ± 1</td>
<td>38 ± 1</td>
</tr>
<tr>
<td>( D_{max} ) (Å)</td>
<td>170</td>
<td>140</td>
</tr>
<tr>
<td>Porod volume estimate (Å³)</td>
<td>227000 ± 10000</td>
<td>190000 ± 20000</td>
</tr>
<tr>
<td>Excluded volume estimate (Å³)</td>
<td>290000 ± 2000</td>
<td>242000 ± 20000</td>
</tr>
<tr>
<td>Dry volume calculated from sequence (Å³)( c )</td>
<td>188384/149667 (3:3/3:2 complex)</td>
<td></td>
</tr>
<tr>
<td><strong>Molecular-mass determination</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( I_0 (cm⁻¹) ) BSA (70,000 Da)</td>
<td>0.050 ± 0.001</td>
<td></td>
</tr>
<tr>
<td>Molecular mass ( M_r ) (Da) [from ( DATMOW )]</td>
<td>130000 ± 4000</td>
<td>124000 ± 5000</td>
</tr>
<tr>
<td>Molecular mass ( M_r ) (Da) [from Porod volume (( V_p/1.6 ))]</td>
<td>134000 ± 9000</td>
<td>112000 ± 6000</td>
</tr>
<tr>
<td>Molecular mass ( M_r ) (Da) [from excluded volume (( V_{ex}/2 ))]</td>
<td>145000 ± 10000</td>
<td>121000 ± 3000</td>
</tr>
<tr>
<td>Calculated monomeric ( M_r ) [from sequence (Da)]</td>
<td>155689/123692 (3:3/3:2 complex)</td>
<td></td>
</tr>
<tr>
<td><strong>Software employed</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary data reduction</td>
<td>RADAVER</td>
<td></td>
</tr>
<tr>
<td>Data processing</td>
<td>PRIMUS/Qt</td>
<td></td>
</tr>
<tr>
<td>( Ab initio ) analysis</td>
<td>DAMMIF</td>
<td></td>
</tr>
<tr>
<td>Validation and averaging</td>
<td>DAMAVER</td>
<td></td>
</tr>
<tr>
<td>Hybrid modeling</td>
<td>CORAL</td>
<td></td>
</tr>
<tr>
<td>Computation of model intensities</td>
<td>CRYSOL</td>
<td></td>
</tr>
<tr>
<td>3D graphics representations</td>
<td>PyMOL, UCSF Chimera</td>
<td></td>
</tr>
</tbody>
</table>

\( a \) Momentum transfer \( s = 4\pi \text{sin} (\theta)/\lambda \).

\( b \) Values reported for selected buffer corrected SEC-SAXS frames.

\( c \) Dry volume determined using the server:
http://www.basic.northwestern.edu/biotools/proteincalc.html
NYILATKOZAT

Alulírott **Nyíri Kinga** kijelentem, hogy ezt a doktori értekezést magam készítettem és abban csak a megadott forrásokat használtam fel. Minden olyan részt, amelyet szó szerint, vagy azonos tartalomban, de átfogalmazva más forrásból átvettem, egyértelműen, a forrás megadásával megjelöltem.

Budapest,………………………..

……………………………..

Nyíri Kinga