

Dissertation zur Erlangung des Doktorgrades  
der Fakultät für Chemie und Pharmazie  
der Ludwig-Maximilians-Universität München



---

Structure of the *Bacillus subtilis* 70S ribosome reveals  
the basis for species-specific stalling

---

Daniel Clemens Michael Sohlen

Aus München, Deutschland

2015

## Erklärung

Die Dissertation wurde im Sinne von § 7 der Promotionsordnung vom 28. November 2011 von Herrn Dr. Daniel Wilson betreut.

## Eidesstattliche Versicherung

Die Dissertation wurde eigenständig und ohne unerlaubte Hilfe erarbeitet.

München, den 22.06.2015

Daniel Clemens Michael Sohmen

Dissertation eingereicht am 25.06.2015

1. Gutachter: Herr Dr. Daniel Wilson

2. Gutachter: Herr Prof. Dr. Roland Beckmann

Mündliche Prüfung am 13.08.2015

# 1 Acknowledgements

First and foremost I am extremely grateful to Dr. Daniel Wilson, for his excellent and continuous support of my work. It has been a privilege to conduct research in your lab as a Ph.D. student. The many insightful discussions, your scientific advice and the warm encouragement I received from you (especially when things got frustrating) let me succeed.

Also, I would not have been able to conclude my research without the wise counsel of Prof. Roland Beckmann, providing an excellent scientific environment, never unwilling to share his resources, manpower or machine.

I also want to express my gratitude to all our collaborators, especially to Prof. Ito Koreaki, Dr. Chiba Shinobu, and Dr. Axel Innis. Their revealing insights and help regarding the MifM project contributed substantially to its successful completion and are very much appreciated.

In general, this thesis would not have been possible without the support of the many people I had the pleasure to meet and work with. Therefore, I wish to express my sincere appreciation to those who have contributed to this thesis and supported me in one way or the other. Especially I'd like to mention by name: Agata Starosta, Alexandra Dönhöfer, Aleksandra Mikolajka, Andrea Gilmozzi, Andreas Anger, Birgitta Beatrix, Bertrand Beckert, Caroline Haas, Charlotte Ungewickell, Christoph Leidig, Eli van der Sluis, Heidi Sieber, Ingegerd Walz, Joanna Musial, Jens Frauenfeld, Julian Deeng, Lukas Bischoff, Matthias Siebert, Otto Berninghausen, Romy Böttcher, Sarah Matheisl, Stephanie Fesser, Stephanie Helfer, Susanne Rieder, Soledad Funes, Sibylle Frankenberg, Stefan Arenz, Stephan Wickles, Thomas Becker, Viter Márquez.

I am proud of what I have achieved with your help. Thank you all.

## 2 Summary

Proteins, which are translated by the large molecular machines called ribosomes (Figure 1), are the essential basis for cellular function. Therefore, the production of proteins makes the ribosome (Figure 2) indispensable and a worthwhile target for detailed biochemical and structural research to obtain novel, insightful knowledge of one of the most important, yet most complex molecular processes within the cell.

By analysing the ribosomal proteome from different archaeal species, we identified three hypothetical proteins that appear to be bona fide ribosomal proteins (r-proteins) of the Crenarchaeota *Sulfolobus acidocaldarius* (*S. acid*) ribosomal large subunit. In general, the analysis presents the first comprehensive two-dimensional PAGE (2D-PAGE) and mass spectrometry (MS) analysis of archaeal ribosomes isolated from the thermophilic Crenarchaeota *Pyrobaculum aerophilum* (*P. aero*) and the thermoacidophilic Crenarchaeota *Sulfolobus acidocaldarius*. Further, we identified possible promiscuous interactions of the r-proteins L7ae and S24e with both ribosomal subunits as well as possible binding positions of these proteins determined by cryo electron microscopy (cryo-EM). Thereby, this survey illustrates the potential of studying the archaeal ribosome to unravel the diversity of the ribosomal proteome, helping to understand the evolution of the ribosome and the development of ribosomal protein function.

To counter the selection of resistant pathogens that are emerging because of the excessive use of antibiotics, there is a continuous need for antibiotic discovery and development. Three analogue compounds of the naturally produced thiopeptide antibiotic amythiamicin D were made available by de novo total synthesis. The characterisation of these compounds *in vitro* and *in vivo* activity revealed enhanced inhibitory activity of the hydroxymethyl derivative (compound 3a) on translation in an *Escherichia coli* (*E. coli*) based *in vitro* translation assay, compared to its parent compound. Further, we provide evidence that translation inhibition by amythiamicin D is due to binding to the bacterial elongation factor Tu (EF-Tu). Taken together, we provide tools and starting points for future studies aiming to understand the characteristics of various substitutions or alterations of the antimicrobial compounds.

The non-canonical single-stranded DNA-binding protein related proteins (Thermo-DBP-RP) found in some hyperthermophilic Crenarchaeota comprise two related groups (1 and 2) of poorly characterized proteins, of which group 1 proteins were suggested to interact with

ribonucleic particles (RNP) and the archaeal ribosome. Here, we present the biochemical investigation of the nucleic acid binding properties and successful crystallization of *Pyrococcus furiosus* (*P. fur*) ThermoDBP-RP1 and *Aeropyrum pernix* (*A. per*) ThermoDBP-RP2, revealing their molecular structure and their mode of oligomerization. The structure provided insights into how the N-terminal domain, common to ThermoDBPs and ThermoDBP-RPs, interacts with nucleic acids. Furthermore, the structures provide indications for the function of ThermoDBP-RP1 in the extremophile organisms, while the physiological relevance of a functional interaction with RNPs or the ribosome remains to be determined.

Expression of the membrane protein biogenesis factor YidC2 from *Bacillus subtilis* (*B. sub*) is regulated by ribosome translation arrest on the upstream MifM leader peptide. By using cryo-electron microscopy (cryo-EM) and single particle analysis we were able to reconstruct the 3D volume of the MifM stalled ribosome complex at 3.9Å resolution. This enabled us to build the first atomic model of a Gram-positive 70S ribosome. In addition, we show for the first time, how interaction of MifM within the ribosomal tunnel stabilizes a specific conformation of the ribosome, blocking an active conformation of the peptidyl transferase centre (PTC). Moreover, we identified a single amino acid determining the species-specificity of MifM dependent ribosome stalling in *Bacillus subtilis*.

The eukaryotic signal recognition particle (SRP) recognizes signal sequences within the nascent polypeptide of secretory and membrane proteins and transiently slows down translation by its Alu domain for efficient targeting of the ribosome-nascent-chain complexes (RNC) to a membrane translocation site. Using cryo-EM we could, for the first time, visualize the 3D structure of a bacterial SRP containing an Alu domain, providing the functional and structural basis for co-translational targeting and SRP mediated translation arrest in Gram-positive bacteria. The unique structure of the *B. sub* SRP Alu RNA shows multiple interactions with the stalk base helices 43 (H43) and 44 (H44) and the  $\alpha$ -sarcin-ricin loop (SRL). Thus, the entire *B. sub* Alu domain acts as a “dock and lock” system to slow down the translation machinery for efficient and proper targeting to a membrane translocation site.

## 3 List of Original Publications

### Publication 1

Viter Márquez, Thomas Fröhlich, Jean-Paul Armache, **Daniel Sohmen**, Alexandra Dönhöfer, Aleksandra Mikolajka, Otto Berninghausen, Michael Thomm, Roland Beckmann, Georg J. Arnold and Daniel N. Wilson (2011), Proteomic Characterization of Archaeal Ribosomes Reveals the Presence of Novel Archaeal-Specific Ribosomal Proteins, *Journal of Molecular Biology*. 405, 1215–1232

### Publication 2

Stefan Gross, Fabian Nguyen, Matthias Bierschenk, **Daniel Sohmen**, Thomas Menzel, Iris Antes, Daniel N. Wilson, and Thorsten Bach (2013), Amythiamicin D and Related Thiopeptides as Inhibitors of the Bacterial Elongation Factor EF-Tu: Modification of the Amino Acid at Carbon Atom C2 of Ring C Dramatically Influences Activity, *ChemMedChem*. 8, 1954–1962

### Publication 3

Homa Ghalei, Holger von Moeller, Detlef Eppers, **Daniel Sohmen**, Daniel N. Wilson, Bernhard Loll and Markus C. Wahl (2014), Entrapment of DNA in an intersubunit tunnel system of a single-stranded DNA-binding protein, *Nucleic Acids Research*. 42, 6698–6708

### Publication 4

**Daniel Sohmen**, Shinobu Chiba, Naomi Shimokawa-Chiba, C. Axel Innis, Otto Berninghausen, Roland Beckmann, Koreaki Ito & Daniel N. Wilson (2015), Structure of the Bacillus subtilis 70S ribosome reveals the basis for species-specific stalling, *Nature Communications*. 6, 6941

### Publication 5

Bertrand Beckert, Alexej Kedrov, **Daniel Sohmen**, Georg Kempf, Klemens Wild, Irmgard Sinning, Henning Stahlberg, Daniel N. Wilson and Roland Beckmann (2015\*), Structural basis for targeting and elongation arrest of Bacillus signal recognition particle, \*Unpublished manuscript

## 4 Contribution Report

This dissertation is based on the work I conducted during my PhD research in the laboratory of Dr. Daniel N. Wilson from 2008 to 2015 at the Gene Centre of the Ludwig-Maximilians University, Munich

### Publication 1 (Márquez et al., 2010)

This paper presents the first comprehensive 2D PAGE and mass spectrometry analysis of archaeal ribosomes isolated from the thermophilic Crenarchaeota *Pyrobaculum aerophilum* and the thermoacidophilic Crenarchaeota *Sulfolobus acidocaldarius*. Three hypothetical proteins were identified that appear to be bona fide r-proteins of the *Sulfolobus acidocaldarius* large subunit. I purified archaeal ribosomes from *Pyrobaculum aerophilum* via sucrose density centrifugation, which were used for analysis by 2D PAGE, LC-MS/MS and cryo-EM.

### Publication 2 (Gross et al., 2013)

This publication reports the de novo total synthesis of three chemically analogous compounds of the naturally produced thiopeptide antibiotic amythiamicin D and their *in vitro* and *in vivo* activity. I purified the EF-Tu proteins and conducted *in vitro* translation experiments. I contributed to writing the manuscript.

### Publication 3 (Ghalei et al., 2014)

This publication reports two novel crystal structures of representatives of ThermoDBP-related proteins from *Aeropyrum pernix* and *Pyrococcus furiosus* and their nucleic acid binding properties. I prepared the *Thermococcus kodakarensis* ribosomes and conducted the ribosome-binding assays. I prepared Figure S3 and helped in writing the manuscript.

### Publication 4 (Sohmen et al., 2015)

This publication reports the first atomic structure of a Gram-positive 70S ribosome and reveals the basis for species-specificity of MifM dependent ribosome stalling in *Bacillus subtilis*. I prepared the *Bacillus subtilis* translation extract, the MifM-stalled ribosomal complex and performed single-particle cryo-EM analysis using direct detector data collected in NeCEN (Leiden, Netherlands). I build the model and helped with the model refinement and validation. I prepared all figures and helped in writing the manuscript.

Publication 5 (Beckert et al., 2015)

This publication reports the first complete molecular model of SRP from the Gram-positive bacterium *Bacillus subtilis* based on cryo-electron microscopy proving the functional and structural basis for co-translational targeting and RNA-driven elongation arrest in prokaryotes. I prepared the ribosomal nascent chain complex (RNC) used for sample preparation and data collection in NeCEN (Leiden, Netherlands). I performed single-particle cryo-EM analysis using direct detector data collected in NeCEN (Leiden, Netherlands) and helped with model building.



## **5 Parts of this Thesis have been Presented at International Conferences:**

‘Ribosomes 2010’

03.05.2010-07.05.2010 Orvieto, Italy, poster presentation

Cold Spring Harbour Laboratory: ‘Translational Control meeting’

04.09.2012 - 08.09.2012 Long Island, USA, poster presentation

EMBO Conference Series: ‘Protein Synthesis and Translational Control’

09.09.2013 - 11.09.2013 Heidelberg, Germany, poster presentation

15th IUBMB – 24th FAOBMB, TSBMB Conference, ‘Regulation of gene expression via nascent chain mediated translational stalling’,

21.10.2014 – 26.10.2014 Taipei, Taiwan, oral presentation

## 6 Abbreviations

AAP	arginine attenuator peptide
AdoMet-DC	S-adenosylmethionine decarboxylase
A-site	aminoacyl tRNA binding site
A-tRNA	tRNA bound to the A-site
anti-SD	anti-Shine-Dalgarno sequence
ATP	adenosine triphosphate
BOM	benzyloxymethyl
CCD camera	charge-coupled device camera
cryo-EM	cryo electron microscopy
DNA	deoxyribonucleic acid
Ec	Escherichia coli
EM	electron microscopy
E-site	tRNA exit site
EF-G	elongation factor G
EF-Tu	elongation factor Tu
FRET	Förster resonance energy transfer
fMet	<i>N</i> -formylmethionine
GTP	guanosine triphosphate
GDP	guanosine diphosphate
GDPNP	guanosine-5'-( $\beta,\gamma$ -imino)- triphosphate
LC tandem mass spectrometry (MS/MS)	liquid chromatography tandem MS
MD simulation	molecular dynamics simulation
MIC	minimal inhibitory concentration
MifM	membrane insertion and folding monitor
mRNA	messenger RNA
tRNA <sub>2</sub> -mRNA	mRNA, A-tRNA and P-tRNA
MS	mass spectrometry
MDa	megadalton
NeCEN	Netherlands Centre for Electron Nanoscopy
NMD	nonsense-mediated decay
NMR	nuclear magnetic resonance spectroscopy

---

ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
P <sub>i</sub>	inorganic phosphate
P-site	peptidyl tRNA binding site
P-tRNA	tRNA bound to the P-site
PTC	peptidyl transferase centre
RAP	ribosome arrest peptide
RNA	ribonucleic acid
rRNA	ribosomal RNA
RNC	ribosome-nascent-chain complexes
RNP	ribonucleic particle
r-proteins	ribosomal proteins
rRNA	ribosomal RNA
RRF	ribosome-recycling factor
S	Svedberg unit
SAM	S-adenosyl-L-methionine
SC-peptide	sucrose control peptide
SecM	secretion monitor
SD	Shine-Dalgarno sequence
SSB	single-stranded DNA-binding protein
ssDNA	single-stranded DNA
SRC	stalled ribosome complexes
sRNA	small RNA
SRL	$\alpha$ -sarcin-ricin loop
SRP	signal recognition particle
Thermo-DBP	non-canonical single-stranded DNA-binding protein
Thermo-DBP-RP	non-canonical single-stranded DNA-binding protein related protein
tRNA	transfer RNA
TM	trans-membrane
uORF	upstream-ORF
2D/3D	two/three-dimensional
30S IC	30S initiation complex
70S IC	70S initiation complex

## 7 Table of Contents

<b>1</b>	<b>Acknowledgements</b> .....	03
<b>2</b>	<b>Summary</b> .....	04
<b>3</b>	<b>List of Original Publication</b> .....	06
<b>4</b>	<b>Contribution Report</b> .....	07
<b>5</b>	<b>Parts of this Thesis have been Presented at International Conferences</b> .....	09
<b>6</b>	<b>Abbreviations</b> .....	10
<b>7</b>	<b>Table of Contents</b> .....	12
<b>8</b>	<b>Introduction</b> .....	14
<b>8.1</b>	<b>The Universal Ribosome</b> .....	14
<b>8.2</b>	<b>Prokaryotic Translation and the Translation Cycle</b> .....	15
8.2.1	Initiation .....	17
8.2.2	Decoding .....	17
8.2.3	Peptide bond formation .....	19
8.2.4	tRNA translocation .....	20
8.2.5	Termination and recycling .....	22
<b>8.3</b>	<b>Stalling and its Biological Relevance</b> .....	24
8.3.1	Mechanisms of RAP mediated regulation of gene expression .....	25
8.3.2	Sensing metabolites & drugs .....	27
8.3.3	Monitoring physiological functions .....	28
<b>9</b>	<b>Cumulative Thesis: Summary of Published Results</b> .....	34
<b>9.1</b>	<b>Paper 1: Proteomic Characterization of Archaeal Ribosomes Reveals the Presence of Novel Archaeal-Specific Ribosomal Proteins</b> .....	34
<b>9.2</b>	<b>Paper 2: Amythiamicin D and Related Thiopeptides as Inhibitors of the Bacterial Elongation Factor EF-Tu: Modification of the Amino Acid at Carbon Atom C2 of Ring C dramatically Influences Activity</b> .....	36
<b>9.3</b>	<b>Paper 3: Entrapment of DNA in an intersubunit tunnel system of a single-stranded DNA-binding protein</b> .....	37

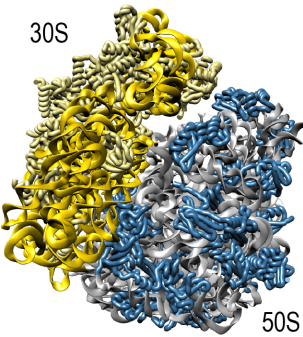
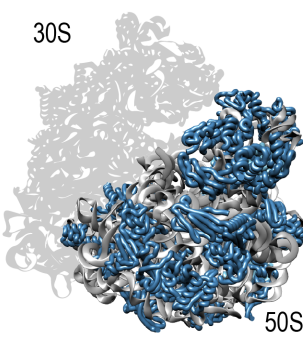
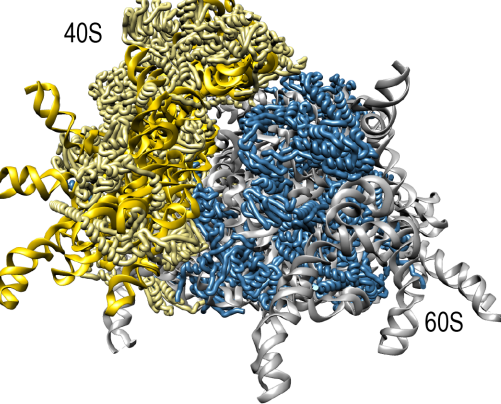
---

9.4	Paper 4: Structure of the <i>Bacillus subtilis</i> 70S ribosome reveals the basis for species-specific stalling .....	39
9.5	Paper 5: Structural basis for targeting and elongation arrest of <i>Bacillus</i> signal recognition particle .....	40
10	Discussion .....	42
10.1	Paper 1: Proteomic Characterization of Archaeal Ribosomes Reveals the Presence of Novel Archaeal-Specific Ribosomal Proteins .....	42
10.2	Paper 2: Amythiamicin D and Related Thiopeptides as Inhibitors of the Bacterial Elongation Factor EF-Tu: Modification of the Amino Acid at Carbon Atom C2 of Ring C dramatically Influences Activity .....	44
10.3	Paper 3: Entrapment of DNA in an intersubunit tunnel system of a single-stranded DNA-binding protein .....	46
10.4	Paper 4: Structure of the <i>Bacillus subtilis</i> 70S ribosome reveals the basis for species-specific stalling .....	48
10.5	Paper 5: Structural basis for targeting and elongation arrest of <i>Bacillus</i> signal recognition particle .....	52
11	References .....	55
12	Publications .....	69

## 8 Introduction

### 8.1 The Universal Ribosome

The ribosome (Figure 1) in its function of decoding genetic information and catalysing peptide bond formation is universally conserved throughout the three domains of life (Bacteria, Archaea and Eukarya). All ribosomes are composed of a small and a large subunit, both of which are built from RNA and proteins, however their composition and size can differ significantly (Melnikov et al., 2012).

Bacteria <i>E. coli</i>	Archaea <i>H. marismortui</i>	Eukarya <i>H. sapiens</i>
<b>a</b>	<b>b</b>	<b>c</b>
		
54 proteins 3 rRNAs	59 proteins 3 rRNAs	79 proteins 4 rRNAs
Large subunit (50S): 33 proteins 23S rRNA – 2904 nts 5S rRNA – 121 nts	Large subunit (50S): 34 proteins 23S rRNA – 2923 nts 5S rRNA – 122 nts	Large subunit (60S): 46 proteins 23S rRNA – 2923 nts 5.8S rRNA – 158 nts 5S rRNA – 122 nts
Small subunit (30S): 21 proteins 16S rRNA – 1542 nts	Small subunit (30S): 25 proteins 16S rRNA – 1472 nts	Small subunit (40S): 33 proteins 16S rRNA – 1800 nts

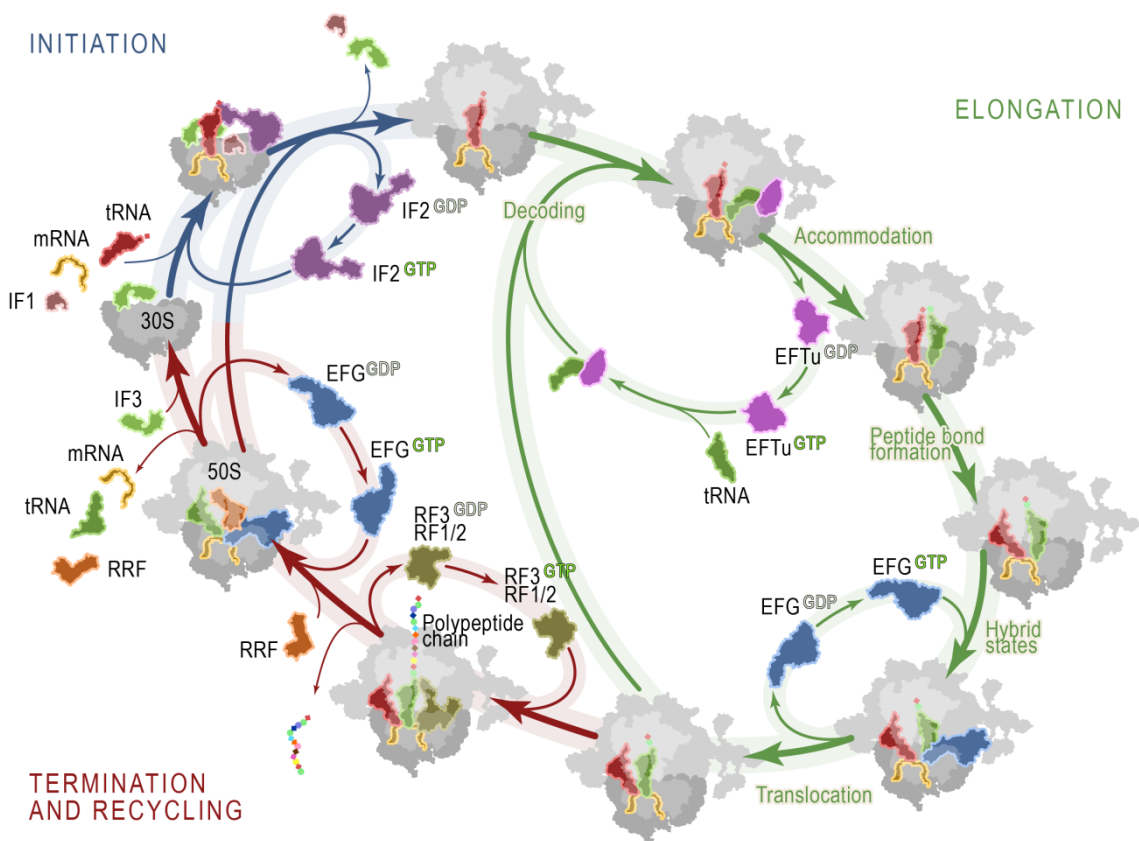
**Figure 1 | The bacterial, archaeal, and eukaryotic ribosome and its composition.** (a) Structure of the bacterial 70S ribosome from *E. coli* (Dunkle et al., 2011) with rRNA/r-proteins colored in gold/yellow and grey/blue for the small and large ribosomal subunit, respectively. (b) Structure of the archaeal large ribosomal subunit from *Haloarcula marismortui* (Gabdulkhakov, Nikonov, & Garber, 2013) with rRNA/r-proteins colored as in a. (c) Structure of the eukaryotic 80S ribosome from *Homo sapiens* (Anger et al., 2013) with rRNA/r-proteins colored as in (a). The small and the large ribosomal subunit are indicated as 30S and 50S or 40S and 60S, respectively.

The archaeal and the bacterial small and large ribosomal subunit sediment with a coefficient of 30 and 50 Svedberg units (S), respectively, and together form the complete prokaryotic ribosome with a sedimentation coefficient of 70S. The sedimentation coefficients for the eukaryotic ribosomal small and large subunit have been estimated at 40S and 60S, respectively, with an estimated sedimentation coefficient of 80S for the whole eukaryotic ribosome. These differences in sedimentation speed reflect the differences in size and mass of the particles and have led the prokaryotic and eukaryotic ribosomes to be simply referred to as 70S and 80S ribosomal particles, respectively. Compared to the bacterial 70S ribosome of *E. coli*, the human 80S ribosome contains, an additional 2MDa of mass. This extra mass is in part due to additional ribosomal RNA (rRNA) as well as due to eukaryote-specific ribosomal proteins and r-protein extensions (Jenner et al., 2012; Melnikov et al., 2012; D.N. Wilson & Cate, 2012). Although the archaeal ribosome has about the size and the mass of the bacterial ribosome, only about half of its protein families have homologous bacterial r-proteins while the other half is specific to archaea and eukarya (Lecompte, Ripp, Thierry, Moras, & Poch, 2002). These differences in size and composition reflect the different evolutionary routes towards cellular complexity and thus the need for a more complex regulation of translation as well as the need to adapt to a diverse variety of living environments. In spite of the differences, all ribosomes possess a conserved core that contains all facets necessary for the basic functions of translation (Melnikov et al., 2012).

## 8.2 Prokaryotic Translation and the Translation Cycle

Protein synthesis, in all living organisms, is carried out on ribosomes in a series of reactions describes in the translation cycle (Figure 2). This translation cycle proceeds through four stages, namely initiation, elongation, termination and recycling, within which the elongation phase can itself be described as a cyclic process (as reviewed in (Voorhees & Ramakrishnan, 2013)). During initiation the ribosome has to assemble on the start codon of the mRNA with the initiator tRNA positioned in the ribosomal P-site, ready to engage in translation elongation. During elongation, the polypeptide chain the product of translation, is produced by the recurring joining of amino acids, which occurs at the peptidyl transferase centre located on the large subunit of the ribosome. With the exception of the first initial methionine, each subsequent amino acid is delivered to the ribosome in the form of an aminoacyl-tRNA by the elongation factor Tu (EF-Tu). The aminoacyl-tRNAs are monitored by the ribosome based on the complementarity between the mRNA codon displayed in the decoding centre of the small ribosomal subunit and the anticodon of the tRNA, in a process termed decoding and proofreading. Correct decoding stimulates the GTPase activity of EF-Tu, releasing it from the

ribosome and allowing the aminoacyl-tRNA to accommodate at the A-site on the large subunit. This allows the fMet (or the nascent polypeptide chain in subsequent elongation cycles) to be transferred from the P-tRNA onto the aminoacyl tRNA in the A-site, and thereby, the growing peptide chain is prolonged by one amino acid. As the nascent chain is prolonged it traverses the large ribosomal subunit's body, through a ribosomal tunnel towards the exit on the solvent side of the ribosome. After transfer of the nascent peptide the ribosome is left with a peptidyl-tRNA in the A-site and a deacylated tRNA in the P-site.



**Figure 2 | Overview of the prokaryotic translation cycle.** The four stages of the translation cycle include: Translation initiation, elongation of the polypeptide chain, translation termination and recycling of the ribosome. Each of these steps is assisted by protein factors. During the initiation phase (blue) IF1, IF2, IF3, mRNA, initiator tRNA and the 30S subunit together form the 30S initiation complex (30S IC). Joining of the 50S subunit releases the initiation factors and renders the ribosome ready to enter the elongation step (green) of the translation cycle. During elongation aminoacyl-tRNA (aa-tRNA) is delivered to the ribosomal A-site in a ternary complex with EF-Tu•GTP. After A-tRNA accommodation, subsequent peptide bond formation is catalysed by the ribosome transferring the nascent peptide chain from the peptidyl-tRNA in the P-site onto the aa-tRNA in the A-site prolonging the nascent peptide by one amino acid. Binding of EF-G•GTP then stabilizes the tRNA hybrid state formation. Thereupon, EF-G hydrolyses GTP and catalyses tRNA<sub>2</sub>-mRNA translocation, which causes movement of the mRNA by one codon and leaves an empty A-site for the next aa-tRNA to be delivered to the ribosome. Translation termination (red) is signalled by a stop codon on the mRNA. The stop codon displayed in the A-site recruits RF1 or RF2 and RF3 releasing the fully translated protein from the ribosome. Cooperative binding of RRF and EF-G leads to splitting of the ribosome, recycling its components for the next round of translation. Figure modified from (Sohmen, Harms, Schlunzen, & Wilson, 2009)



To restore the canonical state of the ribosome for another round of elongation both tRNAs and the mRNA have to be translocated by the action of elongation factor G (EF-G) and GTP hydrolysis from the A- and P-site to the P- and E-site of the ribosome (Holtkamp, Cunha, et al., 2014; Holtkamp, Wintermeyer, & Rodnina, 2014). This leaves an empty A-site, which is ready to accept the next incoming aminoacyl-tRNA. The elongation cycle repeats until a stop codon rather than a sense codon of the mRNA enters the ribosomal A-site. The stop codon signals translation termination, whereby the completed protein gets released from the ribosome by the action of release factors. Termination is followed by factor-aided recycling of the ribosomal machinery to ready its parts to engage the next messenger RNA transcript for another round of translation.

### 8.2.1 Initiation

Formation of the bacterial 30S translation initiation complex (30S IC) requires assembling translation initiation factors 1-3 (IF1, IF2, IF3), messenger RNA and initiator tRNA fMet-tRNA<sup>fMet</sup> on the small ribosomal subunit (Julian et al., 2011; Simonetti et al., 2009). The small subunit can bind mRNA directly by recognising its Shine-Dalgarno sequence (SD; 5'AGGAGGU) (Lim, Kobayashi, & Nakai, 2014; Shine & Dalgarno, 1974) with its complementary 3' end of the 16S rRNA (anti-Shine-Dalgarno sequence (anti-SD)). The SD sequence is located upstream of the start codon, usually spaced by a distance of 7-10 nucleotides from the first position of the start codon, thereby positioning the start codon in the P-site of the 30S subunit.

IF3, IF2 and IF1 bind, independent of the recruitment of the mRNA, to the 30S ribosomal subunit. IF2 promotes binding of fMet-tRNA<sup>fMet</sup> (Milon et al., 2010) and in concert with IF1 and IF3 kinetically control and the fidelity of fMet-tRNA<sup>fMet</sup> during 30S IC formation (Milon, Maracci, Filonava, Gualerzi, & Rodnina, 2012). To form the 70S IC, IF3 has to dissociate to allow docking of the large ribosomal subunit (50S) to the 30S IC (Allen, Zavialov, Gursky, Ehrenberg, & Frank, 2005; Grigoriadou, Marzi, Pan, Gualerzi, & Cooperman, 2007). Docking of the 50S subunit then triggers GTP hydrolysis by IF2 (H. Qin, Grigoriadou, & Cooperman, 2009), making the choice of the start site irreversible (Grigoriadou, Marzi, Kirillov, Gualerzi, & Cooperman, 2007) and subsequent release of IF2 and IF1 from the 70S IC renders the complex ready to enter the translation elongation cycle.

### 8.2.2 Decoding

During translation, the correct or cognate aminoacyl-tRNA (in complex with EF-Tu•GTP) has to be chosen or discriminated from many incorrect or near-/non-cognate aminoacyl-tRNAs by the ribosome (Zaher & Green, 2009a). This process is based on the tRNA anticodon's ability to

base-pair with the mRNA codon presented in the A-site (Demeshkina, Jenner, Westhof, Yusupov, & Yusupova, 2013; Voorhees & Ramakrishnan, 2013). In the first step of this multistep decision, codon-sampling, sampling of the codon by the tRNA is sensed by the ribosome via the bases A1492, A1493 and G530 of the 16S rRNA in the decoding centre on the small subunit. Thereby, A1492 and A1493 probe the codon anticodon interaction via A-minor interactions with the helix formed by the anticodon of the tRNA and the mRNA (Carter et al., 2000; Wimberly et al., 2000). While non-cognate tRNAs will most likely be rejected at this step, recognition of a cognate tRNA being bound triggers an overall conformational change of the 30S subunit (domain closure) (Demeshkina, Jenner, Westhof, Yusupov, & Yusupova, 2012; Ogle, Murphy, Tarry, & Ramakrishnan, 2002) of the ribosome. Interestingly, it has recently been shown in a crystal structure that also a near-cognate tRNA can induce the closed conformation of the 30S subunit (Demeshkina et al., 2012, 2013) and that A1492 and A1493 not only probe the codon anticodon helix but rather restrict the geometry of the base pairing as seen in the case for G•U base pair at the first codon position adopting Watson-Crick geometry rather than the G•U wobble pair geometry. This finding thereby changes our current understanding of the decoding mechanism at the codon-sampling step. It suggests that the formation of a highly restrictive decoding centre destabilizes incorrect base pairing, finally leading to the release of near-cognate tRNAs, rather than the decoding centre promoting the conformational change on the 30S ribosomal subunit only when the formation of Watson–Crick base pairs at the first two positions of the codon–anticodon helix have been detected.

Domain closure of the 30S subunit leads to a distortion of the tRNA body and brings the shoulder region of the small subunit closer to EF-Tu, inducing conformational changes in EF-Tu that finally lead to the activation of the GTPase activity of EF-Tu. The induced interaction with the  $\alpha$ -sarcin-ricin loop (SRL) of the 23S rRNA opens the so-called “hydrophobic gate” and stabilizes the active conformation of a highly conserved histidine which is coordinating a water molecule in the active site, positioned to catalyse hydrolysis of the  $\gamma$ -phosphate group of the bound GTP (Schmeing et al., 2009; Voorhees, Schmeing, Kelley, & Ramakrishnan, 2010). Conversion of GTP to GDP by EF-Tu is associated with conformational changes in EF-Tu that results in lower affinity for the ribosome and therefore dissociation of EF-Tu from the ribosome following release of the  $\gamma$ -phosphate group ( $P_i$ ) of GTP.

After dissociation of EF-Tu, the ribosome again has the chance to reject near- or non-cognate tRNAs based on the properties of the codon anticodon interaction in a second discrimination step, proofreading (Blanchard, Gonzalez, Kim, Chu, & Puglisi, 2004; Gromadski & Rodnina, 2004; Hopfield, 1974; Ninio, 1975), before the tRNA can accommodate. For accommodation the tRNA’s aminoacylated acceptor arm swings into the peptidyl transferase centre, followed by

rapid peptide bond formation and transfer of the nascent peptide chain from the P- to the A-site tRNA and even after incorporation of the aminoacyl tRNA into the growing peptide it seems that the ribosome has means to further iterate probing the identity of the tRNA post peptide bond formation, with a quality control mechanism that leads to release of unwanted peptidyl-tRNAs via sense-codon directed release factor dependent translation termination (Zaher & Green, 2009b). Taken together these tRNA discriminating steps allow translation to proceed with a misincorporation rate of 1 amino acid in  $10^3$ - $10^4$  (Kramer & Farabaugh, 2007).

### 8.2.3 Peptide bond formation

Peptide bond formation is catalysed at the peptidyl transferase center (PTC), located on the large subunit of the ribosome. The PTC consists almost exclusively of RNA, with the N-terminus of L27 being the closest protein moiety to the reactive center by a distance of 8-10Å. Therefore, the ribosome is regarded as a ribozyme, with the RNA being the major contributor to peptide bond catalysis (Nissen, Hansen, Ban, Moore, & Steitz, 2000). Ribosomal RNA creates the environment for the nucleophilic attack of the  $\alpha$ -amino group of the aminoacyl tRNA in the A-site on the carbonyl carbon of the aminoacyl ester of the peptidyl tRNA in the P-site that leads to transfer of the nascent polypeptide bound to the tRNA in the P-site to the aminoacyl tRNA located in the A-site, adding one amino acid to the peptide chain. The ribosome enhances the rate of peptide bond formation by at least six orders of magnitude, utilising an induced-fit mechanism in which substrates and active site residues are repositioned to facilitate the peptidyl transferase reaction (Sievers, Beringer, Rodnina, & Wolfenden, 2004). Proper accommodation of an aminoacyl-tRNA in the A-site induces specific movements of 23S rRNA nucleotides (Ec) 2583–2585 and 2506, thereby reorienting the ester group of the peptidyl tRNA rendering it accessible for a nucleophilic attack (Schmeing, Huang, Strobel, & Steitz, 2005). In the absence of an appropriate A-site substrate, the ester link of the peptidyl tRNA is protected by the conformation of the peptidyl transferase center precluding a nucleophilic attack by a water molecule (Schmeing, Huang, Strobel, et al., 2005). Residues A2451 (Muth, Ortoleva-Donnelly, & Strobel, 2000; Nissen et al., 2000) and A76 of P-tRNA have been implicated to be major contributors to peptide bond catalysis (Koch, Huang, & Sprinzl, 2008). Removal of the 2' OH group of A2451 (Erlacher et al., 2005) or replacement of the 2' OH of A76 by 2' H or 2' F showed a significant decrease in catalysis rate but did not abolish catalysis (Zaher, Shaw, Strobel, & Green, 2011). Therefore, the central importance of these residues in performing the catalysis has to be questioned. Due to its intrinsically high  $pK_a$  and a missing corresponding  $pK_a$  in the reaction or cofactors, like  $Mg^{2+}$ -ions, that could lower its  $pK_a$ , it was reasoned that A76 is unlikely to be acting as a general base in an acid-base catalysis mechanism. Instead it has been

proposed that the 2' hydroxyl group of A76 plays an important role in positioning the substrates for catalysis (Schmeing, Huang, Strobel, et al., 2005), in agreement with the ribosome primarily enhancing peptide bond formation rate by acting as an entropy trap (Sievers et al., 2004). Moreover, the 2' hydroxyl group of A76 could be part of a proton shuttle mechanism (Dorner, Panuschka, Schmid, & Barta, 2003; Polikanov, Steitz, & Innis, 2014; Schmeing, Huang, Kitchen, Strobel, & Steitz, 2005) where at least three protons have to be moved for the peptidyl transfer reaction (Kuhlenkoetter, Wintermeyer, & Rodnina, 2011). High-resolution structures of the ribosome with different substrate and transition state analogs (Hansen, Schmeing, Moore, & Steitz, 2002; Polikanov et al., 2014; Schmeing, Huang, Kitchen, et al., 2005; Schmeing, Huang, Strobel, et al., 2005; Schmeing et al., 2002) have been solved and have provided insight to the peptidyl transfer reaction, but although these structures reveal a network of potentially important interactions, the underlying mechanism of peptide bond catalysis is still not fully understood (Erlacher et al., 2005; Polikanov et al., 2014; Schmeing, Huang, Kitchen, et al., 2005; Trobro & Aqvist, 2006; Wallin & Aqvist, 2010).

#### 8.2.4 tRNA translocation

After peptide bond formation the ribosome is left with a peptidyl tRNA in the A-site and a deacetylated tRNA in the P-site. To initiate a new elongation cycle both tRNAs have to be translocated from the A- and P- to the P- and E-site of the ribosome, and the mRNA has to be moved by one codon. This translocation reaction is mainly driven by large-scale conformational changes and movements of the ribosome, catalysed by elongation factor G (EF-G) (reviewed in (Achenbach & Nierhaus, 2014)). In the PRE translocation state, with tRNAs in the A- and P-site, the ribosomal subunits can spontaneously rotate with respect to one another by 7° and form the rotated PRE state, also referred to as ratcheted state (J. Frank & Agrawal, 2000; Julian et al., 2008; Valle et al., 2003). This counter-clockwise rotation of the subunits to each other facilitates movement of the tRNAs on the large subunit from the A- to the P- and from the P- to the E-position, into hybrid position 1 termed A/P and P/E, respectively, with the first letter of the term referring to the position of the tRNA on the small subunit and the latter referring to the position of the tRNA on the large subunit (Moazed & Noller, 1989). Munro *et al.* (Munro, Altman, O'Connor, & Blanchard, 2007), using single-molecule fluorescence resonance energy transfer (smFRET) measurements identified a second, short lived, hybrid state 2, where only the P-site tRNA is shifted to the P/E position while the A-site tRNA adopts the classical A/A conformation. The rotated PRE state is not stable and the ribosome fluctuates between the classical and the rotated PRE state. While tRNAs on the large subunit will translocate dependent on intersubunit rotation a second movement of the head of the small subunit has to be induced

by the action of EF-G to facilitate translocation of mRNA and tRNAs (tRNA<sub>2</sub>-mRNA) on the small subunit (J. Frank, Jr. & Gonzalez, 2010; Rodnina, Savelsbergh, Katunin, & Wintermeyer, 1997). Bound to the ribosomal complex EF-G contacts h34 of the head of the small subunit and the minor groove of the codon-anticodon duplex in the decoding center on the small subunit, via two highly conserved loops on the tip of domain IV (Ramrath et al., 2013). These interactions lead to the disruption of a network of hydrogen bonds between the decoding centre and the codon-anticodon duplex, inducing a rotation of the head (also called swivelling of the head) of the 30S subunit, turning it by 18° toward the E-site, thereby facilitating translocation of the tRNA<sub>2</sub>-mRNA module (Ermolenko & Noller, 2011; Holtkamp, Wintermeyer, et al., 2014; G. Liu et al., 2014; Salsi, Farah, Dann, & Ermolenko, 2014). This translocation step has been shown to involve an intra-subunit hybrid state, where the mRNA and the anticodon stem loop (ASL) of the tRNAs simultaneously interact with components of the A- and P-site (ap/P) and the P- and E-site (pe/E) on the small subunit (Ramrath et al., 2013; Ratje et al., 2010).

A more recent study by crystallography visualised a previously unknown, compact conformation of EF-G bound to the PRE state of the ribosome in the presence of the antibiotic dityromycin (Lin, Gagnon, Bulkley, & Steitz, 2015). In this conformation, domains III-V of EF-G are rotated away from the A-site on the small subunit (Lin et al., 2015). Additional FRET studies provide further evidence that the compact form of ribosome bound EF-G might resemble an authentic, short lived translocation intermediate preceding the extended conformation that EF-G adopts shortly after its interaction with the ribosome, but still previous to GTP hydrolysis and tRNA<sub>2</sub>-mRNA translocation (Salsi, Farah, Netter, Dann, & Ermolenko, 2015).

The action of EF-G results in a ribosomal complex after translocation (POST-state) containing peptidyl-tRNA in the P site, deacylated tRNA in the E site, and a vacant A site (Lin et al., 2015). Interestingly, the ribosome, as a Brownian machine, has the intrinsic capability to spontaneously translocate tRNAs both ways, forward and backward (Gavrilova & Spirin, 1971). Nevertheless, EF-G is needed for contributing to the efficiency of the translocation process. Moreover, a translational GTPase (EF4/LepA) that is structurally related to EF-G has been identified to catalyse a backwards movement of the tRNA<sub>2</sub>-mRNA module when the ribosome stalls, which is termed back-translocation (Y. Qin et al., 2006; Shoji, Walker, & Fredrick, 2006; Yamamoto et al., 2014).

How exactly EF-G binds to the ribosome and how the EF-G/ribosome interaction results in GTP hydrolysis and translocation are questions that are still not fully answered (Achenbach & Nierhaus, 2014; Yamamoto et al., 2014). The number of recent high-resolution structures of the ribosome in complex with EF-G though (Brilot, Korostelev, Ermolenko, & Grigorieff, 2013; Chen, Feng, Kumar, Ero, & Gao, 2013; Lin et al., 2015; Pulk & Cate, 2013; Tourigny,

Fernandez, Kelley, & Ramakrishnan, 2013) provide additional insights into the proposed mechanism of catalysis. The interaction with the  $\alpha$ -sarcin-ricin loop (SRL) of the 23S rRNA stabilizes the conformation of a highly conserved histidine coordinating a water molecule that is positioned to catalyse hydrolysis of the  $\gamma$ -phosphate group of the bound GTP. This scenario is proposed to be crucial for the activation of the conserved histidine, in a similar manner as previously observed for activation of EF-Tu (Voorhees et al., 2010).

### 8.2.5 Termination and recycling

In most species there are three stop codons, namely UAA, UAG and UGA. These are not recognised by an aminoacyl-tRNA but signal translation termination instead (reviewed in (Dunkle & Cate, 2010)). In eubacteria, translation termination is mediated by two class I release factor proteins, RF1 and RF2, which recognise the stop codons UAA and UAG or UAA and UGA, respectively. RF1 and RF2 are homologous in sequence and three-dimensional structure (Vestergaard et al., 2001) and bind to the ribosome in a way so that domains 2-4 would overlap with a tRNA bound to the ribosomal A-site (A. Korostelev et al., 2008; Laurberg et al., 2008). Thereby, domain 2 of the release factor is positioned to reach to the decoding center on the 30S subunit of the ribosome for stop codon recognition. Domain 2 contains highly conserved protein motifs for stop codon recognition that's bases conformation is strikingly different from a sense codon (A. Korostelev, Zhu, Asahara, & Noller, 2010; A. A. Korostelev, 2011) (Laurberg et al., 2008; A Weixlbaumer et al., 2008). The first base of the stop codon, universally restricted to uridine, is recognized via backbone interactions with the GXXE peptide-motive located on the N-terminal part of helix- $\alpha$ 5, while the identity of the second nucleobase is discriminated via a conserved tripeptide sequence, either PA/VT or SPF for RF1 or RF2, respectively, in which, for a great part, the identity of the Thr and the Ser, within the context of the recognition loop define the specificity (Ito, Uno, & Nakamura, 2000; Young, Edgar, Poole, & Tate, 2010). The third position nucleobase is recognised by a Thr and a Gln in RF1 that can accept a hydrogen-bond from the adenine or donate a hydrogen-bond to the guanine while RF2 has Val instead of Gln that's hydrophobic side chain would be positioned to exclude a needed water molecule from forming an hydrogen-bond with a guanine and therefore discriminates against guanine at this position (A. Korostelev et al., 2008; A. Korostelev et al., 2010; Laurberg et al., 2008; A Weixlbaumer et al., 2008). To catalyse hydrolysis of the ester bond between the terminal A76 of the P-site tRNA and the nascent peptide, the universally conserved GGQ motive of domain 3 is positioned to reach into the PTC, placing Gln230 adjacent to A76. Although, mutations of Gly228 or Gly229 show a drastic decrease in catalytic activity does mutation of Gln230 to any amino acid but proline only show a mild effect on peptide release activity (A. Korostelev et al.,

2008). In line with this observation it was found that the NH group of the protein backbone is within hydrogen bonding distance to the 3' hydroxyl of the ribose of A76 while the side chain of Gln230 is oriented away from the peptidyl-tRNA's ester bond. Therefore it has been proposed that the backbone NH group of Gln230 takes part in stabilizing the tetrahedral transition-state intermediate by establishing an hydrogen-bond to the developing oxyanion and in product stabilization by hydrogen-bonding to the 3'-OH of the deacylated tRNA (A. Korostelev et al., 2008; A. Korostelev et al., 2010; Laurberg et al., 2008; Santos, Zhu, Donohue, Korostelev, & Noller, 2013).

After peptidyl-tRNA hydrolysis, the bacterial class II release factor RF3 stimulates dissociation of RF1 and RF2 from the ribosome (Koutmou, McDonald, Brunelle, & Green, 2014; F. Peske, Kuhlenkoetter, Rodnina, & Wintermeyer, 2014). RF3 binds at the entry of the factor-binding site of the ribosome, similar to EF-G or EF-Tu and contacts the highly conserved SRL of the 23S rRNA, L6, h5 and h15 of the 16S rRNA as well as S12. Like for RF1 and RF2, the solution structure of RF3 differs from its structure bound to the ribosome. Domain II is rotated by 7° and domain III by a whole 55° relative to domain I in solution (Jin, Kelley, & Ramakrishnan, 2011b; J. Zhou, Lancaster, Trakhanov, & Noller, 2012). RF3 is a GTPase and the GTP form of the RF3 stabilizes the rotated state of the ribosome inducing tRNA hybrid conformation (Jin, Kelley, & Ramakrishnan, 2011a). GTP hydrolysis by RF3 promotes the release of the class I RFs as well as itself from the ribosome (H. Gao et al., 2007; Koutmou et al., 2014). Interestingly RF3 has been found to be non-essential in *E. coli* (Grentzmann, Brechemierbaey, Heurgue, Mora, & Buckingham, 1994; Mikuni et al., 1994) and is absent from many groups of bacteria as for *T. thermophilus* (Y. Zhou, Asahara, Gaucher, & Chong, 2012).

The final step in the translation cycle is ribosome recycling. In prokaryotes, ribosome recycling is performed by the ribosome-recycling factor (RRF) in concert with EF-G (F Peske, Rodnina, & Wintermeyer, 2005; Zavialov, Hauryliuk, & Ehrenberg, 2005). Their actions result in splitting of the ribosome into its subunits along with release of the mRNA and deacylated tRNA to make the involved components available for the next round of translation (Hirokawa et al., 2005; Karimi, Pavlov, Buckingham, & Ehrenberg, 1999).

The interaction of RRF with the ribosome has been studied by using a variety of biochemical and structure biological methods, such as chemical probing (Lancaster, Kiel, Kaji, & Noller, 2002), cryo-EM (Agrawal et al., 2004; Gao, Zavialov, Ehrenberg, & Frank, 2007; Gao et al., 2005; Yokoyama et al., 2012) and protein crystallography (Borovinskaya et al., 2007; Dunkle et al., 2011; L. Wang et al., 2012; A. Weixlbaumer et al., 2007; D. N. Wilson et al., 2005). All together, the conducted experiments data analyses suggest a predominate interaction of RRF with the ribosome, with domain I of RRF bound to the ribosomal intersubunit space, interacting

with parts of the ribosomal A- and P-site and with domain II being located in between ribosomal protein S12 and the L11 stalk-base. This interaction stabilizes the ratcheted conformation of the ribosome with the deacylated tRNA in the P/E hybrid state. Subsequent binding of EF-G•GTP leads to splitting of the ribosome by the concerted action of RRF and EF-G, hydrolysing GTP. The proximate interplay with IF3 prohibits re-association of the deacetylated tRNA as well as 70S formation (Hirokawa et al., 2005; Karimi et al., 1999; Singh, Das, Seshadri, Sangeetha, & Varshney, 2005) integrating the step of ribosome recycling into the guided assembly of the 30S translation initiation complex.

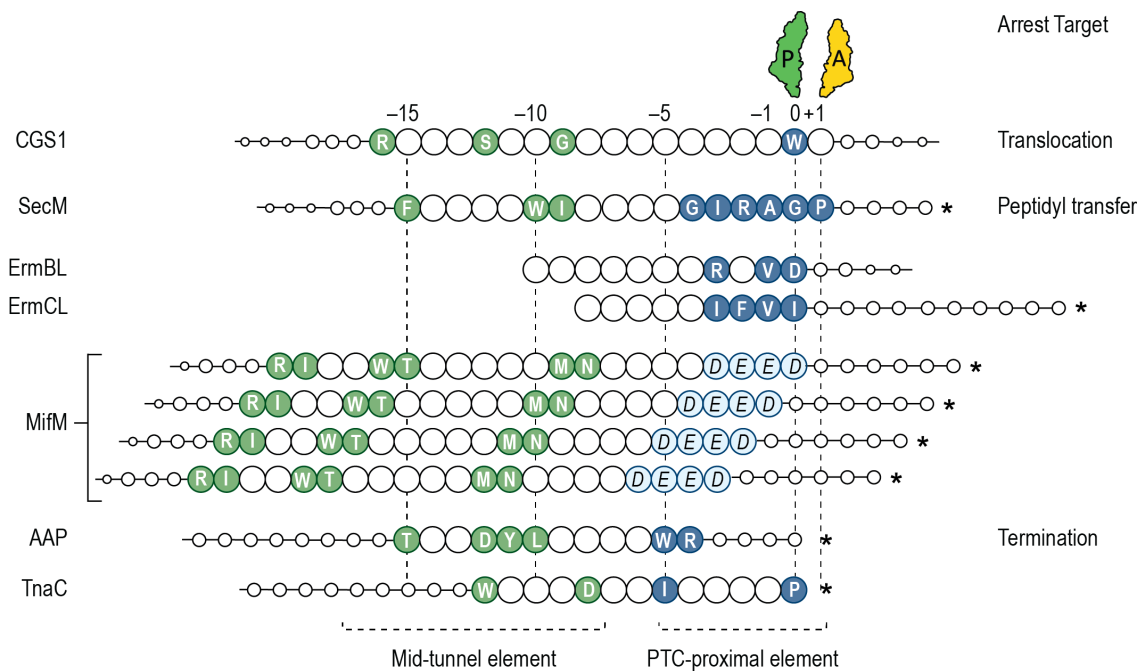
### 8.3 Stalling and its Biological Relevance

The growing product of the ribosome, the nascent peptide, can act as a cis-specific factor from within the ribosomal tunnel to modulate translation, leading to translation slowdown and arrest. Translation arrest is thereby mediated by the specific amino acid sequence of the nascent peptide. These nascent peptides, so-called ribosome arrest peptides (RAP) differ in size and amino acid composition and can be found in a variety of organisms. Each RAP interacts with the ribosome in a unique way and some require for their function the presence of an effector molecule. Nascent peptide mediated translation arrest plays an important role in the regulation of gene expression, and many open reading frames containing arrest motifs have been identified due to their function. They can act as genetic switches, in response to (i) defects in protein translocation and membrane protein biogenesis (Chiba, Lamsa, & Pogliano, 2009; McNicholas, Salavati, & Oliver, 1997; Sarker, Rudd, & Oliver, 2000), (ii) changes in amino acid metabolism (Gollnick & Yanofsky, 1990; Z. Wang, Fang, & Sachs, 1998), (iii) the presence of antibiotic (reviewed in (Ramu, Mankin, & Vazquez-Laslop, 2009)) as well as (iv) other functions of the cell (reviewed in (Ito & Chiba, 2013)).

In recent years significant progress has been made to identify determinants of nascent peptide mediated translation arrest. Amino acid sequences of several ribosomal arrest peptides, critical for stalling have been characterized (Figure 3) (Gong & Yanofsky, 2002; Ito, Chiba, & Pogliano, 2010; Mayford & Weisblum, 1989; Vazquez-Laslop, Thum, & Mankin, 2008) and a number of rRNA and protein residues within the PTC and the ribosomal exit tunnel, involved in ribosome arrest have been identified (Nakatogawa & Ito, 2002; D.N. Wilson & Beckmann, 2011). However, stalling sequences are quite divergent and interact with the ribosome in distinct ways utilising different strategies to render the PTC inactive (Ito & Chiba, 2013). The first cryo-EM structures of stalled ribosome complexes (SRC) at resolutions below 6Å gave insights into the geometric disposition of arrest-essential residues, disclosing the paths of the RAPs from the PTC through the ribosomal exit tunnel and contacts with tunnel wall components (Bhushan et



al., 2011; Bhushan et al., 2010; Seidelt et al., 2009). And more recently published structures of stalled ribosome complexes at resolutions below 4Å, allowing molecular model building and interpretation, now aim to establish a basic framework for understanding mechanistically how RAPs induce translation arrest (Arenz, Meydan, et al., 2014; Arenz, Ramu, et al., 2014; Bischoff, Berninghausen, & Beckmann, 2014).

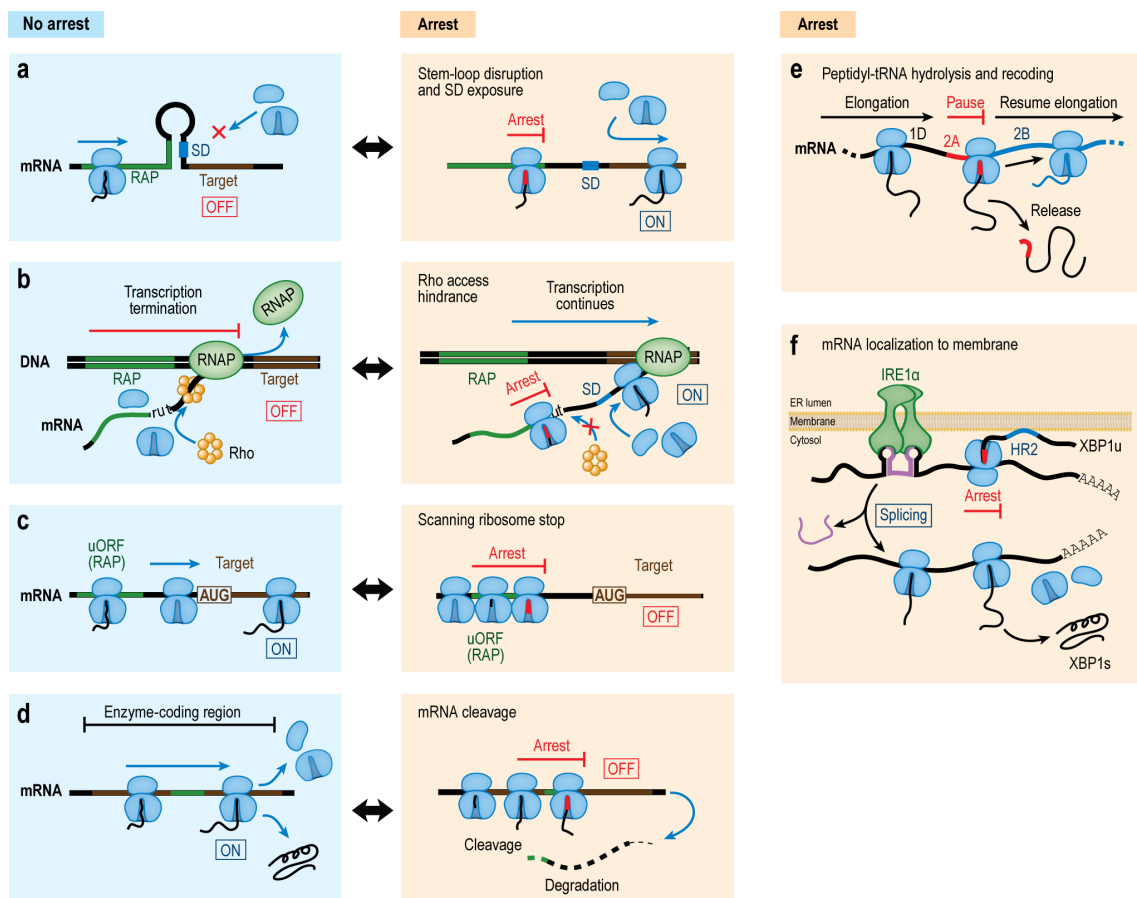


**Figure 3 | Stalling sequences of ribosome arrest peptides (RAP).** Relative position of arrest-critical and arrest-important amino acids identified by comprehensive mutational analysis. The arrest sequence segments are illustrated as circles that represent amino acid residues. Sequences are aligned on the amino acid residue in the ribosomal P-site (P-site position is defined as position 0) in the stalled ribosomal complex. Residues critical for ribosome stalling are highlighted in green (Mid-tunnel elements) and blue (PTC-proximal elements), although some arbitrariness for the boundaries could not be avoided. The arrest-important acidic cluster in MifM is highlighted in light blue with italic letters. Abbreviations: P, P-site peptidyl-tRNA; A, A-site aminoacyl-tRNA, an asterisks indicates a stop codon. Figure modified from (Ito & Chiba, 2013).

### 8.3.1 Mechanisms of RAP mediated regulation of gene expression

RAP mediated translation arrest, stopping the ribosome on a specific position on the mRNA for a prolonged time, can affect translation of a downstream gene(s) (Ito & Chiba, 2013; Lovett & Rogers, 1996; Tenson & Ehrenberg, 2002). Ribosome stalling within a number of leader peptides, for example antibiotic-resistance genes (*ermCL*, *ermAll*, *catA86*, *cmlA*) (Ramu et al., 2009), the secretion monitor (*secM*) (Nakatogawa & Ito, 2001a) and the membrane insertion and folding monitor (*mifM*) (Chiba et al., 2009), leads to disruption of a hairpin structure in the mRNA intergenic region, releasing the ribosomal binding site and thereby inducing expression of the downstream gene(s) (Figure 4 a, 5).

In contrast, ribosome stalling on the TnaC leader peptide blocks access of the Rho transcription termination factor to its binding site on the mRNA, thereby allowing transcription (and translation) of the downstream genes (L. R. Cruz-Vera, Yang, & Yanofsky, 2009; Konan & Yanofsky, 1997, 2000). A mechanism similar to the regulation of the *tnaCAB* operon has been recently proposed for the regulation of the *bmrBCD* operon in *B. subtilis* (Reilman, Mars, van Dijl, & Denham, 2014) (Figure 4 b), regulating *bmrBCD* operon gene expression in response to different antibiotics.



**Figure 4 | Mechanisms of ribosome arrest peptide (RAP) mediated regulation of gene expression.** (a) RAP mediated ribosome stalling on a leader peptide. Stalling of the ribosome leads to disruption of an mRNA secondary structure, exposing the SD region of the downstream gene. (b) RAP mediated ribosome stalling hinders access of the Rho transcription termination factor to its rut binding site. Blocking Rho's access to the mRNA allows transcription of the downstream gene(s). (c) RAP mediated ribosome stalling blocks movement of scanning ribosome to a downstream translation initiation site. (d) RAP mediated ribosome stalling induces mRNA cleavage by an endonuclease. (e) An unusual termination and reinitiation "StopGo" event. A 2A peptide mediated recoding mechanism allows translation of two separate peptides (1D, 2B) by the same ribosome. (f) RAP mediated localization of the mRNA to the membrane via a determinant of a nascent chain (HR2). Localization of the mRNA to the membrane allows efficient splicing of the XBP1u mRNA. Abbreviations: ER, endoplasmic reticulum; RNAP, RNA polymerase; SD, Shine-Dalgarno sequence; uORF, upstream open reading frame. Figure modified from (Ito & Chiba, 2013).

In general, it is found that RAP mediated regulation of gene expression in prokaryotes leads to induction of gene expression by destabilizing mRNA secondary structure, while RAP mediated regulation of gene expression in eukaryotes leads to gene repression by preventing ribosomal scanning for initiation sites (D.N. Wilson & Beckmann, 2011). For example, in eukaryotes, regulation of arginine biosynthesis requires translation of the arginine attenuator peptide (AAP) (Z. Wang & M.S. Sachs, 1997; Z. Wang & M. S. Sachs, 1997). In the presence of arginine, the AAP stalls translation at the step of termination, inhibiting the ribosome preinitiation complex from scanning for downstream ORFs thereby preventing scanning preinitiation complexes to reach the downstream gene (*N. crassa (arg-2)* and *S. cerevisiae (CPAI)*). The mRNA of the AAP stalled ribosomal complex is also subject to nonsense-mediated decay (NMD) in *S. cerevisiae* (Gaba, Jacobson, & Sachs, 2005).

Similarly, polyamine-dependent stalling during translation termination of the S-adenosylmethionine decarboxylase (AdoMet-DC) upstream ORF (uORF) (Hill & Morris, 1993; Law, Raney, Heusner, & Morris, 2001; Mize, Ruan, Low, & Morris, 1998), sucrose-dependent stalling during translation of the *bZIP11* uORF2 (Sucrose control peptide) (Hummel, Rahmani, Smeekens, & Hanson, 2009; Rahmani et al., 2009; Rook, Weisbeek, & Smeekens, 1998; Wiese, Elzinga, Wobbes, & Smeekens, 2004, 2005), and effector molecule independent stalling during translation termination of the UL4 (*gp48*) uORF2 (Cytomegalovirus) (Cao & Geballe, 1996; Degnin, Schleiss, Cao, & Geballe, 1993; Schleiss, Degnin, & Geballe, 1991), inhibit scanning of ribosome preinitiation complexes for downstream ORFs (Figure 4 c).

Expression of cystathione- $\gamma$ -synthase (*cgs1*), involved in regulation of methionine biosynthesis in *A. thaliana* (Onouchi et al., 2005), depends on translation of a RAP encoded within the first exon of the *cgs1* coding sequence, stalling translation in response to free S-adenosyl-L-methionine (SAM). Translation arrest on *cgs1* not only halts ribosome progression but also induces mRNA cleavage in *A. thaliana* (Figure 4 d).

Furthermore, RAP dependent mechanisms involved in gene expression regulation in eukaryotic cells have been shown to affect transcription termination and recoding (Figure 4 e) (Atkins et al., 2007; Ryan, King, & Thomas, 1991; Sharma et al., 2012), mRNA splicing (Figure 4 f) and other events effecting gene expression (Ito & Chiba, 2013).

### 8.3.2 Sensing metabolites & drugs

There are multiple examples of ribosome arrest peptides that control the expression of genes conferring resistance to a range of antibiotics inhibiting protein synthesis. Some of the best studied are the leader peptides of the Erm-type methyltransferases, including ErmCL and ErmBL (Dubnau, 1985; Min, Kwon, Yoon, Shim, & Choi, 2008; Narayanan & Dubnau, 1987).

Sub-lethal concentrations of the effector drug induce expression of a downstream methyltransferase gene. The Erm-type methyltransferases confer resistance to erythromycin and other antibiotics (macrolides, ketolides, lincosamides, streptogramin B) by mono-/dimethylation of A2058 of the 23S rRNA (M. Liu & Douthwaite, 2002). The same regulatory principle applies for the induction of the expression of resistance genes that confer resistance against the ribosome targeting antibiotic chloramphenicol (reviewed in (Lovett, 1996)). Detoxifying or evading the antibiotic by inducing expression of *catA86* or *cmlA*, coding for a chloramphenicol acetyltransferase and an efflux pump, respectively, as this regulatory principle might apply for the control of other substrate induced resistance marker (Ramu et al., 2009; Reilman et al., 2014; Y. Wang, Rotman, Shoemaker, & Salyers, 2005).

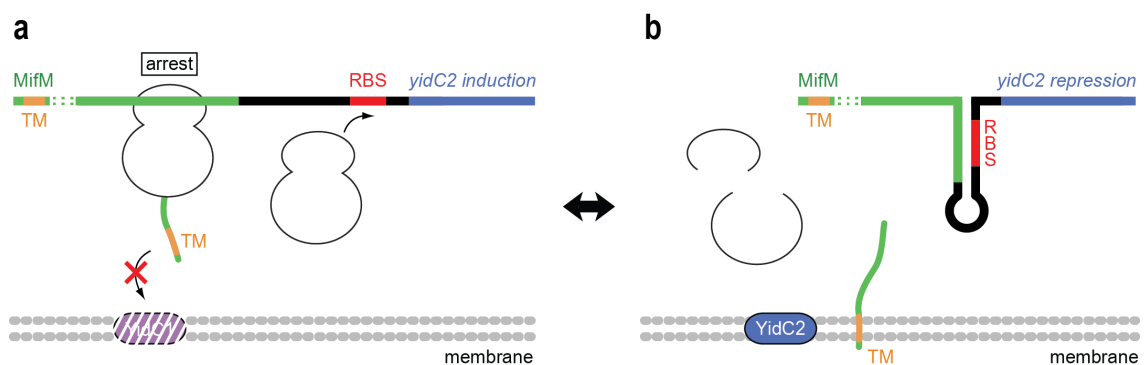
One well-studied example of ribosome arrest peptide mediated regulation of gene expression controlling a metabolic pathway is the *E. coli* *tnaCAB* operon (Gong & Yanofsky, 2003). Elevated intracellular levels of the free amino acid tryptophan (L-Trp) induce expression of the *tnaCAB* operon encoding the L-Trp catabolising enzyme tryptophanase (*tnaA*) and a tryptophan-specific permease (*tnaB*). A recently published cryo-EM structure of the ribosome stalled by the *tnaC* arrest peptide in the presence of tryptophan showed two tryptophan molecules in the ribosomal tunnel coordinated by the nascent peptide and parts of the ribosomal tunnel, explaining how the translating ribosome can act as a small molecule sensor (Bischoff et al., 2014). While ribosome stalling on the arginine attenuator peptide (AAP) (Z. L. Luo & Sachs, 1996) in response to the free amino acid L-arginine leads to down regulation of gene expression of the small subunit of the arginine-specific carbamoyl phosphate synthase. In *Arabidopsis thaliana* (*A. thaliana*), the transcriptional activator basic region-leucine zipper transcription factor 11 (bZIP11) represents a sucrose-level dependent controlled protein. Here, sucrose-sensing is regulated by one of its four uORFs (uORF2) on the 5' leader mRNA which encodes for the conserved sucrose control peptide (SC-peptide). SC-peptide translating ribosomes get stalled at the uORF2 stop codon, only in response to sucrose, efficiently blocking scanning ribosomes from translating key metabolism regulation proteins located further downstream. The molecular details of this stalling event are presently still under investigation, however it is assumed that the interaction of four conserved amino acids (serine29, serine31, leucine35 and tyrosine39) is essential for the peptide-mediated stalling mechanism (Hummel et al., 2009; Rahmani et al., 2009).

### 8.3.3 Monitoring physiological functions

Some RAPs stall translation independent of the presence of an effector molecule. Their ability to modulate translation is mediated by the amino acid sequence identity of the stalling sequence

alone. Nevertheless, effector independent translation arrest can also regulate gene expression in response to environmental and physiological changes.

The N-terminal segment of the *E. coli* secretion monitor (SecM) (Nakatogawa & Ito, 2001a) and the *B. subtilis* membrane insertion and folding monitor (MifM) (Chiba et al., 2009) is long enough to have emerged from the ribosome when translation is stalled by the following C-terminal arrest sequence located in the ribosomal tunnel. The N-terminal segment plays a pivotal role in releasing the translational arrest, as under physiological conditions SecM and MifM mediated translation arrest is abrogated by interaction with the cellular protein translocation machinery.



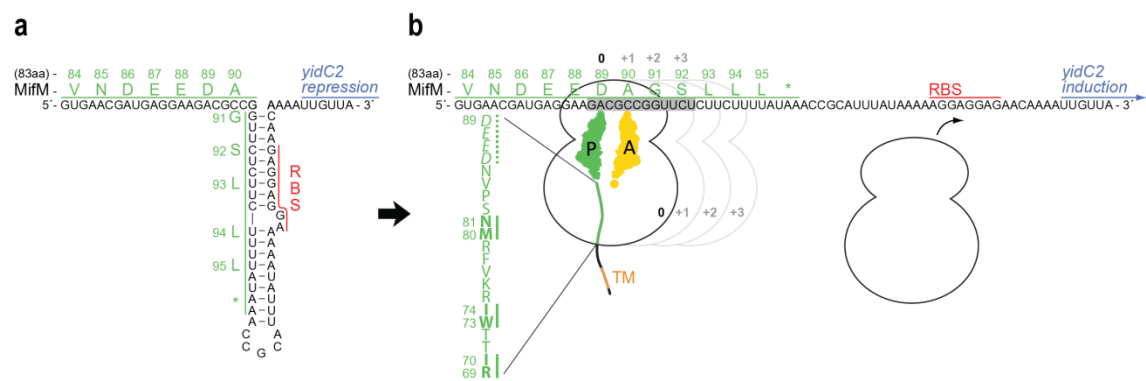
**Figure 5 | Schematic for regulation of *yidC2* expression.** (a) Under YidC1 (*spoIIIJ*)-deficient conditions MifM insertion into the cell membrane is compromised. The prolonged MifM dependent ribosome stalling maintains the unfolded conformation of the *mifM-yidC2* mRNA, which allows ribosome binding to the downstream ribosomal binding site (RBS) and expression of the *yidC2* gene (blue). (b) The expression of YidC2 (*yqjG*) restores the membrane protein insertion capacity of the cell. Interaction of the N-terminal transmembrane (TM) segment (orange) within the MifM nascent chain with YidC2 abrogates ribosome stalling. Subsequent progression and release of the ribosome from the mRNA leads to the reformation of a stem-loop structure within the *mifM-yidC2* mRNA that sequesters the ribosome-binding site (RBS) of the *yidC2* gene leading to repression of the *yidC2* gene (blue).

Truncation of the N-terminal segment causes SecM and MifM to stall translation constitutively (Chiba et al., 2009; Nakatogawa & Ito, 2002) and translation arrest is prolonged when the activity/capacity of the protein translocation machinery is diminished. Prolonged ribosome arrest by SecM leads to upregulation of SecA ATPase (Nakatogawa & Ito, 2001b), a molecular motor to promote protein secretion at a Sec translocon (Chatzi, Sardis, Economou, & Karamanou, 2014; van der Sluis & Driessen, 2006). Prolonged ribosome arrest by MifM leads to upregulation of YidC2 (*yqjG*) a protein insertase from the YidC/Oxa1/Alb3 protein family (Funes, Kauff, van der Sluis, Ott, & Herrmann, 2011), restoring the membrane protein biogenesis capacity of the cell (Figure 5) (Chiba et al., 2009). This suggests that *E. coli* SecM and *B. subtilis* MifM have evolved to monitor the capacity of the protein translocation/insertion machineries of the cell by probing the respective machinery with their N-terminal determinants

for secretion (SecM) or membrane integration (MifM) activity, respectively, engaged in the translocation/insertion reaction.

### 8.3.3.1 The membrane insertion and folding monitor

The *mifm* uORF is found in a subset of species of the Bacillaceae (Chiba et al., 2009). *Bacillus subtilis* (*B. subtilis*) MifM is encoded by a 95 codon long open reading frame 30 nucleotides upstream of YidC2 (*yqjG*) (Chiba et al., 2009). Translation of MifM was shown to stall the ribosome close to the end of the coding sequence. Remarkably, it was found that MifM undergoes elongation arrest at multiple consecutive stall sites with the residues D89, A90, G91 or S92 in the ribosomal P-site (Figure 6) (Chiba & Ito, 2012).



**Figure 6 | Schematic for MifM translation arrest dependent regulation of *yidC2* expression.** (a) Schematic of the *mifM-yidC2* mRNA illustrating the stem-loop structure that sequesters the ribosome-binding site (RBS) of the *yidC2* gene (blue). (b) Multisite ribosome stalling (0, +1, +2 and +3) during translation of MifM maintains the unfolded conformation of the mRNA allowing ribosome binding and induction of *yidC2* expression. The MifM stalling sequence (residues 69–89) is indicated with critical residues in boldface. The amino acids of the acidic cluster are marked italic. The N-terminal trans-membrane (TM) segment (orange) is indicated within the nascent peptide chain. Abbreviations: P, P-site peptidyl-tRNA; A, A-site aminoacyl-tRNA. An asterisk indicates a stop codon.

Moreover, residues critical for stalling of MifM have only been identified in the mid- and lower-tunnel region within a stalling relevant sequence of at least 21 amino acids: **R<sub>69</sub>ITTWIRKVF***RMNSPVNDEED<sub>89</sub>A<sub>90</sub>G<sub>91</sub>S<sub>92</sub>* (amino acids critical for stalling are marked as boldface letters, the cluster of acidic amino acids is marked italic, the four consecutive stall sites are underlined). The importance of the mid-tunnel region for MifM stalling is corroborated by a mutant which is defective in MifM dependent translation arrest, carrying a duplication of seven amino acids in the  $\beta$ -hairpin of ribosomal protein L22 lining the ribosomal tunnel wall close to the mid-tunnel constriction (Chiba et al., 2009). Further, the conserved cluster of acidic amino acids (D<sub>86</sub>EED<sub>89</sub>) was found to be critical for MifM dependent stalling. Single substitutions of these PTC-proximal amino acids showed only minor effects on MifM function, but stalling was lost if all four acidic amino acids were changed to alanine altogether (Chiba & Ito, 2012).

Consistent with finding amino acids critical for stalling only within the mid- and lower-tunnel regions, stalling was hardly affected by insertion of an alanine between the mid-tunnel element and the PTC-proximal acidic cluster (Chiba & Ito, 2012). Most interestingly, MifM has been shown to arrest translation on *B. subtilis* ribosomes but not efficiently on *Escherichia coli* (*E. coli*) ribosomes (Chiba et al., 2011; Vazquez-Laslop & Mankin, 2011). A similar species-specific activity has been shown for the secretion monitor (SecM) where stalling solely works in *E. coli* (Chiba et al., 2011; Vazquez-Laslop & Mankin, 2011). However, MifM and SecM do not exhibit species-exclusive regulation. Replacement of the N-terminal trans-membrane region of MifM with the N-terminal signal-sequence of a Sec dependently secreted protein successfully converts MifM to sense SecA-dependent protein secretion rather than YidC-dependent protein integration into the membrane. (Chiba et al., 2011). MifM stop-codon mutants, programming the A-site with the introduced stop codon, revealed that MifM stalling can as well impair translation at the step of termination, while MifM stalled peptidyl-tRNAs produced from a truncated mRNA template are resistant to treatment with puromycin, indicating inhibition of PTC activity (Chiba & Ito, 2012).

### 8.3.3.2 Mechanism of MifM mediated translation arrest release

MifM mediated ribosomal elongation-arrest is released by co-translational membrane insertion of its N-terminal TM region by SpoIIJ (YidC1) (Chiba et al., 2011; Chiba et al., 2009; Ito et al., 2010). The current model for YidC-dependent membrane protein integration assumes that release of the substrate from the YidC insertase into the membrane is facilitated by the hydrophobic interaction between the TM region and the lipid interior of the membrane, as well as the membrane potential attracting negatively charged residues of the extracellular region of the membrane protein by electrostatic attraction (Ismail, Hedman, Linden, & von Heijne, 2015; Kumazaki, Chiba, et al., 2014). Therefore, it is conceivable that insertion of the N-terminal MifM TM region into the membrane would exert a pulling force on the nascent polypeptide that could physically disrupt further C-terminal arrest-important interactions between residues of MifM and the ribosomal exit tunnel (Cymer, Hedman, Ismail, & von Heijne, 2015; Cymer, Ismail, & von Heijne, 2014; Cymer, von Heijne, & White, 2015).

The pulling force model was first proposed to explain relief of SecM mediated ribosome translation arrest by the pulling force generated by the ATP consuming motor protein SecA, which has been shown to drive protein translocation across the membrane (Butkus, Prundeanu, & Oliver, 2003; Gumbart, Schreiner, Wilson, Beckmann, & Schulten, 2012). But, it was also shown that an engineered stop transfer-membrane anchor sequence is sufficient to release SecM mediated stalling by the pulling force exerted on the nascent peptide, when the membrane anchor reaches the membrane and when it is fully transferred to the height of the membrane leaving the translocon into the lipid bilayer (Ismail, Hedman, Schiller, & von Heijne, 2012).

Due to the observed pulling force profile, SecA is assumed not to be a major contributor to the force generated to overcome the ribosomal translation halt, but the process of membrane protein insertion itself can generate a pulling force strong enough to resume arrested translation. Therefore, SpoIIIJ-mediated membrane insertion of MifM might well provide sufficient pulling force to release elongation arrest.

Though, to relieve stalling by the interaction with the translocon, the stop transfer-membrane anchor sequence has to be positioned at a certain distance from the arrest sequence, with force profiles peaking at sequences around a length of 30 and 40 amino acids in between the membrane anchor and the PTC. Noteworthy, the distance between the C-terminal end of the TM segment of the MifM leader peptide and the amino acid located in the P-site at the first stalling site of the MifM arrest sequence is 54 amino acids, posing a challenge for the pulling force model with the nascent chain envisioned as a simple rope to pull on. Investigation of internal substitution and deletion mutants of the SecM arrest peptide identified a stretch of amino acids in the middle of the peptide required for the export-coupled release of the SecM nascent chain (Nakamori, Chiba, & Ito, 2014). Additional experiments are therefore needed to elucidate the contribution of the force generated by the integration of the TM segment of the MifM leader peptide into the membrane and of the functional interactions of the MifM nascent chain with the ribosome and SpoIIIJ to the regulation of the translation arrest.

### 8.3.3.3 Protein insertases of the YidC/Oxa1/Alb3 family and substrates

*Bacillus subtilis* YidC1 (*spoIIIJ*) and YidC2 (*yqjG*) are homologues of the membrane protein insertase Oxa1p (Murakami, Haga, Takeuchi, & Sato, 2002; Saller, Fusetti, & Driessen, 2009) and belong to the YidC/Oxa1/Alb3 family of proteins. Membrane protein insertases of the YidC/Oxa1/Alb3 protein family are present in the inner membrane of eubacteria (YidC, YidC1/YidC2), the inner mitochondrial membrane (Oxa1/Cox18) and the thylakoid membrane of Chloroplasts (Alb3/Alb4) (Funes et al., 2011), assisting in membrane protein insertion in the lipid bilayer (Samuelson et al., 2000). Most likely due to an early gene duplication event, two YidC genes can be found in the genomes of most Gram-positive bacteria as in species from *Bacillus*, *Lactobacillus*, *Clostridium* and some Actinobacteria. The two paralogues differ in their phylogeny and therefore have been classified as different subgroups designated as YidC1 and YidC2 (Funes et al., 2009; Funes et al., 2011).

Initially mutation of *spoIIIJ* (YidC1) in *Bacillus subtilis* was found to abolish transcription of pre-spore specific genes blocking sporulation after pre-spore engulfment (Errington et al., 1992; Murakami et al., 2002) with no observable phenotype during vegetative growth. Likewise, deletion of *yqjG* (YidC2) did not affect vegetative growth or sporulation and SpoIIIJ and YqjG have been shown to be largely exchangeable in function (Saller et al., 2009), though YqjG



cannot complement SpoIIIJ activity in spore formation and has been implicated to have a specific function in genetic competence development (Saller et al., 2011). Double mutation of *spoIIIJ* and *yqjG* though is lethal.

YidC like insertases of the YidC/Oxa1/Alb3 family of proteins share a conserved membrane region constituted from 5-6 trans-membrane (TM) helices (6TM helices in Gram-negative bacteria) and recent crystal-structures of *E. coli* YidC and YidC2 from *B. halodurans* by Kumazaki et al. (Kumazaki, Chiba, et al., 2014; Kumazaki, Kishimoto, et al., 2014) revealed, that unlike other known proteins and protein complexes involved in membrane protein biogenesis, YidC like insertases do not form a pore in the membrane but a hydrophilic groove in between TM3 and TM5. This hydrophilic groove shows an opening towards the membrane interior and the cytoplasmic side and is meant to provide a hydrophilic environment within the membrane to receive the hydrophilic region of a substrate (termini, loops) guiding its integration (Jiang et al., 2003). YidC has been shown to interact with components of the Sec machinery and it is known for YidC to facilitate integration of membrane proteins into the lipid bilayer in a YidC-dependent (YidC-only) and a Sec-dependent way (Robinson & Woolhead, 2013; Serek et al., 2004) as part of the SecDF-YajC-YidC complex (Nouwen & Driessen, 2002) and the “holo-translocon” (SecYEG-YidC-SecDF-YajC) complex (Schulze et al., 2014) assisting in TM integration and assembly of multimeric membrane protein complexes (Kol, Nouwen, & Driessen, 2008).

## 9 Cumulative Thesis: Summary of Published Results

### 9.1 Paper 1 | Proteomic Characterization of Archaeal Ribosomes Reveals the Presence of Novel Archaeal-Specific Ribosomal Proteins

Viter Márquez, Thomas Fröhlich, Jean-Paul Armache, Daniel Sohmen, Alexandra Dönhöfer, Aleksandra Mikolajka, Otto Berninghausen, Michael Thomm, Roland Beckmann, Georg J. Arnold and Daniel N. Wilson

*Journal of Molecular Biology* 405, 1215–1232 (2011)

The appearance of archaea resembles that of a bacterial cell. Previously, archaea had been confused with bacteria and their genera were found combined under the term prokaryota. The autonomous position of archaea in the tree of life was first uncovered by comparison of 16S rRNA sequences, establishing archaea as an independent phylogenetic domain of life in addition to bacteria and eukarya (Woese & Fox, 1977). However, archaea have a lot in common with their prokaryotic relatives at the molecular level. They have a small circular genome, a compact gene organization, with functionally related genes organized within operons. At the same time archaea also resemble the eukaryotic cell in many ways. Many archaeal proteins are homologous or more similar to their eukaryotic counterpart. Therefore, at the molecular level, archaea are often described as of intermediate complexity between bacteria and eukaryotes (Forterre, 2013; Olsen & Woese, 1997; Zillig, 1991).

The ribosomes of the well-studied bacterial model organism *E. coli* contain 54 r-proteins, 21 in the small subunit and 33 in the large subunit. Further genomic analyses have identified ~57 r-protein families that are almost universally present in the bacterial ribosome. 34 of which are universally conserved having homologues in eukarya and archaea, 23 are bacteria specific. Proteomic together with genomic analyses (Desmond, Brochier-Armanet, Forterre, & Gribaldo, 2010; Lecompte et al., 2002; Yutin, Puigbo, Koonin, & Wolf, 2012) of eukaryotic ribosomes ranging from yeast, to human, identified ~78 r-protein families to be present in the eukaryotic ribosome and 11 r-proteins specific to eukarya. The archaeal ribosome can be compared to the bacterial 70S ribosomal particle in terms of size and rRNA content. It contains 3 rRNAs (5S, 16S and 23S rRNAs) as opposed to 4 rRNAs in the eukaryotic ribosome. The

archaeal ribosome is composed of ~68 r-protein families of which ~33 have homologues to be found only in ribosomes from eukarya, while there are no r-proteins that are bacteria and archaea specific, and only one r-protein that is archaea specific. But while the protein composition of many bacterial and eukaryotic ribosomes has been well characterised, a comprehensive investigation, cataloguing the protein composition of archaeal ribosomes so far was missing.

To close this gap, the composition of the ribosomes of the thermophilic Crenarchaeota *Pyrobaculum aerophilum* and the thermoacidophilic Crenarchaeota *Sulfolobus acidocaldarius* were analysed by 2D gel electrophoresis coupled with LC tandem mass spectrometry (MS/MS) and single particle electron microscopy (EM). These analyses could identify all of the predicted r-proteins almost completely. Only S26e and S30e of the *Sulfolobus acidocaldarius* ribosome were not found in the preparations. Further MS analyses of *S. acidocaldarius* L7ae, L29p, P0 and LX and *P. aerophilum* L7ae, L12p and P0, that were previously found in multiple spots of 2D PAGE preparations were shown to be post-translational modified isoforms in most cases. Additionally three new r-proteins of the *Sulfolobus acidocaldarius* large subunit were discovered and proposed to be renamed L45a, L46a and L47a. L45a, L46a and L47a were shown to be integral parts of the ribosome, embedded deeply into the ribosomal structure as shown by LiCl fractionation and 2D gel quantification. Homology searches in all available (91 at the time published) completely sequenced archaeal genomes revealed that L45a is present in all *Sulfolobus* species with one exception (*Sulfolobus tokodaii* str. 7). L46a is more widely distributed in all *Thermoprotei* and L47a can be found in all *Sulfolobaceae*. A similar phylogenetic distribution has also later been found in an automated computational search performed by Yutin et al. (Yutin et al., 2012).

To identify putative binding sites of the novel r-proteins, 3D models of purified ribosomal 50S subunits from *S. acidocaldarius* and *P. aerophilum* were reconstructed by cryo-EM single particle analysis to a resolution of 27Å and 25Å (0.5 FSC), respectively. Comparison of the 3D models with the ribosomal structure of *Haloarcula marismortui* 50S led to the identification of additional densities in the model of *Sulfolobus acidocaldarius* as potential binding sites for the novel r-proteins. Reproducible appearance of L7ae and S24e in preparations of both, the small and the large subunit of *Sulfolobus acidocaldarius*, were discussed as to hint to possible promiscuous interactions of the proteins with both ribosomal subunit. But unambiguous assignment of density for the discovered novel r-proteins or additional copies of L7ae or S24e was not possible due to the limited resolution. Armache *et al.* (Armache et al., 2013) later indeed discovered L7ae and S24e in both subunits of 3D ribosomal reconstructions of *Pyrococcus furiosus* and *Thermococcus kodakarensis*. Together these findings suggest that there probably

still is a number of r-proteins existing in archaea that have not been described yet and their discovery could help to understand the evolution of the ribosome and the development of r-protein functions.

## 9.2 Paper 2 | Amythiamicin D and Related Thiopeptides as Inhibitors of the Bacterial Elongation Factor EF-Tu: Modification of the Amino Acid at Carbon Atom C2 of Ring C Dramatically Influences Activity

Stefan Gross, Fabian Nguyen, Matthias Bierschenk, Daniel Sohmen, Thomas Menzel, Iris Antes, Daniel N. Wilson, and Thorsten Bach

*ChemMedChem*, 8, 1954–1962 (2013)

The translational GTPase elongation factor Tu (EF-Tu) is a major factor in bacterial protein biosynthesis. EF-Tu delivers aminoacyl-tRNA to the ribosomal A-site and functionally contributes to tRNA selection by the ribosome. On account of its function, EF-Tu is also a potential target for antimicrobial compounds and currently four families of antibiotics of unrelated structure are known to efficiently inhibit EF-Tu actions, with kirromycin, enacycloxin IIa, pulvomycin, and GE2270A as prototypic examples. GE2270A belongs to the thiopeptide antibiotics, a group of ribosomally synthesised and extensively posttranslationally modified, antimicrobial peptides. Multiple chemical modification studies of GE2270A have been conducted to identify structural and functional important parts of the drug. One region that has been indentified is a phenyl-glycine-derived moiety with a  $\alpha$ -hydroxybenzyl group located in the northeastern part of GE2270A. Deletion of the  $\alpha$ -hydroxybenzyl group renders GE2270A inactive. The crystal structure of GE2270A bound to EF-Tu•GTP shows the  $\alpha$ -hydroxybenzyl group to be located close to the helix of the switch II loop in domain D1 of EF-Tu, implicated in catalytic activity of EF-Tu, making this position attractive for further studies. Therefore, non-natural thiopeptides have been utilized to study the influence of this region on translation inhibition. Three analogues of the naturally produced antibiotic amythiamicin D, with different substitutions of the  $\alpha$ -hydroxybenzyl group, that could not be obtained other wise, were made available, for the first time, by de novo total synthesis. One of the derivatives reduced growth of Gram-positive bacteria with a minimal inhibitory concentration (MIC) comparable to amythiamicin D and inhibited translation in cell free *in vitro* translation experiments even more effective than amythiamicin D, while the other analogues were found to be inactive in both,

bacterial growth inhibition and *in vitro* translation experiments. Therefore, binding of amythiamicin D and its derivatives to EF-Tu was investigated by molecular dynamics (MD) simulations. Simulations resulted in binding positions similar to GE2270A only for amythiamicin D and its active derivative, with the analogue's hydroxyl group of its hydroxymethyl substitution suggested to be able to form an additional hydrogen bond with residues of EF-Tu, comparable to the hydroxyl group of the  $\alpha$ -hydroxybenzyl moiety of GE2270A. Therefore his study provides tools and insights for the generation of new non-natural thiopeptides and derivatives by de novo total synthesis to study their physiochemical properties as translation inhibitors and antimicrobial agents.

### 9.3 Paper 3 | Entrapment of DNA in an intersubunit tunnel system of a single-stranded DNA-binding protein

Homa Ghalei, Holger von Moeller, Detlef Eppers, Daniel Sohmen, Daniel N. Wilson, Bernhard Loll and Markus C. Wahl

*Nucleic Acids Research*, 42, 6698–6708 (2014)

Single strand DNA-binding proteins (SSBs) are among the most widespread proteins in eukarya, bacteria and archaea, and are involved in major cellular processes like DNA replication, recombination and repair (reviewed in (Pestryakov & Lavrik, 2008)), which underlines their importance for the cell. SSBs are ubiquitous in nature, except in some hyperthermophilic Crenarchaeota that lack a canonical SSB, but instead encode non-canonical SSBs with unique ssDNA-binding domains termed “ThermoDBPs” (Paytubi et al., 2012).

Two other only poorly characterised groups (1 and 2) of proteins with a broader phylogenetic distribution, related to ThermoDBPs, sharing the ThermoDBP's N-terminal DNA-binding domain (NTD) are known to co-exist with ThermoDBPs and other SSBs. Previously, Rbp18 from *Sulfolobus solfataricus* P2 (*S. solfa*), belonging to group 1 of ThermoDBP related proteins (ThermoDPB-RP1), was found to be associated with box C/D sRNAs as well as with the small subunit of the *S. solfa* ribosome (Ciammaruconi, Gorini, & Londei, 2008). However, no ThermoDPB-RP1 protein was detected in a subsequent proteomic analysis of archaeal ribosomes from *Sulfolobus acidocaldarius* (Marquez et al., 2011) or after enrichment for ssDNA binding proteins from *Thermoproteus tenax* (Paytubi et al., 2012). Therefore, we set out to further characterize the structure and function of ThermoDBP-RP1 and ThermoDBP-RP2

from *Pyrococcus furiosus* (*pfu*ThermoDBP-RP1) and *Aeropyrum pernix* (*ape*ThermoDBP-RP2), respectively.

In our experiments, binding of *pfu*ThermoDBP-RP1 to box C/D sRNAs *in vitro* was efficiently competed by the presence of *E. coli* tRNA. Further, only a weak, salt sensitive interaction of *pfu*ThermoDBP-RP1 and *ape*ThermoDBP-RP2 with archaeal *Thermococcus kodakarensis* (*T. kod*) ribosomes was detected. Both results indicate a nonspecific interaction of *pfu*ThermoDBP-RP1 and *ape*ThermoDBP-RP2 with box C/D sRNAs and the *T. kod* ribosome *in vitro*, yet, they do not rule out a possible interaction of ThermoDBP-RP proteins with box C/D sRNAs and/or the ribosome *in vivo*. Therefore, we looked to investigate the interaction of the ThermoDBP-RP proteins with other nucleic acids. Indeed, we could show, that *ape*ThermoDBP-RP2 specifically binds to mixed sequence or homopyrimidin stretches of ssDNA *in vitro* with high affinity. In addition, we were able to solve the structure of *ape*ThermoDBP-RP2 in complex with single strand oligo-dT DNA at 2.9Å resolution. Moreover, we solved the crystal structures of *pfu*ThermoDBP-RP1 and *ape*ThermoDBP-RP2 at 2.43Å and 2.05Å, respectively. Both proteins were found to form homotetramers, using different mechanisms for oligomerization, revealing how the N-terminal domain of *ape*ThermoDBP-RP2 contacts the nucleic acid to facilitate specific binding of single-stranded DNA molecules within the symmetric homotetrameric protein complex. The ssDNA bound traverses an internal, intersubunit tunnel system in the *ape*ThermoDBP-RP2 tetramer. This mode of binding ssDNA is unique and distinguishes *ape*ThermoDBP-RP2 from canonical SSBs binding ssDNA via interactions on the outside of the ssDNA-binding domain.

The loss of a canonical SSB in this particular group of Crenarchaeota, having acquired special SSBs to protect single stranded stretches of DNA in an intersubunit tunnel system, may have been a necessity to adapt to the particularly challenges in the living environment of these hyperthermophilic organisms.

## 9.4 Paper 4 | Structure of the *Bacillus subtilis* 70S ribosome reveals the basis for species-specific stalling

Daniel Sohmen, Shinobu Chiba, Naomi Shimokawa-Chiba, C. Axel Innis,  
Otto Berninghausen, Roland Beckmann, Koreaki Ito & Daniel N. Wilson

*Nature communications* 6, 6941 (2015)

To uncover the structural basis of MifM dependent translational stalling and to provide insight into the species specificity of MifM arrested translation, a *B. subtilis* MifM-SRC was reconstructed from 305,045 electron microscopy single particle images (cryo-EM) at a resolution of 3.5-3.9Å. Prior to generating a homogeneous MifM-SRC sample suitable for single particle analysis, a cell-free *B. subtilis in vitro* translation system had to be developed. The final reconstruction allowed the first 70S ribosome model of a Gram-positive bacterium to be built. Density for the nascent peptide chain, observed throughout the entire length of the ribosomal exit tunnel allowed modelling of the MifM arrest peptide, localizing the three pairs of MifM residues critical for stalling (Ito et al., 2010), at the sites of contact with components of the ribosomal tunnel - the constriction, the  $\beta$ -hairpin of L22 and H50 at the tunnel exit. Subsequent analysis of sequence and structure conservation of the ribosomal tunnel components between *B. subtilis* and *E. coli* identified a number of amino acid exchanges between *B. subtilis* and *E. coli* L22, suggesting L22 as one of the determinants for species specificity. Genetic replacement of the *B. subtilis* L22  $\beta$ -hairpin with the corresponding sequence from *E. coli* strongly reduced stalling on a MifM-LacZ reporter. Strikingly, reverting a single amino acid (K90M) in the sequence of the  $\beta$ -hairpin of L22 from *E. coli* (K90) back to *B. subtilis* (M90) was shown to be sufficient to restore stalling, thereby identifying M90 of L22 as a key determinant for species specificity of MifM stalling. Comparison of the conformation of nucleotides of the MifM-SRC PTC with the PTC in different states of peptide bond formation (Schmeing, Huang, Kitchen, et al., 2005; Schmeing, Huang, Strobel, et al., 2005; Schmeing et al., 2002; Voorhees, Weixlbaumer, Loakes, Kelley, & Ramakrishnan, 2009) showed that the defined conformation of MifM residue E87 close to the PTC stabilises the uninduced state of the PTC by sterically preventing nucleotides U2506 and U2585 from adopting their induced conformational states, thereby inhibiting peptide bond formation by preventing accommodation of the incoming A-tRNA.

## 9.5 Paper 5 | Structural basis for targeting and elongation arrest of *Bacillus* signal recognition particle

Bertrand Beckert, Alexej Kedrov, Daniel Sohmen, Georg Kempf, Klemens Wild, Irmgard Sinning, Henning Stahlberg, Daniel N. Wilson and Roland Beckmann

*Unpublished manuscript* (2015)

Signal sequence recognition of the nascent peptide on the ribosome by the signal recognition particle (SRP) results in co-translational targeting of membrane and secretory proteins to the protein conducting channel (Akopian, Shen, Zhang, & Shan, 2013; Halic & Beckmann, 2005; Nyathi, Wilkinson, & Pool, 2013). The mammalian and the “long-type” SRP of Gram-positive bacteria (Larsen & Zwieb, 1991; Struck, Vogel, Ulbrich, & Erdmann, 1988) can be divided into two functional domains: The S domain which recognizes the signal sequence while the interaction of the Alu domain with the ribosome imposes an elongation arrest/slow-down by blocking the elongation factor binding site (Halic et al., 2004; Halic, Gartmann, et al., 2006; Mason, Ciuffo, & Brown, 2000; Powers & Walter, 1997; Siegel & Walter, 1985; Walter & Blobel, 1981). However, interaction of the bacterial Alu domain (Rosenblad, Larsen, Samuelsson, & Zwieb, 2009; Zwieb & Bhuiyan, 2010) with the ribosome has been shown to be different from the mammalian in many ways (Kempf, Wild, & Sinning, 2014).

To characterise the prokaryotic SRP/ribosome interaction an SRP-RNC “targeting complex” was assembled from reconstituted *Bacillus subtilis* (*B. sub*) SRP and stalled *B. sub* ribosomes exposing a signal sequence. Next, based on our 3D reconstruction of this *B. sub* SRP-RNC “targeting complex” by cryo-EM and single particle analysis, the complete molecular model of the *B. sub* SRP bound to the ribosome could be constructed, reaching from the ribosomal tunnel exit site to the translation factor binding site. The S domain of the *B. sub* SRP, located at the tunnel exit was found to be structurally similar to the S domain of the *E. coli* SRP bound to the ribosome (Batey, Rambo, Lucast, Rha, & Doudna, 2000; Halic & Beckmann, 2005; Wild, Halic, Sinning, & Beckmann, 2004) with the rod like density of the signal anchor accommodated in the open conformation of the M domain of SRP54 (Halic, Blau, et al., 2006; Janda et al., 2010; von Loeffelholz et al., 2015).

Furthermore, our structure revealed that the *B. sub* SRP Alu domain almost exclusively interacts with the large ribosomal subunit via RNA:RNA interactions contacting the ribosomal stalk base and the  $\alpha$ -sarcin-ricin loop in the factor binding site. In comparison, the Alu domain of the mammalian SRP, based on the interpretation of our additional cryo-EM reconstruction of



the mammalian ribosome “targeting complex” (this study) (Halic et al., 2004; Halic, Gartmann, et al., 2006) assembled from *Canis familiaris* SRP and stalled *Triticum aestivum* ribosomes, only shows interactions with the ribosome via protein:RNA contacts. Thereby, the mammalian SRP Alu RNA contacts uL11 above the stalk base on the large ribosomal subunit.

Moreover, the mammalian Alu domain establishes additional contacts with the small ribosomal subunit RNA via the mammalian-specific SRP9/14 protein heterodimer, which is absent in *B. sub*. Yet the unique structure of the *B. sub* SRP Alu RNA domain allows multiple interactions with the stalk base helices H43 and H44 as well as the SRL, stabilized by additional interactions with uL11 Pro22 and the C-terminal end of uL6. Thus, the entire *B. sub* Alu domain acts as a “dock and lock” system in which the translation factor binding site of the ribosome is locked temporarily in order to slow down the translation machinery for efficient and proper targeting to a membrane translocation site.

## 10 Discussion

### 10.1 Paper 1 | Proteomic Characterization of Archaeal Ribosomes Reveals the Presence of Novel Archaeal-Specific Ribosomal Proteins

Viter Márquez, Thomas Fröhlich, Jean-Paul Armache, Daniel Sohmen, Alexandra Dönhöfer, Aleksandra Mikolajka, Otto Berninghausen, Michael Thomm, Roland Beckmann, Georg J. Arnold and Daniel N. Wilson

*Journal of Molecular Biology* 405, 1215–1232 (2011)

The archaeal ribosomal proteome shows considerable heterogeneity across the archaeal domain of life. Analyses of archaeal genomes implicate that Archaea have undergone progressive loss of ribosomal proteins (r-proteins) as part of a reductive evolution of their translation apparatus (Lecompte et al., 2002). Within the archaeal domain, Crenarchaeota still contain the highest number of r-proteins, while genome analyses show a gradual reduction of the number of r-proteins within the phylum of the Euryarchaeota (Lecompte et al., 2002). Comparison with ribosomal proteomes from Bacteria and Eukarya shows that the r-proteins (excluding the archaeal-specific r-protein LXa) that have been lost in Archaea are also absent in Bacteria but are found in a wide range of Eukarya (Fujita, Baba, & Isono, 1998; Kenmochi et al., 1998; Uechi, Tanaka, & Kenmochi, 2001). The presence of these r-proteins in early divergent representatives of Eukarya and Archaea suggests their existence in the common ancestor of both domains in a more complex ribosome, emphasizing the need for experimental study of the archaeal ribosome to understand the evolution of the ribosome and the development of r-protein functions.

Analyses of the ribosomes from the Crenarchaeota species *Pyrobaculum aerophilum* (*P. aero*) and *Sulfolobus acidocaldarius* (*S. acid*) by 2D PAGE and MS led to the discovery of three new r-proteins: L45a-47a. Their phylogenetic range, revealed by homology searches showed homologues of L45a and L47a in all known species of the *Sulfolobales* and homologues of L46a in all but one species of the *Thermoprotei* genera (this study) (Yutin et al., 2012), while no sequence homology to other known proteins was indicated. The NMR structure of an L46a homologue from *Sulfolobus solfataricus* P2 in solution, determined by Feng *et al.* (Feng et al.,

2014), revealed a novel protein fold and a conserved potential rRNA-binding site of L46a. However, neither computational searches for structural homologues could identify such structural homologues convincingly, nor did L46a's structural features relate to structural similar proteins and their function. Efforts to localize the new r-proteins within the volume of the 50S ribosomal subunit from *S. acid*, reconstructed by cryo-EM single particle analysis, led to ambiguous results due to the limited resolution of the reconstruction.

Reproducible appearance of L7ae (L8e) and S24e in preparations of both the small and the large subunits of *Sulfolobus acidocaldarius* were discussed as to hint to a possible promiscuous interaction of the proteins with both ribosomal subunits. However, unambiguous assignment of density for additional copies of L7ae (L8e) or S24e was again not possible due to the limited resolution and the lack of small ribosomal subunit in the reconstruction. One impediment for studies of many archaeal ribosomes is the lack of protocols for the purification of 70S ribosomal particles. Ribosomes from species from the phylum of the Crenarchaeota, using common bacterial buffer systems, show a strong tendency to dissociate into subunits during preparation. The development of protocols specifically designed for the preparation of 70S ribosomal particles from different archaeal species would be desirable in order to improve future possibilities to study whole archaeal ribosomal complexes. In addition, 50S ribosomal particles have been found to be a less optimal sample for cryo-electron microscopy single particle analysis, limiting the resolution of the reconstruction.

Armache *et al.* (Armache et al., 2013) later indeed discovered promiscuous inter- and intra-subunit binding of archaeal r-proteins and identified secondary binding sites for L7ae (L8e) and S24e on the small and the large subunit, respectively. Unambiguous localisation of proteins was thereby made possible due to the 3D ribosomal reconstructions of the 70S ribosomal particle from the Euryarchaeota *Thermococcus kodakarensis* and *Methanococcus igneus* and the 50S large subunit of the Crenarchaeota *Staphylothermus marinus*. Their interpretation by the atomic model of the *Pyrococcus furiosus* ribosome (Armache et al., 2013) based on a 3D reconstruction at 6.6Å resolution (Becker et al., 2012) enabled Armache et al. to model protein domains.

Together these findings highlight the importance of studying the archaeal ribosome helping to understand the evolution of the ribosome and the development of r-protein functions (Matte-Tailliez, Brochier, Forterre, & Philippe, 2002). In particular structures of higher resolution will be of importance to a mechanistic understanding of the structure/function relations of yet unknown, non-canonical, ancient or lost r-proteins in the context of the modern ribosome.

## 10.2 Paper 2 | Amythiamicin D and Related Thiopeptides as Inhibitors of the Bacterial Elongation Factor EF-Tu: Modification of the Amino Acid at Carbon Atom C2 of Ring C Dramatically Influences Activity

Stefan Gross, Fabian Nguyen, Matthias Bierschenk, Daniel Sohmen, Thomas Menzel, Iris Antes, Daniel N. Wilson, and Thorsten Bach

*ChemMedChem*, 8, 1954–1962 (2013)

Three analogues (compounds 3a-3c) of the naturally produced thiopeptide antibiotic amythiamicin D, which could not be obtained by any other method, have been made available by *de novo* total synthesis. The hydroxymethyl derivative (compound 3a) showed enhanced inhibitory activity, compared to its parent compound, on translation in an *E. coli* based *in vitro* translation assay, suggesting the introduced hydroxyl group at R<sup>1</sup> to contribute to improved binding of the synthetic compound to its target. Although the target of amythiamicin D in the bacterial cell is unknown, strong evidence from *in vitro* translation and rescue experiments is provided for the drug to inhibit the translation cycle by inhibition of the bacterial elongation factor Tu (EF-Tu), the established target of the structurally closely related antimicrobial compound GE2270A (Anborgh & Parmeggiani, 1991). Similarly, as suggested for compound 3a, GE2270A forms a hydrogen bond with its hydroxyl group corresponding to the hydroxyl moiety introduced in compound 3a and the side chain of D216 of EF-Tu, contributing to target binding (Parmeggiani, Krab, Okamura, et al., 2006). Therefore, in the absence of structural studies verifying the binding mode of amythiamicin D or its derivatives to EF-Tu, (though preliminary NMR experiments of the free compounds in combination with *in silico* docking experiments by molecular dynamics simulations were performed) additional translation experiments in the presence of D216A mutant EF-Tu (Doi, Ohtsuki, Shimizu, Ueda, & Sisido, 2007) could help to evaluate the proposed contribution of the hydroxyl moiety of compound 3a to its binding to EF-Tu and its inhibitory activity.

On the other hand, compounds 3b and 3c were found to be completely inactive, showing no translation inhibition activity even at high concentrations. Inactivity of 3b is thereby most likely due to a significantly changed ring conformation caused by the changed stereogenic center at R<sup>1</sup>, abrogating a bridge like hydrogen bond across the large cyclic structure. In compound 3c, the ring structure resembles that of amythiamicin D, however, the relatively large addition of the benzyloxymethyl (BOM) group at R<sup>1</sup> seems to prevent stable binding to EF-Tu. Unstable binding was indicated by the preliminary NMR experiments of the free compound in

combination with molecular dynamics simulations and *in silico* docking experiments, though binding of the BOM group to EF-Tu was in principle not excluded due to a number of charged and polar side chains within the proposed binding area of the group.

Although compound 3a showed enhanced inhibitory activity in cell free *in vitro* translation experiments, its potency to inhibit growth of live cells was diminished by two- to 16-fold compared to amythiamicin D (for the bacterial strains tested). This reduction in growth inhibition activity is possibly due to the introduced, negatively charged group, reducing the drug's propensity to cross the bacterial cell membrane, thereby emphasising that, next to target specificity and binding properties, also properties like cell membrane permeability, solubility and stability can affect a compound activity and should be considered in the experimental setup and for the design of derivative compounds (Bolla et al., 2011).

The total *de novo* synthesis of compounds not only allows examining the effect of various particular substitutions, but also can give access to certain parts and precursor compounds of the drug, which can help to dissect the drug's mechanism of action and to characterise the contribution of its different parts to its activity. Access to precursors can also provide the means to identify potential lead compounds for the development of novel therapeutics (Starosta et al., 2009). In addition, some precursor compounds screened by Starosta *et al.* (Starosta et al., 2009) showed the ability to bind both EF-Tu and the ribosome, suggesting the feasibility of developing antimicrobial drugs with dual inhibitory activity. Such innovative compounds would make it much less likely for bacteria to develop spontaneous resistance by acquiring mutations.

Furthermore, the combination of chemical synthesis and structure-aided design of complex compounds based on the number of crystal (D.N. Wilson, 2014), and most recently cryo-EM structures (Arenz, Meydan, et al., 2014; Arenz, Ramu, et al., 2014; Wong et al., 2014) of antibiotics bound to their drug target, could help to develop semi-synthetic compounds with improved antimicrobial properties (Fischbach & Walsh, 2009). For example, it has been suggested to construct a chimeric antibiotic compound with increased binding affinity to EF-Tu consisting of the head moiety of enacycloxin IIa and the tail moiety of kirromycin (Parmeggiani, Krab, Watanabe, et al., 2006) based on their crystal structures bound to EF-Tu that have been solved previously (Parmeggiani, Krab, Watanabe, et al., 2006). Although, one drawback of compounds, which have been derived from naturally occurring antibiotics, can be that they are likely to be, at least partially, structurally similar to existing antibiotics, and they could, therefore, easily encounter bacteria that already possess a resistance or that could rapidly adapt to become resistant. Therefore, it is important to look at known resistance mechanisms to design effective compounds with a reduced risk of resistance onset by microorganisms.

Moreover, to not only avoid early onset of resistance but to even overcome bacterial resistance it would be necessary to design compounds which are effective against mutants already conferring resistance. In some cases, such compounds, so-called “resistance breakers”, administered in combination with the mother compound can actually restore the antimicrobial activity of the mother compound (Tan, Tillett, & McKay, 2000). This is known as phenotypic conversion.

Taken together, the total *de novo* synthesis of non-natural antibiotics, like the amythiamicin D derivatives and especially the construction of compound 3a with altered biochemical and physiochemical properties provide tools and starting points for future studies trying to understand the characteristics of various substitutions or alterations which build the foundation for the successful development of new derivatives of antimicrobial compounds.

### 10.3 Paper 3 | Entrapment of DNA in an intersubunit tunnel system of a single-stranded DNA-binding protein

Homa Ghalei, Holger von Moeller, Detlef Eppers, Daniel Sohmen, Daniel N. Wilson, Bernhard Loll and Markus C. Wahl

*Nucleic Acids Research*, 42, 6698–6708 (2014)

The single-stranded DNA-binding proteins (SSB) based on the oligonucleotide-binding fold (OB-fold) are considered ubiquitous in nature and essential for genome maintenance (Mushegian & Koonin, 1996). However, it was discovered that thermoproteales, a clade of hyperthermophilic Crenarchaeota, lack canonical SSBs. Instead, they encode a distinct ssDNA-binding protein termed ThermoDBP (X. Luo et al., 2007; Paytubi et al., 2012). Furthermore, two related groups of proteins, termed ThermoDBP related proteins (ThermoDBP-RP) 1 and 2, showing a wider phylogenetic range have been found to co-exist with ThermoDBPs and/or other SSBs. ThermoDBP-RPs share the ThermoDBP N-terminal oligonucleotide-binding domain (NTD), but their function remains elusive.

Rbp18 from *Sulfolobus solfataricus* P2 (*S. solfa*), belonging to the group of ThermoDBP-RP1 proteins, was first purified by Ciammaruconi *et al.* (Ciammaruconi et al., 2008) from cell lysates via biotinylated oligonucleotides complementary to *S. solfa* box C/D sRNAs. Ciammaruconi *et al.* also found Rbp18 to be associated with the small subunit of the *S. solfa* ribosome. In their experiments Rbp18 protein remained stably associated with the

ribosomal subunit even after high salt treatment either with 2 M NH<sub>4</sub>Cl, 2 M KCl or 0.5 M LiCl before sucrose gradient density separation, suggesting Rbp18 to be part of a ribonucleoprotein (RNP) complex and the ribosome.

However, the homologous ThermoDBP-RP protein from *Thermoproteus tenax* (*T. tenax*) could neither be enriched by biotinylated oligonucleotides from *T. tenax* cell lysate (Paytubi et al., 2012) nor was a ThermoDBP-RP protein found to be present in the ribosomal proteome of *Sulfolobus acidocaldarius* ribosomes (Marquez et al., 2011). Further, in our experiments, binding of recombinantly expressed ThermoDBP-RP1 or 2 from *Aeropyrum pernix* and *Pyrococcus furiosus*, respectively, only showed a weak, salt-sensitive interaction with ribosomal particles from *Thermococcus kodakarensis* or its ribosomal subunits, indicating rather unspecific binding of the proteins. Further, binding of *pfu*ThermoDBP-RP1 to box C/D sRNAs *in vitro* was efficiently competed by the presence of *E. coli* tRNA also indicating sequence unspecific binding.

The apo-crystal structure of the *Pyrococcus furiosus* ThermoDBP-RP1 (*pfu*ThermoDBP-RP1) presented in our paper revealed that the orientation of the N-terminal oligonucleotide-binding domain within the *pfu*ThermoDBP-RP1 homotetramer, relative to the C-terminus, does not allow oligonucleotide binding, as shown by comparison with the crystal structure of the truncated NTD from the homologous ThermoDBP-RP1 protein from *Thermoproteus tenax* (*tte*ThermoDBP-RP1, Ttx1576) (Paytubi et al., 2012). Therefore, it can be speculated that the C-terminus of *pfu*ThermoDBP-RP1 exhibits an auto-inhibitory function that blocks the nucleic acid binding site. Additional factors interacting with *pfu*ThermoDBP-RP1 protein might be necessary to induce a structural rearrangement to expose the oligonucleotide-binding surface of the protein to allow amino acid binding.

The previously found interaction of *S. solfa* ThermoDBP-RP1's (Rbp18) with box C/D sRNA suggests it to be part of an RNP complex (Ciammaruconi et al., 2008). Thus, binding to the RNP complex might provide additional interactions necessary to induce an active conformation of the protein, but no interaction with proteins involved in RNP complex formation (Nop56, Nop58, fibrillarin) could be confirmed (Ciammaruconi et al., 2008). However, it should be mentioned that aL7ae, a protein known to interact with the RNP complex, could also not be found interacting with components of the RNP complex in the experimental setup (Ciammaruconi et al., 2008). Therefore, taken together, it still remains unclear whether ThermoDBP-RP1 proteins can functionally interact with box C/D sRNAs, RNPs or the ribosome.

Expression and purification of the *pfu*ThermoDBP-RP1 NTD alone was not feasible, but based on the crystal structure, different C-terminal truncated mutants and mutants defective

in tetramer formation of *pfu*ThermoDBP-RP1 could be designed to verify the postulated auto inhibition and to identify *pfu*ThermoDBP-RP1 substrates.

In contrast to *pfu*ThermoDBP-RP1, we could show that *Aeropyrum pernix* ThermoDBP-RP2 (*ape*ThermoDBP-RP2) protein binds ssDNA with high selectivity over ssRNA, dsDNA or dsRNA *in vitro*. Further, we found that *ape*ThermoDBP-RP2 binds to circular ssDNA lacking free termini, suggesting that *ape*ThermoDBP-RP2 could act as a SSB for example involved in recombination or DNA repair. The crystal structure of *ape*ThermoDBP-RP2 we present revealed that nucleic acid binds to a symmetric *ape*ThermoDBP-RP2 homotetramer in a unique way. While other known SSBs bind ssDNA on their surface or along a channel in the oligonucleotide-binding domain, ssDNA bound to *ape*ThermoDBP-RP2 is found to traverse an internal, intersubunit tunnel system of the ThermoDBP-RP2 homotetramer. These findings indicate a function of ThermoDBP-RP2 to maintain and to protect ssDNA in its single-stranded state from thermal and other types of stress in the particular challenging living conditions of many archaea. This is also explaining the requirement for such specialized SSBs to have evolved in extremophile archaea species. Although, while these results suggest that ThermoDBP-RP2 proteins can function as SSBs *in vitro*, a similar activity *in vivo* remains to be demonstrated.

#### 10.4 Paper 4 | Structure of the *Bacillus subtilis* 70S ribosome reveals the basis for species-specific stalling

Daniel Sohmen, Shinobu Chiba, Naomi Shimokawa-Chiba, C. Axel Innis,  
Otto Berninghausen, Roland Beckmann, Koreaki Ito & Daniel N. Wilson

*Nature Communications* 6, 6941 (2015)

In contrast to other stalling sequences, amino acids critical for stalling of MifM reside at positions starting from -8 to further upstream of the first arrest site (Chiba et al., 2009). In the stalled complex, the invariant N81 and M80 of MifM, at the positions -8 and -9, respectively, localize at the constriction in between the tip of the  $\beta$ -hairpin of L22, R66 of L4 and A751 of the 23S rRNA, suggesting such interaction with the constriction to make a major contribution to the peptide arrest. Consistent with this, we find the electron density for the nascent chain at the constriction to be well resolved and the contact to A751 to be one of the strongest fusions of density between the tunnel wall and MifM. M90, localized at the tip of the  $\beta$ -hairpin of L22, has



been identified as being of critical importance for MifM stalling. In addition, it is considered a key determinant for species specificity, albeit contacts at the tip of the  $\beta$ -hairpin with the 23S rRNA are conserved between *E. coli* and *B. subtilis* (besides a possible M-sulfur aromatic interaction between the side chain of M90 and G748 of the 23S rRNA). Furthermore, no direct contact between MifM and L22 is observed at the constriction, suggesting that the identity of the amino acid at position 90 of L22 indirectly influences the interaction of MifM with A751, possibly through its interaction with G748 or via backbone interaction with the phosphate group of A751. Interaction with the rRNA at the constriction, especially with the 750-loop, has been shown previously to be important for stalling translation of *E. coli* SecM and TnaC: Nucleotide insertion into the 750-loop of the 23S rRNA eliminates translation arrest on both peptides (L. Cruz-Vera, Rajagopal, Squires, & Yanofsky, 2005; L. R. Cruz-Vera, New, Squires, & Yanofsky, 2007; Nakatogawa & Ito, 2002) and mutation of A752 to C or T reduces stalling on TnaC (L. R. Cruz-Vera et al., 2007). Bischoff *et al.* (Bischoff et al., 2014) found a coordinated tryptophan close to the constriction, in concert with the nascent chain, to stabilize an A752 base pair interaction with U2609 as part of a proposed potential relay mechanism involved in PTC inactivation. Prospective mutation studies of the 750-loop region in *B. subtilis* could help to further elucidate the details of the interaction of the  $\beta$ -hairpin of L22 with the 750-loop region of the 23S rRNA and its particular contribution to stalling. Thus, consistent with MifM interacting with the ribosome in a species-specific manner (Chiba et al., 2011), stalling with high efficiency only on ribosomes of the homologous species, *B. subtilis*, replacing the  $\beta$ -hairpin of *B. subtilis* L22 with the corresponding sequence from *E. coli* L22 led to a drastic reduction in stalling efficiency of MifM. Solely reverting *E. coli* K90 back to M90 as in *B. subtilis* L22 almost completely restored stalling efficiency of MifM, disclosing it contributing to the species specificity of MifM stalling. On the contrary, it is still unknown whether mutation of position 90 in *E. coli* L22, or replacing the  $\beta$ -hairpin of *E. coli* L22 with the corresponding sequence from *B. subtilis* L22 can alter the stalling efficiency of *E. coli* ribosomes on MifM. Such further mutational studies of chimeric ribosomes on a heterologous species could reveal further determinants of the species specificity and its molecular mechanism.

The finding that amino acids critical for MifM elongation arrest are only found further upstream of the PTC is accompanied by the finding that translation on MifM can stall the ribosome at more than one stall site. Ribosomes stalled from D89 can spontaneously resume elongation by one amino acid to the next stall site (Chiba & Ito, 2012). This unique mode of multisite stalling, spanning four consecutive stall sites ( $_{89}\text{DAGS}_{92}$ ), has been shown to be dependent on the presence of a conserved cluster of acidic amino acids ( $_{86}\text{DEED}_{89}$ ) including the first stall site. However, multiple residues within the DEED motif have to be converted to A

to abolish stalling completely (Chiba & Ito, 2012), displaying the functional plasticity of the region including E87.

Inspection of our cryo-EM structure revealed the presence of a post-initiation SD-like helix (Yusupova, Jenner, Rees, Moras, & Yusupov, 2006) formed between the MifM mRNA and the 5' end of the 16S rRNA. Internal SD-like sequences within the coding region of mRNAs have been reported before to induce translational pausing in bacteria (Li, Oh, & Weissman, 2012). Since breaking of the SD-like helix is slow, such interaction within the MifM stalled ribosome structure should be discussed as a possible contributor to regulation of multisite stalling. It was also previously shown that a single nucleotide insertion within the MifM leader peptide mRNA, which introduces a frame shift from before the amino acid at position -12, and thereby abrogates critical interactions of MifM with the ribosome at the constriction, abolishes translational stalling (Chiba et al., 2009). This finding indicates that the presence of the SD-like helix alone is insufficient to induce pausing. To test the contribution of the observed 16S rRNA/mRNA interaction to the regulation of multisite stalling of MifM, mutations within the mRNA from three codons upstream of the D89 stall site codon in the P-site, altering the 16S rRNA/mRNA interaction without abrogating critical interactions of MifM with the ribosome, could be a starting point to separate the contribution of the SD-like helix to the regulation of multisite stalling.

Since the aforementioned multiple translational stalling at four different sites would lead to non-homogeneity of the sample and therefore pose a limit to obtain high resolution in cryo-EM, the MifM ribosomal complex in our structure was designed to be homogeneously stalled at D89. Here, E87 has been found in a position incompatible with the induced state of the ribosome, preventing A-tRNA accommodation. To allow A-tRNA accommodation U2585 (*E. coli* numbering is used) has to rotate by 19°. Together with U2585, U2506 has to shift to avoid a steric clash with U2585. The side chain of E87 is found in a position that is incompatible with both movements, thereby interfering with the induced state of the PTC and consequently preventing A-tRNA accommodation. In contrast, in the PTC of the high-resolution cryo-EM structure of the TnaC stalled ribosome (Bischoff et al., 2014) nucleotides U2585 and A2602 adopt conformations incompatible with productive binding of RF2, while an aminoacyl-tRNA in the A-site would not clash with the observed conformation of U2585 or A2602. This interpretation of the structure is thereby corroborated by the finding that placing a sense codon in the A-site can alleviate TnaC stalling (L. R. Cruz-Vera, Gong, & Yanofsky, 2006; Martinez et al., 2014). Within the structure of the MifM ribosomal complex, nucleotide A2602 is also found in a position incompatible with productive docking of the release factors as well as A-tRNA accommodation, thereby explaining the previously observed behaviour of the MifM-

stalled ribosomal complexes to inhibit translation at the step of termination as well as peptide bond formation (Chiba & Ito, 2012). Further, A2602 seems to be one of the major determinants of the ErmCL stalled ribosome interfering with A-tRNA accommodation, while the positions of nucleotides U2585 or U2506 in the ErmCL stalled ribosome do not obviously interfere with A-tRNA binding (although U2585 is definitely found in an extreme position) (Arenz, Meydan, et al., 2014).

With respect to multisite stalling in MifM, it can be speculated that in an E87A mutant background the role of E87 in blocking the induced state of the PTC would be resumed by the following E88, stalling the ribosome on the consecutive stall site. This scenario could indeed explain multisite stalling for the first three consecutive stall sites, assuming a compaction of the nascent chain in the upper part of the tunnel. While for stalling at S92, the side chain of A90 at the -2 position is unlikely to interfere with the induced state of the PTC and A-tRNA accommodation. However, without determining further structures of individual mutant MifM-SRCs, it will be difficult to validate this theoretical scenario. Interestingly, it has been demonstrated that the bacterial secretion monitor (SecM) and the eukaryotic arginine attenuator peptide (AAP) also have the unusual property of interfering with PTC function, stalling ribosomal translation across a window spanning at least four consecutive codons (Tsai, Kornberg, Johansson, Chen, & Puglisi, 2014; Wei, Wu, & Sachs, 2012). However, why arrest peptides with multiple consecutive stall sites have evolved still poses questions. It can be speculated that RAPs might use multisite arrest as a strategy to attain a stalling duration sufficient for the target gene expression and therefore include time as another variable for translation regulation.

The rapidly developing method of cryo-EM has, especially lately, enormously contributed to the understanding of several arrest-peptides and their underlying molecular stalling mechanisms. Even though our cryo-EM structure, supported by biochemical data, can already provide an insightful basis for comprehending the MifM stalled ribosome and its regulation, we can only reveal a snapshot of the stalling event. Why nature has chosen such mode of regulation for the membrane protein biogenesis factor YidC2 and therefore high efficiency stalling in a species-specific manner still requires further investigation. Moreover, the extend to which the SD interaction contributes to translational stalling or if stalling is induced similarly at the three additional stalling sites are only a small selection of the questions which remain to be elucidated.

## 10.5 Paper 5 | Structural basis for targeting and elongation arrest of *Bacillus* signal recognition particle

Bertrand Beckert, Alexej Kedrov, Daniel Sohmen, Georg Kempf, Klemens Wild, Irmgard Sinning, Henning Stahlberg, Daniel N. Wilson and Roland Beckmann

*Unpublished manuscript* (2015)

The presented 3D reconstruction of the *B. sub* SRP-RNC “targeting complex” from cryo-EM data, collected using a conventional CCD detector resulted in a reconstruction of  $\sim 7\text{\AA}$  resolution. The volume shows continuous ligand density that allowed the complete molecular model of the *B. sub* SRP to be built. However, further attempts to obtain a higher resolved structure using data collected on K2 Summit and Falcon 2 direct electron detectors did not reveal any further details of the ligand. Instead, only weakly represented density for SRP was present in the reconstructions, although the overall resolution of the ribosome could be improved by  $\sim 2\text{\AA}$ . The weak density for SRP could also not be attributed to different sub-states of the ligand, as further computational sorting failed, suggesting an intrinsic flexibility of ribosome bound SRP. Inspection of the density and local resolution determination of the structure further revealed that within the structure the local resolution for the S domain is lower (reflecting its higher flexibility) than for the Alu domain, although high affinity binding to the ribosomal complex has been shown to be dependent on interactions with the S domain (Lutcke, High, Romisch, Ashford, & Dobberstein, 1992; Romisch, Webb, Lingelbach, Gausepohl, & Dobberstein, 1990). Flexibility of the S domain of SRP has already been observed before in cryo-EM reconstructions of the SRP receptor-SRP “docking complex” (Estrozi, Boehringer, Shan, Ban, & Schaffitzel, 2011; Halic, Gartmann, et al., 2006; von Loeffelholz et al., 2015) as well as for the free state of the SRP core (Rosendal, Wild, Montoya, & Sinning, 2003). The flexibility of the SRP S domain is thereby provided via a flexible hinge region in the SRP RNA and the flexible domain architecture of Ffh (Rosendal et al., 2003), uncoupling Alu and S domain. It can be assumed that the flexibility of the S domain is important for SRP function, enabling SRP to interact with a variety of substrates (Noriega et al., 2014; Walter, Ibrahimi, & Blobel, 1981) and to adapt to structural changes of the complex when adopting different states in the SRP cycle to successfully engage the translocon (Halic & Beckmann, 2005). Additional high-resolution X-ray and cryo-EM structures of the SRP-RNC complex in different functional states should provide the missing details of the structural rearrangements within SRP and the ribosome during the SRP cycle.

The SRP Alu domain, upon signal sequence recognition, locates in a defined space at the ribosomal subunit interface, thereby blocking the factor binding site, imposing elongation arrest. In higher eukaryotes, the Alu domain comprises the 5' and 3' extremities of the SRP RNA bound to the SRP9/14 heterodimer. SRP9/14 stabilize the complex fold of the Alu RNA and contribute to ribosome binding by interacting with the small ribosomal subunit (Halic et al., 2004; Weichenrieder et al., 2001; Weichenrieder, Wild, Strub, & Cusack, 2000). However, homologues of SRP9/14 have not been found in archaeal or bacterial (Andersen et al., 2006). The DNA-binding protein HU1 has been suggested to interact with the *B. sub* Alu domain (Nakamura, Yahagi, Yamazaki, & Yamane, 1999; Yamazaki, Yahagi, Nakamura, & Yamane, 1999). HU1 was found to bind to *B. sub* SRP RNA *in vitro* and was also found associated with the SRP RNA after immunoprecipitation of HU1 from cell lysates (Nakamura et al., 1999). Further, its functional interaction with the protein sorting machinery was suggested: Depletion of HU1 from the cell led to the upregulation of the intracellular SRP RNA level and to accumulation of secretory protein precursors indicating defective translocation (Yamazaki et al., 1999). Yet, in contrast to the previously published results, in our case HU1 did neither bind to *B. sub* SRP RNA during *in vitro* reconstitution, nor was it found in *in vivo* pullouts. Consequently, we do not consider HU1 to be part of the *B. sub* SRP. Further support for this assumption comes from the recently solved crystal structures of the complete *B. subtilis* Alu domain (Kempf et al., 2014) and the chimeric Alu domain of the archaea *Pyrococcus hirokoshii* together with human SRP9/14 (Bousset et al., 2014). The detailed interpretations of both structures showed that extra RNA sequences at the SRP RNA 5' and 3' ends form an additional helix (H1) in the Alu domain, which stabilizes the closed conformation of the Alu domain instead of SRP9/14. Further, the *B. sub* SRP was found fully folded and stably bound to the ribosome in our reconstruction. Noteworthy, the human SRP9/14 heterodimer was still able to bind to the *Pyrococcus hirokoshii* Alu domain in an essentially identical manner as observed previously in the structure of the human Alu domain (Weichenrieder et al., 2000). Our structure revealed that the *B. sub* SRP Alu domain binds to the ribosome in a similar manner to the mammalian Alu domain, blocking the elongation factor binding site. However, the *B. sub* Alu domain interacts with the ribosome via an entirely different set of interactions. The mammalian Alu domain interacts via the Alu RNA with uL11 on top of the stalk base on the large subunit and with h4/15 of the small subunit via SRP9/14, while the *B. sub* SRP Alu domain interacts with the ribosome mainly via RNA:RNA interactions with H43/H44 of the stalk base and the  $\alpha$ -sarcin-ricin loop (H95) of the large subunit. Therefore, although the fold and the structure of the Alu RNA have been conserved, it appears that the mammalian and the bacterial Alu domain have developed distinctive ways of interacting with the ribosome. Especially the interaction

with the small subunit via SRP9/14 found in the mammalian SRP might provide additional control over elongation factor independent, intrinsic intersubunit movements of the ribosome. A posing question that could be addressed in future studies using single particle FRET or related techniques.

Interestingly, the common Alu domain RNA core resembles the structure of the canonical tRNA (Chang, Newitt, Hsu, Bernstein, & Maraia, 1997) and in several protozoa such as trypanosomes, a tRNA-like RNA has been found to be associated with a significantly reduced Alu domain suggesting a potential general evolutionary relationship of the Alu domain with tRNA (Beja, Ullu, & Michaeli, 1993; L. Liu et al., 2003). This could be one possible explanation why bacterial SRP interacts with the ribosome almost exclusively via RNA:RNA contacts. However, evolution and interaction of the Alu domain with the ribosome are still only poorly understood. A number of SRPs with phylogenetically shared divergent features and of variable composition have been identified by genomic analysis in all domains of life (Rosenblad et al., 2009; Zwieb & Bhuiyan, 2010). Additional reconstructions of distinct SRP-RNC complexes will help to better understand the evolution of the SRP-RNC interaction and its function in protein targeting.

## 11 References Publications

- Achenbach, J., & Nierhaus, K. H. (2014). The mechanics of ribosomal translocation. *Biochimie*. doi: 10.1016/j.biochi.2014.12.003
- Agrawal, R., Sharma, M., Kiel, M., Hirokawa, G., Booth, T., Spahn, C., . . . Frank, J. (2004). Visualization of ribosome-recycling factor on the *Escherichia coli* 70S ribosome: Functional implications. *Proc. Natl Acad. Sci. USA*, *101*(24), 8900–8905.
- Akopian, D., Shen, K., Zhang, X., & Shan, S. O. (2013). Signal recognition particle: an essential protein-targeting machine. *Annu Rev Biochem*, *82*, 693-721. doi: 10.1146/annurev-biochem-072711-164732
- Allen, G., Zavialov, A., Gursky, R., Ehrenberg, M., & Frank, J. (2005). The cryo-EM structure of a translation initiation complex from *Escherichia coli*. *Cell*, *121*(5), 703-712.
- Anborgh, P. H., & Parmeggiani, A. (1991). New Antibiotic That Acts Specifically on the GTP-Bound Form of Elongation Factor-Tu. *EMBO J*, *10*(4), 779-784.
- Andersen, E. S., Rosenblad, M. A., Larsen, N., Westergaard, J. C., Burks, J., Wower, I. K., . . . Zwieb, C. (2006). The tmRDB and SRPDB resources. *Nucleic Acids Res*, *34*(Database issue), D163-168. doi: 10.1093/nar/gkj142
- Anger, A. M., Armache, J. P., Berninghausen, O., Habeck, M., Subklewe, M., Wilson, D. N., & Beckmann, R. (2013). Structures of the human and *Drosophila* 80S ribosome. *Nature*, *497*(7447), 80-85.
- Arenz, S., Meydan, S., Starosta, A. L., Berninghausen, O., Beckmann, R., Vazquez-Laslop, N., & Wilson, D. N. (2014). Drug sensing by the ribosome induces translational arrest via active site perturbation. *Mol Cell*, *56*(3), 446-452. doi: 10.1016/j.molcel.2014.09.014
- Arenz, S., Ramu, H., Gupta, P., Berninghausen, O., Beckmann, R., Vazquez-Laslop, N., . . . Wilson, D. N. (2014). Molecular basis for erythromycin-dependent ribosome stalling during translation of the ErmBL leader peptide. *Nat Commun*, *5*, 3501. doi: 10.1038/ncomms4501
- Armache, J. P., Anger, A. M., Marquez, V., Franckenberg, S., Frohlich, T., Villa, E., . . . Wilson, D. N. (2013). Promiscuous behaviour of archaeal ribosomal proteins: Implications for eukaryotic ribosome evolution. *Nucleic Acids Res*, *41*(2), 1284-1293.
- Atkins, J. F., Wills, N. M., Loughran, G., Wu, C. Y., Parsawar, K., Ryan, M. D., . . . Nelson, C. C. (2007). A case for "StopGo": reprogramming translation to augment codon meaning of GGN by promoting unconventional termination (Stop) after addition of glycine and then allowing continued translation (Go). *RNA*, *13*(6), 803-810. doi: 10.1261/rna.487907
- Batey, R. T., Rambo, R. P., Lucast, L., Rha, B., & Doudna, J. A. (2000). Crystal structure of the ribonucleoprotein core of the signal recognition particle. *Science*, *287*(5456), 1232-1239.
- Becker, T., Franckenberg, S., Wickles, S., Shoemaker, C. J., Anger, A. M., Armache, J. P., . . . Beckmann, R. (2012). Structural basis of highly conserved ribosome recycling in eukaryotes and archaea. *Nature*, *482*(7386), 501-506.
- Beja, O., Ullu, E., & Michaeli, S. (1993). Identification of a tRNA-like molecule that copurifies with the 7SL RNA of *Trypanosoma brucei*. *Mol Biochem Parasitol*, *57*(2), 223-229.
- Bhushan, S., Hoffmann, T., Seidelt, B., Frauenfeld, J., Mielke, T., Berninghausen, O., . . . Beckmann, R. (2011). SecM-stalled ribosomes adopt an altered geometry at the peptidyltransferase center. *PLoS Biol.*, *19*(1), e1000581.
- Bhushan, S., Meyer, H., Starosta, A. L., Becker, T., Mielke, T., Berninghausen, O., . . . Beckmann, R. (2010). Structural basis for translational stalling by human

- cytomegalovirus (hCMV) and fungal arginine attenuator peptide (AAP). *Mol. Cell*, 40(1), 138-146.
- Bischoff, L., Berninghausen, O., & Beckmann, R. (2014). Molecular basis for the ribosome functioning as an L-tryptophan sensor. *Cell Rep*, 9(2), 469-475. doi: 10.1016/j.celrep.2014.09.011
- Blanchard, S. C., Gonzalez, R. L., Kim, H. D., Chu, S., & Puglisi, J. D. (2004). tRNA selection and kinetic proofreading in translation. *Nat. Struct. Mol. Biol.*, 11(10), 1008-1014.
- Bolla, J. M., Alibert-Franco, S., Handzlik, J., Chevalier, J., Mahamoud, A., Boyer, G., . . . Pages, J. M. (2011). Strategies for bypassing the membrane barrier in multidrug resistant Gram-negative bacteria. *FEBS Lett*, 585(11), 1682-1690. doi: 10.1016/j.febslet.2011.04.054
- Borovinskaya, M. A., Pai, R. D., Zhang, W., Schuwirth, B. S., Holton, J. M., Hirokawa, G., . . . Cate, J. H. (2007). Structural basis for aminoglycoside inhibition of bacterial ribosome recycling. *Nat Struct Mol Biol*, 14(8), 727-732.
- Bousset, L., Mary, C., Brooks, M. A., Scherrer, A., Strub, K., & Cusack, S. (2014). Crystal structure of a signal recognition particle Alu domain in the elongation arrest conformation. *RNA*, 20(12), 1955-1962. doi: 10.1261/rna.047209.114
- Brilot, A. F., Korostelev, A. A., Ermolenko, D. N., & Grigorieff, N. (2013). Structure of the ribosome with elongation factor G trapped in the pretranslocation state. *Proc Natl Acad Sci U S A*, 110(52), 20994-20999. doi: 10.1073/pnas.1311423110
- Butkus, M. E., Prundeanu, L. B., & Oliver, D. B. (2003). Translocon "pulling" of nascent SecM controls the duration of its translational pause and secretion-responsive secA regulation. *J. Bacteriol.*, 185(22), 6719-6722.
- Cao, J. H., & Geballe, A. P. (1996). Coding sequence-dependent ribosomal arrest at termination of translation. *Mol Cell Biol*, 16(2), 603-608.
- Carter, A. P., Clemons, W. M., Brodersen, D. E., Morgan-Warren, R. J., Wimberly, B. T., & Ramakrishnan, V. (2000). Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics. *Nature*, 407(6802), 340-348.
- Chang, D. Y., Newitt, J. A., Hsu, K., Bernstein, H. D., & Maraja, R. J. (1997). A highly conserved nucleotide in the Alu domain of SRP RNA mediates translation arrest through high affinity binding to SRP9/14. *Nucleic Acids Res*, 25(6), 1117-1122.
- Chatzi, K. E., Sardis, M. F., Economou, A., & Karamanou, S. (2014). SecA-mediated targeting and translocation of secretory proteins. *Biochim Biophys Acta*, 1843(8), 1466-1474. doi: 10.1016/j.bbamcr.2014.02.014
- Chen, Y., Feng, S., Kumar, V., Ero, R., & Gao, Y. G. (2013). Structure of EF-G-ribosome complex in a pretranslocation state. *Nat Struct Mol Biol*, 20(9), 1077-1084. doi: 10.1038/nsmb.2645
- Chiba, S., & Ito, K. (2012). Multisite Ribosomal Stalling: A Unique Mode of Regulatory Nascent Chain Action Revealed for MifM. *Mol. Cell*.
- Chiba, S., Kanamori, T., Ueda, T., Akiyama, Y., Pogliano, K., & Ito, K. (2011). Recruitment of a species-specific translational arrest module to monitor different cellular processes. *Proc. Natl Acad. Sci. USA*, 108(15), 6073-6078.
- Chiba, S., Lamsa, A., & Pogliano, K. (2009). A ribosome-nascent chain sensor of membrane protein biogenesis in *Bacillus subtilis*. *EMBO J.*, 28(22), 3461-3475.
- Ciammaruconi, A., Gorini, S., & Londei, P. (2008). A bifunctional archaeal protein that is a component of 30S ribosomal subunits and interacts with C/D box small RNAs. *Archaea*, 2(3), 151-158.
- Cruz-Vera, L., Rajagopal, S., Squires, C., & Yanofsky, C. (2005). Features of ribosome-peptidyl-tRNA interactions essential for tryptophan induction of tna operon expression. *Mol. Cell*, 19(3), 333-343.
- Cruz-Vera, L. R., Gong, M., & Yanofsky, C. (2006). Changes produced by bound tryptophan in the ribosome peptidyl transferase center in response to TnaC, a nascent leader peptide. *Proc Natl Acad Sci U S A*, 103(10), 3598-3603.



- Cruz-Vera, L. R., New, A., Squires, C., & Yanofsky, C. (2007). Ribosomal features essential for tna operon induction: tryptophan binding at the peptidyl transferase center. *J. Bacteriol.*, *189*(8), 3140-3146.
- Cruz-Vera, L. R., Yang, R., & Yanofsky, C. (2009). Tryptophan inhibits *Proteus vulgaris* TnaC leader peptide elongation, activating tna operon expression. *J Bacteriol*, *191*(22), 7001-7006. doi: 10.1128/JB.01002-09
- Cymer, F., Hedman, R., Ismail, N., & von Heijne, G. (2015). Exploration of the Arrest Peptide Sequence Space Reveals Arrest-enhanced Variants. *J Biol Chem*, *290*(16), 10208-10215. doi: 10.1074/jbc.M115.641555
- Cymer, F., Ismail, N., & von Heijne, G. (2014). Weak pulling forces exerted on Nin-orientated transmembrane segments during co-translational insertion into the inner membrane of *Escherichia coli*. *FEBS Lett*, *588*(10), 1930-1934. doi: 10.1016/j.febslet.2014.03.050
- Cymer, F., von Heijne, G., & White, S. H. (2015). Mechanisms of integral membrane protein insertion and folding. *J Mol Biol*, *427*(5), 999-1022. doi: 10.1016/j.jmb.2014.09.014
- Degnin, C. R., Schleiss, M. R., Cao, J., & Geballe, A. P. (1993). Translational inhibition mediated by a short upstream open reading frame in the human cytomegalovirus gpUL4 (gp48) transcript. *J. Virol.*, *67*(9), 5514-5521.
- Demeshkina, N., Jenner, L., Westhof, E., Yusupov, M., & Yusupova, G. (2012). A new understanding of the decoding principle on the ribosome. *Nature*, *484*(7393), 256-259.
- Demeshkina, N., Jenner, L., Westhof, E., Yusupov, M., & Yusupova, G. (2013). New structural insights into the decoding mechanism: Translation infidelity via a G.U pair with Watson-Crick geometry. *FEBS Lett*.
- Desmond, E., Brochier-Armanet, C., Forterre, P., & Gribaldo, S. (2010). On the last common ancestor and early evolution of eukaryotes: reconstructing the history of mitochondrial ribosomes. *Res. Microbiol.*, *162*(1), 53-70.
- Doi, Y., Ohtsuki, T., Shimizu, Y., Ueda, T., & Sisido, M. (2007). Elongation factor Tu mutants expand amino acid tolerance of protein biosynthesis system. *J Am Chem Soc*, *129*(46), 14458-14462. doi: 10.1021/ja075557u
- Dorner, S., Panuschka, C., Schmid, W., & Barta, A. (2003). Mononucleotide derivatives as ribosomal P-site substrates reveal an important contribution of the 2'-OH to activity. *Nucleic Acids Res*, *31*(22), 6536-6542.
- Dubnau, D. (1985). Induction of ermC requires translation of the leader peptide. *EMBO J*, *4*(2), 533-537.
- Dunkle, J. A., & Cate, J. H. (2010). Ribosome structure and dynamics during translocation and termination. *Annu Rev Biophys*, *39*, 227-244. doi: 10.1146/annurev.biophys.37.032807.125954
- Dunkle, J. A., Wang, L., Feldman, M. B., Pulk, A., Chen, V. B., Kapral, G. J., . . . Cate, J. H. (2011). Structures of the bacterial ribosome in classical and hybrid states of tRNA binding. *Science*, *332*(6032), 981-984.
- Erlacher, M. D., Lang, K., Shankaran, N., Wotzel, B., Huttenhofer, A., Micura, R., . . . Polacek, N. (2005). Chemical engineering of the peptidyl transferase center reveals an important role of the 2'-hydroxyl group of A2451. *Nucleic Acids Res.*, *33*(5), 1618-1627.
- Ermolenko, D. N., & Noller, H. F. (2011). mRNA translocation occurs during the second step of ribosomal intersubunit rotation. *Nat Struct Mol Biol*, *18*(4), 457-462. doi: 10.1038/nsmb.2011
- Errington, J., Appleby, L., Daniel, R. A., Goodfellow, H., Partridge, S. R., & Yudkin, M. D. (1992). Structure and function of the spoIIIJ gene of *Bacillus subtilis*: a vegetatively expressed gene that is essential for sigma G activity at an intermediate stage of sporulation. *J Gen Microbiol*, *138*(12), 2609-2618.
- Estrozi, L. F., Boehringer, D., Shan, S. O., Ban, N., & Schaffitzel, C. (2011). Cryo-EM structure of the *E. coli* translating ribosome in complex with SRP and its receptor. *Nat Struct Mol Biol*, *18*(1), 88-90. doi: 10.1038/nsmb.1952

- Feng, Y., Song, X., Lin, J., Xuan, J., Cui, Q., & Wang, J. (2014). Structure determination of archaea-specific ribosomal protein L46a reveals a novel protein fold. *Biochem Biophys Res Commun*, 450(1), 67-72. doi: 10.1016/j.bbrc.2014.05.077
- Fischbach, M. A., & Walsh, C. T. (2009). Antibiotics for emerging pathogens. *Science*, 325(5944), 1089-1093.
- Forterre, P. (2013). The common ancestor of archaea and eukarya was not an archaeon. *Archaea*, 2013, 372396. doi: 10.1155/2013/372396
- Frank, J., & Agrawal, R. K. (2000). A ratchet-like inter-subunit reorganization of the ribosome during translocation. *Nature*, 406, 318-322.
- Frank, J., Jr., & Gonzalez, R. L. (2010). Structure and dynamics of a processive Brownian motor: the translating ribosome. *Annu Rev Biochem*, 79, 381-412.
- Fujita, K., Baba, T., & Isono, K. (1998). Genomic Analysis of the Genes Encoding Ribosomal Proteins in Eight Eubacterial Species and *Saccharomyces cerevisiae*. *Genome Inform Ser Workshop Genome Inform*, 9, 3-12.
- Funes, S., Hasona, A., Bauerschmitt, H., Grubbauer, C., Kauff, F., Collins, R., . . . Herrmann, J. M. (2009). Independent gene duplications of the YidC/Oxa/Alb3 family enabled a specialized cotranslational function. *Proc Natl Acad Sci U S A*, 106(16), 6656-6661. doi: 10.1073/pnas.0809951106
- Funes, S., Kauff, F., van der Sluis, E. O., Ott, M., & Herrmann, J. M. (2011). Evolution of YidC/Oxa1/Alb3 insertases: three independent gene duplications followed by functional specialization in bacteria, mitochondria and chloroplasts. *Biol Chem*, 392(1-2), 13-19. doi: 10.1515/BC.2011.013
- Gaba, A., Jacobson, A., & Sachs, M. S. (2005). Ribosome occupancy of the yeast CPA1 upstream open reading frame termination codon modulates nonsense-mediated mRNA decay. *Mol Cell*, 20(3), 449-460. doi: 10.1016/j.molcel.2005.09.019
- Gabdulkhakov, A., Nikonov, S., & Garber, M. (2013). Revisiting the Haloarcula marismortui 50S ribosomal subunit model. *Acta Crystallogr D Biol Crystallogr*, 69(Pt 6), 997-1004. doi: 10.1107/S0907444913004745
- Gao, H., Zhou, Z., Rawat, U., Huang, C., Bouakaz, L., Wang, C., . . . Song, H. (2007). RF3 induces ribosomal conformational changes responsible for dissociation of class I release factors. *Cell*, 129(5), 929-941.
- Gao, N., Zavialov, A. V., Ehrenberg, M., & Frank, J. (2007). Specific interaction between EF-G and RRF and its implication for GTP-dependent ribosome splitting into subunits. *J Mol Biol*, 374(5), 1345-1358. doi: 10.1016/j.jmb.2007.10.021
- Gao, N., Zavialov, A. V., Li, W., Sengupta, J., Valle, M., Gursky, R. P., . . . Frank, J. (2005). Mechanism for the disassembly of the posttermination complex inferred from cryo-EM studies. *Mol. Cell*, 18(6), 663-674.
- Gavrilova, L. P., & Spirin, A. S. (1971). Stimulation of 'non-enzymic' translocation in ribosome by p-chloromercuribenzoate. *FEBS Lett.*, 17(2), 324-326.
- Gollnick, P., & Yanofsky, C. (1990). Transfer Rnatrp Translation of Leader Peptide Codon-12 and Other Factors That Regulate Expression of the Tryptophanase Operon. *Journal Of Bacteriology*, 172(6), 3100-3107.
- Gong, F., & Yanofsky, C. (2002). Instruction of translating ribosome by nascent peptide. *Science*, 297(5588), 1864-1867.
- Gong, F., & Yanofsky, C. (2003). A transcriptional pause synchronizes translation with transcription in the tryptophanase operon leader region. *J Bacteriol*, 185(21), 6472-6476.
- Grentzmann, G., Brechemierbaey, D., Heurgue, V., Mora, L., & Buckingham, R. H. (1994). Localization and characterization of the gene encoding release factor RF3 in *Escherichia coli*. *Proc Natl Acad Sci USA*, 91(13), 5848-5852.
- Grigoriadou, C., Marzi, S., Kirillov, S., Gualerzi, C. O., & Cooperman, B. S. (2007). A quantitative kinetic scheme for 70 S translation initiation complex formation. *J Mol Biol*, 373(3), 562-572.

- Grigoriadou, C., Marzi, S., Pan, D., Gualerzi, C. O., & Cooperman, B. S. (2007). The translational fidelity function of IF3 during transition from the 30 S initiation complex to the 70 S initiation complex. *J. Mol. Biol.*, *373*(3), 551-561.
- Gromadski, K. B., & Rodnina, M. V. (2004). Kinetic determinants of high-fidelity tRNA discrimination on the ribosome. *Mol Cell*, *13*(2), 191-200.
- Gumbart, J., Schreiner, E., Wilson, D. N., Beckmann, R., & Schulten, K. (2012). Mechanisms of SecM-mediated stalling in the ribosome. *Biophys J*, *103*(2), 331-341. doi: 10.1016/j.bpj.2012.06.005
- Halic, M., Becker, T., Pool, M., Spahn, C., Grassucci, R., Frank, J., & Beckmann, R. (2004). Structure of the signal recognition particle interacting with the elongation-arrested ribosome. *Nature*, *427*(6977), 808-814.
- Halic, M., & Beckmann, R. (2005). The signal recognition particle and its interactions during protein targeting. *Curr. Opin. Struct. Biol.*, *15*(1), 116-125.
- Halic, M., Blau, M., Becker, T., Mielke, T., Pool, M. R., Wild, K., . . . Beckmann, R. (2006). Following the signal sequence from ribosomal tunnel exit to signal recognition particle. *Nature*, *444*(7118), 507-511.
- Halic, M., Gartmann, M., Schlenker, O., Mielke, T., Pool, M. R., Sinning, I., & Beckmann, R. (2006). Signal recognition particle receptor exposes the ribosomal translocon binding site. *Science*, *312*(5774), 745-747.
- Hansen, J. L., Schmeing, T. M., Moore, P. B., & Steitz, T. A. (2002). Structural insights into peptide bond formation. *Proc. Natl. Acad. Sci. USA*, *99*(18), 11670-11675.
- Hill, J. R., & Morris, D. R. (1993). Cell-specific translational regulation of S-adenosylmethionine decarboxylase mRNA. Dependence on translation and coding capacity of the cis-acting upstream open reading frame. *J Biol Chem*, *268*(1), 726-731.
- Hirokawa, G., Nijman, R., Raj, V., Kaji, H., Igarashi, K., & Kaji, A. (2005). The role of ribosome recycling factor in dissociation of 70S ribosomes into subunits. *RNA*, *11*(8), 1317-1128.
- Holtkamp, W., Cunha, C. E., Peske, F., Konevega, A. L., Wintermeyer, W., & Rodnina, M. V. (2014). GTP hydrolysis by EF-G synchronizes tRNA movement on small and large ribosomal subunits. *EMBO J*, *33*(9), 1073-1085. doi: 10.1002/embj.201387465
- Holtkamp, W., Wintermeyer, W., & Rodnina, M. V. (2014). Synchronous tRNA movements during translocation on the ribosome are orchestrated by elongation factor G and GTP hydrolysis. *BioEssays*, *36*(10), 908-918. doi: 10.1002/bies.201400076
- Hopfield, J. J. (1974). Kinetic proofreading: a new mechanism for reducing errors in biosynthesis process requiring high specificity. *Proc. Natl. Acad. Sci. USA*, *71*(10), 4135-4139.
- Hummel, M., Rahmani, F., Smeekens, S., & Hanson, J. (2009). Sucrose-mediated translational control. *Ann Bot*, *104*(1), 1-7. doi: 10.1093/aob/mcp086
- Ismail, N., Hedman, R., Linden, M., & von Heijne, G. (2015). Charge-driven dynamics of nascent-chain movement through the SecYEG translocon. *Nat Struct Mol Biol*, *22*(2), 145-149. doi: 10.1038/nsmb.2940
- Ismail, N., Hedman, R., Schiller, N., & von Heijne, G. (2012). A biphasic pulling force acts on transmembrane helices during translocon-mediated membrane integration. *Nat Struct Mol Biol*, *19*(10), 1018-1022. doi: 10.1038/nsmb.2376
- Ito, K., & Chiba, S. (2013). Arrest peptides: cis-acting modulators of translation. *Annu. Rev. Biochem.*, *82*, 171-202.
- Ito, K., Chiba, S., & Pogliano, K. (2010). Divergent stalling sequences sense and control cellular physiology. *Biochem Biophys Res Commun*, *393*(1), 1-5.
- Ito, K., Uno, M., & Nakamura, Y. (2000). A tripeptide 'anticodon' deciphers stop codons in messenger RNA. *Nature*, *403*(6770), 680-684.
- Janda, C. Y., Li, J., Oubridge, C., Hernandez, H., Robinson, C. V., & Nagai, K. (2010). Recognition of a signal peptide by the signal recognition particle. *Nature*, *465*(7297), 507-510. doi: 10.1038/nature08870

- Jenner, L., Melnikov, S., de Loubresse, N. G., Ben-Shem, A., Iskakova, M., Urzhumtsev, A., . . . Yusupov, M. (2012). Crystal structure of the 80S yeast ribosome. *Curr Opin Struct Biol*, [Epub ahead of print].
- Jiang, F., Chen, M., Yi, L., de Gier, J. W., Kuhn, A., & Dalbey, R. E. (2003). Defining the regions of *Escherichia coli* YidC that contribute to activity. *J Biol Chem*, *278*(49), 48965-48972. doi: 10.1074/jbc.M307362200
- Jin, H., Kelley, A. C., & Ramakrishnan, V. (2011a). Crystal structure of the hybrid state of ribosome in complex with the guanosine triphosphatase release factor 3. *Proc Natl Acad Sci U S A*, *108*(38), 15798-15803. doi: 10.1073/pnas.1112185108
- Jin, H., Kelley, A. C., & Ramakrishnan, V. (2011b). Crystal structure of the hybrid state of ribosome in complex with the guanosine triphosphatase release factor 3. *Proc. Natl Acad. Sci. U S A*, Epub.
- Julian, P., Konevega, A. L., Scheres, S. H., Lazaro, M., Gil, D., Wintermeyer, W., . . . Valle, M. (2008). Structure of ratcheted ribosomes with tRNAs in hybrid states. *Proc Natl Acad Sci U S A*, *105*(44), 16924-16927.
- Julian, P., Milon, P., Agirrezabala, X., Lasso, G., Gil, D., Rodnina, M. V., & Valle, M. (2011). The Cryo-EM structure of a complete 30S translation initiation complex from *Escherichia coli*. *PLoS Biol*, *9*(7), e1001095. doi: 10.1371/journal.pbio.1001095
- Karimi, R., Pavlov, M., Buckingham, R., & Ehrenberg, M. (1999). Novel roles for classical factors at the interface between translation termination and initiation. *Mol. Cell*, *3*, 601-609.
- Kempf, G., Wild, K., & Sinning, I. (2014). Structure of the complete bacterial SRP Alu domain. *Nucleic Acids Res*, *42*(19), 12284-12294. doi: 10.1093/nar/gku883
- Kenmochi, N., Kawaguchi, T., Rozen, S., Davis, E., Goodman, N., Hudson, T. J., . . . Page, D. C. (1998). A map of 75 human ribosomal protein genes. *Genome Res*, *8*(5), 509-523.
- Koch, M., Huang, Y., & Sprinzl, M. (2008). Peptide-bond synthesis on the ribosome: no free vicinal hydroxy group required on the terminal ribose residue of peptidyl-tRNA. *Angew Chem Int Ed Engl*, *47*(38), 7242-7245. doi: 10.1002/anie.200801511
- Kol, S., Nouwen, N., & Driessen, A. J. (2008). Mechanisms of YidC-mediated insertion and assembly of multimeric membrane protein complexes. *J Biol Chem*, *283*(46), 31269-31273. doi: 10.1074/jbc.R800029200
- Konan, K. V., & Yanofsky, C. (1997). Regulation of the *Escherichia coli* tna operon: nascent leader peptide control at the tnaC stop codon. *J Bacteriol*, *179*(5), 1774-1779.
- Konan, K. V., & Yanofsky, C. (2000). Rho-dependent transcription termination in the tna operon of *Escherichia coli*: roles of the boxA sequence and the rut site. *J Bacteriol*, *182*(14), 3981-3988.
- Korostelev, A., Asahara, H., Lancaster, L., Laurberg, M., Hirschi, A., Zhu, J., . . . Noller, H. F. (2008). Crystal structure of a translation termination complex formed with release factor RF2. *Proc Natl Acad Sci U S A*, *105*(50), 19684-19689.
- Korostelev, A., Zhu, J., Asahara, H., & Noller, H. F. (2010). Recognition of the amber UAG stop codon by release factor RF1. *EMBO J*, *29*(15), 2577-2585. doi: 10.1038/emboj.2010.139
- Korostelev, A. A. (2011). Structural aspects of translation termination on the ribosome. *RNA*, *17*(8), 1409-1421. doi: 10.1261/rna.2733411
- Koutmou, K. S., McDonald, M. E., Brunelle, J. L., & Green, R. (2014). RF3:GTP promotes rapid dissociation of the class 1 termination factor. *RNA*, *20*(5), 609-620. doi: 10.1261/rna.042523.113
- Kramer, E. B., & Farabaugh, P. J. (2007). The frequency of translational misreading errors in *E. coli* is largely determined by tRNA competition. *RNA*, *13*(1), 87-96. doi: 10.1261/rna.294907
- Kuhlenkoetter, S., Wintermeyer, W., & Rodnina, M. V. (2011). Different substrate-dependent transition states in the active site of the ribosome. *Nature*, *476*(7360), 351-354. doi: 10.1038/nature10247

- Kumazaki, K., Chiba, S., Takemoto, M., Furukawa, A., Nishiyama, K., Sugano, Y., . . . Nureki, O. (2014). Structural basis of Sec-independent membrane protein insertion by YidC. *Nature*, *509*(7501), 516-520. doi: 10.1038/nature13167
- Kumazaki, K., Kishimoto, T., Furukawa, A., Mori, H., Tanaka, Y., Dohmae, N., . . . Nureki, O. (2014). Crystal structure of Escherichia coli YidC, a membrane protein chaperone and insertase. *Sci Rep*, *4*, 7299. doi: 10.1038/srep07299
- Lancaster, L., Kiel, M. C., Kaji, A., & Noller, H. F. (2002). Orientation of ribosome recycling factor from directed hydroxyl radical probing. *Cell*, *111*, 129-140.
- Larsen, N., & Zwieb, C. (1991). SRP-RNA sequence alignment and secondary structure. *Nucleic Acids Res*, *19*(2), 209-215.
- Laurberg, M., Asahara, H., Korostelev, A., Zhu, J., Trakhanov, S., & Noller, H. F. (2008). Structural basis for translation termination on the 70S ribosome. *Nature*, *454*(7206), 852-857.
- Law, G. L., Raney, A., Heusner, C., & Morris, D. R. (2001). Polyamine regulation of ribosome pausing at the upstream open reading frame of S-adenosylmethionine decarboxylase. *J. Biol. Chem.*, *276*(41), 38036-38043.
- Lecompte, O., Ripp, R., Thierry, J. C., Moras, D., & Poch, O. (2002). Comparative analysis of ribosomal proteins in complete genomes: an example of reductive evolution at the domain scale. *Nucleic Acids Res.*, *30*(24), 5382-5390.
- Li, G. W., Oh, E., & Weissman, J. S. (2012). The anti-Shine-Dalgarno sequence drives translational pausing and codon choice in bacteria. *Nature*, *484*(7395), 538-541.
- Lim, K., Kobayashi, I., & Nakai, K. (2014). Alterations in rRNA-mRNA interaction during plastid evolution. *Mol Biol Evol*, *31*(7), 1728-1740. doi: 10.1093/molbev/msu120
- Lin, J., Gagnon, M. G., Bulkley, D., & Steitz, T. A. (2015). Conformational changes of elongation factor G on the ribosome during tRNA translocation. *Cell*, *160*(1-2), 219-227. doi: 10.1016/j.cell.2014.11.049
- Liu, G., Song, G., Zhang, D., Zhang, D., Li, Z., Lyu, Z., . . . Qin, Y. (2014). EF-G catalyzes tRNA translocation by disrupting interactions between decoding center and codon-anticodon duplex. *Nat Struct Mol Biol*, *21*(9), 817-824. doi: 10.1038/nsmb.2869
- Liu, L., Ben-Shlomo, H., Xu, Y. X., Stern, M. Z., Goncharov, I., Zhang, Y., & Michaeli, S. (2003). The trypanosomatid signal recognition particle consists of two RNA molecules, a 7SL RNA homologue and a novel tRNA-like molecule. *J Biol Chem*, *278*(20), 18271-18280. doi: 10.1074/jbc.M209215200
- Liu, M., & Douthwaite, S. (2002). Activity of the ketolide telithromycin is refractory to Erm monomethylation of bacterial rRNA. *Antimicrob Agents Chemother.*, *46*(6), 1629-1633.
- Lovett, P. S. (1996). Translation attenuation regulation of chloramphenicol resistance in bacteria - A review. *Gene*, *179*(1), 157-162.
- Lovett, P. S., & Rogers, E. J. (1996). Ribosome regulation by the nascent peptide. *Microbiol. Rev.*, *60*(2), 366-385.
- Luo, X., Schwarz-Linek, U., Botting, C. H., Hensel, R., Siebers, B., & White, M. F. (2007). CC1, a novel crenarchaeal DNA binding protein. *J Bacteriol*, *189*(2), 403-409. doi: 10.1128/JB.01246-06
- Luo, Z. L., & Sachs, M. S. (1996). Role of an upstream open reading frame in mediating arginine-specific translational control in Neurospora crassa. *J Bacteriol*, *178*(8), 2172-2177.
- Lutcke, H., High, S., Romisch, K., Ashford, A. J., & Dobberstein, B. (1992). The methionine-rich domain of the 54 kDa subunit of signal recognition particle is sufficient for the interaction with signal sequences. *EMBO J*, *11*(4), 1543-1551.
- Marquez, V., Frohlich, T., Armache, J. P., Sohmen, D., Donhofer, A., Mikolajka, A., . . . Wilson, D. N. (2011). Proteomic characterization of archaeal ribosomes reveals the presence of novel archaeal-specific ribosomal proteins. *J. Mol. Biol.*, *405*(5), 1215-1232.

- Martinez, A. K., Gordon, E., Sengupta, A., Shirole, N., Klepacki, D., Martinez-Garriga, B., . . . Cruz-Vera, L. R. (2014). Interactions of the TnaC nascent peptide with rRNA in the exit tunnel enable the ribosome to respond to free tryptophan. *Nucleic Acids Res*, *42*(2), 1245-1256. doi: 10.1093/nar/gkt923
- Mason, N., Ciuffo, L. F., & Brown, J. D. (2000). Elongation arrest is a physiologically important function of signal recognition particle. *EMBO J*, *19*(15), 4164-4174. doi: 10.1093/emboj/19.15.4164
- Matte-Tailliez, O., Brochier, C., Forterre, P., & Philippe, H. (2002). Archaeal phylogeny based on ribosomal proteins. *Mol Biol Evol*, *19*(5), 631-639.
- Mayford, M., & Weisblum, B. (1989). ermC leader peptide. Amino acid sequence critical for induction by translational attenuation. *J Mol Biol*, *206*(1), 69-79.
- McNicholas, P., Salavati, R., & Oliver, D. (1997). Dual regulation of *Escherichia coli* secA translation by distinct upstream elements. *J Mol Biol*, *265*(2), 128-141.
- Melnikov, S., Ben-Shem, A., Garreau de Loubresse, N., Jenner, L., Yusupova, G., & Yusupov, M. (2012). One core, two shells: bacterial and eukaryotic ribosomes. *Nat. Struct. Mol. Biol.*, *19*(6), 560-567.
- Mikuni, O., Ito, K., Moffat, J., Matsumura, K., McCaughan, K., Nobukuni, T., . . . Nakamura, Y. (1994). Identification of the *prfC* gene, which encodes peptide-chain-release factor 3 of *Escherichia coli*. *Proc. Natl Acad. Sci. USA*, *91*(13), 5798-5802.
- Milon, P., Carotti, M., Konevega, A. L., Wintermeyer, W., Rodnina, M. V., & Gualerzi, C. O. (2010). The ribosome-bound initiation factor 2 recruits initiator tRNA to the 30S initiation complex. *EMBO Rep*.
- Milon, P., Maracci, C., Filonava, L., Gualerzi, C. O., & Rodnina, M. V. (2012). Real-time assembly landscape of bacterial 30S translation initiation complex. *Nat Struct Mol Biol*, *19*(6), 609-615. doi: 10.1038/nsmb.2285
- Min, Y. H., Kwon, A. R., Yoon, E. J., Shim, M. J., & Choi, E. C. (2008). Translational attenuation and mRNA stabilization as mechanisms of erm(B) induction by erythromycin. *Antimicrob Agents Chemother*, *52*(5), 1782-1789.
- Mize, G. J., Ruan, H., Low, J. J., & Morris, D. R. (1998). The inhibitory upstream open reading frame from mammalian S-adenosylmethionine decarboxylase mRNA has a strict sequence specificity in critical positions. *J Biol Chem*, *273*(49), 32500-32505.
- Moazed, D., & Noller, H. F. (1989). Intermediate states in the movement of transfer RNA in the ribosome. *Nature*, *342*, 142-148.
- Munro, J. B., Altman, R. B., O'Connor, N., & Blanchard, S. C. (2007). Identification of two distinct hybrid state intermediates on the ribosome. *Mol. Cell*, *25*(4), 505-517.
- Murakami, T., Haga, K., Takeuchi, M., & Sato, T. (2002). Analysis of the *Bacillus subtilis* spoIIJ gene and its Parologue gene, yqjG. *J Bacteriol*, *184*(7), 1998-2004.
- Mushegian, A. R., & Koonin, E. V. (1996). A minimal gene set for cellular life derived by comparison of complete bacterial genomes. *Proc Natl Acad Sci U S A*, *93*(19), 10268-10273.
- Muth, G. W., Ortoleva-Donnelly, L., & Strobel, S. A. (2000). A single adenosine with a neutral pKa in the ribosomal peptidyl transferase center. *Science*, *289*(5481), 947-950.
- Nakamori, K., Chiba, S., & Ito, K. (2014). Identification of a SecM segment required for export-coupled release from elongation arrest. *FEBS Lett*, *588*(17), 3098-3103. doi: 10.1016/j.febslet.2014.06.038
- Nakamura, K., Yahagi, S., Yamazaki, T., & Yamane, K. (1999). *Bacillus subtilis* histone-like protein, HBSu, is an integral component of a SRP-like particle that can bind the Alu domain of small cytoplasmic RNA. *J Biol Chem*, *274*(19), 13569-13576.
- Nakatogawa, H., & Ito, K. (2001a). Secretion monitor, SecM, undergoes self-translation arrest in the cytosol. *Mol. Cell*, *7*(1), 185-192.
- Nakatogawa, H., & Ito, K. (2001b). Secretion monitor, SecM, undergoes self-translation arrest in the cytosol. *Mol Cell*, *7*(1), 185-192.

- Nakatogawa, H., & Ito, K. (2002). The ribosomal exit tunnel functions as a discriminating gate. *Cell*, *108*(5), 629-636.
- Narayanan, C. S., & Dubnau, D. (1987). Demonstration of erythromycin-dependent stalling of ribosomes on the ermC leader transcript. *J. Biol. Chem.*, *262*(4), 1766-1771.
- Ninio, J. (1975). Kinetic amplification of enzyme discrimination. *Biochimie*, *57*, 587-595.
- Nissen, P., Hansen, J., Ban, N., Moore, P. B., & Steitz, T. A. (2000). The structural basis of ribosome activity in peptide bond synthesis. *Science*, *289*(5481), 920-930.
- Noriega, T. R., Tsai, A., Elvekrog, M. M., Petrov, A., Neher, S. B., Chen, J., . . . Walter, P. (2014). Signal recognition particle-ribosome binding is sensitive to nascent chain length. *J Biol Chem*, *289*(28), 19294-19305. doi: 10.1074/jbc.M114.563239
- Nouwen, N., & Driessen, A. J. (2002). SecDFyajC forms a heterotetrameric complex with YidC. *Mol Microbiol*, *44*(5), 1397-1405.
- Nyathi, Y., Wilkinson, B. M., & Pool, M. R. (2013). Co-translational targeting and translocation of proteins to the endoplasmic reticulum. *Biochim Biophys Acta*, *1833*(11), 2392-2402. doi: 10.1016/j.bbamcr.2013.02.021
- Ogle, J. M., Murphy, F. V., Tarry, M. J., & Ramakrishnan, V. (2002). Selection of tRNA by the ribosome requires a transition from an open to a closed form. *Cell*, *111*(5), 721-732.
- Olsen, G. J., & Woese, C. R. (1997). Archaeal genomics: An overview. *Cell*, *89*(7), 991-994.
- Onouchi, H., Nagami, Y., Haraguchi, Y., Nakamoto, M., Nishimura, Y., Sakurai, R., . . . Naito, S. (2005). Nascent peptide-mediated translation elongation arrest coupled with mRNA degradation in the CGS1 gene of Arabidopsis. *Genes Dev*, *19*(15), 1799-1810. doi: 10.1101/gad.1317105
- Parmeggiani, A., Krab, I. M., Okamura, S., Nielsen, R. C., Nyborg, J., & Nissen, P. (2006). Structural basis of the action of pulvomycin and GE2270 A on elongation factor Tu. *Biochemistry*, *45*(22), 6846-6857.
- Parmeggiani, A., Krab, I. M., Watanabe, T., Nielsen, R. C., Dahlberg, C., Nyborg, J., & Nissen, P. (2006). Enacyloxin IIa pinpoints a binding pocket of elongation factor Tu for development of novel antibiotics. *J Biol Chem*, *281*(5), 2893-2900.
- Paytubi, S., McMahon, S. A., Graham, S., Liu, H., Botting, C. H., Makarova, K. S., . . . White, M. F. (2012). Displacement of the canonical single-stranded DNA-binding protein in the Thermoproteales. *Proc Natl Acad Sci U S A*, *109*(7), E398-405. doi: 10.1073/pnas.1113277108
- Peske, F., Kuhlkoetter, S., Rodnina, M. V., & Wintermeyer, W. (2014). Timing of GTP binding and hydrolysis by translation termination factor RF3. *Nucleic Acids Res*, *42*(3), 1812-1820. doi: 10.1093/nar/gkt1095
- Peske, F., Rodnina, M., & Wintermeyer, W. (2005). Sequence of steps in ribosome recycling as defined by kinetic analysis. *Mol. Cell*, *18*(4), 403-412.
- Pestryakov, P. E., & Lavrik, O. I. (2008). Mechanisms of single-stranded DNA-binding protein functioning in cellular DNA metabolism. *Biochemistry (Mosc)*, *73*(13), 1388-1404.
- Polikanov, Y. S., Steitz, T. A., & Innis, C. A. (2014). A proton wire to couple aminoacyl-tRNA accommodation and peptide-bond formation on the ribosome. *Nat Struct Mol Biol*, *21*(9), 787-793. doi: 10.1038/nsmb.2871
- Powers, T., & Walter, P. (1997). Co-translational protein targeting catalyzed by the Escherichia coli signal recognition particle and its receptor. *EMBO J*, *16*(16), 4880-4886. doi: 10.1093/emboj/16.16.4880
- Pulk, A., & Cate, J. H. (2013). Control of ribosomal subunit rotation by elongation factor G. *Science*, *340*(6140), 1235970. doi: 10.1126/science.1235970
- Qin, H., Grigoriadou, C., & Cooperman, B. S. (2009). Interaction of IF2 with the ribosomal GTPase-associated center during 70S initiation complex formation. *Biochemistry*, *48*(22), 4699-4706.
- Qin, Y., Polacek, N., Vesper, O., Staub, E., Einfeldt, E., Wilson, D. N., & Nierhaus, K. H. (2006). The highly conserved LepA is a ribosomal elongation factor that back-translocates the ribosome. *Cell*, *127*(4), 721-733.

- Rahmani, F., Hummel, M., Schuurmans, J., Wiese-Klinkenberg, A., Smeekens, S., & Hanson, J. (2009). Sucrose control of translation mediated by an upstream open reading frame-encoded peptide. *Plant Physiol*, *150*(3), 1356-1367. doi: 10.1104/pp.109.136036
- Ramrath, D. J., Lancaster, L., Sprink, T., Mielke, T., Loerke, J., Noller, H. F., & Spahn, C. M. (2013). Visualization of two transfer RNAs trapped in transit during elongation factor G-mediated translocation. *Proc Natl Acad Sci U S A*, *110*(52), 20964-20969. doi: 10.1073/pnas.1320387110
- Ramu, H., Mankin, A., & Vazquez-Laslop, N. (2009). Programmed drug-dependent ribosome stalling. *Mol Microbiol*, *71*(4), 811-824.
- Ratje, A. H., Loerke, J., Mikolajka, A., Brunner, M., Hildebrand, P. W., Starosta, A. L., . . . Spahn, C. M. (2010). Head swivel on the ribosome facilitates translocation by means of intra-subunit tRNA hybrid sites. *Nature*, *468*(7324), 713-716.
- Reilman, E., Mars, R. A., van Dijl, J. M., & Denham, E. L. (2014). The multidrug ABC transporter BmrC/BmrD of *Bacillus subtilis* is regulated via a ribosome-mediated transcriptional attenuation mechanism. *Nucleic Acids Res*, *42*(18), 11393-11407. doi: 10.1093/nar/gku832
- Robinson, P. J., & Woolhead, C. A. (2013). Post-translational membrane insertion of an endogenous YidC substrate. *Biochim Biophys Acta*, *1833*(12), 2781-2788. doi: 10.1016/j.bbamcr.2013.07.003
- Rodnina, M. V., Savelsbergh, A., Katunin, V. I., & Wintermeyer, W. (1997). Hydrolysis of GTP by elongation factor G drives tRNA movement on the ribosome. *Nature*, *385*(6611), 37-41.
- Romisch, K., Webb, J., Lingelbach, K., Gausepohl, H., & Dobberstein, B. (1990). The 54-kD protein of signal recognition particle contains a methionine-rich RNA binding domain. *J Cell Biol*, *111*(5 Pt 1), 1793-1802.
- Rook, F., Weisbeek, P., & Smeekens, S. (1998). The light-regulated *Arabidopsis* bZIP transcription factor gene ATB2 encodes a protein with an unusually long leucine zipper domain. *Plant Mol Biol*, *37*(1), 171-178.
- Rosenblad, M. A., Larsen, N., Samuelsson, T., & Zwieb, C. (2009). Kinship in the SRP RNA family. *RNA Biol*, *6*(5), 508-516.
- Rosendal, K. R., Wild, K., Montoya, G., & Sinning, I. (2003). Crystal structure of the complete core of archaeal signal recognition particle and implications for interdomain communication. *Proc Natl Acad Sci U S A*, *100*(25), 14701-14706. doi: 10.1073/pnas.2436132100
- Ryan, M. D., King, A. M., & Thomas, G. P. (1991). Cleavage of foot-and-mouth disease virus polyprotein is mediated by residues located within a 19 amino acid sequence. *J Gen Virol*, *72* (Pt 11), 2727-2732.
- Saller, M. J., Fusetti, F., & Driessen, A. J. (2009). *Bacillus subtilis* SpoIIIJ and YqjG function in membrane protein biogenesis. *J Bacteriol*, *191*(21), 6749-6757. doi: 10.1128/JB.00853-09
- Saller, M. J., Otto, A., Berrelkamp-Lahpor, G. A., Becher, D., Hecker, M., & Driessen, A. J. (2011). *Bacillus subtilis* YqjG is required for genetic competence development. *Proteomics*, *11*(2), 270-282. doi: 10.1002/pmic.201000435
- Salsi, E., Farah, E., Dann, J., & Ermolenko, D. N. (2014). Following movement of domain IV of elongation factor G during ribosomal translocation. *Proc Natl Acad Sci U S A*, *111*(42), 15060-15065. doi: 10.1073/pnas.1410873111
- Salsi, E., Farah, E., Netter, Z., Dann, J., & Ermolenko, D. N. (2015). Movement of elongation factor G between compact and extended conformations. *J Mol Biol*, *427*(2), 454-467. doi: 10.1016/j.jmb.2014.11.010
- Samuelson, J. C., Chen, M., Jiang, F., Moller, I., Wiedmann, M., Kuhn, A., . . . Dalbey, R. E. (2000). YidC mediates membrane protein insertion in bacteria. *Nature*, *406*(6796), 637-641. doi: 10.1038/35020586



- Santos, N., Zhu, J., Donohue, J. P., Korostelev, A. A., & Noller, H. F. (2013). Crystal structure of the 70S ribosome bound with the Q253P mutant form of release factor RF2. *Structure*, *21*(7), 1258-1263. doi: 10.1016/j.str.2013.04.028
- Sarker, S., Rudd, K. E., & Oliver, D. (2000). Revised translation start site for secM defines an atypical signal peptide that regulates Escherichia coli secA expression. *J Bacteriol*, *182*(19), 5592-5595.
- Schleiss, M. R., Degnin, C. R., & Geballe, A. P. (1991). Translational control of human cytomegalovirus gp48 expression. *J Virol*, *65*(12), 6782-6789.
- Schmeing, T. M., Huang, K. S., Kitchen, D. E., Strobel, S. A., & Steitz, T. A. (2005). Structural insights into the roles of water and the 2' hydroxyl of the P site tRNA in the peptidyl transferase reaction. *Mol. Cell*, *20*(3), 437-448.
- Schmeing, T. M., Huang, K. S., Strobel, S. A., & Steitz, T. A. (2005). An induced-fit mechanism to promote peptide bond formation and exclude hydrolysis of peptidyl-tRNA. *Nature*, *438*(7067), 520-524.
- Schmeing, T. M., Seila, A. C., Hansen, J. L., Freeborn, B., Soukup, J. K., Scaringe, S. A., . . . Steitz, T. A. (2002). A pre-translocational intermediate in protein synthesis observed in crystals of enzymatically active 50S subunits. *Nat. Struct. Biol.*, *9*(3), 225-230.
- Schmeing, T. M., Voorhees, R. M., Kelley, A. C., Gao, Y. G., Murphy, F. V. t., Weir, J. R., & Ramakrishnan, V. (2009). The crystal structure of the ribosome bound to EF-Tu and aminoacyl-tRNA. *Science*, *326*(5953), 688-694.
- Schulze, R. J., Komar, J., Botte, M., Allen, W. J., Whitehouse, S., Gold, V. A., . . . Collinson, I. (2014). Membrane protein insertion and proton-motive-force-dependent secretion through the bacterial holo-translocon SecYEG-SecDF-YajC-YidC. *Proc Natl Acad Sci U S A*, *111*(13), 4844-4849. doi: 10.1073/pnas.1315901111
- Seidelt, B., Innis, C. A., Wilson, D. N., Gartmann, M., Armache, J. P., Villa, E., . . . Beckmann, R. (2009). Structural insight into nascent polypeptide chain-mediated translational stalling. *Science*, *326*(5958), 1412-1415.
- Serek, J., Bauer-Manz, G., Struhalla, G., van den Berg, L., Kiefer, D., Dalbey, R., & Kuhn, A. (2004). Escherichia coli YidC is a membrane insertase for Sec-independent proteins. *EMBO J*, *23*(2), 294-301. doi: 10.1038/sj.emboj.7600063
- Sharma, P., Yan, F., Doronina, V. A., Escuin-Ordinas, H., Ryan, M. D., & Brown, J. D. (2012). 2A peptides provide distinct solutions to driving stop-carry on translational recoding. *Nucleic Acids Res*, *40*(7), 3143-3151. doi: 10.1093/nar/gkr1176
- Shine, J., & Dalgarno, L. (1974). The 3'-terminal sequence of *E. coli* 16S rRNA: Complementarity to nonsense triplets and ribosome binding sites. *Proc. Natl Acad. Sci. USA*, *71*(4), 1342-1346.
- Shoji, S., Walker, S. E., & Fredrick, K. (2006). Reverse translocation of tRNA in the ribosome. *Mol Cell*, *24*(6), 931-942.
- Siegel, V., & Walter, P. (1985). Elongation arrest is not a prerequisite for secretory protein translocation across the microsomal membrane. *J Cell Biol*, *100*(6), 1913-1921.
- Sievers, A., Beringer, M., Rodnina, M. V., & Wolfenden, R. (2004). The ribosome as an entropy trap. *Proc. Natl. Acad. Sci. U S A*, *101*(21), 7897-7901.
- Simonetti, A., Marzi, S., Jenner, L., Myasnikov, A., Romby, P., Yusupova, G., . . . Yusupov, M. (2009). A structural view of translation initiation in bacteria. *Cell Mol Life Sci*, *66*(3), 423-436.
- Singh, N., Das, G., Seshadri, A., Sangeetha, R., & Varshney, U. (2005). Evidence for a role of initiation factor 3 in recycling of ribosomal complexes stalled on mRNAs in Escherichia coli. *Nucleic Acids Res.*, *33*(17), 5591-5601.
- Sohmen, D., Harms, J. M., Schlunzen, F., & Wilson, D. N. (2009). Enhanced Snapshot: Antibiotic inhibition of protein synthesis II. *Cell*, *139*(1), 212-212 e211. doi: 10.1016/j.cell.2009.08.009

- Starosta, A. L., Qin, H., Mikolajka, A., Leung, G. Y., Schwinghammer, K., Nicolaou, K. C., . . . Wilson, D. N. (2009). Identification of distinct thiopeptide-antibiotic precursor lead compounds using translation machinery assays. *Chem Biol*, *16*(10), 1087-1096.
- Struck, J. C., Vogel, D. W., Ulbrich, N., & Erdmann, V. A. (1988). The *Bacillus subtilis* scRNA is related to the 4.5S RNA from *Escherichia coli*. *Nucleic Acids Res*, *16*(6), 2719.
- Tan, Y. T., Tillett, D. J., & McKay, I. A. (2000). Molecular strategies for overcoming antibiotic resistance in bacteria. *Mol Med Today*, *6*(8), 309-314.
- Tenson, T., & Ehrenberg, M. (2002). Regulatory nascent peptides in the ribosomal tunnel. *Cell*, *108*(5), 591-594.
- Tourigny, D. S., Fernandez, I. S., Kelley, A. C., & Ramakrishnan, V. (2013). Elongation factor G bound to the ribosome in an intermediate state of translocation. *Science*, *340*(6140), 1235490. doi: 10.1126/science.1235490
- Trobro, S., & Aqvist, J. (2006). Analysis of predictions for the catalytic mechanism of ribosomal peptidyl transfer. *Biochemistry*, *45*(23), 7049-7056. doi: 10.1021/bi0605383
- Tsai, A., Kornberg, G., Johansson, M., Chen, J., & Puglisi, J. D. (2014). The Dynamics of SecM-Induced Translational Stalling. *Cell Rep*. doi: 10.1016/j.celrep.2014.04.033
- Uechi, T., Tanaka, T., & Kenmochi, N. (2001). A complete map of the human ribosomal protein genes: assignment of 80 genes to the cytogenetic map and implications for human disorders. *Genomics*, *72*(3), 223-230. doi: 10.1006/geno.2000.6470
- Valle, M., Zavialov, A., Sengupta, J., Rawat, U., Ehrenberg, M., & Frank, J. (2003). Locking and unlocking of ribosomal motions. *Cell*, *114*(1), 123-134.
- van der Sluis, E. O., & Driessen, A. J. (2006). Stepwise evolution of the Sec machinery in Proteobacteria. *Trends Microbiol*, *14*(3), 105-108. doi: 10.1016/j.tim.2006.01.009
- Vazquez-Laslop, N., & Mankin, A. S. (2011). Picky nascent peptides do not talk to foreign ribosomes. *Proc Natl Acad Sci U S A*, *108*(15), 5931-5932. doi: 10.1073/pnas.1103011108
- Vazquez-Laslop, N., Thum, C., & Mankin, A. S. (2008). Molecular mechanism of drug-dependent ribosome stalling. *Mol Cell*, *30*(2), 190-202.
- Vestergaard, B., Van, L., Andersen, G., Nyborg, J., Buckingham, R., & Kjeldgaard, M. (2001). Bacterial polypeptide release factor RF2 is structurally distinct from eukaryotic eRF1. *Mol. Cell*, *8*(6), 1375-1382.
- von Loeffelholz, O., Jiang, Q., Ariosa, A., Karuppasamy, M., Huard, K., Berger, I., . . . Schaffitzel, C. (2015). Ribosome-SRP-FtsY cotranslational targeting complex in the closed state. *Proc Natl Acad Sci U S A*, *112*(13), 3943-3948. doi: 10.1073/pnas.1424453112
- Voorhees, R. M., & Ramakrishnan, V. (2013). Structural basis of the translational elongation cycle. *Annu. Rev. Biochem.*, *82*, 203-236.
- Voorhees, R. M., Schmeing, T. M., Kelley, A. C., & Ramakrishnan, V. (2010). The mechanism for activation of GTP hydrolysis on the ribosome. *Science*, *330*(6005), 835-838.
- Voorhees, R. M., Weixlbaumer, A., Loakes, D., Kelley, A. C., & Ramakrishnan, V. (2009). Insights into substrate stabilization from snapshots of the peptidyl transferase center of the intact 70S ribosome. *Nat Struct Mol Biol*, *16*(5), 528-533.
- Wallin, G., & Aqvist, J. (2010). The transition state for peptide bond formation reveals the ribosome as a water trap. *Proc Natl Acad Sci U S A*, *107*(5), 1888-1893. doi: 10.1073/pnas.0914192107
- Walter, P., & Blobel, G. (1981). Translocation of proteins across the endoplasmic reticulum III. Signal recognition protein (SRP) causes signal sequence-dependent and site-specific arrest of chain elongation that is released by microsomal membranes. *J Cell Biol*, *91*(2 Pt 1), 557-561.
- Walter, P., Ibrahimi, I., & Blobel, G. (1981). Translocation of proteins across the endoplasmic reticulum. I. Signal recognition protein (SRP) binds to in-vitro-assembled polysomes synthesizing secretory protein. *J Cell Biol*, *91*(2 Pt 1), 545-550.

- Wang, L., Pulk, A., Wasserman, M. R., Feldman, M. B., Altman, R. B., Cate, J. H., & Blanchard, S. C. (2012). Allosteric control of the ribosome by small-molecule antibiotics. *Nat. Struct. Mol. Biol.*, *19*(9), 957-963.
- Wang, Y., Rotman, E. R., Shoemaker, N. B., & Salyers, A. A. (2005). Translational control of tetracycline resistance and conjugation in the *Bacteroides* conjugative transposon CTnDOT. *J Bacteriol*, *187*(8), 2673-2680. doi: 10.1128/JB.187.8.2673-2680.2005
- Wang, Z., Fang, P., & Sachs, M. S. (1998). The evolutionarily conserved eukaryotic arginine attenuator peptide regulates the movement of ribosomes that have translated it. *Mol Cell Biol*, *18*(12), 7528-7536.
- Wang, Z., & Sachs, M. S. (1997). Arginine-specific regulation mediated by the *Neurospora crassa* arg-2 upstream open reading frame in a homologous, cell-free in vitro translation system. *J Biol Chem*, *272*(1), 255-261.
- Wang, Z., & Sachs, M. S. (1997). Ribosome stalling is responsible for arginine-specific translational attenuation in *Neurospora crassa*. *Mol Cell Biol*, *17*(9), 4904-4913.
- Wei, J., Wu, C., & Sachs, M. S. (2012). The arginine attenuator peptide interferes with the ribosome peptidyl transferase center. *Mol Cell Biol*, *32*(13), 2396-2406. doi: 10.1128/MCB.00136-12
- Weichenrieder, O., Stehlin, C., Kapp, U., Birse, D. E., Timmins, P. A., Strub, K., & Cusack, S. (2001). Hierarchical assembly of the Alu domain of the mammalian signal recognition particle. *RNA*, *7*(5), 731-740.
- Weichenrieder, O., Wild, K., Strub, K., & Cusack, S. (2000). Structure and assembly of the Alu domain of the mammalian signal recognition particle. *Nature*, *408*(6809), 167-173. doi: 10.1038/35041507
- Weixlbaumer, A., Jin, H., Neubauer, C., Voorhees, R., Petry, S., Kelley, A., & Ramakrishnan, V. (2008). Insights into translational termination from the structure of RF2 bound to the ribosome. *Science*, *322*(5903), 953-956.
- Weixlbaumer, A., Petry, S., Dunham, C. M., Selmer, M., Kelley, A. C., & Ramakrishnan, V. (2007). Crystal structure of the ribosome recycling factor bound to the ribosome. *Nat. Struct. Mol. Biol.*, *14*(8), 733-737.
- Wiese, A., Elzinga, N., Wobbes, B., & Smeekens, S. (2004). A conserved upstream open reading frame mediates sucrose-induced repression of translation. *Plant Cell*, *16*(7), 1717-1729. doi: 10.1105/tpc.019349
- Wiese, A., Elzinga, N., Wobbes, B., & Smeekens, S. (2005). Sucrose-induced translational repression of plant bZIP-type transcription factors. *Biochem Soc Trans*, *33*(Pt 1), 272-275. doi: 10.1042/BST0330272
- Wild, K., Halic, M., Sinning, I., & Beckmann, R. (2004). SRP meets the ribosome. *Nat Struct Mol Biol*, *11*(11), 1049-1053. doi: 10.1038/nsmb853
- Wilson, D. N. (2014). Ribosome-targeting antibiotics and bacterial resistance mechanisms. *Nat. Rev. Microbiol.*, *12*(1), 35-48.
- Wilson, D. N., & Beckmann, R. (2011). The ribosomal tunnel as a functional environment for nascent polypeptide folding and translational stalling. *Curr. Opin. Struct. Biol.*, *21*(2), 1-10.
- Wilson, D. N., & Cate, J. H. D. (2012). The structure and function of the eukaryotic ribosome. *Cold. Spring. Harb. Perspect. Biol.*, *4*(5), 1-17.
- Wilson, D. N., Schluenzen, F., Harms, J. M., Yoshida, T., Ohkubo, T., Albrecht, R., . . . Fucini, P. (2005). X-ray crystallography study on ribosome recycling: the mechanism of binding and action of RRF on the 50S ribosomal subunit. *EMBO J.*, *24*(2), 251-260.
- Wimberly, B. T., Brodersen, D. E., Clemons, W. M., Morgan-Warren, R. J., Carter, A. P., Vonrhein, C., . . . Ramakrishnan, V. (2000). Structure of the 30S ribosomal subunit. *Nature*, *407*, 327-339.
- Woese, C. R., & Fox, G. E. (1977). Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc. Natl Acad. Sci. U S A*, *74*(11), 5088-5090.

- Wong, W., Bai, X. C., Brown, A., Fernandez, I. S., Hanssen, E., Condrón, M., . . . Scheres, S. H. (2014). Cryo-EM structure of the Plasmodium falciparum 80S ribosome bound to the anti-protozoan drug emetine. *Elife*, *3*. doi: 10.7554/eLife.03080
- Yamamoto, H., Qin, Y., Achenbach, J., Li, C., Kijek, J., Spahn, C. M., & Nierhaus, K. H. (2014). EF-G and EF4: translocation and back-translocation on the bacterial ribosome. *Nat Rev Microbiol*, *12*(2), 89-100. doi: 10.1038/nrmicro3176
- Yamazaki, T., Yahagi, S., Nakamura, K., & Yamane, K. (1999). Depletion of Bacillus subtilis histone-like protein, HBSu, causes defective protein translocation and induces upregulation of small cytoplasmic RNA. *Biochem Biophys Res Commun*, *258*(1), 211-214.
- Yokoyama, T., Shaikh, T. R., Iwakura, N., Kaji, H., Kaji, A., & Agrawal, R. K. (2012). Structural insights into initial and intermediate steps of the ribosome-recycling process. *EMBO J*, *31*(7), 1836-1846. doi: 10.1038/emboj.2012.22
- Young, D. J., Edgar, C. D., Poole, E. S., & Tate, W. P. (2010). The codon specificity of eubacterial release factors is determined by the sequence and size of the recognition loop. *RNA*, *16*(8), 1623-1633. doi: 10.1261/rna.2117010
- Yusupova, G., Jenner, L., Rees, B., Moras, D., & Yusupov, M. (2006). Structural basis for messenger RNA movement on the ribosome. *Nature*, *444*(7117), 391-394.
- Yutin, N., Puigbo, P., Koonin, E. V., & Wolf, Y. I. (2012). Phylogenomics of prokaryotic ribosomal proteins. *PLoS ONE*, *7*(5), e36972. doi: 10.1371/journal.pone.0036972
- Zaher, H. S., & Green, R. (2009a). Fidelity at the molecular level: lessons from protein synthesis. *Cell*, *136*(4), 746-762.
- Zaher, H. S., & Green, R. (2009b). Quality control by the ribosome following peptide bond formation. *Nature*, *457*(7226), 161-166.
- Zaher, H. S., Shaw, J. J., Strobel, S. A., & Green, R. (2011). The 2'-OH group of the peptidyl-tRNA stabilizes an active conformation of the ribosomal PTC. *EMBO J*, *30*(12), 2445-2453. doi: 10.1038/emboj.2011.142
- Zavialov, A. V., Hauryliuk, V. V., & Ehrenberg, M. (2005). Splitting of the posttermination ribosome into subunits by the concerted action of RRF and EF-G. *Mol. Cell*, *18*(6), 675-686.
- Zhou, J., Lancaster, L., Trakhanov, S., & Noller, H. F. (2012). Crystal structure of release factor RF3 trapped in the GTP state on a rotated conformation of the ribosome. *RNA*, *18*(2), 230-240. doi: 10.1261/rna.031187.111
- Zhou, Y., Asahara, H., Gaucher, E. A., & Chong, S. (2012). Reconstitution of translation from Thermus thermophilus reveals a minimal set of components sufficient for protein synthesis at high temperatures and functional conservation of modern and ancient translation components. *Nucleic Acids Res*, *40*(16), 7932-7945. doi: 10.1093/nar/gks568
- Zillig, W. (1991). Comparative biochemistry of Archaea and Bacteria. *Curr Opin Genet Dev*, *1*(4), 544-551.
- Zwieb, C., & Bhuiyan, S. (2010). Archaea signal recognition particle shows the way. *Archaea*, *2010*, 485051. doi: 10.1155/2010/485051

## 12 Publications

### Publication 1

Viter Márquez, Thomas Fröhlich, Jean-Paul Armache, **Daniel Sohmen**, Alexandra Dönhöfer, Aleksandra Mikolajka, Otto Berninghausen, Michael Thomm, Roland Beckmann, Georg J. Arnold and Daniel N. Wilson (2011), Proteomic Characterization of Archaeal Ribosomes Reveals the Presence of Novel Archaeal-Specific Ribosomal Proteins, *Journal of Molecular Biology*. 405, 1215–1232

### Publication 2

Stefan Gross, Fabian Nguyen, Matthias Bierschenk, **Daniel Sohmen**, Thomas Menzel, Iris Antes, Daniel N. Wilson, and Thorsten Bach (2013), Amythiamicin D and Related Thiopeptides as Inhibitors of the Bacterial Elongation Factor EF-Tu: Modification of the Amino Acid at Carbon Atom C2 of Ring C Dramatically Influences Activity, *ChemMedChem*. 8, 1954–1962

### Publication 3

Homa Ghalei, Holger von Moeller, Detlef Eppers, **Daniel Sohmen**, Daniel N. Wilson, Bernhard Loll and Markus C. Wahl (2014), Entrapment of DNA in an intersubunit tunnel system of a single-stranded DNA-binding protein, *Nucleic Acids Research*. 42, 6698–6708

### Publication 4

**Daniel Sohmen**, Shinobu Chiba, Naomi Shimokawa-Chiba, C. Axel Innis, Otto Berninghausen, Roland Beckmann, Koreaki Ito & Daniel N. Wilson (2015), Structure of the Bacillus subtilis 70S ribosome reveals the basis for species-specific stalling, *Nature Communications*. 6, 6941

### Publication 5

Bertrand Beckert, Alexej Kedrov, **Daniel Sohmen**, Georg Kempf, Klemens Wild, Irmgard Sinning, Henning Stahlberg, Daniel N. Wilson and Roland Beckmann (2015\*), Structural basis for targeting and elongation arrest of Bacillus signal recognition particle, \*Unpublished manuscript



# Proteomic Characterization of Archaeal Ribosomes Reveals the Presence of Novel Archaeal-Specific Ribosomal Proteins

Viter Márquez<sup>1,2,†</sup>, Thomas Fröhlich<sup>1,3,†</sup>, Jean-Paul Armache<sup>1,2</sup>, Daniel Sohmen<sup>1,2</sup>, Alexandra Dönhöfer<sup>1,2</sup>, Aleksandra Mikolajka<sup>1,2</sup>, Otto Berninghausen<sup>1,2</sup>, Michael Thomm<sup>4</sup>, Roland Beckmann<sup>1,2</sup>, Georg J. Arnold<sup>1,3\*</sup> and Daniel N. Wilson<sup>1,2\*</sup>

<sup>1</sup>Gene Center and Department of Biochemistry, Ludwig-Maximilians-Universität, Feodor Lynen Str. 25, 81377 Munich, Germany

<sup>2</sup>Center for Integrated Protein Science Munich, Ludwig-Maximilians-Universität, Munich, Germany

<sup>3</sup>Laboratory for Functional Genome Analysis, Gene Center and Department of Biochemistry, Ludwig-Maximilians-Universität, Feodor Lynen Str. 25, 81377 Munich, Germany

<sup>4</sup>Lehrstuhl für Mikrobiologie, Regensburg University, Universitätsstraße 32, 93053 Regensburg, Germany

Received 29 September 2010;  
received in revised form  
19 November 2010;  
accepted 27 November 2010  
Available online  
4 December 2010

Edited by J. Doudna

## Keywords:

archaea;  
cryo electron microscopy;  
*Pyrobaculum aerophilum*;  
ribosomal protein;  
*Sulfolobus acidocaldarius*;  
mass spectrometry

Protein synthesis occurs in macromolecular particles called ribosomes. All ribosomes are composed of RNA and proteins. While the protein composition of bacterial and eukaryotic ribosomes has been well-characterized, a systematic analysis of archaeal ribosomes has been lacking. Here we report the first comprehensive two-dimensional PAGE and mass spectrometry analysis of archaeal ribosomes isolated from the thermophilic *Pyrobaculum aerophilum* and the thermoacidophilic *Sulfolobus acidocaldarius* Crenarchaeota. Our analysis identified all 66 ribosomal proteins (r-proteins) of the *P. aerophilum* small and large subunits, as well as all but two (62 of 64; 97%) r-proteins of the *S. acidocaldarius* small and large subunits that are predicted genomically. Some r-proteins were identified with one or two lysine methylations and N-terminal acetylations. In addition, we identify three hypothetical proteins that appear to be *bona fide* r-proteins of the *S. acidocaldarius* large subunit. Dissociation of r-proteins from the *S. acidocaldarius* large subunit indicates that the novel r-proteins establish tighter interactions with the large subunit than some integral r-proteins. Furthermore, cryo electron microscopy reconstructions of the *S. acidocaldarius* and *P. aerophilum* 50S subunits allow for a tentative localization of the binding site of the novel r-proteins. This study illustrates not only the potential diversity of the archaeal ribosomes but also the necessity to experimentally analyze the archaeal ribosomes to ascertain their protein

\*Corresponding authors. D. N. Wilson is to be contacted at Gene Center, Feodor Lynen Str. 25, 81377 Munich, Germany. G. J. Arnold, Laboratory for Functional Genome Analysis, Gene Center and Department of Biochemistry, Ludwig-Maximilians-Universität, Feodor Lynen Str. 25, 81377 Munich, Germany. E-mail addresses: [wilson@lmb.uni-muenchen.de](mailto:wilson@lmb.uni-muenchen.de); [arnold@lmb.uni-muenchen.de](mailto:arnold@lmb.uni-muenchen.de).

† V.M. and T.F. contributed equally to this work.

Abbreviations used: r-protein, ribosomal protein; 2D, two-dimensional; MS/MS, tandem mass spectrometry; EM, electron microscopy; HSW, high-salt-washed; MS, mass spectrometry; EF-1 $\alpha$ , elongation factor 1 $\alpha$ ; CLP, chromatin-like protein; USP, universal stress protein; EF-G, elongation factor G; FA, formic acid; MALDI, matrix-assisted laser desorption/ionization; TFA, trifluoroacetic acid; PDB, Protein Data Bank.

composition. The discovery of novel archaeal r-proteins and factors may be the first step to understanding how archaeal ribosomes cope with extreme environmental conditions.

© 2010 Elsevier Ltd. All rights reserved.

## Introduction

Protein synthesis occurs in large macromolecular particles called ribosomes (reviewed by Schmeing and Ramakrishnan<sup>1</sup>). Ribosomes are composed of RNA and proteins. In bacteria, the 70S ribosome can be split into small (30S) and large (50S) ribosomal subunits. Although the active sites of small and large subunits responsible for decoding and peptide bond formation, respectively, are composed predominantly of RNA, the contribution of ribosomal proteins (r-proteins) is not to be overlooked.<sup>2,3</sup> At the decoding site, r-protein S12 contributes to the fidelity of the codon-anticodon interaction of the mRNA-tRNA duplex,<sup>4</sup> whereas at the peptidyltransferase center, the N-terminal extension of the r-protein L27 of bacterial ribosomes and the loop of L10e in eukaryotic ribosomes contacts the terminal CCA end of the P-site tRNA.<sup>5-8</sup> R-proteins comprise the tRNA binding sites, stabilizing their interaction with and passage through the ribosome during translation.<sup>8-13</sup> Moreover, many r-proteins play critical roles in facilitating various aspects of translation ranging from biogenesis to recruitment of translation factors and chaperones (reviewed by Wilson and Nierhaus<sup>2</sup> and Brodersen and Nissen<sup>3</sup>).

The *Escherichia coli* 70S ribosome contains 54 r-proteins: 21 in the small subunit (S1-S21) and 33 in the large subunit (L1-L36). R-proteins were originally numbered according to their position on two-dimensional (2D) PAGE; as a consequence, large acidic proteins have small numbers, and small basic proteins have large numbers.<sup>14</sup> L7 is the N-acetylated version of L12; L8 was later found to be the pentameric complex of L10·(L7/L12)<sub>4</sub>; and L26 was later re-assigned as S20. The genes for 48 of the 54 *E. coli* r-proteins are present in the genomes of all bacteria, whereas S1, S21, and S22 of the small subunit, as well as L25 and L30 of the large subunit, are missing in some bacteria. In fact, S22 associates with the ribosome during stationary phase and is therefore considered a ribosomal factor rather than a *bona fide* r-protein. Chloroplast ribosomes have homologues to all the *E. coli* r-proteins, except for L25 and L30, but in addition have six plastid-specific r-proteins,<sup>15,16</sup> one of which (PSRP1) was subsequently shown to be a ribosomal factor rather than an r-protein.<sup>17,18</sup> Mitochondrial ribosomes are even more diverse, with 81 r-proteins identified in the human mitochondrial 55S ribosome, 33 r-proteins identified in the small subunit, and 48 r-proteins identified in the large subunit.<sup>19-21</sup>

Proteomic analysis of eukaryotic ribosomes ranging from yeast,<sup>22,23</sup> fruit flies,<sup>24</sup> and plants<sup>25</sup> to

rodents<sup>26-28</sup> and humans,<sup>29,30</sup> together with genomic analysis,<sup>31</sup> has identified ~78 r-proteins, 34 of which have homologues in bacteria. In contrast, no systematic study cataloguing the protein composition of archaeal ribosomes has been performed. Archaeal r-proteins have been characterized individually (e.g., see Kimura *et al.*<sup>32</sup> and Auer *et al.*<sup>33,34</sup>), leading to the identification of archaeal-specific r-protein LX,<sup>35</sup> or collectively, using 2D-PAGE.<sup>36-39</sup> However, assignment of individual r-proteins by comparing 2D-PAGE profiles with data obtained from *E. coli* ribosomes is often ambiguous due to different buffer/running conditions, as well as the increased complexity of archaeal ribosomes. Genomic analysis indicates that archaeal ribosomes are intermediate in terms of composition between bacterial ribosomes and eukaryotic ribosomes, containing up to 68 r-protein families, of which 34 are common to bacteria and eukaryotes, 33 are present in eukaryotes only, and 1 (LX) is archaeal-specific.<sup>31</sup> Ten of the r-protein families exhibit a heterogeneous distribution within archaea, with the ribosomes from Euryarchaeota predicted to have fewer r-proteins than their counterparts in Crenarchaeota (Fig. 1).<sup>31</sup>

Using 2D-PAGE and LC tandem mass spectrometry (MS/MS), we identified all 66 r-proteins of the small and large subunits of the thermophilic *Pyrobaculum aerophilum*. In addition to all but two (62 of 64; 97%) r-proteins of the *Sulfolobus acidocaldarius* small and large subunits, we identified three hypothetical proteins that are present in the thermoacidophilic *S. acidocaldarius* large subunit. These proteins interact more tightly with the large subunit than some integral r-proteins, suggesting that they are *bona fide* r-proteins. Cryo electron microscopy (EM) reconstructions of the *S. acidocaldarius* and *P. aerophilum* 50S subunits identified multiple potential binding sites for the novel r-proteins. In fact, in total, nine hypothetical proteins with pI values of >9 were identified within the small and large ribosomal subunits of these two Crenarchaeota, suggesting that the number of novel r-proteins in archaeal species may far exceed the predictions based on genomic analyses.

## Results and Discussion

### Near-complete proteomic characterization of Crenarchaeota r-proteins

Since genomic analyses indicate that, compared to Crenarchaeota, Euryarchaeota lineages appear to

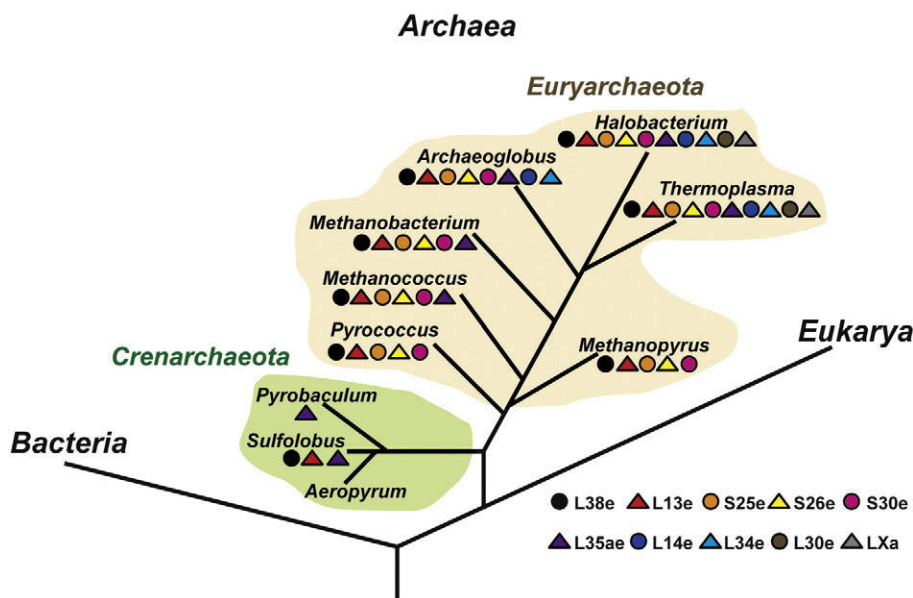


Fig. 1. Heterogeneous distribution of archaeal r-proteins (modified from Lecompte *et al.*<sup>31</sup>).

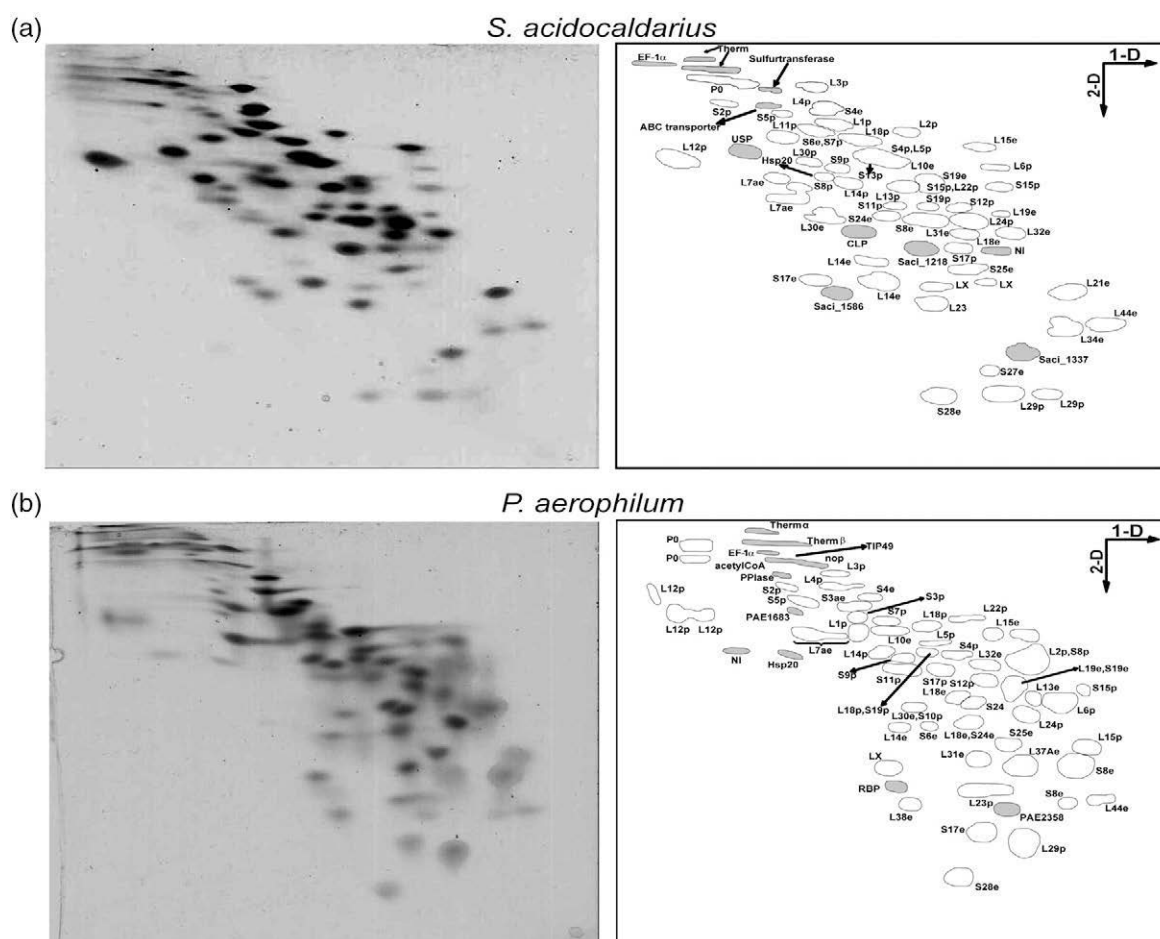
lack many r-protein families (Fig. 1),<sup>31</sup> we selected two Crenarchaeota with completely sequenced genomes—*S. acidocaldarius* and *P. aerophilum*—from which ribosomes were isolated and their r-protein content was characterized. For example, the genome of the euryarchaeon *Haloarcula marismortui* predicts a total of 56 (25 small subunits/31 large subunits) r-proteins in the *H. marismortui* 70S ribosome, whereas the 70S ribosomes of the Crenarchaeota *S. acidocaldarius* and *P. aerophilum* are predicted to comprise 64 (28 and 36) and 66 (28 and 38) r-proteins, respectively. *S. acidocaldarius* is a thermoacidophilic crenarchaeon that grows optimally at 75–80 °C and pH 2–3, whereas *P. aerophilum* grows optimally at 100 °C and pH 7.0 (see [Materials and Methods](#)). Both Crenarchaeota can grow aerobically and, in fact, the species name *aerophilum* translates as “air-loving,” reflecting its ability to respire aerobically. Nevertheless, both archaea are slow-growing, and large-scale fermentation was required to obtain sufficient yields for the isolation and characterization of the ribosomes. Furthermore, ribosome yields were reduced by the necessity for repeated high-salt washing and centrifugation steps through sucrose cushions to obtain clarified ribosomal pellets.

The high-salt-washed (HSW) ribosomal fractions from *S. acidocaldarius* and *P. aerophilum* were analyzed by 2D-PAGE using the method of Kaltschmidt and Wittmann, and protein spots were identified by mass spectrometry (MS) (Fig. 2, [Tables 12](#)).<sup>14</sup> From 2D-PAGE, it was possible to clearly identify spots corresponding to 51 of 64 (80%) r-proteins of the *S. acidocaldarius* 70S ribosome and 51 of 66 (78%) r-proteins of the *P. aerophilum* 70S ribosome. Missing r-proteins, such as S27ae, S27e, and S30e in the small subunit, as well as L24e, L39e, and L40e in the large

subunit, are between 6 kDa and 8 kDa and therefore are not resolved under these running conditions, which were optimized for the separation of higher-molecular-weight r-proteins. However, *S. acidocaldarius* and *P. aerophilum* ribosomal subunits were subsequently purified using sucrose gradients, and LC-MS/MS analysis of the purified fractions led to further identification of 26 of 28 (93%) and 36 of 36 (100%) r-proteins of the *S. acidocaldarius* 30S and 50S subunits, respectively, and 28 of 28 (100%) and 37 of 38 (97%) r-proteins of the *P. aerophilum* small and large subunits ([Tables 1 and 2](#)). Thus, in summary, only two r-proteins (S26e and S30e) from the *S. acidocaldarius* 30S subunit were not detected in any of the analyses.

Curiously, some r-proteins were assigned to multiple distinct spots on 2D-PAGE, such as L7ae, L29p, P0, and LX in *S. acidocaldarius* (Fig. 2a), and L7ae, L12p, and P0 in *P. aerophilum* (Fig. 2b). Consistently, we find that both LX and L29p are monomethylated on Lys residues 66 and 29, respectively (Fig. 3a and b). The stalk proteins P0 (L10 in bacteria) and L12p (P1–P2 in eukaryotes) are known to be modified in other organisms. In fact, L7 was originally mistakenly identified as a unique r-protein and was subsequently shown to be the N-terminally acetylated form of L12.<sup>40</sup> *E. coli* L12 is also methylated at K81 position.<sup>41</sup> Likewise, in eukaryotes, P1 is acetylated<sup>42</sup> and phosphorylated,<sup>42,43</sup> and P2 is phosphorylated.<sup>44</sup> P0 is phosphorylated at the C-terminus in yeast;<sup>45</sup> in *E. coli*, the N-terminal Met of L10 is removed, and Lys residues 37 and 105 are acetylated.<sup>40,46</sup> We also find that L12p is N-terminally acetylated; however, unlike in *E. coli*, the N-terminal methionine is not cleaved (Fig. S1). In addition, we could identify two distinct methylation





**Fig. 2.** Two-dimensional PAGE of crude ribosomes from *S. acidocaldarius* and *P. aerophilum*. Coomassie-blue-stained 2D gel of HSW ribosomes from (a) *S. acidocaldarius* and (b) *P. aerophilum* (left) showing schematic spot assignments based on LC-MS/MS analysis (right). The directions of the first dimension (1D; based on charge) and of the second dimension (2D; based on mass) are indicated with arrows. Spots identified as proteins not previously assigned as r-proteins are shaded gray. NI indicates that the protein was not identified within the spot.

sites in L12p, namely K39 and K89 (Fig. S1). To our knowledge, distinct forms of L7ae have not been previously reported. We can show that the N-terminal methionine of L7ae is cleaved and that Ser2 is N-terminally acetylated (Fig. S2). In addition, Lys residues 74, 108, and 116 are monomethylated (Fig. 3c; Fig. S2). In archaea, L7ae is a subunit of RNase P, H/ACA, and C/D snoRNPs, as well as a component of the ribosome;<sup>47</sup> thus, it will be interesting to investigate whether posttranslational modifications play a role in regulating the involvement of L7ae in these diverse complexes.

### Identification of potential ribosome-associated proteins

In addition to the expected r-proteins, a number of additional proteins were identified in the 2D gels of the HSW ribosomal fractions, as well as in the LC-

MS/MS of the purified 30S and 50S fractions from *S. acidocaldarius* and *P. aerophilum* (Fig. 2, Table 3). The presence of these proteins can reflect a stable interaction with the pelleted ribosomal particles or can reflect that the proteins pellet or migrate similarly to ribosomal particles, due to their being associated with membrane fragments [e.g., ABC transporter and nitrate reductase (NarGHI)] or their being part of a large macromolecular complex [e.g., the 19S proteasome or the  $\alpha$ -subunit and the  $\beta$ -subunit of the thermosome (eukaryotic chaperonin CCT/TRiC), which forms a hexadecameric  $(\alpha\beta)_4$   $(\alpha\beta)_4$  complex of  $\sim 1$  MDa].<sup>48</sup> Consistent with their size, the thermosome subunits were detected in the HSW ribosomal fraction, as well as the purified 30S subunit but not in the purified 50S subunit fraction (Table 3). Moreover, the archaeal thermosome has been reported to be an RNA binding protein involved in rRNA processing.<sup>49</sup> Other proteins

**Table 1.** R-proteins of *S. acidocaldarius* identified by LC-MS and 2D-PAGE

Protein family	LC-MS	2D	Amino acids	gi <sup>a</sup>	Molecular mass (kDa)	pI
<i>Small subunit r-protein</i>						
S2p	✓	✓	225	70605938	25.7	5.8
S3ae	✓	—	197	70606435	22.6	10.2
S3p	✓	—	231	70606406	25.4	10.1
S4e	✓	✓	244	70606400	27.8	10.1
S4p	✓	✓	176	70605931	20.4	10.3
S5p	✓	✓	214	70606392	23.9	9.8
S6e	✓	✓	213	70606621	23.7	10.0
S7p	✓	✓	195	70606489	22.1	10.2
S8e	✓	✓	128	70606557	14.4	10.4
S8p	✓	✓	133	70606397	14.9	9.9
S9p	✓	✓	138	70605936	15.8	10.2
S10p	✓	—	102	70606487	11.9	10.5
S11p	✓	✓	132	70605932	14.1	11.5
S12p	✓	✓	147	70606491	16.1	11.3
S13p	✓	✓	170	70605930	19.4	10.5
S14p <sup>b</sup>	✓	—	63	68567046	7.6	10.2
S15p	✓	✓	153	70606619	17.8	10.5
S17e	✓	✓	82	70606477	9.6	5.3
S17p	✓	✓	108	162139951	12.3	9.9
S19e	✓	✓	154	70607213	18.0	10.6
S19p	✓	✓	140	70606408	16.4	10.6
S24e	✓	✓	118	70606643	13.3	10.3
S25e	✓	✓	109	70606631	12.3	10.5
S26e	—	—	95	68567938	10.8	9.9
S27ae	✓	—	67	70606642	7.8	9.9
S27e	✓	✓	66	70607041	7.3	10.4
S28e	✓	✓	84	70606501	9.5	10.5
S30e	—	—	54	68567754	6.2	11.8
<i>Large subunit r-protein</i>						
L1p	✓	✓	221	162139946	24.9	10.4
L2p	✓	✓	238	70606409	25.2	11.4
L3p	✓	✓	342	70606412	38.5	10.5
L4p	✓	✓	266	70606411	29.2	10.7
L5p	✓	✓	178	70606399	20.2	10.8
L6p	✓	✓	189	3914762	21.2	9.8
L7ae	✓	✓	126	70607262	13.7	8.2
L10e	✓	✓	176	70606018	19.9	10.4
L11p	✓	✓	170	70607203	18.2	9.5
L12p	✓	✓	105	70607200	11.1	4.8
L13p	✓	✓	148	70605935	16.7	10.9
L14e	✓	✓	96	70606603	10.7	9.9
L14p	✓	✓	138	70606402	15.2	10.9
L15e	✓	✓	217	70606432	26.0	11.7
L15p	✓	—	144	70606390	16.4	10.5
L18e	✓	✓	110	70605934	12.5	10.8
L18p	✓	✓	197	70606393	22.1	10.3
L19e	✓	✓	150	70606394	17.8	11.3
L21e	✓	✓	103	70606441	11.8	10.9
L22p	✓	✓	156	70606407	18.1	10.4
L23p	✓	✓	82	70606410	9.4	10.2
L24e	✓	—	62	70606500	7.1	10.4
L24p	✓	✓	134	70606401	15.5	10.6
L29p	✓	✓	69	70606405	8.1	10.5
L30e	✓	✓	104	70606493	11.5	9.8
L30p	✓	✓	156	70606391	17.9	10.2
L31e	✓	✓	129	70607210	15.2	10.8
L32e	✓	✓	131	70606395	15.2	11.1
L34e	✓	✓	87	70606386	10.1	10.5
L37e <sup>b</sup>	✓	✓	61	70606453	6.9	11.9
L37ae	✓	—	70	70606423	8.1	10.4
L39e <sup>b</sup>	✓	—	53	70607211	6.3	12.6
L40e	✓	✓	56	70606562	6.4	11.0
L44e	✓	✓	95	15921177	11.1	10.7

**Table 1 (continued)**

Protein family	LC-MS	2D	Amino acids	gi <sup>a</sup>	Molecular mass (kDa)	pI
LX	✓	✓	86	70607208	10.1	10.1
P0	✓	✓	335	70607201	36.5	9.1

<sup>a</sup> gi refers to the GenInfo identifier for retrieval from NCBI.  
<sup>b</sup> Identified with only a single peptide.

that are better known to be associated with the translational machinery, such as elongation factor 1 $\alpha$  (EF-1 $\alpha$ ) and the translation initiation factors IF-6 and Sui1 (eIF1), were also identified (Table 3). In addition, the ribosome biogenesis factors CBF5 (pseudo-uridine synthetase) and fibrillar (rRNA 2'-O-methyltransferase), and rRNA processing proteins such as NOP56/58 were identified (Table 3). Fibrillar and NOP56/58, together with the archaeal r-protein L7ae, interact with C/D box sRNAs to form an RNP complex involved in 2'-OH ribose methylation, whereas CBF5, L7ae, NOP10, and Gar1 interact with H/ACA sRNAs to direct rRNA pseudo-uridylation.<sup>47</sup>

Four proteins containing putative DNA/RNA-binding domains, which may be present due to interaction with nucleic acids [i.e., chromatin-like protein (CLP), TBP-interacting protein TIP49, transcription factor NusA-like protein, and zinc-ribbon RNA-binding domain protein], were identified (Table 3). Interestingly, we identified six and eight hypothetical proteins in the *S. acidocaldarius* and *P. aerophilum* ribosomal fractions, respectively, that have no known motifs and no sequence homology to any protein with a known function, yet many had pI values of >9.0, characteristic of r-proteins (Tables 1 and 2). While most of the spots for the additional proteins were significantly less intense than the r-protein spots (Fig. 2), there were a number of exceptions—particularly evident in the *S. acidocaldarius* 2D gel were the spots for hypothetical proteins Saci\_1218, Saci\_1337, and Saci\_1586; CLP; and the universal stress protein (USP), all of which appeared to be stoichiometric with, or even more intense, than the majority of r-proteins (Fig. 2a). We thus considered the possibility that these proteins may be novel integral r-proteins of the *S. acidocaldarius* ribosome.

### Identification and distribution of novel *S. acidocaldarius* large subunit r-proteins

In order to determine if the potential ribosome-associated proteins were *bona fide* r-proteins and whether they are integral to the small subunit or the large subunit, we applied the HSW ribosomal fractions onto 10–40% sucrose gradients. In contrast

**Table 2.** R-proteins of *P. aerophilum* identified by LC-MS and 2D-PAGE

Protein family	LC-MS	2D	Amino acids	gi <sup>a</sup>	Molecular mass (kDa)	pI
<i>Small subunit r-protein</i>						
S2p	✓	✓	208	18312202	23.9	9.0
S3ae	✓	✓	221	18314093	24.7	10.3
S3p	✓	✓	218	27734520	25.0	10.4
S4e	✓	✓	238	18313980	26.7	10.4
S4p	✓	✓	159	26399437	18.5	10.9
S5p	✓	✓	218	18312458	24.2	10.1
S6e	✓	✓	148	18312675	15.9	9.9
S7p	✓	✓	223	18312139	25.3	9.6
S8e	✓	✓	131	18313995	14.3	11.2
S8p	✓	✓	130	18313095	14.9	9.9
S9p	✓	✓	146	18312094	16.4	10.6
S10p	✓	✓	106	18313679	12.3	11.7
S11p	✓	✓	133	18313881	14.1	11.4
S12p	✓	✓	147	20140085	16.5	10.9
S13p	✓	—	152	18313810	16.7	10.6
S14p	✓	—	54	18313094	6.4	10.5
S15p	✓	✓	151	18314103	17.7	10.6
S17e	✓	✓	71	18312190	8.2	10.3
S17p	✓	✓	146	18160326	17.0	10.5
S19e	✓	✓	158	18313782	17.7	10.7
S19p	✓	✓	158	18312838	18.2	10.6
S24e	✓	✓	121	18313342	14.2	11.0
S25e	✓	✓	110	18313166	12.5	10.6
S26e	✓	—	98	18313151	11.3	10.7
S27ae	✓	—	65	18313341	7.6	10.3
S27e	✓	—	67	18313776	7.3	10.4
S28e	✓	✓	77	18314007	8.6	10.4
S30e	✓	—	55	18313356	6.2	12.4
<i>Large subunit r-protein</i>						
L1p	✓	✓	222	18313826	24.7	10.7
L2p	✓	✓	246	18312189	26.2	10.9
L3p	✓	✓	338	18313001	37.4	10.8
L4p	✓	✓	283	18313002	31.1	10.4
L5p	✓	✓	179	18314167	20.1	10.9
L6p	✓	✓	196	18313300	21.9	9.8
L7ae	✓	✓	151	18314009	16.1	9.4
L10e	✓	✓	180	18314160	20.2	10.5
L11p <sup>b</sup>	✓	—	167	18313825	18.6	9.6
L12p	✓	✓	110	18312824	11.5	4.5
L13e	✓	✓	159	18312773	18.1	11.6
L13p	✓	—	187	18312093	21.8	10.7
L14e	✓	✓	103	18312232	11.4	10.2
L14p	✓	✓	144	18313850	15.8	11.3
L15e	✓	✓	191	18312915	22.9	11.5
L15p	✓	✓	156	18314064	17.4	10.9
L18e	✓	✓	122	18312092	13.3	12.1
L18p	✓	✓	205	18313098	23.4	10.5
L19e	✓	✓	147	18313097	17.4	11.2
L21e	✓	—	100	18313998	11.5	10.7
L22p	✓	✓	168	18312876	19.6	11.1
L23p	—	✓	81	18313003	9.3	10.3
L24e	✓	—	58	18314008	6.7	10.6
L24p	✓	✓	123	18313979	14.2	11.8
L29p	✓	✓	75	18312872	9.0	11.6
L30e <sup>b</sup>	✓	✓	102	18312088	10.9	9.2
L30p	✓	—	178	18312459	20.3	10.3
L31e	✓	✓	91	18313818	10.6	10.9
L32e	✓	✓	152	18313096	18.0	11.7
L34e	✓	—	84	18313920	9.5	12.2
L37e	—	—	52	NA <sup>c</sup>	12.3	5.4
L37ae	✓	✓	105	18313176	11.8	10.8
L38e	✓	✓	67	18312825	7.9	9.8
L39e <sup>d</sup>	✓	—	51	29427908	6.0	13.4
L40e	✓	—	53	18313942	6.4	11.2

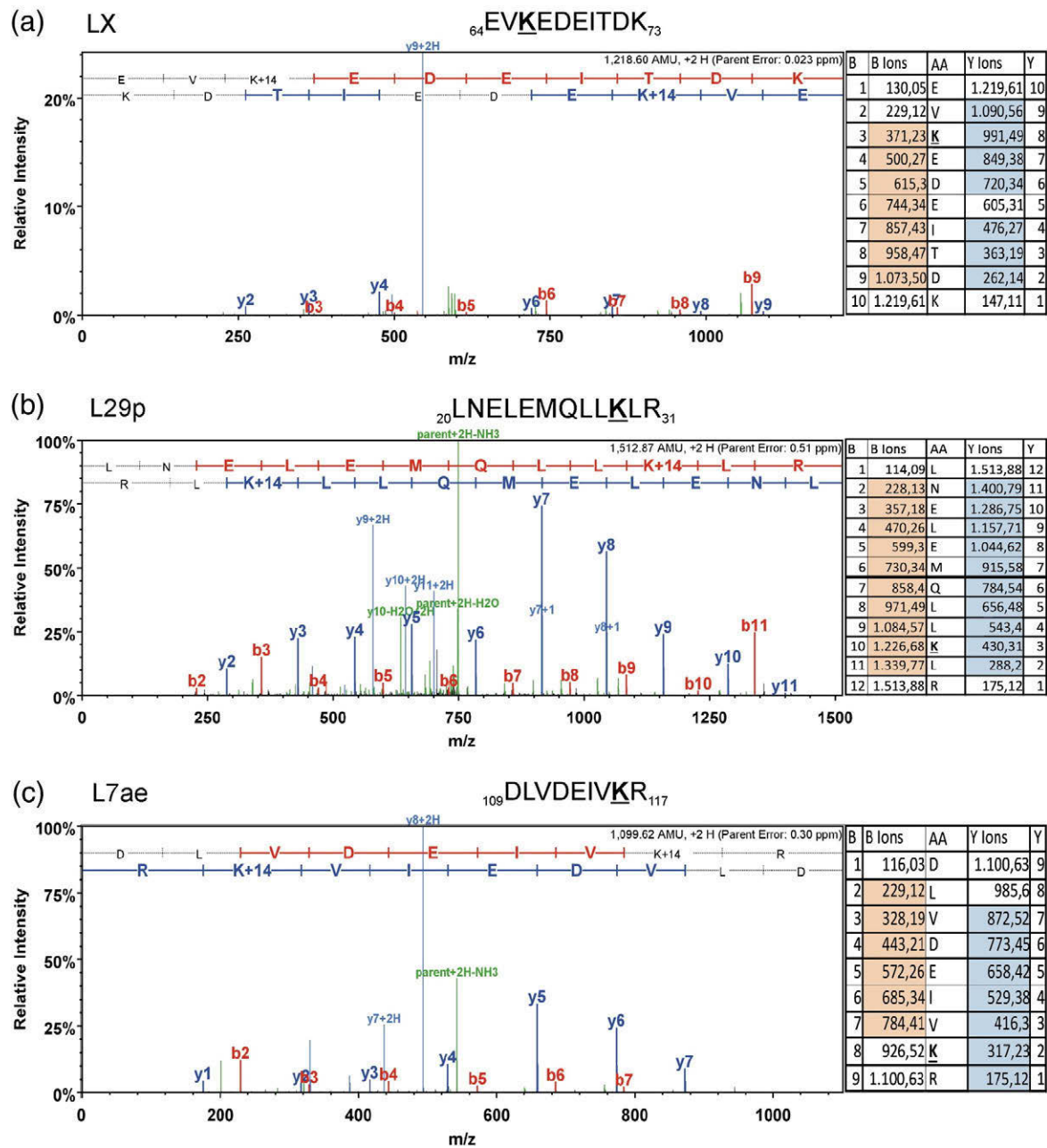
**Table 2 (continued)**

Protein family	LC-MS	2D	Amino acids	gi <sup>a</sup>	Molecular mass (kDa)	pI
L44e	✓	✓	91	18313777	10.6	10.9
LX	✓	✓	78	18313820	8.9	9.7
P0	✓	✓	345	18313827	38.0	9.9

<sup>a</sup> gi refers to the GenInfo identifier for sequence retrieval from NCBI.  
<sup>b</sup> Identified with only a single peptide  
<sup>c</sup> Not annotated in NCBI, but reading frame identified by Lecompte *et al.*<sup>31</sup>  
<sup>d</sup> Inferred by sequence homology identity to *Pyrobaculum calidifontis* JCM 11548.

to *E. coli*, where a low level of magnesium was necessary to split the 70S ribosomes into the component subunits (Fig. 4a), 30S and 50S subunits and little or no 70S ribosomes were observed for *S. acidocaldarius* and *P. aerophilum* even with low-salt and high -magnesium (30 mM) concentrations (data not shown). Similar observations have been reported previously for a variety of *Desulfurococcus* ribosomes, with 70S particles only being observed when cross-linking agents, such as formaldehyde, were used on actively translating ribosomes.<sup>50</sup> The fractions corresponding to the *S. acidocaldarius* and *P. aerophilum* 30S and 50S subunits were collected, pooled, and pelleted (Fig. 4a–c). The purified subunits were analyzed by negative-stain EM, as performed previously for archaeal subunits.<sup>51</sup> The images indicate that the purified subunits were homogeneous, although some contamination of the 30S subunits by chaperonins was evident, consistent with LC-MS/MS analysis (Fig. 4d). The purified *S. acidocaldarius* 30S and 50S subunits were then analyzed by 2D-PAGE, and the protein spots were identified by LC-MS/MS (Fig. 4e and f).

The r-protein assignments were consistent with those of the 2D gels of the high-salt ribosomal fraction and corresponded to the expected respective subunit composition (Fig. 2a). However, L7ae was also found at lower stoichiometry in the 30S fraction (Fig. 4e), which may reflect a promiscuous interaction with the 30S subunit, although we cannot exclude that it also results from the aforementioned involvement of L7ae in other biogenesis or processing complexes. Similarly, S24e is found at a surprisingly higher stoichiometry in the 50S subunit compared to the 30S subunit (Fig. 4e and f), possibly reflecting an as yet uncharacterized interaction of this protein with both ribosomal subunits. In contrast, it was immediately apparent that the spots for the CLP and the USP were absent. We also did not detect Rbp18 (Saci\_1216), which has been proposed to interact with the *Sulfolobus solfataricus* 30S subunit,<sup>52</sup> suggesting that CLP, USP, and Rbp18 are not integral r-proteins but may instead be only transiently associated with the



**Fig. 3.** MS identification of r-protein modifications. MS/MS spectra (left) and fragmentation tables (right) used for the identification of the monomethylation sites of r-proteins (a) LX, (b) L29p, and (c) L7ae. Monomethylated lysines within the peptide sequences are underlined and marked in boldface. Detected b-ions are highlighted in red, and y-ions are highlighted in blue.

ribosome. In contrast, the three hypothetical proteins Saci\_1218, Saci\_1337, and Saci\_1586 were present in the 2D gel of the *S. acidocaldarius* 50S subunit (but not in the 30S subunit), with intensities paralleling those of *bona fide* r-proteins (Fig. 4f). Saci\_1218, Saci\_1337, and Saci\_1586 are composed of 125 aa (14.6 kDa), 75 aa (8.8 kDa), and 80 aa (9.3 kDa), with *pI* values of 9.3, 9.9, and 8.0, respectively. We propose to rename Saci\_1218,

Saci\_1337, and Saci\_1586 as L45a, L46a, and L47a, respectively, continuing the numerical nomenclature of known archaeal r-proteins.

At the time of writing, the genomes of 91 archaeal species had been completely sequenced. Homology searches against these genomes revealed that L45a (Saci\_1218) has homologues present in all *Sulfolobus* species, with the exception of *Sulfolobus tokodaii* str. 7 (Fig. 5a). In contrast, L46a (Saci\_1337) appears to be

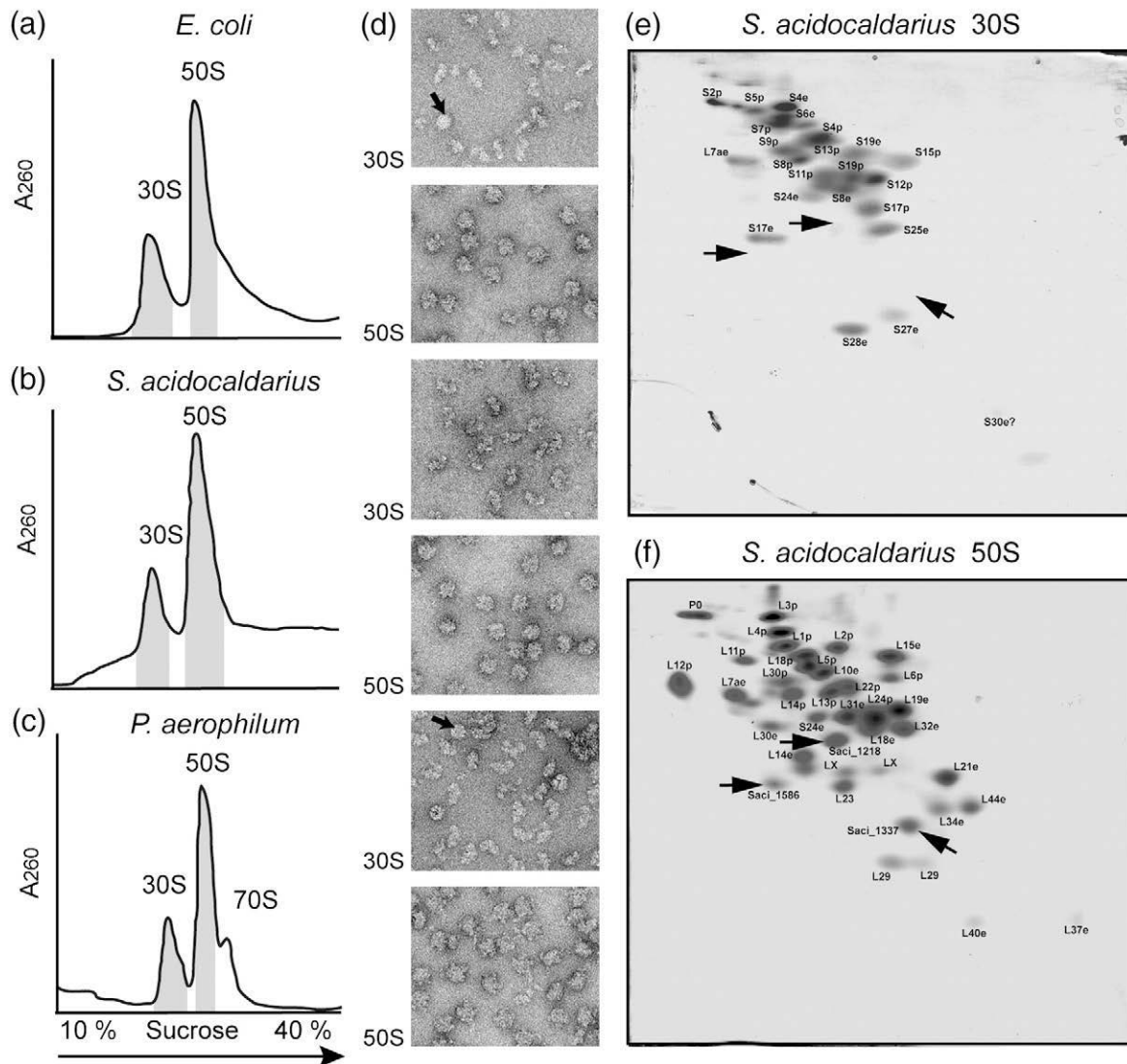
**Table 3.** Potential ribosome-associated proteins identified by LC-MS/MS and 2D-PAGE

Protein group	Protein description	Species	gi <sup>a</sup>	LC-MS	2D	
Translation-related factors	Elongation factor 1 $\alpha$ (EF-1 $\alpha$ )	PAE	18313751	30S	✓	
		SAC	70606488	—	✓	
	Translation initiation factor IF-6	PAE	18313819	30S/50S	—	
		SAC	70607209	50S	—	
	Translation initiation factor Sui1	PAE	18314004	30S	—	
		SAC	70606655	30S	—	
Ribosome biogenesis factors	Cbf5 (pseudo-uridine synthetase)	SAC	212373368	30S	—	
		PAE	18313923	30S	—	
	Fibrillarin (rRNA 2'-O-methyltransferase)	SAC	70607104	30S/50S	—	
		PAE	18313924	30S/50S	✓	
	NOP56 subunit (pre-rRNA processing protein)	SAC	70607105	50S	—	
		PAE	18314094	30S/50S	—	
Dim2p-like rRNA processing protein	SAC	70606751	30S	—		
	PAE	18313194	30S/50S	—		
RNA/DNA binding proteins	Chromatin-like protein (CLP)	SAC	70607080	30S/50S	✓	
		PAE	18312811	—	✓	
	TBP-interacting protein TIP49	SAC	162139950	30S	—	
	Transcription factor NusA-like protein	PAE	18312108	30S	✓	
	Zn-ribbon RNA binding protein (RBP)	SAC	70606567	30S	—	
		PAE	18313875	30S	—	
Proteases, heat shock proteins, and chaperones	AAA <sup>+</sup> protease	PAE	18312191	30S	✓	
		PAE	18313227	30S	✓	
	FKBP-type PPIase	SAC	70607398	—	✓	
		PAE	18313186	30S	—	
	Proteasome, $\alpha$ -subunit	SAC	70606428	50S	—	
		PAE	18313108	30S	✓	
	Thermosome (chaperonin), $\alpha$ -subunit	PAE	18313954	30S	✓	
		SAC	70606473	—	✓	
	Thermosome (chaperonin), $\beta$ -subunit	SAC	70607374	—	✓	
		SAC	70606787	—	✓	
	Other	ABC transporter	PAE	18160913	30S	✓
			PAE	18313383	30S	—
Acetyl-CoA acetyltransferase		PAE	18161785	30S/50S	—	
Acetyl-CoA acetyltransferase-associated protein		PAE	18161786	30S/50S	—	
Nitrate reductase, $\alpha$ -subunit (narG)		PAE	18159558	50S	—	
Nitrate reductase, $\beta$ -subunit (narH)		PAE	18159558	50S	—	
Succinate dehydrogenase subunit		SAC	70607907	—	✓	
Sulfurtransferase		PAE	18312079	50S	—	
PAE0659 (pI=9.3)		PAE	18312296	30S	—	
PAE0944 (pI=6.1)		PAE	18312573	30S/50S	—	
PAE1347 (pI=6.5)	PAE	18160290	30S/50S	✓		
PAE1683 (pI=9.2)	PAE	18160397	30S	—		
PAE1820 (pI=8.1)	PAE	18160809	50S	✓		
PAE2358 (pI=9.3)	PAE	18313855	30S/50S	—		
PAE3143 (pI=7.8)	PAE	18161642	30S/50S	—		
PAE3432 (pI=9.3)	SAC	70606685	30S	—		
Saci_0899 (pI=6.2)	SAC	70606987	50S	✓		
Saci_1218 (pI=9.3)	SAC	70607006	30S	—		
Saci_1237 (pI=9.4)	SAC	70607095	50S	✓		
Saci_1337 (pI=9.9)	SAC	70607187	30S	—		
Saci_1435 (pI=9.5)	SAC	70607322	50S	✓		
Saci_1586 (pI=8.0)	SAC	70607322	50S	✓		

<sup>a</sup> gi refers to the GenInfo identifier for sequence retrieval from NCBI.

more widely distributed, with homologues being found in all Thermoprotei genera (Fig. 5b) and with L47a (Saci\_1586) being found in all Sulfolobaceae (Fig. 5c). Initial homology searches did not reveal a homologue of L47a within the genome of *S. tokodaii* str. 7; however, careful inspection of the genome context of the neighboring genes of L47a, namely *asnC* and *rpoM*, revealed the presence of an unannotated open reading frame, which encodes a protein with 64% identity to L47a (Fig. 5g). While most archaeal r-proteins are organized either within

the major conserved operons ( $\alpha$ , *str*, *spc*, and *S10*) as in bacteria or in one of 10 archaeal-unique operons,<sup>53</sup> 11 archaeal r-proteins are not associated with any specific genomic context. Likewise, L45a and L46a do not appear to be within conserved operon structures, nor associated with any particular genes, although we note that the gene (Saci\_1216) encoding the 30S subunit binding protein Rbp18,<sup>52</sup> the ribosome biogenesis factor Gar1, and Ser-tRNA synthetase are in close proximity to L45a, L46a, and L47a (Fig. 5d–f), respectively.



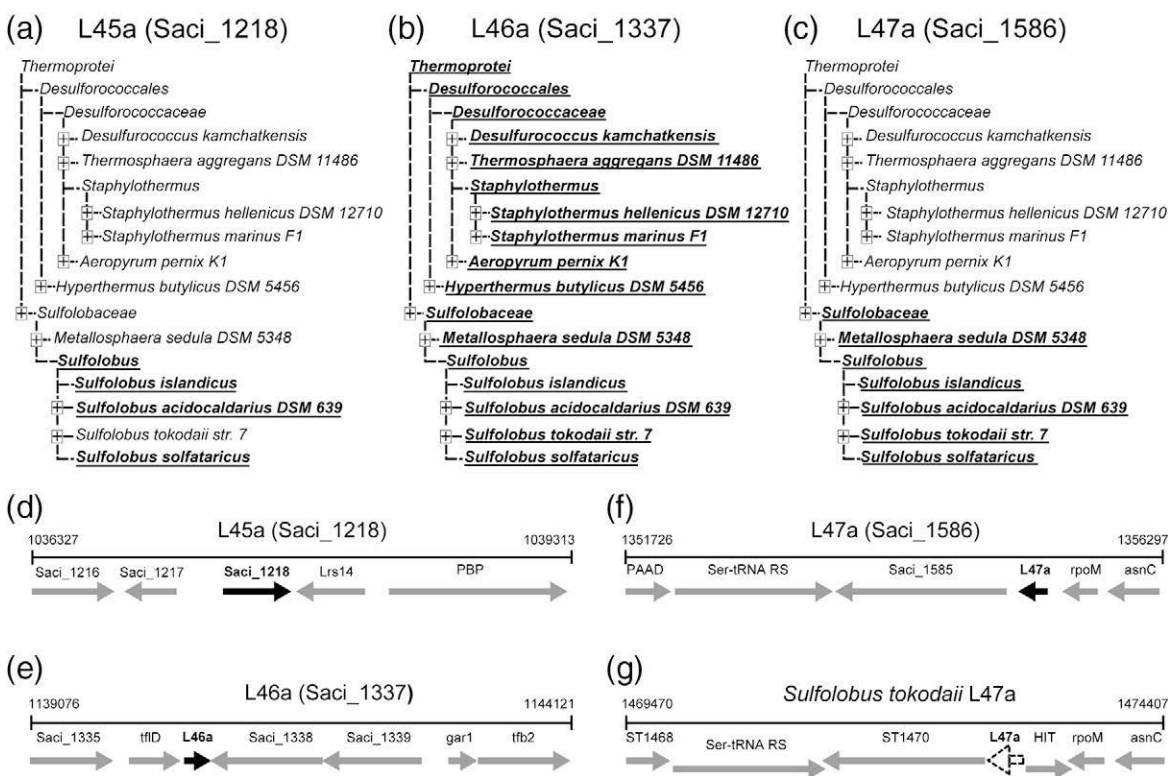
**Fig. 4.** Characterization of purified archaeal ribosomal subunits. Sucrose density gradient centrifugation profiles of HSW ribosomes from (a) *E. coli* as control, compared with (b) *S. acidocaldarius* and (c) *P. aerophilum* (left). (d) Negative-stain EM of 30S and 50S fractions isolated from (a)–(c) at a magnification of 90,000 $\times$ . The arrow indicates chaperonin particles in ribosome preparations. Two-dimensional PAGE of *S. acidocaldarius* (e) 30S and (f) 50S subunits. The positions of the novel r-proteins Saci\_1218, Saci\_1337, and Saci\_1586 are indicated by arrows in (e) and (f).

R-proteins identified in archaeal-unique operons are often found together with components of the transcriptional machinery (e.g., L30e is found together with genes encoding the transcription elongation factor NusA and the DNA-dependent RNA polymerase subunits RpoA, RpoB, and RpoH). This finding has been suggested to reflect the increasing coordination in the regulation of transcription and translation that has evolved in archaeal (and eukaryotic) lineages since the separation from bacterial phyla.<sup>53</sup> L45a neighbors the gene for a transcription regulator (Lrs14), whereas L46a is downstream of tflD, encoding the TATA binding protein of the transcription factor TFIID. Similarly,

L47a is often found downstream of genes encoding the transcriptional regulator AsnC and the DNA-directed RNA polymerase subunit M (RpoM).

#### **NTPase activities of *S. acidocaldarius* small and large subunits**

Ribosomes from higher eukaryotes, such as rat and bovine liver and rabbit reticulocytes, have been documented to harbor intrinsic ATPase activities.<sup>54,55</sup> This prompted us to examine the purified archaeal ribosomes to assess whether any intrinsic ATP or GTP activities could be detected. As controls, the intrinsic ATPase/GTPase activities of



**Fig. 5.** Distribution and genomic organization of Saci r-proteins. Subsection of the archaeal taxonomic tree showing the relative distributions of (a) Saci\_1218 (L45a), (b) Saci\_1337 (L46a), and (c) Saci\_1586 (L47a). Genomic organization of (d) Saci\_1218 (L45a), (e) Saci\_1337 (L46a), and (f) Saci\_1586 (L47a). (g) Identification of a homologue of Saci\_1586 in *S. tokodaii* str. 7 based on genomic context. Lrs14, transcription regulator Lrs14-like protein; PBP, periplasmic binding protein; tflD, TATA binding protein of the transcription factor TFIID; ttf2, transcription initiation factor 2; PAAD, phenylacetic acid degradation protein; Ser-tRNA RS, Ser-tRNA aminoacyl synthetase; RpoM, DNA-dependent RNA polymerase subunit M; Lrp, transcription regulator Lrp; HIT, histidine triad nucleotide binding protein.

*E. coli* 70S ribosomes, as well as those of the purified *E. coli* elongation factor G (EF-G), were tested in the presence and in the absence of *E. coli* 70S ribosomes. As seen in Fig. S3, no intrinsic ATPase or GTPase activity was detected for *E. coli* 70S ribosomes, nor for *S. acidocaldarius* 30S or 50S subunits, at 37 °C within 1 h. In comparison, high GTPase activity was detected within 1 h when *E. coli* EF-G was incubated together with *E. coli* 70S ribosomes. Similarly, after 24-h incubations, no intrinsic ATPase/GTPase activities were detected for *E. coli* 70S ribosomes. However, low ATPase and GTPase activities were detected for both *S. acidocaldarius* 30S or 50S subunits at levels similar to the intrinsic GTPase

activity of *E. coli* EF-G (Fig. S3). Similar experiments were performed with *Thermus thermophilus* 70S ribosomes and *S. acidocaldarius* 30S or 50S subunits at 60 °C, and no intrinsic ATPase or GTPase activity was detected after 1 h (data not shown). Longer incubations were not possible because of the elevated background signal due to a spontaneous cleavage of NTPs at 60 °C. Although we cannot completely rule out the presence of an intrinsic ATPase/GTPase activity of the archaeal subunits, we believe that the low activities detected are more consistent with the presence of residual contaminating factors, such as AAA<sup>+</sup> proteases or EF-1 $\alpha$ , that were detected in the samples by MS (Table 3). Such a

**Fig. 6.** Two-dimensional gels for the incremental washing of *S. acidocaldarius* 50S subunits. (a) Coomassie-blue-stained 2D-PAGE of 1–4 M LiCl wash fractions from purified *S. acidocaldarius* 50S subunits, as well as the remaining r-proteins in the core 50S subunit after treatment with 4 M LiCl. (b) Scheme illustrating the subset of *S. acidocaldarius* 50S subunit r-proteins that are removed with increasing concentrations of LiCl (red, orange, and lime), as well as those remaining in the core after 4 M washing (blue). (c) Two views of a model for the *S. acidocaldarius* 50S subunits with r-proteins removed by 1.0 M, 1.5–2 M, and 3–4 M washing, as well as those remaining after 4 M washing, shown in red, orange, lime, and blue, respectively.

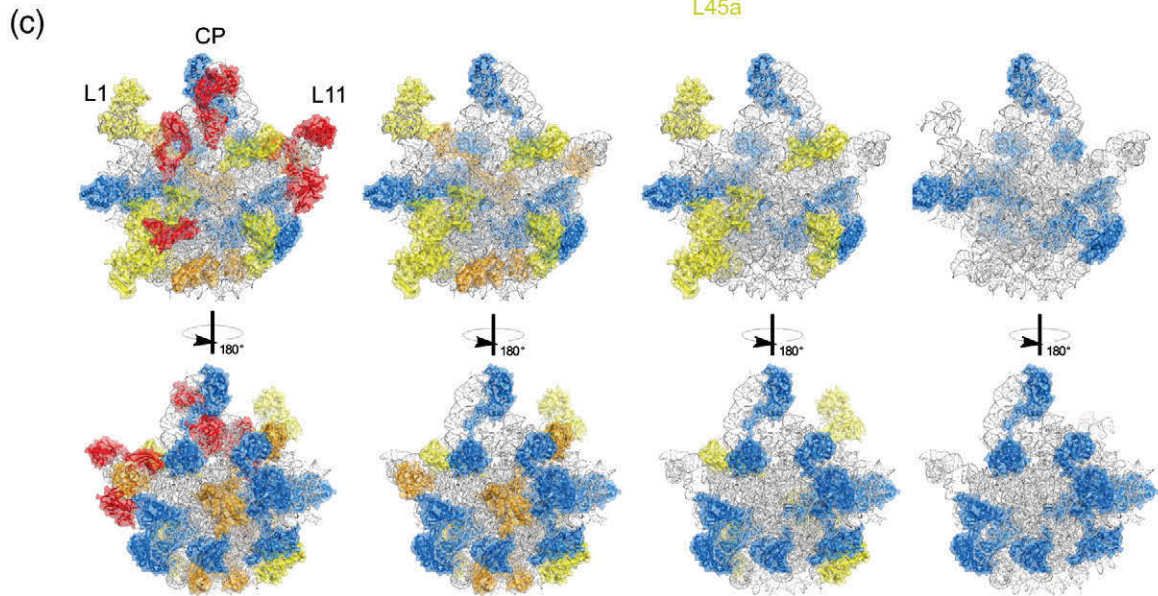
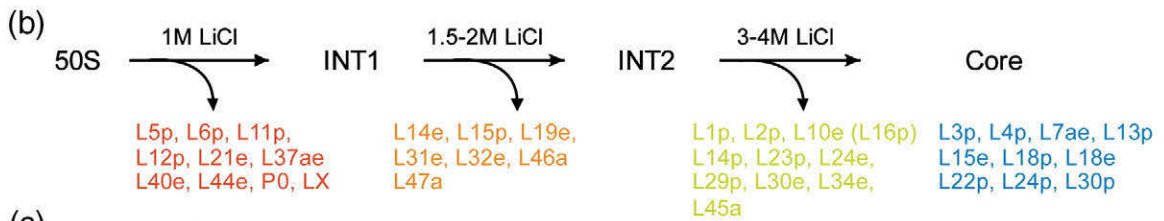
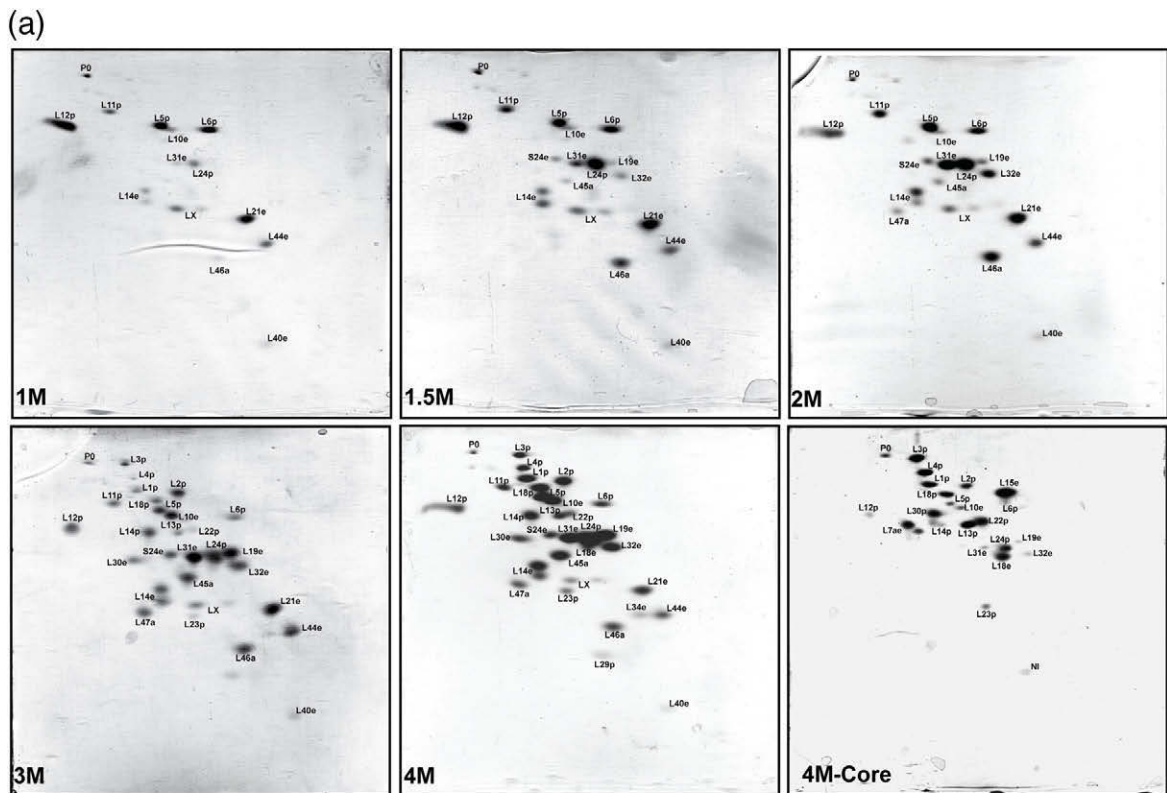


Fig. 6 (legend on previous page)

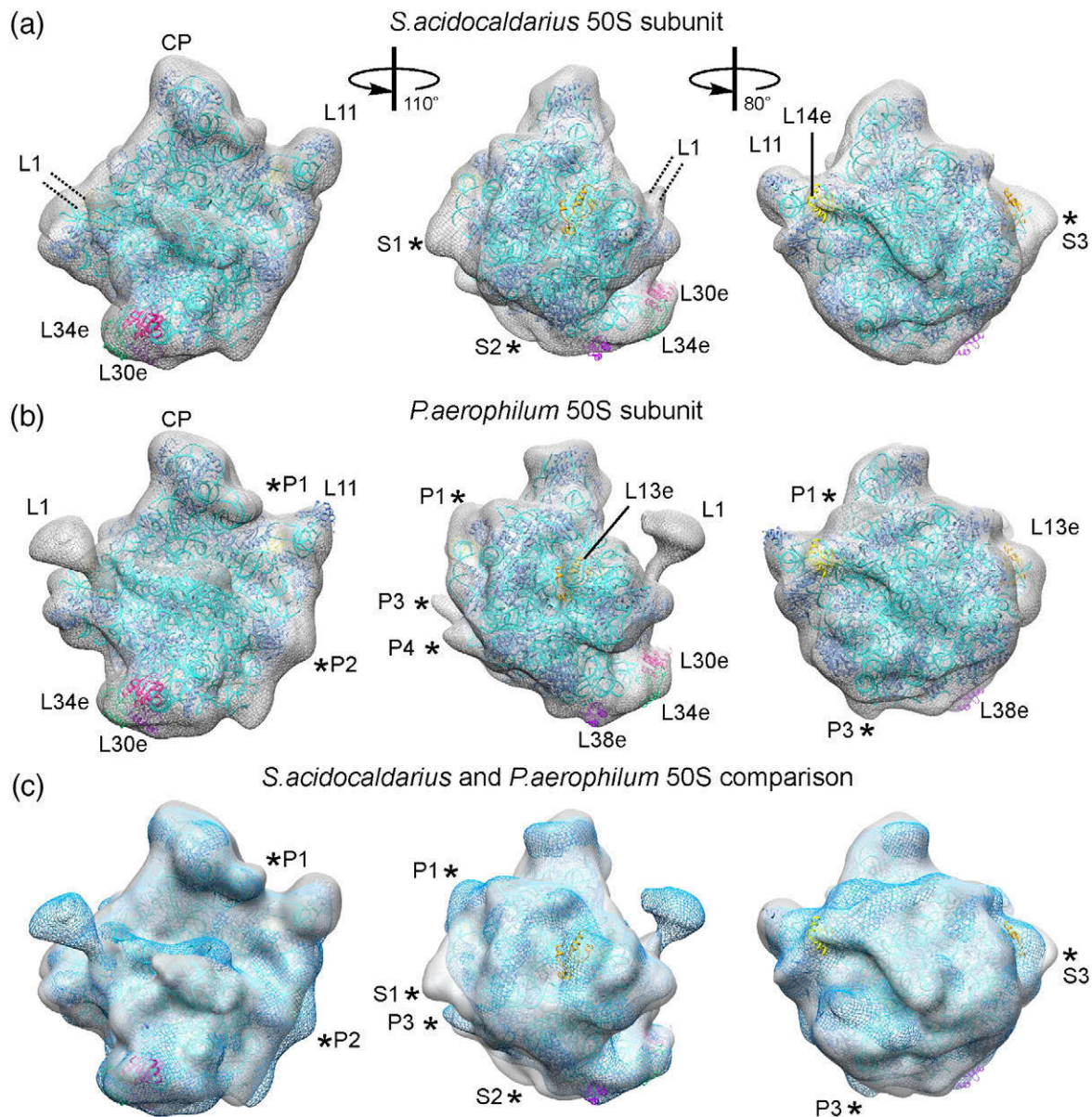


conclusion would be consistent with the lack of any highly conserved ATP and/or GTP binding motifs in any of the archaeal r-proteins.

### Stepwise dissociation of archaeal 50S ribosomal subunits by LiCl

To investigate the interaction and stability of the novel archaeal r-proteins L45a, L46a, and L47a on the

50S subunit relative to other known r-proteins, we washed the *S. acidocaldarius* 50S subunit with increasing concentrations of LiCl (1 M, 1.5 M, 2 M, 3 M, or 4 M), precipitated the r-proteins that dissociated from the particles, and analyzed them by 2D-PAGE (Fig. 6a). In addition, 2D-PAGE was also performed on the core r-proteins that remain attached to the rRNA after washing with 4 M LiCl (Fig. 6a). A representation summarizing the order of



**Fig. 7.** Cryo-EM reconstructions of 50S subunits from *S. acidocaldarius* and *P. aerophilum*. Three different views of the cryo-EM maps (gray mesh) of 50S subunits from (a) *S. acidocaldarius* and (b) *P. aerophilum*, with a fitted crystal structure of the *H. marismortui* 50S subunit (PDB code 3CC2)<sup>63</sup> shown with RNA (cyan) and r-proteins (blue) as ribbons, with additional archaeal/eukaryotic-specific r-proteins included from Armache *et al.*<sup>8</sup> (c) Comparison of the density maps of *S. acidocaldarius* (gray) and *P. aerophilum* (blue mesh). The density for r-proteins L30e and L34e is present in both *S. acidocaldarius* and *P. aerophilum* 50S subunits, whereas the density for L13e and L38e is present in only *P. aerophilum*. Potential locations for additional r-proteins in *S. acidocaldarius* and *P. aerophilum* are indicated with asterisks and labeled S1–S3 and P1–P4, respectively.

r-protein release by LiCl is shown schematically in Fig. 6b and graphically on a molecular model for the *S. acidocaldarius* 50S subunit in Fig. 6c. Surprisingly, although many integral r-proteins (L5p, L6p, L11p, L12p, L21e, L37ae, L44e, LX, and P0) were washed from the 50S subunit with 1 M LiCl, the r-proteins L45a, L46a, and L47a were not. L46a was removed with 1.5 M LiCl, whereas L45a and L47a only started to appear in the supernatant with 2–3 M LiCl washing, suggesting a tighter association with the 50S subunit (Fig. 6a). Generally, the trend is that the 3–4 M LiCl wash and core fractions contain predominantly universally conserved families [L1p, L2p, L3p, L4p, L13p, L14p, L16p (L10e), L18p, L22p, L24p, L29p, and L30p], with only a few archaeal/eukaryotic-specific families present (L7ae, L15e, L18e, L30e, L31e, and L34e). In contrast, the reverse trend is observed in the 1–2 M LiCl wash fractions, which contain more archaeal/eukaryotic-specific r-proteins than universal r-proteins. The dissociating universal r-proteins include L11p, L12p, and P0 (L10p) that form the stalk region, which is known to be easily removable from the ribosome.<sup>56,57</sup> Studies using bacterial ribosomes have revealed that the susceptibility of r-proteins to removal by LiCl washing corresponds inversely to the order of assembly of r-proteins *in vivo* and from reconstitution studies.<sup>58</sup> Two-dimensional PAGE analysis of the *S. solfataricus* 50S core r-proteins following treatment with 3 M LiCl/6 M urea revealed a number of r-proteins; however, no direct identification of the r-proteins was performed.<sup>39</sup> Although the *in vitro* reconstitution of *S. solfataricus* and *Haloferax mediterranei* large ribosomal subunits was successfully performed, the order and dependency of r-proteins were not investigated.<sup>59,60</sup> Nevertheless, the findings here are generally consistent with the LiCl treatment<sup>61</sup> and *in vivo* analysis of the yeast 60S subunit assembly.<sup>62</sup>

### Cryo-EM reconstruction of archaeal large ribosomal subunits

Cryo-EM reconstructions of the *S. acidocaldarius* and *P. aerophilum* 50S subunits were determined to 27 Å and 25 Å (0.5 FSC) resolutions, respectively (Fig. 7). The resolution was limited by the need to use 50S subunits, rather than 70S ribosomes, in part due to an orientation bias of the 50S subunits on the cryogrids. Nevertheless, there was an excellent fit of the crystal structure of the archaeal *H. marismortui* 50S subunit<sup>63</sup> into both cryo-EM maps (Fig. 7a and b). The 23S rRNA sequences of *S. acidocaldarius* and *P. aerophilum* are very similar to that of *H. marismortui*; however, *S. acidocaldarius* and *P. aerophilum* 50S subunits contain additional r-proteins that are not present in *H. marismortui* (Fig. 1).<sup>31</sup> Consistently, both maps had additional regions of density that were not occupied by the *H. marismortui*

50S subunit model (Fig. 7). Extra density is located on both maps on the interface side of the 50S subunit, just below the L1 stalk, consistent with the positions of L30e and L34e in eukaryotic 80S ribosomes<sup>8,64</sup> (Fig. 7a and b). Similarly, additional density is located between the L11 stalk and the central protuberance on the solvent side of the large subunit where L14e is located.<sup>8</sup> The *P. aerophilum* 50S subunit has two r-proteins (L13e and L38e) that are not found in *S. acidocaldarius* or *H. marismortui* 50S subunits (Fig. 1).<sup>31</sup> Additional density (which is absent in the *S. acidocaldarius* 50S map) is observed in the *P. aerophilum* 50S map in the two positions where L13e and L38e are located in eukaryotic 80S ribosomes<sup>8</sup> (Fig. 7c). The location of the archaeal-specific r-protein LX, which is present in both *S. acidocaldarius* and *P. aerophilum*, cannot be localized at this resolution. We could, however, identify at least three regions (S1–S3) with the remaining additional density in the *S. acidocaldarius* 50S map that are not present in the *P. aerophilum* 50S map, which could potentially be due to the additional *S. acidocaldarius* L45a, L46a, and L47a identified in this study; however, a higher resolution will be required to confirm this. Four regions (P1–P4) of additional density that are specific for the *P. aerophilum* 50S map were also observed (Fig. 7b and c), possibly reflecting the presence of additional as yet unknown *P. aerophilum* r-proteins, candidates for which were identified by LC-MS/MS in the 50S fraction (e.g. PAE0659 and PAE2358) (Table 3).

## Materials and Methods

### Growth of *S. acidocaldarius* and *P. aerophilum*

*S. acidocaldarius* (DSM 639<sup>T</sup>) and *P. aerophilum* IM2<sup>T</sup> (DSM 7523<sup>T</sup>) were obtained from the culture collection of the Institute of Microbiology and Archaeal Center, University of Regensburg. *S. acidocaldarius* was grown under oxic conditions in modified ALLEN medium at 75 °C and pH 3.0,<sup>65,66</sup> as described previously.<sup>67</sup> As substrate, 0.05% yeast extract, 0.2% saccharose, and 0.2% peptone were added. *P. aerophilum* was grown in BS medium at 97 °C, as previously described.<sup>68</sup> The medium was covered with a gas-phase H<sub>2</sub>/CO<sub>2</sub> (80:20 vol/vol, 250 kPa), thereby using hydrogen as electron donor and nitrate (0.1%) as electron acceptor. Mass cultivation for both strains was carried out in 300 L of enamel-protected fermenters (HTE; BioEngineering, Wald, Switzerland). The cells were harvested by centrifugation (Padberg, Lahr, Germany), shock-frozen in liquid nitrogen, and stored at –80 °C until further use.

### Preparation of ribosomal particles

Archaeal ribosomes were isolated and purified similarly as previously reported,<sup>50,69</sup> but with some modifications. Briefly, cell pellets were dissolved in Tico buffer [20 mM

Hepes (pH 7.5), 10 mM Mg(OAc)<sub>2</sub>, 30 mM NH<sub>4</sub>OAc, and 4 mM β-mercaptoethanol] at 4 °C and subsequently disrupted with Microfluidizer (Microfluidics M-110L Pneumatic) at 18,000 psi. The crude homogenate was centrifuged twice at 30,000g at 4 °C for 30 min in order to obtain the S30 fraction. A crude ribosomal fraction was obtained by centrifugation at 100,000g for 5 h at 4 °C and by dissolution of the pellet in an equal volume of HSW buffer [20 mM Hepes, 10 mM Mg(OAc)<sub>2</sub>, 500 mM NH<sub>4</sub>OAc, and 4 mM β-mercaptoethanol (pH 7.5)]. Large debris was removed by centrifuging the crude ribosomes for 5 min at 18,000g at 4 °C. The clear supernatant was diluted 10-fold in HSW buffer and layered on top of 1.3 vol of 25% (wt/vol) sucrose cushion prepared in HSW buffer and centrifuged at 100,000g for 7 h at 4 °C. The pellet was resuspended in a minimal volume of Tico buffer and subsequently purified using sucrose density gradient centrifugation (10–40% sucrose in Tico buffer) at 46,000g for 17 h at 4 °C. Fractions corresponding to 50S and 30S subunits were separately pooled, pelleted at 140,000g for 12 h at 4 °C, and resuspended in a minimal volume of Tico buffer. *E. coli* ribosomes were prepared in accordance with Blaha *et al.*<sup>70</sup>

### Extraction of total r-proteins and 2D-PAGE

The total proteins from HSW ribosomes and purified ribosomal subunits were extracted by acetic acid in accordance with Nierhaus and Dohme<sup>71</sup>. Lyophilized proteins were further processed for LC-MS/MS analysis and 2D-PAGE. Around 2 μg of total proteins was necessary for LC-MS/MS, whereas 5–10 μg of total proteins was required for 2D-PAGE. Two-dimensional electrophoresis was performed as described by Kaltschmidt and Wittmann.<sup>14</sup>

### Analysis of ribosomal particles by stepwise LiCl washing

Stepwise washing of purified 50S ribosomal subunits from *S. acidocaldarius* was performed in accordance with Homann and Nierhaus, with minor modifications.<sup>72</sup> Briefly, five A<sub>260</sub> units of ribosomal particles were incubated in a total volume of 100 μl with 10 mM Tris-HCl (pH 7.8), 10 mM MgCl<sub>2</sub>, 0.1 mM ethylenediaminetetraacetic acid, and appropriate concentrations of LiCl (0.6 M, 1 M, 1.5 M, 2 M, 3 M, and 4 M) for 4 h at 4 °C, with moderate shaking. The incubation reaction was further centrifuged at 160,000g for 12 h at 4 °C, and the dissociated proteins (supernatants) were precipitated with trichloroacetic acid. Lyophilized proteins were treated as described above for LC-MS/MS analysis and 2D-PAGE. In those cases where the core particle was analyzed (4 M LiCl core), the pellet obtained after ultracentrifugation was dissolved in Tico buffer, and the core proteins were extracted by acetic acid as described previously by Nierhaus and Dohme.<sup>71</sup>

### Mass spectrometry

#### Excision and tryptic in-gel digestion of 2D gel spots

Spots of interest were automatically excised using a Proteiner robot (Bruker Daltonics, Leipzig, Germany) and transferred onto 96-well reaction plates (Intavis,

Köln, Germany). Digests were performed using a DigestPro MS digest robot (Intavis) using the following a protocol: (i) washing with 60 μl of 100% CH<sub>3</sub>CN; (ii) rinsing with 45 μl of 50 mM NH<sub>4</sub>HCO<sub>3</sub>; (iii) washing with 60 μl of CH<sub>3</sub>CN; (iv) 20-min incubation with 30 μl of 10 mM dithiothreitol in water at 65 °C; (v) 20-min incubation with 30 μl of 50 mM iodoacetamide in water; (vi) two 20-min washes with 60 μl of 50 mM NH<sub>4</sub>HCO<sub>3</sub>; (vii) one 20-min wash in 60 μl of CH<sub>3</sub>CN; (viii) 15-min wash in 45 μl of CH<sub>3</sub>CN; (ix) addition of 90 ng of modified porcine trypsin (Promega, Madison, WI, USA) in 15 μl of 50 mM NH<sub>4</sub>HCO<sub>3</sub> and incubation at 37 °C for 6 h; (x) addition of 15 μl of 2.5% formic acid (FA); and (xi) collection of the supernatant peptide extracts on 96-well collection plates (Intavis).

### Matrix-assisted laser desorption/ionization MS-based identification of 2D gel spots

Matrix-assisted laser desorption/ionization (MALDI) targets were prepared automatically using a DigestPro MS robot (Intavis) equipped with ZipTip C-18 reversed-phase tips (Millipore, Billerica, MA, USA) for desalting. The following protocol was used: (i) activation of ZipTips with 20 μl of 50% CH<sub>3</sub>CN and 0.1% trifluoroacetic acid (TFA); (ii) washing with 20 μl of 0.1% TFA; (iii) loading of the tryptic digest; (iv) washing with 20 μl of 0.1% TFA; and (v) elution of peptides onto the MALDI target plate with 1–2 μl of matrix solution (8 mg/ml CHCA, 65% CH<sub>3</sub>CN, and 0.1% TFA). MALDI TOF/TOF MS was performed on a 4800 MALDI TOF/TOF Analyzer (Applied Biosystems). For precursor ion scans, a mass range between 800 Da and 4000 Da was chosen, and a minimum signal-to-noise ratio of 20 was set for peak detection. The MS/MS spectra of the 20 most intense precursors within each spot were generated by 500 laser shots.

### LC-MS/MS analysis of protein samples on the LTQ ion trap

Peptide samples were diluted in 40 μl of 0.1% FA and centrifuged for 15 min at 14,000 rpm at room temperature. A multidimensional liquid chromatography system (Ettan MDLC; GE Healthcare) connected upstream of the spectrometer was used for separation. Peptide samples were loaded onto a trap column (C18 PepMap100, 5 μm particle size, 100 Å, 300 μm × 5 mm column size; LC Packings Dionex) at a flow rate of 10 μl/min and subsequently separated by an RP column (C18 PepMap 100, 3 μm bead size, 75 μm i.d., 15 cm length; LC Packings) with an 80-min gradient from 0% solvent B to 30% solvent B (solvent A: 0.1% FA; solvent B: 84% CH<sub>3</sub>CN/0.1% FA), followed by a 30-min gradient to 60% solvent B at a flow rate of 280 nl/min. Electrospray ionization was performed with a distal coated SilicaTip (FS-360-20-10-D-20; New Objective) at a needle voltage of 1.2 kV. MS/MS of peptide samples was performed on a linear ion-trap mass spectrometer (LTQ; Thermo Electron). MS and MS/MS analyses were performed using cycles of one MS scan (mass range *m/z* 300–1600) and three subsequent data-dependent MS/MS scans ("Dynamic Exclusion™ activated"; 35% collision energy).

### LC-MS/MS analysis of protein samples on the Orbitrap XL instrument

Chromatographic separation of peptides was performed as described previously for LTQ ion-trap measurements, but using a Reprosil-Pur C18 separation column (Reprosil-Pur C18 AQ, 3  $\mu\text{m}$ , 150 nm  $\times$  75  $\mu\text{m}$ ; Dr. Maisch). Electrospray ionization was performed with a distal coated SilicaTip (FS-360-20-10-D-20; New Objective) and a needle voltage of 1.4 kV. In order to detect low-molecular-mass ammonium ions, we performed MS and MS/MS analyses using cycles of one MS scan (mass range  $m/z$  300–2000) and three subsequent data-dependent CID MS/MS scans, followed by three HCD MS/MS scans ("Dynamic Exclusion™ activated"; 35% collision energy).

### Database search and data analysis

MS/MS data were searched with Mascot version 2.1.03 (Matrix Science, Boston, MA, USA) using the NCBI nr 20091028 database and the following parameters: (i) enzyme: trypsin; (ii) fixed modification: carbamidomethyl (C); (iii) variable modifications: oxidation (M), methyl (K), and *N*-acetyl (protein); (iv) peptide mass tolerance: 2 Da for electrospray ionization measurements and 100 ppm for MALDI measurements; (v) MS/MS mass tolerance: 0.8 Da; (vi) peptide charges: 1+, 2+, and 3+; (vii) instrument: electrospray ionization trap; and (viii) allow up to one missed cleavage. Mascot results were further validated with the Scaffold software V 2.6 (Proteome Software, Inc., Portland, OR, USA). Protein identification was accepted if it could be established at >99.0% probability. In order to determine the number of false-positive identifications, we repeated MASCOT searches and Scaffold evaluations with a randomized version of the NCBI nr database. For the generation of this so-called "decoy database" consisting of random sequences with the same average amino acid composition, the decoy PerlScript (Matrix Science) was used.

### Malachite Green GTPase activity assays

GTPase and ATPase activities were measured using the Malachite Green Phosphate Kit (BioAssay) that quantifies the green complex formed between Malachite Green, molybdate, and free orthophosphate, as described previously.<sup>73</sup> All reactions contained 30–90 nM *E. coli* 70S ribosomes, *S. acidocaldarius* 30S or 50S subunits, 20  $\mu\text{M}$  GTP or ATP, and/or 30–90 nM EF-G. Reactions were transferred onto 96-well microtiter plates and incubated at 37 °C or 60 °C for various lengths of time. Color formation was measured on Tecan-Infinite M1000 microplate reader at 650 nm. Reactions performed in the absence of ribosomes were used as background signal to account for the spontaneous hydrolysis of ATP or GTP.

### Electron microscopy

#### Negative-stain EM

Ribosomal particles were resuspended in Tico buffer to a final concentration of 0.5–1  $A_{260}/\text{ml}$ . One drop of each sample was deposited on carbon-coated grids. After 45 s, the grids were washed with distilled water. The grids were then stained with three drops of 2% aqueous uranyl

acetate for 15 s, and excess liquid was removed again by touching the grids with filter paper. Micrographs were taken with a Morgagni transmission electron microscope (FEI; 80 kV) equipped with a wide-angle 1000 CCD camera.

#### Cryo-EM and single-particle reconstruction

As described previously,<sup>74</sup> *S. acidocaldarius* and *P. aerophilum* 50S subunits were applied to carbon-coated holey grids. Images were collected on a Tecnai G2 Spirit TEM at 120 kV at a nominal magnification of 90,000 $\times$  using an Eagle 4096  $\times$  4096-pixel CCD camera (FEI), resulting in a pixel size of 3.31 Å/pixel. The data were analyzed by determining the contrast transfer function with CTFFIND<sup>75</sup> and were further processed with the SPIDER software package,<sup>76</sup> using the *H. marismortui* 50S subunit [Protein Data Bank (PDB) code 3CC2]<sup>63</sup> filtered to between 20 Å and 25 Å as initial reference. For the final *S. acidocaldarius* and *P. aerophilum* 50S subunit reconstructions, 9301 and 9183 particles were used, respectively.

#### Modeling and figure preparation

The model for the *S. acidocaldarius* 50S subunit utilized the crystal structure of the archaeon *H. marismortui* 50S subunit (PDB code 3CC2),<sup>63</sup> with additional missing proteins added using the location of the homologues in the eukaryotic 80S ribosome.<sup>8,77</sup> The models were fitted to the maps using Chimera,<sup>78</sup> and Figs. 5c and 6 were prepared using PyMOL (The PyMOL Molecular Graphics System, version 1.3, Schrödinger LLC) and Chimera, respectively.

#### Accession numbers

The cryo-EM maps of the *S. acidocaldarius* and *P. aerophilum* 50S subunits have been deposited in the three-dimensional EM database under accession numbers EMD-1797 and EMD-1797, respectively.

Supplementary materials related to this article can be found online at [doi:10.1016/j.jmb.2010.11.055](https://doi.org/10.1016/j.jmb.2010.11.055)

### Acknowledgements

This research was supported by Deutsche Forschungsgemeinschaft grants SFB594 and SFB646 (to R.B.) and WI3285/1-1 (to D.N.W.).

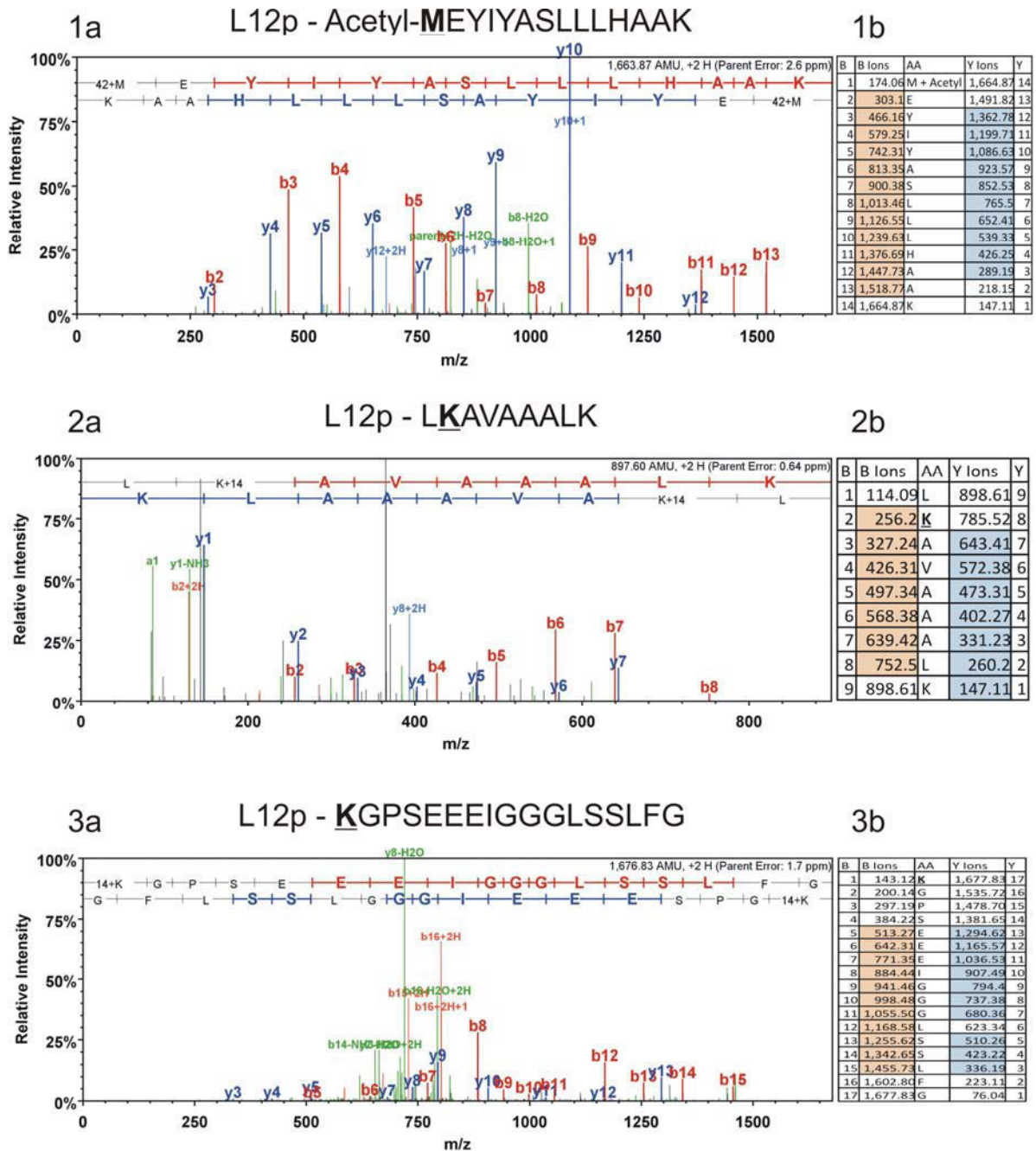
### References

- Schmeing, T. M. & Ramakrishnan, V. (2009). What recent ribosome structures have revealed about the mechanism of translation. *Nature*, **461**, 1234–1242.
- Wilson, D. N. & Nierhaus, K. H. (2005). Ribosomal proteins in the spotlight. *Crit. Rev. Biochem. Mol. Biol.* **40**, 243–267.
- Brodersen, D. & Nissen, P. (2005). The social life of ribosomal proteins. *FEBS J.* **272**, 2098–2108.

4. Ogle, J. M. & Ramakrishnan, V. (2005). Structural insights into translational fidelity. *Annu. Rev. Biochem.* **74**, 129–177.
5. Maguire, B. A., Beniaminov, A. D., Ramu, H., Mankin, A. S. & Zimmermann, R. A. (2005). A protein component at the heart of an RNA machine: the importance of protein I27 for the function of the bacterial ribosome. *Mol. Cell*, **20**, 427–435.
6. Voorhees, R. M., Weixlbaumer, A., Loakes, D., Kelley, A. C. & Ramakrishnan, V. (2009). Insights into substrate stabilization from snapshots of the peptidyl transferase center of the intact 70S ribosome. *Nat. Struct. Mol. Biol.* **16**, 528–533.
7. Bhushan, S., Meyer, H., Starosta, A., Becker, T., Mielke, T., Berninghausen, O. *et al.* (2010). Structural basis for translational stalling by human cytomegalovirus (hCMV) and fungal arginine attenuator peptide (AAP). *Mol. Cell*, **40**, 138–146.
8. Armache, J.-P., Jarasch, A., Anger, A. M., Villa, E., Becker, T., Blushan, S. *et al.* (2010). Localization of eukaryote-specific ribosomal proteins: implications for structure, function and evolution. *Proc. Natl Acad. Sci. USA*, **107**, 19754–19759.
9. Meskauskas, A. & Dinman, J. D. (2001). Ribosomal protein L5 helps anchor peptidyl-tRNA to the P-site in *Saccharomyces cerevisiae*. *RNA*, **7**, 1084–1096.
10. Yusupov, M. M., Yusupova, G. Z., Baucom, A., Lieberman, K., Earnest, T. N., Cate, J. H. & Noller, H. F. (2001). Crystal structure of the ribosome at 5.5 Å resolution. *Science*, **292**, 883–896.
11. Selmer, M., Dunham, C., Murphy, F. T., Weixlbaumer, A., Petry, S., Kelley, A. *et al.* (2006). Structure of the 70S ribosome complexed with mRNA and tRNA. *Science*, **313**, 1935–1942.
12. Jenner, L., Demeshkina, N., Yusupova, G. & Yusupov, M. (2010). Structural rearrangements of the ribosome at the tRNA proofreading step. *Nat. Struct. Mol. Biol.* **17**, 1072–1078.
13. Fischer, N., Konevega, A. L., Wintermeyer, W., Rodnina, M. V. & Stark, H. (2010). Ribosome dynamics and tRNA movement by time-resolved electron cryomicroscopy. *Nature*, **466**, 329–333.
14. Kaltschmidt, E. & Wittmann, H. G. (1970). Ribosomal proteins: VII. 2D polyacrylamide gel electrophoresis for fingerprinting of ribosomal proteins. *Anal. Biochem.* **36**, 401–412.
15. Yamaguchi, K. & Subramanian, A. R. (2000). The plastid ribosomal proteins. Identification of all the proteins in the 50S subunit of an organelle ribosome (chloroplast). *J. Biol. Chem.* **275**, 28466–28482.
16. Yamaguchi, K., von Knoblauch, K. & Subramanian, A. R. (2000). The plastid ribosomal proteins. Identification of all the proteins in the 30S subunit of an organelle ribosome (chloroplast). *J. Biol. Chem.* **275**, 28455–28465.
17. Sharma, M. R., Wilson, D. N., Datta, P. P., Barat, C., Schluenzen, F., Fucini, P. & Agrawal, R. K. (2007). Cryo-EM study of the spinach chloroplast ribosome reveals the structural and functional roles of plastid-specific ribosomal proteins. *Proc. Natl Acad. Sci. USA*, **104**, 19315–19320.
18. Sharma, M. R., Donhofer, A., Barat, C., Marquez, V., Datta, P. P., Fucini, P. *et al.* (2010). PSRP1 is not a ribosomal protein, but a ribosome-binding factor that is recycled by the ribosome-recycling factor (RRF) and elongation factor G (EF-G). *J. Biol. Chem.* **285**, 4006–4014.
19. Koc, E. C., Burkhart, W., Blackburn, K., Moseley, A. & Spremulli, L. L. (2001). The small subunit of the mammalian mitochondrial ribosome. Identification of the full complement of ribosomal proteins present. *J. Biol. Chem.* **276**, 19363–19374.
20. Koc, E. C., Burkhart, W., Blackburn, K., Moyer, M. B., Schlatzer, D. M., Moseley, A. & Spremulli, L. L. (2001). The large subunit of the mammalian mitochondrial ribosome. Analysis of the complement of ribosomal proteins present. *J. Biol. Chem.* **276**, 43958–43969.
21. Smits, P., Smeitink, J. A., van den Heuvel, L. P., Huynen, M. A. & Ettema, T. J. (2007). Reconstructing the evolution of the mitochondrial ribosomal proteome. *Nucleic Acids Res.* **35**, 4686–4703.
22. Lee, S., Berger, S., Martinovic, S., Pasa-Tolic, L., Anderson, G., Shen, Y. *et al.* (2002). Direct mass spectrometric analysis of intact proteins of the yeast large ribosomal subunit using capillary LC/FTICR. *Proc. Natl Acad. Sci. USA*, **99**, 5942–5947.
23. Link, A., Eng, J., Schieltz, D., Carmack, E., Mize, G., Morris, D. *et al.* (1999). Direct analysis of protein complexes using mass spectrometry. *Nat. Biotechnol.* **17**, 676–682.
24. Alonso, J. & Santaren, J. F. (2006). Characterization of the *Drosophila melanogaster* ribosomal proteome. *J. Proteome Res.* **5**, 2025–2032.
25. Giavalisco, P., Wilson, D., Kreidler, T., Lehrach, H., Klose, J., Gobom, J. & Fucini, P. (2005). High heterogeneity within the ribosomal proteins of the *Arabidopsis thaliana* 80S ribosome. *Plant Mol. Biol.* **57**, 577–591.
26. Louie, D. F., Resing, K. A., Lewis, T. S. & Ahn, N. G. (1996). Mass spectrometric analysis of 40S ribosomal proteins from rat-1 fibroblasts. *J. Biol. Chem.* **271**, 28189–28198.
27. Wool, I. G., Chan, Y. L. & Glück, A. (1995). Structure and evolution of mammalian ribosomal proteins. *Biochem. Cell Biol.* **73**, 933–947.
28. Sugihara, Y., Honda, H., Iida, T., Morinaga, T., Hino, S., Okajima, T. *et al.* (2010). Proteomic analysis of rodent ribosomes revealed heterogeneity including ribosomal proteins L10-like, L22-like 1, and L39-like. *J. Proteome Res.* **9**, 1351–1366.
29. Vladimirov, S. N., Ivanov, A. V., Karpova, G. G., Musolyamov, A. K., Egorov, T. A., Thiede, B. *et al.* (1996). Characterization of the human small-ribosomal-subunit proteins by N-terminal and internal sequencing, and mass spectrometry. *Eur. J. Biochem.* **239**, 144–149.
30. Odintsova, T. I., Muller, E. C., Ivanov, A. V., Egorov, T. A., Bienert, R., Vladimirov, S. N. *et al.* (2003). Characterization and analysis of posttranslational modifications of the human large cytoplasmic ribosomal subunit proteins by mass spectrometry and Edman sequencing. *J. Protein Chem.* **22**, 249–258.
31. Lecompte, O., Ripp, R., Thierry, J. C., Moras, D. & Poch, O. (2002). Comparative analysis of ribosomal proteins in complete genomes: an example of reductive evolution at the domain scale. *Nucleic Acids Res.* **30**, 5382–5390.
32. Kimura, M., Arndt, E., Hatakeyama, T. & Kimura, J. (1989). Ribosomal proteins in Halobacteria. *Can. J. Microbiol.* **35**, 195–199.

33. Auer, J., Lechner, K. & Bock, A. (1989). Gene organization and structure of two transcriptional units from *Methanococcus* coding for ribosomal proteins and elongation factors. *Can. J. Microbiol.* **35**, 200–204.
34. Auer, J., Spicker, G. & Bock, A. (1989). Organization and structure of the *Methanococcus* transcriptional unit homologous to the *Escherichia coli* "spectinomycin operon." Implications for the evolutionary relationship of 70S and 80S ribosomes. *J. Mol. Biol.* **209**, 21–36.
35. Ramirez, C., Louie, K. A. & Matheson, A. T. (1991). A small basic ribosomal protein from the extreme thermophilic archaeobacterium *Sulfolobus solfataricus* that has no equivalent in *Escherichia coli*. *FEBS Lett.* **284**, 39–41.
36. Strom, A. R. & Visentin, L. P. (1973). Acidic ribosomal proteins from the extreme halophile, *Halobacterium cutirubrum*. The simultaneous separation, identification and molecular weight determination. *FEBS Lett.* **37**, 274–280.
37. Londei, P., Teichner, A., Cammarano, P., De Rosa, M. & Gambacorta, A. (1983). Particle weights and protein composition of the ribosomal subunits of the extremely thermoacidophilic archaeobacterium *Caldariella acidophila*. *Biochem. J.* **209**, 461–470.
38. Teixido, J., Altamura, S., Londei, P. & Amils, R. (1989). Structural and functional exchangeability of 5S RNA species from the eubacterium *E. coli* and the thermoacidophilic archaeobacterium *Sulfolobus solfataricus*. *Nucleic Acids Res.* **17**, 845–851.
39. Altamura, S., Caprini, E., Sanchez, M. E. & Londei, P. (1991). Early assembly proteins of the large ribosomal subunit of the thermophilic archaeobacterium *Sulfolobus*. Identification and binding to heterologous rRNA species. *J. Biol. Chem.* **266**, 6195–6200.
40. Pettersson, I., Hardy, S. J. S. & Liljas, A. (1976). The ribosomal protein L8 is a complex of L7/L12 and L10. *FEBS Lett.* **6**, 135–138.
41. Arnold, R. J. & Reilly, J. P. (2002). Analysis of methylation and acetylation in *E. coli* ribosomal proteins. *Methods Mol. Biol.* **194**, 205–210.
42. Gauci, S., Helbig, A. O., Slijper, M., Krijgsveld, J., Heck, A. J. & Mohammed, S. (2009). Lys-N and trypsin cover complementary parts of the phosphoproteome in a refined SCX-based approach. *Anal. Chem.* **81**, 4493–4501.
43. Yu, L. R., Zhu, Z., Chan, K. C., Issaq, H. J., Dimitrov, D. S. & Veenstra, T. D. (2007). Improved titanium dioxide enrichment of phosphopeptides from HeLa cells and high confident phosphopeptide identification by cross-validation of MS/MS and MS/MS/MS spectra. *J. Proteome Res.* **6**, 4150–4162.
44. Moser, K. & White, F. M. (2006). Phosphoproteomic analysis of rat liver by high capacity IMAC and LC-MS/MS. *J. Proteome Res.* **5**, 98–104.
45. Rodriguez-Gabriel, M. A., Remacha, M. & Ballesta, J. P. (1998). Phosphorylation of ribosomal protein P0 is not essential for ribosome function but can affect translation. *Biochemistry*, **37**, 16620–16626.
46. Zhang, J., Sprung, R., Pei, J., Tan, X., Kim, S., Zhu, H. et al. (2009). Lysine acetylation is a highly abundant and evolutionarily conserved modification in *Escherichia coli*. *Mol. Cell. Proteomics*, **8**, 215–225.
47. Omer, A. D., Ziesche, S., Decatur, W. A., Fournier, M. J. & Dennis, P. P. (2003). RNA-modifying machines in archaea. *Mol. Microbiol.* **48**, 617–629.
48. Ditzel, L., Lowe, J., Stock, D., Stetter, K. O., Huber, H., Huber, R. & Steinbacher, S. (1998). Crystal structure of the thermosome, the archaeal chaperonin and homolog of CCT. *Cell*, **93**, 125–138.
49. Ruggero, D., Ciammaruconi, A. & Londei, P. (1998). The chaperonin of the archaeon *Sulfolobus solfataricus* is an RNA-binding protein that participates in ribosomal RNA processing. *EMBO J.* **17**, 3471–3477.
50. Londei, P., Altamura, S., Cammarano, P. & Petrucci, L. (1986). Differential features of ribosomes and of poly (U)-programmed cell-free systems derived from sulfur-dependent archaeobacterial species. *Eur. J. Biochem.* **157**, 455–462.
51. Henderson, E., Oakes, M., Clark, M. W., Lake, J. A., Matheson, A. T. & Zillig, W. (1984). A new ribosome structure. *Science*, **225**, 510–512.
52. Ciammaruconi, A., Gorini, S. & Londei, P. (2008). A bifunctional archaeal protein that is a component of 30S ribosomal subunits and interacts with C/D box small RNAs. *Archaea*, **2**, 151–158.
53. Wang, J., Dasgupta, I. & Fox, G. E. (2009). Many nonuniversal archaeal ribosomal proteins are found in conserved gene clusters. *Archaea*, **2**, 241–251.
54. Rodnina, M. V., Serebryanik, A. I., Ovcharenko, G. V. & Elskaya, V. (1994). ATPase strongly bound to higher eukaryotic ribosomes. *Eur. J. Biochem.* **225**, 305–310.
55. Ogata, K., Ohno, R., Terao, K., Iwasaki, K. & Endo, Y. (2000). Some properties and the possible role of intrinsic ATPase of rat liver 80S ribosomes in peptide bond elongation. *J. Biochem.* **127**, 221–231.
56. Hamel, E., Koka, M. & Nakamoto, T. (1972). Requirement of an *E. coli* 50S ribosomal protein component for effective interaction of the ribosome with T and G factors and with guanosine triphosphate. *J. Biol. Chem.* **247**, 805–814.
57. Uchiumi, T., Honma, S., Endo, Y. & Hachimori, A. (2002). Ribosomal proteins at the stalk region modulate functional rRNA structures in the GTPase center. *J. Biol. Chem.* **277**, 41401–41409.
58. Nierhaus, K. H. (1991). The assembly of prokaryotic ribosomes. *Biochimie*, **73**, 739–755.
59. Londei, P., Teixido, J., Acca, M., Cammarano, P. & Amils, R. (1986). Total reconstitution of active large ribosomal subunits of the thermoacidophilic archaeobacterium *Sulfolobus solfataricus*. *Nucleic Acids Res.* **14**, 2269–2285.
60. Sanchez, M. E., Urena, D., Amils, R. & Londei, P. (1990). *In vitro* reassembly of active large ribosomal subunits of the halophilic archaeobacterium *Haloferax mediterranei*. *Biochemistry*, **29**, 9256–9261.
61. El-Baradi, T. T., Raue, H. A., De Regt, C. H. & Planta, R. J. (1984). Stepwise dissociation of yeast 60S ribosomal subunits by LiCl and identification of L25 as a primary 26S rRNA binding protein. *Eur. J. Biochem.* **144**, 393–400.
62. Kruiswijk, T., Planta, R. J. & Krop, J. M. (1978). The course of the assembly of ribosomal subunits in yeast. *Biochim. Biophys. Acta*, **517**, 378–389.
63. Ban, N., Nissen, P., Hansen, J., Moore, P. B. & Steitz, T. A. (2000). The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science*, **289**, 905–920.
64. Halic, M., Becker, T., Frank, J., Spahn, C. M. & Beckmann, R. (2005). Localization and dynamic

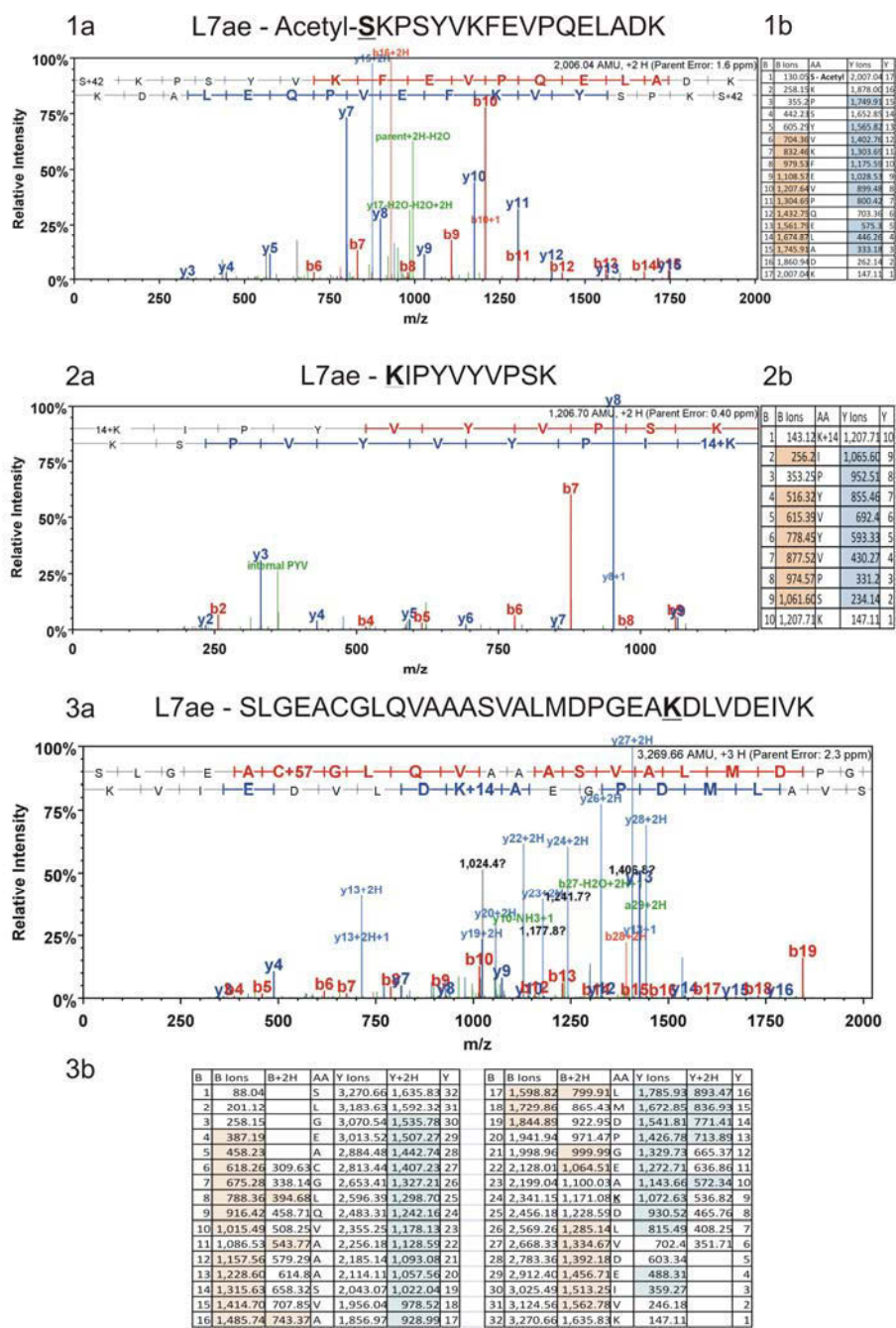
- behavior of ribosomal protein L30e. *Nat. Struct. Mol. Biol.* **12**, 467–468.
65. Allen, M. B. (1959). Studies with *Cyanidium caldarium*, an anomalously pigmented chlorophyte. *Arch. Mikrobiol.* **32**, 270–277.
  66. Brock, T. D., Brock, K. M., Belly, R. T. & Weiss, R. L. (1972). *Sulfolobus*: a new genus of sulfur-oxidizing bacteria living at low pH and high temperature. *Arch. Mikrobiol.* **84**, 54–68.
  67. Huber, G., Spinnler, C., Gambacorta, A. & Stetter, K. O. (1989). *Metallosphaera sedula* gen. and sp. nov. represents a new genus of aerobic, metal-mobilizing, thermoacidophilic archaeobacteria. *Syst. Appl. Microbiol.* **12**, 38–47.
  68. Volkl, P., Huber, R., Drobner, E., Rachel, R., Burggraf, S., Trincone, A. & Stetter, K. O. (1993). *Pyrobaculum aerophilum* sp. nov., a novel nitrate-reducing hyperthermophilic archaeum. *Appl. Environ. Microbiol.* **59**, 2918–2926.
  69. Bommer, U., Burkhardt, N., Jünemann, R., Spahn, C. M. T., Triana-Alonso, F. J. & Nierhaus, K. H. (1996). Ribosomes and polysomes. In (Graham, J. & Rickwoods, D., eds), pp. 271–301, IRL Press at Oxford University Press, Oxford, UK.
  70. Blaha, G., Stelzl, U., Spahn, C. M. T., Agrawal, R. K., Frank, J. & Nierhaus, K. H. (2000). Preparation of functional ribosomal complexes and the effect of buffer conditions on tRNA positions observed by cryoelectron microscopy. *Methods Enzymol.* **317**, 292–309.
  71. Nierhaus, K. H. & Dohme, F. (1974). Total reconstitution of functionally active 50S ribosomal subunits from *E. coli*. *Proc. Natl Acad. Sci. USA*, **71**, 4713–4717.
  72. Homann, H. E. & Nierhaus, K. H. (1971). Ribosomal proteins. Protein compositions of biosynthetic precursors and artificial subparticles from ribosomal subunits in *Escherichia coli* K 12. *Eur. J. Biochem.* **20**, 249–257.
  73. Starosta, A. L., Qin, H., Mikolajka, A., Leung, G. Y., Schwinghammer, K., Nicolaou, K. C. *et al.* (2009). Identification of distinct thiopeptide–antibiotic precursor lead compounds using translation machinery assays. *Chem. Biol.* **16**, 1087–1096.
  74. Wagenknecht, T., Frank, J., Boublik, M., Nurse, K. & Ofengand, J. (1988). Direct localization of the tRNA–anticodon interaction site on the *Escherichia coli* 30S ribosomal subunit by electron microscopy and computerized image averaging. *J. Mol. Biol.* **203**, 753–760.
  75. Mindell, J. A. & Grigorieff, N. (2003). Accurate determination of local defocus and specimen tilt in electron microscopy. *J. Struct. Biol.* **142**, 334–347.
  76. Frank, J., Radermacher, M., Penczek, P., Zhu, J., Li, Y., Ladjadj, M. & Leith, A. (1996). SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J. Struct. Biol.* **116**, 190–199.
  77. Armache, J.-P., Jarasch, A., Anger, A. M., Villa, E., Becker, T., Bhushan, S. *et al.* (2010). Cryo-EM structure of a translating eukaryotic 80S ribosome at 5.5 Å resolution. *Proc. Natl Acad. Sci. USA*, **107**, 19748–19753.
  78. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. & Ferrin, T. E. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612.



**Figure S1: Mass spectrometric identification of L12p modifications**

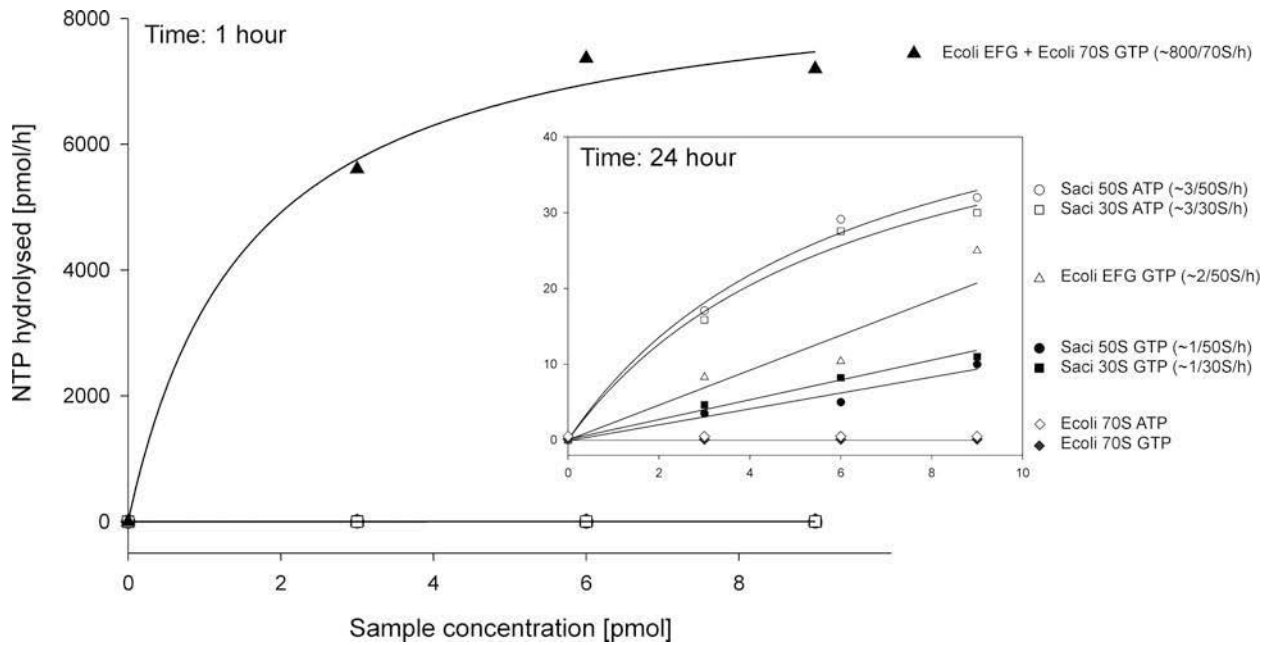
MS/MS spectra (panels a) and fragmentation tables (panels b) used for the identification of N-terminal acetylation (panel 1a and 1b), mono-methylation site K39 (2a and 2b) and mono-methylation site K89 of ribosomal protein L12p. Mono-methylated lysines within the peptide sequences are underlined and marked as bold. Detected b ions are highlighted in red and y ions are highlighted in blue.





**Figure S2: Mass spectrometric identification of L7ae modifications**

MS/MS spectra (panels a) and fragmentation tables (panels b) used for the identification of N-terminal acetylation (panel 1a and 1b), mono-methylation site K74 (2a and 2b) and mono-methylation site K108 of ribosomal protein L7ae. Mono-methylated lysines within the peptide sequences are underlined and marked as bold. Detected b ions are highlighted in red and y ions are highlighted in blue.



**Figure S3: GTPase and ATPase activities of archaeal ribosomal subunits**

The amount of ATP or GTP hydrolyzed per hour (at 37 °C) by *S. acidocaldarius* 30S (squares) or 50S subunits (circles) or *E. coli* EF-G, alone (open triangle) or in the presence of *E. coli* 70S ribosomes (closed triangle) is presented as function of the ribosome/factor concentration. No ATP or GTP activity was seen for *E. coli* 70S ribosomes or *S. acidocaldarius* 30S or 50S subunits within 1 hour, whereas rapid GTP hydrolysis was observed with *E. coli* 70S ribosomes in the presence of *E. coli* EF-G. Similarly, no ATPase or GTPase activities were detected at 60 °C for *S. acidocaldarius* 30S and 50S subunits (data not shown). However, after 24 hours (see inset), low GTPase and ATPase activities were detected for *S. acidocaldarius* 30S and 50S subunits, at levels similar to that observed for the intrinsic GTPase activity of *E. coli* EF-G. After 24 hours, no ATPase/GTPase activity was detected for *E. coli* 70S ribosomes (diamonds).

DOI: 10.1002/cmdc.201300323

# Amythiamicin D and Related Thiopeptides as Inhibitors of the Bacterial Elongation Factor EF-Tu: Modification of the Amino Acid at Carbon Atom C2 of Ring C Dramatically Influences Activity

Stefan Gross,<sup>[a]</sup> Fabian Nguyen,<sup>[b]</sup> Matthias Bierschenk,<sup>[c]</sup> Daniel Sohmen,<sup>[b]</sup> Thomas Menzel,<sup>[d]</sup> Iris Antes,<sup>\*[c]</sup> Daniel N. Wilson,<sup>\*[b]</sup> and Thorsten Bach<sup>\*[a]</sup>

Three analogues of amythiamicin D, which differ in the substitution pattern at the methine group adjacent to C2 of the thiazole ring C, were prepared by de novo total synthesis. In amythiamicin D, this carbon atom is (*S*)-isopropyl substituted. Two of the new analogues carry a hydroxymethyl in place of the isopropyl group, one at an *S*- (compound **3a**) and the other at an *R*-configured stereogenic center (**3b**). The third analogue, **3c**, contains a benzyloxymethyl group at an *S*-configured stereogenic center. Compounds **3b** and **3c** showed no inhibitory effect toward various bacterial strains, nor did they influence the translation of firefly luciferase. In stark contrast, compound **3a** inhibited the growth of Gram-positive bacteria *Staphylococ-*

*cus aureus* (strains NCTC and Mu50) and *Listeria monocytogenes* EGD. In the firefly luciferase assay it proved more potent than amythiamicin D, and rescue experiments provided evidence that translation inhibition is due to binding to the bacterial elongation factor Tu (EF-Tu). The results were rationalized by structural investigations and by molecular dynamics simulations of the free compounds in solution and bound to the EF-Tu binding site. The low affinity of compound **3b** was attributed to the absence of a critical hydrogen bond, which stabilizes the conformation required for binding to EF-Tu. Compound **3c** was shown not to comply with the binding properties of the binding site.

## Introduction

The bacterial elongation factor Tu (EF-Tu) was first described in 1966<sup>[1]</sup> and plays a crucial role in bacterial protein biosynthesis.<sup>[2]</sup> EF-Tu is a guanosine triphosphatase (GTPase) and displays high binding affinity for aminoacyl transfer RNA (aa-tRNA) in the GTP-bound form.<sup>[3]</sup> During translation, EF-Tu is responsible for the delivery of aa-tRNA to the ribosome in the form of a ternary complex with GTP. Upon delivering the correct aa-tRNA to the ribosome—that is, specific for the mRNA codon—GTP hy-

drolisis occurs, allowing dissociation of EF-Tu from the ribosome and accommodation of the aa-tRNA. Peptide bond formation then occurs, incorporating the newly delivered amino acid into the growing peptide chain.<sup>[4]</sup>

EF-Tu has been established as a validated drug target; it is a ubiquitous enzyme essential for bacterial protein biosynthesis.<sup>[5]</sup> EF-Tu differs significantly from the human elongation factor eEF-1, guaranteeing desirable target specificity. Currently, four structurally distinct compound classes are known to inhibit EF-Tu efficiently, for which prototypical examples are kirromycin, enacycloxin Ila, pulvomycin, and GE2270 A (**1**). It has been shown that the binding sites of pulvomycin and GE2270 A are similar,<sup>[6]</sup> whereas kirromycin<sup>[7]</sup> and enacycloxin Ila<sup>[8]</sup> possess a different mode of action. GE2270 A and pulvomycin hinder the formation of the ternary complex between EF-Tu, GTP, and aa-tRNA by binding to domain D2 of the enzyme. Binding of GE2270 A and related thiopeptides does not influence the GTPase activity of EF-Tu,<sup>[9]</sup> in accordance with the fact that binding of GE2270 A to EF-Tu does not significantly affect the catalytically active domain D1 of the protein.<sup>[6]</sup> Structural information about the interaction between GE2270 A and EF-Tu is based on crystallographic data obtained for EF-Tu-GDPNP-GE2270 A (GDPNP = guanosine-5'-( $\beta$ , $\gamma$ -imino)-triphosphate)<sup>[6b]</sup> and EF-Tu-GDP-GE2270 A (GDP = guanosine diphosphate).<sup>[6a]</sup> Binding occurs essentially along the cyclic thiopeptide, whereas peripheral substituents are less important for binding. All known naturally occurring GE2270 analogues,<sup>[10]</sup>

[a] Dr. S. Gross, Prof. Dr. T. Bach

Lehrstuhl für Organische Chemie I, Technische Universität München  
Lichtenbergstr. 4, 85747 Garching (Germany)  
E-mail: thorsten.bach@ch.tum.de

[b] F. Nguyen, D. Sohmen, Dr. D. N. Wilson

Gene Center, Department for Biochemistry, and  
Center for integrated Protein Science Munich (CiPSM)  
Ludwig Maximilian University of Munich  
Feodor-Lynen-Str. 25, 81377 Munich (Germany)  
E-mail: wilson@genzentrum.lmu.de

[c] M. Bierschenk, Prof. Dr. I. Antes

Protein Modelling, Department for Life Sciences, and  
Center for integrated Protein Science Munich (CiPSM)  
Technische Universität München  
Emil-Erlenmeyer-Forum 8, 85354 Freising (Germany)  
E-mail: antes@wzw.tum.de

[d] Dr. T. Menzel

Lehrstuhl für Organische Chemie II, Technische Universität München  
Lichtenbergstr. 4, 85747 Garching (Germany)

Supporting information for this article is available on the WWW under  
<http://dx.doi.org/10.1002/cmdc.201300323>.

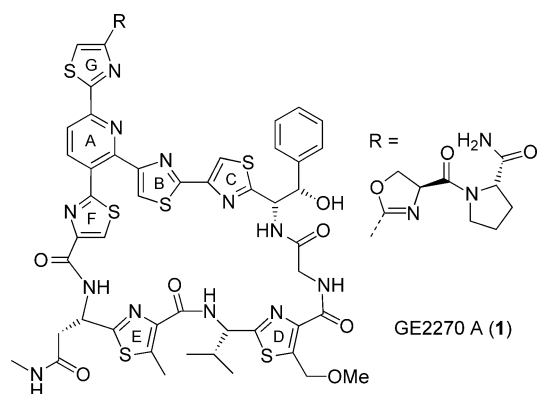


Figure 1. Ring numbering in the thiazolyl peptide GE2270 A (1).

which vary in the substituents at rings D and G or at the asparagine-derived amide in the southwestern part of the molecule (Figure 1), show an inhibitory effect on EF-Tu.<sup>[9–11]</sup> Because hydrolysis of the oxazoline (R at ring G) to a carboxylic acid is facile, several modification studies were devoted to derivatization at this site. In an early study the acid and reduced derivatives of it were converted by conventional nucleophilic displacement reactions into a plethora of compounds.<sup>[12]</sup> It was shown that these modifications can lead to a higher aqueous solubility while retaining biological activity. More recently, the same thiazole carboxylic acid was degraded by a Curtius rearrangement to deliver the respective 4-aminothiazole.<sup>[13]</sup> Derivatives of this compound showed promising activity toward several multi-resistant bacterial strains.<sup>[14]</sup> Optimization studies led to the discovery of two compounds with excellent physicochemical properties, including high water solubility.<sup>[15]</sup> Further development resulted in an investigational drug (LFF571), which shows superior activity against *Clostridium difficile*, a bacterium responsible for severe infections of the large intestine.<sup>[16]</sup>

Chemical modifications were also performed at the phenylglycine-derived northeastern part of GE2270 A located at position C2 of thiazole ring C.<sup>[17]</sup> Removal of the  $\alpha$ -hydroxybenzyl group led to a complete loss of activity, whereas some activity was retained in other derivatives in which the phenyl group was still present. It was thus concluded that this part is inserted into a lipophilic cavity of EF-Tu. Crystallographic data confirmed this view regarding the position of the phenyl group,<sup>[6]</sup> but also indicated a hydrogen bond between the hydroxy group of the phenylglycine to amino acid Glu226 of EF-Tu. The proximity of this binding site to domain D1 of EF-Tu and in particular to an  $\alpha$  helix (His85–Ala97), the C-terminal end of which is located in direct proximity to the catalytically active site, makes this position attractive for further studies.

The research presented herein addresses the latter issue by studying non-natural thiopeptides<sup>[18]</sup> for the first time, which were synthesized *de novo* and modified at this critical site. Compounds **3** (Figure 2) are analogues of amythiamicin D (**2**). In the amythiamicins,<sup>[19]</sup> which are closely related to the GE factors, the residue at C2 of thiazole ring C is derived from (*S*)-valine. Hydrogen bond formation with Glu226 of EF-Tu is

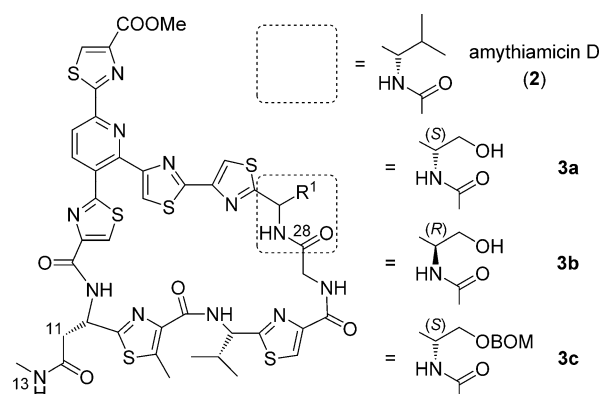


Figure 2. Molecules **2** and **3 a–c**, of which the antibiotic activity and mode of action on EF-Tu was investigated in this study.

therefore unfeasible. However, there is evidence that the amythiamicins also inhibit EF-Tu,<sup>[11,20]</sup> although they have been studied less thoroughly than GE2270 A. Compound **3 a** bears a hydroxymethyl group at an *S*-configured stereogenic center, mimicking the  $\alpha$ -hydroxybenzyl group present in GE2270 A. Compound **3 b** is an epimer of compound **3 a**, exhibiting the non-natural *R* configuration at the stereogenic center. Compound **3 c** has a lipophilic benzyloxymethyl (BOM) group instead of the polar hydroxy group.

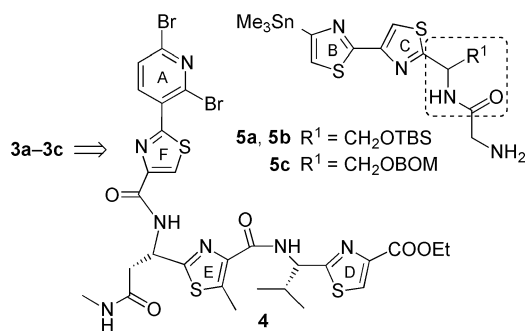
## Results and Discussion

### *De novo* synthesis of thiopeptides **3 a–3 c**

The synthetic strategy<sup>[21]</sup> toward compounds **3** was based on earlier work that had culminated in the total synthesis of GE2270 A<sup>[22]</sup> and amythiamicins C and D.<sup>[23]</sup> Key to this strategy was the use of 2,6-dibromo-3-iodopyridine as pivotal building block, to which the various thiazolyl fragments were coupled in successive order.

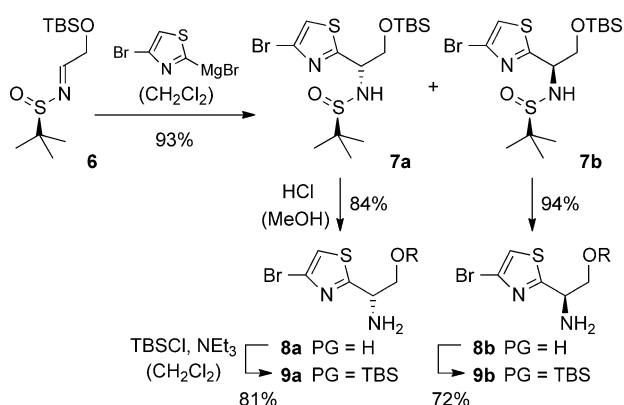
In this context, the synthesis of the southern fragment **4** of compounds **3** was reported earlier,<sup>[23]</sup> and its synthesis is not discussed further. It is available from (*S*)-valine in seven steps and with an overall yield of 22%. While the southern trithiazolyl part (rings D–F) and the northern thiazolyl fragment (ring G) of compounds **2** and **3** are identical, synthetic access to the eastern part (rings B and C) had to be individually secured by preparation of compounds **5** (Scheme 1, TBS = *tert*-butyldimethylsilyl).

The Grignard addition to chiral sulfinyl imines has been established by Ellman et al. as a useful method to generate chiral amines,<sup>[24]</sup> and it was previously shown that 4-bromthiazolyl-2-magnesium bromide adds efficiently to various imines derived from enantiomerically pure *tert*-butylsulfonamide.<sup>[23]</sup> The auxiliary can be cleaved after addition by acidic methanolysis. A putative starting material for the desired target compound was consequently an appropriately protected glycolaldehyde. Upon monosilylation of glycol, the resulting primary alcohol was oxidized to the respective aldehyde by Swern oxidation, which in turn was immediately converted into the known<sup>[25]</sup> imine **6** by



**Scheme 1.** Retrosynthetic disconnection of compounds **3 a–3 c** leading to different eastern fragments **5 a–5 c**, which were to be individually synthesized. See Figure 2 for the configuration in the dashed (----) box.

treatment with commercially available (*R*)-*tert*-butylsulfonamide (see Supporting Information for details). Grignard addition of the above-mentioned thiazolyl magnesium bromide (prepared by bromine magnesium exchange of 2,4-dibromothiazole with isopropyl magnesium bromide in THF/Et<sub>2</sub>O<sup>[26]</sup>) proceeded smoothly and delivered the desired products **7** as a mixture of diastereomers (Scheme 2). Gratifyingly, their separation by

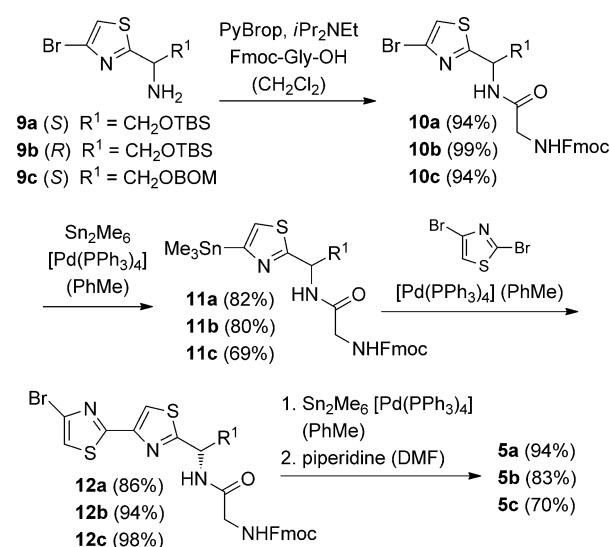


**Scheme 2.** Preparation of enantiomeric amino alcohols **8 a** and **8 b** via the respective diastereomeric sulfonil amines **7 a** and **7 b** and their conversion into the O-protected amino alcohols **9 a** and **9 b**.

chromatography was facile, and products **7 a** and **7 b** were employed as individual diastereomers in the next step. Given that both enantiomeric amino alcohols **8 a** and **8 b** were required for our further investigations, the Grignard addition was not thoroughly optimized. When performed at  $-78^{\circ}\text{C}$  in a 6:1 solvent mixture of CH<sub>2</sub>Cl<sub>2</sub> and ethers (THF/Et<sub>2</sub>O), there was no notable diastereoselectivity. As previously observed,<sup>[23]</sup> a larger fraction of CH<sub>2</sub>Cl<sub>2</sub> was required to achieve diastereoselectivity. In the present case, the diastereomeric ratio (**7 a**/**7 b**) improved to 73:27 in a 14:1 mixture of CH<sub>2</sub>Cl<sub>2</sub>/ether. The configuration at the newly formed stereogenic center was determined by the Mosher method<sup>[27]</sup> upon removal of the chiral auxiliary (Supporting Information). The major enantiomer of the diastereoselective addition was shown to be *S*-configured. Based on the conventional model for the Grignard addition to chiral sulfonil imines,<sup>[24]</sup> the result appears surprising, but the unusual behav-

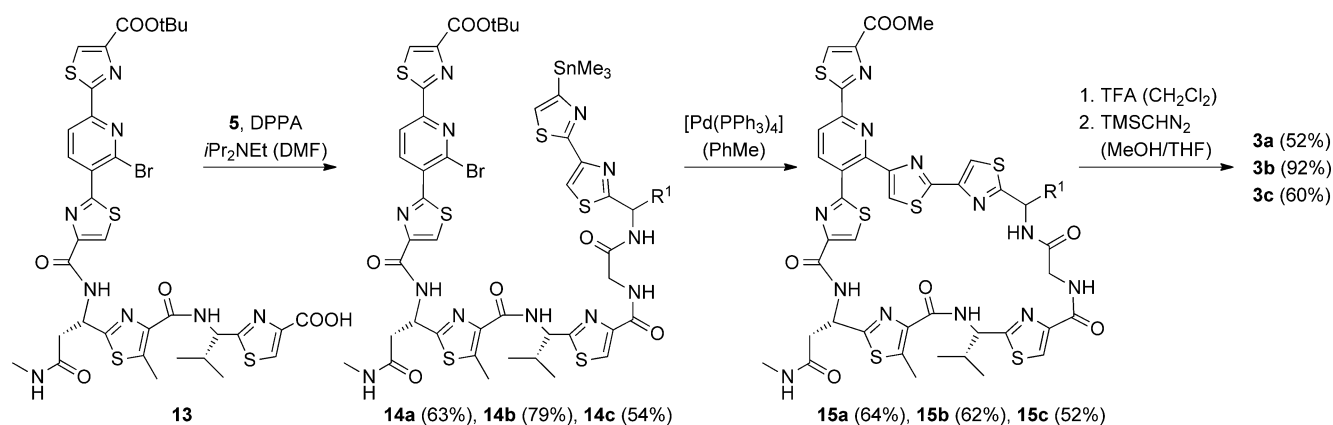
ior of imine **6** had been previously discussed by Barrow et al.<sup>[25]</sup>

Because removal of the auxiliary also led to silyl deprotection, the silyl group was installed again by treatment of alcohols **8** with TBSCl/NEt<sub>3</sub>. Introduction of the BOM group was facile at this stage and delivered the required amine **9 c** from amino alcohol **8 a** (Supporting Information). Peptide coupling of amines **9** to 9-fluorenylmethoxycarbonyl (Fmoc)-protected glycine (Fmoc-Gly-OH) was achieved by treatment of the coupling partners with bromotri(pyrrolidino)phosphonium hexafluorophosphate (PyBrop).<sup>[28]</sup> To convert bromides **10** into suitable nucleophiles for a regioselective cross-coupling reaction with the second thiazole fragment,<sup>[29]</sup> a stannylation was performed with hexamethylditin at 100 °C in toluene using tetrakis(triphenylphosphine)palladium as the catalyst.<sup>[30]</sup> Stannanes **11** underwent a smooth Stille cross-coupling<sup>[31]</sup> with 2,4-dibromothiazole,<sup>[32]</sup> which was used in slight excess (1.4 equiv), yielding the desired dithiazoles **12** in high yields (Scheme 3).



**Scheme 3.** Synthesis of stannylated compounds **5 a–5 c** from the respective bromothiazoles **9 a–9 c**, employing regioselective Stille cross-coupling reactions on 2,4-dibromothiazole as the key step.

4-Bromodithiazoles **12** were converted into the corresponding stannanes by another palladium-catalyzed stannyl debromination. Removal of the Fmoc protecting group with piperidine delivered the free amines **5** in high yields. The order of events for the incorporation of building blocks **5** into a pyridine fragment and the choice of this fragment was based on previous experience (Scheme 4). In the synthesis of GE2270 A,<sup>[22]</sup> it had been found that a macrocyclization by Stille cross-coupling after preceding amide bond formation provides significantly higher yields than an initial Stille cross-coupling followed by macrolactamization. In the synthesis of amythiamicin C,<sup>[23]</sup> it was found that amide bond formation with the acid derived from ester **4** is feasible, but the subsequent ring closure, which requires a regioselective Stille cross-coupling at C2 of the pyridine core, is not sufficiently selective. Therefore,



**Scheme 4.** Peptide bond formation between pyridine core fragment **13** and the individual building blocks **5a–5c** (cf. Schemes 1–3) followed by a macrocyclization via Stille cross-coupling and two functional group transformation steps toward products **3a–3c**. The hydrolysis of the *tert*-butyl ester **15b** with TFA was followed by addition of HF-py to complete the desilylation.

ester **4** was converted according to known methodology<sup>[23]</sup> into the previously described pyridine fragment **13**. Coupling of the two fragments to amides **14** was achieved with diphenylphosphoryl azide (DPPA)<sup>[33]</sup> in the presence of Hünig's base. We were pleased to note that the Stille cross-coupling protocol was also successful for substrates **14** and delivered macrolactams **15** in moderate to good yields. Cleavage of the *tert*-butyl ester in ring G was possible with trifluoroacetic acid (TFA), which also led to cleavage of the TBS ether. Because TBS deprotection was relatively slow for substrate **15a** leading to side reactions and a diminished yield, deprotection of substrate **15b** was performed by an initial treatment with TFA and subsequent addition of HF-pyridine. Yields were significantly higher with this modification than the TFA deprotection used for **15a**. Conversion of the free acids into the desired methyl esters **3** was performed with trimethylsilyl diazomethane.

### Biological activity and translation assays

To assess the biological activities of amythiamicin D (**2**) and the synthetic derivatives **3a–3c**, the minimal inhibitory concentration (MIC) was determined for the Gram-positive bacteria *Staphylococcus aureus* (strains NCTC and Mu50), *Listeria monocytogenes* EGD and *Streptococcus pyogenes* ATCC 10231, as well as the Gram-negative bacterium *Pseudomonas aeruginosa* PA01 and compared with control antibiotics kirromycin and the aminoglycoside kanamycin (Table 1).<sup>[34]</sup> Amythiamicin D was previously shown to inhibit the growth of many Gram-positive bacteria, but not Gram-negative bacteria such as *E. coli*.<sup>[19]</sup> We consistently found that compound **2** displays excellent biological activity against *S. aureus* and *L. monocytogenes* with an MIC<sub>50</sub> value of 0.32 μM, yet displays no activity against *S. pyogenes* or *P. aeruginosa*. A similar activity profile was observed for the synthetic derivative **3a**, although the MIC values were somewhat higher for the *S. aureus* and *L. monocytogenes* strains, when compared with **2**. In contrast, compounds **3b** and **3c**

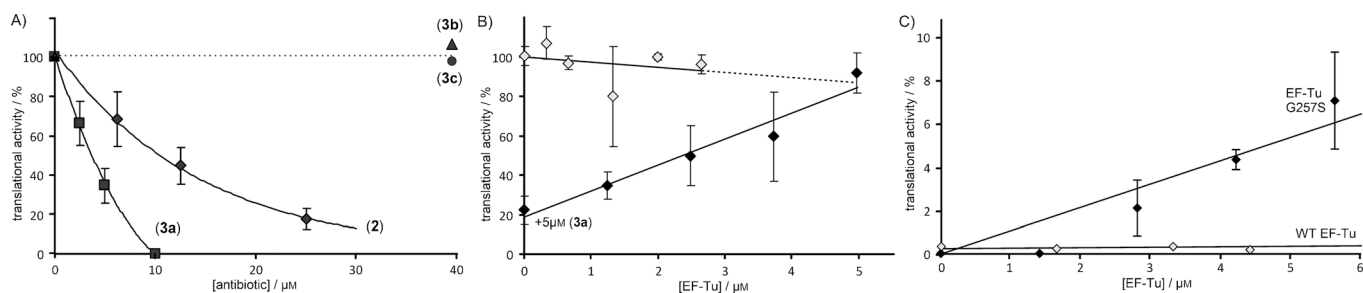
**Table 1.** Minimal inhibitory concentrations of synthetic derivatives **3a–3c** relative to amythiamicin D (**2**), kirromycin, and kanamycin.

Strain	MIC [μM]					
	kanamycin	kirromycin	<b>2</b>	<b>3a</b>	<b>3b</b>	<b>3c</b>
<i>S. aureus</i> NCTC 8325	2.56	> 100	0.32	0.64	> 100	> 100
<i>S. aureus</i> Mu50	> 100	> 100	0.32	1.28	> 100	> 100
<i>L. monocytogenes</i> EGD-e	12.8	2.56	0.32	5.12	> 100	> 100
<i>S. pyogenes</i> ATCC 10231	100	0.32	> 100	> 100	> 100	> 100
<i>P. aeruginosa</i> PA01	> 100	> 100	> 100	> 100	> 100	> 100

displayed no activity against any of the strains tested, even at concentrations up to 100 μM (Table 1).

Although amythiamicins display poor activity against Gram-negative organisms such as *P. aeruginosa* (Table 1) and *E. coli*,<sup>[19]</sup> inhibitory effects of amythiamicins on translation of poly(U)-directed poly(Phe) synthesis using an *E. coli* in vitro translation system has been demonstrated.<sup>[20]</sup> This indicates that the inactivity of amythiamicins in blocking bacterial growth can be related to other factors such as membrane penetration rather than a lack of effect on translation. Therefore, to directly assess the effect of amythiamicin derivatives **3a–3c** on protein synthesis, we monitored the translation of firefly luciferase (Fluc) in an *E. coli* cell-free in vitro translation system in the presence of increasing concentrations of amythiamicin D (**2**) or one of the synthetic derivatives **3a–3c** (Figure 3A). As expected,<sup>[20]</sup> **2** is a potent inhibitor of protein synthesis with a half-inhibitory concentration (IC<sub>50</sub> value) of ~10 μM. Surprisingly, however, the synthetic derivative **3a** displayed improved inhibitory activity, with IC<sub>50</sub> ~4 μM, at least twofold better than amythiamicin D. In contrast, derivatives **3b** and **3c** were completely inactive, even at high drug concentrations up to 40 μM. This suggests that the lack of effect of thiopeptides **3b** and **3c** on the growth of *S. aureus* and *L. monocytogenes* strains (Table 1) may indeed be due to an inability to inhibit protein synthesis, rather than an inability to penetrate the cell wall.

The structural similarity between amythiamicins and GE2270 A (**1**) suggests that amythiamicins also inhibit transla-



**Figure 3.** Translation inhibition by amythiamicin D and synthetic derivatives. A) Effect of increasing concentrations of amythiamicin D (**2**,  $\blacklozenge$ ) and the amythiamicin D derivative **3a** ( $\blacksquare$ ), **3b** ( $\blacktriangle$ ), and **3c** ( $\bullet$ ) on the translation of firefly luciferase. B) Effect of increasing concentrations of EF-Tu in the presence of 5  $\mu\text{M}$  amythiamicin D derivatives **3a** ( $\blacklozenge$ ) and absence of drug ( $\diamond$ ). C) Effect of increasing concentrations of wild-type EF-Tu ( $\diamond$ ) or EF-Tu-G257S ( $\blacklozenge$ ) in the presence of 30  $\mu\text{M}$  amythiamicin D derivatives **3a** ( $\blacklozenge$ ). The firefly luminescence in the absence of drug was assigned as 100%, and error bars display the standard deviation from the mean for three individual experiments.

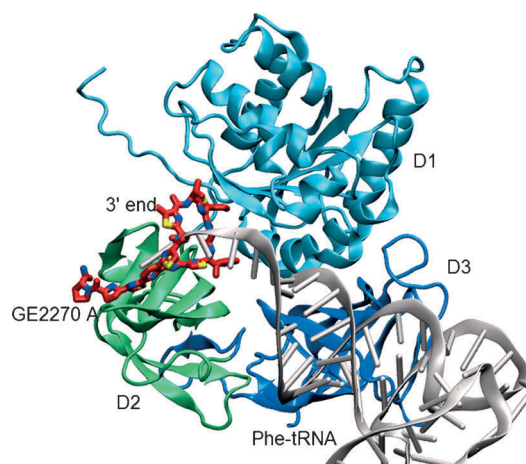
tion by interaction with EF-Tu, consistent with the observation that resistance to amythiamicins results from mutation with *tufA*, the gene for EF-Tu.<sup>[20]</sup> We reasoned that if EF-Tu is the target of amythiamicin derivative **3a**, then increasing the concentration of EF-Tu within the in vitro translation assays should relieve the inhibition of translation resulting from amythiamicins. As is apparent in Figure 3B, the presence of **3a** at 5  $\mu\text{M}$  decreases translation to 20%, which can be restored to 100% by the addition of an extra 5  $\mu\text{M}$  recombinantly purified EF-Tu protein. As a control, the presence of additional EF-Tu was shown to have no significant effect on translation of Fluc in the absence of the drug.

The rescue of translation by the presence of additional EF-Tu is analogous to the target overexpression mechanisms that are used by some bacteria to obtain resistance to various antibiotics.<sup>[35]</sup> Consistently, when the same experiment was performed in the presence of higher saturating concentrations of **3a** (30  $\mu\text{M}$ ), no rescue was observed (Figure 3C).

Next, we rationalized that if amythiamicins do indeed target EF-Tu, then titrating in an EF-Tu mutant protein that is resistant to thiopeptides should rescue translation to some extent, even in the presence of saturating drug concentrations. It has been reported that mutation of glycine at position 257 to serine in *E. coli* EF-Tu confers resistance to the thiopeptide GE2270 A by allowing simultaneous binding of the antibiotic and aa-tRNA to EF-Tu.<sup>[36]</sup> We found that EF-Tu-G257S protein alone inhibits translation by  $\sim 60\%$  at 1  $\mu\text{M}$  (Supporting Information). Nevertheless, in the presence of 30  $\mu\text{M}$  **3a**, it was indeed possible to rescue translation with the EF-Tu-G257S mutant, but not with wild-type EF-Tu. These findings not only support that the amythiamicin D derivative **3a** binds to EF-Tu, but also suggests that it interacts with EF-Tu in a manner analogous to GE2270 A, as expected based on the similarity in their chemical structures.

### Structural investigations

The experimental structure of GE2270 A bound to EF-Tu-GDPNP<sup>[6b]</sup> was used to investigate the structural basis of the binding behavior of the synthesized compounds; GE2270 A binds competitively to the aa-tRNA binding site of



**Figure 4.** Structure of EF-Tu-GDPNP (blue/green ribbons) with bound GE2270 A (red sticks) (PDB ID: 2C77).<sup>[6b]</sup> The additionally shown Phe-tRNA strand (grey) was extracted from the corresponding superimposed EF-Tu-GDPNP:Phe-tRNA structure (PDB ID: 1TTT),<sup>[37]</sup> and superimposition was performed based on the EF-Tu backbone.

EF-Tu (Figure 4). The binding site of GE2270 A is located at the interface between domains D1 and D2, where the aa-tRNA binds with its 3' end. GE2270 A predominantly binds to D2 with the upper part of its thiazolyl ring contacting D1 in EF-Tu-GDPNP. Through binding of the antibiotic the interface geometry is widened, leading to a gap between the two domains in which the antibiotic compound is bound. Consequently, the binding site is blocked and thus binding of the 3' end of the aa-tRNA is inhibited by steric hindrance (Figure 4).

To elucidate the structural basis for the different binding behaviors of compounds **2** and **3a–3c** we first investigated their structures in aqueous solution by preliminary NMR experiments (Supporting Information) and by more extensive molecular dynamics (MD) and docking simulations. In previous work,<sup>[38]</sup> it was established by a detailed NMR study of amythiamicin D that a single hydrogen bond is present between the carbonyl group at C28 and the proton at N13 (Figure 2), significantly stabilizing the ring conformation. The same hydrogen bond is present in the crystal structure of the EF-Tu-GDPNP-GE2270 A complex (Figure 4),<sup>[6b]</sup> indicating that the stabilized

ring conformation in solution is retained upon binding to EF-Tu. As a result of the fixed conformation, the diastereotopic protons at the methylene group C11 of the side chain between rings E and F result in distinct peaks, which are clearly separated in the  $^1\text{H}$  NMR spectrum (see  $^1\text{H}$  NMR data in the Supporting Information). In stark contrast, a significant peak broadening was detected for the same protons in the  $^1\text{H}$  NMR spectrum of compound **3b**. Only one peak was recorded at room temperature, which was significantly broadened upon cooling (Supporting Information). Although this observation could be explained by other biophysical effects, it is most likely to indicate that the unusual *R*-stereogenic center at  $\text{R}^1$  in compound **3b** leads to significant changes in the ring conformation and to a higher degree of ring flexibility, averaging out the environment of the two protons. The higher flexibility, in turn, appears to be due to the lack of a hydrogen bond between the carbonyl group at C28 and the proton at N13 in **3b**.

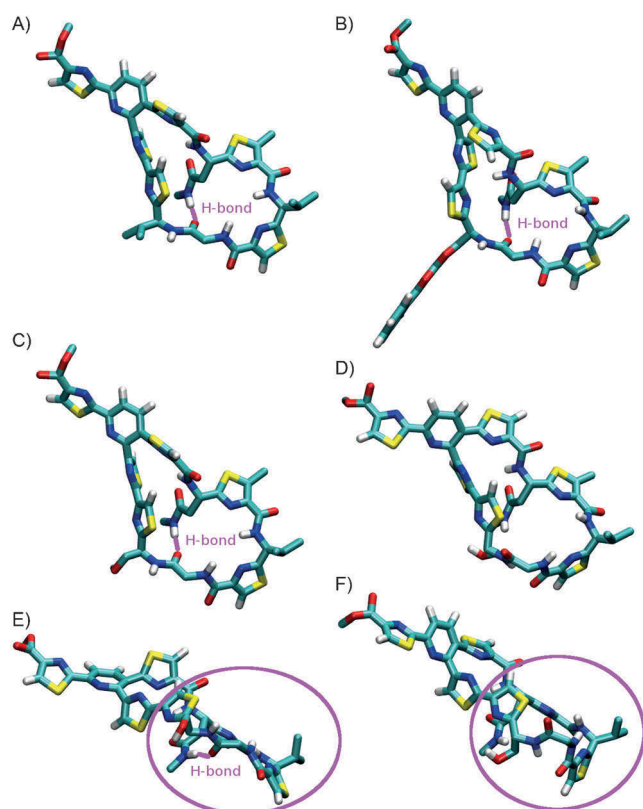
To further investigate this topic we performed MD simulations of the free compounds in solution and bound to the EF-Tu binding site. The solution results are shown in Figure 5. For compounds **2**, **3a**, and **3c** the same predominant ring conformation was observed (Figure 5A–C,E and Supporting Information (SI) figure SI4) for all three compounds. This conformation is very similar to the bound conformation of GE2270 A in EF-Tu (Figure 4) and is in agreement with the NMR results, as the

ring is stabilized in all three cases by the same hydrogen bond between the C28 carbonyl group and the proton at N13 as also observed by Lewis et al.<sup>[38]</sup> In compound **3b** the *R*-stereogenic center at C2 of the thiazole ring C leads to a geometrical inversion of its bonds to its neighboring ring atoms and its side chain. This causes a  $180^\circ$  rotation of the neighboring peptide group containing C28, and thus the ring stabilizing the hydrogen bond between N13 and the C28 carbonyl group is broken (Figure 5D, SI figure SI4d). Therefore, significant alterations in the overall ring conformation can be observed, together with a higher flexibility in its side chains. This leads to a different and less planar overall conformation than the ring conformation of compounds **2**, **3a**, and **3c** (see Figure 5E,F). Therefore compound **3b** has different steric and interaction requirements for successful protein binding. As the bound conformation of GE2270 A is close to the solution structures of compounds **2**, **3a**, and **3c** and also features the conserved hydrogen bond, it can be assumed that adopting this conformation in solution is a prerequisite for stable binding of the compounds to EF-Tu, which may explain the nonbinding behavior of compound **3b** (Figure 3).

To further substantiate this hypothesis, docking-based MD simulations were performed for all four compounds using the docking software DynaDock.<sup>[39]</sup> As the algorithm is MD based, it allows full flexibility of the whole system and a proper treatment of the changes in the protein binding site upon ligand binding. This is important in the case of EF-Tu, as the binding site is very flexible, and large movements can be observed upon ligand binding. For the docking simulations the compounds were placed within the binding site based on their solution conformation of the ring system. Conformational sampling of all side chains and rotation and translation of the whole compound was then performed. In a second step, MD simulations were performed for the best-docked conformations to investigate the stability of the predicted complexes.

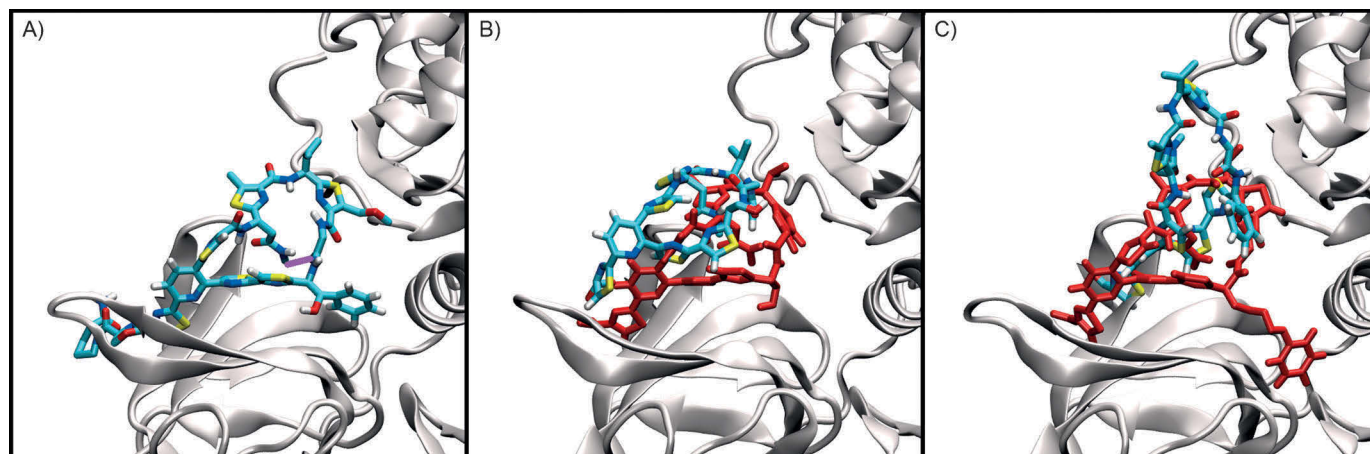
The final equilibrated structures after MD refinement are provided in Figure 6 and in the Supporting Information (figures SI1 and SI2). In the case of compounds **2** and **3a**, stable complexes with a bound ring conformation close to GE2270 A could be obtained, indicating stable, strong binding to EF-Tu (see SI figure SI1c,d). In both cases a strong hydrogen bond network was formed between the ring system and the neighboring residues Asn285 and Gln98, which is further stabilized by an additional hydrogen bond between His67 and Gln98. The ring-stabilizing intramolecular C28O–N13 hydrogen bond as observed for GE2270 A was retained (SI figure SI6). In addition, for the hydroxy group at  $\text{R}^1$  in compound **3a**, fluctuating hydrogen bonds were observed with Glu226, Thr239, and Gln98 (SI figure SI2a).

This hydrogen bond pattern is consistent with the pattern observed for the experimental structure of the EF-Tu-GDPNP-GE2270 A complex.<sup>[6b]</sup> In both cases the hydroxy groups of  $\text{R}^1$  (Figure 1) form additional hydrogen bonds, but the hydrogen bond pattern of the hydroxy group is more restricted in GE2270 A due to the steric lock position by the additional phenyl ring. The hydroxy group in compound **3a** is very flexible and is thus able to form alternative hydrogen bonds



**Figure 5.** Solution structures of compounds A) **2**, B) **3c**, C,E) **3a**, and D,F) **3b** after 1 ns MD simulation. In E) and F) compounds **3a** and **3b** are shown with the ring plane being rotated by  $90^\circ$  with respect to the perspective shown in parts C) and D).





**Figure 6.** Stable binding of A) GE2270 A (1) and unstable binding of compounds B) **3b** and C) **3c** after 1 ns MD simulations. The final ligand conformations of the simulations are shown in stick representation colored by atom type. The starting conformations of compounds **3b** and **3c** are provided in red for comparison. The ring-stabilizing hydrogen bond is shown in magenta.

throughout the surrounding lipophilic sub-pocket. Consequently, its position fluctuates between the corresponding hydrogen bond forming residues.

The nonbinding behavior of compound **3c** is more difficult to explain: During the initial placement of compound **3c** only one favorable binding mode was found in which the BOM side chain is located in a stretched conformation inside a side cavity of the binding site (Figure 6c, red ligand conformation). However, due to the large amount of charged and polar side chains alongside this cavity, the bound conformation is not stable during the following MD simulation, leading to an altered ring conformation and movement of the BOM side chain out of its cavity (Figure 6C, SI figure SI5). Therefore, it is unlikely that BOM moves into this cavity during the real binding process, explaining its nonbinding behavior.

In the case of compounds **3b** and **3c**, the MD simulations showed that the complexes are not stable, as the compounds slowly move out of the binding site (Figure 6B,C, SI figure SI5). In the case of compound **3b** this is due to inversion of the binding geometry at the *R*-stereogenic center and thus formation of a bulky ring conformation, which is not compatible with the rather flat  $\beta$ -sheet-based surface of the binding site and its hydrogen bonding pattern, thus leading to a movement of the ligand out of the binding site (Figure 6B, SI figures SI1b and SI5), consistent with the lack of EF-Tu binding observed experimentally (Figure 3).

## Conclusions

Based on an expedient synthetic access, three analogues (**3a–3c**) of amythiamicin D (**2**) have become available, which cannot be obtained by degradation studies or any other means. Translation studies revealed that compound **3a** shows enhanced potency in EF-Tu inhibition relative to the natural product. The activity of both amythiamicin D and compound **3a** could be nicely corroborated based on MD simulations: Compounds **3b** and **3c** were inactive in both antibacterial and in vitro translation assays. The nonbinding behavior of **3b** can

be explained by a ring conformation that is—due to the absence of a crucial hydrogen bond—different from those of the other thiopeptides. The conformational change is induced by the stereogenic center at the crucial methine carbon atom adjacent to C2 of thiazole ring C, which is distinct from the **2** and **3a** *R* configuration. The inactivity of **3c** is remarkable because it was envisioned that the lipophilic BOM group would be complementary to the hydrophobic pocket between domains D2 and D3. However, due to the size, electrostatic properties, and flexibility of the BOM group, it appears that it does not comply with the binding properties of the binding site.

The present study is the first to have used de novo synthesized non-natural thiopeptides. It provides unambiguous evidence that readily available synthetic analogues of the amythiamicins and the GE factors bind efficiently to the previously described binding site of EF-Tu. In this regard, compound **3a** can serve as a versatile scaffold to study the effect of various substitution patterns on the activity of EF-Tu and on conformational changes induced by non-natural ligands.

## Experimental Section

**Synthetic studies:** Experimental procedures and characterization data (including NMR spectra) for all new compounds are provided in the Supporting Information. Analytical data for the three amythiamicin derivatives **3a–3c** are given below.

**Thiopeptide 3a.**  $R_f=0.35$  ( $\text{CH}_2\text{Cl}_2/\text{MeOH}=10:1$ , UV);  $[\alpha]_D^{20}=+33.0 \text{ cm}^3 \text{ g}^{-1} \text{ dm}^{-1}$  ( $c=0.10$  in  $\text{CHCl}_3$ );  $^1\text{H NMR}$  ( $\text{CDCl}_3$ , 500 MHz):  $\delta=0.93$  (d,  $^3J=6.5$  Hz, 3H), 1.00 (d,  $^3J=6.7$  Hz, 3H), 2.19 (m, 1H), 2.40 (brs, 2H), 2.62 (brs, 3H), 2.68 (s, 3H), 4.01 (s, 3H), 4.12 (brs, 2H), 4.28 (d,  $^2J=17.1$  Hz, 1H), 4.63 (d,  $^2J=17.1$  Hz, 1H), 5.28–5.34 (m, 2H), 5.47 (s, 1H), 6.76 (brs, 1H), 7.34 (s, 1H), 7.92 (s, 1H), 7.99–8.10 (m, 2H), 8.06 (d,  $^3J=7.5$  Hz, 1H), 8.11 (s, 1H), 8.24–8.35 (m, 4H), 8.37 ppm (d,  $^3J=7.5$  Hz, 1H);  $^{13}\text{C NMR}$  ( $\text{CDCl}_3$ , 126 MHz):  $\delta=12.6$ , 18.5, 19.0, 26.3, 35.0, 38.3, 42.8, 49.1, 52.7, 54.7, 55.0, 64.2, 115.8, 118.9, 121.7, 123.8, 125.4, 128.2, 130.6, 139.7, 141.3, 142.2, 148.2, 148.6, 148.8, 150.0, 150.7, 154.0, 160.3, 160.7, 161.4, 161.7, 161.8, 165.1, 166.0, 168.7, 169.2, 169.8, 170.3, 170.4 ppm; IR (ATR):  $\tilde{\nu}=2956$  (w), 2924 (m), 2850 (w), 1654 (m), 1542 (m), 766 (w), 667  $\text{cm}^{-1}$

(w); MS (ESI):  $m/z$  (%) = 1041 (45), 1019 (100); HRMS (ESI):  $m/z$  =  $C_{41}H_{39}N_{12}O_8S_6$  [ $M+H$ ]<sup>+</sup>, calcd: 1019.1333, found: 1019.1332.

**Thiopeptide 3b.**  $R_f$  = 0.35 ( $CH_2Cl_2/MeOH$  = 10:1, UV);  $[\alpha]_D^{20}$  = +33.0  $cm^3 g^{-1} dm^{-1}$  ( $c$  = 0.10 in  $CHCl_3$ );  $^1H$  NMR ( $CDCl_3$ , 500 MHz):  $\delta$  = 0.93 (d,  $^3J$  = 6.5 Hz, 3H), 1.00 (d,  $^3J$  = 6.7 Hz, 3H), 2.19 (m, 1H), 2.40 (brs, 2H), 2.62 (brs, 3H), 2.68 (s, 3H), 4.01 (s, 3H), 4.12 (brs, 2H), 4.28 (d,  $^2J$  = 17.1 Hz, 1H), 4.63 (d,  $^2J$  = 17.1 Hz, 1H), 5.28–5.34 (m, 2H), 5.47 (s, 1H), 6.76 (brs, 1H), 7.34 (s, 1H), 7.92 (s, 1H), 7.99–8.10 (m, 2H), 8.06 (d,  $^3J$  = 7.5 Hz, 1H), 8.11 (s, 1H), 8.24–8.35 (m, 4H), 8.37 ppm (d,  $^3J$  = 7.5 Hz, 1H);  $^{13}C$  NMR ( $CDCl_3$ , 126 MHz):  $\delta$  = 12.6, 18.5, 19.0, 26.3, 35.0, 38.3, 42.8, 49.1, 52.7, 54.7, 55.0, 64.2, 115.8, 118.9, 121.7, 123.8, 125.4, 128.2, 130.6, 139.7, 141.3, 142.2, 148.2, 148.6, 148.8, 150.0, 150.7, 154.0, 160.3, 160.7, 161.4, 161.7, 161.8, 165.1, 166.0, 168.7, 169.2, 169.8, 170.3, 170.4 ppm; IR (ATR):  $\tilde{\nu}$  = 2956 (w), 2924 (m), 2850 (w), 1654 (m), 1542 (m), 766 (w), 667  $cm^{-1}$  (w); MS (ESI):  $m/z$  (%) = 1041 (45), 1019 (100); HRMS (ESI):  $m/z$  =  $C_{41}H_{39}N_{12}O_8S_6$  [ $M+H$ ]<sup>+</sup>, calcd: 1019.1333, found: 1019.1332.

**Thiopeptide 3c.**  $R_f$  = 0.20 ( $CH_2Cl_2/MeOH$  = 20:1, UV);  $[\alpha]_D^{20}$  = +137  $cm^3 g^{-1} dm^{-1}$  ( $c$  = 0.09 in  $CHCl_3$ );  $^1H$  NMR ( $CDCl_3$ , 500 MHz):  $\delta$  = 0.88 (d,  $^3J$  = 6.8 Hz, 3H), 0.96 (dd,  $^2J$  = 17.5 Hz,  $^3J$  = 5.9 Hz, 1H), 1.00 (d,  $^3J$  = 6.8 Hz, 3H), 2.28 (m, 1H), 2.63 (d,  $^3J$  = 4.8 Hz, 3H), 2.65 (s, 3H), 2.70 (dd,  $^2J$  = 17.5 Hz,  $^3J$  = 3.5 Hz, 1H), 3.59 (dd,  $^2J$  = 17.6 Hz,  $^3J$  = 3.3 Hz, 1H), 3.94 (dd,  $^2J$  = 10.7 Hz,  $^3J$  = 3.6 Hz, 1H), 4.01 (dd,  $^2J$  = 10.7 Hz,  $^3J$  = 4.0 Hz, 1H), 4.02 (s, 3H), 4.59 (d,  $^2J$  = 11.7 Hz, 1H), 4.67 (d,  $^2J$  = 11.7 Hz, 1H), 4.79 (dd,  $^2J$  = 17.6 Hz,  $^3J$  = 9.7 Hz, 1H), 4.84 (d,  $^2J$  = 6.8 Hz, 1H), 4.89 (d,  $^2J$  = 6.8 Hz, 1H), 5.23 (dd,  $^3J$  = 7.9, 4.7 Hz, 1H), 5.39 (m, 2H), 6.80 (d,  $^3J$  = 4.8 Hz, 1H), 6.91 (d,  $^3J$  = 6.0 Hz, 1H), 7.27 (s, 1H), 7.33–7.42 (m, 5H), 7.71 (dd,  $^3J$  = 9.7, 3.3 Hz, 1H), 8.11 (s, 1H), 8.13 (d,  $^3J$  = 8.1 Hz, 1H), 8.26 (s, 1H), 8.36 (s, 1H), 8.38 (s, 1H), 8.38 (d,  $^3J$  = 8.1 Hz, 1H), 8.76 (d,  $^3J$  = 7.9 Hz, 1H), 8.99 ppm (d,  $^3J$  = 9.2 Hz, 1H);  $^{13}C$  NMR ( $CDCl_3$ , 126 MHz):  $\delta$  = 12.3, 18.0, 18.4, 26.2, 34.7, 38.3, 41.1, 48.2, 52.6, 53.3, 56.1, 69.9, 70.8, 95.9, 115.0, 118.7, 123.2, 123.8, 125.2, 127.7, 127.9, 128.3, 128.8, 130.5, 137.1, 140.4, 140.5, 142.1, 148.2, 148.4, 148.8, 150.2, 150.3, 150.5, 154.5, 159.8, 161.2, 161.3, 161.8, 162.0, 164.8, 167.6, 168.5, 169.0, 169.7, 170.7 ppm; IR (ATR):  $\tilde{\nu}$  = 2928 (w), 1658 (s), 1543 (s), 1494 (m), 1210 (m), 1051 (m), 751  $cm^{-1}$  (s); MS (ESI):  $m/z$  (%) = 1161 (50), 1139 (100), 1107 (4); HRMS (ESI):  $m/z$  =  $C_{49}H_{47}N_{12}O_9S_6$  [ $M+H$ ]<sup>+</sup>, calcd: 1139.1907, found: 1139.1887.

**Determination of MIC values:**<sup>[40]</sup> Substances were added in different concentrations to 1 mL B-broth [yeast extract (5.0 g), tryptic peptone (10.0 g), NaCl (5.0 g),  $K_2HPO_4$  (1.0 g),  $H_2O$  (1.0 L)] including a control with DMSO. The tubes were inoculated with  $1 \times 10^6$  bacteria per mL (OD: 0.6–0.8) and incubated overnight with shaking at 37 °C. The OD<sub>600</sub> value of 1:10 diluted overnight cultures was measured to determine the minimal inhibitory concentration (MIC). All experiments were conducted at least in triplicate, and DMSO served as control.

**Reagents, bacterial strains, and vectors:** The *E. coli* EF-Tu gene (*tufA*) cloned into pPROEX-Ht-b was a kind gift from Prof. Knud Nierhaus (MPIMG, Berlin).

**In vitro translation assays:** In vitro translation assays were performed using a homemade *E. coli* strain BL21 (Invitrogen) S12 lysate-based system, as described previously.<sup>[41]</sup> As before,<sup>[42]</sup> translation of the firefly luciferase (Fluc) reporter was monitored by measuring the luciferase activity in the presence and absence of antibiotic and/or EF-Tu.

**Protein expression and purification:** The QuikChange Site-directed Mutagenesis kit (Stratagene) was used to generate *E. coli* EF-Tu mutant G257S in pPROEX-Ht-b. Wild-type EF-Tu and the G257S

mutant were overexpressed in *E. coli* strain BL21(DE3) and purified using the N-terminal histidine tags via Ni-NTA affinity columns (QIAGEN) and subsequent gel-filtration chromatography on HiLoad 16/60 Superdex 75 prep grade column (Amersham-Pharmacia) in a buffer containing 20 mM HEPES (pH 7.8) and 150 mM NaCl.

**Computational studies:** Details of the methods employed for in silico experiments as well as additional figures (SI1–SI6) and a brief discussion of the simulation details are provided in the Supporting Information.

## Acknowledgements

This work was supported by the EMBO young investigator program and the Deutsche Forschungsgemeinschaft (WI3285/1-2 to D.N.W.). S.G. acknowledges support by the TUM Graduate School. I.A. and M.B. acknowledge support by the Deutsche Forschungsgemeinschaft (CiPSM cluster of excellence and SFB 1035).

**Keywords:** antibiotics • molecular modeling • natural products • protein biosynthesis • structure–activity relationships

- [1] J. Lucas-Lenard, F. Lipmann, *Proc. Natl. Acad. Sci. USA* **1966**, *55*, 1562–1566.
- [2] Reviews: a) R. Berisio, A. Ruggiero, L. Vitagliano, *Isr. J. Chem.* **2010**, *50*, 71–79; b) I. M. Krab, A. Parmeggiani, *Prog. Nucleic-Acid Res. Mol. Biol.* **2002**, *71*, 513–551.
- [3] H. Berchtold, L. Reshetnikova, C. O. A. Reiser, N. K. Schirmer, M. Sprinzl, R. Hilgenfeld, *Nature* **1993**, *365*, 126–132.
- [4] For an overview on protein biosynthesis see: D. Voet, C. W. Pratt, J. G. Voet, *Fundamentals in Biochemistry: Life at the Molecular Level*, 4th ed., Wiley, Hoboken, **2012**, pp. 962–1012.
- [5] A. Parmeggiani, P. Nissen, *FEBS Lett.* **2006**, *580*, 4576–4581.
- [6] a) S. E. Heffron, F. Journak, *Biochemistry* **2000**, *39*, 37–45; b) A. Parmeggiani, I. M. Krab, S. Okamura, R. C. Nielsen, J. Nyborg, P. Nissen, *Biochemistry* **2006**, *45*, 6846–6857. Protein numbering for EF-Tu was taken from the latter publication.
- [7] A. Parmeggiani, I. M. Krab, T. Watanabe, R. C. Nielsen, C. Dahlberg, J. Nyborg, P. Nissen, *J. Biol. Chem.* **2006**, *281*, 2893–2900.
- [8] L. Vogeley, G. J. Palm, J. R. Mesters, R. Hilgenfeld, *J. Biol. Chem.* **2001**, *276*, 17149–17155.
- [9] P. H. Anborgh, A. Parmeggiani, *J. Biol. Chem.* **1993**, *268*, 24622–24628.
- [10] a) E. Selva, G. Beretta, N. Montanini, G. S. Saddler, L. Gastaldo, P. Ferrari, R. Lorenzetti, P. Landini, F. Ripamonti, B. P. Goldstein, M. Berti, L. Montanaro, M. Denaro, *J. Antibiot.* **1991**, *44*, 693–701; b) J. Kettenring, L. Colombo, P. Ferrari, P. Tavecchia, M. Nebuloni, K. Vékey, G. G. Gallo, E. Selva, *J. Antibiot.* **1991**, *44*, 702–715; c) E. Selva, P. Ferrari, M. Kurz, P. Tavecchia, L. Colombo, S. Stella, E. Restelli, B. P. Goldstein, F. Ripamonti, M. Denaro, *J. Antibiot.* **1995**, *48*, 1039–1042; d) P. Tavecchia, P. Gentili, M. Kurz, C. Sottani, R. Bonfichi, E. Selva, S. Lociuro, E. Restelli, R. Ciabatti, *Tetrahedron* **1995**, *51*, 4867–4890.
- [11] E. Selva, N. Montanini, S. Stella, A. Soffientini, L. Gastaldo, M. Denaro, *J. Antibiot.* **1997**, *50*, 22–26.
- [12] J. Clough, S. Chen, E. M. Gordon, C. Hackbarth, S. Lam, J. Trias, R. J. White, G. Candiani, S. Donadio, G. Romanò, R. Ciabatti, J. W. Jacobs, *Bioorg. Med. Chem. Lett.* **2003**, *13*, 3409–3414.
- [13] M. J. LaMarche, J. A. Leeds, J. Dzink-Fox, K. Gunderson, P. Krastel, K. Memmert, M. A. Patane, E. M. Rann, E. Schmitt, S. Tiamfook, B. Wang, *J. Med. Chem.* **2011**, *54*, 2517–2521.
- [14] J. A. Leeds, M. J. LaMarche, J. T. Brewer, S. M. Bushell, G. Deng, J. M. Dewhurst, J. Dzink-Fox, E. Gangl, A. Jain, L. Lee, M. Lilly, K. Manni, S. Mullin, G. Neckermann, C. Osborne, D. Palestrant, M. A. Patane, A. Raimondi, S. Ranjitkar, E. M. Rann, M. Sachdeva, J. Shao, S. Tiamfook, L. Whitehead, D. Yu, *Antimicrob. Agents Chemother.* **2011**, *55*, 5277–5283.
- [15] M. J. LaMarche, J. A. Leeds, K. Amaral, J. T. Brewer, S. M. Bushell, J. M. Dewhurst, J. Dzink-Fox, E. Gangl, J. Goldovitz, A. Jain, S. Mullin, G. Neck-

- ermann, C. Osborne, D. Palestrant, M. A. Patane, E. M. Rann, M. Sachdeva, J. Shao, S. Tiamfook, L. Whitehead, D. Yu, *J. Med. Chem.* **2011**, *54*, 8099–8109.
- [16] J. A. Leeds, M. Sachdeva, S. Mullin, J. Dzink-Fox, M. J. LaMarche, *J. Med. Chem.* **2012**, *55*, 2376–2387.
- [17] S. Lociuero, P. Tavecchia, E. Marzorati, P. Landini, B. P. Goldstein, M. Denaro, R. Ciabatti, *J. Antibiot.* **1997**, *50*, 344–349.
- [18] Reviews: a) M. C. Bagley, J. W. Dale, E. A. Merritt, X. Xiong, *Chem. Rev.* **2005**, *105*, 685–714; b) R. A. Hughes, C. J. Moody, *Angew. Chem.* **2007**, *119*, 8076–8101; *Angew. Chem. Int. Ed.* **2007**, *46*, 7930–7954.
- [19] a) K. Shimanaka, N. Kinoshita, H. Iinuma, M. Hamada, T. Takeuchi, *J. Antibiot.* **1994**, *47*, 668–674; b) K. Shimanaka, Y. Takahashi, H. Iinuma, H. Naganawa, T. Takeuchi, *J. Antibiot.* **1994**, *47*, 1145–1152; c) K. Shimanaka, Y. Takahashi, H. Iinuma, H. Naganawa, T. Takeuchi, *J. Antibiot.* **1994**, *47*, 1153–1159.
- [20] K. Shimanaka, H. Iinuma, M. Hamada, *J. Antibiot.* **1995**, *48*, 182–184.
- [21] Selected references to completed total syntheses of thiopeptides: a) K. C. Nicolaou, B. S. Safina, M. Zak, S. H. Lee, M. Nevalainen, M. Bella, A. A. Estrada, C. Funke, F. J. Zécri, S. Bulat, *J. Am. Chem. Soc.* **2005**, *127*, 11159–11175; b) K. C. Nicolaou, M. Zak, B. S. Safina, A. A. Estrada, S. H. Lee, M. Nevalainen, *J. Am. Chem. Soc.* **2005**, *127*, 11176–11183; c) R. A. Hughes, S. P. Thompson, L. Alcaraz, C. J. Moody, *J. Am. Chem. Soc.* **2005**, *127*, 15644–15651; d) K. C. Nicolaou, B. Zou, D. H. Dethé, D. B. Li, D. Y.-K. Chen, *Angew. Chem.* **2006**, *118*, 7950–7956; *Angew. Chem. Int. Ed.* **2006**, *45*, 7786–7792; e) K. C. Nicolaou, D. H. Dethé, G. Y. C. Leung, B. Zou, D. Y.-K. Chen, *Chem. Asian J.* **2008**, *3*, 413–429; f) T. i. Mori, S. Higashibayashi, T. Goto, M. Kohno, Y. Satouchi, K. Shinko, K. Suzuki, S. Suzuki, H. Tohmiya, K. Hashimoto, M. Nakata, *Chem. Asian J.* **2008**, *3*, 984–1012; g) D. Lefranc, M. A. Ciufolini, *Angew. Chem.* **2009**, *121*, 4262–4265; D. Lefranc, M. A. Ciufolini, *Angew. Chem. Int. Ed.* **2009**, *48*, 4198–4201; h) D. Lefranc, M. A. Ciufolini, *Nat. Prod. Rep.* **2010**, *27*, 330–342; i) V. S. Aulakh, M. A. Ciufolini, *J. Am. Chem. Soc.* **2011**, *133*, 5900–5904; j) K. C. Nicolaou, *Angew. Chem.* **2012**, *124*, 12582–12604; *Angew. Chem. Int. Ed.* **2012**, *51*, 12414–12436.
- [22] a) H. M. Müller, O. Delgado, T. Bach, *Angew. Chem.* **2007**, *119*, 4855–4858; *Angew. Chem. Int. Ed.* **2007**, *46*, 4771–4774; b) O. Delgado, H. M. Müller, T. Bach, *Chem. Eur. J.* **2008**, *14*, 2322–2339.
- [23] C. Ammer, T. Bach, *Chem. Eur. J.* **2010**, *16*, 14083–14093.
- [24] a) G. Liu, D. A. Cogan, J. A. Ellman, *J. Am. Chem. Soc.* **1997**, *119*, 9913–9914; b) G. Liu, D. A. Cogan, T. D. Owens, T. P. Tang, J. A. Ellman, *J. Org. Chem.* **1999**, *64*, 1278–1284; c) D. A. Cogan, G. Liu, J. Ellman, *Tetrahedron* **1999**, *55*, 8883–8904; d) J. A. Ellman, T. D. Owens, T. P. Tang, *Acc. Chem. Res.* **2002**, *35*, 984–995.
- [25] J. C. Barrow, P. L. Ngo, J. M. Pellicore, H. G. Selnick, P. G. Nanterment, *Tetrahedron Lett.* **2001**, *42*, 2051–2054.
- [26] A. Spieß, G. Heckmann, T. Bach, *Synlett* **2004**, 131–133.
- [27] a) J. A. Dale, H. S. Mosher, *J. Am. Chem. Soc.* **1973**, *95*, 512–519; b) T. R. Hoye, C. S. Jeffrey, F. Shao, *Nat. Protoc.* **2007**, *2*, 2451–2458.
- [28] E. Frérot, J. Coste, A. Pantaloni, M.-N. Dufour, P. Jouin, *Tetrahedron* **1991**, *47*, 259–270.
- [29] S. Gross, S. Heuser, C. Ammer, G. Heckmann, T. Bach, *Synthesis* **2011**, 199–206.
- [30] H. Azizian, C. Eaborn, A. Pidcock, *J. Organomet. Chem.* **1981**, *215*, 49–58.
- [31] a) D. Milstein, J. K. Stille, *J. Am. Chem. Soc.* **1979**, *101*, 4992–4998; b) J. K. Stille, *Angew. Chem.* **1986**, *98*, 504–519; *Angew. Chem. Int. Ed. Engl.* **1986**, *25*, 508–524; c) V. Farina, V. Krishnamurthy, W. J. Scott, *Org. React.* **1997**, *50*, 1–652.
- [32] a) T. Bach, S. Heuser, *J. Org. Chem.* **2002**, *67*, 5789–5795; b) G. Heckmann, T. Bach, *Angew. Chem.* **2005**, *117*, 1223–1226; *Angew. Chem. Int. Ed.* **2005**, *44*, 1199–1201.
- [33] a) Y. Hamada, S. Shioiri, *Tetrahedron Lett.* **1982**, *23*, 235–236; b) L. Qian, Z. Sun, T. Deffo, K. Bowman Mertes, *Tetrahedron Lett.* **1990**, *31*, 6469–6472; c) S. Jayaprakash, G. Pattenden, M. S. Viljoen, C. Wilson, *Tetrahedron* **2003**, *59*, 6637–6646.
- [34] M. Misumi, N. Tanaka, *Biochem. Biophys. Res. Commun.* **1980**, *92*, 647–654.
- [35] I. Chopra, *J. Antimicrob. Chemother.* **1998**, *41*, 584–588.
- [36] A. M. Zuurmond, J. Martien de Graaf, L. N. Olsthoorn-Tielemans, B. Y. van Duyl, V. G. Möhrle, F. Jurnak, J. R. Mesters, R. Hilgenfeld, B. Kraal, *J. Mol. Biol.* **2000**, *304*, 995–1005.
- [37] P. Nissen, M. Kjeldgaard, S. Thirup, G. Polekhina, L. Reshetnikova, B. F. Clark, J. Nyborg, *Science* **1995**, *270*, 1464–1472.
- [38] R. J. Lewis, R. A. Hughes, L. Alcaraz, S. P. Thompson, C. J. Moody, *Chem. Commun.* **2006**, 4215–4217.
- [39] I. Antes, *Proteins* **2010**, *78*, 1084–1104.
- [40] M. B. Nodwell, H. Menz, S. F. Kirsch, S. A. Sieber, *ChemBioChem* **2012**, *13*, 1439–1446.
- [41] T. W. Kim, H. C. Kim, I. S. Oh, D. M. A. Kim, *Biotechnol. Bioprocess Eng.* **2008**, *13*, 464–469.
- [42] a) A. L. Starosta, H. Qin, A. Mikolajka, G. Y. Leung, K. Schwinghammer, K. C. Nicolaou, D. Y. Chen, B. S. Cooperman, D. N. Wilson, *Chem. Biol.* **2009**, *16*, 1087–1096; b) A. L. Starosta, V. V. Karpenko, A. V. Shishkina, A. Mikolajka, N. V. Sumbatyan, F. Schluenzen, G. A. Korshunova, A. A. Bogdanov, D. N. Wilson, *Chem. Biol.* **2010**, *17*, 504–514.

Received: July 26, 2013

Published online on September 17, 2013

## Supporting Information

© Copyright Wiley-VCH Verlag GmbH & Co. KGaA, 69451 Weinheim, 2013

### **Amythiamicin D and Related Thiopeptides as Inhibitors of the Bacterial Elongation Factor EF-Tu: Modification of the Amino Acid at Carbon Atom C2 of Ring C Dramatically Influences Activity**

Stefan Gross,<sup>[a]</sup> Fabian Nguyen,<sup>[b]</sup> Matthias Bierschenk,<sup>[c]</sup> Daniel Sohmen,<sup>[b]</sup> Thomas Menzel,<sup>[d]</sup>  
Iris Antes,<sup>\*[c]</sup> Daniel N. Wilson,<sup>\*[b]</sup> and Thorsten Bach<sup>\*[a]</sup>

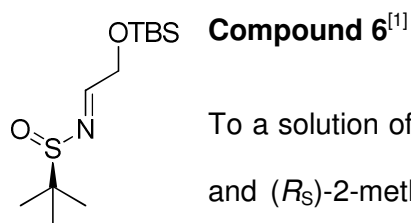
cmdc\_201300323\_sm\_miscellaneous\_information.pdf

## Table of contents

Compound 6.....	1
Compounds 7a and 7b .....	2
Compound 8a.....	3
Compound 8b.....	4
Compound 9a.....	4
Compound 9b.....	4
Configurational assignments by formation of Mosher amides from amine 9a .....	5
Compound 10a.....	6
Compound 10b.....	7
Compound 11a.....	7
Compound 11b.....	8
Compound 12a.....	8
Compound 12b.....	9
Compound 5a.....	9
Compound 5b.....	10
Compound 14a.....	10
Compound 14b.....	11
Macrocycle 15a .....	12
Macrocycle 15b .....	13
Thiopeptide 3a .....	14
Thiopeptide 3b .....	15
Compound 9c.....	17

Compound 10c.....	18
Compound 11c.....	19
Compound 12c.....	19
Compound 5c.....	20
Compound 14c.....	21
Macrocycle 15c .....	22
Thiopeptide 3c.....	23
Compounds 5a and 5b, <sup>1</sup> H-NMR .....	25
Compounds 5a and 5b, <sup>13</sup> C-NMR .....	26
Compound 5c, <sup>1</sup> H-NMR .....	27
Compound 5c, <sup>13</sup> C-NMR.....	28
Thiopeptide 3a, <sup>1</sup> H-NMR.....	29
Thiopeptide 3a, <sup>13</sup> C-NMR .....	30
Thiopeptide 3b, <sup>1</sup> H-NMR.....	31
Thiopeptide 3b, <sup>13</sup> C-NMR .....	32
Thiopeptide 3c, <sup>1</sup> H-NMR.....	34
Thiopeptide 3c, <sup>13</sup> C-NMR .....	35
Materials and methods for <i>in silico</i> studies .....	36
References.....	39
Translational assays.....	47

**General remarks:** All reactions involving water-sensitive chemicals were carried out in flame-dried glassware with magnetic stirring under argon. Tetrahydrofuran (THF), diethyl ether (Et<sub>2</sub>O) and dichloromethane (CH<sub>2</sub>Cl<sub>2</sub>) were dried using a SPS-800 solvent purification system (M. Braun). Chemicals were either commercially available or prepared according to the cited references. TLC was performed on silica coated glass plates (0.25 mm silica gel 60 F<sub>254</sub>) with detection by UV (254 nm) or KMnO<sub>4</sub> (0.5% in water) with subsequent heating. Flash chromatography was performed on silica gel 60 (Merck, 230-400 mesh) with the indicated eluent. Common solvents for chromatography such as pentane (P), ethyl acetate (EtOAc), diethyl ether, dichloromethane and methanol (MeOH) were distilled prior to use. IR: JASCO IR-4100 (ATR). MS and HRMS: Finnigan MAT 8200 (MS-EI) and Finnigan MAT 95S (HRMS-EI). <sup>1</sup>H and <sup>13</sup>C NMR: Bruker AV-200 and Bruker AV-360, recorded at 296 K. Chemical shifts are reported relative to tetramethylsilane. The multiplicities of the <sup>13</sup>C NMR signal were determined by DEPT experiments, assignments are based on COSY, HMBC and HMQC experiments. <sup>1</sup>H or <sup>13</sup>C NMR signals are usually assigned using significant short sections of the molecular formula. Optical rotations were measured using a Perkin-Elmer 241 MC Polarimeter.

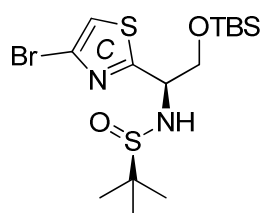
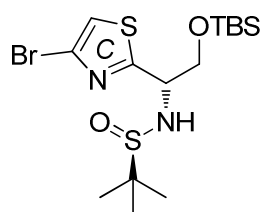


To a solution of 2-(*tert*-butyldimethylsilyloxy)acetaldehyd (3.19 g, 18.3 mmol) and (*R*<sub>S</sub>)-2-methyl-2-propanesulfonamide (2.01 g, 16.6 mmol) in CH<sub>2</sub>Cl<sub>2</sub> (30 mL) was added anhydrous CuSO<sub>4</sub> (5.29 g, 33.2 mmol) and the resulting suspension was stirred overnight at room temperature. The mixture was diluted with CH<sub>2</sub>Cl<sub>2</sub> (50 mL) and filtered through celite. The solvent was removed in vacuo and the residue was purified by flash chromatography (P/EtOAc = 8:1) yielding sulfonamide **6** (4.24 g, 15.3 mmol, 92%) as a colorless oil. *R*<sub>f</sub> = 0.35 (P/EtOAc = 8:1, UV), [α]<sub>D</sub><sup>20</sup> = -157 (*c* = 1.28 in CHCl<sub>3</sub>); <sup>1</sup>H-NMR (CDCl<sub>3</sub>, 360 MHz): δ(ppm) = 0.10 [s, 6 H, Si(CH<sub>3</sub>)<sub>2</sub>], 0.91 [s, 9 H, SiC(CH<sub>3</sub>)<sub>3</sub>], 1.20 [s, 9 H,

<sup>[1]</sup> J. C. Barrow, P. L. Ngo, J. M. Pellicore, H. G. Selnick, P. G. Nantermet, *Tetrahedron Lett.* **2001**, *42*, 2051-2054.

SC(CH<sub>3</sub>)<sub>3</sub>], 4.54 (d, <sup>3</sup>J = 3.0 Hz, 2 H, CH<sub>2</sub>OTBS), 8.06 (t, <sup>3</sup>J = 3.0 Hz, 1 H, N=CH); <sup>13</sup>C-NMR (CDCl<sub>3</sub>, 91 MHz): δ(ppm) = -4.9 [q, Si(CH<sub>3</sub>)<sub>2</sub>], 18.8 [s, SiC(CH<sub>3</sub>)<sub>3</sub>], 22.8 [q, SC(CH<sub>3</sub>)<sub>3</sub>], 26.2 [q, SiC(CH<sub>3</sub>)<sub>3</sub>], 57.3 [s, SC(CH<sub>3</sub>)<sub>3</sub>], 65.9 (t, CH<sub>2</sub>OTBS), 169.2 (d, N=CH).

### Compounds 7a and 7b

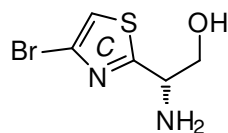


To a solution of 2,4-Dibromothiazole (4.73 g, 19.5 mmol) in THF (10 mL) was added <sup>i</sup>PrMgBr (4.90 mL, 19.5 mmol, 4.0 M in Et<sub>2</sub>O) at 0 °C and the mixture was stirred for 2 h. The obtained *Grignard* reagent was then added to a solution of sulfinamide **6** (3.00 g, 10.5 mmol) in CH<sub>2</sub>Cl<sub>2</sub> (200 mL), cooled to -78 °C. The mixture was allowed to warm up to room temperature overnight, before the reaction was quenched by addition of sat. NH<sub>4</sub>Cl solution (200 mL). The layers were separated and the aqueous layer was extracted with CH<sub>2</sub>Cl<sub>2</sub> (2 × 100 mL). The combined extracts were washed with brine (1 × 200 mL) and dried over Na<sub>2</sub>SO<sub>4</sub>. After filtration the solvent was removed under reduced pressure. Purification by flash chromatography (P/EtOAc = 4:1) yielded the major diastereoisomer **7a** (3.15 g, 7.13 mmol, 66%), the minor diastereoisomer **7b** (1.18 g, 2.67 mmol, 25%) and mixed fractions (86.2 mg, 0.20 mmol, 2%) as yellow oils. Diastereoisomer **7a**: *R*<sub>f</sub> = 0.25 (P/EtOAc = 4:1, UV); [α]<sub>D</sub><sup>20</sup> = -49.7 (*c* = 1.15 in CHCl<sub>3</sub>); IR (ATR):  $\tilde{\nu}$  [cm<sup>-1</sup>] = 2942 (m, NH), 2856 (m), 1470 (m), 1252 (m), 1077 (s), 834 (s), 776 (m), 757 (m); <sup>1</sup>H-NMR (CDCl<sub>3</sub>, 360 MHz): δ(ppm) = -0.07 (s, 3 H, SiCH<sub>3</sub>), -0.04 (s, 3 H, SiCH<sub>3</sub>), 0.81 [s, 9 H, SiC(CH<sub>3</sub>)<sub>3</sub>], 1.30 [s, 9 H, SC(CH<sub>3</sub>)<sub>3</sub>], 4.08 (dd, <sup>2</sup>J = 9.8 Hz, <sup>3</sup>J = 3.8 Hz, 1 H, CHHOTBS), 4.16 (dd, <sup>2</sup>J = 9.8 Hz, <sup>3</sup>J = 3.7 Hz, 1 H, CHHOTBS), 4.65 (d, <sup>3</sup>J = 6.5 Hz, 1 H, NH), 4.77-4.82 (m, 1 H, CHNH), 7.16 (s, 1 H, C: 5-H); <sup>13</sup>C-NMR (CDCl<sub>3</sub>, 91 MHz): δ(ppm) = -5.7 (q, SiCH<sub>3</sub>), -5.4 (q, SiCH<sub>3</sub>), 18.1 [s, SiC(CH<sub>3</sub>)<sub>3</sub>], 22.6 [q, SC(CH<sub>3</sub>)<sub>3</sub>], 25.7 [q, SiC(CH<sub>3</sub>)<sub>3</sub>], 56.4 [s, SC(CH<sub>3</sub>)<sub>3</sub>], 59.1 (d, CHNH), 66.1 (t, CH<sub>2</sub>OTBS), 117.4 (d, C: C5), 125.1 (s, C: C4), 173.8 (s, C: C2); MS (EI, 70 eV): *m/z* (%) = 386 (10) [(M(<sup>81</sup>Br)-C<sub>4</sub>H<sub>8</sub>)<sup>+</sup>], 384 (10) [(M(<sup>79</sup>Br)-C<sub>4</sub>H<sub>8</sub>)<sup>+</sup>], 329 (10), 327 (10), 264 (22), 190 (25), 89 (38),



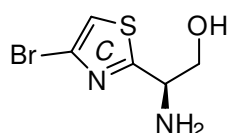
75 (58), 73 (54), 57 (100) [C<sub>4</sub>H<sub>9</sub><sup>+</sup>]; HRMS (EI): m/z = C<sub>11</sub>H<sub>21</sub><sup>79</sup>BrN<sub>2</sub>O<sub>2</sub>S<sub>2</sub>Si [(M(<sup>79</sup>Br)-C<sub>4</sub>H<sub>8</sub>)<sup>+</sup>], calcd.: 383.9997, found: 383.9998; Diastereoisomer **7b**: *R*<sub>f</sub> = 0.15 (P/EtOAc = 4:1, UV); [α]<sub>D</sub><sup>20</sup> = -1.30 (*c* = 1.10 in CHCl<sub>3</sub>); <sup>1</sup>H-NMR (CDCl<sub>3</sub>, 360 MHz): δ(ppm) = -0.06 (s, 3 H, SiCH<sub>3</sub>), -0.01 (s, 3 H, SiCH<sub>3</sub>), 0.82 [s, 9 H, SiC(CH<sub>3</sub>)<sub>3</sub>], 1.23 [s, 9 H, SC(CH<sub>3</sub>)<sub>3</sub>], 3.89 (dd, <sup>2</sup>*J* = 9.8 Hz, <sup>3</sup>*J* = 5.0 Hz, 1 H, CHHOTBS), 4.04 (dd, <sup>2</sup>*J* = 9.8 Hz, <sup>3</sup>*J* = 3.6 Hz, 1 H, CHHOTBS), 4.65 (d, <sup>2</sup>*J* = 6.5 Hz, 1 H, NH), 4.75-4.80 (m, 1 H, CHNH), 7.18 (s, 1 H, C: 5-H); <sup>13</sup>C-NMR (CDCl<sub>3</sub>, 91 MHz): δ(ppm) = -5.7 (q, SiCH<sub>3</sub>), -5.6 (q, SiCH<sub>3</sub>), 18.1 [s, SiC(CH<sub>3</sub>)<sub>3</sub>], 22.5 [q, SC(CH<sub>3</sub>)<sub>3</sub>], 25.7 [q, SiC(CH<sub>3</sub>)<sub>3</sub>], 56.7 [s, SC(CH<sub>3</sub>)<sub>3</sub>], 57.1 (d, CHNH), 66.4 (t, CH<sub>2</sub>OTBS), 117.6 (d, C: C5), 124.5 (s, C: C4), 172.2 (s, C: C2).

### Compound 8a



To a solution of sulfinamide **7a** (5.64 g, 12.8 mmol) in MeOH (100 mL) was added 4 M HCl in dioxane (6.40 mL, 25.6 mmol) and the mixture was stirred for 1 h at room temperature. The solvent was removed under reduced pressure and the residue was taken up in EtOAc (150 mL). 2 M NaOH (150 mL) was added, the layers were separated and the aqueous layer was extracted with EtOAc (2 × 100 mL). The combined organic layers were washed with brine (1 × 100 mL) and dried over Na<sub>2</sub>SO<sub>4</sub>. After filtration and evaporation of the solvent the crude product was purified by flash chromatography (CH<sub>2</sub>Cl<sub>2</sub>/MeOH = 10:1) to yield **8a** (2.41 g, 10.8 mmol, 84%) as a pale-yellow solid. *R*<sub>f</sub> = 0.35 (CH<sub>2</sub>Cl<sub>2</sub>/MeOH = 10:1); [α]<sub>D</sub><sup>20</sup> = -31.7 (*c* = 1.03 in CHCl<sub>3</sub>); M.p. 112 °C; IR (ATR):  $\tilde{\nu}$  [cm<sup>-1</sup>] = 3323 (m), 3250 (m, NH), 3091 (m), 2924 (m), 1582 (m), 1471 (s), 1092 (m), 1049 (m), 967 (m), 767 (m); <sup>1</sup>H-NMR (CDCl<sub>3</sub>, 360 MHz): δ(ppm) = 2.11 (br. s, 3 H, NH<sub>2</sub> and OH), 3.89 (m<sub>c</sub>, 2 H, CH<sub>2</sub>OH), 4.30 (virt. t, <sup>3</sup>*J* ≅ 5.4 Hz, 1 H, CHNH<sub>2</sub>), 7.20 (s, 1 H, C: 5-H); <sup>13</sup>C-NMR (CDCl<sub>3</sub>, 91 MHz): δ(ppm) = 55.2 (d, CHNH<sub>2</sub>), 66.9 (t, CH<sub>2</sub>OH), 117.2 (d, C: C5), 124.6 (s, C: C4), 175.9 (s, C: C2); MS (EI, 70 eV): m/z (%) = 224 (1) [M(<sup>81</sup>Br)<sup>+</sup>], 222 (1) [M(<sup>79</sup>Br)<sup>+</sup>], 193 (100) [(M(<sup>81</sup>Br)-CH<sub>2</sub>OH)<sup>+</sup>], 191 (95) [(M(<sup>79</sup>Br)-CH<sub>2</sub>OH)<sup>+</sup>], 166 (25), 164 (23),

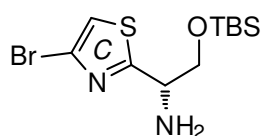
139 (10), 137 (10); HRMS (EI):  $m/z = C_4H_4^{79}BrN_2S [(M(^{79}Br)-CH_2OH)^+]$ , calcd.: 190.9279, found.: 190.9284.



#### Compound 8b

This compound was prepared from sulfonamide **7b** according to the procedure described above. Aminoalcohol **8b** (0.56 g, 2.51 mmol, 94%)

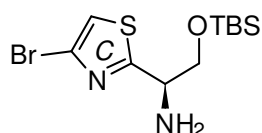
was obtained as a pale-yellow solid.  $[\alpha]_D^{20} = +32.5$  ( $c = 1.03$  in  $CHCl_3$ ).



#### Compound 9a

A solution of aminoalcohol **8a** (223 mg, 1.00 mmol) in  $CH_2Cl_2$  (8 mL) were added  $NEt_3$  (0.42 mL, 303 mg, 3.00 mmol) and TBSCl (332 mg,

2.20 mmol) at 0 °C. The reaction was allowed to warm up to room temperature overnight and it was then quenched by the addition of sat.  $NaHCO_3$  (10 mL). The mixture was extracted with  $CH_2Cl_2$  (3 × 20 mL). The combined organic layers were washed with brine (1 × 30 mL) and dried over  $Na_2SO_4$ . Filtration and evaporation of the solvent yielded the crude product that was purified by flash chromatography (P/EtOAc = 10:1). The title compound **9a** (273 mg, 0.81 mmol, 81%) was obtained as a yellow oil.  $R_f = 0.30$  (P/EtOAc = 4:1, UV);  $^1H$ -NMR ( $CDCl_3$ , 360 MHz):  $\delta$ (ppm) = 0.01 (s, 3 H,  $SiCH_3$ ), 0.05 (s, 3 H,  $SiCH_3$ ), 0.87 [s, 9 H,  $SiC(CH_3)_3$ ], 1.94 (br. s, 2 H,  $NH_2$ ), 3.76 (dd,  $^2J = 9.8$  Hz,  $^3J = 6.4$  Hz, 1 H,  $CHHOTBS$ ), 3.96 (dd,  $^2J = 9.8$  Hz,  $^3J = 4.1$  Hz, 1 H,  $CHHOTBS$ ), 4.33 (dd,  $^3J = 6.4$ , 4.1 Hz, 1 H,  $CHNH_2$ ), 7.15 (s, 1 H, C: 5-H).



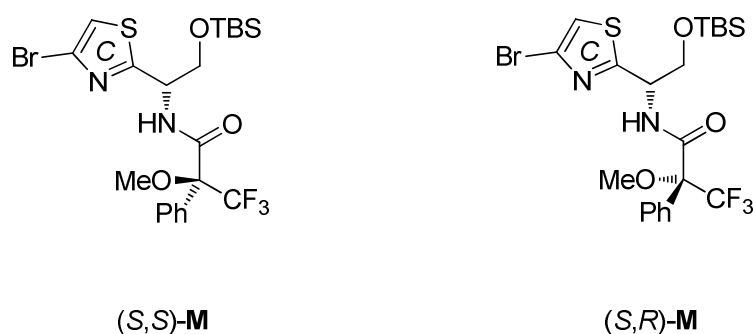
#### Compound 9b

This compound was prepared from **8b** according to the procedure described above for **9a**. Aminoalcohol **9b** (0.11 g, 0.32 mmol, 72%)

was obtained as a yellow oil.

### Configurational assignments by formation of Mosher amides from amine **9a**<sup>[2]</sup>

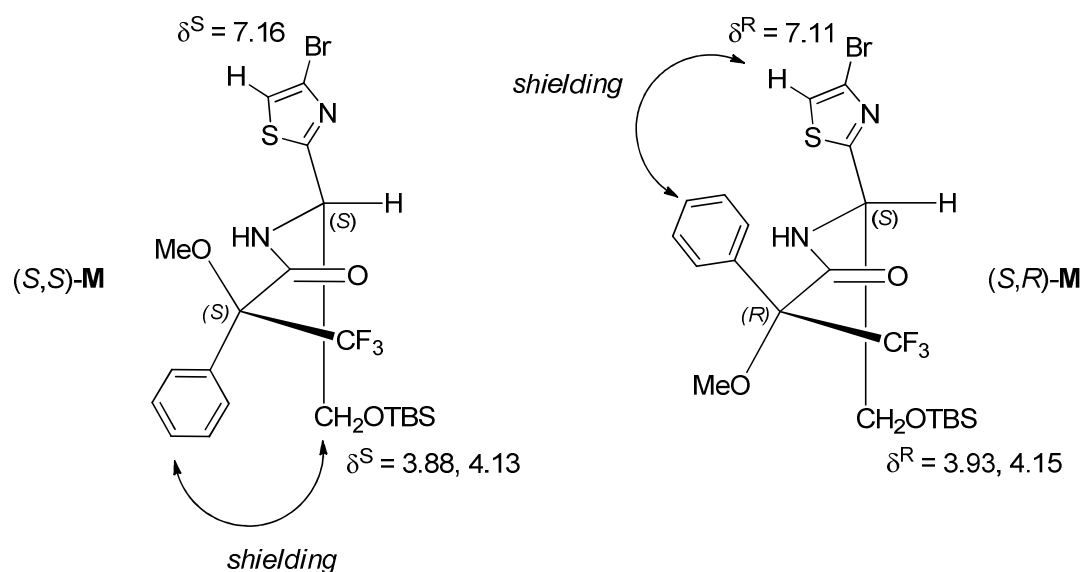
To a solution of **9a** (7.90 mg, 23.4  $\mu$ mol) in dry  $\text{CDCl}_3$  (0.7 mL) were added pyridine (5.37  $\mu$ L, 5.56 mg, 70.2  $\mu$ mol) and (*R*)-MTPACI (8.76  $\mu$ L, 11.8 mg, 46.8  $\mu$ mol). After 1 h at room temperature conversion into (*S,S*)-**M** was complete. The mixture was transferred into a NMR tube and subjected to  $^1\text{H-NMR}$  analysis. The diastereoisomer (*S,R*)-**M** was obtained following the identical procedure, reacting **9a** with (*R*)-MTPACI. Data of relevant signals is given in the table below (Table S1).



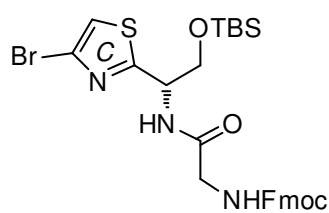
**Table S1.**  $\delta^{\text{S}}$  = chemical shifts of (*S,S*)-**M**,  $\delta^{\text{R}}$  = chemical shifts of (*S,R*)-**M**,  $\delta\delta^{\text{S,R}} = \delta^{\text{S}} - \delta^{\text{R}}$ .

assignment	multiplicity	$\delta^{\text{S}}$ [ppm]	$\delta^{\text{R}}$ [ppm]	$\delta\delta^{\text{S,R}}$ [ppm]
SiCH <sub>3</sub>	s	-0.10	-0.08	-0.02
SiCH <sub>3</sub>	s	-0.04	+0.01	-0.05
SiC(CH <sub>3</sub> ) <sub>3</sub>	s	+0.80	+0.81	-0.01
CHHOTBS	dd	+3.88	+3.93	-0.05
CHHOTBS	dd	+4.13	+4.15	-0.02
C: 5-H	s	+7.16	+7.11	+0.05

<sup>[2]</sup> D. A. Allen, A. E. Tomaso, O. P. Priest, *J. Chem. Ed.* **2008**, *85*, 698-700.



**Figure S1.** Shielding effects in *Mosher* amides (*S,S*)-**M** and (*S,R*)-**M**.

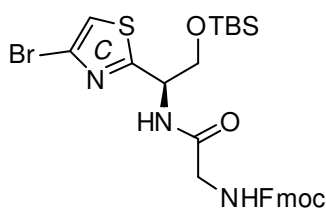


#### Compound 10a

To a solution of amine **9a** (1.98 g, 5.87 mmol) in  $\text{CH}_2\text{Cl}_2$  (100 mL) were added  $i\text{Pr}_2\text{NEt}$  (2.99 mL, 2.27 g, 17.6 mmol), Fmoc-Gly-OH (2.09 g, 7.04 mmol) and PyBrOP (3.28 g, 7.04 mmol) at 0 °C.

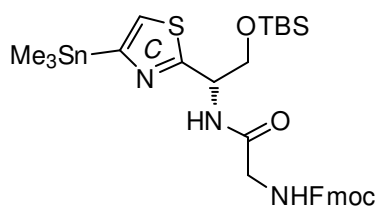
After stirring for 4 h at 0 °C the reaction was quenched by the addition of sat.  $\text{NaHCO}_3$  (100 mL). The layers were separated and the aqueous layer was extracted with  $\text{CH}_2\text{Cl}_2$  (2 × 50 mL). Drying over  $\text{Na}_2\text{SO}_4$ , filtration, evaporation of the solvent and purification by flash chromatography ( $\text{P/EtOAc} = 2:1$ ) yielded the title compound **10a** (3.39 g, 5.50 mmol, 94%) as a colorless foam.  $R_f = 0.30$  ( $\text{P/EtOAc} = 2:1$ , UV);  $[\alpha]_D^{20} = -0.7$  ( $c = 1.06$  in  $\text{CHCl}_3$ ); IR (ATR):  $\tilde{\nu}$  [ $\text{cm}^{-1}$ ] = 2943 (w), 2856 (w), 1679 (m), 1510 (m), 1471 (w), 1251 (s), 1102 (m), 833 (s), 780 (m), 737 (s);  $^1\text{H-NMR}$  ( $\text{CDCl}_3$ , 360 MHz):  $\delta$ (ppm) =  $-0.06$  (s, 3 H,  $\text{SiCH}_3$ ),  $-0.01$  (s, 3 H,  $\text{SiCH}_3$ ), 0.82 [s, 9 H,  $\text{SiC}(\text{CH}_3)_3$ ], 3.86 ( $m_c$ , 1 H,  $\text{CHHOTBS}$ ), 3.98 [ $m_c$ , 2 H,  $\text{C}(=\text{O})\text{CH}_2\text{NH}$ ], 4.16 (dd,  $^2J = 10.0$  Hz,  $^3J = 3.2$  Hz, 1 H,  $\text{CHHOTBS}$ ), 4.24 (t,  $^3J = 6.7$  Hz, 1 H, Fmoc:  $\text{CHCH}_2\text{O}$ ), 4.44 (d,  $^3J = 6.7$  Hz, 2 H, Fmoc:  $\text{CHCH}_2\text{O}$ ), 5.35 ( $m_c$ , 1 H,  $\text{CHNH}$ ), 5.42 [br. s, 1 H,  $\text{C}(=\text{O})\text{CH}_2\text{NH}$ ], 7.00 (br. d,  $^3J = 7.2$  Hz, 1 H,  $\text{CHNH}$ ), 7.13 (s, 1 H, C: 5-H), 7.31 (virt. t,  $^3J \cong$

7.4 Hz, 2 H, Fmoc: 2-H and 7-H), 7.41 (virt. t,  $^3J \cong 7.4$  Hz, 2 H, Fmoc: 3-H and 6-H), 7.58 (d,  $^3J = 7.3$  Hz, 2 H, Fmoc: 1-H and 8-H), 7.77 (d,  $^3J = 7.5$  Hz, 2 H, Fmoc: 4-H and 5-H);  $^{13}\text{C}$ -NMR ( $\text{CDCl}_3$ , 91 MHz):  $\delta(\text{ppm}) = -5.7$  (q,  $\text{SiCH}_3$ ),  $-5.6$  (q,  $\text{SiCH}_3$ ), 18.1 [s,  $\text{SiC}(\text{CH}_3)_3$ ], 25.7 [q,  $\text{SiC}(\text{CH}_3)_3$ ], 44.6 (d, Fmoc:  $\text{CHCH}_2\text{O}$ ), 47.1 [t,  $\text{C}(=\text{O})\text{CH}_2\text{NH}$ ], 53.1 (d,  $\text{CHNH}$ ), 64.7 (t,  $\text{CH}_2\text{OTBS}$ ), 67.4 (t, Fmoc:  $\text{CHCH}_2\text{O}$ ), 117.1 (d, C: C5), 120.0 (d, Fmoc: C4 and C5), 124.6 (s, C: C4), 125.0 (d, Fmoc: C1 and C8), 127.1 (d, Fmoc: C2 and C7), 127.8 (d, Fmoc: C3 and C6), 141.3 (s, Fmoc: C4a und C4b), 143.7 (s, Fmoc: C8a and C8b), 156.6 [s,  $\text{CH}_2\text{OC}(=\text{O})\text{NH}$ ], 168.7 [s,  $\text{C}(=\text{O})\text{CH}_2\text{NH}$ ], 170.8 (s, C: C2); MS (ESI):  $m/z$  (%) = 640 (100)  $[(\text{M}^{81}\text{Br})+\text{Na}]^+$ , 638 (100)  $[(\text{M}^{79}\text{Br})+\text{Na}]^+$ , 618 (13)  $[(\text{M}^{81}\text{Br})+\text{H}]^+$ , 616 (13)  $[(\text{M}^{79}\text{Br})+\text{H}]^+$ , 536 (1)  $[(\text{M}-\text{Br})^+]$ , 441 (4), 179 (2); HRMS (ESI):  $m/z = \text{C}_{28}\text{H}_{34}^{79}\text{BrN}_3\text{O}_4\text{SSiNa} [(\text{M}^{79}\text{Br})+\text{Na}]^+$ , calcd.: 638.1120, found: 638.1107.



#### Compound 10b

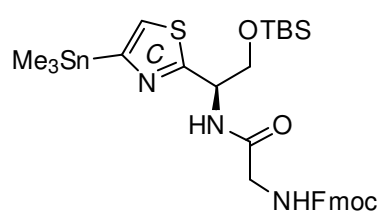
This compound was prepared from **9b** according to the procedure described above. Compound **10b** (0.73 g, 1.18 mmol, 99%) was obtained as a colorless foam.  $[\alpha]_D^{20} = +0.5$  ( $c = 0.44$  in  $\text{CHCl}_3$ ).



#### Compound 11a

To a solution of bromothiazole **10a** (1.54 g, 2.50 mmol) in degassed toluene (30 mL) were added  $\text{Pd}(\text{PPh}_3)_4$  (290 mg, 0.25 mmol) and hexamethylditin (2.08 mL, 3.28 g, 10.0 mmol). After stirring for 1.5 h at 100 °C the mixture was concentrated in vacuo. The residue was purified by flash chromatography (P/EtOAc = 2:1) to yield stannane **11a** (1.43 g, 2.04 mmol, 82%) as a colorless foam.  $R_f = 0.30$  (P/EtOAc = 2:1, UV);  $^1\text{H}$ -NMR ( $\text{CDCl}_3$ , 360 MHz):  $\delta(\text{ppm}) = -0.11$  (s, 3 H,  $\text{SiCH}_3$ ),  $-0.05$  (s, 3 H,  $\text{SiCH}_3$ ), 0.34 [s, 9 H,  $\text{Sn}(\text{CH}_3)_3$ ], 0.80 [s, 9 H,  $\text{SiC}(\text{CH}_3)_3$ ], 3.86 ( $m_c$ , 1 H,  $\text{CHHOTBS}$ ), 4.01 [ $m_c$ , 2 H,  $\text{C}(=\text{O})\text{CH}_2\text{NH}$ ], 4.17 (dd,  $^2J = 9.7$  Hz,  $^3J$

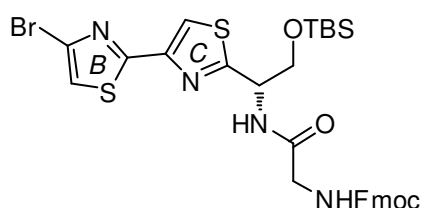
= 3.1 Hz, 1 H, CHHOTBS), 4.23 (t,  $^3J = 7.1$  Hz, 1 H, Fmoc: CHCH<sub>2</sub>O), 4.42 (d,  $^3J = 7.1$  Hz, 2 H, Fmoc: CHCH<sub>2</sub>O), 5.43 (m<sub>c</sub>, 1 H, CHNH), 5.51 [br. s, 1 H, C(=O)CH<sub>2</sub>NH], 7.16 (br. d,  $^3J = 7.4$  Hz, 1 H, CHNH), 7.27 (s, 1 H, C: 5-H), 7.31 (virt. t,  $^3J \cong 7.3$  Hz, 2 H, Fmoc: 2-H and 7-H), 7.40 (virt. t,  $^3J \cong 7.4$  Hz, 2 H, Fmoc: 3-H and 6-H), 7.60 (d,  $^3J = 7.3$  Hz, 2 H, Fmoc: 1-H and 8-H), 7.77 (d,  $^3J = 7.5$  Hz, 2 H, Fmoc: 4-H and 5-H). After <sup>1</sup>H-NMR stannane **11a** was quickly subjected to the next step.



### Compound 11b

This compound was prepared from **10b** according to the procedure described above. Compound **11b** (0.23 g, 0.33 mmol, 80%) was obtained as a colorless foam. After <sup>1</sup>H-NMR

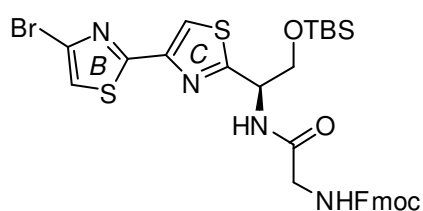
stannane **11b** was quickly subjected to the next step.



### Compound 12a

Stannane **11a** (1.40 g, 2.00 mmol) was dissolved in degassed toluene (30 mL). Pd(PPh<sub>3</sub>)<sub>4</sub> (230 mg, 0.20 mmol) and 2,4-dibromothiazole (0.68 g, 2.80 mmol) were sequentially added and the solution was stirred for 16 h at 90 °C. Sat. NaHCO<sub>3</sub> (30 mL) was added and the mixture was extracted with EtOAc (3 × 30 mL). The combined organic layers were dried over Na<sub>2</sub>SO<sub>4</sub> and after filtration the solvents were removed under reduced pressure. Flash chromatography (P/EtOAc = 2:1) yielded the bithiazole **12a** (1.20 g, 1.71 mmol, 86%) as a colorless solid. *R*<sub>f</sub> = 0.25 (P/EtOAc = 2:1, UV); M.p. 82 °C; [α]<sub>D</sub><sup>20</sup> = -7.2 (*c* = 1.01 in CHCl<sub>3</sub>); IR (ATR):  $\tilde{\nu}$  [cm<sup>-1</sup>] = 2846 (w), 1716 (s, C=O), 1678 (s, C=O), 1513 (s), 1450 (s), 1252 (m), 1106 (w), 839 (w), 752 (m); <sup>1</sup>H-NMR (CDCl<sub>3</sub>, 360 MHz): δ(ppm) = -0.07 (s, 3 H, SiCH<sub>3</sub>), -0.01 (s, 3 H, SiCH<sub>3</sub>), 0.81 [s, 9 H, SiC(CH<sub>3</sub>)<sub>3</sub>], 3.89 (m<sub>c</sub>, 1 H, CHHOTBS), 4.01 (m<sub>c</sub>, 2 H, C(=O)CH<sub>2</sub>NH), 4.20-4.26 (m, 2 H, CHHOTBS and Fmoc: CHCH<sub>2</sub>O), 4.42-4.47 (m, 2 H, Fmoc: CHCH<sub>2</sub>O), 5.39 (m<sub>c</sub>, 1 H, CHNH), 5.49 [br. s, 1 H, C(=O)CH<sub>2</sub>NH], 7.15 (br. d,  $^3J =$

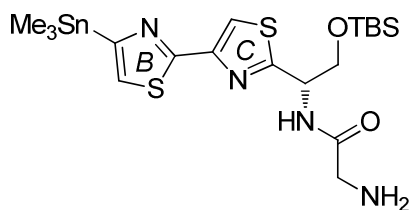
8.3 Hz, 1 H, CHNH), 7.17 (s, 1 H, B: 5-H), 7.30 (virt. t,  $^3J \cong 7.4$  Hz, 2 H, Fmoc: 2-H and 7-H), 7.40 (virt. t,  $^3J \cong 7.5$  Hz, 2 H, Fmoc: 3-H and 6-H), 7.59 (d,  $^3J = 7.2$  Hz, 2 H, Fmoc: 1-H and 8-H), 7.77 (d,  $^3J = 7.5$  Hz, 2 H, Fmoc: 4-H and 5-H), 7.92 (s, 1 H, C: 5-H);  $^{13}\text{C}$ -NMR ( $\text{CDCl}_3$ , 91 MHz):  $\delta(\text{ppm}) = -5.6$  (q,  $\text{SiCH}_3$ ),  $-5.6$  (q,  $\text{SiCH}_3$ ), 18.1 [s,  $\text{SiC}(\text{CH}_3)_3$ ], 25.7 [q,  $\text{SiC}(\text{CH}_3)_3$ ], 44.8 (d, Fmoc:  $\text{CHCH}_2\text{O}$ ), 47.1 [t,  $\text{C}(=\text{O})\text{CH}_2\text{NH}$ ], 53.1 (d, CHNH), 64.7 (t,  $\text{CH}_2\text{OTBS}$ ), 67.5 (t, Fmoc:  $\text{CHCH}_2\text{O}$ ), 117.0 (d, C: C5), 117.4 (d, B: C5), 120.0 (d, Fmoc: C4 and C5), 125.0 (d, Fmoc: C1 and C8), 126.0 (s, B: C4), 127.1 (d, Fmoc: C2 and C7), 127.8 (d, Fmoc: C3 and C6), 141.3 (s, Fmoc: C4a and C4b), 143.6 (s, Fmoc: C8a and C8b), 147.9 (s, C: C4), 156.6 [s,  $\text{CH}_2\text{OC}(=\text{O})\text{NH}$ ], 163.2 (s, B: C2), 168.7 [s,  $\text{C}(=\text{O})\text{CH}_2\text{NH}$ ], 170.7 (s, C: C2); MS (ESI):  $m/z$  (%) = 834 (10), 832 (20), 806 (30), 804 (55), 723 (70) [( $\text{M}^{81}\text{Br}+\text{Na}$ ) $^+$ ], 721 (70) [( $\text{M}^{79}\text{Br}+\text{Na}$ ) $^+$ ], 701 (16) [( $\text{M}^{81}\text{Br}+\text{H}$ ) $^+$ ], 699 (14) [( $\text{M}^{79}\text{Br}+\text{H}$ ) $^+$ ], 413 (100); HRMS (ESI):  $m/z = \text{C}_{31}\text{H}_{35}^{79}\text{BrN}_4\text{O}_4\text{S}_2\text{Si}$  [( $\text{M}^{79}\text{Br}+\text{H}$ ) $^+$ ], calcd.: 699.1131, found: 699.1118.



### Compound 12b

This compound was prepared from **11b** according to the procedure described above. Compound **12b** (0.53 g, 0.76 mmol, 94%) was obtained as colorless solid.  $[\alpha]_D^{20} = +6.4$

( $c = 0.76$  in  $\text{CHCl}_3$ ).



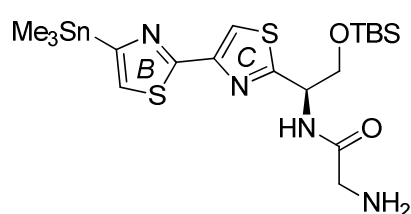
### Compound 5a

To a solution of bithiazole **12a** (140 mg, 200  $\mu\text{mol}$ ) in degassed toluene (3 mL) was added  $\text{Pd}(\text{PPh}_3)_4$  (23.1 mg, 20.0  $\mu\text{mol}$ , 10 mol%) and hexamethylditin (166  $\mu\text{L}$ , 262 mg,

800  $\mu\text{mol}$ ). After stirring for 2.5 h at 90  $^\circ\text{C}$  the solvent was removed under reduced pressure and the residue was purified by flash chromatography ( $\text{P}/\text{EtOAc} = 2:1$ ). The obtained stannane was dissolved in DMF (3 mL). Piperidine (0.3 mL) was added and the solution was stirred for 3 h at room temperature. The solvent was removed in vacuo and the crude product

was then purified by flash chromatography (CH<sub>2</sub>Cl<sub>2</sub>/MeOH = 10:1) to yield **5a** (105 mg, 187 μmol, 94% over 2 steps) as a pale-yellow oil. *R*<sub>f</sub> = 0.15 (CH<sub>2</sub>Cl<sub>2</sub>/MeOH = 10:1, UV); [α]<sub>D</sub><sup>20</sup> = -25.5 (*c* = 1.30 in CHCl<sub>3</sub>); IR (ATR):  $\tilde{\nu}$  [cm<sup>-1</sup>] = 2952 (m), 2928 (m), 2851 (w), 1721 (m), 1678 (s, C=O), 1503 (s), 1465 (m), 1281 (m), 1256 (s), 1096 (s), 1023 (w), 834 (s), 776 (s); <sup>1</sup>H-NMR (CDCl<sub>3</sub>, 360 MHz): δ(ppm) = -0.06 (s, 3 H, SiCH<sub>3</sub>), 0.01 (s, 3 H, SiCH<sub>3</sub>), 0.38 [s, 9 H, Sn(CH<sub>3</sub>)<sub>3</sub>], 0.83 [s, 9 H, SiC(CH<sub>3</sub>)<sub>3</sub>], 1.59 (br. s, 2 H, NH<sub>2</sub>), 3.47 (s, 2 H, CH<sub>2</sub>NH<sub>2</sub>), 3.92 (dd, <sup>2</sup>*J* = 9.8 Hz, <sup>3</sup>*J* = 4.9 Hz, 1 H, CHHOTBS), 4.27 (dd, <sup>2</sup>*J* = 9.8 Hz, <sup>3</sup>*J* = 3.2 Hz, 1 H, CHHOTBS), 5.43 (ddd, <sup>3</sup>*J* = 8.2, 4.9, 3.2 Hz, 1 H, CHNH), 7.37 (s, 1 H, B: 5-H), 7.89 (s, 1 H, C: 5-H), 8.34 (br. d, <sup>3</sup>*J* = 8.2 Hz, 1 H, CHNH); <sup>13</sup>C-NMR (CDCl<sub>3</sub>, 91 MHz) δ(ppm) = -8.8 [q, Sn(CH<sub>3</sub>)<sub>3</sub>], -5.6 [q, Si(CH<sub>3</sub>)<sub>2</sub>], 18.1 [s, SiC(CH<sub>3</sub>)<sub>3</sub>], 25.7 [q, SiC(CH<sub>3</sub>)<sub>3</sub>], 44.6 (t, CH<sub>2</sub>NH<sub>2</sub>), 52.7 (d, CHNH), 64.9 (t, CH<sub>2</sub>OTBS), 115.7 (d, C: C5), 126.1 (d, B: C5), 149.6 (s, C: C4), 161.0 (s, B: C4), 163.4 (s, B: C2), 170.7 (s, C: C2), 172.1 (s, C=O); MS (ESI): *m/z* (%) = 603 (1), 563 (39) [(M(<sup>120</sup>Sn)+H)<sup>+</sup>], 282 (100) [(M(<sup>120</sup>Sn)+H<sub>2</sub>)<sup>2+</sup>]; HRMS (ESI): *m/z* = C<sub>19</sub>H<sub>35</sub>N<sub>4</sub>O<sub>2</sub>S<sub>2</sub>Si<sup>120</sup>Sn [(M(<sup>120</sup>Sn)+H)<sup>+</sup>], calcd.: 563.0987, found: 563.0978.

### Compound 5b



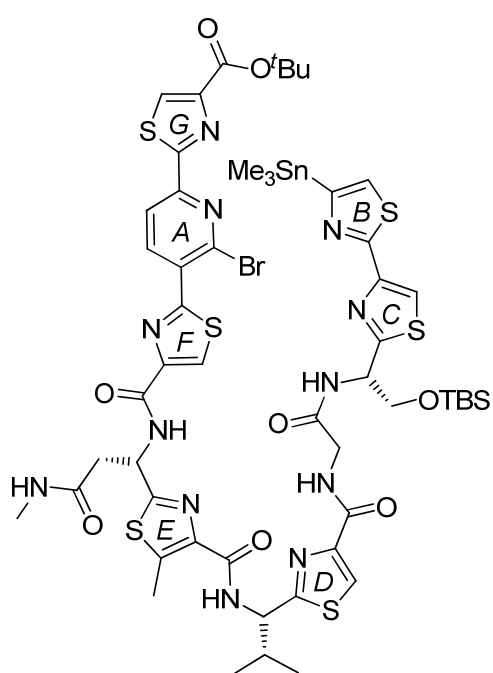
The corresponding enantiomer **5b** (166 μmol, 83% over 2 steps) was obtained from **12a** using the identical procedure as described above for eastern building block **5a**; [α]<sub>D</sub><sup>20</sup> = +28.5 (*c* = 1.10 in CHCl<sub>3</sub>).

### Compound 14a

To a solution of building block **13** (81.7 mg, 90.4 μmol) in THF (0.90 mL) and <sup>t</sup>BuOH (1.80 mL) was dropwise added 1 M LiOH (0.90 mL, 0.90 mmol) and the mixture was stirred at room temperature for 3 h. The solvents were removed under reduced pressure and the residue was taken up in water (10 mL). The mixture was adjusted to pH 3 by addition of 10% citric acid and then extracted with CH<sub>2</sub>Cl<sub>2</sub> (3 × 20 mL). After drying over Na<sub>2</sub>SO<sub>4</sub> and filtration



the solvent was removed in vacuo. The obtained carboxylic acid was dissolved in DMF (10 mL) and then **5a** (66.2 mg, 118  $\mu\text{mol}$ ), *i*Pr<sub>2</sub>NEt (61.6  $\mu\text{L}$ , 46.8 mg, 362  $\mu\text{mol}$ ) and DPPA (38.9  $\mu\text{L}$ , 49.8 mg, 181  $\mu\text{mol}$ ) were added at 0 °C. The reaction was allowed to warm up to room temperature overnight before it was quenched by the addition of sat. NH<sub>4</sub>Cl (20 mL). The mixture was extracted with CH<sub>2</sub>Cl<sub>2</sub> (3  $\times$  20 mL) and the combined organic layers were dried over Na<sub>2</sub>SO<sub>4</sub>. Filtration, evaporation of the solvent in vacuo and purification by flash chromatography (EtOAc) yielded organostannane **14a** (80.4 mg, 56.6  $\mu\text{mol}$ , 63%) as a pale-yellow oil. *R*<sub>f</sub> = 0.50 (EtOAc, UV); <sup>1</sup>H-NMR (CDCl<sub>3</sub>, 500 MHz):  $\delta$ (ppm) = -0.09 (s, 3 H, SiCH<sub>3</sub>),

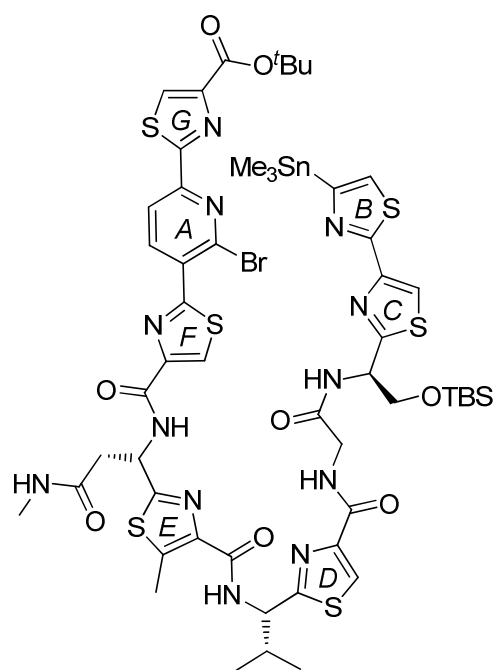


-0.03 (s, 3 H, SiCH<sub>3</sub>), 0.36 [s, 9 H, Sn(CH<sub>3</sub>)<sub>3</sub>], 0.79 [s, 9 H, SiC(CH<sub>3</sub>)<sub>3</sub>], 1.00 [virt. t, <sup>3</sup>*J*  $\cong$  6.2 Hz, 6 H, CH(CH<sub>3</sub>)<sub>2</sub>], 1.64 [s, 9 H, OC(CH<sub>3</sub>)<sub>3</sub>], 2.36 [m<sub>c</sub>, 1 H, CH(CH<sub>3</sub>)<sub>2</sub>], 2.67 (d, <sup>3</sup>*J* = 4.7 Hz, 3 H, NHCH<sub>3</sub>), 2.75 (s, 3 H, E: 5-CH<sub>3</sub>), 2.92 [dd, <sup>2</sup>*J* = 15.8 Hz, <sup>3</sup>*J* = 5.0 Hz, 1 H, CHHC(=O)NHMe], 3.18 [dd, <sup>2</sup>*J* = 15.8 Hz, <sup>3</sup>*J* = 4.5 Hz, 1 H, CHHC(=O)NHMe], 3.90 (dd, <sup>2</sup>*J* = 10.0 Hz, <sup>3</sup>*J* = 5.1 Hz, 1 H, CHHOTBS), 4.11-4.23 [m, 2 H, CHHOTBS and C(=O)CHHNH], 4.47 [dd, <sup>2</sup>*J* = 16.7 Hz, <sup>3</sup>*J* = 6.6 Hz, 1 H, C(=O)CHHNH], 5.31-5.37 (m, 2 H, C: CHNH and D: CHNH), 5.74 (m<sub>c</sub>, 1 H, E:

CHNH), 6.60 (d, <sup>3</sup>*J* = 4.7 Hz, 1 H, NHCH<sub>3</sub>), 7.20 (d, <sup>3</sup>*J* = 7.9 Hz, 1 H, C: CHNH), 7.35 (s, 1 H, C: 5-H), 7.91 (s, 1 H, B: 5-H), 8.06-8.10 (m, 1 H, D: CHNH), 8.09 (s, 1 H, D: 5-H), 8.16-8.22 [m, 1 H, C(=O)CHHNH], 8.21 (s, 1 H, G: 5-H), 8.34 (s, 1 H, F: 5-H), 8.39 (d, <sup>3</sup>*J* = 8.2 Hz, 1 H, A: 5-H), 8.74 (d, <sup>3</sup>*J* = 8.2 Hz, 1 H, A: 4-H), 9.58 (d, <sup>3</sup>*J* = 8.8 Hz, 1 H, E: CHNH). After <sup>1</sup>H-NMR stannane **14a** was quickly subjected to the cyclization step.

## Compound 14b

This compound was prepared as described above for **14a** starting from compounds **13** and **5b** yielding compound **14b** (97.4 mg, 68.6  $\mu\text{mol}$ , 79%) as a pale-yellow oil.  $R_f = 0.50$  (EtOAc,



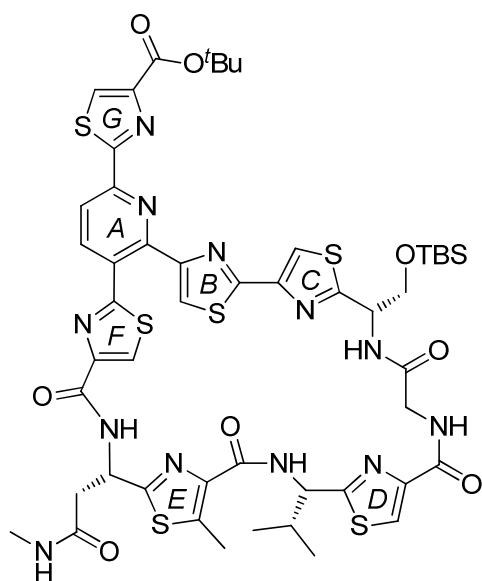
UV),  $^1\text{H-NMR}$  ( $\text{CDCl}_3$ , 500 MHz):  $\delta(\text{ppm}) = -0.08$  (s, 3 H,  $\text{SiCH}_3$ ),  $-0.01$  (s, 3 H,  $\text{SiCH}_3$ ), 0.35 [s, 9 H,  $\text{Sn}(\text{CH}_3)_3$ ], 0.81 [s, 9 H,  $\text{SiC}(\text{CH}_3)_3$ ], 1.00 [d,  $^3J = 6.6$  Hz, 3 H,  $\text{CH}(\text{CH}_3)_2$ ], 1.02 [d,  $^3J = 6.9$  Hz, 3 H,  $\text{CH}(\text{CH}_3)_2$ ], 1.64 [s, 9 H,  $\text{OC}(\text{CH}_3)_3$ ], 2.37 [m<sub>c</sub>, 1 H,  $\text{CH}(\text{CH}_3)_2$ ], 2.58 (d,  $^3J = 4.7$  Hz, 3 H,  $\text{NHCH}_3$ ), 2.75

(s, 3 H, E: 5- $\text{CH}_3$ ), 2.96 [dd,  $^2J = 15.8$  Hz,  $^3J = 4.8$  Hz, 1 H,  $\text{CHHC}(\text{=O})\text{NHMe}$ ], 3.18 [dd,  $^2J = 15.8$  Hz,  $^3J = 4.5$  Hz, 1 H,  $\text{CHHC}(\text{=O})\text{NHMe}$ ], 3.93 (dd,  $^2J = 9.9$  Hz,  $^3J = 4.9$  Hz, 1 H,  $\text{CHHOTBS}$ ), 4.15 [dd,  $^2J = 16.8$  Hz,  $^3J = 4.6$  Hz, 1 H,  $\text{C}(\text{=O})\text{CHHNH}$ ], 4.22 (dd,  $^2J = 9.9$  Hz,  $^3J = 3.2$  Hz, 1 H,  $\text{CHHOTBS}$ ), 4.52 [dd,  $^2J = 16.8$  Hz,  $^3J = 6.5$  Hz, 1 H,  $\text{C}(\text{=O})\text{CHHNH}$ ], 5.33-5.39 (m, 2 H, C:  $\text{CHNH}$  and D:  $\text{CHNH}$ ), 5.77 (m<sub>c</sub>, 1 H, E:  $\text{CHNH}$ ), 6.58 (m<sub>c</sub>, 1 H,  $\text{NHCH}_3$ ), 7.24-7.26 (m, 1 H, NH, C:  $\text{CHNH}$ ), 7.33 (s, 1 H, C: 5-H), 7.82 (s, 1 H, B: 5-H), 8.08 (s, 1 H, D: 5-H), 8.06-8.10 (m, 1 H, D:  $\text{CHNH}$ ), 8.19-8.22 [m, 1 H,  $\text{C}(\text{=O})\text{CHHNH}$ ], 8.20 (s, 1 H, G: 5-H), 8.33 (s, 1 H, F: 5-H), 8.37 (d,  $^3J = 8.2$  Hz, 1 H, A: 5-H), 8.73 (d,  $^3J = 8.2$  Hz, 1 H, A: 4-H), 9.53 (d,  $^3J = 8.7$  Hz, 1 H, E:  $\text{CHNH}$ ). After  $^1\text{H-NMR}$  stannane *epi-78a* was quickly subjected to the cyclization step.

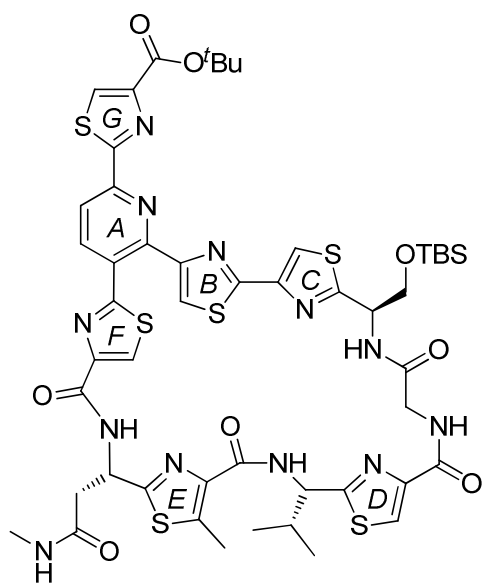
### Macrocycle 15a

To a solution of compound **14a** (19.9 mg, 14.0  $\mu\text{mol}$ ) in degassed toluene (14 mL) was added  $\text{Pd}(\text{PPh}_3)_4$  (3.24 mg, 2.80  $\mu\text{mol}$ , 20 mol%) and the mixture was stirred for 20 h at 85  $^\circ\text{C}$ . The solvent was removed in vacuo and the crude product was purified by flash chromatography (EtOAc/MeOH = 60:1) yielding macrocycle **15a** (10.5 mg, 8.96  $\mu\text{mol}$ , 64%) as a pale-yellow wax.  $R_f = 0.50$  (EtOAc/MeOH = 60:1, UV);  $^1\text{H-NMR}$  ( $\text{CDCl}_3$ , 500 MHz):

$\delta(\text{ppm}) = 0.01$  (s, 3 H, SiCH<sub>3</sub>),  $0.03$  (s, 3 H, SiCH<sub>3</sub>),  $0.88$  [d,  $^3J = 6.8$  Hz, 3 H, CH(CH<sub>3</sub>)<sub>2</sub>],  $0.92$  [s, 9 H, SiC(CH<sub>3</sub>)<sub>3</sub>],  $1.00$  [d,  $^3J = 6.8$  Hz, 3 H, CH(CH<sub>3</sub>)<sub>2</sub>],  $1.01$  [m<sub>c</sub>, 1 H, CHHC(=O)NHMe],



$1.65$  [s, 9 H, OC(CH<sub>3</sub>)<sub>3</sub>],  $2.28$  [m<sub>c</sub>, 1 H, CH(CH<sub>3</sub>)<sub>2</sub>],  $2.64$  (d,  $^3J = 6.0$  Hz, 3 H, NHCH<sub>3</sub>),  $2.65$  (s, 3 H, E: 5-CH<sub>3</sub>),  $2.70$  [dd,  $^2J = 17.1$  Hz,  $^3J = 3.5$  Hz, 1 H, CHHC(=O)NHMe],  $3.85$ - $3.91$  [m, 1 H, C(=O)CHHNH],  $3.90$ - $3.95$  (m, 1 H, CHHOTBS),  $4.05$  (dd,  $^2J = 10.3$  Hz,  $^3J = 3.5$  Hz, 1 H, CHHOTBS),  $4.99$  [dd,  $^2J = 17.5$  Hz,  $^3J = 9.4$  Hz, 1 H, C(=O)CHHNH],  $5.24$  (dd,  $^3J = 7.9$ ,  $4.6$  Hz, 1 H, D: CHNH),  $5.35$  (m<sub>c</sub>, 1 H, C: CHNH),  $5.41$  (m<sub>c</sub>, 1 H, E: CHNH),  $6.75$ - $6.85$  (m, 2 H, NHCH<sub>3</sub> and C: CHNH),  $7.25$  (s, 1 H, C: 5-H),  $7.76$  [dd,  $^3J = 9.4$ ,  $2.8$  Hz, 1 H, C(=O)CHHNH],  $8.11$  (d,  $^3J = 8.2$  Hz, 1 H, A: 4-H),  $8.11$  (s, 1 H, D: 5-H),  $8.20$  (s, 1 H, B: 5-H),  $8.25$  (s, 1 H, G: 5-H),  $8.37$  (s, 1 H, F: 5-H),  $8.38$  (d,  $^3J = 8.2$  Hz, 1 H, A: 5-H),  $8.78$  (d,  $^3J = 7.9$  Hz, 1 H, D: CHNH),  $8.97$  (d,  $^3J = 9.2$  Hz, 1 H, E: CHNH). This compound was obtained as a mixture together with triphenylphosphine oxide and was directly used in the next step.



### Macrocycle 15b

Starting from compound **14b** macrocycle **15b** (25.7 mg, 21.9  $\mu\text{mol}$ , 62%) was obtained as a pale-yellow wax, using the procedure given above for macrocycle **15a**.  $R_f = 0.50$  (EtOAc/MeOH = 60:1, UV),  $^1\text{H-NMR}$  (CDCl<sub>3</sub>, 500 MHz):  $\delta(\text{ppm}) = -0.09$  (s, 3 H, SiCH<sub>3</sub>),  $-0.06$  (s, 3 H, SiCH<sub>3</sub>),  $0.84$  [s, 9 H, SiC(CH<sub>3</sub>)<sub>3</sub>],  $0.95$  [d,  $^3J = 6.8$  Hz, 3 H, CH(CH<sub>3</sub>)<sub>2</sub>],  $1.03$  [d,  $^3J = 6.7$  Hz, 3 H, CH(CH<sub>3</sub>)<sub>2</sub>],  $1.65$  [s, 9 H, OC(CH<sub>3</sub>)<sub>3</sub>],  $2.20$  [m<sub>c</sub>, 1 H, CH(CH<sub>3</sub>)<sub>2</sub>],  $2.26$ - $2.37$  [m, 2 H,

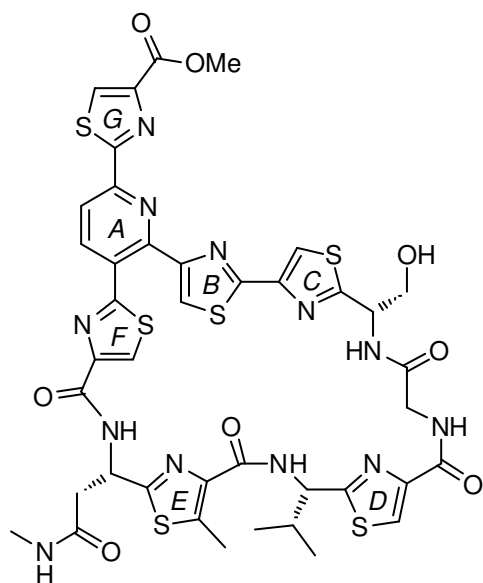
$\text{CH}_2\text{C}(=\text{O})\text{NHMe}$ ], 2.56 (d,  $^3J = 4.5$  Hz, 3 H,  $\text{NHCH}_3$ ), 2.68 (s, 3 H, E: 5- $\text{CH}_3$ ), 4.03-4.12 [m, 2 H,  $\text{CH}_2\text{OTBS}$ ], 4.24 [dd,  $^2J = 17.8$  Hz,  $^3J = 3.4$  Hz, 1 H,  $\text{C}(=\text{O})\text{CHHNH}$ ], 4.53 [dd,  $^2J = 17.8$  Hz,  $^3J = 5.6$  Hz, 1 H,  $\text{C}(=\text{O})\text{CHHNH}$ ], 5.26-5.31 (m, 2 H, C:  $\text{CHNH}$  and D:  $\text{CHNH}$ ), 5.45 (m<sub>c</sub>, 1 H, E:  $\text{CHNH}$ ), 6.57 (br. s, 1 H,  $\text{NHCH}_3$ ), 6.85 (d,  $^3J = 6.8$  Hz, 1 H, C:  $\text{CHNH}$ ), 7.39 (s, 1 H, C: 5-H), 7.91 (s, 1 H, B: 5-H), 8.08 (d,  $^3J = 9.0$  Hz, 1 H, A: 4-H), 8.08 (s, 1 H, D: 5-H), 8.11-8.18 (m, 2 H, D:  $\text{CHNH}$  and  $\text{C}(=\text{O})\text{CH}_2\text{NH}$ ], 8.20 (s, 1 H, F: 5-H), 8.24 (d,  $^3J = 9.8$  Hz, 1 H, E:  $\text{CHNH}$ ), 8.33 (s, 1 H, G: 5-H), 8.41 (d,  $^3J = 9.0$  Hz, 1 H, A: 5-H). This compound was obtained as a mixture together with triphenylphosphine oxide and was directly used in the next step.

### Thiopeptide 3a

To a solution of **15a** (7.50 mg, 6.37  $\mu\text{mol}$ ) in  $\text{CH}_2\text{Cl}_2$  (1 mL) was added TFA (0.2 mL) and the mixture was stirred for 6 h at room temperature. The solvent was removed under reduced pressure and TFA was co-evaporated with toluene (2  $\times$  1 mL). The residue was dried in vacuo and dissolved in THF (0.2 mL) and MeOH (0.2 mL).  $\text{TMSCHN}_2$  (4.14  $\mu\text{L}$ , 8.28  $\mu\text{mol}$ , 2.0 M in  $\text{Et}_2\text{O}$ ) was added and the solution was stirred for 1 h at room temperature. The

solvents were removed in vacuo and the crude product was purified by flash chromatography ( $\text{CH}_2\text{Cl}_2/\text{MeOH} = 10:1$ ) to yield thiopeptide **3a** (3.40 mg, 3.34  $\mu\text{mol}$ , 52% over 2 steps) as a colorless solid.  $R_f = 0.25$  ( $\text{CH}_2\text{Cl}_2/\text{MeOH} = 10:1$ , UV);  $[\alpha]_D^{20} = +172$  ( $c = 0.26$  in  $\text{CHCl}_3$ ); IR (ATR):  $\tilde{\nu}$  [ $\text{cm}^{-1}$ ] = 3280 (w), 2939 (w), 1725 (w), 1658 (s, C=O), 1539 (m), 1492 (m), 1224 (m), 911 (w), 728 (m);  $^1\text{H-NMR}$  ( $\text{CDCl}_3$ , 500 MHz):  $\delta$ (ppm) = 0.91 [d,  $^3J = 6.8$  Hz, 3 H,  $\text{CH}(\text{CH}_3)_2$ ], 1.01 [d,  $^3J = 6.9$  Hz, 3 H,  $\text{CH}(\text{CH}_3)_2$ ], 1.28

[m<sub>c</sub>, 1 H,  $\text{CHHC}(=\text{O})\text{NHMe}$ ], 2.31 [m<sub>c</sub>, 1 H,  $\text{CH}(\text{CH}_3)_2$ ], 2.65 (d,  $^3J = 4.7$  Hz, 3 H,  $\text{NHCH}_3$ ),

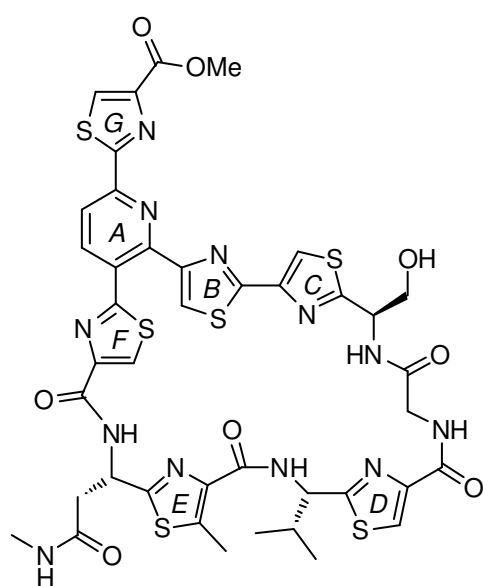


2.67 (s, 3 H, E: 5-CH<sub>3</sub>), 2.68-2.72 [m, 1 H, CHHC(=O)NHMe], 3.37 (br. s, 1 H, OH), 3.96 [dd, <sup>2</sup>J = 17.5 Hz, <sup>3</sup>J = 3.5 Hz, 1 H, C(=O)CHHNH], 4.01-4.05 (m, 1 H, CHHOH), 4.02 (s, 3 H, OCH<sub>3</sub>), 4.12 (dd, <sup>2</sup>J = 11.6 Hz, <sup>3</sup>J = 3.2 Hz, 1 H, CHHOH), 4.95 [dd, <sup>2</sup>J = 17.5 Hz, <sup>3</sup>J = 9.2 Hz, 1 H, C(=O)CHHNH], 5.26 (dd, <sup>3</sup>J = 8.1, 5.0 Hz, 1 H, D: CHNH), 5.37 (m<sub>c</sub>, 1 H, C: CHNH), 5.42 (virt. dt, <sup>3</sup>J ≅ 8.8, 4.3 Hz, 1 H, E: CHNH), 6.79 (d, <sup>3</sup>J = 4.7 Hz, 1 H, NHCH<sub>3</sub>), 7.30 (s, 1 H, C: 5-H), 7.39 (br. d, <sup>3</sup>J = 6.0 Hz, 1 H, C: CHNH), 7.95 [dd, <sup>3</sup>J = 9.2, 3.5 Hz, 1 H, C(=O)CHHNH], 8.13 (d, <sup>3</sup>J = 8.1 Hz, 1 H, A: 4-H), 8.14 (s, 1 H, D: 5-H), 8.16 (s, 1 H, B: 5-H), 8.35 (s, 1 H, G: 5-H), 8.36 (s, 1 H, F: 5-H), 8.39 (d, <sup>3</sup>J = 8.1 Hz, 1 H, A: 5-H), 8.66 (d, <sup>3</sup>J = 8.1 Hz, 1 H, D: CHNH), 8.78 (br. d, <sup>3</sup>J = 8.8 Hz, 1 H, E: CHNH); <sup>13</sup>C-NMR (CDCl<sub>3</sub>, 126 MHz): δ(ppm) = 12.4 (q, E: 5-CH<sub>3</sub>), 18.2 [q, CH(CH<sub>3</sub>)<sub>2</sub>], 18.4 [q, CH(CH<sub>3</sub>)<sub>2</sub>], 26.3 (q, NHCH<sub>3</sub>), 34.7 [d, CH(CH<sub>3</sub>)<sub>2</sub>], 38.6 [t, CH<sub>2</sub>C(=O)NHMe], 41.8 [t, C(=O)CH<sub>2</sub>NH], 48.2 (d, E: CHNH), 52.7 (q, OCH<sub>3</sub>), 55.0 (d, C: CHNH), 56.1 (d, D: CHNH), 63.8 (t, CH<sub>2</sub>OH), 115.2 (d, C: C5), 118.8 (d, A: C5), 122.9 (d, B: C5), 124.1 (d, D: C5), 125.3 (d, F: C5), 127.8 (s, A: C3), 130.6 (d, G: C5), 140.3 (d, A: C4), 140.9 (s, E: C5), 142.1 (s, E: C4), 148.2 (s, G: C4), 148.3 (s, D: C4), 148.9 (s, C: C4), 150.1 (s, F: C4), 150.3 (s, A: C2), 150.5 (s, A: C6), 154.6 (s, B: C4), 159.9 (s, B: C2), 161.2 [s, D: C(=O)NH], 161.7 [s, E: C(=O)NH], 161.8 [s, C(=O)OMe], 161.9 [s, F: C(=O)NH], 164.9 (s, F: C2), 167.1 (s, E: C2), 168.8 (s, D: C2), 169.0 (s, G: C2), 169.2 [s, C(=O)CH<sub>2</sub>NH], 169.8 [s, C(=O)NHMe], 170.6 (s, C: C2); MS (ESI): m/z (%) = 1019 (100) [(M+H)<sup>+</sup>]; HRMS (ESI): m/z = C<sub>41</sub>H<sub>39</sub>N<sub>12</sub>O<sub>8</sub>S<sub>6</sub> [(M+H)<sup>+</sup>], calcd.: 1019.1333, found: 1019.1311.

### Thiopeptide 3b

To a solution of **15b** (24.7 mg, 21.0 μmol) in CH<sub>2</sub>Cl<sub>2</sub> (3 mL) was added TFA (0.6 mL) and the mixture was stirred for 3 h at room temperature. One drop of 70% HF·Py complex was added and the mixture was stirred for another 1 h at room temperature. The solvent was removed under reduced pressure and TFA was co-evaporated with toluene (2 × 1 mL). The residue was dried in vacuo and dissolved in THF (0.6 mL) and MeOH (0.6 mL). TMSCHN<sub>2</sub> (13.7 μL, 27.3 μmol, 2.0 M in Et<sub>2</sub>O) was added and the solution was stirred for 1 h at room

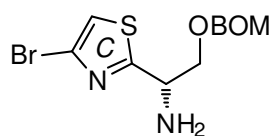
temperature. The solvents were removed in vacuo and the crude product was purified by flash chromatography (CH<sub>2</sub>Cl<sub>2</sub>/MeOH = 10:1) to yield thiopeptide **3b** (19.8 mg, 19.4 μmol, 92% over 2 steps) as a colorless solid. *R*<sub>f</sub> = 0.35 (CH<sub>2</sub>Cl<sub>2</sub>/MeOH = 10:1, UV); [α]<sub>D</sub><sup>20</sup> = +33.0 (*c* = 0.10 in CHCl<sub>3</sub>); IR (ATR):  $\tilde{\nu}$  [cm<sup>-1</sup>] = 2956 (w), 2924 (m), 2850 (w), 1654 (m, C=O), 1542 (m), 766 (w), 667 (w); <sup>1</sup>H-NMR (CDCl<sub>3</sub>, 500 MHz): δ(ppm) = 0.93 [d, <sup>3</sup>*J* = 6.5 Hz, 3 H, CH(CH<sub>3</sub>)<sub>2</sub>], 1.00 [d, <sup>3</sup>*J* = 6.7 Hz, 3 H, CH(CH<sub>3</sub>)<sub>2</sub>], 2.19 [m<sub>c</sub>, 1 H, CH(CH<sub>3</sub>)<sub>2</sub>], 2.40 [br. s, 2 H,



CH<sub>2</sub>C(=O)NHMe], 2.62 (br. s, 3 H, NHCH<sub>3</sub>), 2.68 (s, 3 H, E: 5-CH<sub>3</sub>), 4.01 (s, 3 H, OCH<sub>3</sub>), 4.12 (br. s, 2 H, CH<sub>2</sub>OH), 4.28 [d, <sup>2</sup>*J* = 17.1 Hz, 1 H, C(=O)CHHNH], 4.63 [d, <sup>2</sup>*J* = 17.1 Hz, 1 H, C(=O)CHHNH], 5.28-5.34 (m, 2 H, C: CHNH and D: CHNH), 5.47 (s, 1 H, E: CHNH), 6.76 (br. s, 1 H, NHCH<sub>3</sub>), 7.34 (s, 1 H, C: 5-H), 7.92 (s, 1 H, B: 5-H), 7.99-8.10 [m, 2 H, C: CHNH and C(=O)CH<sub>2</sub>NH], 8.06 (d, <sup>3</sup>*J* = 7.5 Hz, 1 H, A: 4-H), 8.11 (s, 1 H, D: 5-H), 8.24-8.35 (m, 4 H, D: CHNH

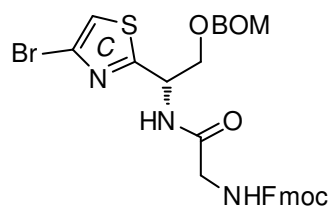
and E: CHNH and F: 5-H and G: 5-H), 8.37 (d, <sup>3</sup>*J* = 7.5 Hz, 1 H, A: 5-H); <sup>13</sup>C-NMR (CDCl<sub>3</sub>, 126 MHz): δ(ppm) = 12.6 (q, E: 5-CH<sub>3</sub>), 18.5 [q, CH(CH<sub>3</sub>)<sub>2</sub>], 19.0 [q, CH(CH<sub>3</sub>)<sub>2</sub>], 26.3 (q, NHCH<sub>3</sub>), 35.0 [d, CH(CH<sub>3</sub>)<sub>2</sub>], 38.3 [t, CH<sub>2</sub>C(=O)NHMe], 42.8 [t, C(=O)CH<sub>2</sub>NH], 49.1 (d, E: CHNH), 52.7 (q, OCH<sub>3</sub>), 54.7 (d, C: CHNH), 55.0 (d, D: CHNH), 64.2 (t, CH<sub>2</sub>OH), 115.8 (d, C: C5), 118.9 (d, A: C5), 121.7 (d, B: C5), 123.8 (d, D: C5), 125.4 (d, F: C5), 128.2 (s, A: C3), 130.6 (d, G: C5), 139.7 (d, A: C4), 141.3 (s, E: C5), 142.2 (s, E: C4), 148.2 (s, G: C4), 148.6 (s, A: C2 and D: C4), 148.8 (s, C: C4), 150.0 (s, F: C4), 150.7 (s, A: C6), 154.0 (s, B: C4), 160.3 (s, B: C2), 160.7 [s, D: C(=O)NH], 161.4 [s, E: C(=O)NH], 161.7 [s, C(=O)OMe], 161.8 [s, F: C(=O)NH], 165.1 (s, F: C2), 166.0 (s, E: C2), 168.7 (s, D: C2), 169.2 (s, G: C2), 169.8 [s, C(=O)CH<sub>2</sub>NH], 170.3 [s, C(=O)NHMe], 170.4 (s, C: C2); MS (ESI): *m/z* (%) = 1041 (45) [(M+Na)<sup>+</sup>], 1019 (100) [(M+H)<sup>+</sup>]; HRMS (ESI): *m/z* = C<sub>41</sub>H<sub>39</sub>N<sub>12</sub>O<sub>8</sub>S<sub>6</sub> [(M+H)<sup>+</sup>], calcd.: 1019.1333, found: 1019.1332.

### Compound 9c



Aminoalcohol **8a** (0.56 g, 2.51 mmol) was dissolved in  $\text{CH}_2\text{Cl}_2$  (30 mL) and cooled to 0 °C.  $\text{NEt}_3$  (0.42 mL, 0.30 g, 3.01 mmol) and  $\text{Boc}_2\text{O}$  (0.60 g, 2.76 mmol) were added and the reaction was allowed to warm up to room temperature overnight. The mixture was washed with 1 N  $\text{KHSO}_4$  (1 × 30 mL) and brine (1 × 30 mL). The organic layer was dried over  $\text{Na}_2\text{SO}_4$  and solids were removed by filtration. The solvent was removed under reduced pressure and the residue was purified by flash chromatography (P/EtOAc = 4:1). The obtained *N*-Boc-protected aminoalcohol (0.57 g, 1.76 mmol) was dissolved in THF (4 mL). At room temperature  $i\text{Pr}_2\text{NEt}$  (0.90 mL, 0.68 g, 5.28 mmol), BOMCl (0.44 mL, 0.47 g, 3.17 mmol) and TBAI (196 mg, 0.53 mmol) were added and the mixture was stirred for 16 h. The reaction was quenched by the addition of MeOH (2 mL) and stirred for another 1 h. It was then diluted by the addition of EtOAc (50 mL) and washed sequentially with 1 N  $\text{KHSO}_4$  (1 × 50 mL), sat.  $\text{NaHCO}_3$  (1 × 50 mL) and brine (1 × 50 mL). The organic layer was dried over  $\text{Na}_2\text{SO}_4$ , filtered and concentrated in vacuo. The crude product was purified by flash chromatography (P/EtOAc = 4:1). The double protected aminoalcohol (0.59 g, 1.33 mmol) was dissolved in  $\text{CH}_2\text{Cl}_2$  (8 mL) and TFA (2 mL) was added. After stirring for 5 h at room temperature the reaction was quenched by the addition of sat.  $\text{NaHCO}_3$  (10 mL). The layers were separated and the aqueous phase was extracted with  $\text{CH}_2\text{Cl}_2$  (2 × 10 mL). The combined extracts were dried over  $\text{Na}_2\text{SO}_4$ , filtered and concentrated under reduced pressure. The residue was purified by flash chromatography to yield the *O*-BOM-protected aminoalcohol **9c** (196 mg, 0.57 mmol, 23% over 3 steps) as a yellow oil.  $R_f$  = 0.20 (P/EtOAc = 2:1, UV);  $^1\text{H-NMR}$  ( $\text{CDCl}_3$ , 360 MHz):  $\delta$ (ppm) = 2.26 (br. s, 2 H,  $\text{NH}_2$ ), 3.83 (dd,  $^2J$  = 9.8 Hz,  $^3J$  = 6.7 Hz, 1 H,  $\text{CHHOBOM}$ ), 3.99 (dd,  $^2J$  = 9.8 Hz,  $^3J$  = 4.0 Hz, 1 H,  $\text{CHHOBOM}$ ), 4.49 (dd,  $^3J$  = 6.7, 4.0 Hz, 1 H,  $\text{CHNH}_2$ ), 4.57

(s, 2 H, OCH<sub>2</sub>Ph), 4.78 (d, <sup>2</sup>J = 6.7 Hz, 1 H, OCHHO), 4.80 (d, <sup>2</sup>J = 6.7 Hz, 1 H, OCHHO), 7.18 (s, 1 H, C: 5-H), 7.27-7.39 (m, 5 H, Ph).

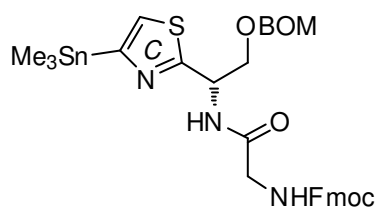


### Compound 10c

A solution of amine **9c** (185 mg, 0.54 mmol) in CH<sub>2</sub>Cl<sub>2</sub> (10 mL) was cooled to 0 °C and *i*Pr<sub>2</sub>NEt (0.28 mL, 209 mg, 1.62 mmol), Fmoc-Gly-OH (193 mg, 0.65 mmol) and PyBrOP (303 mg, 0.65 mmol) were sequentially added. After stirring for 4 h at this temperature the reaction was quenched by the addition of sat. NaHCO<sub>3</sub> (10 mL). The layers were separated, the aqueous layer was extracted with CH<sub>2</sub>Cl<sub>2</sub> (2 × 10 mL) and the combined extracts were dried over Na<sub>2</sub>SO<sub>4</sub>. Filtration and evaporation of the solvent yielded the crude product that was purified by flash chromatography (P/EtOAc = 1:1). The title compound **10c** (317 mg, 0.51 mmol, 94%) was obtained as a colorless foam. *R*<sub>f</sub> = 0.40 (P/EtOAc = 1:1, UV); [α]<sub>D</sub><sup>20</sup> = -8.7 (*c* = 1.00 in CHCl<sub>3</sub>); IR (ATR):  $\tilde{\nu}$  [cm<sup>-1</sup>] = 3304 (w), 1791 (m), 1677 (s, C=O), 1523 (m), 1248 (m), 1046 (w), 737 (m); <sup>1</sup>H-NMR (CDCl<sub>3</sub>, 360 MHz): δ(ppm) = 3.83-3.94 [m, 3 H, CHHOBOM and C(=O)CH<sub>2</sub>NH], 4.17-4.25 (m, 2 H, Fmoc: CHCH<sub>2</sub>O and CHHOBOM), 4.42 (d, <sup>3</sup>J = 6.9 Hz, 2 H, Fmoc: CHCH<sub>2</sub>O), 4.45 (s, 2 H, OCH<sub>2</sub>Ph), 4.70 (m<sub>c</sub>, 2 H, OCH<sub>2</sub>O), 5.30 [br. s, 1 H C(=O)CH<sub>2</sub>NH], 5.50 (dd, <sup>3</sup>J = 8.3, 4.2 Hz, 1 H, CHNH), 7.14 (s, 1 H, C: 5-H), 7.17 (m<sub>c</sub>, 1 H, CHNH), 7.24-7.34 (m, 7 H, Ph and Fmoc: 2-H and 7-H), 7.40 (virt. t, <sup>3</sup>J ≅ 7.4 Hz, 2 H, Fmoc: 3-H and 6-H), 7.57 (d, <sup>3</sup>J = 7.3 Hz, 2 H, Fmoc: 1-H and 8-H), 7.77 (d, <sup>3</sup>J = 7.5 Hz, 2 H, Fmoc: 4-H and 5-H); <sup>13</sup>C-NMR (CDCl<sub>3</sub>, 91 MHz): δ(ppm) = 44.5 (d, Fmoc: CHCH<sub>2</sub>O), 47.0 [t, C(=O)CH<sub>2</sub>NH], 51.4 (d, CHNH), 67.3 (t, Fmoc: CHCH<sub>2</sub>O), 69.4 (t, CH<sub>2</sub>OBOM), 69.8 (t, OCH<sub>2</sub>Ph), 94.9 (t, OCH<sub>2</sub>O), 117.1 (d, C: C5), 120.0 (d, Fmoc: C4 and C5), 124.8 (s, C: C4), 125.0 (d, Fmoc: C1 and C8), 127.1 (d, Fmoc: C2 and C7), 127.7 (d, Ph: C4), 127.9 (d, Ph: C2 and C6 and Fmoc: C3 and C6), 128.5 (d, Ph: C3 and C5), 137.2 (s, Ph: C1), 141.3 (s, Fmoc: C4a and C4b), 143.6 (s, Fmoc: C8a and C8b), 156.5 [s, CH<sub>2</sub>OC(=O)NH], 168.9 [s, C(=O)CH<sub>2</sub>NH], 170.3 (s, C: C2); MS (ESI): *m/z* (%) = 646 (27) [(M(<sup>81</sup>Br)+Na)<sup>+</sup>], 644 (27)



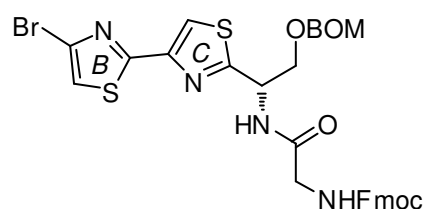
$[(M(^{79}\text{Br})+\text{Na})^+]$ , 624 (100)  $[(M(^{81}\text{Br})+\text{H})^+]$ , 622 (13)  $[(M(^{79}\text{Br})+\text{H})^+]$ , 445 (2), 443 (2); HRMS (ESI):  $m/z = \text{C}_{30}\text{H}_{29}^{79}\text{BrN}_3\text{O}_5\text{S} [(M(^{79}\text{Br})+\text{H})^+]$ , calcd.: 622.1006, found: 622.1000.



### Compound 11c

To a solution of bromothiazole **10c** (288 mg, 0.46 mmol) in degassed toluene (6 mL) were added  $\text{Pd}(\text{PPh}_3)_4$  (53.5 mg, 46.3  $\mu\text{mol}$ ) and hexamethylditin (0.38 mL, 0.60 g, 1.84 mmol).

After stirring for 1.5 h at 100 °C the mixture was concentrated in vacuo. The residue was purified by flash chromatography (P/EtOAc = 2:1) to yield stannane **11c** (224 mg, 0.32 mmol, 69%) as a colorless foam.  $R_f = 0.40$  (P/EtOAc = 1:1, UV);  $^1\text{H-NMR}$  ( $\text{CDCl}_3$ , 500 MHz):  $\delta$ (ppm) = 0.33 [s, 9 H,  $\text{Sn}(\text{CH}_3)_3$ ], 3.87-3.97 (m, 3 H,  $\text{CHHOBOM}$  and  $\text{C}(=\text{O})\text{CH}_2\text{NH}$ ), 4.19 (dd,  $^2J = 10.1$  Hz,  $^3J = 4.2$  Hz, 1 H,  $\text{CHHOBOM}$ ), 4.23 (t,  $^3J = 7.4$  Hz, 1 H, Fmoc:  $\text{CHCH}_2\text{O}$ ), 4.41 (d,  $^3J = 7.4$  Hz, 2 H, Fmoc:  $\text{CHCH}_2\text{O}$ ), 4.45 (s, 2 H,  $\text{OCH}_2\text{Ph}$ ), 4.70 (d,  $^2J = 6.8$  Hz, 1 H,  $\text{OCHHO}$ ), 4.72 (d,  $^2J = 6.8$  Hz, 1 H,  $\text{OCHHO}$ ), 5.39 [br. s, 1 H,  $\text{C}(=\text{O})\text{CH}_2\text{NH}$ ], 5.59 (m, 1 H,  $\text{CHNH}$ ), 7.12 (br. d,  $^3J = 6.8$  Hz, 1 H,  $\text{CHNH}$ ), 7.28 (s, 1 H, C: 5-H), 7.28-7.35 (m, 7 H, Ph and Fmoc: 2-H and 7-H), 7.40 (virt. t,  $^3J \cong 7.4$  Hz, 2 H, Fmoc: 3-H and 6-H), 7.58 (d,  $^3J = 7.3$  Hz, 2 H, Fmoc: 1-H and 8-H), 7.76 (d,  $^3J = 7.5$  Hz, 2 H, Fmoc: 4-H and 5-H). After  $^1\text{H-NMR}$  stannane **11c** was quickly subjected to the next step.

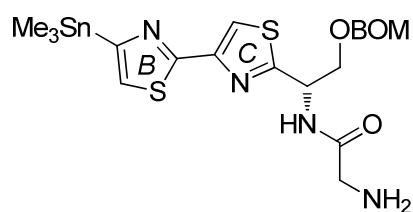


### Compound 12c

To solution of stannane **11c** (210 mg, 0.30 mmol) in degassed toluene (5 mL) were added  $\text{Pd}(\text{PPh}_3)_4$  (34.3 mg, 29.7  $\mu\text{mol}$ ) and 2,4-dibromothiazole (102 mg, 0.42 mmol).

After stirring for 16 h at 90 °C the reaction was quenched by the addition of sat.  $\text{NaHCO}_3$  (5 mL). The mixture was extracted with EtOAc (3  $\times$  20 mL) and the combined organic layers were dried over  $\text{Na}_2\text{SO}_4$ . After filtration, the solvents were removed under reduced pressure. Flash chromatography (P/EtOAc = 1:1) yielded bithiazole **12c** (207 mg, 0.29 mmol, 98%) as

a colorless foam.  $R_f = 0.40$  (P/EtOAc = 1:1, UV);  $[\alpha]_D^{20} = -13.7$  ( $c = 1.00$  in  $\text{CHCl}_3$ ); IR (ATR):  $\tilde{\nu}$  [ $\text{cm}^{-1}$ ] = 3302 (w), 2923 (w), 1669 (s, C=O), 1516 (m), 1449 (m), 1253 (m), 1167 (w), 1042 (s), 736 (s), 698 (w);  $^1\text{H-NMR}$  ( $\text{CDCl}_3$ , 360 MHz):  $\delta$ (ppm) = 3.85-3.98 (m, 3 H,  $\text{CHHOBOM}$  and  $\text{C(=O)CH}_2\text{NH}$ ), 4.21 (t,  $^3J = 7.0$  Hz, 1 H, Fmoc:  $\text{CHCH}_2\text{O}$ ), 4.28 (dd,  $^2J = 10.2$  Hz,  $^3J = 3.9$  Hz, 1 H,  $\text{CHHOBOM}$ ), 4.40-4.44 (m, 2 H, Fmoc:  $\text{CHCH}_2\text{O}$ ), 4.45 (s, 2 H,  $\text{OCH}_2\text{Ph}$ ), 4.70 (d,  $^2J = 6.8$  Hz, 1 H,  $\text{OCHHO}$ ), 4.73 (d,  $^2J = 6.8$  Hz, 1 H,  $\text{OCHHO}$ ), 5.45 [br. s, 1 H,  $\text{C(=O)CH}_2\text{NH}$ ], 5.54 (m, 1 H,  $\text{CHNH}$ ), 7.16 (s, 1 H, B: 5-H), 7.20-7.31 (m, 8 H, Ph and  $\text{CHNH}$  and Fmoc: 2-H and 7-H), 7.39 (virt. t,  $^3J \cong 7.5$  Hz, 2 H, Fmoc: 3-H and 6-H), 7.56 (d,  $^3J = 7.5$  Hz, 2 H, Fmoc: 1-H and 8-H), 7.77 (d,  $^3J = 7.5$  Hz, 2 H, Fmoc: 4-H and 5-H), 7.92 (s, 1 H, C: 5-H);  $^{13}\text{C-NMR}$  ( $\text{CDCl}_3$ , 91 MHz):  $\delta$ (ppm) = 44.7 (d, Fmoc:  $\text{CHCH}_2\text{O}$ ), 47.1 [t,  $\text{C(=O)CH}_2\text{NH}$ ], 51.5 (d,  $\text{CHNH}$ ), 67.4 (t, Fmoc:  $\text{CHCH}_2\text{O}$ ), 69.4 (t,  $\text{CH}_2\text{OBOM}$ ), 69.8 (t,  $\text{OCH}_2\text{Ph}$ ), 95.0 (t,  $\text{OCH}_2\text{O}$ ), 117.5 (d, C: C5), 117.4 (d, B: C5), 120.0 (d, Fmoc: C4 and C5), 125.0 (d, Fmoc: C1 and C8), 126.0 (s, B: C4), 127.1 (d, Fmoc: C2 and C7), 127.8 (d, Ph: C2 and C6, Fmoc: C3 and C6), 127.9 (d, Ph: C4), 128.5 (d, Ph: C3 and C5), 137.2 (s, Ph: C1), 141.3 (s, Fmoc: C4a and C4b), 143.6 (s, Fmoc: C8a and C8b), 148.1 (s, C: C4), 156.6 [s,  $\text{CH}_2\text{OC(=O)NH}$ ], 163.1 (s, B: C2), 168.9 [s,  $\text{C(=O)CH}_2\text{NH}$ ], 170.2 (s, C: C2); MS (ESI):  $m/z$  (%) = 745 (4)  $[(\text{M}^{81}\text{Br})+\text{K}]^+$ , 743 (4)  $[(\text{M}^{79}\text{Br})+\text{K}]^+$ , 729 (80)  $[(\text{M}^{81}\text{Br})+\text{Na}]^+$ , 727 (80)  $[(\text{M}^{79}\text{Br})+\text{Na}]^+$ , 707 (100)  $[(\text{M}^{81}\text{Br})+\text{H}]^+$ , 705 (100)  $[(\text{M}^{79}\text{Br})+\text{H}]^+$ ; HRMS (ESI):  $m/z = \text{C}_{33}\text{H}_{29}^{79}\text{BrN}_4\text{O}_5\text{S}_2\text{Na}$   $[(\text{M}^{79}\text{Br})+\text{Na}]^+$ , calcd.: 727.0660, found: 727.0660.



### Compound 5c

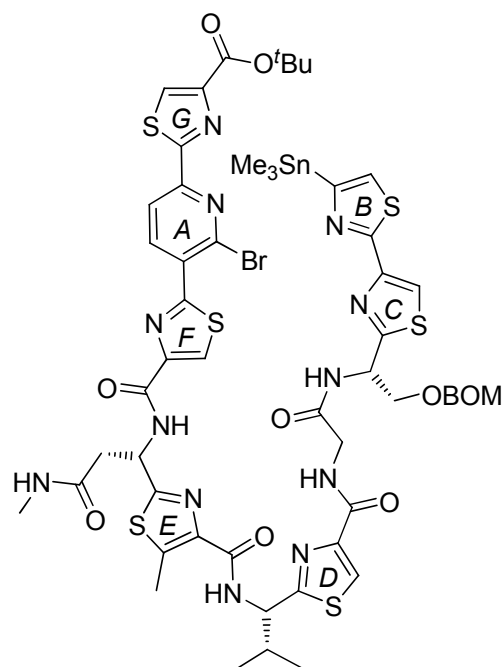
Compound **5c** was prepared as described above for **5a**, starting from bromide **12c**. It was obtained as a pale-yellow oil (84.3  $\mu\text{mol}$ , 73%).  $R_f = 0.25$  ( $\text{CH}_2\text{Cl}_2/\text{MeOH} = 10:1$ , UV);

$[\alpha]_D^{20} = -27.5$  ( $c = 0.45$  in  $\text{CHCl}_3$ ); IR (ATR):  $\tilde{\nu}$  [ $\text{cm}^{-1}$ ] = 2950 (w), 1672 (s, C=O), 1504 (s), 1262 (m), 1046 (s), 1026 (m), 770 (s);  $^1\text{H-NMR}$  ( $\text{CDCl}_3$ , 360 MHz):  $\delta$ (ppm) = 0.38 [s, 9 H,  $\text{Sn}(\text{CH}_3)_3$ ], 1.53 (br. s, 2 H,  $\text{NH}_2$ ), 3.45 (m, 2 H,  $\text{CH}_2\text{NH}_2$ ), 3.98 (dd,  $^2J = 9.9$  Hz,  $^3J = 4.6$  Hz,

1 H, *CHHOBOM*), 4.31 (dd,  $^2J = 9.9$  Hz,  $^3J = 4.1$  Hz, 1 H, *CHHOBOM*), 4.51 (s, 2 H, *OCH<sub>2</sub>Ph*), 4.77 (s, 2 H, *OCH<sub>2</sub>O*), 5.60 (m<sub>c</sub>, 1 H, *CHNH*), 7.26-7.31 (m, 5 H, Ph), 7.37 (s, 1 H, B: 5-H), 7.93 (s, 1 H, C: 5-H), 8.25 (d,  $^3J = 8.7$  Hz, 1 H, *CHNH*);  $^{13}\text{C-NMR}$  ( $\text{CDCl}_3$ , 91 MHz):  $\delta(\text{ppm}) = -8.8$  [q,  $\text{Sn}(\text{CH}_3)_3$ ], 44.9 (t,  $\text{CH}_2\text{NH}_2$ ), 51.0 (d, *CHNH*), 69.4 (t,  $\text{CH}_2\text{OBOM}$ ), 69.5 (t, *OCH<sub>2</sub>Ph*), 94.7 (t, *OCH<sub>2</sub>O*), 115.8 (d, C: C5), 126.2 (d, B: C5), 127.7 (d, Ph: C4), 127.9 (d, Ph: C2 and C6), 128.4 (d, Ph: C3 and C5), 137.4 (s, Ph: C1), 149.9 (s, C: C4), 161.1 (s, B: C4), 163.3 (s, B: C2), 170.1 (s, C: C2), 172.7 (s, C=O); MS (ESI):  $m/z$  (%) = 569 (100) [ $(\text{M}^{120}\text{Sn}+\text{H})^+$ ], 539 (17), 449 (23); HRMS (ESI):  $m/z = \text{C}_{21}\text{H}_{29}\text{N}_4\text{O}_3\text{S}_2^{120}\text{Sn}$  [ $(\text{M}^{120}\text{Sn}+\text{H})^+$ ], calcd.: 569.0698, found: 569.0699.

### Compound 14c

This compound was prepared as described above starting from compounds **13** and **5c** to yield compound **14c** (29.3 mg, 20.5  $\mu\text{mol}$ , 54%) as a pale-yellow oil.  $R_f = 0.40$  (EtOAc, UV),  $^1\text{H-NMR}$  ( $\text{CDCl}_3$ , 360 MHz):  $\delta(\text{ppm}) = 0.36$  [s, 9 H,  $\text{Sn}(\text{CH}_3)_3$ ], 0.96 [d,  $^3J = 6.5$  Hz, 3 H,



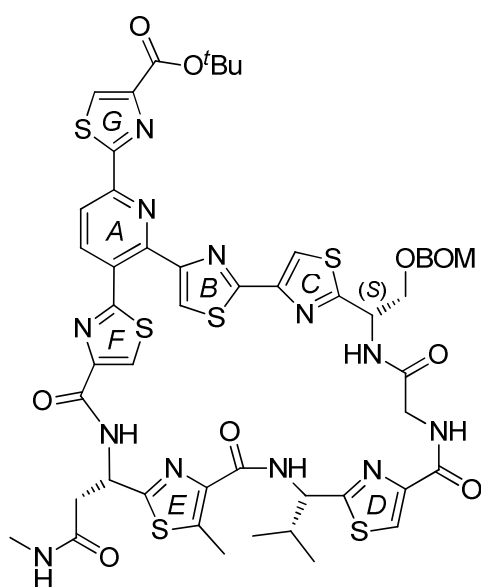
CH( $\text{CH}_3$ )<sub>2</sub>], 0.99 [d,  $^3J = 6.7$  Hz, 3 H, CH( $\text{CH}_3$ )<sub>2</sub>], 1.63 [s, 9 H, OC( $\text{CH}_3$ )<sub>3</sub>], 2.33 [m<sub>c</sub>, 1 H, CH( $\text{CH}_3$ )<sub>2</sub>], 2.65 (d,  $^3J = 4.8$  Hz, 3 H, NHCH<sub>3</sub>), 2.73 (s, 3 H, E: 5-CH<sub>3</sub>), 2.95 [dd,  $^2J = 15.7$  Hz,  $^3J = 5.2$  Hz, 1 H, CHHC(=O)NHMe], 3.20 [dd,  $^2J = 15.7$  Hz,  $^3J = 4.5$  Hz, 1 H, CHHC(=O)NHMe], 3.93 (dd,  $^2J = 10.1$  Hz,  $^3J = 4.5$  Hz, 1 H, *CHHOBOM*), 4.11 [dd,  $^2J = 16.9$  Hz,  $^3J = 4.5$  Hz, 1 H, C(=O)CHNH], 4.22 (dd,  $^2J = 10.1$  Hz,  $^3J = 4.2$  Hz, 1 H, *CHHOBOM*), 4.46 (s, 2 H, *OCH<sub>2</sub>Ph*), 4.48 [dd,  $^2J = 16.9$  Hz,  $^3J = 6.5$  Hz, 1 H, C(=O)CHNH], 4.71 (s, 2 H, *OCH<sub>2</sub>O*), 5.31 (m<sub>c</sub>,

1 H, D: *CHNH*), 5.52 (m<sub>c</sub>, 1 H, C: *CHNH*), 5.74 (m<sub>c</sub>, 1 H, E: *CHNH*), 6.72 (d,  $^3J = 4.8$  Hz, 1 H, NHCH<sub>3</sub>), 7.20-7.30 (m, 5 H, Ph), 7.34 (s, 1 H, C: 5-H), 7.62 (m<sub>c</sub>,  $^3J = 1$  H, NH), 7.93 (s, 1 H,

B: 5-H), 8.05 (s, 1 H, D: 5-H), 8.11 (m<sub>c</sub>, 1 H, NH), 8.19-8.22 (m, 1 H, NH), 8.20 (s, 1 H, G: 5-H), 8.33 (s, 1 H, F: 5-H), 8.36 (d, <sup>3</sup>J = 8.4 Hz, 1 H, A: 4-H), 8.73 (d, <sup>3</sup>J = 8.4 Hz, 1 H, A: 5-H), 9.56 (m<sub>c</sub>, 1 H, NH). After <sup>1</sup>H-NMR stannane **14c** was quickly subjected to the cyclization step.

### Macrocycle **15c**

Starting from compound **14c** macrocycle **15c** (11.1 mg, 9.40 μmol, 52%) was obtained as a pale-yellow wax, using the procedure given above for compound **14a**. *R*<sub>f</sub> = 0.50



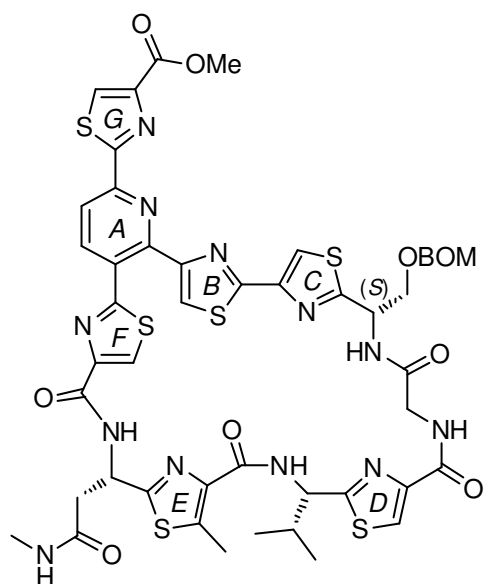
(EtOAc/MeOH = 60:1); <sup>1</sup>H-NMR (CDCl<sub>3</sub>, 360 MHz):

δ(ppm) = 0.89 [d, <sup>3</sup>J = 6.7 Hz, 3 H, CH(CH<sub>3</sub>)<sub>2</sub>], 1.00 [d, <sup>3</sup>J = 6.9 Hz, 3 H, CH(CH<sub>3</sub>)<sub>2</sub>], 1.01-1.05 [m, 1 H, CHHC(=O)NHMe], 1.66 [s, 9 H, OC(CH<sub>3</sub>)<sub>3</sub>], 2.27 [m<sub>c</sub>, 1 H, CH(CH<sub>3</sub>)<sub>2</sub>], 2.62 (d, <sup>3</sup>J = 4.7 Hz, 3 H, NHCH<sub>3</sub>), 2.65 (s, 3 H, E: 5-CH<sub>3</sub>), 2.71 [dd, <sup>2</sup>J = 16.9 Hz, <sup>3</sup>J = 3.6 Hz, 1 H, CHHC(=O)NHMe], 3.63 [dd, <sup>2</sup>J = 17.6 Hz, <sup>3</sup>J = 3.2 Hz, 1 H, C(=O)CHHNH], 3.93 (dd, <sup>2</sup>J = 10.6 Hz, <sup>3</sup>J = 3.6 Hz, 1 H, CHHOBOM), 4.01 (dd, <sup>2</sup>J = 10.6 Hz, <sup>3</sup>J =

4.1 Hz, 1 H, CHHOBOM), 4.55 (d, <sup>2</sup>J = 11.7 Hz, 1 H, OCHHPh), 4.62 (d, <sup>2</sup>J = 11.7 Hz, 1 H, OCHHPh), 4.79 (d, <sup>2</sup>J = 6.5 Hz, 1 H, OCHHO), 4.80-4.84 [m, 1 H, C(=O)CHHNH], 4.84 (d, <sup>2</sup>J = 6.5 Hz, 1 H, OCHHO), 5.23 (dd, <sup>3</sup>J = 7.9, 4.6 Hz, 1 H, D: CHNH), 5.36-5.44 (m, 2 H, C: CHNH and E: CHNH), 6.72 (d, <sup>3</sup>J = 4.7 Hz, 1 H, NHCH<sub>3</sub>), 7.20 (d, <sup>3</sup>J = 6.0 Hz, 1 H, C: CHNH), 7.26 (s, 1 H, C: 5-H)\*, 7.29-7.39 (m, 5 H, Ph), 8.10 (s, 1 H, D: 5-H), 8.11 (d, <sup>3</sup>J = 8.3 Hz, 1 H, A: 4-H), 8.21 (s, 1 H, G: 5-H), 8.25 (s, 1 H, B: 5-H), 8.37 (s, 1 H, F: 5-H), 8.39 (d, <sup>3</sup>J = 8.3 Hz, 1 H, A: 5-H), 8.77 (d, <sup>3</sup>J = 7.9 Hz, 1 H, D: CHNH), 8.98 (d, <sup>3</sup>J = 9.1 Hz, 1 H, E: CHNH). \*Signal is overlain by CHCl<sub>3</sub>. This compound was obtained as a mixture together with triphenylphosphine oxide and was directly used in the next step. The triphenylphosphine oxide multiplet overlays one NH signal.

### Thiopeptide 3c

To a solution of **15c** (8.49 mg, 7.19  $\mu\text{mol}$ ) in  $\text{CH}_2\text{Cl}_2$  (1 mL) was added TFA (0.2 mL) and the mixture was stirred for 3 h at room temperature. The solvent was removed under reduced pressure and TFA was co-evaporated with toluene ( $2 \times 1$  mL). The residue was dried in vacuo and dissolved in THF (0.2 mL) and MeOH (0.2 mL).  $\text{TMSCHN}_2$  (4.68  $\mu\text{L}$ , 9.35  $\mu\text{mol}$ , 2.0 M in  $\text{Et}_2\text{O}$ ) was added and the solution was stirred for 1 h at room temperature. The solvents were removed in vacuo and the crude product was purified by flash chromatography ( $\text{CH}_2\text{Cl}_2/\text{MeOH} = 10:1$ ) followed by reversed phase HPLC to remove residual triphenylphosphine oxide. Thiopeptide **3c** (4.91 mg, 4.31  $\mu\text{mol}$ , 60% over 2 steps) was obtained as a colorless solid.  $R_f = 0.20$  ( $\text{CH}_2\text{Cl}_2/\text{MeOH} = 20:1$ , UV);  $[\alpha]_D^{20} = +137$  ( $c = 0.09$  in

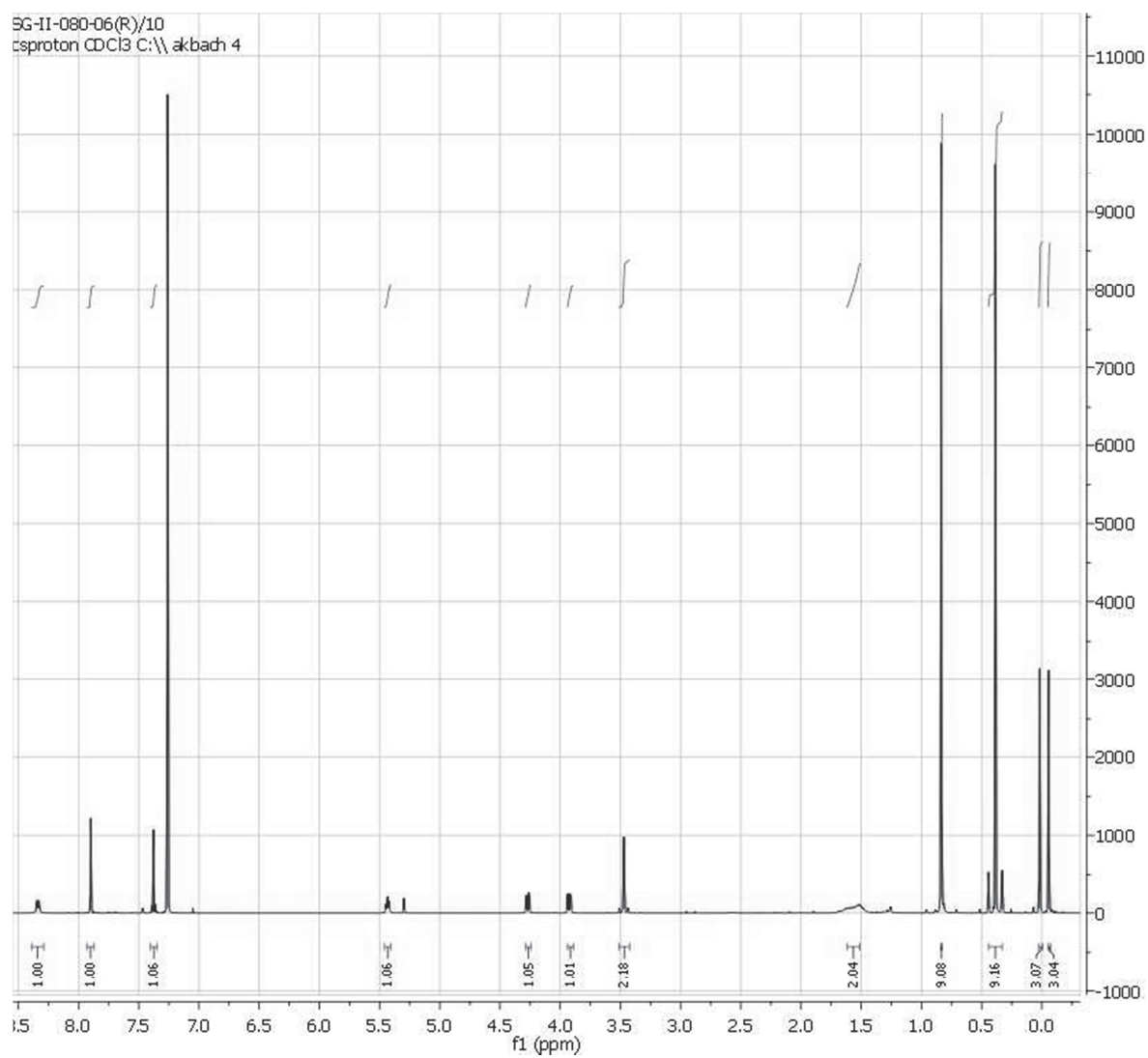
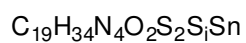
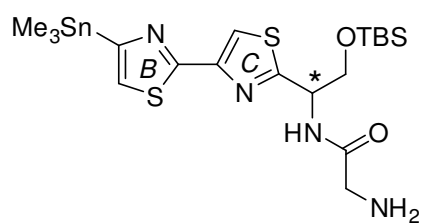


$\text{CHCl}_3$ ); IR (ATR):  $\tilde{\nu}$  [ $\text{cm}^{-1}$ ] = 2928 (w), 1658 (s, C=O), 1543 (s), 1494 (m), 1210 (m), 1051 (m), 751 (s);  $^1\text{H-NMR}$  ( $\text{CDCl}_3$ , 500 MHz):  $\delta$ (ppm) = 0.88 [d,  $^3J = 6.8$  Hz, 3 H,  $\text{CH}(\text{CH}_3)_2$ ], 0.96 [dd,  $^2J = 17.5$  Hz,  $^3J = 5.9$  Hz, 1 H,  $\text{CHHC}(=\text{O})\text{NHMe}$ ], 1.00 [d,  $^3J = 6.8$  Hz, 3 H,  $\text{CH}(\text{CH}_3)_2$ ], 2.28 [ $m_c$ , 1 H,  $\text{CH}(\text{CH}_3)_2$ ], 2.63 (d,  $^3J = 4.8$  Hz, 3 H,  $\text{NHCH}_3$ ), 2.65 (s, 3 H, E: 5- $\text{CH}_3$ ), 2.70 [dd,  $^2J = 17.5$  Hz,  $^3J = 3.5$  Hz, 1 H,  $\text{CHHC}(=\text{O})\text{NHMe}$ ], 3.59 [dd,  $^2J = 17.6$  Hz,  $^3J = 3.3$

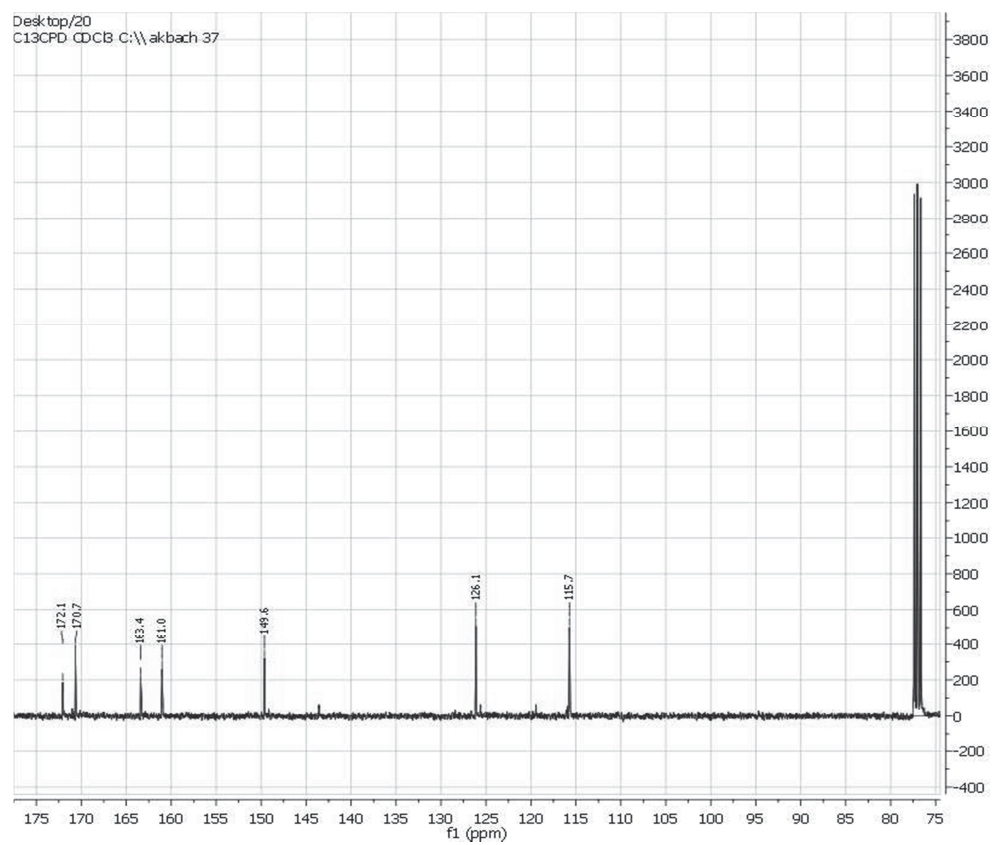
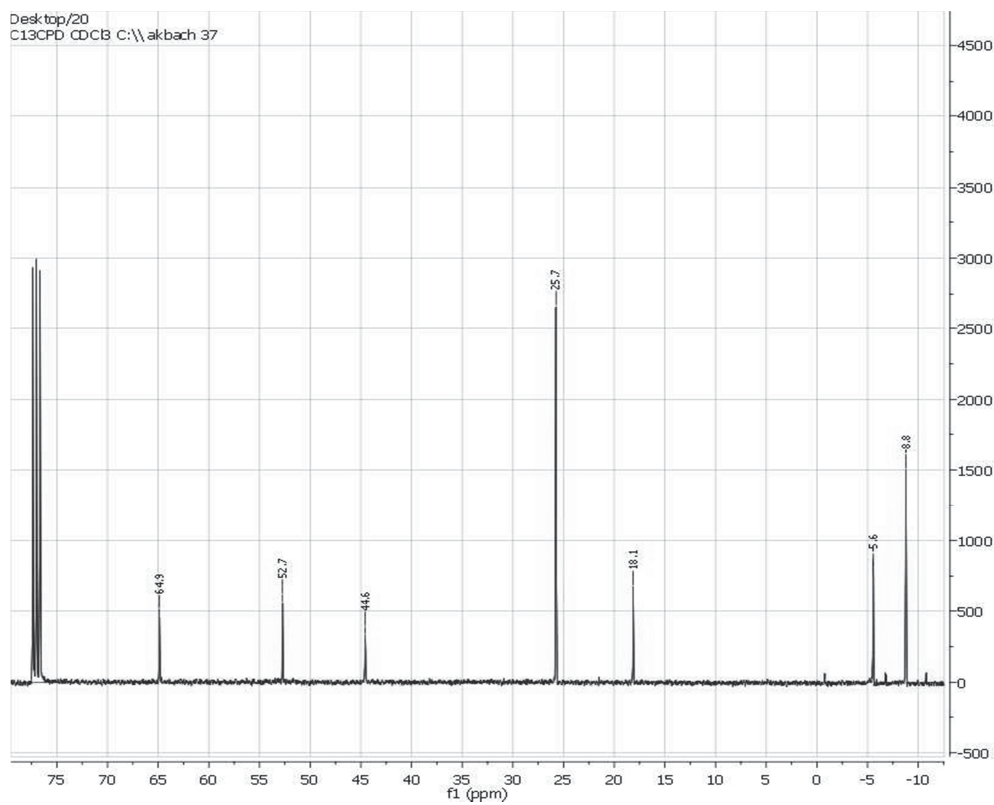
Hz, 1 H,  $\text{C}(=\text{O})\text{CHHNH}$ ], 3.94 (dd,  $^2J = 10.7$  Hz,  $^3J = 3.6$  Hz, 1 H,  $\text{CHHOBOM}$ ), 4.01 (dd,  $^2J = 10.7$  Hz,  $^3J = 4.0$  Hz, 1 H,  $\text{CHHOBOM}$ ), 4.02 (s, 3 H,  $\text{OCH}_3$ ), 4.59 (d,  $^2J = 11.7$  Hz, 1 H,  $\text{OCHHPh}$ ), 4.67 (d,  $^2J = 11.7$  Hz, 1 H,  $\text{OCHHPh}$ ), 4.79 [dd,  $^2J = 17.6$  Hz,  $^3J = 9.7$  Hz, 1 H,  $\text{C}(=\text{O})\text{CHHNH}$ ], 4.84 (d,  $^2J = 6.8$  Hz, 1 H,  $\text{OCHHO}$ ), 4.89 (d,  $^2J = 6.8$  Hz, 1 H,  $\text{OCHHO}$ ), 5.23 (dd,  $^3J = 7.9$ , 4.7 Hz, 1 H, D:  $\text{CHNH}$ ), 5.39 ( $m_c$ , 2 H, C:  $\text{CHNH}$  and E:  $\text{CHNH}$ ), 6.80 (d,  $^3J = 4.8$  Hz, 1 H,  $\text{NHCH}_3$ ), 6.91 (d,  $^3J = 6.0$  Hz, 1 H, C:  $\text{CHNH}$ ), 7.27 (s, 1 H, C: 5-H), 7.33-7.42 (m, 5 H, Ph), 7.71 [dd,  $^3J = 9.7$ , 3.3 Hz, 1 H,  $\text{C}(=\text{O})\text{CHHNH}$ ], 8.11 (s, 1 H, D: 5-H), 8.13 (d,  $^3J = 8.1$  Hz, 1 H, A: 4-H), 8.26 (s, 1 H, B: 5-H), 8.36 (s, 1 H, G: 5-H), 8.38 (s, 1 H, F: 5-H), 8.38

(d,  $^3J = 8.1$  Hz, 1 H, A: 5-H), 8.76 (d,  $^3J = 7.9$  Hz, 1 H, D: CHNH), 8.99 (d,  $^3J = 9.2$  Hz, 1 H, E: CHNH);  $^{13}\text{C}$ -NMR ( $\text{CDCl}_3$ , 126 MHz):  $\delta(\text{ppm}) = 12.3$  (q, E: 5- $\text{CH}_3$ ), 18.0 [q,  $\text{CH}(\text{CH}_3)_2$ ], 18.4 [q,  $\text{CH}(\text{CH}_3)_2$ ], 26.2 (q,  $\text{NHCH}_3$ ), 34.7 [d,  $\text{CH}(\text{CH}_3)_2$ ], 38.3 [t,  $\text{CH}_2\text{C}(=\text{O})\text{NHMe}$ ], 41.1 [t,  $\text{C}(=\text{O})\text{CH}_2\text{NH}$ ], 48.2 (d, E: CHNH), 52.6 (q,  $\text{OCH}_3$ ), 53.3 (d, C: CHNH), 56.1 (d, D: CHNH), 69.9 (t,  $\text{CH}_2\text{OBOM}$ ), 70.8 (t,  $\text{OCH}_2\text{Ph}$ ), 95.9 (t,  $\text{OCH}_2\text{O}$ ), 115.0 (d, C: C5), 118.7 (d, A: C5), 123.2 (d, B: C5), 123.8 (d, D: C5), 125.2 (d, F: C5), 127.7 (s, A: C3), 127.9 (d, Ph: C3 and C5), 128.3 (d, Ph: C4), 128.8 (d, Ph: C2 and C6), 130.5 (d, G: C5), 137.1 (s, Ph: C1), 140.4 (d, A: C4), 140.5 (s, E: C5), 142.1 (s, E: C4), 148.2 (s, G: C4), 148.4 (s, D: C4), 148.8 (s, C: C4), 150.2 (s, F: C4), 150.3 (s, A: C2), 150.5 (s, A: C6), 154.5 (s, B: C5), 159.8 (s, B: C2), 161.2 [s, D:  $\text{C}(=\text{O})\text{NH}$ ], 161.3 [s, E:  $\text{C}(=\text{O})\text{NH}$ ], 161.8 [s,  $\text{C}(=\text{O})\text{OMe}$ ], 162.0 [s, F:  $\text{C}(=\text{O})\text{NH}$ ], 164.8 (s, F: C2), 167.6 (s, E: C2), 168.5 [s,  $\text{C}(=\text{O})\text{CH}_2\text{NH}$ ], 169.0 (s, D: C2 and G: C2), 169.7 [s,  $\text{C}(=\text{O})\text{NHMe}$ ], 170.7 (s, C: C2); MS (ESI):  $m/z$  (%) = 1161 (50)  $[(\text{M}+\text{Na})^+]$ , 1139 (100)  $[(\text{M}+\text{H})^+]$ , 1107 (4)  $[(\text{M}-\text{OCH}_3)^+]$ ; HRMS (ESI):  $m/z = \text{C}_{49}\text{H}_{47}\text{N}_{12}\text{O}_9\text{S}_6$   $[(\text{M}+\text{H})^+]$ , calcd.: 1139.1907, found: 1139.1887.

# Compounds 5a and 5b, <sup>1</sup>H-NMR

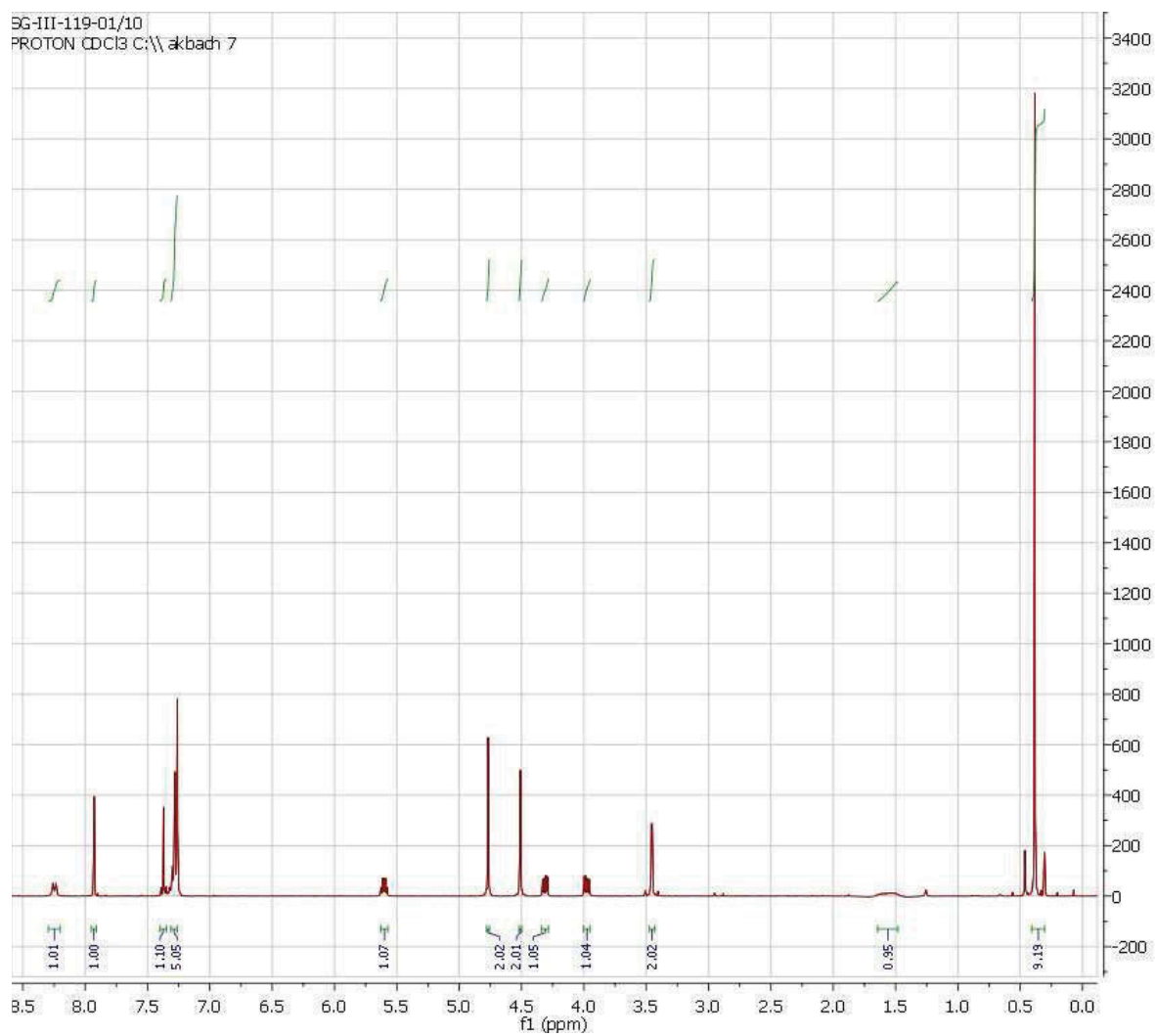
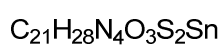
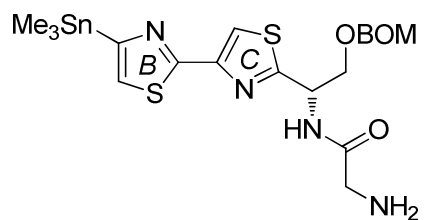


# Compounds 5a and 5b, $^{13}\text{C}$ -NMR

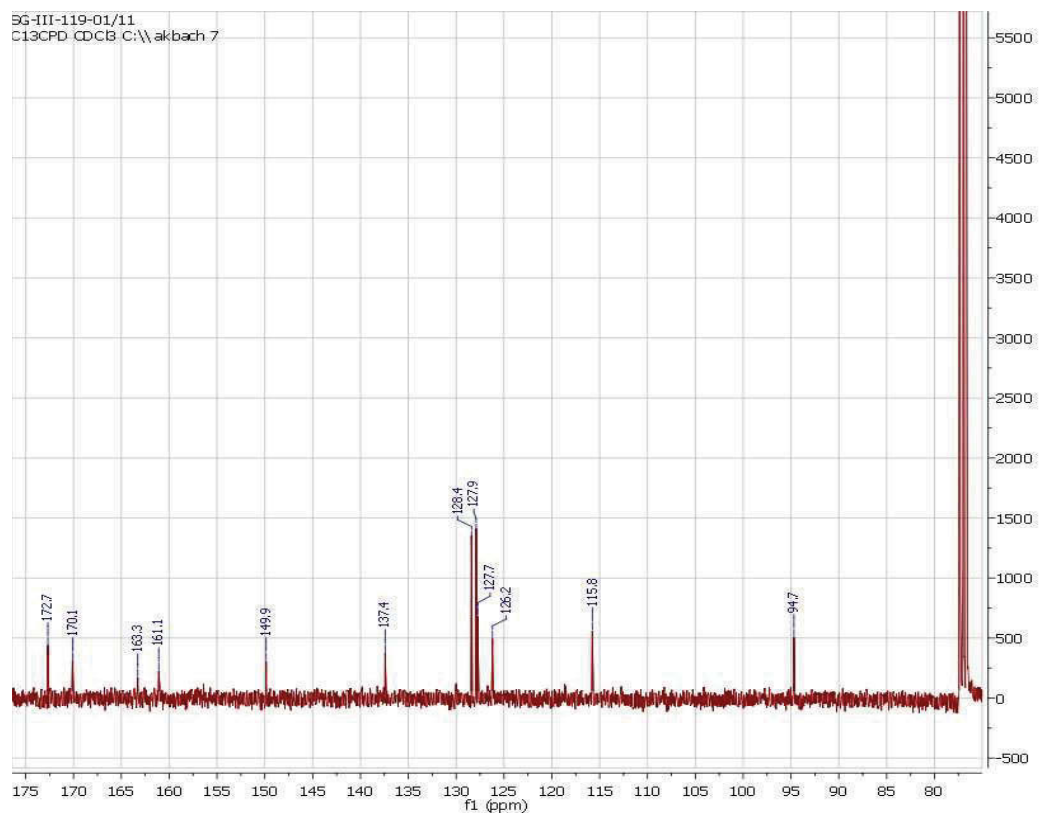
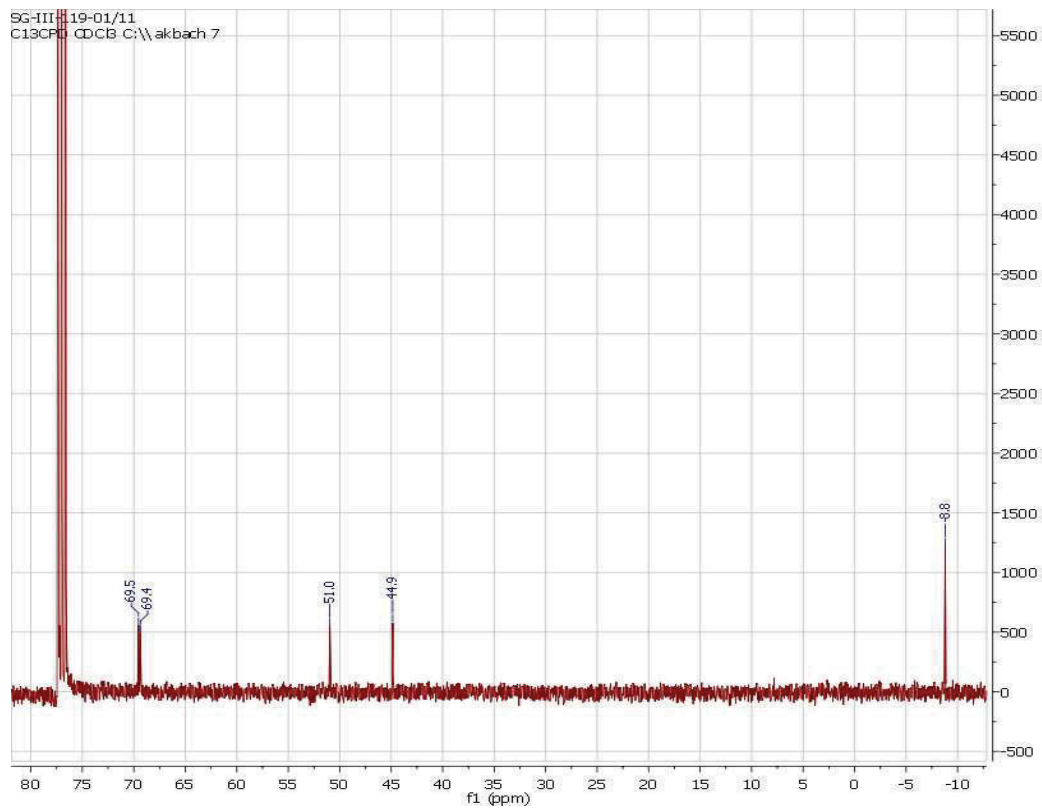




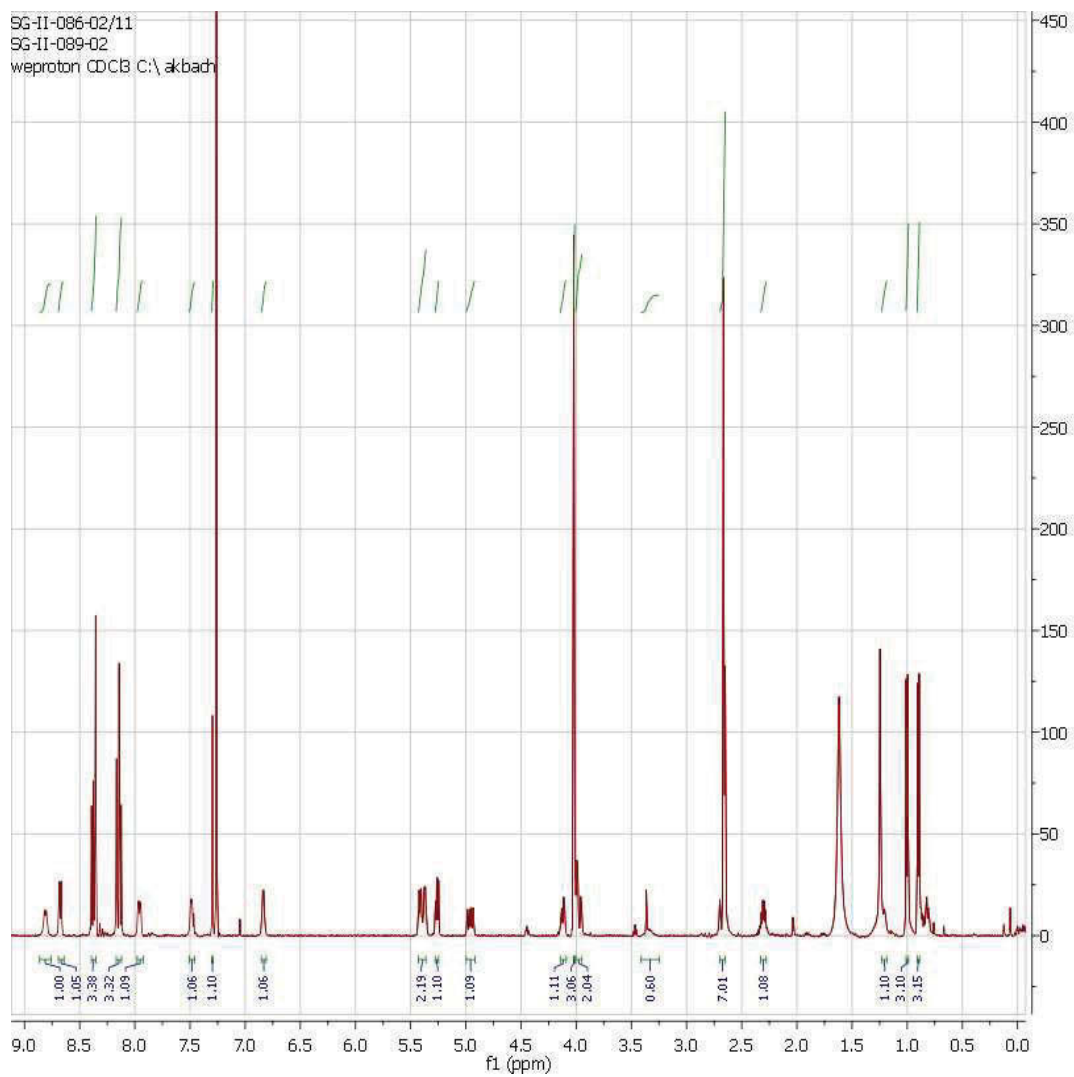
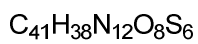
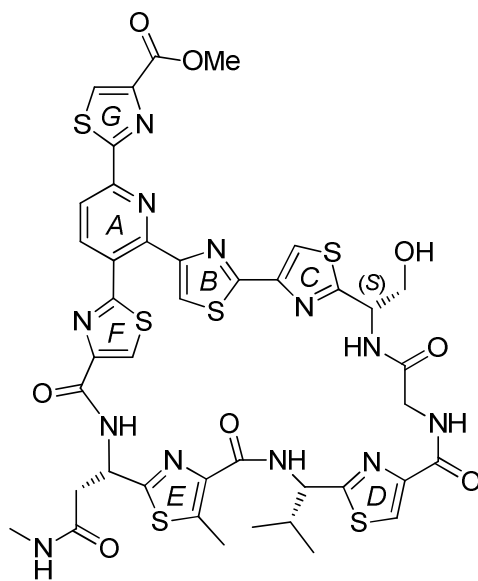
# Compound 5c, <sup>1</sup>H-NMR



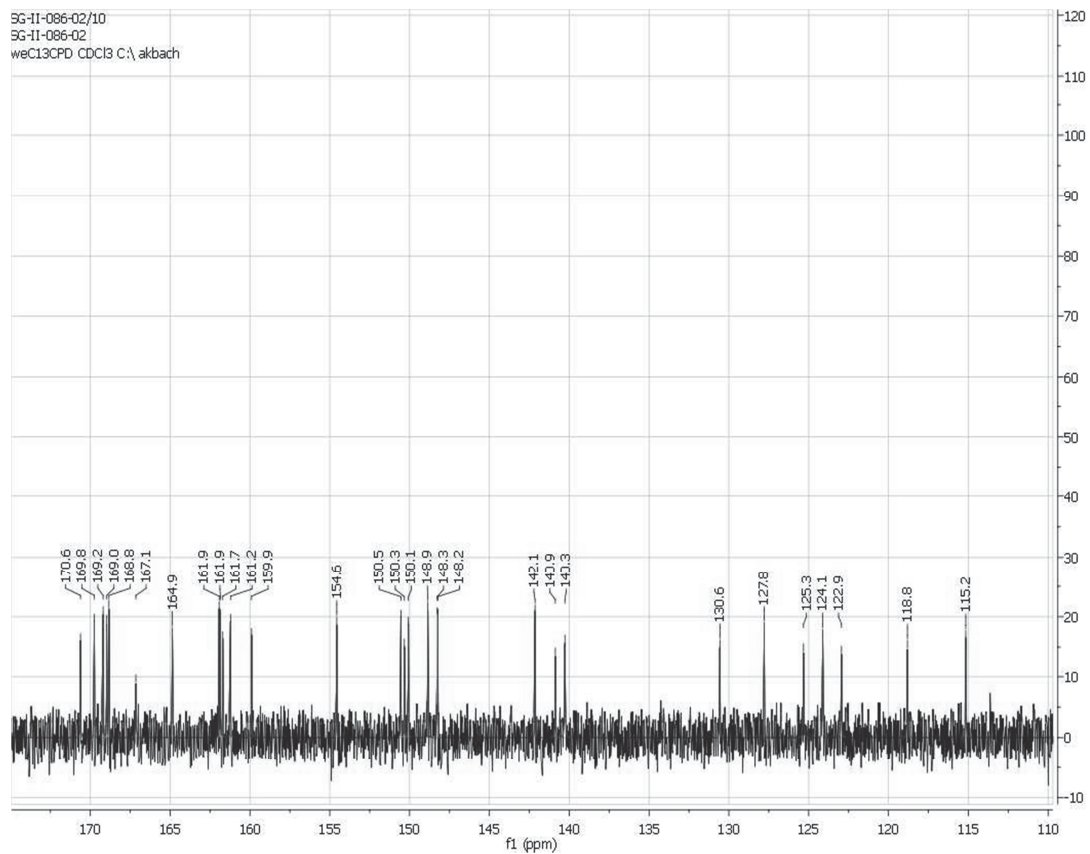
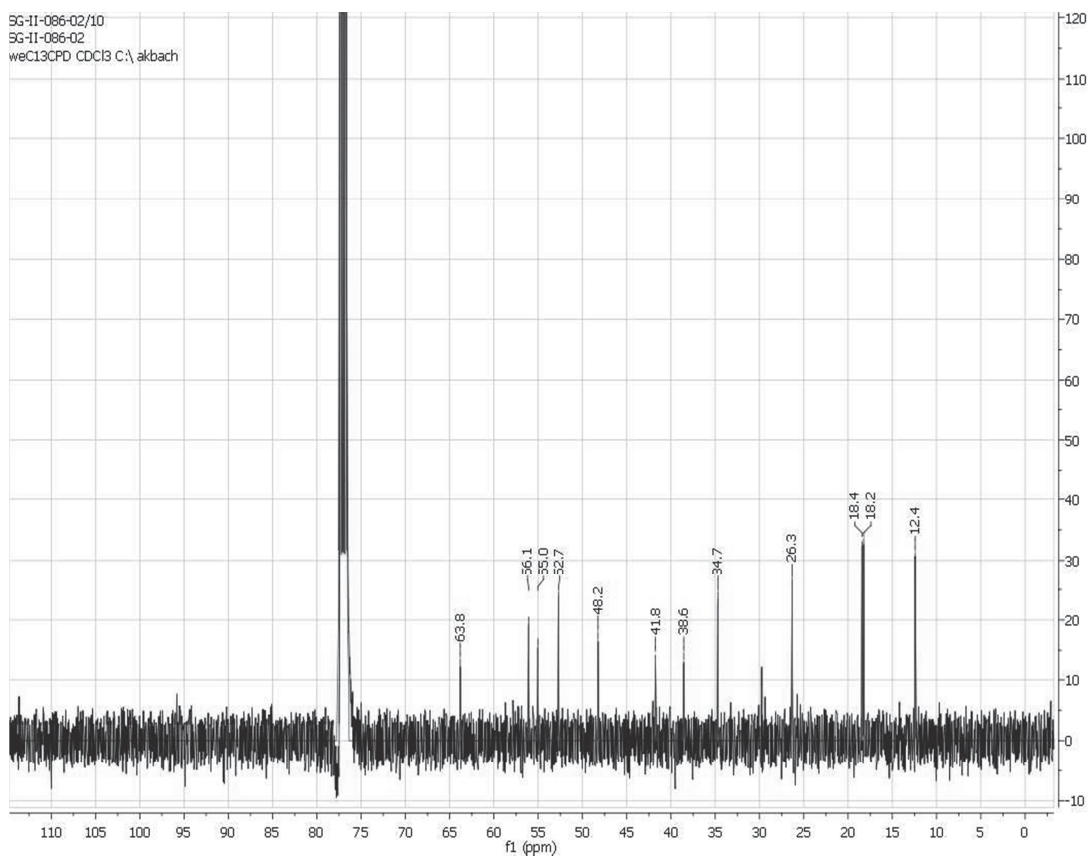
# Compound 5c, <sup>13</sup>C-NMR



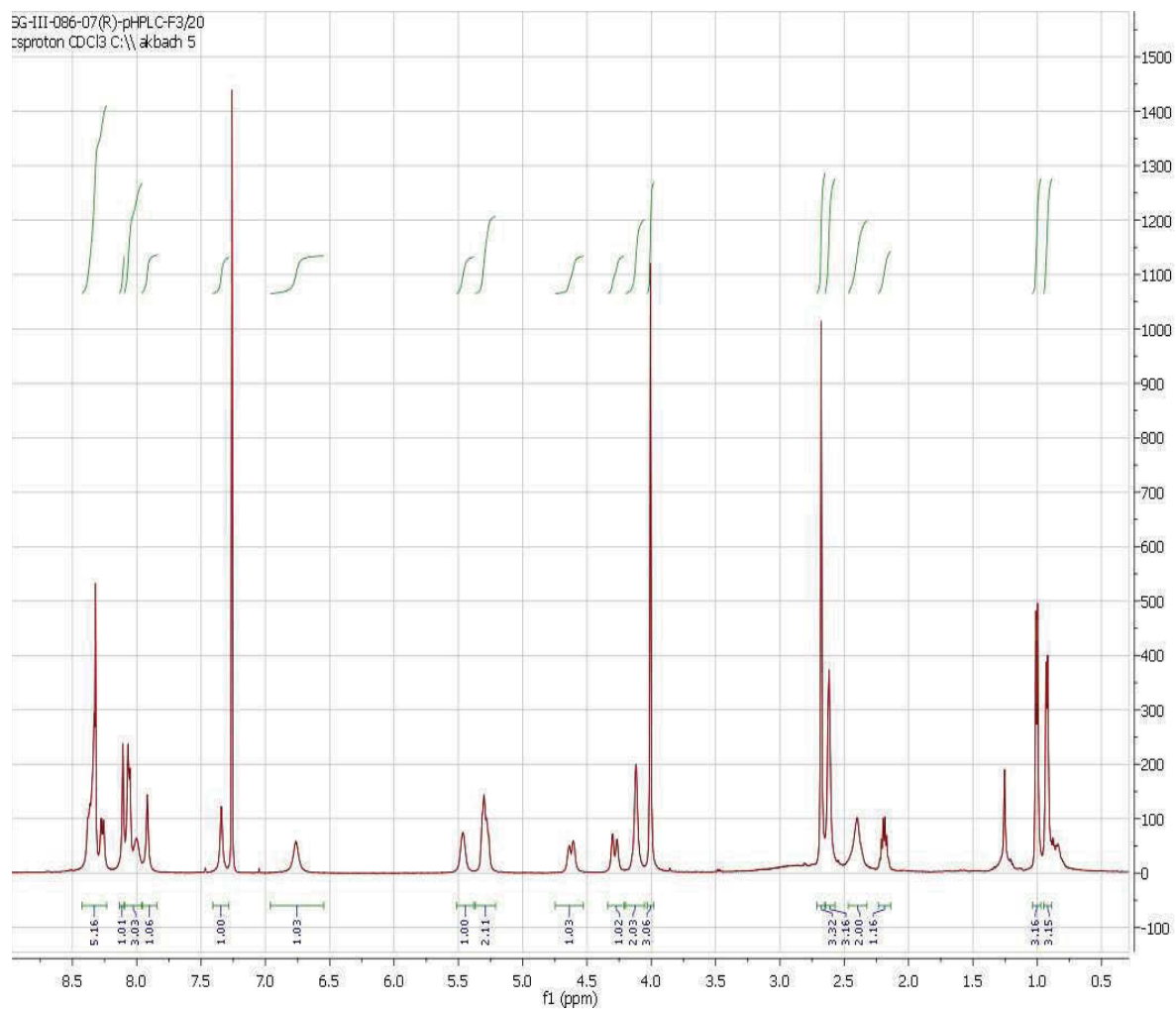
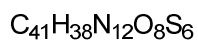
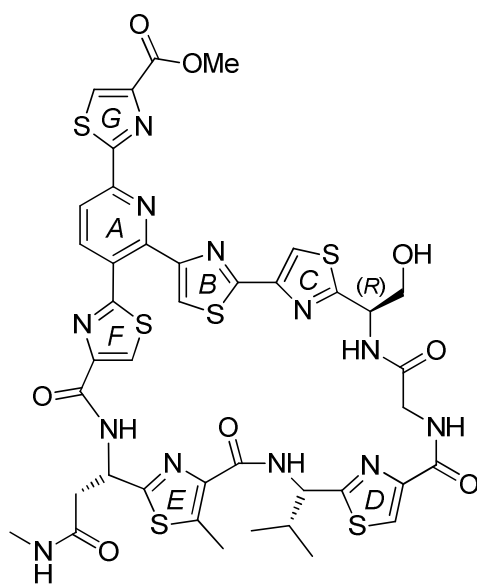
Thiopeptide 3a, <sup>1</sup>H-NMR



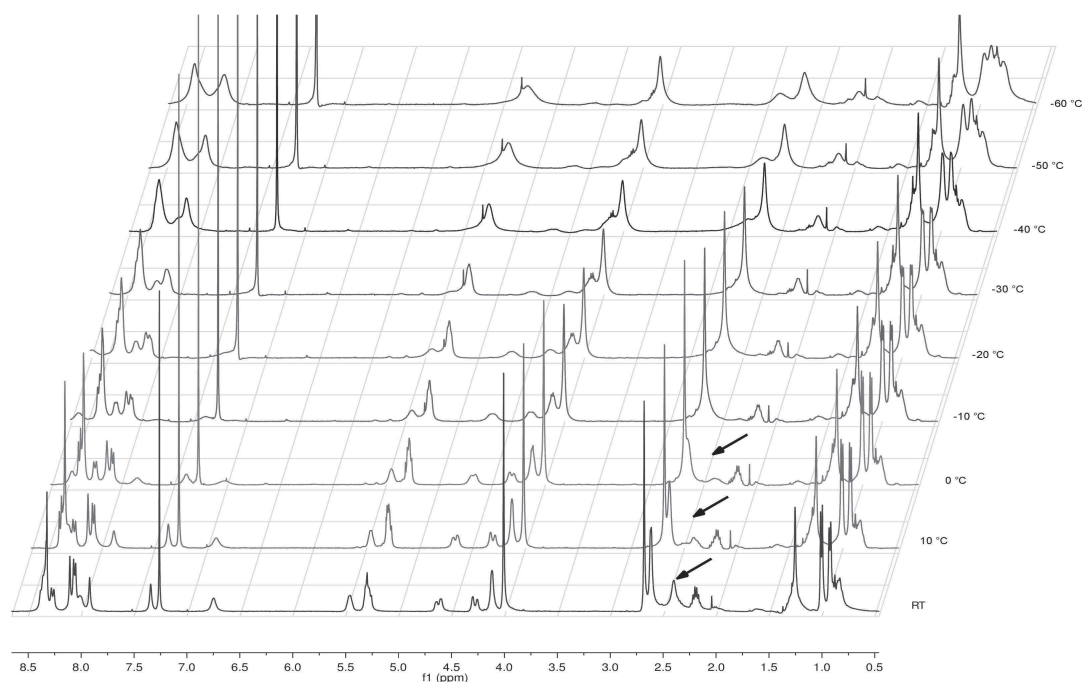
# Thiopeptide 3a, <sup>13</sup>C-NMR



# Thiopeptide 3b, <sup>1</sup>H-NMR

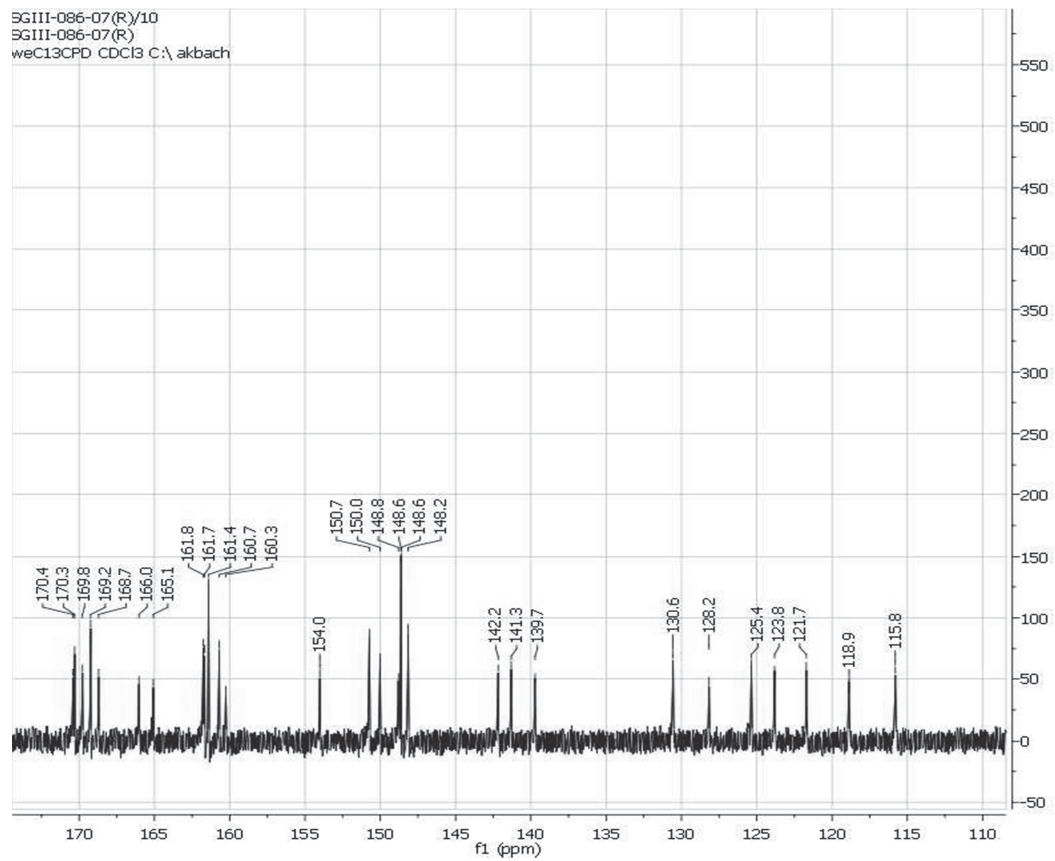
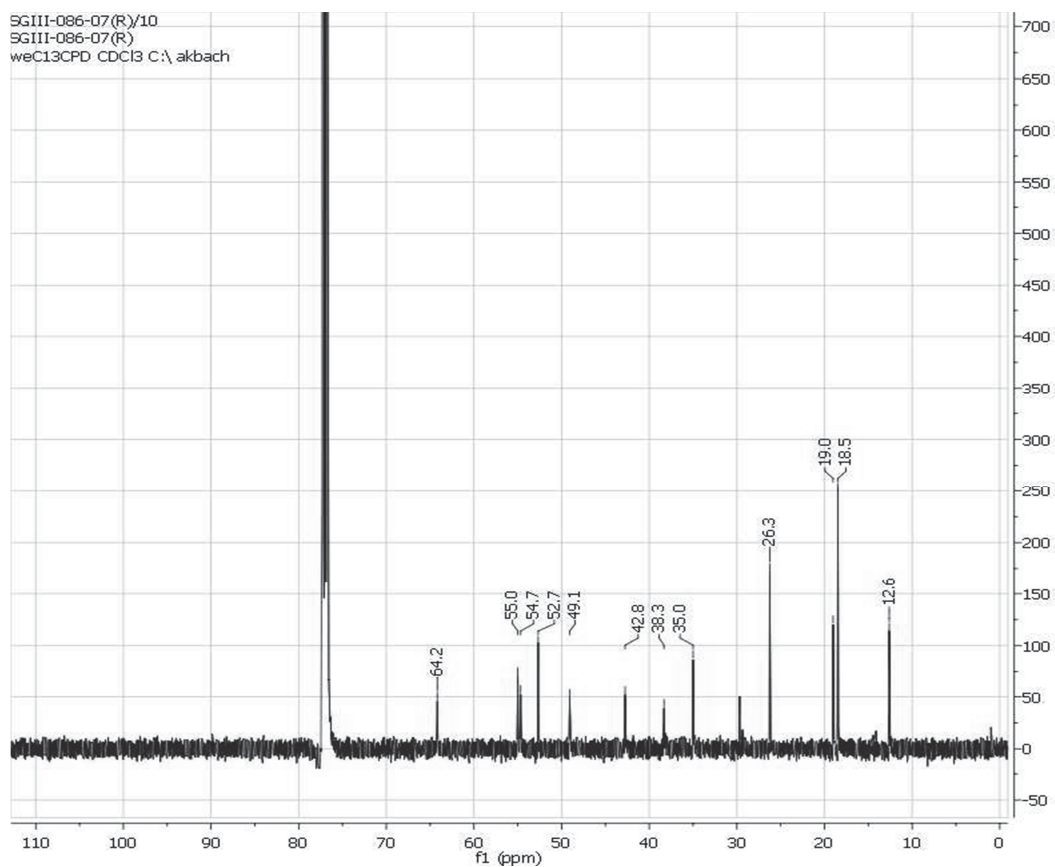


<sup>1</sup>H-NMR Spectra of thiopeptide **3b** at different temperatures in CDCl<sub>3</sub>:

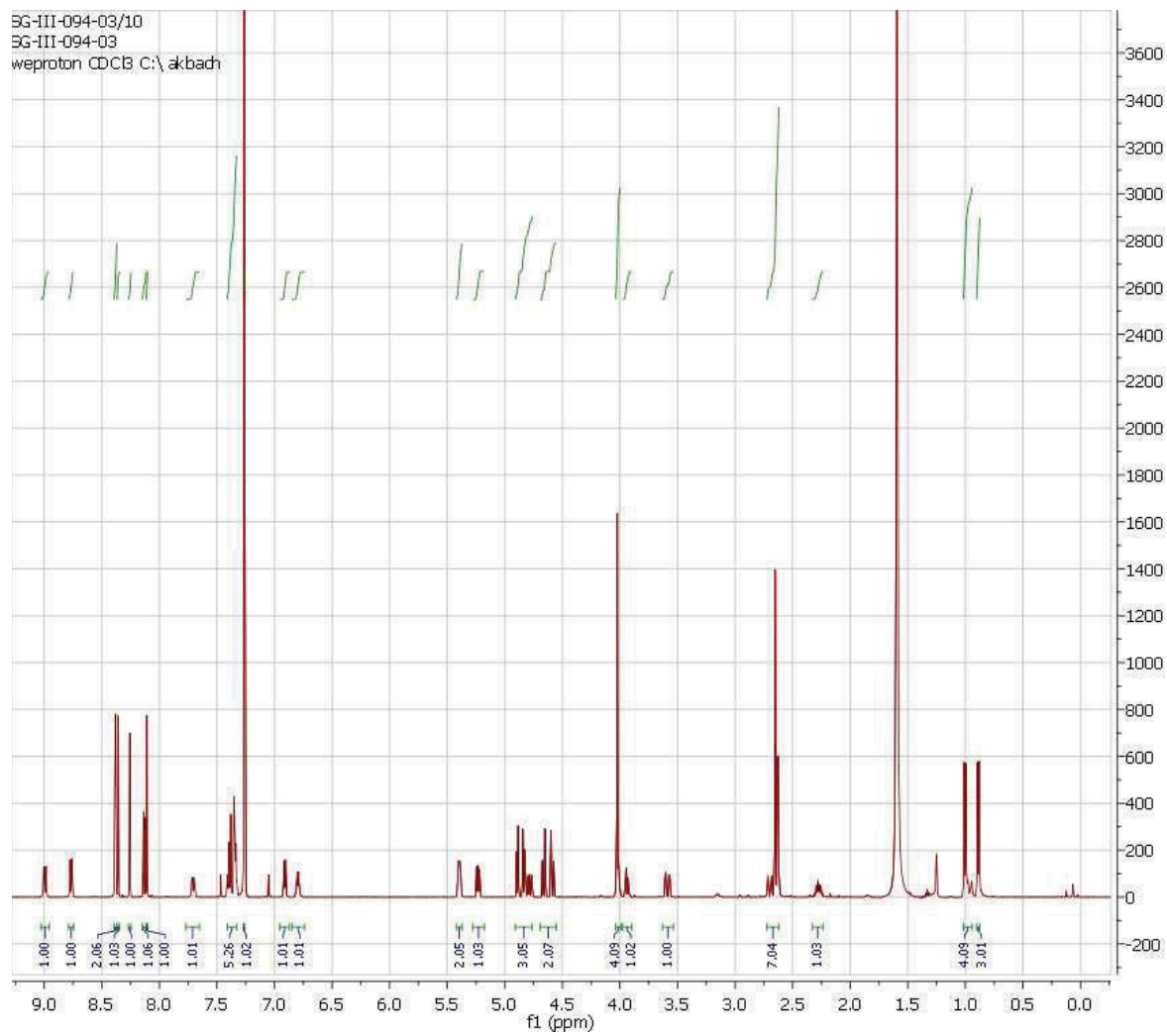
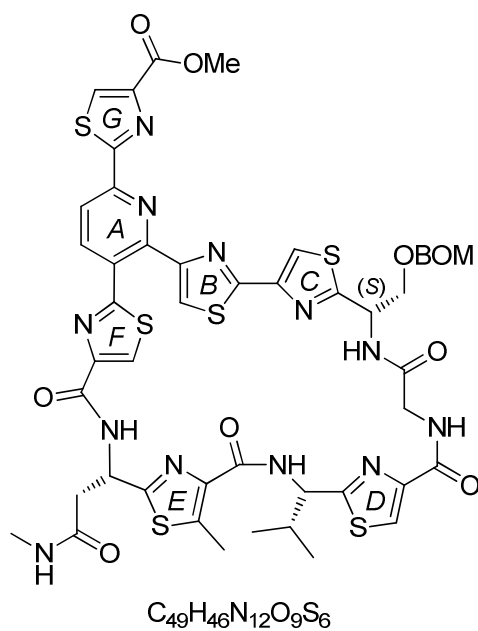


For the two diastereotopic protons at the methylene group C11 of the side chain between ring E and F only one broad singlet at 2.40 ppm can be observed at room temperature whereas at 10 °C the spectrum provides clear evidence for a separation of the signal (arrows). The observation supports the hypothesis that **3b** lacks hydrogen bonding between the proton at N13 and the oxygen atom at C28 and thus adopts a more flexible conformation than **2**, **3a** and **3c**.

# Thiopeptide 3b, <sup>13</sup>C-NMR

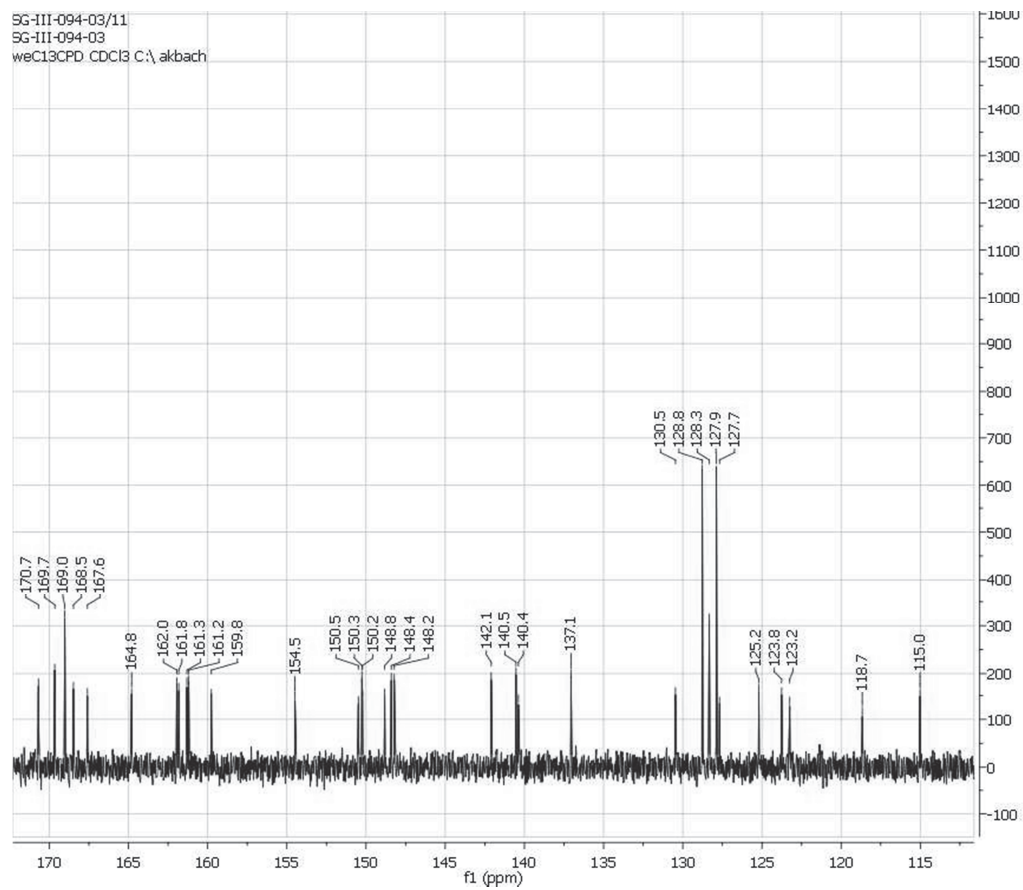
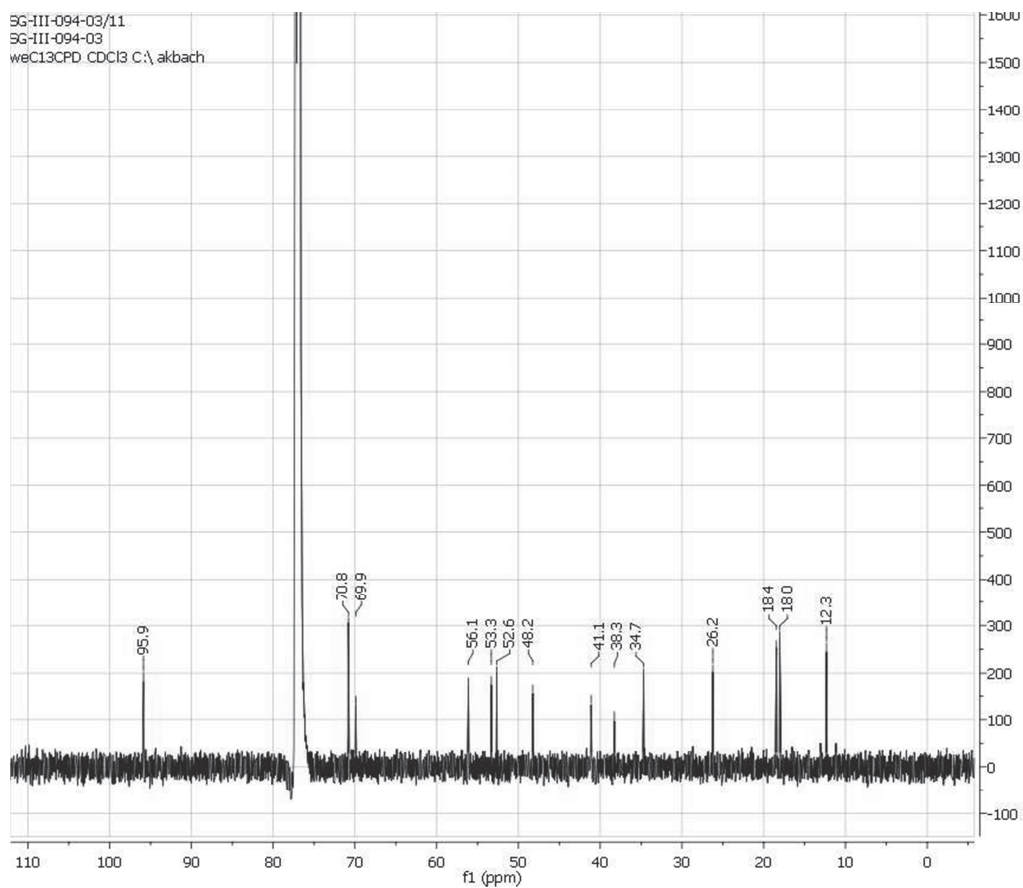


# Thiopeptide 3c, <sup>1</sup>H-NMR





# Thiopeptide 3c, <sup>13</sup>C-NMR



## Materials and methods for *in silico* studies

*Computational investigations:* As structural basis for the computational studies the experimental structure of EF-TU-GDPNP with bound GE2270A (PDB-ID: 2C77) was used.<sup>[1]</sup> Calculations were performed by the program DynaCell<sup>[4]</sup> together with the GROMOS 53A1 force field<sup>[6]</sup> and illustrations were created using VMD.<sup>[7]</sup> The 3D structures of the compounds **2** and **3a-c** were constructed using the programs ChemDraw and ChemDraw3D<sup>[8]</sup>. Force field parameters for the compounds were assigned based on GROMOS53A1 compatible topologies created by PRODRG.<sup>[9]</sup> Afterwards we manually checked all atom types, protonation states, and charges and adjusted parameters if necessary. In addition, we checked the consistency of the topology with respect to the compounds geometry using an in-house protocol of combined energy minimizations and short MD simulations, which was especially developed for the assignment of consistent ligand parameters (Topologies are available from the authors upon request).

*Simulation settings:* All simulations were performed using the simulation module of our in-house modeling program DynaCell. The same general simulation conditions were applied for the initial equilibration, the simulations of the compounds in solution, in complex with EF-Tu, and the OPMD refinement simulations. As energy function the GROMOS 53A1 force field was used<sup>[6]</sup> and a General Born implicit solvent model was applied.<sup>[10]</sup> Due to the use of the implicit solvent model no cutoffs were applied for the electrostatic and van-der-Waals interactions. Based on this energy function the leapfrog algorithm<sup>[11]</sup> was used and a time step of 2 fs (1 fs during heatup and for the OPMD refinement) was chosen. Due to the use of an implicit solvent model, additional stochastic forces were applied to keep the system stable<sup>[12]</sup>, a stochastic force constant of 10 kJ/(mol·nm) was used and a cutoff for the stochastic force calculations of 3 Å was chosen. The SHAKE algorithm<sup>[13]</sup> was applied to constrain all bond lengths. During heatup the temperature was controlled by velocity scaling, afterwards the temperature was kept at 300 K using the Berendsen's weak coupling

method<sup>[4]</sup> with a coupling constant of 0.005 (the coupling constant was especially optimized for the settings of the simulations, i.e. solvent model, stochastic dynamics conditions, etc.)<sup>[4]</sup>.

*Solution simulations:* All simulations in solution were performed using the general simulation conditions described above. The systems were heated up to 300 K over a period of 100 ps and afterwards the simulation was continued for 1000ps and afterwards extended to 5000 ps. As the analysis of the 5 ns trajectories showed that after 500 ps stable ring conformations were observed in all cases the most common structure (largest structural cluster) of the last 200 ps of simulation (800-1000 ps) was taken as starting structure for the docking studies, as at that time point the ring conformations had reached convergence and as an independent sampling of the side chain conformations was performed during in the docking calculations anyhow, convergence of these was not required.

*Docking calculations:* The bound complex of GE2270A was heated up over 100 ps to 300 K and afterwards equilibrated at 300 K for another 400 ps using the simulation conditions described above. Subsequently, the ligand was cut out and the final protein structure of the complex was used as starting structure for the docking procedure. The solution structures of the compounds **2** and **3a-c** were placed into the EF-Tu binding site with the ring system close to the bound conformation of GE2270 A. Afterwards random sampling was performed of the compounds position (rigid body rotation (10° steps) and translation ( $< \pm 2 \text{ \AA}$ ) and its side chain conformations until 200 docking poses with an overlap with the protein of less than 80% were found. All poses were refined for 100 ps using the OPMD method<sup>[4]</sup>. Afterwards ligand positional clustering was performed with a cutoff of 2 Å and the cluster representatives (the poses within the cluster with the best pepscore interaction energy) were ranked according to their pepscore-based interaction energy.<sup>[4]</sup> For the best ranked poses stochastic dynamics simulations were performed for 2 ns until equilibration was reached.

## Discussion of the technical aspects of the *in silico* studies

*Solution studies:* In Figures SI3 and SI4 the results of the simulations of the compounds in implicit solvent are shown. In SI3 it can be observed that the simulations converge very fast, between 500 and 1000ps the central ring conformation stabilizes and the side chain conformations become stable within 2000 ps. In Figure SI4 it can be observed that the intra-molecular hydrogen bond between the C28 carbonyl group and the proton at N13, which stabilizes the EF-Tu bound ring conformation, is retained during the simulations in solution in all compounds except 3c.

*EF-Tu complex docking and simulation studies:* In Figures SI5 and SI6 the corresponding results are provided for the MD refinement simulations of the best scored docking poses. In Figure SI5a the backbone RMSD of the protein with respect to its energy minimized X-Ray structure (PDB-ID: 2C77) is shown. It can be seen that the protein has reached a stable conformation after 1000ps in all four cases. In Figure SI5b the all-atom RMSD of the docked compounds from their starting docking poses are given. Also in this case the complexes reach a local equilibrium after 1000 ps. However for two compounds (3b and 3c) the RMSD of this equilibrium is rather high. The reason for this is that in these cases the docked conformations are not stable and the compounds start moving out of the binding site very quickly due to unfavorable interactions (see also discussion in the main paper and Figure 6). They stabilize in an intermediate position half out of the binding site, in which only stabilizing interactions are retained (Figure 6), before dissolving in solution. As this last step might take a long time, we stopped the simulations at that point, as our interests are focused on the stability of the bound complexes and by then it is already clear that no stable binding occurs. In Figure SI6 the hydrogen binding pattern of the intra-molecular hydrogen bond between the C28 carbonyl group and the proton at N13 during the time course of the simulations is given. Here the same trend can be observed as for the solution simulations, namely that for the compounds 2, 3a, and 3c this hydrogen bond is rather stable, but for compound 3b it is not. This underlines the importance of this interaction for the formation of stable bound

complexes. Overall the hydrogen bonds are less stable in the EF-Tu complex than in solution.

## References

[1] Parmeggiani, A.; Krab, I. M.; Okamura, S.; Nielsen, R. C.; Nyborg, J.; Nissen, P., Structural basis of the action of pulvomycin and GE2270 A on elongation factor Tu. *Biochemistry* **2006**, *45* (22), 6846-57.

[2] Nissen, P.; Kjeldgaard, M.; Thirup, S.; Polekhina, G.; Reshetnikova, L.; Clark, B. F.; Nyborg, J., Crystal structure of the ternary complex of Phe-tRNAPhe, EF-Tu, and a GTP analog. *Science* **1995**, *270* (5241), 1464-72.

[3] Lewis, R. J.; Hughes, R. A.; Alcaraz, L.; Thompson, S. P.; Moody, C. J., Solution structures of thiopeptide antibiotics. *Chem. Commun. (Camb)* **2006**, (40), 4215-7.

[4] Antes, I., DynaDock: A new molecular dynamics-based algorithm for protein-peptide docking including receptor flexibility. *Proteins* **2010**, *78*, 1084-1104

[5] Kjeldgaard, M.; Nissen, P.; Thirup, S.; Nyborg, J., The crystal structure of elongation factor EF-Tu from *Thermus aquaticus* in the GTP conformation. *Structure* **1993**, *1* (1), 35-50.

[6] Oostenbrink, C.; Villa, A.; Mark, A. E.; Van, W. F., A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6. *J. Comput. Chem.* **2004**, *25*, 1656-1676.

[7] Humphrey, W.; Dalke, A.; Schulten, K., VMD: visual molecular dynamics. *J Mol Graph* **1996**, *14*, 33-8, 27-8.

[8] Cousins, K. R., Computer review of ChemDraw Ultra 12.0. *J. Am. Chem. Soc.* **2011**, *133* (21), 8388.

<sup>[9]</sup> Schuettelkopf, A. W.; van Aalten, D. M. F., PRODRG - a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallographica D* **2004**, *60*, 1355-1363.

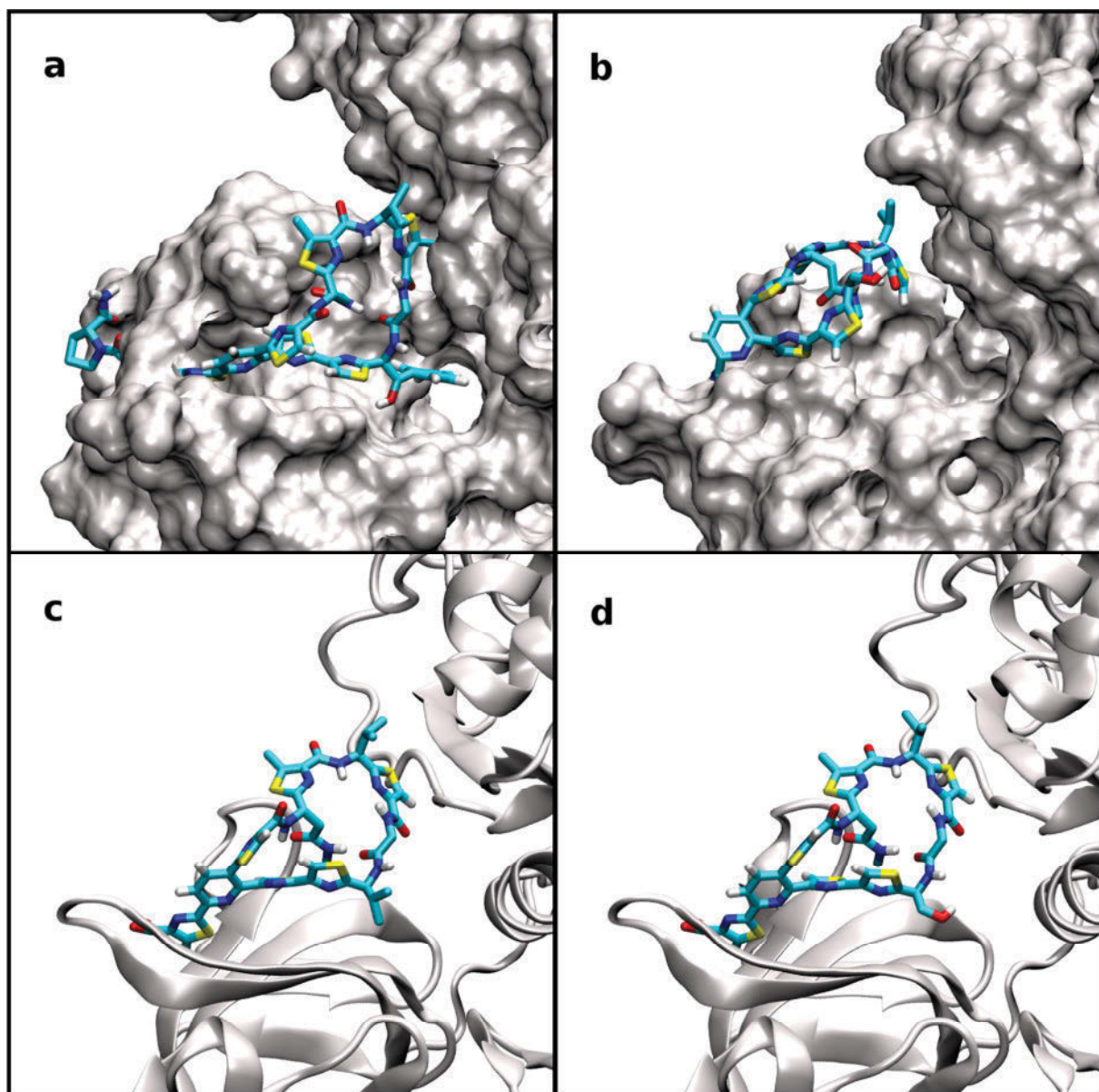
<sup>[10]</sup> Gallicchio, E.; Levy, R. M., AGBNP: an analytic implicit solvent model suitable for molecular dynamics simulations and high-resolution modeling. *J. Comput. Chem.* **2004**, *25* (4), 479-499.

<sup>[11]</sup> Gunsteren, W. F. v.; Berendsen, H. J. C., A leap-frog algorithm for stochastic dynamics. *Molecular Simulation* **1988**, *1*, 173-185.

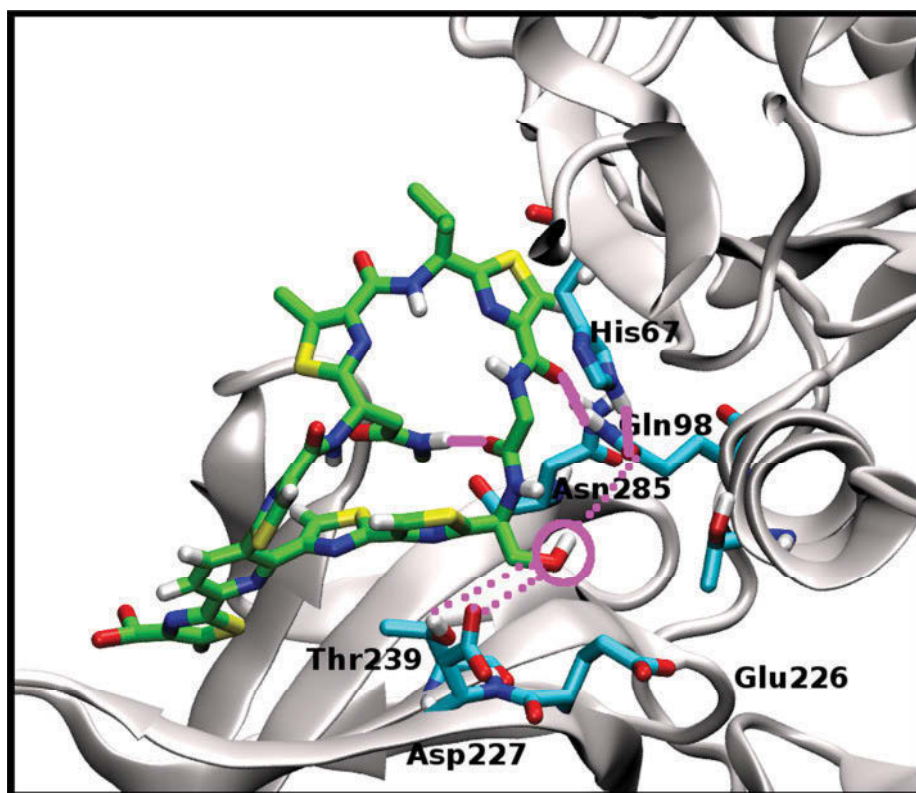
<sup>[12]</sup> Fan, H.; Mark, A. E.; Zhu, J.; Honig, B., Comparative study of generalized Born models: protein dynamics. *Proc. Natl. Acad. Sci. U S A* **2005**, *102* (19), 6760--6764.

<sup>[13]</sup> Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H., *J. Comput. Phys.* **1977**, *23*, 327-341.

<sup>[14]</sup> Berendsen, H.; Postma, J.; Gunsteren, W. v.; DiNola, A.; Haak, J., *J. Chem. Phys.* **1984**, *81*, 3684-3690.

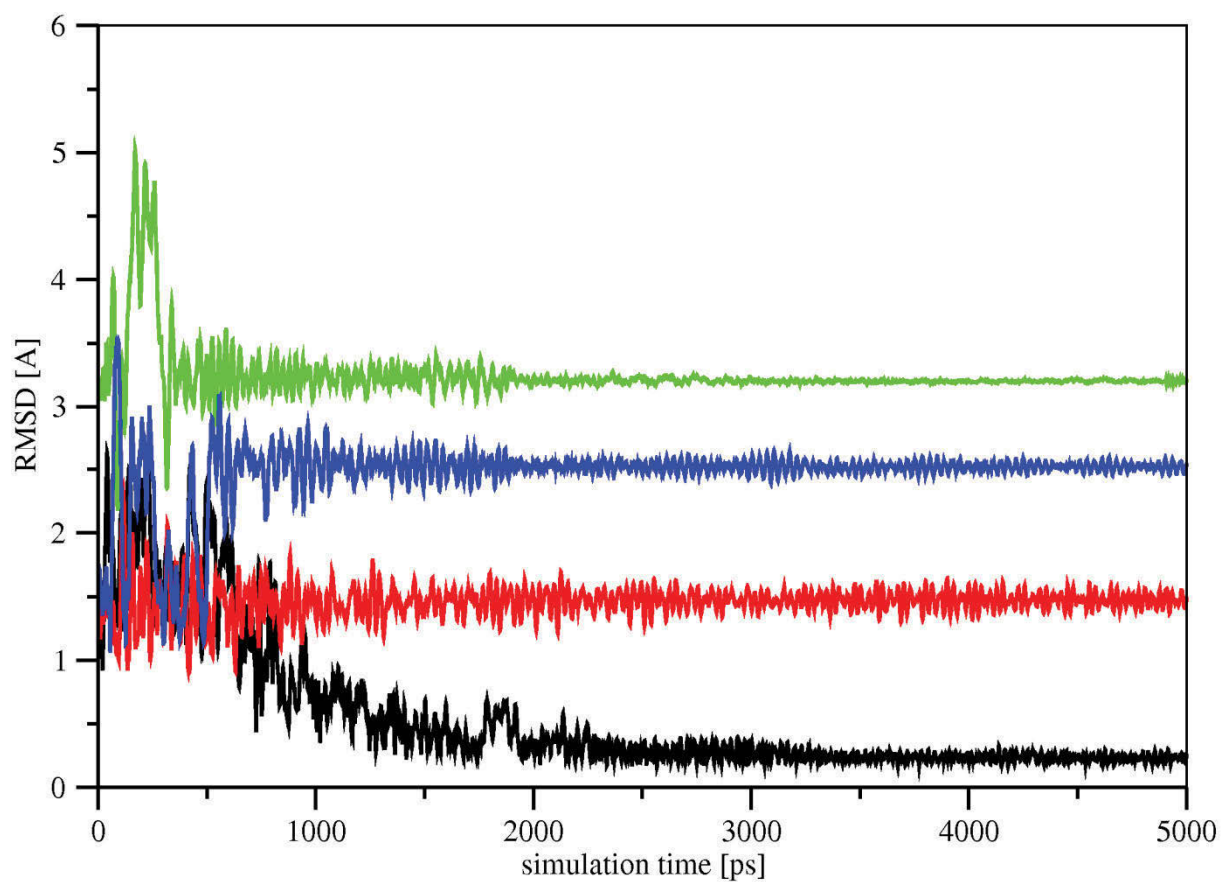


**Figure S11:** Bound structures of compounds GE2270 A (a) (X-ray structure), and **3b** (b), **2** (c), and **3a** (d) after 1 ns of molecular dynamics simulation (largest structural cluster within the last 200 ps of simulation). The ligands are shown in atom colored licorice representation and the binding site is given either in surface representation for (a) and (b) or cartoon representation for (c) and (d).

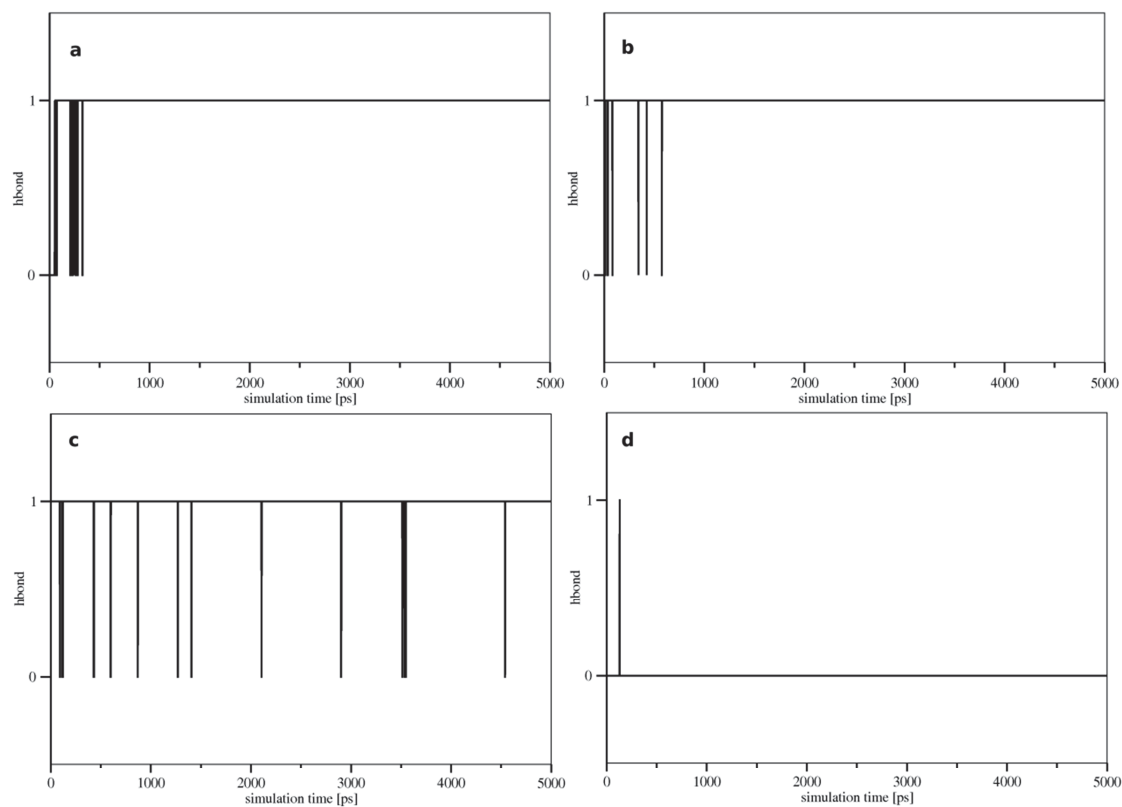


**Figure SI2:** Equilibrated bound structure of compound **3a** after 1 ns of MD simulation (largest structural cluster within the last 200 ps of simulation). The ligand and important residues in the binding site are shown in atom-colored licorice representation and the binding site is given in cartoon representation, Hydrogen bonds are shown in magenta. Switching hydrogen bonds of the hydroxyl group are given as pointed line.

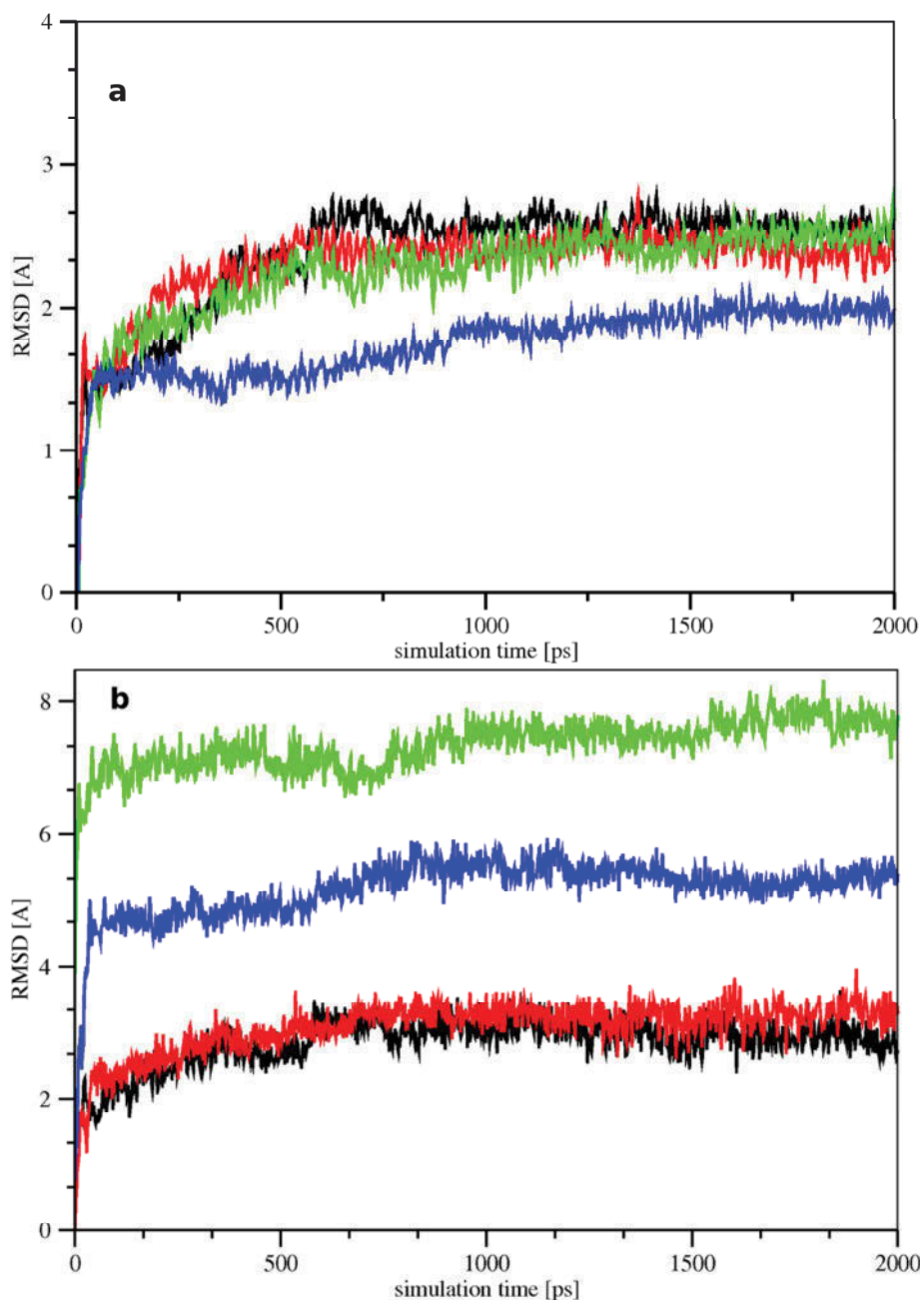




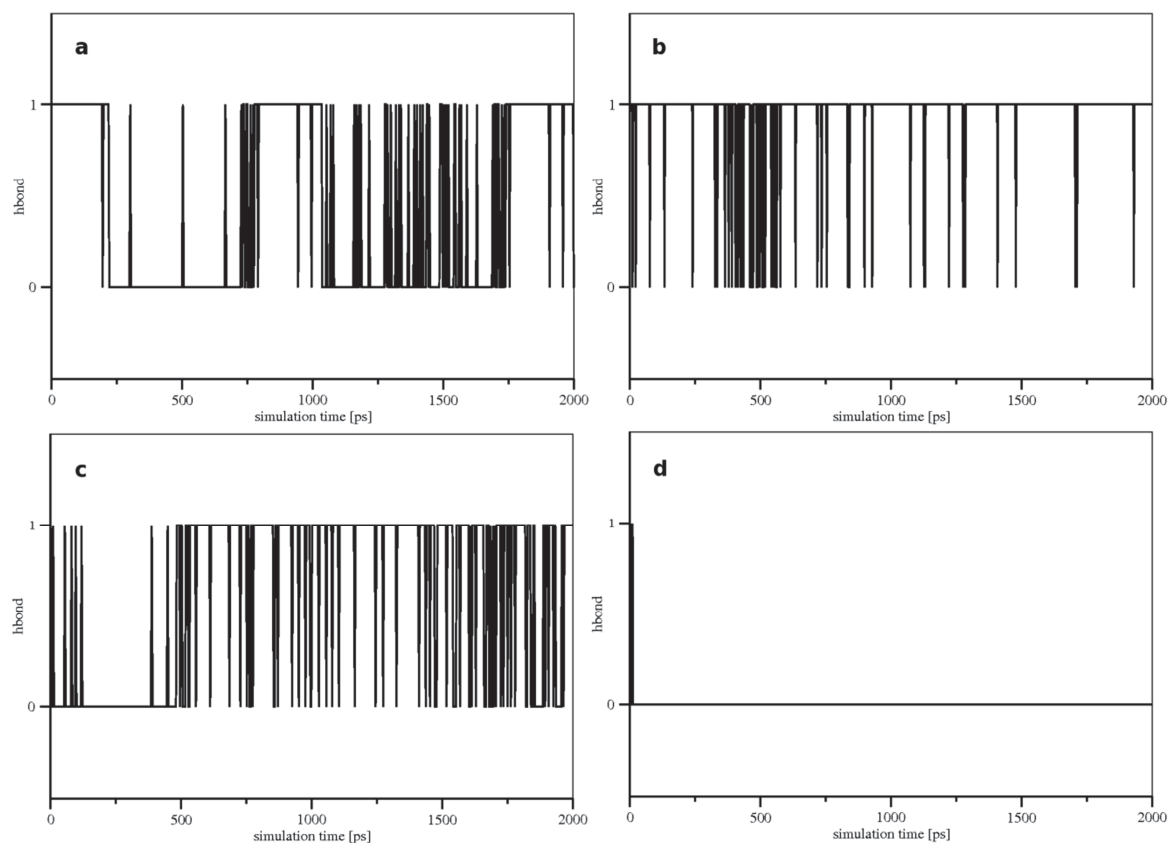
**Figure S13:** All-atom RMSD [Å] from the starting structures (energy minimised structures based on the EF-Tu complex bound ring conformation with minimized side chains) for the compounds **2** (red), **3a** (black), **3b** (blue), and **3c** (green) during the time course of the simulations in solution (starting after 100ps of heatup from 0 K to 300 K).



**Figure S14:** Hydrogen bond pattern for the intramolecular hydrogen bond between the C28 carbonyl group and the proton at N13 of the compounds **2** (a), **3a** (b), **3c** (c), and **3b** (d) during the time course of the simulations in solution (starting after 100 ps of heatup from 0 K to 300 K). 1 = hydrogen bond exists, 0 = no hydrogen bond, hydrogen bonds were calculated using VMD<sup>[7]</sup> default parameters.

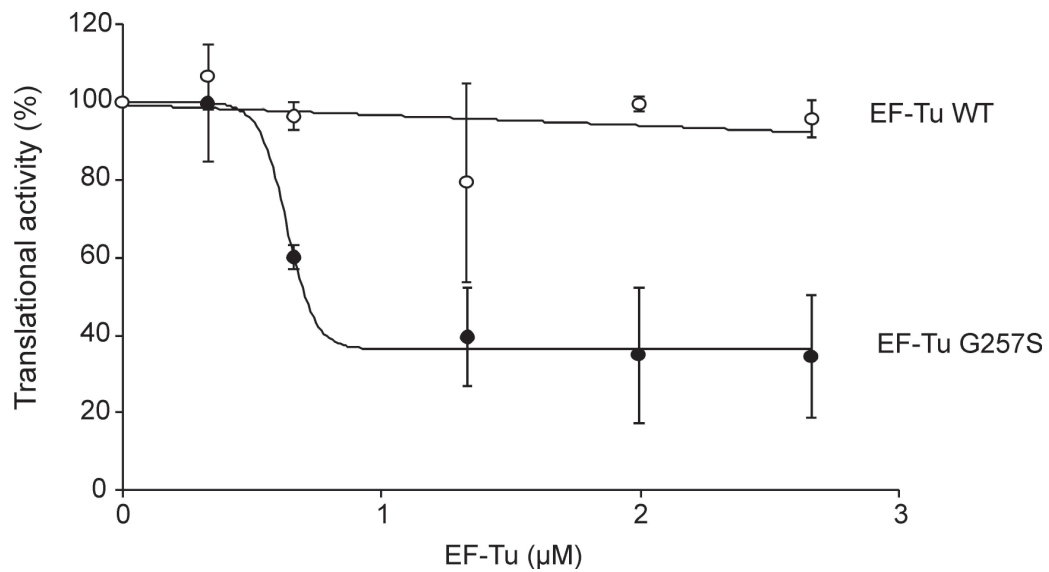


**Figure S15:** RMSD [Å] from the starting structures for the EF-Tu bound compounds **2** (red), **3a** (black), **3b** (blue), and **3c** (green) during the time course of the 2 ns refinement simulations of the EF-Tu complexes (starting after 500 ps of heatup from 0 K to 300 K and primary equilibration). (a) Backbone RMSD [Å] of the protein from its energy minimized X-ray conformation, (b) all-atom RMSD [Å] of the compounds from the docked conformations (the first 100 ps of OPMD refinement are not shown).



**Figure S16:** Hydrogen bond pattern for the intra-molecular hydrogen bond between the C28 carbonyl group and the proton at N13 of the compounds **2** (a), **3a** (b), **3c** (c), and **3b** (d) during the time course of the simulations in complex with EF-Tu (starting after 500 ps of heatup from 0 K to 300 K and primary equilibration). 1 = hydrogen bond exists, 0 = no hydrogen bond, hydrogen bonds were calculated using VMD<sup>[7]</sup> default parameters.

## Translational assays



**Supplementary figure.** Effect of EF-Tu G257S mutant on *in vitro* translation: *In vitro* translation of firefly luciferase in the presence of increasing concentrations of wildtype (WT) EF-Tu (open circles) and EF-Tu-G257S (closed circles). The firefly luminescence in the absence of EF-Tu was assigned as 100% and error bars display the standard deviation from the mean for three individual experiments.

# Entrapment of DNA in an intersubunit tunnel system of a single-stranded DNA-binding protein

Homa Ghalei<sup>1,2,†</sup>, Holger von Moeller<sup>1,†</sup>, Detlef Eppers<sup>1</sup>, Daniel Sohmen<sup>3</sup>, Daniel N. Wilson<sup>3</sup>, Bernhard Loll<sup>1</sup> and Markus C. Wahl<sup>1,\*</sup>

<sup>1</sup>Freie Universität Berlin, Fachbereich Biologie, Chemie, Pharmazie, Institut für Chemie und Biochemie, AG Strukturbiochemie, Takustr. 6, 14195 Berlin, Germany, <sup>2</sup>Max-Planck-Institute for Biophysical Chemistry, Department of Cellular Biochemistry, Am Fassberg 11, 37077 Göttingen, Germany and <sup>3</sup>Gene Center, Department of Biochemistry and Center of integrated Protein Science Munich (CiPSM), University of Munich, Feodor-Lynenstr. 25, 81377 Munich, Germany

Received December 27, 2013; Revised March 16, 2014; Accepted March 17, 2014

## ABSTRACT

Instead of a classical single-stranded deoxyribonucleic acid (DNA)-binding protein (SSB), some hyperthermophilic crenarchaea harbor a non-canonical SSB termed ThermoDBP. Two related but poorly characterized groups of proteins, which share the ThermoDBP N-terminal DNA-binding domain, have a broader phylogenetic distribution and co-exist with ThermoDBPs and/or other SSBs. We have investigated the nucleic acid binding properties and crystal structures of representatives of these groups of ThermoDBP-related proteins (ThermoDBP-RPs) 1 and 2. ThermoDBP-RP 1 and 2 oligomerize by different mechanisms and only ThermoDBP-RP2 exhibits strong single-stranded DNA affinity *in vitro*. A crystal structure of ThermoDBP-RP2 in complex with DNA reveals how the NTD common to ThermoDBPs and ThermoDBP-RPs can contact the nucleic acid in a manner that allows a symmetric homotetrameric protein complex to bind single-stranded DNA molecules asymmetrically. While single-stranded DNA wraps around the surface or binds along channels of previously investigated SSBs, it traverses an internal, intersubunit tunnel system of a ThermoDBP-RP2 tetramer. Our results indicate that some archaea have acquired special SSBs for genome maintenance in particularly challenging environments.

## INTRODUCTION

Deoxyribonucleic acid (DNA) replication, repair, recombination and telomere maintenance require the transient unwinding of duplex DNA (1,2). To maintain and protect DNA in the single-stranded (ss) state, all organisms harbor essential single-stranded DNA-binding proteins (SSBs) that bind ssDNA with high affinity and low sequence specificity. Classical SSBs harbor one of four distinct DNA-binding domains: the oligonucleotide/oligosaccharide-binding (OB) fold; the K homology (KH) domain; the ribonucleic acid (RNA) recognition motif (RRM) or the whirly domain (3). The sequences of the proteins are poorly conserved in either group (4) and different SSBs adopt a variety of oligomeric states (5), thereby bringing several ssDNA-binding domains together (6).

Recently, a group of hyperthermophilic crenarchaea has been found to lack a classical SSB and instead to contain a distinct SSB, termed ThermoDBP (7). The crystal structure of the ssDNA-binding N-terminal domain (NTD) of *Thermoproteus tenax* (*tte*) ThermoDBP, comprising a four-stranded  $\beta$ -sheet packed against four  $\alpha$ -helices (7), differs markedly from the classical SSB DNA-binding domains, and *tte*ThermoDBP was predicted to dimerize in a parallel fashion *via* a C-terminal leucine zipper (7). However, the exact mode of oligomerization and the mechanism of DNA binding of ThermoDBPs so far remained elusive.

ThermoDBPs share sequence similarity with domain of unknown function (DUF) 2258 proteins, which have a broader phylogenetic distribution in archaea (*Thermoproteales*, *Sulfolobales*, *Desulfurococcales*, *Thermococci* and *Archaeoglobi*). DUF2258 proteins can be divided into two groups that we term ThermoDBP-related proteins (ThermoDBP-RPs) 1 and 2. Both groups contain a

\*To whom correspondence should be addressed. Tel: +49 (0)30 838-53456; Fax: +49 (0)30 838-56981; Email: mwahl@zedat.fu-berlin.de

<sup>†</sup>These authors contributed equally to this work.

Present addresses:

Homa Ghalei, Department of Cancer Biology, The Scripps Research Institute, Jupiter, FL 33458, USA.

Holger von Moeller, moloX GmbH, Takustr. 6, 14195 Berlin, Germany.

ThermoDBP-like NTD but differ from ThermoDBPs and from each other in their C-terminal regions (Supplementary Figure S1). The ThermoDBP-like NTD suggests that ThermoDBP-RPs might also be SSBs. However, *tte*ThermoDBP-RP1 was not recovered from cell lysates via a biotinylated oligodeoxynucleotide used for isolation of *tte*ThermoDBP (7). Furthermore, *Sulfolobus solfataricus* (*ss*) ThermoDBP-RP1 was reported to associate with box C/D small (s) RNAs and with the 30S ribosomal subunit (8), but the protein was not detected in a subsequent proteomic characterization of *ss* ribosomes (9). We therefore set out to further characterize the structures and functions of ThermoDBP-RPs.

Our analyses show that ThermoDBP-RP2s are unconventional SSBs that bind ssDNA in an asymmetric fashion at an internal, intersubunit tunnel system of a symmetric protein tetramer. The apo-structure of ThermoDBP-RP1 suggests that DNA binding by this group of proteins may be auto-inhibited, explaining their weak nucleic acid affinities observed *in vitro*. Structural comparisons reveal that the eukaryotic DNA polymerase  $\eta$  has acquired a domain that closely resembles the N-terminal ssDNA-binding domain of ThermoDBPs and ThermoDBP-RPs but employs a different molecular surface for binding to double-stranded (ds) DNA.

## MATERIALS AND METHODS

### Protein production

Coding sequences of *Pyrococcus furiosus* (*pfu*) ThermoDBP-RP1 and *Aeropyrum pernix* (*ape*) ThermoDBP-RP2 were polymerase chain reaction (PCR)-amplified from *P. furiosus* and *A. pernix* genomic DNA. The PCR fragments encoding full-length ThermoDBP-RP genes were cloned into pET-M11 (EMBL, Heidelberg) via *NcoI/BamHI* restriction sites in frame with an N-terminal His<sub>6</sub>-tag followed by a Tobacco Etch Virus (TEV) protease cleavage site. For expression, cells were grown in auto-inducing medium (10) at 37°C to an OD<sub>600</sub> of 0.8, transferred to 18°C and harvested after 60 h. Cell pellets were resuspended in 20 mM HEPES-NaOH, pH 7.5, 500 mM NaCl and 5 mM  $\beta$ -mercaptoethanol supplemented with 10  $\mu$ g/ml DNaseI, 3 mM MgCl<sub>2</sub> and 100  $\mu$ g/ml lysozyme. Cells were lysed by four passes through a microfluidizer at 80 kPsi. The cell lysates were clarified by centrifugation. For purification, clarified lysates were loaded onto Ni<sup>2+</sup>-NTA resin equilibrated with lysis buffer. Beads were washed with one bed volume of 20 mM HEPES-NaOH, pH 7.5, 2 M LiCl and 5 mM  $\beta$ -mercaptoethanol followed by washing with 5 bed volumes of 20 mM HEPES-NaOH, pH 7.5, 500 mM NaCl, 20 mM imidazole and 5 mM  $\beta$ -mercaptoethanol. Proteins were eluted with 20 mM HEPES-NaOH, pH 7.5, 200 mM NaCl, 5 mM  $\beta$ -mercaptoethanol and 300 mM imidazole. Eluates were mixed with TEV protease (mass ratio protease/protein 1:50) and dialyzed overnight at 4°C against 20 mM HEPES-NaOH, pH 7.5, 100 mM NaCl and 2 mM  $\beta$ -mercaptoethanol. The dialyzed samples were again loaded on Ni<sup>2+</sup>-NTA resin equilibrated with dialysis buffer to remove the TEV protease and uncleaved protein. The flow-throughs were further purified by ion

exchange chromatography. Samples were loaded on a MonoQ column (GE Healthcare) equilibrated with dialysis buffer. Proteins were eluted with a linear gradient from 100 mM to 500 mM NaCl over 50 column volumes. Fractions containing the proteins of interest were loaded on Superdex 75 (*pfu*ThermoDBP-RP1) or Superdex 200 (*ape*ThermoDBP-RP2) size-exclusion chromatography columns (GE Healthcare) pre-equilibrated with 10 mM Tris-HCl, pH 7.6, 50 mM NaCl and 1 mM DTT. Fractions containing pure protein were pooled and concentrated to 20 mg/ml using Amicon concentrators (Millipore). To prepare selenomethionine (SeMet)-labeled proteins, *E. coli* BL21(DE3) cells were grown in SeMet auto-inducing medium (10). The other steps for expression and protein purification were the same as described above.

### Electrophoretic gel mobility shift assays

Box C/D small RNAs (sRNAs) were synthesized by *in vitro* transcription using T7 RNA polymerase and PCR amplified templates from *P. furiosus* genomic DNA. Transcription products were purified on a 6% urea-TBE denaturing polyacrylamide gel followed by phenol-chloroform extraction and ethanol precipitation. All other oligonucleotides (Supplementary Table S1) were chemically synthesized (Dharmacon and MWG/Operon). To prepare DNA or RNA duplexes, complementary strands were mixed in equimolar ratios, heated to 80°C and slowly cooled (0.2°C/min) in a thermocycler to 20°C. The samples were then separated on agarose gels and duplex bands were extracted and ethanol precipitated. For electrophoretic gel mobility shift assays, oligonucleotides were 5'-end labeled with  $\gamma$ -[<sup>32</sup>P]-ATP. Labeled oligonucleotides were mixed with recombinant proteins in 20 mM Tris-HCl, pH 8.0, 150 mM NaCl, 2 mM DTT and incubated at 20°C or 80°C for 10 min. To compete RNA binding, 0.5 g/l *E. coli* transferRNAs (tRNAs) were added to the reactions. Samples were then fractionated on a 6% (60:1) polyacrylamide gel and visualized using a PhosphorImager (Typhoon 8600, GE Healthcare). Binding of *ape*ThermoDBP-RP2 to single-stranded and circular M13 phage DNA was tested on a 1% agarose gel. Increasing concentrations of the protein were mixed with 100 ng DNA and incubated at 4°C for 1 h. DNA was visualized by ethidium bromide staining.

### Ribosome binding assays

Ribosomes and ribosomal subunits from *Thermococcus kodakarensis* were prepared as described (9). In a total reaction volume of 150  $\mu$ l, 50 pmoles of 30S, 50S or 70S *T. kodakarensis* ribosomal subunits or ribosomes were incubated with 400 pmoles of His<sub>6</sub>-tagged *pfu*ThermoDBP-RP1 or *ape*ThermoDBP-RP2 in a buffer containing 10 mM Tris-HCl, pH 7.5, 10 mM MgCl<sub>2</sub>, 1 mM DTT, 45 mM NH<sub>4</sub>Cl and the varying concentrations of NaCl. The mixture was incubated for 20 min at 37°C and was then placed on 90  $\mu$ l of a 20% sucrose cushion. The cushion was centrifuged for 30 min at 78 000 rpm in a TLA 100 rotor. After centrifugation the supernatant was removed and the pellet was resuspended. The input sample before centrifugation, the supernatant and the pellet were ana-

lyzed by sodium dodecyl sulphate-polyacrylamide gel electrophoresis followed by Western blotting. The membranes were stained by amido black to check the protein content in each lane and were then treated by anti-His<sub>6</sub> antibody for detection of ThermoDBP-RPs.

### Isothermal titration calorimetry

Isothermal titration calorimetry (ITC) experiments were performed at 25°C on an iTC200 microcalorimeter equipped with a 300- $\mu$ l syringe (Microcal, GE Healthcare). DNA oligonucleotides and *ape*ThermoDBP-RP2 were dialysed against the same buffer overnight. *ape*ThermoDBP-RP2 was titrated at 10  $\mu$ M concentration in a cell volume of 250  $\mu$ l with 50–100  $\mu$ M of different DNA oligonucleotides in 16 injections of 2.5  $\mu$ l volumes with 5-min intervals. The released heat was obtained by integrating the calorimetric output curves. Binding parameters were calculated using the Origin5 software using the ‘One Set of Sites’ curve fitting model provided by the software.

### Analytical size-exclusion chromatography

Analytical size-exclusion chromatography was conducted on Superdex 75 (*pfu*ThermoDBP-RP1) or Superdex 200 (*ape*ThermoDBP-RP2) PC 3.2/30 analytical size-exclusion columns (GE Healthcare) in 20 mM HEPES-NaOH, pH 7.5, 100 mM NaCl. Exclusion volumes ( $V_o$ ) were estimated using Blue Dextran (approximate molecular mass 2 MDa). Bovine  $\gamma$ -globulin (158 kDa), rabbit lactate dehydrogenase (140 kDa), chicken conalbumin (75 kDa), bovine serum albumin (67 kDa), chicken ovalbumin (45 kDa) and horse myoglobin (17 kDa) were used as molecular mass standards. Linear regression of plots of log (molecular mass) versus the ratios of the observed elution and exclusion volumes ( $V_e/V_o$ ) allowed calculation of the apparent molecular masses of *pfu*ThermoDBP-RP1 and *ape*ThermoDBP-RP2 based on the observed elution volumes of the proteins.

### Crystallization of ThermoDBP-RPs and of an *ape*ThermoDBP-RP2-dT<sub>10</sub> complex

Crystallization of *pfu*ThermoDBP-RP1 was carried out at 20°C using the sitting drop vapor diffusion method. *pfu*ThermoDBP-RP1 crystallized in two different crystal forms. Form 1 crystals of SeMet-labeled *pfu*ThermoDBP-RP1 were obtained by mixing 1  $\mu$ l of SeMet-labeled protein solution with 1  $\mu$ l of reservoir solution containing 50 mM sodium cacodylate, pH 6.4, 2.3 M (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> and 10 mM MgSO<sub>4</sub>. Prior to flash-cooling in liquid nitrogen, form 1 crystals of *pfu*ThermoDBP-RP1 were soaked in reservoir buffer supplemented with 25% (v/v) glycerol. Form 2 crystals of SeMet-labeled *pfu*ThermoDBP-RP1 were obtained by mixing 0.5  $\mu$ l of protein solution with 1  $\mu$ l of reservoir solution containing 0.1 M Tris-HCl, pH 8.5, 0.2 M Li<sub>2</sub>SO<sub>4</sub> and 40% (v/v) PEG400. These crystals were flash-cooled in a 100 K cryogenic stream in their mother liquor. Crystals of *ape*ThermoDBP-RP2 were produced by mixing equal volumes of SeMet-labeled *ape*ThermoDBP-RP2 with a reservoir solution containing 0.1 M imidazole, pH 8.0, 0.4 M NaH<sub>2</sub>PO<sub>4</sub>, 1.6 M K<sub>2</sub>HPO<sub>4</sub>, 0.2 M NaCl and 0.25 M glycine.

Crystals were cryo-protected by addition of 25% (v/v) glycerol to the mother liquor.

For co-crystallization with ssDNA, *ape*ThermoDBP-RP2 (17 mg/ml) was mixed with a 10-fold molar excess of dT<sub>10</sub> in 10 mM Tris-HCl, pH 7.6, 50 mM NaCl and 1 mM DTT. The mixture was incubated for 5 min at 80°C. Crystallization experiments were performed in a sitting drop format at 18°C. Crystals appeared after three days over a reservoir solution composed of 200 mM NaOAc, pH 4.3, 1.3 M NaH<sub>2</sub>PO<sub>4</sub> and 0.7 M K<sub>2</sub>HPO<sub>4</sub>. For cryo-protection, crystals were transferred to mother liquor supplemented with 15% (v/v) glycerol and subsequently flash-cooled in liquid nitrogen.

### Crystallographic procedures

Diffraction data were collected at beamline BL14.2 of the BESSY II storage ring (Berlin, Germany; Supplementary Table S2). Data were processed with the XDS package (11). The structure of *pfu*ThermoDBP-RP1 form 1 crystals was solved by the multiple-wavelength anomalous dispersion method using the program SHARP/autoSHARP (12). The structure of the second crystal form of *pfu*ThermoDBP-RP1 was solved by molecular replacement using the structure coordinates of the crystal form 1 as a search model with the program MOLREP (13). The structure of *ape*ThermoDBP-RP2 was solved *via* the single-wavelength anomalous dispersion method using the program HKL2MAP (14,15). Finally, the crystal structure of an *ape*ThermoDBP-RP2-dT<sub>10</sub> complex was solved by molecular replacement using the structure coordinates of an *ape*ThermoDBP-RP2 tetramer as a search model with the program PHASER (16). Model building was done manually with COOT (17). For all the structures, refinement was done with phenix.refine (18,19) or Refmac5 (20) including TLS refinement (21). Intermediate and final structures were evaluated with MOLPROBITY (22). All structure figures were drawn using PYMOL (23). Electrostatic surfaces were calculated with APBS (24).

## RESULTS

### Nucleic acid binding by ThermoDBP-RPs

To investigate the possible cellular functions of ThermoDBP-RPs, we first tested the reported interactions of the proteins with box C/D sRNAs and ribosomes or ribosomal subunits *in vitro*. In electrophoretic mobility shift assays (EMSA), *pfu*ThermoDBP-RP1 and *ape*ThermoDBP-RP2 bound to box C/D sRNAs but the interactions were efficiently competed by the addition of unlabeled *Escherichia coli* tRNAs (Supplementary Figure S2), suggesting that the proteins do not recognize sequence or structural features of box C/D sRNAs with high specificity. At low salt concentrations (20–120 mM NaCl), the proteins were pelleted by 30S and 50S ribosomal subunits and by 70S ribosomes (Supplementary Figure S3A; left and middle panels; lanes 3, 6 and 9). Increasing the salt concentration to 200 mM abrogated binding to the subunits and ribosomes, suggesting that these interactions are only marginally stable (Supplementary Figure S3B; left and middle panels; lanes 3, 6 and 9).



While these results do not rule out the possibility that ThermoDBP-RP proteins interact with box C/D sRNAs or ribosomes *in vivo*, they prompted us to investigate if the proteins interact also with other nucleic acids. Indeed, EMSAs (Figure 1A and B) and isothermal titration calorimetry (ITC; Supplementary Figure S4) showed that *ape*ThermoDBP-RP2 bound to a 21-mer ssDNA of mixed sequence or to 21-mer homo-pyrimidine ssDNAs with apparent  $K_d$  values of about 50–100 nM but very weakly or not at all to homo-purine ssDNAs, DNA duplexes, ssRNAs or an RNA duplex of the same length. A longer (45-mer) ssDNA was gradually shifted in EMSAs (Figure 1C), indicating a stepwise binding of more than one *ape*ThermoDBP-RP2 molecule or complex. Under the same experimental conditions, *pfu*ThermoDBP-RP1 did not bind to any of the 21-mer nucleic acids (Figure 1A and B) but weakly bound to a 45-mer ssDNA (about 50% shift at 50–100  $\mu$ M protein concentration; Figure 1C). Increasing the temperature or time of incubation had no effect on any of these interactions. These data suggest that at least ThermoDBP-RP2 proteins can act as SSBs.

### Crystal structures and modes of oligomerization of *pfu*ThermoDBP-RP1 and *ape*ThermoDBP-RP2

We next determined the crystal structures of *pfu*ThermoDBP-RP1 and *ape*ThermoDBP-RP2 at 2.43 and 2.05 Å resolution, respectively (Figure 1A, B and Supplementary Table S2). As expected from sequence analysis (Supplementary Figure S1), *pfu*ThermoDBP-RP1 contains an NTD similar to the *tte*ThermoDBP NTD (Figure 2C; root-mean-square deviation (r.m.s.d.) of 2.0 Å for 81 pairs of  $C\alpha$  atoms). Compared to the *pfu*ThermoDBP-RP1 NTD, the *tte*ThermoDBP NTD is expanded by two additional N-terminal  $\beta$ -strands and an extra helix between *pfu*ThermoDBP-RP1 helices  $\alpha$ 1 and  $\alpha$ 2 (Figure 2C and Supplementary Figure S1). The C-terminal part of *pfu*ThermoDBP-RP1 forms an amphipathic  $\alpha$ -helix. The C-terminal  $\alpha$ 3-helices of four *pfu*ThermoDBP-RP1 molecules associate as an antiparallel, four-helix bundle, positioning two pairs of NTDs at the opposite ends of a central four-helix rod (Figure 2A). About 14 000 Å<sup>2</sup> of solvent accessible surface area is buried upon tetramerization, and an interface analysis (25) suggested that the observed tetramer is stable in solution. In agreement with this analysis, *pfu*ThermoDBP-RP1 eluted from an analytical size-exclusion column consistent with the size of a tetramer (apparent molecular mass 65.5 kDa; calculated tetramer molecular mass 68.9 kDa).

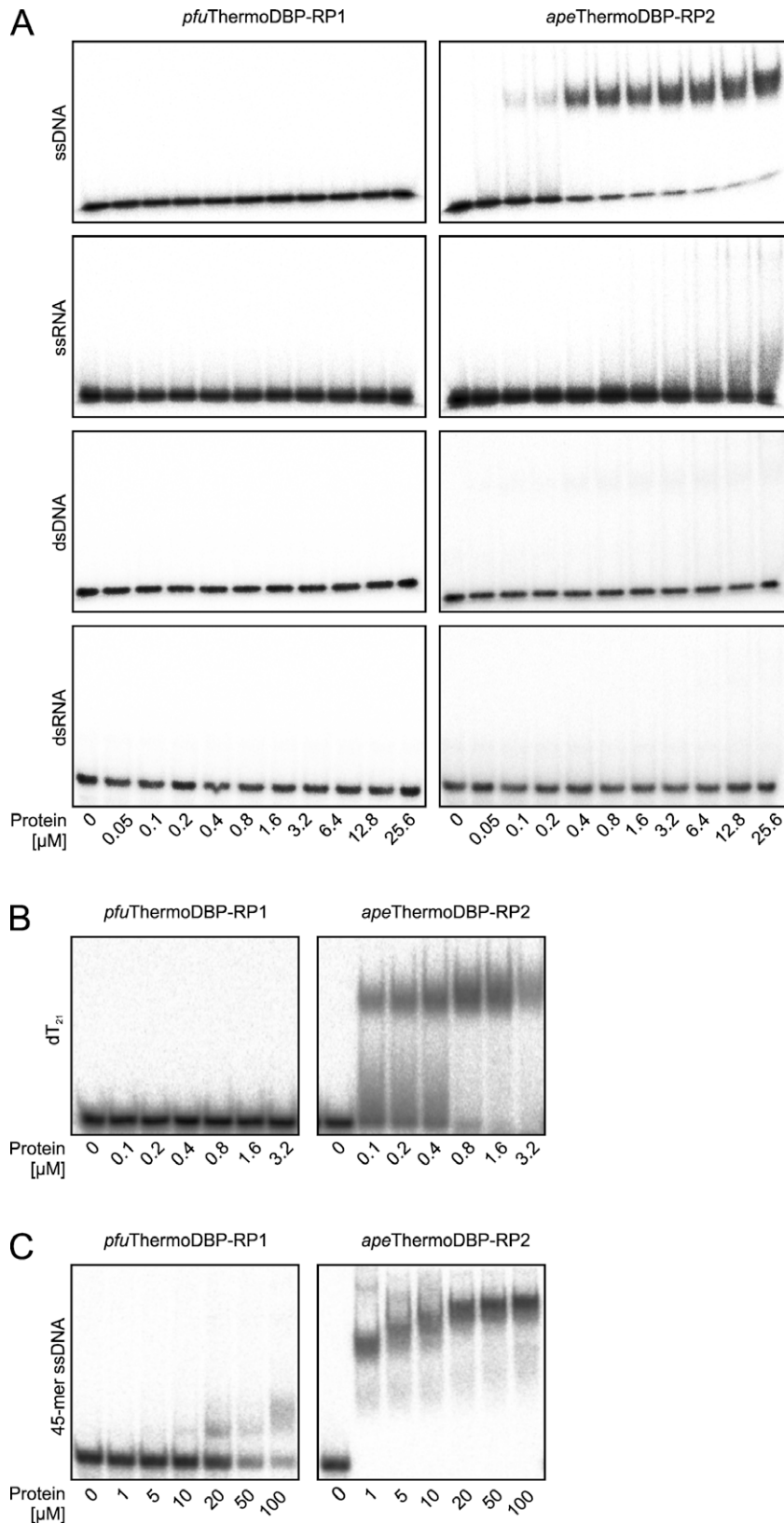
*ape*ThermoDBP-RP2 contains two globular domains, an NTD and a C-terminal domain (CTD) that are connected by a bent,  $\alpha$ -helical linker (helix  $\alpha$ 3; Figure 2B). The NTD is again similar to those of *tte*ThermoDBP and of *pfu*ThermoDBP-RP1 (r.m.s.d. of 2.1 Å and 0.8 Å, respectively, for 87  $C\alpha$  atoms; Figure 2C). The *ape*ThermoDBP-RP2 CTD is composed of a five-stranded, anti-parallel  $\beta$ -sheet sandwiched between two  $\alpha$ -helices (Figure 2B). *ape*ThermoDBP-RP2 also crystallized as a tetramer, which was arranged as a dimer of tightly intertwined dimers (Figure 2B). Analytical gel filtration analyses indicated

that *ape*ThermoDBP-RP2 also forms tetramers in solution (apparent molecular mass 111.3 kDa; calculated tetramer molecular mass 105.4 kDa). Within a tight dimer (subunits A/A' and B/B'; Figure 2B), each molecule cradles the NTD of the partner molecule between its own CTD and the connecting helix  $\alpha$ 3 (interface I; Figure 2B), thereby burying  $\sim$ 3300 Å<sup>2</sup> of solvent accessible surface area. Each molecule of a tight dimer interacts *via* its own NTD with the NTD of one molecule of a neighboring tight dimer (burying  $\sim$ 1300 Å<sup>2</sup> of solvent accessible surface area; interface II; Figure 2B), and *via* its own CTD with the CTD of the other molecule of the neighboring tight dimer (burying  $\sim$ 1600 Å<sup>2</sup> of solvent accessible surface area; interface III; Figure 2B). Only interface I within a tight dimer is predicted to be stable in isolation, while isolated interfaces II and III (between tight dimers) are predicted to be unstable (25). Consistently, interface I is largely hydrophobic (0.24 salt bridges/100 Å<sup>2</sup>) while interfaces II and III are more hydrophilic (1.5 and 0.38 salt bridges/100 Å<sup>2</sup>, respectively). Based on the interface analyses, we expect that *pfu*ThermoDBP-RP1 forms permanent tetramers but that interfaces II and III in *ape*ThermoDBP-RP2 may intermittently break either partly (forming open tetramers) or entirely (forming isolated tight dimers).

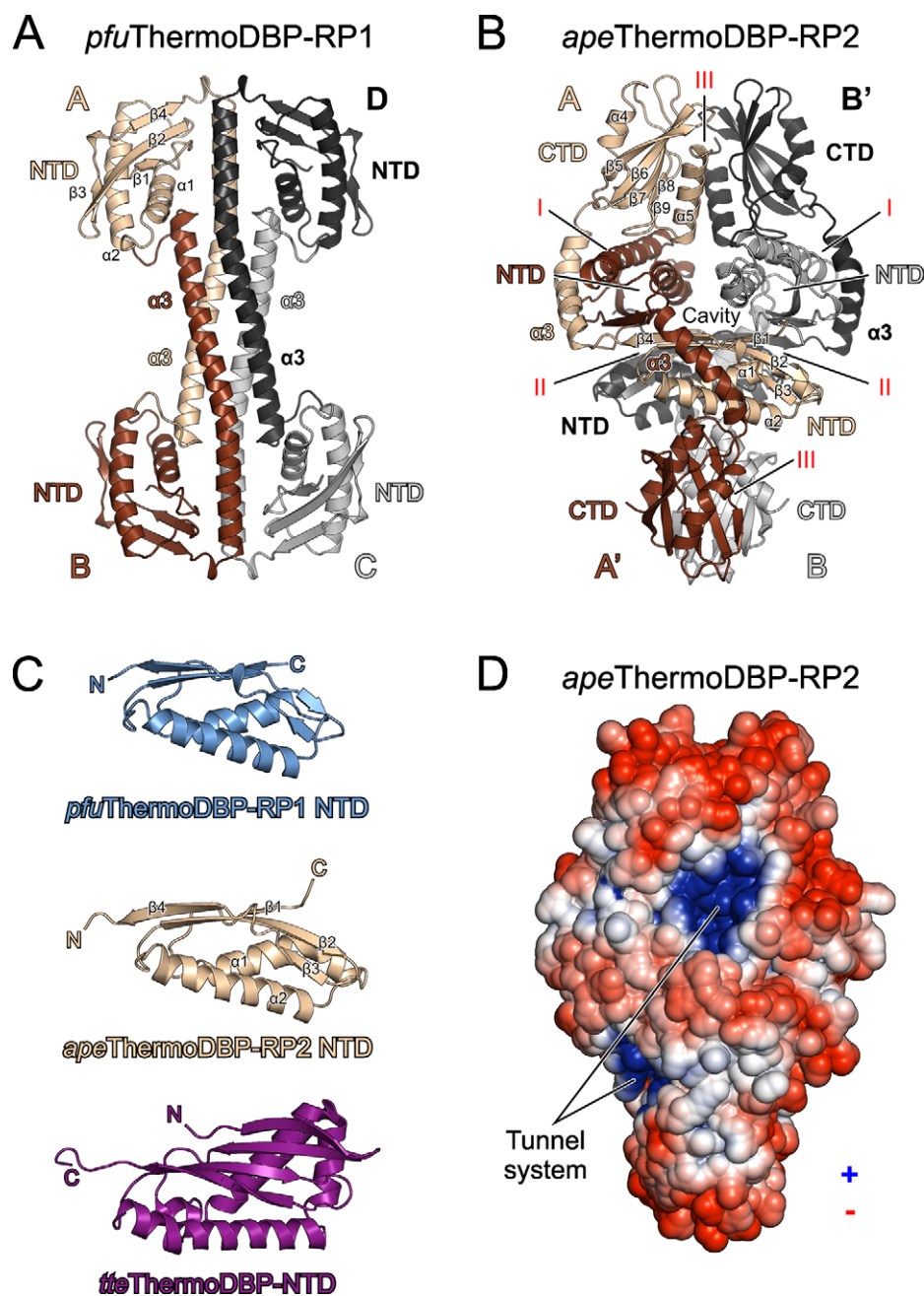
### Crystal structure of a ThermoDBP-RP2-DNA complex

The *ape*ThermoDBP-RP2 tetramer in the crystal exhibits largely electronegative outer surfaces and an interior tunnel system that is lined with electropositive surface potential (Figure 2D), possibly suitable for ssDNA binding. To reveal how *ape*ThermoDBP-RP2 binds ssDNA, we determined its crystal structure in complex with a decameric oligo-dT DNA (dT<sub>10</sub>) at 2.9 Å resolution (Figure 3A, B and Supplementary Table S2). The crystals contained two crystallographically independent *ape*ThermoDBP-RP2 tetramers per asymmetric unit, each associated with two dT<sub>10</sub> molecules (Figure 3A). We will focus our description on one of these virtually identical complexes with better defined electron density.

All 10 nucleotides of the first DNA strand (DNA1) could be traced, whereas only seven nucleotides of the second DNA strand (DNA2) were visible in the electron density. As shown in Figure 3A–C, the two DNA molecules meander in the 5'-to-3' direction from the bottom of the complex into the interior of the *ape*ThermoDBP-RP2 tetramer, where they are cradled between the juxtaposed NTDs of the four subunits. The central portions of both DNA molecules are bound in the same orientation across equivalent surfaces of two non-contacting NTDs of two tight dimers; *i.e.* DNA1 lies across the NTD of the A subunit (NTD<sup>A</sup>), while DNA2 binds across NTD<sup>B'</sup> (Figure 3B). Superimposing the subunit B'-DNA2 sub-complex on the subunit A-DNA1 sub-complex revealed that nucleotides 5–9 of DNA1 and the first five nucleotides visible in DNA2 are bound in a similar fashion across NTD<sup>A</sup> and NTD<sup>B'</sup>, respectively (Figure 3B). We attributed the same numbers to those nucleotides of the two DNA molecules that occupy equivalent positions on NTD<sup>A</sup> and NTD<sup>B'</sup> and engage in similar contacts, irrespective of the total length of the oligos (ten nucleotides). Thus, visible nucleotides of DNA1 were numbered dT1–



**Figure 1.** Nucleic acid binding properties of ThermoDBP-RP proteins. **(A)** Electrophoretic gel mobility shift assays testing the nucleic acid binding capabilities and preferences of *pfu*ThermoDBP-RP1 and *ape*ThermoDBP-RP2. ssDNA–d(ACTGCTAGAGATTTCCACAT); ssRNA–r(ACTGCTAGAGATTTCCACAT); dsDNA and dsRNA additionally contained the corresponding complementary strands. **(B)** Binding of *pfu*ThermoDBP-RP1 and *ape*ThermoDBP-RP2 to a 21-mer homo-pyrimidine (dT<sub>21</sub>) DNA. **(C)** Binding of *pfu*ThermoDBP-RP1 and *ape*ThermoDBP-RP2 to a 45-mer ssDNA–d(CTTGCTAGGACGGATCCCTCGAGGTTTTTTTTTTTTTTTTTTTTTTT).

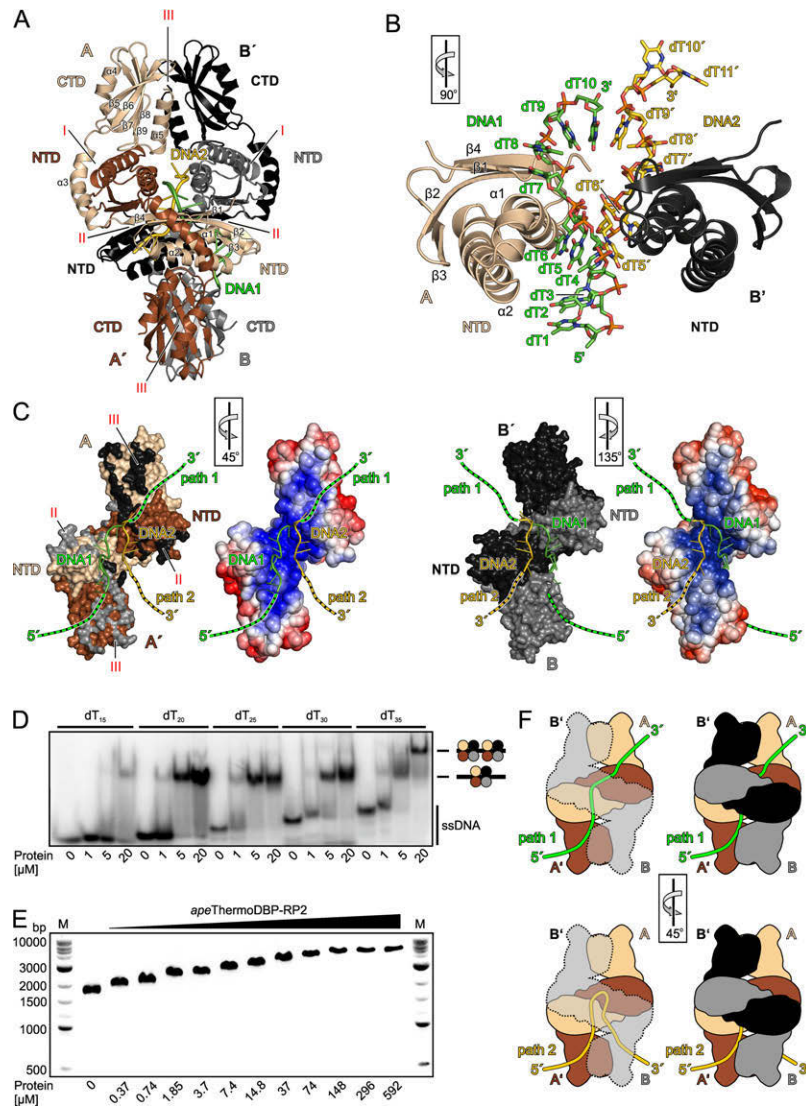


**Figure 2.** Structures and oligomerization of ThermoDBP-RP proteins. (A) Ribbon plot of the *pfuThermoDBP-RP1* tetramer (subunits A–D). Subunits A–D are colored in beige, brown, black and gray (or shaded differently). The NTD and the C-terminal  $\alpha 3$  helices forming an antiparallel coiled-coil are labeled. (B) Ribbon plot of the *apeThermoDBP-RP2* tetramer. Monomers are colored/shaded as in A. Subunits A/A' and B/B' form the tight dimers that associate into tetramers. NTDs, connecting helices  $\alpha 3$  and CTDs are labeled. Roman numerals (I–III) denote the different interfaces through which the tetramer forms. (C) Comparison of the NTDs of *pfuThermoDBP-RP1* (top), *apeThermoDBP-RP2* (center) and *tteThermoDBP-NTD* (bottom). Secondary structure elements are labeled. (D) Electrostatic surface representation of an *apeThermoDBP-RP2* tetramer (electrostatic coloring/shading scaled in the range of  $\pm 5$  kT/e). Entries/exits of an intersubunit tunnel system, which is lined with electropositive surface potential, are indicated.

dT10 while visible nucleotides of DNA2 were numbered dT5'–dT11', with nucleotides dT5–dT9 of DNA1 and nucleotides dT5'–dT9' of DNA2 being bound in an equivalent fashion on NTD<sup>A</sup> and NTD<sup>B</sup>, respectively (Figure 3B and Supplementary Figure S6). The two DNAs meet in the center of the *apeThermoDBP-RP2* tetramer, with the 3'-terminal nucleobase (dT10) of DNA1 stacking on the fifth nucleobase (dT9') of DNA2 (Figure 3B and Sup-

plementary Figure S6). The 3'-terminal two nucleotides of DNA2 (dT10' and dT11') are deflected from the center of the tetramer towards the upper part of the complex (Figure 3B and C).

We observed contacts of the protein subunits to the sugar-phosphate backbone as well as to the nucleobases of the DNA molecules (Supplementary Table S3; Figure 3B and Supplementary Figure S6). In addition to the in-



**Figure 3.** Binding of ssDNA by *apeThermoDBP-RP2*. **(A)** Ribbon plot of the *apeThermoDBP-RP2*-dT<sub>10</sub> complex with the DNAs in cartoon representation (DNA1: green/dark; DNA2: yellow/light). Monomers are labeled and colored/shaded as in Figure 2B. **(B)** Close-up view of two neighboring NTDs of the *apeThermoDBP-RP2* (from subunits A and B'; ribbons) in contact with two dT<sub>10</sub> molecules (DNA1 and DNA2 with nucleotides dT1–dT10 and dT5'–dT11', respectively; ball-and-stick; coloring/shading as in A). The two DNA strands run across the α1 helices of the respective NTDs. The rotation symbol in this and the following panels indicates the orientation relative to (A). **(C)** Views on the DNA-binding surfaces of the two tight dimers (A/A' left; B/B' right) with the associated DNA1 and DNA2. Relative orientations of the two tight dimers with respect to the tetramer are indicated by the rotation symbols. Two surface representations are shown for each tight dimer. Representations on the left show coloring/shading by molecules (as in Figure 2B). Black and gray patches on tight dimer A/A' indicate contact residues to subunits B' and B, respectively, of their neighboring tight dimer (belonging to interfaces II or III as labeled); beige and brown/light and dark patches on tight dimer B/B' indicate contact residues to subunits A and A', respectively, of their neighboring tight dimer. Representations on the right show surfaces of the tight dimers colored/shaded by electrostatic surface potential (as in Figure 2A). DNA1 is proposed to outline part of the path that an elongated ssDNA molecule could take through the tetramer. This molecule could either traverse the tetramer diagonally (green/dark dashed extensions; path 1), or turn back along the path that DNA2 is bound in opposite orientation (yellow/light dashed extensions; path 2). Both putative paths are lined with electropositive surface potential. **(D)** Electrophoretic gel mobility shift assay testing the binding of increasing amounts of *apeThermoDBP-RP2* to oligo-dT ssDNAs of increasing length. Interpretation of bands is given on the right. dT<sub>15</sub>, dT<sub>20</sub>, dT<sub>25</sub> and dT<sub>30</sub> show only one shifted band, a second, slower migrating band appears with dT<sub>30</sub> at higher *apeThermoDBP-RP2* concentrations, indicating that a ssDNA molecule of 35 nucleotides can accommodate two *apeThermoDBP-RP2* tetramers. **(E)** Electrophoretic gel mobility shift assay testing the binding of *apeThermoDBP-RP2* to circular M13 phage ssDNA on a 1% agarose gel. DNA was visualized by ethidium bromide staining. Upon addition of increasing concentrations of the protein (indicated below the gel), the DNA migrates progressively slower, indicating that increasing numbers of *apeThermoDBP-RP2* molecules bind to the circularly closed DNA. **(F)** Model for binding of long ssDNAs to an *apeThermoDBP-RP2* tetramer along path 1 (top) and path 2 (bottom). Left panels: Views on the ssDNAs running across the *apeThermoDBP-RP2* A/A' tight dimer with the other dimers shown as a semi-transparent outline. Right panels: Same views with the B/B' tight dimer as solid surfaces.

teractions of the central portions of DNA1 and DNA2 with NTD<sup>A</sup> and NTD<sup>B</sup>, respectively, they both engage in additional contacts to all other protein subunits of the *apeThermoDBP-RP2* tetramer. These interactions again involve primarily residues from the NTDs of the other subunits (Supplementary Table S3 and Supplementary Figure S6). Thus, the NTD of one protein subunit provides the main binding platform for one ssDNA molecule and NTDs from the other subunits engage in different contacts to this DNA molecule and thereby guide it asymmetrically through the symmetric protein tetramer.

In DNA1, dT1 and dT2 stack on each other and on N211 of helix  $\alpha_6$  from CTD<sup>B</sup> (the only CTD-DNA interaction seen in the entire structure). The side chain of dT3 is rotated outwards, introducing a first kink in the DNA backbone. The kinked backbone is stabilized by contacts to K63 and L64, originating from the loop following helix  $\alpha_2$  of NTD<sup>B</sup>. The following dT4, dT5 and dT6 again form a continuous stack that is capped by F23 from helix  $\alpha_1$  of NTD<sup>A</sup>. This helix is wedged between nucleotides dT6 and dT7, introducing a second kink in the nucleic acid. dT8 stacks on dT7 and the two nucleotides are sandwiched between the central portions of the NTD<sup>A</sup> and NTD<sup>B</sup>  $\alpha_1$  helices. NTD<sup>B</sup> helix  $\alpha_1$  imposes a third kink and guides the 3'-terminal two-nucleotide stack (dT9–dT10) towards the center of the *apeThermoDBP-RP2* tetramer, where the backbone of dT10 is held between K17 and R20 from helix  $\alpha_1$  of NTD<sup>A</sup>. The conformation and path through the *apeThermoDBP-RP2* tetramer of the second DNA strand is similar except for its last two nucleotides, dT10' and dT11' (see above). While the protein imposes a multiply kinked structure on the DNA ligands, the *apeThermoDBP-RP2* tetramer itself did not undergo any significant conformational changes upon DNA binding (r.m.s.d.  $\sim 0.9$  Å for 880 C $\alpha$ -pairs between DNA-free and DNA-bound forms).

### Molecular basis for the nucleic acid-binding preferences of *apeThermoDBP-RP2*

The kinks in the DNA backbone are incompatible with continuous Watson-Crick base pairing in a DNA duplex and the size of the central tunnel system in an *apeThermoDBP-RP2* tetramer could not accommodate duplex DNA (or RNA), rendering *apeThermoDBP-RP2* specific for single-stranded nucleic acids. Purines could be accommodated at most positions but a purine at positions 9/9' would clash with the  $\alpha_1$  helices of NTD<sup>B</sup>/NTD<sup>A</sup>, respectively, explaining why *apeThermoDBP-RP2* can bind ssDNA of mixed sequence but fails to bind homo-purine sequences (Supplementary Figure S4). Alternatively or additionally, homo-purine sequences may be excluded because of their tendency to form higher-order structures, such as dG-quadruplexes (26) or poly-dA parallel helices (27), which could not be accommodated in the restricted internal tunnel system of the *apeThermoDBP-RP2* tetramer. Although the protein interacts with several nucleic acid bases, none of these contacts would be expected to further restrict the nucleic acid sequence that can be bound. Furthermore, a 2'-hydroxyl group on nucleotides 2/2' would clash with the base of nucleotides 4/4', a 2'-hydroxyl group on nucleotides 4/4' would sterically interfere with the ribose of nucleotides 5/5'

and a 2'-hydroxyl on nucleotides 6/6' would clash with the backbone of Arg20 of NTD<sup>A</sup>, explaining the preference of *apeThermoDBP-RP2* for ssDNA over ssRNA.

### Extent of an *apeThermoDBP-RP2* binding site and engagement of ssDNA lacking free termini

The precise stoichiometry and ssDNA binding mode seen in the present crystal structure may be a consequence of the short DNA oligomers and high DNA concentrations used during the crystallization experiment. To estimate the number of nucleotides occupied by one *apeThermoDBP-RP2* tetramer on ssDNA in solution, we conducted EMSA experiments with increasing length oligo-dT ssDNAs (Figure 3D). Up to a length of 30 nucleotides only a single shifted band was detected, indicating binding of one *apeThermoDBP-RP2* tetramer (Figure 3D). However, starting with a length of 35 nucleotides, a second, slower migrating band appeared at higher *apeThermoDBP-RP2* concentrations (Figure 3D), indicative of two *apeThermoDBP-RP2* tetramers being accommodated on a dT<sub>35</sub> ssDNA. These results suggest that one *apeThermoDBP-RP2* tetramer occupies a stretch of 17 to 18 nucleotides on a ssDNA target, very similar to the combined number of nucleotides from DNA1 and DNA2 that we see bound to an *apeThermoDBP-RP2* tetramer in the crystal structure (17 nucleotides).

Physiologically, *apeThermoDBP-RP2* may engage extended ssDNA stretches that lack free ends. While we also obtained crystals of *apeThermoDBP-RP2* with longer DNA oligomers, they diffracted poorly (about 5 Å resolution) and the electron densities for the DNA ligand(s) were poorly defined. To test whether *apeThermoDBP-RP2* is capable of binding ssDNA molecules that lack free termini, we performed EMSAs with single-stranded circular M13 phage DNA (Figure 3E). Upon addition of increasing concentrations of *apeThermoDBP-RP2*, the ssDNA was progressively retarded on the gels, indicative of the protein occupying multiple binding sites in a stepwise manner. Due to the large size of the M13 DNA (7249 nucleotides) and the limited number of nucleotides required to accommodate one *apeThermoDBP-RP2* tetramer, it is likely that at each increment in protein concentration multiple additional *apeThermoDBP-RP2* tetramers bind.

### DNA-binding may be auto-inhibited in the *pfuThermoDBP-RP1* tetramer

Residues of *apeThermoDBP-RP2* that contact the DNA molecules in the present structure are conserved in *pfuThermoDBP-RP1* (Supplementary Figure S1). To investigate why the latter protein nevertheless binds poorly to ssDNA (Figure 1B and C), we superimposed the NTD<sup>A</sup> domain of *apeThermoDBP-RP2* in complex with the DNA1 molecule on an NTD of the *pfuThermoDBP-RP1* tetramer (Supplementary Figure S5A, left panel). The superposition revealed that in *pfuThermoDBP-RP1*, the putative DNA-binding surface on the NTD is occluded by the C-terminal tail of another *pfuThermoDBP-RP1* subunit (Supplementary Figure S5A; close-up in left panel). Residues 128–139 of subunit B clash with a DNA molecule modeled on

the NTD of subunit A. Thus, the present crystal structure of *pfu*ThermoDBP-RP1 may represent an auto-inhibitory conformation with respect to ssDNA binding. We attempted to test whether the NTD of *pfu*ThermoDBP-RP1 alone interacts more strongly with nucleic acids but failed to produce the protein in soluble form. The weak affinity for longer DNA oligonucleotides seen with *pfu*ThermoDBP-RP1 (Figure 1C) may indicate that the protein can bind ssDNA only in special contexts, where the presumed auto-inhibition is relieved, for example by other interacting proteins.

## DISCUSSION

### ThermoDBP-RP2 proteins can act as non-canonical SSBs

We have demonstrated that the NTD, conserved among ThermoDBPs and ThermoDBP-RPs, is a versatile ssDNA-binding domain in ThermoDBP-RP2 proteins *in vitro*. ssDNA can come to lie on various sites and in different orientations on multiple copies of this domain in an oligomeric protein. While these results show that ThermoDBP-RP2 proteins can function as SSBs *in vitro*, a similar activity *in vivo* remains to be demonstrated.

The fold and mode of ssDNA binding of ThermoDBP-RP2 differ markedly from previously characterized SSBs (3). Canonical SSBs typically bind ssDNA or ssRNA on their outer surfaces with the bases pointing towards the protein and the backbone solvent exposed. In the case of bacteriophage T4 gene 32 protein, ssDNA binds along a deep channel formed between two subdomains (28). In contrast, a ThermoDBP-RP2 tetramer binds ssDNA inside an internal, intersubunit tunnel system. This unprecedented binding mode may be particularly useful to protect ssDNA in challenging environments such as high temperatures.

### Molecular mechanism of ssDNA engagement

Our crystal structure of an *ape*ThermoDBP-RP2-DNA complex shows directly that the protein can bind short or nicked ssDNAs. Such a function could be required to intermittently hold on to DNA ends during recombination or repair events. It is even imaginable that such an activity could be used to detect linear, foreign DNA molecules as part of a simple immune system, but there is no direct evidence supporting this latter idea.

On the other hand, we could show that *ape*ThermoDBP-RP2 can bind efficiently to a circular ssDNA molecule that lacks free termini, suggesting that it could also substitute for canonical SSBs in general ssDNA-binding functions. We suggest that the short oligomers bound in our co-crystal structure also mark the possible paths that a single, long ssDNA molecule would take through the tetramer. The DNA1 molecule most likely outlines a major part of the binding site, as it represents that largest coherent stretch of DNA that could be traced in the electron density. The 5'-end of that molecule emerges from the internal tunnel system and lies on the surface of the *ape*ThermoDBP-RP2 tetramer. Additional nucleotides on this side of the DNA most likely simply extend into the solvent. Additional nucleotides at 3'-end would come to lie in positions that in our crystal structure are occupied by the DNA2

molecule. Topologically, the 3'-end of DNA1 could be directly connected to the 3'-terminal two nucleotides (dT10' and dT11') of the DNA2 molecule. Thus, these terminal two nucleotides of DNA2 point out a possible exit path for an extended ssDNA molecule between the CTDs of subunits A and B and the NTDs of subunits A' and B (path 1; green dashed line or tube in Figure 3C and F). Alternatively, an extended DNA1 molecule could bind along the path outlined by nucleotides dT9'–dT5' of the DNA2 molecule but in opposite direction (path 2; yellow dashed line or tube in Figure 3C and F). Although in this scenario it would traverse the same surfaces of NTD<sup>A</sup> and NTD<sup>B'</sup> in opposite orientation as seen for DNA2 in the present structure, such a binding mode is not entirely unlikely given the versatile contacts that are seen between different NTDs and portions of the bound DNA molecules in our crystal structure. It is also supported by the quasi-continuous stacking interaction observed between the 3'-terminal nucleotide of DNA1 (dT10) and an internal nucleotide (dT9') of DNA2. Both putative paths for extended ssDNA molecules through the *ape*ThermoDBP-RP2 tetramer are lined with positive electrostatic surface potential on the protein (Figure 3C). However, due to ensuing steric hindrance, it would be impossible to bind two extended ssDNA molecules to an *ape*ThermoDBP-RP2 tetramer at the same time (e.g. along both of the two putative paths through the internal tunnel system).

Another immediate question is how an *ape*ThermoDBP-RP2 tetramer can engage a ssDNA stretch that lacks free 3'-ends and thus could not be threaded end-on into the tunnel system of the tetramer. Given the predicted lability of interfaces II and III of the *ape*ThermoDBP-RP2 tetramer, we envision that the protein complex can open up along these interfaces to engage extended ssDNA molecules at its center. The complex would not have to completely dissociate into tight dimers; opening of one interface III at the top or bottom of the tetramer would suffice to allow accommodation of an extended ssDNA molecule at the internal tunnel system.

### Proteins bearing ThermoDBP-RP2-like NTDs

Although the ThermoDBP-RP1 NTD closely resembles the NTD of ThermoDBP-RP2 proteins, the physiological function of ThermoDBP-RP1 proteins remains unclear. Our *in vitro* binding studies do not support, but also do not ultimately rule out, a specific ribosome or box C/D sRNA association of these proteins, as had been reported previously (8). Weak DNA binding and the observation that the equivalent of the main ssDNA-binding surface of the ThermoDBP-RP2 NTD is obstructed in the tetrameric organization of *pfu*ThermoDBP-RP1 suggest that the proteins may require special activatory mechanisms to engage ssDNA or other nucleic acids. Such activation (or relief of auto-inhibition) may be provided by other interacting proteins that force the *pfu*ThermoDBP-RP1 NTDs into a different relative arrangement with respect to the central four-helix bundle.

Gel filtration and sequence analyses suggested that ThermoDBPs form parallel dimers *via* a C-terminal leucine zipper motif (7). If these putative, parallel ThermoDBP dimers

resemble a parallel dimer of our present *pfu*ThermoDBP-RP1 structure (e.g. subunits A and D; Figure 2A and Supplementary Figure S5A), the main ThermoDBP-RP2-like ssDNA binding surface of the ThermoDBP NTDs would remain unobstructed (Supplementary Figure S5A; right panel). However, most of the ssDNA-contacting residues of the *ape*ThermoDBP-RP2 NTDs are poorly conserved in ThermoDBPs (Supplementary Figure S1), suggesting that ssDNA could also be bound *via* a different surface in the latter proteins. The NTDs in the presumed parallel ThermoDBP dimers also would not cluster as seen in the *ape*ThermoDBP-RP2 tetramer and must bind ssDNA non-cooperatively or cooperate in a different fashion in ssDNA binding than observed in *ape*ThermoDBP-RP2. Irrespective of the exact mode of ssDNA binding in ThermoDBPs, the observation of strong ssDNA binding by full-length *tte*ThermoDBP (7) indicates that ThermoDBPs are not auto-inhibited.

A search of the Protein Data Bank with the Dali server (29) using the NTD of *ape*ThermoDBP-RP2 revealed its similarity to a C-terminal dsDNA-binding domain in the eukaryotic DNA polymerase eta (r.m.s.d. of 1.75 Å for 97 C $\alpha$  atoms; Supplementary Figure S5B). Unlike the binding of the *ape*ThermoDBP-RP2 NTDs to ssDNA, the DNA polymerase eta dsDNA-binding domain employs its  $\beta$ -sheet surface to engage dsDNA along its major groove. Together, our analyses show that the ThermoDBP-RP2-like NTD provides a structural scaffold that can be adapted in diverse ways for ssDNA binding (as in the cases of ThermoDBPs and ThermoDBP-RP2s) or for dsDNA binding (as in the case of the C-terminal domain of DNA polymerase eta).

## ACCESSION NUMBERS

PDB IDs: 4PSL, 4PSM, 4PSN and 4PSO.

## SUPPLEMENTARY DATA

Supplementary Data is available at NAR Online, including [1–5].

## ACKNOWLEDGMENTS

We thank C. Alings for help in crystallization experiments. We accessed beamlines of the BESSY II (Berliner Elektronenspeicherring-Gesellschaft für Synchrotronstrahlung II) storage ring (Berlin, Germany) *via* the Joint Berlin MX-Laboratory sponsored by the Helmholtz Zentrum Berlin für Materialien und Energie, the Freie Universität Berlin, the Humboldt-Universität zu Berlin, the Max-Planck-Gesellschaft, and the Leibniz-Institut für Molekulare Pharmakologie. Funding for open access charge: Freie Universität Berlin Kaiserswerther.

*Conflict of interest statement.* None declared.

## REFERENCES

1. Pestryakov, P.E. and Lavrik, O.I. (2008) Mechanisms of single-stranded DNA-binding protein functioning in cellular DNA metabolism. *Biochemistry (Mosc)*, **73**, 1388–1404.

2. Shereda, R.D., Kozlov, A.G., Lohman, T.M., Cox, M.M. and Keck, J.L. (2008) SSB as an organizer/mobilizer of genome maintenance complexes. *Crit. Rev. Biochem. Mol. Biol.*, **43**, 289–318.
3. Dickey, T.H., Altschuler, S.E. and Wuttke, D.S. (2013) Single-stranded DNA-binding proteins: multiple domains for multiple functions. *Structure*, **21**, 1074–1084.
4. Sancar, A., Williams, K.R., Chase, J.W. and Rupp, W.D. (1981) Sequences of the *ssb* gene and protein. *Proc. Natl. Acad. Sci. U.S.A.*, **78**, 4274–4278.
5. Theobald, D.L., Mitton-Fry, R.M. and Wuttke, D.S. (2003) Nucleic acid recognition by OB-fold proteins. *Annu. Rev. Biophys. Biomol. Struct.*, **32**, 115–133.
6. Lohman, T.M. and Ferrari, M.E. (1994) Escherichia coli single-stranded DNA-binding protein: multiple DNA-binding modes and cooperativities. *Annu. Rev. Biochem.*, **63**, 527–570.
7. Paytubi, S., McMahon, S.A., Graham, S., Liu, H., Botting, C.H., Makarova, K.S., Koonin, E.V., Naismith, J.H. and White, M.F. (2012) Displacement of the canonical single-stranded DNA-binding protein in the Thermoproteales. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, E398–E405.
8. Ciammaruconi, A., Gorini, S. and Londei, P. (2008) A bifunctional archaeal protein that is a component of 30S ribosomal subunits and interacts with C/D box small RNAs. *Archaea*, **2**, 151–158.
9. Marquez, V., Frohlich, T., Armache, J.P., Sohmen, D., Donhofer, A., Mikolajka, A., Berninghausen, O., Thomm, M., Beckmann, R., Arnold, G.J. *et al.* (2011) Proteomic characterization of archaeal ribosomes reveals the presence of novel archaeal-specific ribosomal proteins. *J. Mol. Biol.*, **405**, 1215–1232.
10. Studier, F.W. (2005) Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.*, **41**, 207–234.
11. Kabsch, W. (2010) XDS. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 125–132.
12. Vonrhein, C., Blanc, E., Roversi, P. and Bricogne, G. (2007) Automated structure solution with autoSHARP. *Methods Mol. Biol.*, **364**, 215–230.
13. Vagin, A. and Teplyakov, A. (2000) An approach to multi-copy search in molecular replacement. *Acta Crystallogr. D Biol. Crystallogr.*, **56**, 1622–1624.
14. Sheldrick, G.M. (2010) Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 479–485.
15. Pape, T. and Schneider, T.R. (2004) HKL2MAP: a graphical user interface for macromolecular phasing with SXELX programs. *J. Appl. Cryst.*, **37**, 843–844.
16. McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C. and Read, R.J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.*, **40**, 658–674.
17. Emsley, P., Lohkamp, B., Scott, W.G. and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 486–501.
18. Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W. *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 213–221.
19. Afonine, P.V., Grosse-Kunstleve, R.W., Echols, N., Headd, J.J., Moriarty, N.W., Mustyakimov, M., Terwilliger, T.C., Urzhumtsev, A., Zwart, P.H. and Adams, P.D. (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D Biol. Crystallogr.*, **68**, 352–367.
20. Murshudov, G.N., Skubak, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A., Winn, M.D., Long, F. and Vagin, A.A. (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol. Crystallogr.*, **67**, 355–367.
21. Winn, M.D., Isupov, M.N. and Murshudov, G.N. (2001) Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr. D Biol. Crystallogr.*, **57**, 122–133.
22. Chen, V.B., Arendall, W.B. III, Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S. and Richardson, D.C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 12–21.
23. DeLano, W.L. (2002) The PyMOL molecular graphics system. DeLano Scientific, San Carlos, CA.

24. Baker, N.A., Sept, D., Joseph, S., Holst, M.J. and McCammon, J.A. (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U.S.A.*, **98**, 10037–10041.
25. Krissinel, E. and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.*, **372**, 774–797.
26. Gellert, M., Lipsett, M.N. and Davies, D.R. (1962) Helix formation by guanylic acid. *Proc. Natl. Acad. Sci. U.S.A.*, **48**, 2013–2018.
27. Saenger, W., Riecke, J. and Suck, D. (1975) A structural model for the polyadenylic acid single helix. *J. Mol. Biol.*, **93**, 529–534.
28. Shamo, Y., Friedman, A.M., Parsons, M.R., Konigsberg, W.H. and Steitz, T.A. (1995) Crystal structure of a replication fork single-stranded DNA binding protein (T4 gp32) complexed to DNA. *Nature*, **376**, 362–366.
29. Holm, L. and Rosenstrom, P. (2010) Dali server: conservation mapping in 3D. *Nucleic Acids Res.*, **38**, W545–W 549.



## Supplementary Information

### Entrapment of DNA in an intersubunit tunnel system of a single-stranded DNA-binding protein

Homa Ghalei<sup>1,2,†,#</sup>, Holger von Moeller<sup>1,a,#</sup>, Detlef Eppers<sup>1</sup>, Daniel Sohmen<sup>3</sup>, Daniel N. Wilson<sup>3</sup>, Bernhard Loll<sup>1</sup>, and Markus C. Wahl<sup>1,\*</sup>

<sup>1</sup> Freie Universität Berlin, Fachbereich Biologie, Chemie, Pharmazie, Institut für Chemie und Biochemie, AG Strukturbiochemie, Takustr. 6, 14195 Berlin, Germany.

<sup>2</sup> Max-Planck-Institute for Biophysical Chemistry, Department of Cellular Biochemistry, Am Fassberg 11, 37077 Göttingen, Germany.

<sup>3</sup> Gene Center, Department of Biochemistry and Center of integrated Protein Science Munich (CiPSM), University of Munich, Feodor-Lynenstr. 25, 81377 Munich, Germany

<sup>†</sup> Current address: Department of Cancer Biology, The Scripps Research Institute, Jupiter, FL 33458, USA.

<sup>a</sup> Current address: moloX GmbH, Takustr. 6, 14195 Berlin, Germany

<sup>#</sup> These authors contributed equally to this work.

<sup>\*</sup> Corresponding author:  
Email: mwahl@zedat.fu-berlin.de (M.C.W.)

## Supplementary Tables

**Supplementary Table S1.** DNA and RNA oligonucleotides of mixed sequence used in binding assays

Nucleic acid	Sequence
ssDNA <sup>+</sup>	5'-d(ATGTGGAAAATCTCTAGCAGT)-3'
ssDNA <sup>-</sup>	5'-d(ACTGCTAGAGATTTTCCACAT)-3'
ssRNA <sup>+</sup>	5'-r(AUGUGGAAAUCUCUAGCAGU)-3'
ssRNA <sup>-</sup>	5'-r(ACUGCUAGAGAUUUUCCACAU)-3'
45-mer ssDNA	5'-d(CTTGCTAGGACGGATCCCTCGAGGTTTTTTTTTTTTTTTTTTTT)-3'
mixed 20-mer ssDNA	5'-d(CATGGTCAGTTAGCAGGTTC)-3'
sR2 sRNA	5'-r(GGGGGAUGAUGAGUUUUUCCCUCACUCUGA UUAGUGAUGAGGAGCCGAUGCACUGACCUC)-3'
sR12 sRNA	5'-r(GGAGGGGAUGAUGAGCGUUUACCGGUCUGAGUUGUGAUGAUA CUGGCACUGUCUGACCUUCC)-3'

**Supplementary Table S2.** Diffraction data collection, phasing and refinement statistics

	<i>pfu</i> ThermoDBP-RP1			<i>ape</i> ThermoDBP-RP2	<i>ape</i> ThermoDBP-RP2-dT <sub>10</sub>	
Data Collection	Form 1		Form 2			
Wavelength (Å)	0.97962 (peak)	0.98003 (inflection)	0.97185 (remote)	1.0385	0.98000	0.91841
Space Group	<i>P</i> 6 <sub>4</sub>	<i>P</i> 6 <sub>4</sub>	<i>P</i> 6 <sub>4</sub>	<i>C</i> 222 <sub>1</sub>	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>
Unit Cell Parameters (Å)						
a	136.6	136.6	136.6	61.3	61.2	61.2
b	136.6	136.6	136.6	199.3	108.6	108.6
c	122.1	122.1	122.1	114.9	154.9	154.9
Resolution (Å) <sup>a</sup>	35.00 - 3.75 (3.85 - 3.75)	35.00 - 3.75 (3.85 - 3.75)	35.00 - 3.50 (3.59 - 3.50)	35.00 - 2.43 (2.49 - 2.43)	35.00 - 2.05 (2.10 - 2.05)	30.00 - 2.90 <sup>f</sup> (3.00 - 2.90)
Reflections						
Unique <sup>a</sup>	26034 (1914)	26103 (1940)	32116 (2387)	150600 (10942)	125145 (9242)	46414 (4433)
Completeness (%) <sup>a</sup>	99.6 (99.6)	99.6 (99.5)	99.7 (99.6)	99.3 (98.9)	99.9 (100.0)	99.8 (99.9)
Redundancy <sup>a</sup>	5.9 (5.9)	5.9 (5.9)	5.9 (5.9)	5.6 (5.7)	7.4 (5.9)	4.5 (4.6)
<i>I</i> / $\sigma$ ( <i>I</i> ) <sup>a</sup>	10.8 (2.9)	11.1 (2.7)	16.1 (2.9)	11.5 (4.2)	15.5 (3.8)	11.07 (1.23)
<i>R</i> <sub>meas</sub> <sup>a,b</sup>	0.149 (0.665)	0.152 (0.722)	0.094 (0.659)	0.105 (0.670)	0.090 (0.614)	0.194 (1.725)
<b>Phasing</b>						
Heavy Atoms	24				12	
FOM <sup>c</sup>	0.41					
Phasing Power <sup>c</sup> (isomorphous anomalous)	0.0 0.792	0.591 0.098	1.102 1.190			
<i>R</i> <sub>cullis</sub> <sup>c</sup>	0.792	0.976	0.792			
Overall CC <sup>d</sup>					0.168	
Pseudo free CC <sup>d</sup>					0.631	
CC <sub>all</sub> /CC <sub>weak</sub> <sup>a, d</sup>					0.278/0.165	
<b>Refinement</b>						
Resolution (Å) <sup>a</sup>	34.16 - 3.50 (3.59 - 3.50)			33.41 - 2.43 (2.52 - 2.43)	32.73 - 2.05 (2.08 - 2.05)	30.00 - 2.90 <sup>f</sup> (3.00 - 2.90)
Reflections						
Number <sup>a</sup>	15568 (1148)			26694 (2612)	63473 (2581)	46426 (2541)
Completeness (%) <sup>a</sup>	100.0 (100.0)			99.0 (99.0)	96.7 (93.0)	99.9 (100.0)
Test Set (%) <sup>a</sup>	5.0			5.0	5.0	5.0
<i>R</i> <sub>work</sub> <sup>a, e</sup>	20.7 (27.3)			22.3 (26.7)	17.9 (24.2)	26.7 (37.5)
<i>R</i> <sub>free</sub> <sup>a, e</sup>	26.7 (35.0)			27.6 (36.5)	21.6 (29.8)	31.1 (40.6)

Refined Residues						
Protein	590			586	881	1754
Water molecules	31			68	289	38
Ligands molecules	3 sulfate			4 sulfate	2 glycerol, 6 imidazole	1 phosphate
DNA						4 strands
Mean B-Factors (Å <sup>2</sup> )						
Wilson	69.5			37.5	28.8	56.2
Protein	69.4			61.3	37.4	59.3
Water molecules	39.9			46.6	40.8	44.72
Ligand molecules	164.0			112.6	53.3	88.7
DNA						98.2
Ramachandran Plot <sup>g</sup> (%)						
Favored	97.42			98.96	99.31	97.60
Outliers	0.17			0.0	0.0	0.0
R.m.s.d. Geometry						
Bond Lengths (Å)	0.009			0.005	0.008	0.004
Bond Angles (°)	1.158			0.992	1.082	0.909

<sup>a</sup> Data for the highest resolution shell in parentheses

<sup>b</sup>  $R_{\text{meas}} = \frac{\sum_h [n/(n-1)]^{1/2} \sum_i |I_h - I_{h,i}|}{\sum_h \sum_i I_{h,i}}$ , where  $I_h$  is the mean intensity of symmetry-equivalent reflections and  $n$  is the redundancy.

<sup>c</sup> Figure of merit (FOM), phasing power and  $R_{\text{cullis}}$  as output by SHARP (1)

<sup>d</sup> Correlation coefficients as output by ShelxD/E (2,3)

<sup>e</sup>  $R = \frac{\sum_{hkl} ||F_{\text{obs}}| - |F_{\text{calc}}||}{\sum_{hkl} |F_{\text{obs}}|}$ ;  $R_{\text{work}} - hkl \notin T$ ;  $R_{\text{free}} - hkl \in T$ ; T – test set

<sup>f</sup> observed anisotropy with resolution limits along  $a^* = 2.6 \text{ \AA}$ ,  $b^* = 2.9 \text{ \AA}$  and  $c^* = 3.1 \text{ \AA}$

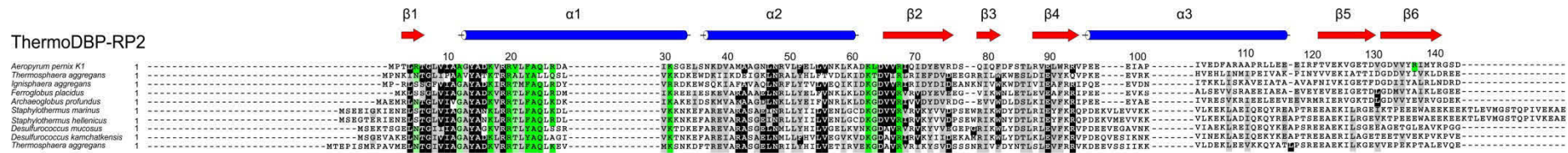
<sup>g</sup> Calculated with MolProbity (4)

**Supplementary Table S3.** List of protein-DNA contacts

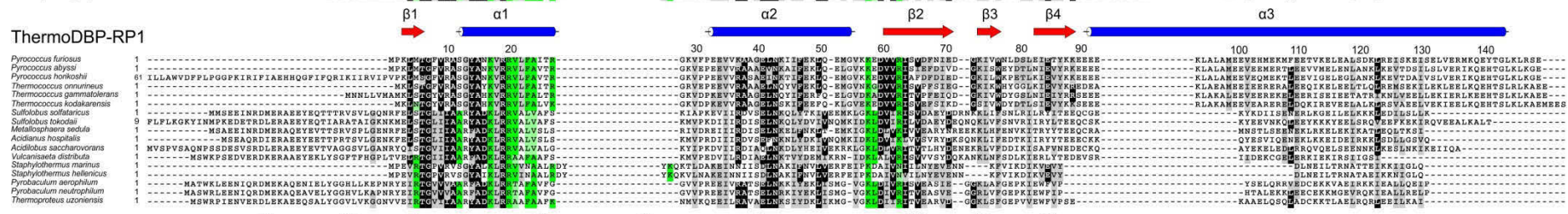
	<b>DNA1</b>	<b>DNA2</b>
<b>Sugars:</b>	Asn211 <sup>B</sup> - O4' dT1	Arg27 <sup>B'</sup> - O3' dT11'
	Asn211 <sup>B</sup> - sugar dT1	Val21 <sup>A'</sup> - sugar dT7'
	Val207 <sup>B</sup> - sugar dT2	Arg204 <sup>A'</sup> - O4' dT10'
	Leu64 <sup>A'</sup> - sugar dT4	
	Ala12 <sup>A''</sup> - sugar dT4	
	Arg27 <sup>A</sup> - O3' dT5	
	Lys31 <sup>A</sup> - sugar dT5	
	Arg20 <sup>A</sup> - sugar dT6	
	Val21 <sup>A</sup> - sugar dT7	
	Ala24 <sup>B</sup> - sugar dT9	
	Arg20 <sup>B'</sup> - O3' dT10	
<b>Phosphates:</b>	Arg137 <sup>B</sup> - O1' dT2	Arg68 <sup>B'</sup> - O1' dT6'
	Lys203 <sup>B</sup> - O1' dT3	Arg27 <sup>A'</sup> - O1' dT6'
	Arg68 <sup>B</sup> - O1' dT5	Lys17 <sup>A'</sup> - O1' dT8'
	Arg27 <sup>A</sup> - O2' dT6	
	Lys17 <sup>A</sup> - O1' dT8	
<b>Base:</b>	Asn211 <sup>B</sup> - base dT1	Arg204 <sup>A'</sup> - O2' dT10'
	Arg204 <sup>B</sup> - O2' dT6	Phe23 <sup>A'</sup> - base dT6'
	Phe23 <sup>A</sup> - base dT6	Ala24 <sup>A</sup> - base dT6
	Ala24 <sup>A</sup> - base dT7	Gln25 <sup>B'</sup> - O2, N3 dT8'
	Gln25 <sup>A</sup> - base dT8	Arg5 <sup>B'</sup> - O4, base dT6'
	Arg5 <sup>B</sup> - O2' dT8	Lys17 <sup>B'</sup> - O2, N3 dT9'
	Lys17 <sup>B</sup> - N3, O2' dT9	Val21 <sup>B'</sup> - base dT7'
	Val21 <sup>B</sup> - base dT9	

# Supplementary Figures

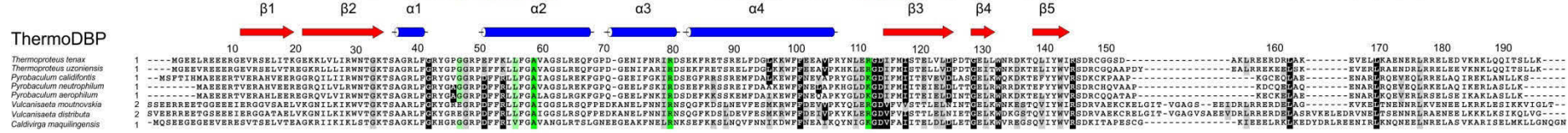
## ThermoDBP-RP2



## ThermoDBP-RP1

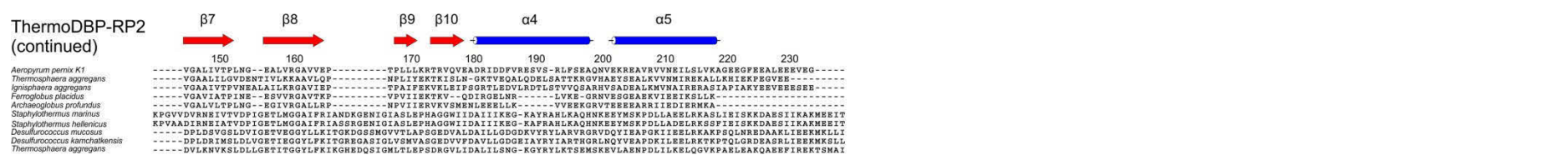


## ThermoDBP

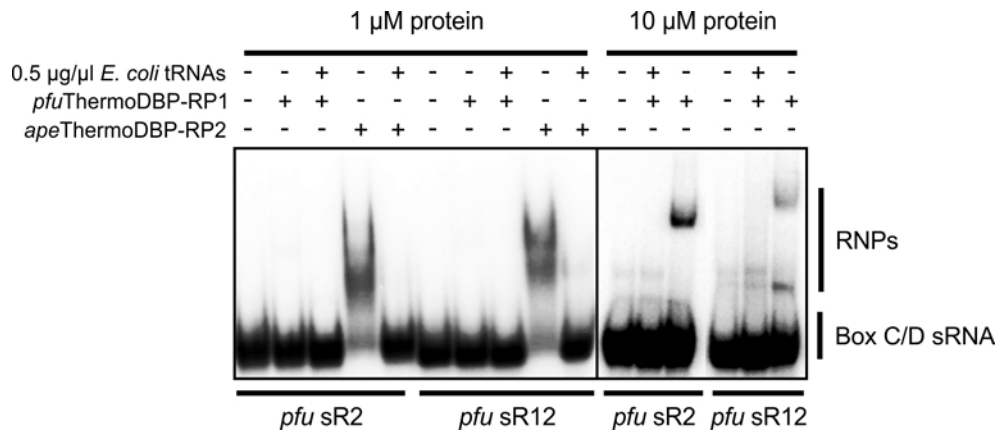


NTD

## ThermoDBP-RP2 (continued)

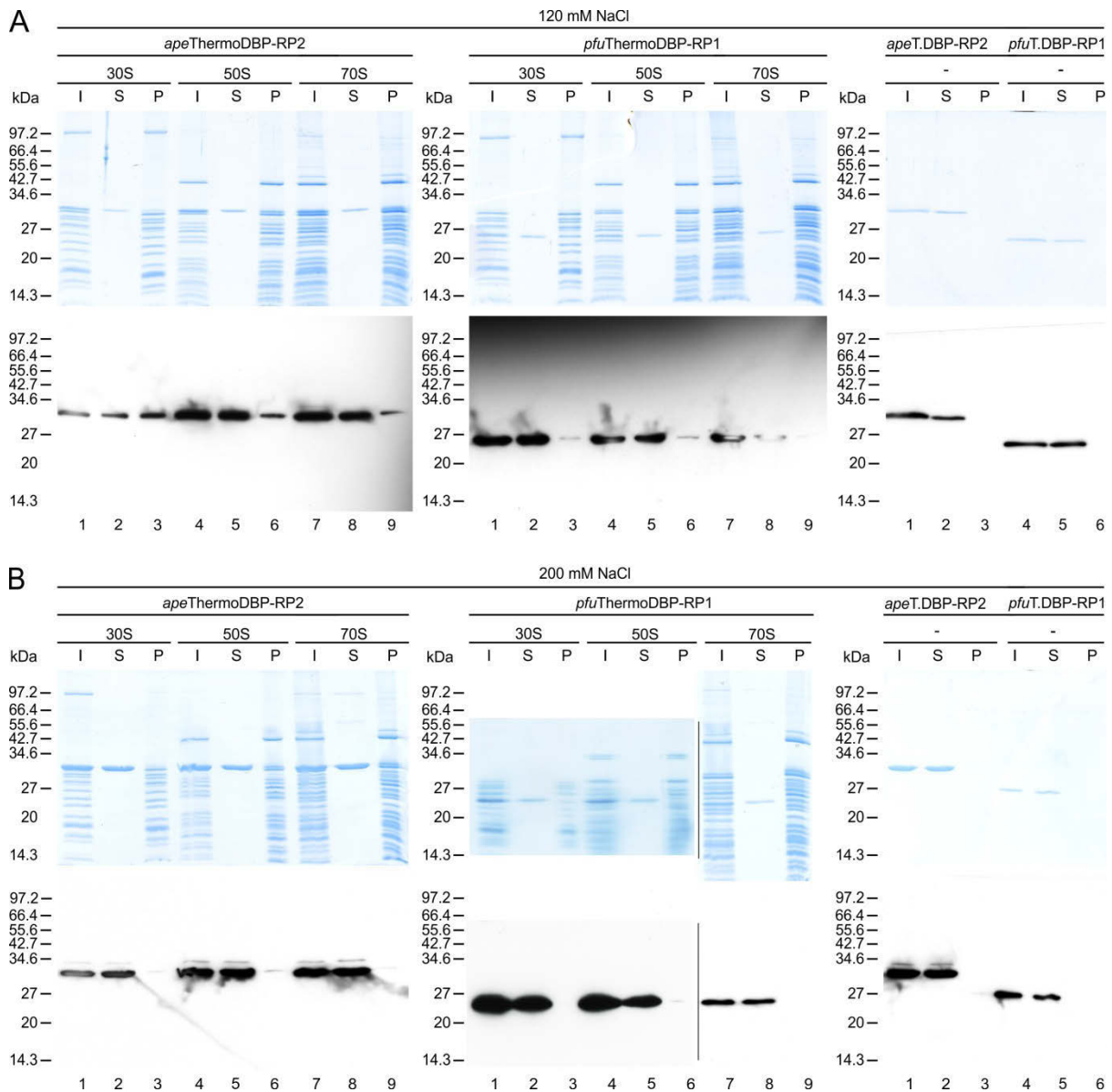


**Supplementary Figure S1.** Multiple sequence alignment. Multiple structure-based sequence alignment of the proteins containing the DUF2258 domain. Secondary structure elements of *ape*ThermoDBP-RP2, *pfu*ThermoDBP-RP1 and *tte*ThermoDBP are shown ( $\alpha$ -helices – blue;  $\beta$  strands – red). Darker background indicates higher conservation with respect to *ape*ThermoDBP-RP2. Residue with green background contact DNA in the *ape*ThermoDBP-RP2–dT<sub>10</sub> structure. Regions corresponding to the NTDs are indicated. Proteins listed: *Aeropyrum pernix* - APE\_1866.1, *Ignisphaera aggregans* - IGAG-0238, *Thermosphaera aggregans* - TAGG\_0476, *Desulfurococcus fermentans* - DESFE\_1124, *Desulfurococcus kamchatkensis* - DKAM\_1006, *Desulfurococcus mucosus* - DESMU\_0072, *Thermosphaera aggregans* - TAGG\_0069, *Staphylothermus marinus* - SMAR\_1160, SMAR\_0213, *Staphylothermus hellenicus* - SHELL\_1311, SHELL\_0606, *Thermogladius cellulolyticus* - TCELL\_0577, *Hyperthermus butylicus* - HBUT\_1217, *Pyrolobus fumarii* - PYRFU\_0856, *Acidilobus saccharovorans* - ASAC\_0853, ASAC\_0999, *Vulcanisaeta moutnovskia* - VMUT\_0250, *Pyrobaculum aerophilum* - PAE3173, *Pyrobaculum sp. 1860* - P186\_1371, *Pyrobaculum arsenaticum* - PARS\_1656, *Pyrobaculum islandicum* - PISL\_0538, *Thermoproteus neutrophilus* - TNEU\_1621, *Pyrobaculum calidifontis* - PCAL\_1811, *Thermoproteus uzoniensis* - TUZN\_1647, *Vulcanisaeta distributa* - VDIS\_1809, *Sulfolobus islandicus* - LS215\_1226, *Sulfolobus islandicus* - M1425\_1129, *Sulfolobus solfataricus* - SSO1098, SSOL\_2070, Ssol98\_010100004655, *Sulfolobus tokodaii* - ST0853, *Sulfolobus acidocaldarius* - SACI\_1216, *Acidianus hospitalis* - Ahos\_0932, *Metallosphaera sedula* - MSED\_1043, *Metallosphaera yellowstonensis* - METMK1DRAFT\_00029730, *Archaeoglobus profundus* - ARCPR\_1830, *Ferroglobus placidus* - FERP\_2293, *Thermococcus sp. AM4* - TAM4\_671, *Thermococcus gammatolerans* - TGAM\_0941, *Thermococcus zilligii* - TZILA\_05110, *Thermococcus kodakarensis* - TK0813, *Thermococcus onnurineus* - TON\_0781, *Thermococcus sibiricus* - TSIB\_0836, *Thermococcus litoralis* - OCC\_01989, *Thermococcus barophilus* - TERMP\_00659, *Pyrococcus yayanosii* - PYCH\_06080, *Pyrococcus horikoshii* - PH1118, *Pyrococcus abyssi* - PAB1631, *Pyrococcus furiosus* - PF1044

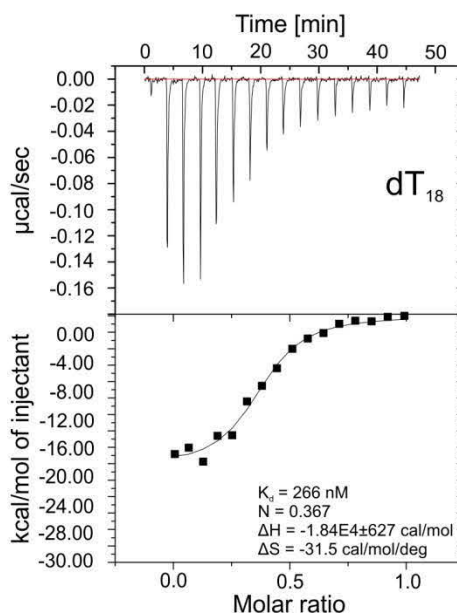
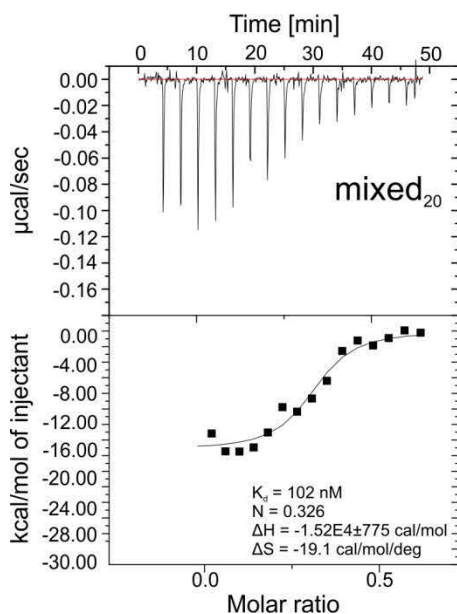
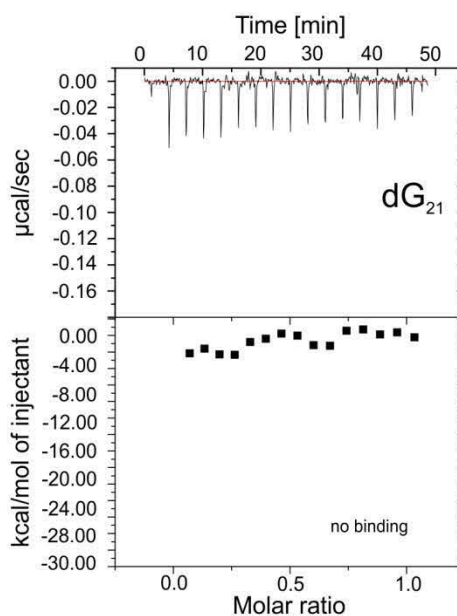
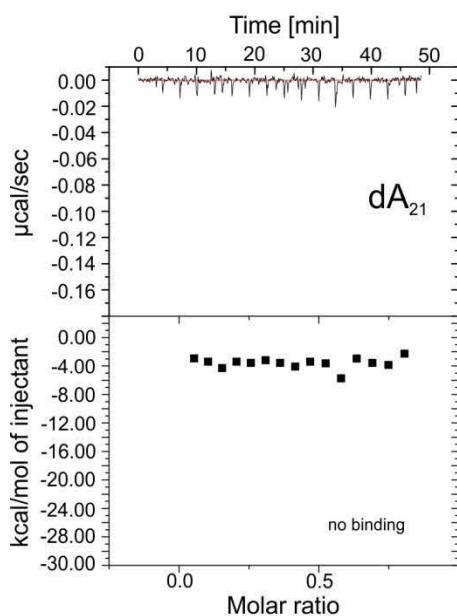
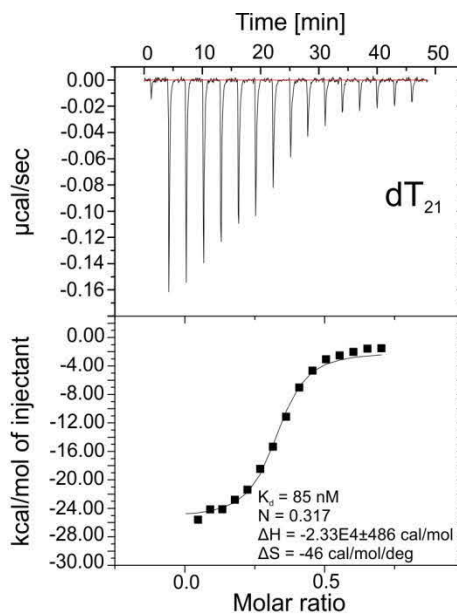
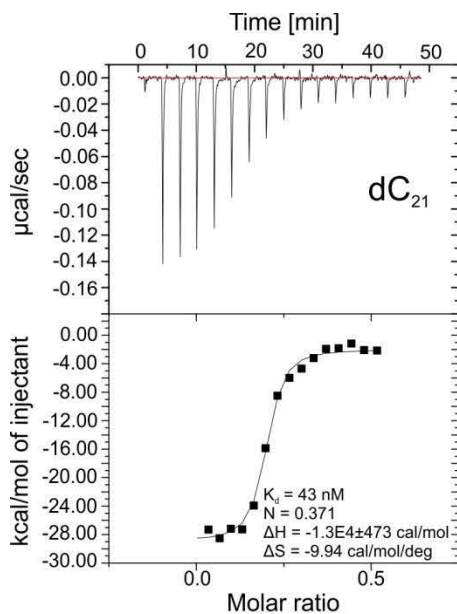


**Supplementary Figure S2.** Binding of ThermoDBP-RPs to box C/D sRNAs. Electrophoretic gel mobility shift assays testing the binding of *pfu*ThermoDBP-RP1 and of *ape*ThermoDBP-RP2 to sR2 and sR12 box C/D sRNAs from *Pyrococcus furiosus*.

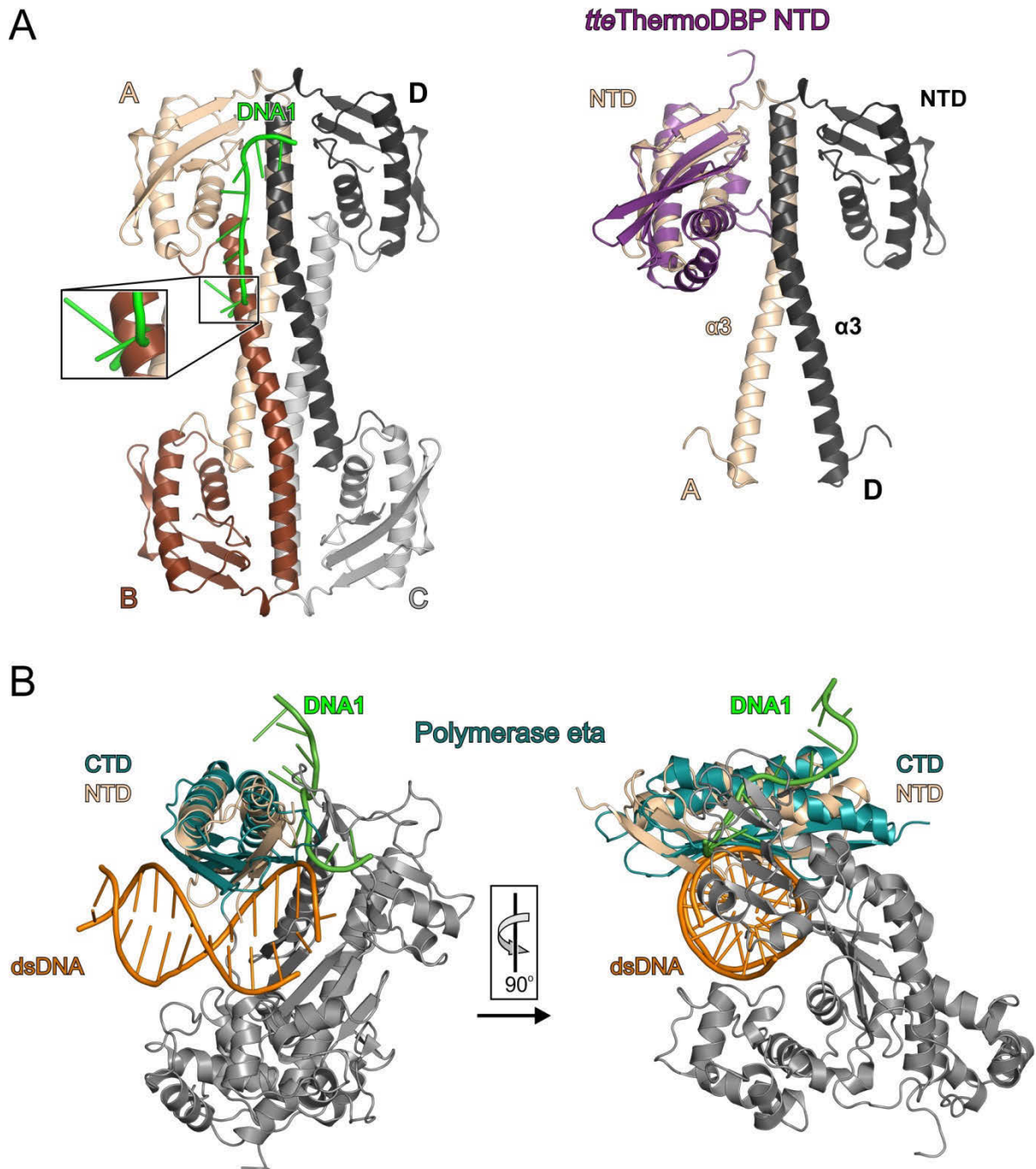




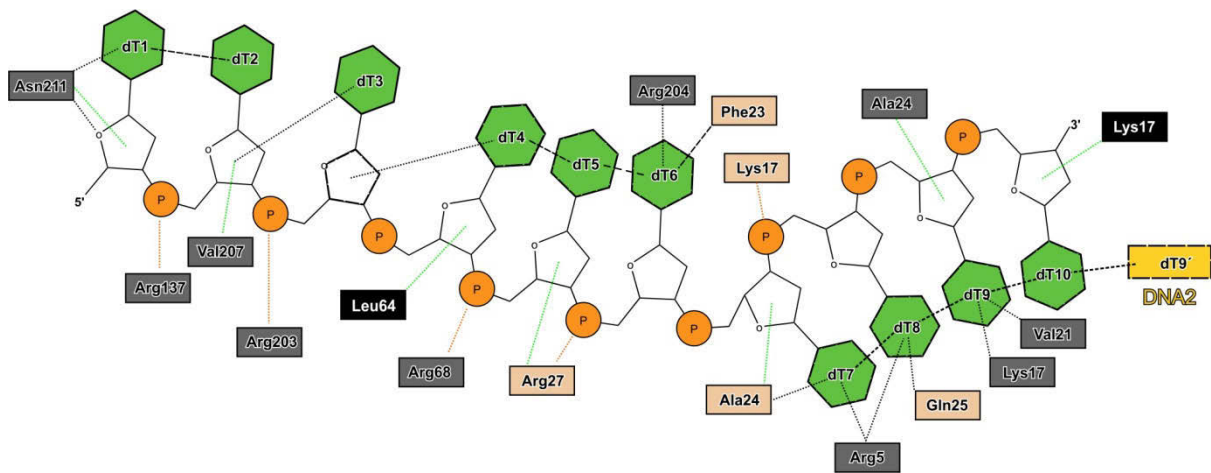
**Supplementary Figure S3.** Interactions of ThermoDBP-RPs with ribosomal subunits and ribosomes. Interactions were probed by co-sedimentation binding assays of ThermoDBP-RPs and 30S, 50S and 70S ribosomal particles of *Thermococcus kodakarensis*. Ribosomal particles were isolated from *T. kodakarensis* lysates as previously described (5) and incubated in the presence of 120 mM NaCl (**A**) or 200 mM NaCl (**B**) together with purified ThermoDBP-RPs from *A. pernix* and *P. furiosus* and subsequently separated by ultracentrifugation. Upper panels: Proteins were separated by SDS-PAGE and stained with Coomassie Brilliant Blue. Lower panels: Proteins, after separation by SDS-PAGE, were transferred onto a nitrocellulose membrane for immuno-detection using an anti-His<sub>6</sub> antibody. Signals for proteins co-sedimented with ribosomal particles are detected in the pellet fractions (P). Whole-reaction-samples and samples after ribosomal particles separation are labeled (IN) and (SN), respectively.



**Supplementary Figure S4.** Quantification of *ape*ThermoDBP-RP2-ssDNA interactions by ITC. Both the raw data and the integrated data are shown. Oligonucleotide sequences are listed in Supplementary Table S3. Data were fitted based on the “One Set of Sites” model. The stoichiometry (N) gives the number of DNA molecules bound per *ape*ThermoDBP-RP2 monomer.



**Supplementary Figure S5.** Structural comparisons. **(A)** Cartoon view of *pfu*ThermoDBP-RP1 superposed with *tte*ThermoDBP NTD (left panel; *tte*ThermoDBP NTD – magenta; PDB ID 3TEK) and one ssDNA molecule from the *ape*ThermoDBP-RP2 co-crystal structure (right panel; DNA - green). Additional helices of *tte*ThermoDBP NTD do not clash with the helical extensions of *pfu*ThermoDBP-RP1 (left). However, a ssDNA bound on subunit A of a *pfu*ThermoDBP-RP1 tetramer in the same conformation as observed in the *ape*ThermoDBP-RP2 structure would clash with helix  $\alpha 3$  of subunit B. **(B)** Superposition of yeast polymerase eta (grey; C-terminal domain – cyan; PDB ID 3OHB) with the NTD of *ape*ThermoDBP-RP2 (beige). A ssDNA molecule as observed in the *ape*ThermoDBP-RP2 structure is colored in green. dsDNA (orange) binds to a different site on the polymerase eta CTD, which is structurally related to the NTDs of ThermoDBP and ThermoDBP-RP proteins.



**Supplementary Figure S6.** Schematic representation of the DNA contacts. Schematic representation of the interactions of *apeThermoDBP-RP2* with one  $dT_{10}$  oligonucleotide. Color coding as in Figure 3B and C. Orange dashed lines indicate protein-phosphate contacts, green dashed lines are protein-sugar contacts and red (polar) and black (hydrophobic) dashed lines represent protein-nucleobase interactions. Black slashed lines represent stacking interactions.

## Supplementary References

1. Vonrhein, C., Blanc, E., Roversi, P. and Bricogne, G. (2007) Automated structure solution with autoSHARP. *Methods Mol Biol*, **364**, 215-230.
2. Pape, T. and Schneider, T.R. (2004) HKL2MAP: a graphical user interface for macromolecular phasing with SXELX programs. *J Appl Cryst*, **37**, 843-844.
3. Sheldrick, G.M. (2010) Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr D Biol Crystallogr*, **66**, 479-485.
4. Chen, V.B., Arendall, W.B., 3rd, Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S. and Richardson, D.C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr*, **66**, 12-21.
5. Marquez, V., Frohlich, T., Armache, J.P., Sohmen, D., Donhofer, A., Mikolajka, A., Berninghausen, O., Thomm, M., Beckmann, R., Arnold, G.J. *et al.* (2011) Proteomic characterization of archaeal ribosomes reveals the presence of novel archaeal-specific ribosomal proteins. *J Mol Biol*, **405**, 1215-1232.

ARTICLE

Received 12 Jan 2015 | Accepted 16 Mar 2015 | Published 23 Apr 2015

DOI: 10.1038/ncomms7941

OPEN

# Structure of the *Bacillus subtilis* 70S ribosome reveals the basis for species-specific stalling

Daniel Sohmen<sup>1</sup>, Shinobu Chiba<sup>2</sup>, Naomi Shimokawa-Chiba<sup>2</sup>, C. Axel Innis<sup>3,4</sup>, Otto Berninghausen<sup>1</sup>, Roland Beckmann<sup>1,5</sup>, Koreaki Ito<sup>2</sup> & Daniel N. Wilson<sup>1,5</sup>

Ribosomal stalling is used to regulate gene expression and can occur in a species-specific manner. Stalling during translation of the MifM leader peptide regulates expression of the downstream membrane protein biogenesis factor YidC2 (YqjG) in *Bacillus subtilis*, but not in *Escherichia coli*. In the absence of structures of Gram-positive bacterial ribosomes, a molecular basis for species-specific stalling has remained unclear. Here we present the structure of a Gram-positive *B. subtilis* MifM-stalled 70S ribosome at 3.5–3.9 Å, revealing a network of interactions between MifM and the ribosomal tunnel, which stabilize a non-productive conformation of the PTC that prevents aminoacyl-tRNA accommodation and thereby induces translational arrest. Complementary genetic analyses identify a single amino acid within ribosomal protein L22 that dictates the species specificity of the stalling event. Such insights expand our understanding of how the synergism between the ribosome and the nascent chain is utilized to modulate the translome in a species-specific manner.

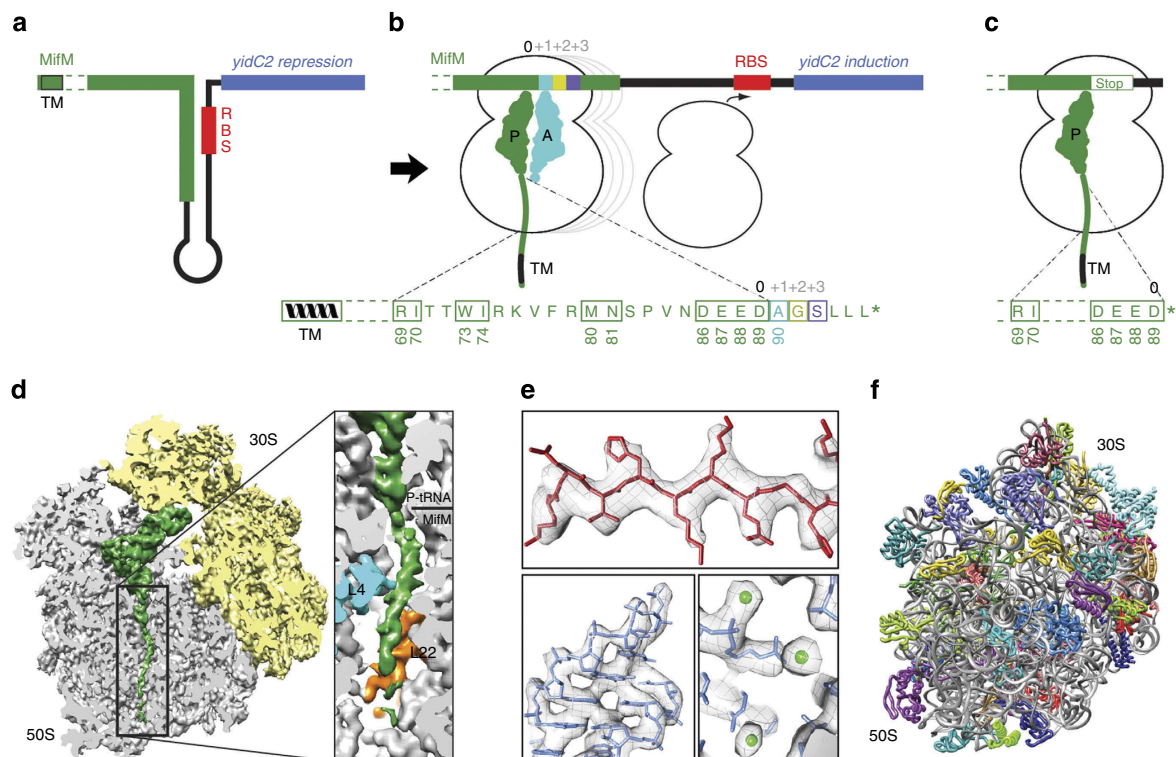
<sup>1</sup>Gene Center and Department for Biochemistry, University of Munich, Feodor-Lynen-Street 25, Munich 81377, Germany. <sup>2</sup>Faculty of Life Sciences, Kyoto Sangyo University, Motoyama, Kamigamo, Kita-Ku, Kyoto 603-8555, Japan. <sup>3</sup>Institut Européen de Chimie et Biologie, Université de Bordeaux, Pessac, France. <sup>4</sup>Institut National de la Santé et de la Recherche Médicale (U869), Bordeaux, France. <sup>5</sup>Center for integrated Protein Science Munich (CiPSM), University of Munich, Feodor-Lynen-Street 25, Munich 81377, Germany. Correspondence and requests for materials should be addressed to D.N.W. (email: wilson@lmb.uni-muenchen.de).

The ribosome is the major protein-synthesizing machine in the cell that converts the genetic information present within the codons of the mRNA into an amino-acid sequence within the polypeptide chain<sup>1</sup>. Ribosomes can transiently pause or even become stalled for prolonged periods during the translation of mRNAs due to many intrinsic factors, such as secondary structure within the mRNA<sup>2</sup> or the presence of particular amino-acid sequences within the nascent polypeptide chain, such as polyproline sequences<sup>3,4</sup>. In many cases, ribosome stalling occurs during translation of short upstream open reading frames, so-called leader peptides, which is utilized by both bacteria and eukaryotes to regulate expression of downstream genes<sup>5,6</sup>. These regulatory events can be ligand-dependent, as exemplified by the ErmCL and TnaC leader peptides that require the presence of a macrolide antibiotic or free tryptophan, respectively, to induce ribosome stalling<sup>7,8</sup>, or ligand independent, with well-characterized examples including the *E. coli* SecM and *B. subtilis* MifM leader peptides, where translation of the amino-acid sequence is sufficient to induce the translational arrest<sup>9–11</sup>.

In *B. subtilis*, the MifM leader peptide is located upstream of the *yidC2* (*yqjG*) gene (Fig. 1a), which cannot be translated independently of MifM because the ribosome-binding site of *yidC2* is sequestered within a stem-loop structure in the mRNA (Fig. 1a)<sup>10</sup>. Ribosome stalling during translation of MifM leads to unwinding of the stem-loop structure and maintains the ribosome-binding site accessible for ribosome binding and induction of YidC2 expression (Fig. 1b). YidC2 is a homologue

of the constitutively expressed SpoIIIJ (YidC1), a protein involved in membrane protein insertion and folding<sup>12</sup>. The *B. subtilis* MifM leader peptide is 95 amino-acids long and comprises a C-terminal region (residues 69–89) that is critical for ribosome stalling as well as an N-terminal transmembrane (TM) segment that targets the MifM peptide for membrane insertion, presumably via SpoIIIJ<sup>10</sup>. Interaction between SpoIIIJ and the TM segment of MifM as it emerges from the ribosomal tunnel is thought to prevent ribosome stalling by providing a pulling force on the MifM nascent chain<sup>11</sup>, analogous to SecA relief of SecM stalling<sup>13–15</sup>. Subsequently, canonical translation termination and ribosome-recycling ensues, leading to rapid refolding of the mRNA and repression of YidC2 expression (Fig. 1a). In contrast, when cellular levels of SpoIIIJ are low, ribosome stalling occurs on MifM, maintaining the unfolded conformation of the mRNA and thereby promoting expression of YidC2. In this manner, *B. subtilis* ensures that sufficient levels of SpoIIIJ or YidC2 are present in the membrane to direct membrane protein insertion and/or protein folding<sup>10,16</sup>.

Biochemical studies have demonstrated that four major ribosome-stalling sites are present in the MifM leader peptide; the first occurring when the codon for residue D89 is present in the P-site, then ribosomes stall at the following three codon positions corresponding to residues A90, G91 and S92 (Fig. 1b)<sup>6</sup>. Mutagenesis studies have identified six residues (R69, I70, W73, I74, M80 and N81) as well as the negatively charged DEED sequence (residues 86–89) within the C-terminal region of MifM (Fig. 1b) that are important for ribosome stalling<sup>6,10</sup>. Despite the



**Figure 1 | Cryo-EM structure and molecular model of the *B. subtilis* MifM-stalled ribosome complex.** (a,b) Schematic of the *mifM-yidC2* mRNA illustrating the N-terminal transmembrane (TM) segment (black, helix) and C-terminal stalling region (green) of the MifM leader peptide with the stem-loop structure that sequesters the ribosome-binding site (RBS) of the *yidC2* gene (blue). In (b) the multisite ribosome stalling (0, +1, +2 and +3) during translation of MifM maintains the unfolded conformation of the mRNA allowing ribosome binding and induction of *yidC2* expression. The MifM stalling sequence (residues 69–89) is shown with critical residues boxed in green. Asterisks indicate stop codons. (c) MifM-stalled ribosome complex used for cryo-EM. (d) Transverse section of the cryo-EM structure of the MifM-SRC (30S, yellow; 50S, grey) showing P-tRNA and MifM nascent chain (green) within the ribosomal tunnel and enlargement where ribosomal proteins L4 (cyan) and L22 (orange) are coloured. (e) Electron density (grey mesh) for selected regions of large subunit ribosomal protein and rRNA of the MifM-SRC. (f) Molecular model of the *B. subtilis* 70S ribosome.



high conservation of the ribosomal tunnel, translational stalling by MifM occurs on *B. subtilis* ribosomes, but not efficiently on *E. coli* ribosomes<sup>11</sup>; however, a molecular basis for this species specificity has remained unclear.

Here we have developed a *B. subtilis* cell-free *in vitro* translation system, which was subsequently used to generate a *B. subtilis* MifM-stalled ribosome complex (MifM-SRC). We present a cryoelectron microscopy (cryo-EM) reconstruction of the MifM-SRC at 3.5–3.9 Å resolution, which was used to build the first molecular model of a Gram-positive 70S ribosome. The only contacts of the MifM nascent chain with the ribosome that are not conserved between *B. subtilis* and *E. coli* are located within the lower region of the exit tunnel where they involve ribosomal protein L22. Consistently, our complementary genetic mutagenesis studies identified a single amino-acid residue, Met90 of L22, which modulates the specificity of MifM-dependent stalling. Interaction of the MifM nascent chain within the ribosomal tunnel positions the sidechain of Glu88 of MifM in a manner that prevents accommodation of an incoming aminoacyl-transfer RNA (tRNA) at the peptidyltransferase centre (PTC) of the ribosome, thus providing a structural basis for the MifM-dependent translational arrest.

## Results

**Cryo-EM structure of *B. subtilis* MifM-SRC.** To understand the structural basis for the mechanism of MifM-dependent translational stalling and provide structural insight into the species specificity of MifM stalling, we set out to structurally characterize the *B. subtilis* MifM-SRC. Unlike for *E. coli*, translation systems for *B. subtilis* are not commercially available. Therefore, to generate complexes of translating *B. subtilis* ribosomes, we developed and optimized a *B. subtilis* S12 lysate-based *in vitro* coupled transcription–translation system (Supplementary Fig. 1 and Methods). This system was used to prepare *B. subtilis* MifM-SRCs by translation of a template encoding the *B. subtilis* MifM leader peptide (Supplementary Fig. 1 and Methods). Homogeneity of the MifM-SRC was increased by replacing Ala90 of MifM with a UAA stop codon (Supplementary Fig. 1), which arrests ribosomes at the first stall site (Fig. 1c), and thereby prevents further translation and subsequent arrest at downstream stall sites<sup>6</sup>. Like the wildtype MifM-SRC, the MifM-Ala90UAA stalled ribosomes are refractory to the action of puromycin, and are also not subject to the action of the termination release factors RF1 and RF2<sup>6,10</sup>. This MifM-SRC (Fig. 1c) was separated from non-translating ribosomes and polysomes using sucrose gradient centrifugation and affinity chromatography, as performed previously for other SRCs<sup>17,18</sup>. A single MifM-peptidyl-tRNA product, and no free MifM peptide, was detected in the final MifM-SRC (Supplementary Fig. 1), indicating the homogeneity of the sample and its suitability for structural analysis. A single-particle reconstruction of the MifM-SRC was obtained from 305,045 particles using cryo-EM (Fig. 1d). The MifM-SRC has an average resolution of 3.9 Å, with local resolution calculations<sup>19</sup> indicating that the resolution of the core of the 50S subunit reaches 3.5 Å (Supplementary Fig. 2). This is consistent with the features of the cryo-EM map, such as strand separation in  $\beta$ -sheets, the pitch of  $\alpha$ -helices as well as density for many side chains of the ribosomal proteins (Fig. 1e). In addition, the rRNA backbone and nucleotides are well resolved, as well as the position of many putative magnesium ions (Fig. 1e). The distinct features of the electron density (Supplementary Figs 3 and 4), together with the high homology to *E. coli* and *T. thermophilus* 70S ribosomal structures<sup>20,21</sup>, allowed a molecular model of the *B. subtilis* 70S ribosome to be built (Fig. 1f, Table 1 and Supplementary Figs 5 and 6), encompassing a total of 4,579 of the

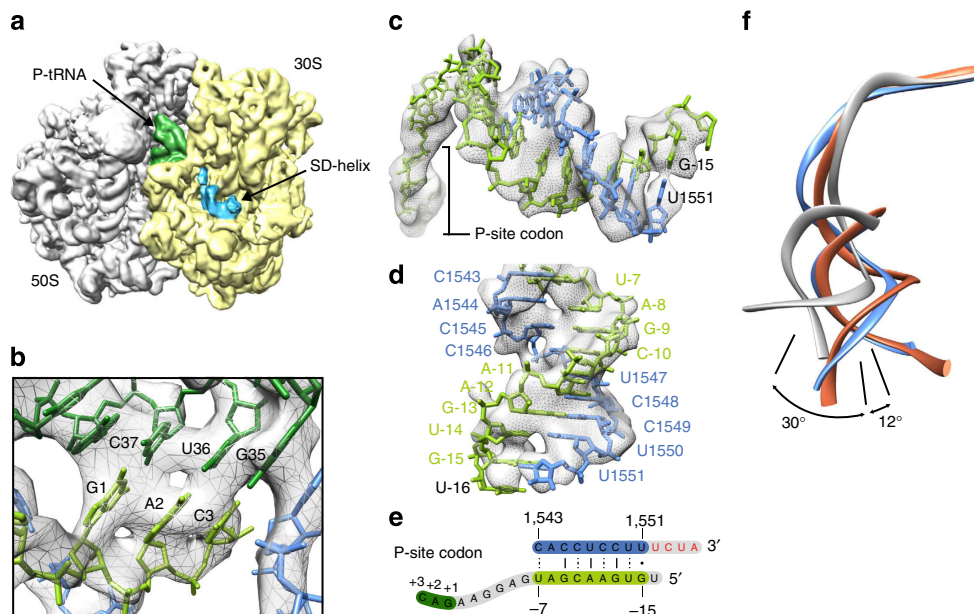
**Table 1 | Refinement and Model Statistics.**

<i>Data collection and refinement</i>	
Particles	305,045
Pixel size (Å)	1.108
Defocus range ( $\mu\text{m}$ )	1.0–2.5
Voltage (kV)	300
Electron dose ( $\text{e}^-/\text{Å}^{-2}$ )	28
Map sharpening B factor ( $\text{Å}^2$ )	–124.35
Resolution (Å, 0.143 FSC)	3.9
<i>Model Composition</i>	
Non-hydrogen atoms	135,413
Protein residues	5660
RNA bases	4675
<i>R.m.s. deviations</i>	
Bonds (Å)	0.0094
Angles (Å)	1.46
<i>Validation (proteins)</i>	
Molprobrity score (79th percentile)	1.94
Clashscore, all atoms (80th percentile)	8.38
Good rotamers (%)	99.85
<i>Ramachandran plot</i>	
Favored (%)	91.95
Outliers (%)	1.98
<i>Validation (RNA)</i>	
Correct sugar puckers (%)	97.75
Good backbone conformation (%)	100

FSC, Fourier Shell Correlation; R.m.s., root-mean square.

4,602 (99.5%) nucleotides of the 5S, 16S and 23S rRNAs, 19 small subunit ribosomal proteins and 29 of the 32 large subunit proteins (L1, L7/L12 and L9 were excluded). The slightly lower resolution of the small subunit (4.0 Å, Supplementary Fig. 2) restricted the modelling of the small subunit ribosomal proteins to a backbone trace, whereas sidechains were modelled for the large subunit proteins.

The majority of the rRNA in the core of *B. subtilis* 70S ribosome is structurally conserved with *E. coli* and *T. thermophilus* ribosomes, however, specific regions resemble only *E. coli* (h26, h33, h44 in the 30S; H27, H68 and H79 in the 50S) or *T. thermophilus* (h9 in the 30S; H15 and H63 in the 50S), whereas others appear to be *Bacillus* specific (h6, h10, h17 in the 30S; H25 and H54 in the 50S; Supplementary Fig. 8). As expected from the genome sequence, the *B. subtilis* 70S ribosome lacks density for S21 (Supplementary Fig. 9), which is present in *E. coli* but absent in *T. thermophilus*. Surprisingly, density for L9 and L25 (also referred to as the general stress protein CTC) is also absent in the *B. subtilis* 70S ribosome (Supplementary Fig. 9), despite L9 and L25 being encoded in the *B. subtilis* genome. Only one tRNA is present in the MifM-SRC, located at the P-site, where it makes codon–anticodon interactions with P-site codon of the mRNA (Fig. 2a,b). No density for the mRNA is observed in the A-site, nor within the downstream mRNA channel, whereas in contrast, the upstream 5' end of the mRNA can be unambiguously traced from the P-site through the E-site to the back of 30S subunit where it establishes a number of non-canonical base-pairs with the 3' end of the 16S rRNA resulting in a 8-basepair helix (Fig. 2c–e), analogous to the Shine–Dalgarno (SD)-anti-SD helices observed during elongation<sup>22,23</sup> (Fig. 2f). The CCA-end of the P-tRNA adopts its canonical position at the PTC of the ribosome, with clear density for the MifM nascent polypeptide chain extending from the ribose of A76 and traversing the entire



**Figure 2 | The path of the mRNA through the *B. subtilis* MifM-SRC.** (a) Location of Shine-Dalgarno(SD)-anti-SD-like helix on 70S ribosome (30S, yellow; 50S, grey, P-tRNA, green). (b) Codon-anticodon interaction between P-tRNA (dark green) and mRNA (pale green). (c) Electron density (grey mesh) and molecular model for the path of the *mifM* mRNA (green) from the P-site to back of the 30S subunit (d) where it forms an 8 base pair SD-anti-SD-like helix with the 3' end of the 16S rRNA (blue). (e) Schematic for the non-canonical base-pairing observed in the SD-anti-SD-like helix of the MifM-SRC. (f) Comparison of SD-anti-SD-like helix observed in *B. subtilis* MifM-SRC (blue) with initiation SD-anti-SD helix (grey) and post-initiation SD-anti-SD helix (red) observed previously<sup>22</sup>.

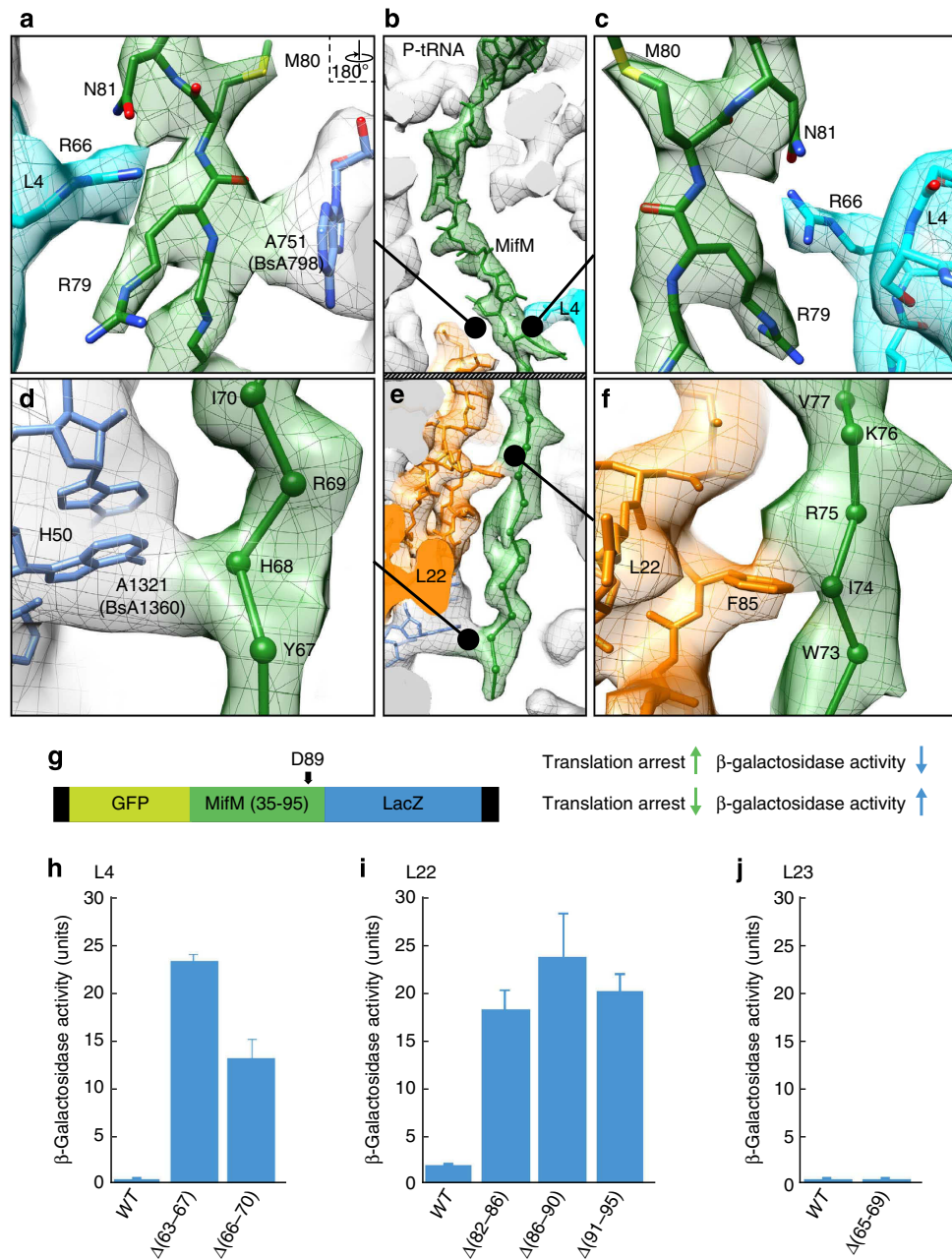
length of the ribosomal exit tunnel (Fig. 1b). Local resolution calculations indicate that the MifM nascent chain exhibits some degree of flexibility (Supplementary Fig. 2), particularly in regions that do not appear to establish contacts with components of the ribosomal tunnel.

**MifM interactions with the ribosome.** In the upper region of the tunnel, the density for the MifM nascent chain is well resolved at the sites of contact, namely adjacent to the PTC and at the constriction, enabling side chains of MifM residues E87/D89 and R79-N81 to be modelled (Fig. 3a–c). In contrast, the MifM nascent chain becomes fragmented in the lower region of the tunnel, indicating a higher degree of flexibility, and was therefore filtered to 4 Å to obtain continuous uninterrupted electron density (Fig. 3d–f). For the less resolved regions, that is, S82-E86 and W73-F78, we present a model for the backbone trace and acknowledge that the register for MifM may become lost after the constriction in the lower regions of the tunnel. The model localizes the three pairs of critical MifM residues, M80-N81, W73-I74 and R69-I70, at the sites of contact with components of the ribosomal tunnel (Fig. 3b,f).

In addition to the PTC (discussed later), four main interactions are observed between MifM and the ribosome: Two contacts are observed at the constriction, which sandwich the MifM nascent chain between Arg66 of L4 on one side of the tunnel and A751 (*E. coli* numbering is used throughout the text) of the 23S rRNA on the other (Fig. 3a–c). The contact to A751 is one of the strongest fusions of density between the tunnel wall and MifM, and appears to involve the backbone in the vicinity of R79 of MifM (Fig. 3a), whereas Arg79 and Asn81 coordinate interaction with the sidechain of Arg66 of L4. Consistently, the N81A mutation that abrogates MifM-stalling<sup>6,10</sup> would also abolish the potential hydrogen bond formed with Arg66 of L4. Two contacts are observed within the lower region of the tunnel, one between Phe85 of L22 and Ile74 of MifM (Fig. 3d) and another between

A1321 in H50 of the 23S rRNA in the vicinity of His68 of MifM (Fig. 3f). These contacts are likely to be important since insertions within L22 compromise MifM-dependent stalling, as do mutations of Trp73, Ile74 or His68 to alanine<sup>10</sup>. However, because the H68A only partially affected YidC2 induction using the *mifM-yidC2'-lacZ* reporter<sup>10</sup>, the latter contact with A1321 may instead reflect the importance of the neighbouring MifM residues Arg69 and Ile70 for stalling<sup>10</sup>. While a fully extended backbone conformation for MifM in the lower tunnel region does not bring these residues into direct proximity of A1321, because of the limited resolution for this region of the nascent chain we cannot rule out that the side chain of Arg69 can reach down to directly contact A1321. In agreement with the observed contacts between MifM and the ribosome, we could demonstrate using a *B. subtilis in vivo* GFP-MifM-LacZ (translational fusion) reporter (Fig. 3g) that both L22 and L4, but not L23 (the only other ribosomal protein that contributes to the exit tunnel), are important for MifM stalling (Fig. 3h–j): In a wildtype *B. subtilis* strain, translation of the GFP-MifM-LacZ reporter leads to stalling within the MifM sequence and therefore negligible  $\beta$ -galactosidase is produced (Fig. 3h,j), whereas *B. subtilis* strains bearing internal deletions within the tunnel lumen regions of L22 and L4 lead to an increase in  $\beta$ -galactosidase activity, consistent with a reduction in MifM stalling. In contrast,  $\beta$ -galactosidase activities remained low when deletions were made in L23 (Fig. 3j), consistent with the absence of a direct interaction between MifM and L23 in the MifM-SRC.

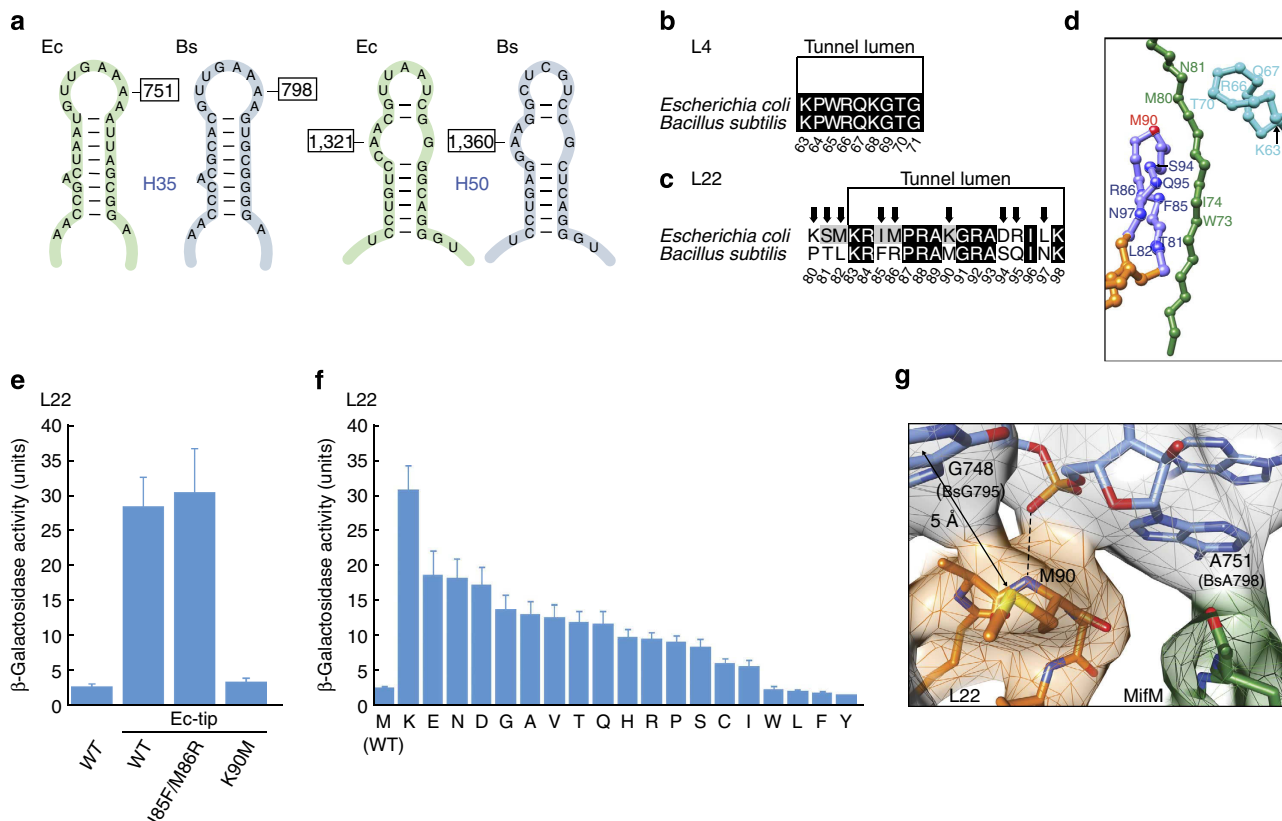
**Species specificity of MifM stalling.** To understand the species specificity of MifM stalling, we analysed the sequence and structure conservation of the ribosomal tunnel between *B. subtilis* and *E. coli*, revealing that the rRNA components are highly conserved, including nucleotides A751 and A1321 (Fig. 4a) and, similarly, the luminal region of L4 is invariant between *B. subtilis* and *E. coli* (Fig. 4b). In contrast, the  $\beta$ -hairpin of L22 exhibits a



**Figure 3 | Interactions of the MifM nascent chain with components of the ribosomal tunnel. (a–f)** Contacts between MifM nascent chain (green) and the ribosome in the (a–c) upper and (d–f) lower region of the ribosomal tunnel. Electron density (mesh) is coloured for MifM (green), L4 (cyan), L22 (orange) and rRNA (grey). In (d–f) the cryo-EM map was filtered to 4 Å resolution and the nascent chain modelled as a backbone trace. **(g)** Schematic for the GFP–MifM–LacZ reporter used to monitor translational arrest, where stalling prevents β-galactosidase production in *B. subtilis* *in vivo*. **(h–j)** β-galactosidase activity from the GFP–MifM–LacZ reporter (as in **g**) expressed in *B. subtilis* strains bearing wildtype (WT) or truncated versions of ribosomal proteins **(h)** L4, **(i)** L22 or **(j)** L23. Error bars indicate the s.d. of three independent biological replicates.

number of amino-acid differences between *B. subtilis* and *E. coli* (Fig. 4b–d), suggesting that the specificity of MifM stalling may be conferred by L22. To test this possibility, we utilized the GFP–MifM–LacZ reporter assay and demonstrated that replacing the β-hairpin of *B. subtilis* L22 with corresponding sequence from *E. coli* L22 led to an increase in β-galactosidase levels (Fig. 4e), consistent with a reduction in stalling at the MifM-derived sequence. Strikingly, reverting a single amino acid (K90M) in the β-hairpin of L22 from the *E. coli* sequence (K90) back to *B. subtilis* (M90) was sufficient to restore the low levels of β-galactosidase activity indicative of efficient stalling at MifM, whereas other reversions, such as the IM to FR double mutation

at positions 85–86 of L22 retained the elevated β-galactosidase levels (Fig. 4e). These findings suggest that the nature of the amino acid at position 90 of L22 can markedly influence the efficiency and specificity of stalling. Therefore, we introduced all 19 amino-acid substitutions at position 90 of *B. subtilis* L22 and monitored the β-galactosidase activity using the GFP–MifM–LacZ reporter (Fig. 4f). A wide range of β-galactosidase activities was observed, with K90, as present in *E. coli* L22, producing the highest levels, whereas hydrophobic or aromatic amino acids, such as F, Y, L and I, in position 90 exhibited β-galactosidase activities similar to the wildtype L22 with M90 (Fig. 4f). While γ-proteobacteria, such as *E. coli* usually have K90 and most



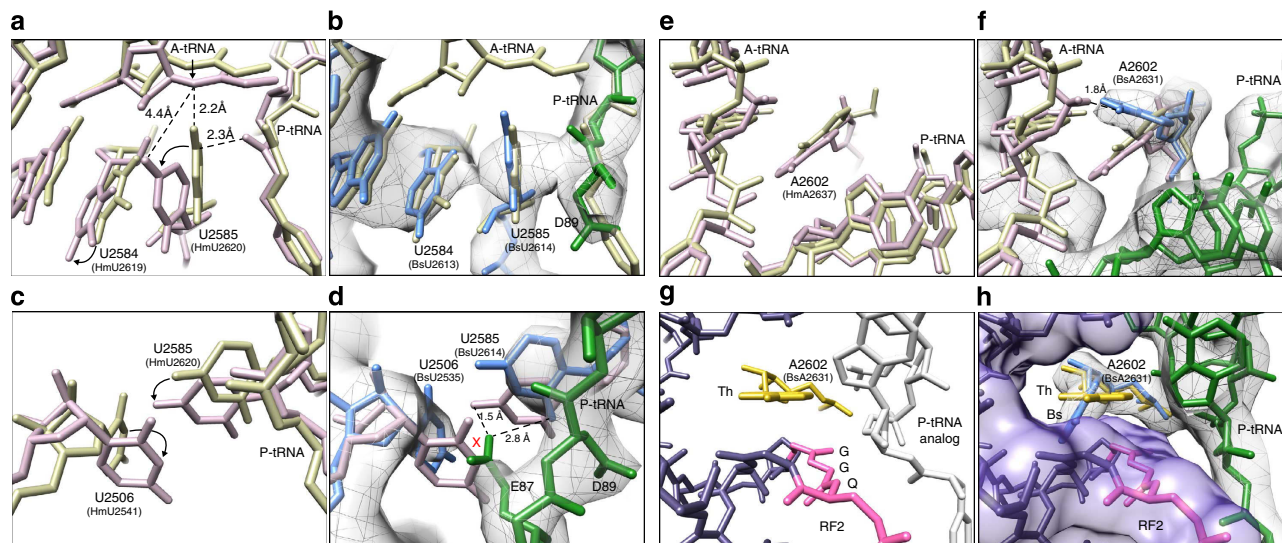
**Figure 4 | Residue M90 in L22 contributes to the species specificity of MifM stalling.** (a–c) Conservation between *B. subtilis* and *E. coli* of (a) 23S rRNA nucleotides A751 (A798 in *B. subtilis*) and A1321 (A1360 in *B. subtilis*) within helices H35 and H50, respectively, as well as the tunnel lumen region of ribosomal proteins (b) L4 and (c) L22. In (b,c) similar and identical residues are shaded grey and black, respectively. (d) Overview of relative positions of MifM to tunnel lumen residues of L4 and L22. (e,f)  $\beta$ -galactosidase activity from the GFP-MifM-LacZ reporter (as in Fig. 3g) expressed in *B. subtilis* strains bearing wildtype (WT) L22 compared with (e) Ec-tip mutants where *B. subtilis* residues 80–98 are substituted with the equivalent *E. coli* residues (see d) and then additionally reverted to *B. subtilis* residues by substitutions I85F/M86R or K90M, or (f) L22 mutants where all possible amino-acid substitutions at position M90 of *B. subtilis* L22 were generated. Error bars indicate the s.d. of three independent biological replicates. (g) Interaction between the side chain of M90 of L22 with the base of G748 and potential hydrogen bonding (dashed line) between the backbone of M90 and the phosphate-oxygen of A751.

*Bacillus* species have M90, other bacterial lineages generally contain positively charged amino acids such as R, Q, K or H at the equivalent position (Supplementary Fig. 10), and are therefore unlikely to promote efficient MifM stalling.

M90 of L22 does not directly contact MifM but rather forms a Met–Sulfur aromatic interaction with G748 of the 23S rRNA (Fig. 4g), suggesting that the influence of this residue on MifM stalling may be indirect. One possibility is that M90 influences the interaction of MifM with nucleotide A751, either indirectly through G748, or more directly since the backbone nitrogen of M90 is within hydrogen bonding distance of the phosphate-oxygen of A751 (Fig. 4g). In fact, the entire tip of the  $\beta$ -hairpin of L22 forms an intricate network of interactions with the 750-loop of 23S rRNA helix H35 (Supplementary Fig. 11), suggesting that M90 mutations which alter the interaction with G748 as well as the conformation of the L22 tip could lead to rearrangements in the A750-loop. While the A751 region is structurally similar between *B. subtilis* and *E. coli*, the resolution does not allow us to exclude subtle changes or shifts in the positions of the nucleotides that could influence interaction with the MifM nascent chain. In addition, M90 mutations could alter the conformation of the  $\beta$ -hairpin of L22 and thereby influence the interaction between MifM and F85 of L22 (Fig. 3f). Support for the interplay between the  $\beta$ -hairpin of L22 and MifM residues located below the constriction comes from rescue experiments: Random

mutagenesis of the  $\beta$ -hairpin (positions 80–89 and 91–98) of *B. subtilis* L22-M90K was performed to search for second site mutations that restore MifM stalling and induce expression of a YidC2'–LacZ fusion, which led to the identification of mutations at position G91 of L22 that restored induction by up to  $\sim 70\%$  of the wildtype level, as well as lesser effects at positions R92, A93 and F85 (Supplementary Fig. 12). Second site mutations that could restore MifM stalling and induce expression of a YidC2'–LacZ fusion in the presence of *B. subtilis* L22-M90K were also identified within the MifM nascent chain. In this latter case, random mutagenesis was performed at MifM positions 72–81, leading to the identification of the T72R mutation that restored induction by up to  $\sim 60\%$  of the wildtype level, with lesser effects also being observed at positions R75, K76 and F78 (Supplementary Fig. 12).

**MifM inactivation of the PTC.** To understand how the interaction of the MifM nascent chain with components of exit tunnel prevents stable binding of the A-tRNA and therefore leads to inhibition of peptide bond formation, we compared the conformation of nucleotides at the PTC in different states of peptide bond formation<sup>24–27</sup> with the PTC of the MifM-SRC (Fig. 5a–c). U2584 and U2585 shift by 2–3 Å on A-tRNA accommodation, that is, when moving from the unaccommodated (uninduced) to the accommodated (induced) state<sup>24–26</sup> (Fig. 5a). In the MifM-



**Figure 5 | MifM stabilizes the uninduced state of the PTC to inhibit A-tRNA accommodation.** (a) A-tRNA accommodation leads to conformational changes in U2584 and U2585 shifting the PTC from an uninduced (yellow) to an induced state (salmon)<sup>24–26</sup>. (b) The PTC of the MifM-SRC (grey density with rRNA (blue) and P-tRNA (green)) resembles that of the uninduced state (yellow). (c) Rotation of U2506 is required to accommodate the shift in U2585 that occurs on A-tRNA accommodation<sup>24–26</sup>. (d) E87 of MifM nascent chain (green) occupies the position of U2506. (e) A-tRNA accommodation leads to a slight shift in the position of A2602. (f) The position of A2602 (blue) in the MifM-SRC encroaches on the A-tRNA binding site and is distinct from the position of A2602 in the uninduced (yellow) or induced states (pink)<sup>24–26</sup>. (g) A distinct conformation of A2602 (yellow) is required to position the GGQ motif of RF2 at the PTC to catalyse peptidyl-tRNA hydrolysis<sup>29–32</sup>. (h) The position of A2602 (blue) in the MifM-SRC is incompatible with the canonical binding position of RF2 at the PTC<sup>29–32</sup>.

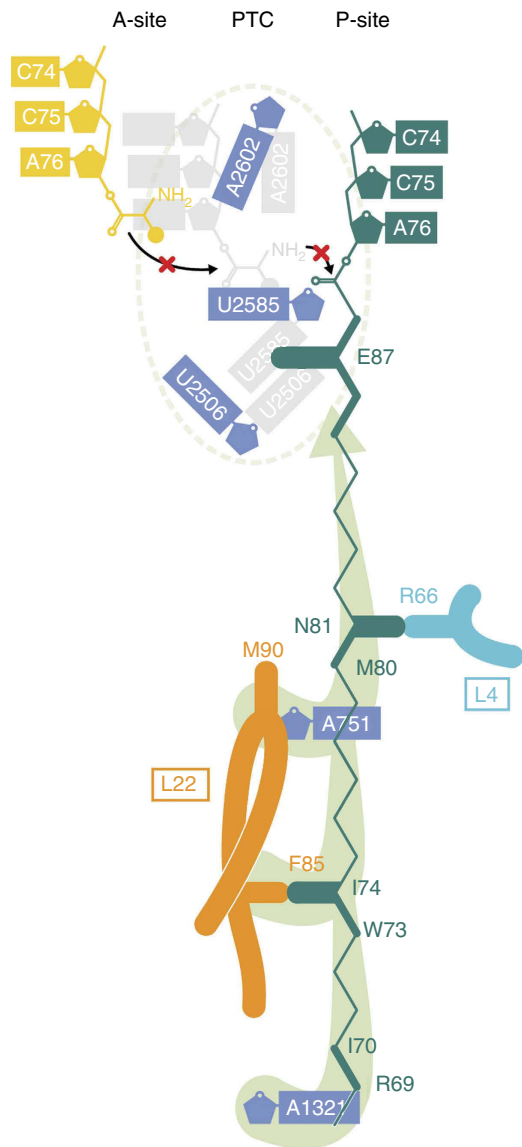
SRC, U2584 and U2585 resemble the uninduced state, consistent with the absence of A-tRNA (Fig. 5b). The shift in U2585 that occurs during A-tRNA accommodation also requires a corresponding rotation of U2506 (Fig. 5c). In the MifM-SRC, the side chain of E87 overlaps in position with the induced conformation of U2506 (Fig. 5d), suggesting that the MifM nascent chain prevents A-tRNA accommodation by sterically preventing PTC nucleotides U2506, and thereby also U2585, from adopting their induced conformational states. E87 comprises part of the DEED sequence that is critical for MifM stalling<sup>2</sup>, however, single mutation of any of the amino acids within the DEED motif, including E87A, does not markedly reduce stalling<sup>6</sup>. To completely abolish stalling, multiple D/E to A mutations within this motif are required<sup>6</sup>, indicating some functional redundancy. One could speculate that within the context of the E87A mutation the role of E87 in blocking the U2506 shift is assumed by a neighbouring D/E amino acids, however, without determining structures of the individual mutant MifM-RNCs it would be difficult to address this question.

In addition, during normal translation A2602 undergoes a slight shift on A-tRNA accommodation (Fig. 5e), whereas in the MifM-SRC, A2602 appears to adopt a defined position (Fig. 5f), which would restrict the entry of the incoming aminoacyl-tRNA to the A-site of the PTC. The conformation of the PTC of the MifM-SRC also provides an explanation as to why the MifM-stalled ribosomes are resistant to the action of the RF1 and RF2 when the A90 position is a stop codon<sup>6,10</sup>: Peptidyl-tRNA hydrolysis by RF1 and RF2 requires the accurate placement of the GGQ motif in domain 3 at the PTC of the ribosome, which induces specific conformations of PTC nucleotides, such as A2602<sup>28–32</sup> (Fig. 5g). This conformational change would be incompatible with the uninduced state of the PTC observed in the MifM-SRC (Fig. 5h). Indeed, the conformation of A2602 in the MifM-SRC is similar to that observed in the cryo-EM structure of the TnaC-SRC<sup>33</sup>, where translational arrest occurs because RF2-dependent peptidyl-tRNA hydrolysis is inhibited<sup>7</sup>.

Although our structure does not provide direct insight into the mechanism of PTC inhibition when the ribosome stalls at subsequent downstream residues, we note that the SD-like helix formed between the MifM mRNA and the 5' end of the 16S rRNA (Fig. 2) is likely to contribute to multisite stalling, as SD-like sequences within the open reading frames of mRNAs have been reported to induce translational pausing in bacteria<sup>34</sup>. However, the introduction of frameshifts within the MifM leader peptide abolished translational stalling at all sites<sup>10</sup>, indicating that the SD-like helix alone is insufficient to induce pausing independently of the MifM nascent chain-induced stall, possibly because the helix cannot form in the absence of stalling or that SD-dependent pausing is actually negligible<sup>35</sup>.

## Discussion

Collectively, our biochemical and structural findings lead us to propose a model for the mechanism and specificity of MifM stalling (Fig. 6): Residue M90 of L22 contributes to the species specificity of stalling, either through structural constraint given to A751 or residue F85 of L22. In the absence of any obvious relay of conformational changes to the PTC through the rRNA, we propose that these interactions, together with contacts between MifM and A1321 in 23S rRNA helix H50 and R66 of L4, promote a defined conformation of the MifM nascent chain such that residue E87 of MifM interacts with U2506 and U2585, thereby stabilizing the uninduced state of the PTC and preventing accommodation of the incoming A-tRNA (Fig. 6). In addition, we envisage that the pulling force exerted on the MifM nascent chain by interaction of the MifM TM with the YidC1/2 translocator during membrane insertion would release the MifM-dependent stalling by disrupting interactions between MifM and the ribosomal tunnel as well as displacing the E87 side chain of MifM and thus allowing U2506 and U2585 to adopt the induced state of the PTC that is necessary for aminoacyl-tRNA accommodation at the A-site. The mechanism of ligand-



**Figure 6 | Molecular basis for the specificity and mechanism of MifM-dependent translation arrest.** Model illustrating how interactions of MifM within the ribosomal tunnel stabilize a specific conformation of E87 that prevents conversion of the uninduced state to the induced state and thereby blocks aminoacyl-tRNA binding at the A-site of the PTC.

independent translation arrest by MifM as described here is structurally distinct from previously characterized peptide-dependent stalling systems, such as Erm<sup>36,37</sup>, TnaC<sup>33,38</sup> and SecM<sup>18</sup>, illustrating the plasticity of the ribosome to employ diverse mechanisms to sense the nascent chain in the tunnel and evoke silencing of the PTC.

## Methods

**Strain and plasmid construction.** *B. subtilis* strains and plasmids used in this study are listed in Supplementary Tables 1 and 2, respectively. *B. subtilis* strains were derivatives of PY79 (wildtype)<sup>39</sup> and constructed by transformation with either plasmid or purified *B. subtilis* chromosomal DNA. Plasmids were constructed by standard cloning methods, fusion PCR and site-directed mutagenesis<sup>40</sup> using primers and template plasmids listed in Supplementary Tables 2 and 3. Constructions of plasmids pCH735, pCH835, pCH913, pCH1142 and pEB71 were described previously (references shown in Supplementary Table 2). Plasmid pCH1517 was constructed by cloning an SphI-BglII-digested PCR fragment amplified from pCH913 using primers SP26 and SP27 into the SphI-BamHI site of pCH735. Plasmid pCH1570 was constructed by cloning a

BamHI-SphI fragment of pCH1567 into pCH1142. Plasmid pCH1567 was constructed by cloning a XbaI-SphI-digested PCR fragment amplified from PY79 chromosomal DNA using primers SP66 and SP67 into pCH1557. Plasmid pCH1557 was constructed as follows. A PCR fragment amplified from PY79 chromosomal DNA using primers SP68 and SP69 and another PCR fragment amplified from pEB71 using primers SP70 and SP71 were fused by the following PCR using primers SP68 and SP72, digested with both BamHI and XbaI and then cloned into pUC118. Plasmid pCH1587 was constructed as follows. A PCR fragment amplified from PY79 chromosomal DNA using primers SP28 and SP29 and another PCR fragment amplified from pEB71 using primers SP30 and SP31 were fused by the following PCR using primers SP28 and SP31, digested with both SacI and SphI and then cloned into pCH1142. Plasmid pCH1901 was constructed as follows. A PCR fragment amplified from pCH1587 using primers SP34 and SP35 and another PCR fragment amplified from pCH1587 using primers SP36 and SP37 were fused by the following PCR using primers SP34 and SP37, digested with both XhoI and SphI and then cloned into pCH1590. Other plasmids were constructed by site-directed mutagenesis using primers and templates shown in Supplementary Tables 2 and 3.

**Bacillus subtilis S12 translation extract.** The *B. subtilis* S12 translation extract was prepared following a procedure described for *E. coli* S12 translation extract<sup>41,42</sup>, with some modifications. Briefly, cells (*B. subtilis* strain 168) were grown to OD<sub>600</sub> 4.5 in an 'INFORCE HT minifors' bench-top fermenter in 2 × YPTG medium (16 g l<sup>-1</sup> peptone, 10 g l<sup>-1</sup> yeast extract, 5 g l<sup>-1</sup> NaCl, 22 mM NaH<sub>2</sub>PO<sub>4</sub>, 40 mM Na<sub>2</sub>HPO<sub>4</sub>, 19.8 g l<sup>-1</sup> glucose (sterile filtered)), with extra glucose feeding (feed: 10) at 37 °C while maintaining pH 7.0 and oxygen level (60%). Cells were collected at 5,000 × g at 4 °C for 15 min and subsequently washed 3 × in cold Buffer A (10 mM Tris-acetate (pH 8.2), 14 mM magnesium acetate, 60 mM potassium glutamate, 1 mM dithiothreitol and 6 mM 2-mercaptoethanol). Cells were snap-frozen in liquid nitrogen and stored at -80 °C. About 15 g of cells were thawed on ice, resuspended in 10 ml of cold buffer B (buffer A without 2-mercaptoethanol) and broken open in an 'microfluidics model 1101 lab homogenizer', 3 × at 15,000 psi. The lysate was cleared subsequently at 12,000 × g and 4 °C for 10 min and incubated for 30 min at 37 °C in a water bath. The cell extract was aliquoted, snap frozen and stored at -80 °C.

**PCR and *in vitro* transcription.** PCR reaction (Phusion Master Mix (NEB), 98 °C 5 min (98 °C 5 s; 52 °C 5 s; 72 °C 20 s) 28 × 72 °C 10 s, 12 °C 5 min) was prepared with MifMfor and MifMrev (Supplementary Table 3) and template DNA containing the sequence: (5'-TAATACGACTCACTATAGGGCGAATTGGCGGAAGGCCGTCAAGGCCACGTTGCTTGTCCATTAATTAACGTTTAACTTTAAGAAGGAGATA TACCAATGGGTTCATCACCATCACCATCACCATCAGATTACGATATTCC AACGACCCCTGGAAGTTCTGTTCCAGGGACCCGGTACCATGTTTGTGGAA TCGATAAATGACGTTTTATTCTTAGTCGATTTTTTACAAATATTCTTCTCT GCTCTAACGGCAATCGGGATTGCATTCCTTACGGGAGTGCCGTGCGG GCGAGCAATGGAAATCAAAAACGAAACAGATGGGCCCTACCCATACGATG TTCCAGATTACGCTGACTTCTTATTATTATATATCATCGATTACAAC TGGATACGTAAGGCTTCCGCATGAATTCGCCTGTGAACGATGAGGAAG ACGCGGTTCTCTTCTTTATAA-3'; underlined are the T7 promoter region, ribosomal binding site, start codon, 8 × His-tag and stop codon, respectively). PCR product was purified via spin columns and *in vitro* transcription reaction was set up using 1 μg PCR product per 50 μl reaction volume and T7 RNA polymerase. RNA was purified by LiCl/ethanol precipitation.

**Preparation of the MifM-SRC.** A volume of 1,360 μl of reaction mix containing 240 mM HEPES-KOH (pH 8.2), 60 mM glucose, 1.2 mM ATP, 0.85 mM CTP, GTP, and UTP, 2 mM DTT, 0.17 mg ml<sup>-1</sup> *E. coli* total tRNA mixture (from strain MRE600), 90 mM potassium glutamate, 80 mM of ammonium acetate, 8 mM magnesium acetate, 20 mM potassium phosphate dibasic (pH 7.2), 34 μg ml<sup>-1</sup> L-5-formyl-5, 6, 7, 8-tetrahydrofolic acid (folinic acid), 2.1 mM (each) amino-acid mix, 2 mM cysteine, 2% (w/v) PEG 8000 was preheated to 30 °C before the addition of 540 μl (27% (v/v)) cell extract (S12 extract). The whole reaction was incubated at 30 °C for 2 min before the addition of 100 μl *mifM* mRNA (Supplementary Fig. 1) and further incubated at 30 °C for 33 min, shaking at 1,000 r.p.m.

**Purification of the MifM-SRC.** *In vitro* translation reactions (4 × 500 μl) were loaded onto 500 μl sucrose cushion (750 mM sucrose) in Buffer C (50 mM HEPES, 250 mM KOAc, 10 mM MgOAc, 0.1% DDM, 1/1,000 complete protease inhibitor (Roche), 0.2 U ml<sup>-1</sup> RNasin, pH 7.2 at 4 °C) and centrifuged for 150 min (45,000 r.p.m., 4 °C) in a Beckman Coulter TLA 120.2 fixed-angle rotor. Pellets were resuspended in 4 × 300 μl 250 mM sucrose in ice cold buffer C and loaded onto a Talon metal affinity chromatography column (750 μl resin) pre-equilibrated in 10 ml buffer C containing 10 μg μl<sup>-1</sup> bulk tRNA. The column was washed with 25 ml buffer C until no significant absorption (A<sub>260</sub>) could be detected in the wash fractions. The MifM-SRC, bound to the Talon matrix by MifM's N-terminal 8 × His-tag, was eluted in 4 × 500 μl buffer C containing 150 mM Imidazole. The eluate was loaded onto 10–40% sucrose gradients (prepared with buffer C) and centrifuged for 13 h in a Beckman coulter SW40 Ti swinging bucket rotor (20,000 r.p.m., 4 °C). Gradients were separated on a Biocomp Gradient Station *ip*

and fractions containing 70S ribosomal particles were collected and pelleted for 3 h in a Beckman Coulter TLA 120.2 fixed-angle rotor (45,000 r.p.m., 4 °C). The MifM-SRC pellet was resuspended in 28  $\mu$ l 30 mM sucrose in buffer C, on ice (69 OD ml<sup>-1</sup>), aliquoted and snap-frozen. Samples were further analysed by SDS-PAGE and western blotting.

**MifM-SRC/SRP sample and cryogrid preparation.** For grid preparation 2.5  $\mu$ l *in vitro* reconstituted *B. subtilis* signal recognition particle (20 pmol  $\mu$ l<sup>-1</sup> in SRP buffer: 25 mM HEPES, 100 mM KOAc, 10 mM MgOAc, 1 mM DTT, 10% Glycerol, pH 7.2 at 4 °C) were diluted to a final volume of 44.2  $\mu$ l with buffer E (50 mM HEPES, 250 mM KOAc, 2 mM 2-mercaptoethanol, 0.06% Nikkol, pH 7.2 at 4 °C) and activated by incubation at 30 °C for 10 min. Subsequently, 5.8  $\mu$ l (9.6 pmol) of MifM-SRC was added and incubated for 10 min at 30 °C. Finally the prepared mix was diluted with 22.5  $\mu$ l buffer F (50 mM HEPES, 250 mM KOAc, 2 mM 2-mercaptoethanol, pH 7.2 at 4 °C) for optimal grid coverage.

**Cryoelectron microscopy and single particle reconstruction.** The MifM-SRC/SRP sample was applied to 2 nm precoated Quantifoil R3/3 holey carbon supported grids and vitrified using a Vitrobot Mark IV (FEI Company). Data collection was performed using the EPU software (FEI) at NeCEN (Leiden, Netherlands) on a Titan Krios transmission electron microscope (FEI, Eindhoven, Netherlands) equipped with a Falcon II direct electron detector. Data were collected with the microscope set to 300 kV, a magnification of 125,085x (pixel size: 1.108 Å) in a defocus range of 1.0–2.5  $\mu$ m (Table 1). The data were provided as a series of seven frames (grid was pre-exposed for 55 ms, 4 e<sup>-</sup> Å<sup>-2</sup> before recording; dose per recorded frame: 8 e<sup>-</sup> Å<sup>-2</sup>, exposure-time/ recorded frame: 110 ms). Frames 1–3 (accumulated dose of 28 e<sup>-</sup> Å<sup>-2</sup>) were summed after alignment using Motion Correction software<sup>43</sup>. Images were processed using a frequency-limited refinement protocol that helps prevent over-fitting<sup>44</sup>, specifically by truncation of high frequencies (in this case at 8 Å). As reported and expected<sup>44</sup>, we found that using this processing regime the 0.143 Fourier Shell Correlation (FSC) value provides a good indicator for the true average resolution of the map (Supplementary Fig. 2). In addition, the local resolution of the final map was calculated using ResMap<sup>19</sup>. Power spectra and defocus values were determined using the SPIDER TF ED command and recorded images were manually inspected for good areas and power-spectra quality. Data were processed further using the SPIDER software package<sup>45</sup>, in combination with an automated workflow as described previously<sup>46</sup>. After initial, automated particle selection based on the program SIGNATURE<sup>47</sup>, initial alignment was performed with 529,488 particles, using an *E. coli* 70S ribosome as reference structure<sup>36</sup>. After removal of noisy particles (101,109 particles; 19%), the data set was further sorted using an incremental K-means-like method of unsupervised 3D sorting<sup>48</sup>. Particles lacking density for P-tRNA (75,984 particles; 14%) and those that did not improve in resolution during refinement were omitted from the data set. The major subpopulation (305,045 particles; 58%) showed the presence of stoichiometric densities for P-tRNA and could be refined to an average resolution of 3.9 Å (0.143 FSC) and a local resolution extending to 3.5 Å for the core of the ribosome as computed using ResMap<sup>19</sup> (Supplementary Fig. 2) and was used for modelling. Subsequently, this main volume could be sorted further into a subvolume containing the signal recognition particle bound to the ribosome (158,726 particles; 30%), which will be described elsewhere (Beckert *et al.*, unpublished). The final volumes were subjected to the program EM-BFACTOR<sup>49</sup> in order to apply an automatically determined negative B-factor for sharpening of the map.

**Molecular modelling and map-docking procedures.** The *B. subtilis* 5S, 16S and 23S rRNA sequences were taken from GeneBank; Gene ID 2914271, 936774 and 939981, respectively. Structure-based sequence alignments were generated using Sequence to Structure (S2S)<sup>50</sup> based on X-ray structure derived models of the small and large ribosomal subunit of *E. coli* (PDB ID 4KIX/Y)<sup>51</sup>. Molecular dynamics flexible fitting (MDFF)<sup>52</sup> in VMD<sup>53</sup> was used for initial fitting and refinement of the rRNA models into the electron density. The resulting models of the *B. subtilis* rRNAs were manually inspected and adjusted according to features of the electron-density using Coot<sup>54</sup>, followed by refinement using eraser<sup>55</sup> and the real-space refine tool in PHENIX<sup>56</sup> (Table 1). A total of 29 of the 32 large subunit proteins (L1, L7/L12 and L9 were excluded) from the *B. subtilis* 50S subunit were generated using the homology with the equivalent *E. coli* and *T. thermophilus* protein templates (PDB codes 4KIX and 3I8I, respectively). The models were initially fitted to the density using Chimera<sup>57</sup>, followed by real-space refinement using PHENIX<sup>56</sup> (Table 1). For the 19 proteins of the *B. subtilis* 30S subunit, homology models were generated based on protein templates from *T. thermophilus* (PDB code 3V2C) using HHPred<sup>58</sup> and fitted to the density using Chimera<sup>57</sup> and refined in Coot<sup>54</sup>. The resolution of the core of the large subunit (3.5 Å, Supplementary Fig. 2) enabled the majority of the sidechains for the large ribosomal proteins to be modelled (Supplementary Fig. 3), whereas the lower resolution of the core of the small subunit (4.0 Å, Supplementary Fig. 2) limited the molecular modelling to the backbone of the small subunit ribosomal proteins. The fit of the atomic models was validated as described<sup>59</sup> by calculating the FSC between a theoretical density derived from the atomic models for the *B. subtilis* 70S ribosome (generated using the CP FROM PDB SPIDER command) and the unsharpened cryo-EM map of

MifM-SRC (Supplementary Fig. 4). The volumes were multiplied (using the MU SPIDER command) with a 'soft' mask generated from the theoretical model density (using the TH M and FQ NP SPIDER commands) previous to calculating their Fourier shell correlation (using the RF 3 SPIDER command).

**Figure preparation.** Figures showing electron densities and atomic models were generated using UCSF Chimera<sup>57</sup>.

**$\beta$ -galactosidase assays.** *B. subtilis* cells were cultured at 37 °C in LB media for  $\beta$ -galactosidase activity assay<sup>60</sup>. Cells were collected from cultures of OD<sub>600</sub> = 0.5–1.0. To test  $\beta$ -galactosidase activities on agar plates, cells were grown on Difco sporulation medium (DSM) agar plates containing 60  $\mu$ g ml<sup>-1</sup> 5-bromo-4-chloro-3-indolyl- $\beta$ -D-galactoside (X-gal) at 37 °C<sup>6</sup>.

**Isolation of suppressor mutants.** Plasmid library for isolation of intragenic suppressor mutants of the *rplV(M90K)* mutant were prepared as follows. The second point mutation was introduced into either of the 80–89th or the 91st–98th codons of *rplV(M90K)* on the original plasmid pCH1897 by site-directed mutagenesis using random mixed mutagenic primers listed in Supplementary Table 3 (SP38–SP55). *B. subtilis* strain SCB3378 (*gfp-mifM*<sup>35–95</sup>-*yidC2'-lacZcat*, *rplV94*) was transformed with the mutagenized plasmid library and then blue colonies were isolated on DSM agar plates including 3  $\mu$ g ml<sup>-1</sup> kanamycin and 60  $\mu$ g ml<sup>-1</sup> X-gal at 37 °C. The original strain SCB3378 harbours the *rplV94* mutation, which has a seven codon insertion within the *rplV*, resulting in arrest-defective phenotype. Strain SCB3378 was made from SCB824 (*gfp-mifM*<sup>35–95</sup>-*yidC2'-lacZcat*, *rplV*<sup>+</sup>) by isolating a spontaneous erythromycin resistant mutant. Suppressor *mifM* mutants were isolated as follows: Plasmid library was prepared by site-directed mutagenesis using template plasmid pCH835 (*gfp-mifM*<sup>35–95</sup>-*yidC2'-lacZcat*) and random mixed mutagenic primers listed in Supplementary Table 3 (SP56–65) that are designed to introduce a point mutation into either one of the 72nd–81st codons of *mifM*. *B. subtilis* strain SCB3348 (*rplV(M90K)*) was transformed with the mutagenized plasmid library and then blue colonies were isolated on DSM-chloramphenicol-X-gal agar plates at 37 °C.

## References

- Voorhees, R. M. & Ramakrishnan, V. Structural basis of the translational elongation cycle. *Annu. Rev. Biochem.* **82**, 203–236 (2013).
- Wen, J. D. *et al.* Following translation by single ribosomes one codon at a time. *Nature* **452**, 598–603 (2008).
- Peil, L. *et al.* Distinct XPPX sequence motifs induce ribosome stalling, which is rescued by the translation elongation factor EF-P. *Proc. Natl Acad. Sci. USA* **110**, 15265–15270 (2013).
- Starosta, A. L. *et al.* Translational stalling at polyproline stretches is modulated by the sequence context upstream of the stall site. *Nucleic Acids Res.* **42**, 10711–10719 (2014).
- Wilson, D. N. & Beckmann, R. The ribosomal tunnel as a functional environment for nascent polypeptide folding and translational stalling. *Curr. Opin. Struct. Biol.* **21**, 1–10 (2011).
- Chiba, S. & Ito, K. Multisite ribosomal stalling: a unique mode of regulatory nascent chain action revealed for MifM. *Mol. Cell* **47**, 863–872 (2012).
- Gong, F. & Yanofsky, C. Instruction of translating ribosome by nascent peptide. *Science* **297**, 1864–1867 (2002).
- Vazquez-Laslop, N., Thum, C. & Mankin, A. S. Molecular mechanism of drug-dependent ribosome stalling. *Mol. Cell* **30**, 190–202 (2008).
- Nakatogawa, H. & Ito, K. The ribosomal exit tunnel functions as a discriminating gate. *Cell* **108**, 629–636 (2002).
- Chiba, S., Lamsa, A. & Pogliano, K. A ribosome-nascent chain sensor of membrane protein biogenesis in *Bacillus subtilis*. *EMBO J.* **28**, 3461–3475 (2009).
- Chiba, S. *et al.* Recruitment of a species-specific translational arrest module to monitor different cellular processes. *Proc. Natl Acad. Sci. USA* **108**, 6073–6078 (2011).
- Dalbey, R. E., Wang, P. & Kuhn, A. Assembly of bacterial inner membrane proteins. *Annu. Rev. Biochem.* **80**, 161–187 (2011).
- Butkus, M. E., Prundeanu, L. B. & Oliver, D. B. Translocon 'pulling' of nascent SecM controls the duration of its translational pause and secretion-responsive secA regulation. *J. Bacteriol.* **185**, 6719–6722 (2003).
- Gumbart, J., Schreiner, E., Wilson, D. N., Beckmann, R. & Schulten, K. Mechanisms of SecM-mediated stalling in the ribosome. *Biophys. J.* **103**, 331–341 (2012).
- Ismail, N., Hedman, R., Schiller, N. & von Heijne, G. A biphasic pulling force acts on transmembrane helices during translocon-mediated membrane integration. *Nat. Struct. Mol. Biol.* **19**, 1018–1022 (2012).
- Chiba, S. & Ito, K. MifM monitors total YidC activities of *Bacillus subtilis* including that of YidC2, the target of regulation. *J. Bacteriol.* **197**, 99–107 (2014).

17. Bhushan, S. *et al.* alpha-Helical nascent polypeptide chains visualized within distinct regions of the ribosomal exit tunnel. *Nat. Struct. Mol. Biol.* **17**, 313–317 (2010).
18. Bhushan, S. *et al.* SecM-stalled ribosomes adopt an altered geometry at the peptidyltransferase center. *PLoS Biol.* **19**, e1000581 (2011).
19. Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nat. Methods* **11**, 63–65 (2014).
20. Schuwirth, B. *et al.* Structures of the bacterial ribosome at 3.5 Å resolution. *Science* **310**, 827–834 (2005).
21. Selmer, M. *et al.* Structure of the 70S ribosome complexed with mRNA and tRNA. *Science* **313**, 1935–1942 (2006).
22. Yusupova, G., Jenner, L., Rees, B., Moras, D. & Yusupov, M. Structural basis for messenger RNA movement on the ribosome. *Nature* **444**, 391–394 (2006).
23. Kaminishi, T. *et al.* A snapshot of the 30S ribosomal subunit capturing mRNA via the Shine-Dalgarno interaction. *Structure* **15**, 289–297 (2007).
24. Schmeing, T. M. *et al.* A pre-translocational intermediate in protein synthesis observed in crystals of enzymatically active 50S subunits. *Nat. Struct. Mol. Biol.* **9**, 225–230 (2002).
25. Schmeing, T. M., Huang, K. S., Kitchen, D. E., Strobel, S. A. & Steitz, T. A. Structural insights into the roles of water and the 2' hydroxyl of the P site tRNA in the peptidyl transferase reaction. *Mol. Cell* **20**, 437–448 (2005).
26. Schmeing, T. M., Huang, K. S., Strobel, S. A. & Steitz, T. A. An induced-fit mechanism to promote peptide bond formation and exclude hydrolysis of peptidyl-tRNA. *Nature* **438**, 520–524 (2005).
27. Voorhees, R. M., Weixlbaumer, A., Loakes, D., Kelley, A. C. & Ramakrishnan, V. Insights into substrate stabilization from snapshots of the peptidyl transferase center of the intact 70S ribosome. *Nat. Struct. Mol. Biol.* **16**, 528–533 (2009).
28. Polacek, N. *et al.* The critical role of the universally conserved A2602 of 23S ribosomal RNA in the release of the nascent peptide during translation termination. *Mol. Cell* **11**, 103–112 (2003).
29. Weixlbaumer, A. *et al.* Insights into translational termination from the structure of RF2 bound to the ribosome. *Science* **322**, 953–956 (2008).
30. Laurberg, M. *et al.* Structural basis for translation termination on the 70S ribosome. *Nature* **454**, 852–857 (2008).
31. Korostelev, A. *et al.* Crystal structure of a translation termination complex formed with release factor RF2. *Proc. Natl Acad. Sci. USA* **105**, 19684–19689 (2008).
32. Korostelev, A. A. Structural aspects of translation termination on the ribosome. *RNA* **17**, 1409–1421 (2011).
33. Seidelt, B. *et al.* Structural insight into nascent polypeptide chain-mediated translational stalling. *Science* **326**, 1412–1415 (2009).
34. Li, G. W., Oh, E. & Weissman, J. S. The anti-Shine-Dalgarno sequence drives translational pausing and codon choice in bacteria. *Nature* **484**, 538–541 (2012).
35. Borg, A. & Ehrenberg, M. Determinants of the rate of mRNA translocation in bacterial protein synthesis. *J. Mol. Biol.* (2014).
36. Arenz, S. *et al.* Molecular basis for erythromycin-dependent ribosome stalling during translation of the ErmBL leader peptide. *Nat. Commun.* **5**, 3501 (2014).
37. Arenz, S. *et al.* Drug Sensing by the ribosome induces translational arrest via active site perturbation. *Mol. Cell* **56**, 446–452 (2014).
38. Bischoff, L., Berninghausen, O. & Beckmann, R. Molecular basis for the ribosome functioning as an L-tryptophan sensor. *Cell Rep.* **9**, 469–475 (2014).
39. Youngman, P., Perkins, J. B. & Losick, R. A novel method for the rapid cloning in *Escherichia coli* of *Bacillus subtilis* chromosomal DNA adjacent to Tn917 insertions. *Mol. Gen. Genet.* **195**, 424–433 (1984).
40. Sawano, A. & Miyawaki, A. Directed evolution of green fluorescent protein by a new versatile PCR strategy for site-directed and semi-random mutagenesis. *Nucleic Acids Res.* **28**, E78 (2000).
41. Kim, T. W. *et al.* Simple procedures for the construction of a robust and cost-effective cell-free protein synthesis system. *J. Biotechnol.* **126**, 554–561 (2006).
42. Kim, T.-W., Kim, H.-C., Oh, I.-S. & Kim, D.-M. A highly efficient and economical cell-free protein synthesis system using the S12 extract of *Escherichia coli*. *Biotechnol. Bioproc. E.* **13**, 464–469 (2009).
43. Li, X. *et al.* Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat. Methods* **10**, 584–590 (2013).
44. Scheres, S. H. & Chen, S. Prevention of overfitting in cryo-EM structure determination. *Nat. Methods* **9**, 853–854 (2012).
45. Frank, J. *et al.* SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J. Struct. Biol.* **116**, 190–199 (1996).
46. Becker, T. *et al.* Structural basis of highly conserved ribosome recycling in eukaryotes and archaea. *Nature* **482**, 501–506 (2012).
47. Chen, J. Z. & Grigorieff, N. SIGNATURE: a single-particle selection system for molecular electron microscopy. *J. Struct. Biol.* **157**, 168–173 (2007).
48. Loerke, J., Giesebrecht, J. & Spahn, C. M. Multiparticle cryo-EM of ribosomes. *Methods Enzymol.* **483**, 161–177 (2010).
49. Fernandez, J. J., Luque, D., Caston, J. R. & Carrascosa, J. L. Sharpening high resolution information in single particle electron cryomicroscopy. *J. Struct. Biol.* **164**, 170–175 (2008).
50. Jossinet, F. & Westhof, E. Sequence to Structure (S2S): display, manipulate and interconnect RNA data from sequence to structure. *Bioinformatics* **21**, 3320–3321 (2005).
51. Pulk, A. & Cate, J. H. Control of ribosomal subunit rotation by elongation factor G. *Science* **340**, 1235970 (2013).
52. Trabuco, L. G., Villa, E., Mitra, K., Frank, J. & Schulten, K. Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. *Structure* **16**, 673–683 (2008).
53. Humphrey, W., Dalke, A. & Schulten, K. VMD—visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).
54. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
55. Chou, F. C., Sripakdeevong, P., Dibrov, S. M., Hermann, T. & Das, R. Correcting pervasive errors in RNA crystallography through enumerative structure prediction. *Nat. Methods* **10**, 74–76 (2013).
56. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).
57. Pettersen, E. F. *et al.* UCSF chimera; a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
58. Hildebrand, A., Remmert, M., Biegert, A. & Soding, J. Fast and accurate automatic structure prediction with HHpred. *Proteins* **77**(Suppl 9): 128–132 (2009).
59. Chen, S. *et al.* High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy. *Ultramicroscopy* **135**, 24–35 (2013).
60. Rubio, A., Jiang, X. & Pogliano, K. Localization of translocation complex components in *Bacillus subtilis*: enrichment of the signal recognition particle receptor at early sporulation septa. *J. Bacteriol.* **187**, 5000–5002 (2005).

## Acknowledgements

This research was supported by grants from the Deutsche Forschungsgemeinschaft (WI3285/3-1 to D.W.; FOR1805 and GRK1721 to D.N.W. and R.B.), European Research Council (to R.B.), MEXT and JSPS Grant-in-Aid for Scientific Research (20523517, 25291006, 26116008 to S.C. and 20247020 to K.I.), the Private University Strategic Research Foundation Support Program from MEXT (S1101029 to K.I.), the Fondation pour la Recherche Médicale (AJE201133 to C.A.I.) and the Conseil Régional d'Aquitaine (2012 13 01 009 to C.A.I.) and the European Union (PCIG14-GA-2013-631479 to C.A.I.). We thank Rishi Matadeen and Sacha DeCarlo for EM data collection, Naoka Katagatsume and Shiori Yamamoto for preliminary data collection, Tomoe Takino and Chika Tsutsumi, for technical support and Andreas Anger for help with molecular modelling of the 30S subunit.

## Author contributions

D.S. prepared translation extract, MifM-SRC complex performed single-particle cryo-EM analysis and model building. S.C. and N.S.-C. performed *in vivo* *Bacillus* studies. D.S. and A.I. performed the model refinement and validation. O.B. helped with sample preparation. All authors interpreted the data and helped with the manuscript preparation. D.S. and D.N.W. wrote the manuscript.

## Additional information

**Accession codes.** The cryo-EM map and associated atomic coordinates have been deposited in the EMD and PDB with the accession numbers EMD-6306 and PDB ID 3J9W, respectively.

**Supplementary Information** accompanies this paper at <http://www.nature.com/naturecommunications>

**Competing financial interests:** The authors declare no competing financial interests.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

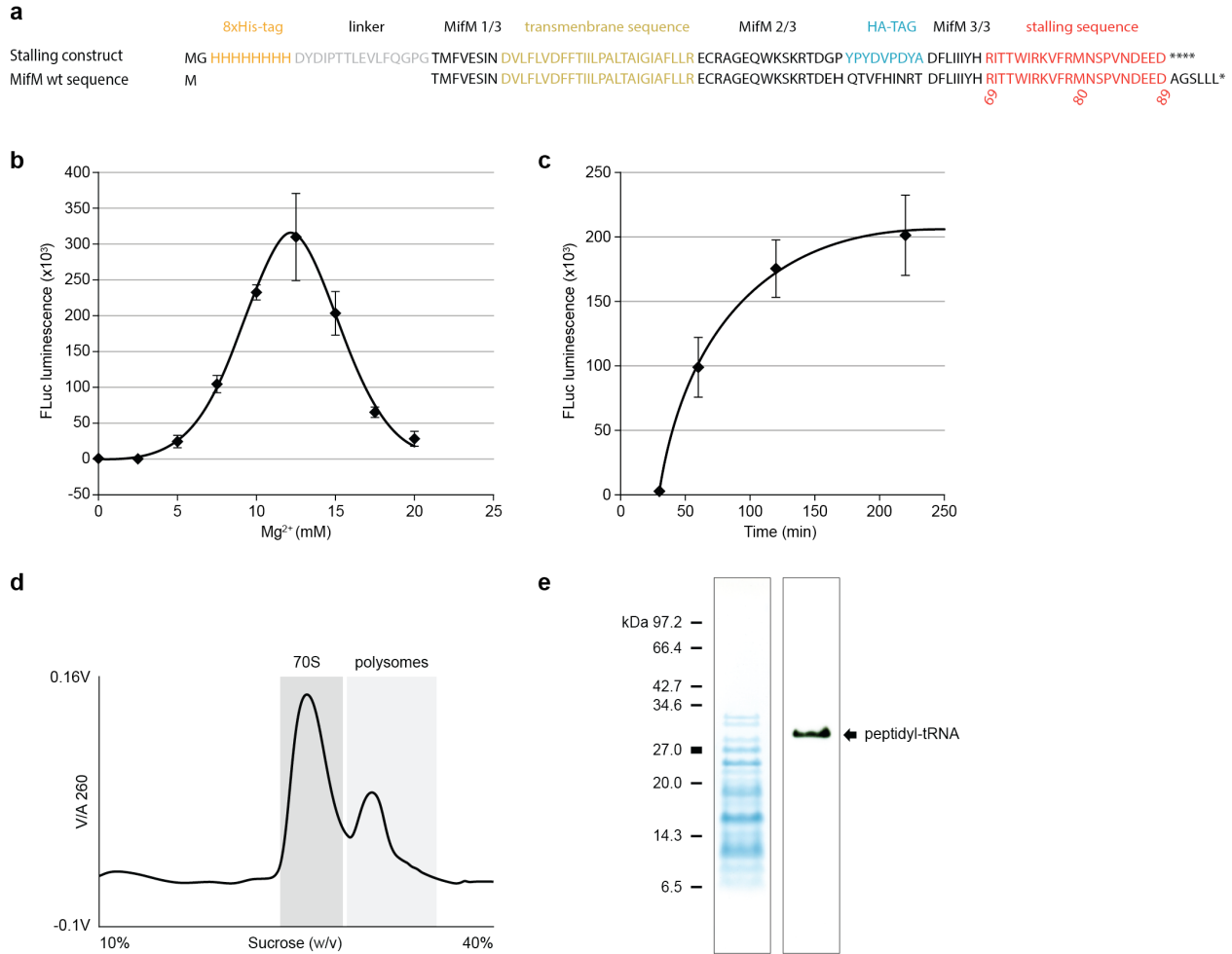
**How to cite this article:** Sohnen, D. *et al.* Structure of the *Bacillus subtilis* 70S ribosome reveals the basis for species-specific stalling. *Nat. Commun.* **6**:6941 doi: 10.1038/ncomms7941 (2015).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

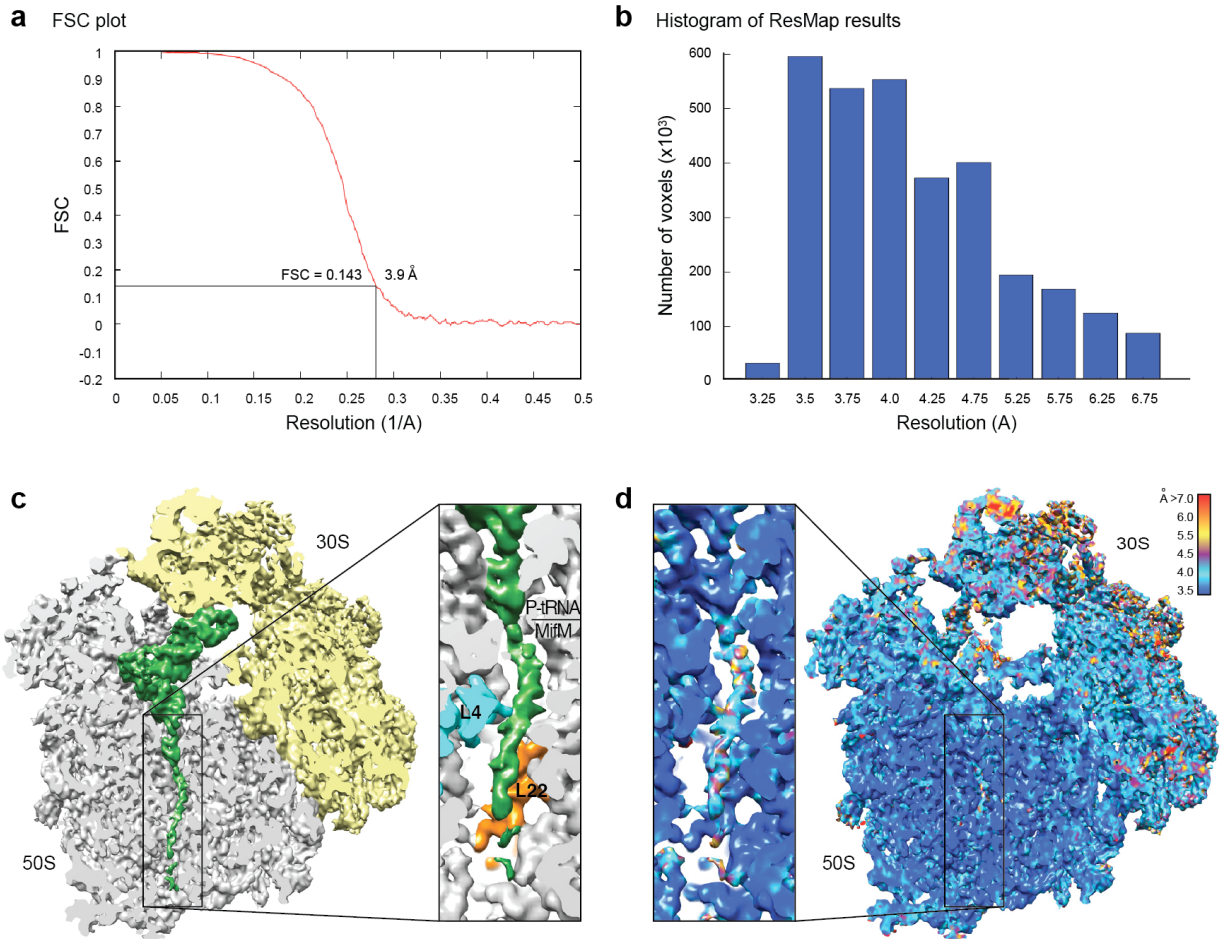


## Supplementary Figures



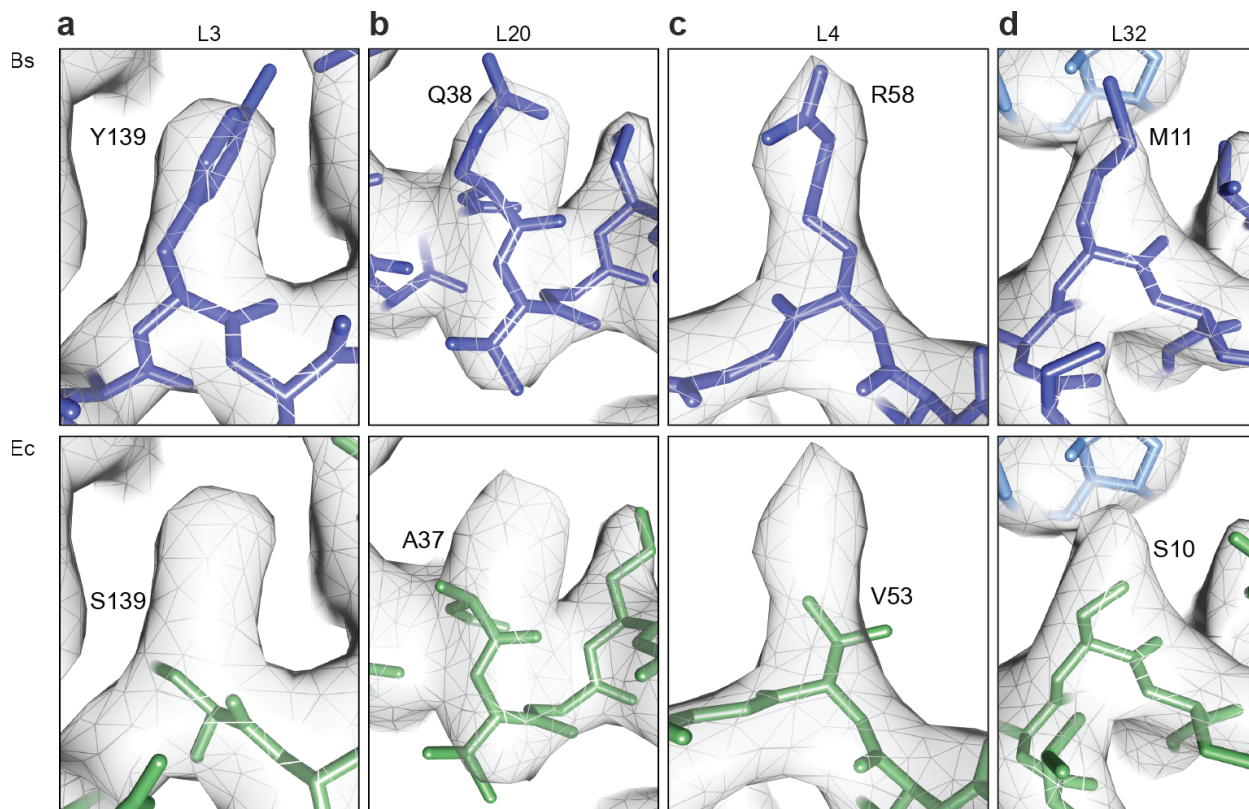
### Supplementary Figure 1: Preparation of the MifM-SRC.

**a**, Comparison of the amino acid sequence of wildtype MifM with the stalling construct used to generate the MifM-SRC. The MifM stalling window (red), transmembrane segment (gold), 8xHis-tag for purification (orange) and HA-tag (cyan) for detection by Western blotting are coloured. **b,c**, Optimization of **b**, magnesium ( $Mg^{2+}$ ) concentration and **c**, time (min) for translation of firefly luciferase (Fluc) in the *B. subtilis* S12 lysate-based *in vitro* translation system. Fluc activity was monitored using luminescence and the error bars represent the standard deviation from the mean for triplicate reactions. **d**, Sucrose gradient profile of translation reaction of MifM stalling construct (from a) used in the optimized *B. subtilis* S12 lysate *in vitro* translation system. Fractions from the 70S peak were collected to remove polysomes. **e**, SDS-PAGE and HA-tag Western blotting revealed a single band for the MifM peptidyl-tRNA and no free MifM peptide.



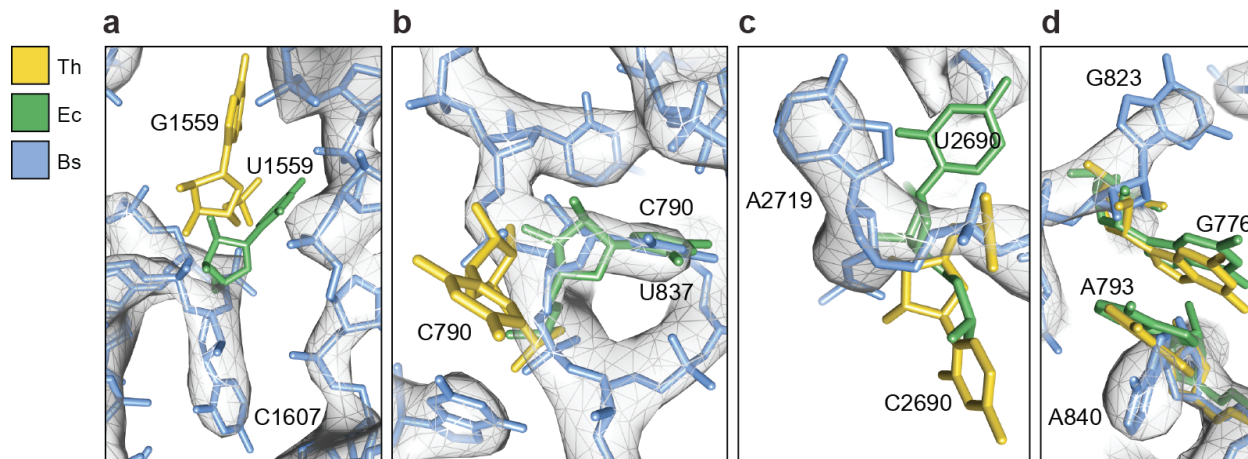
## Supplementary Figure 2: Average and local resolution of the cryo-EM map of the MifM-SRC

**a**, Average resolution of the MifM-SRC was 3.9 Å using the Fourier shell correlation (FSC) cut-off value of 0.143, which was employed since the microscopy images were processed in the absence of spatial frequencies higher than 8 Å<sup>-1</sup>. **b**, Histogram generated by ResMap<sup>2</sup> showing that the local resolution of the final MifM-SRC map reaches resolutions of 3.5 Å. **c**, Transverse section of the cryo-EM structure of the MifM-SRC (30S, yellow; 50S, grey) showing P-tRNA and MifM nascent chain (green) within the ribosomal tunnel and enlargement where ribosomal proteins L4 (cyan) and L22 (orange) are coloured. **d**, Same views as **c** but coloured according to local resolution as determined using ResMap<sup>2</sup>.



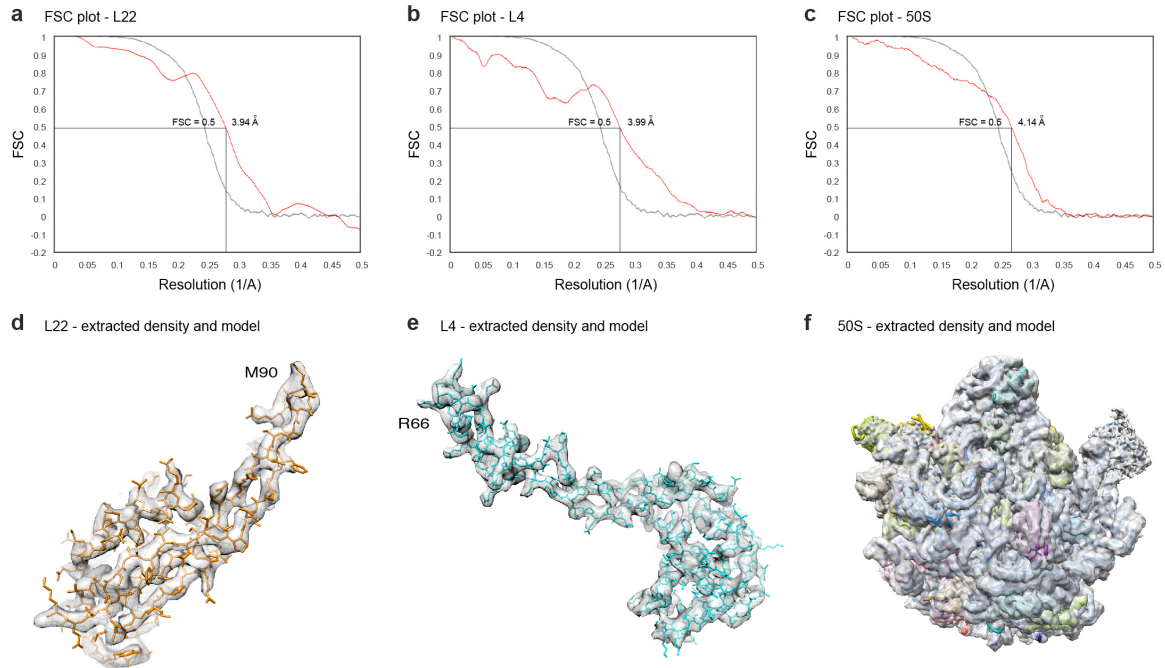
**Supplementary Figure 3: Differences in amino acid sequence for ribosomal proteins of the *B. subtilis*, *E. coli* and *T. thermophilus* 50S subunit**

**a-d**, Comparison of selected regions where single amino acid positions differ between *B. subtilis* (blue) and *E. coli* (green, PDB4KIX/Y)<sup>3,4</sup>. The electron density for the MifM-SRC is shown as a grey mesh.



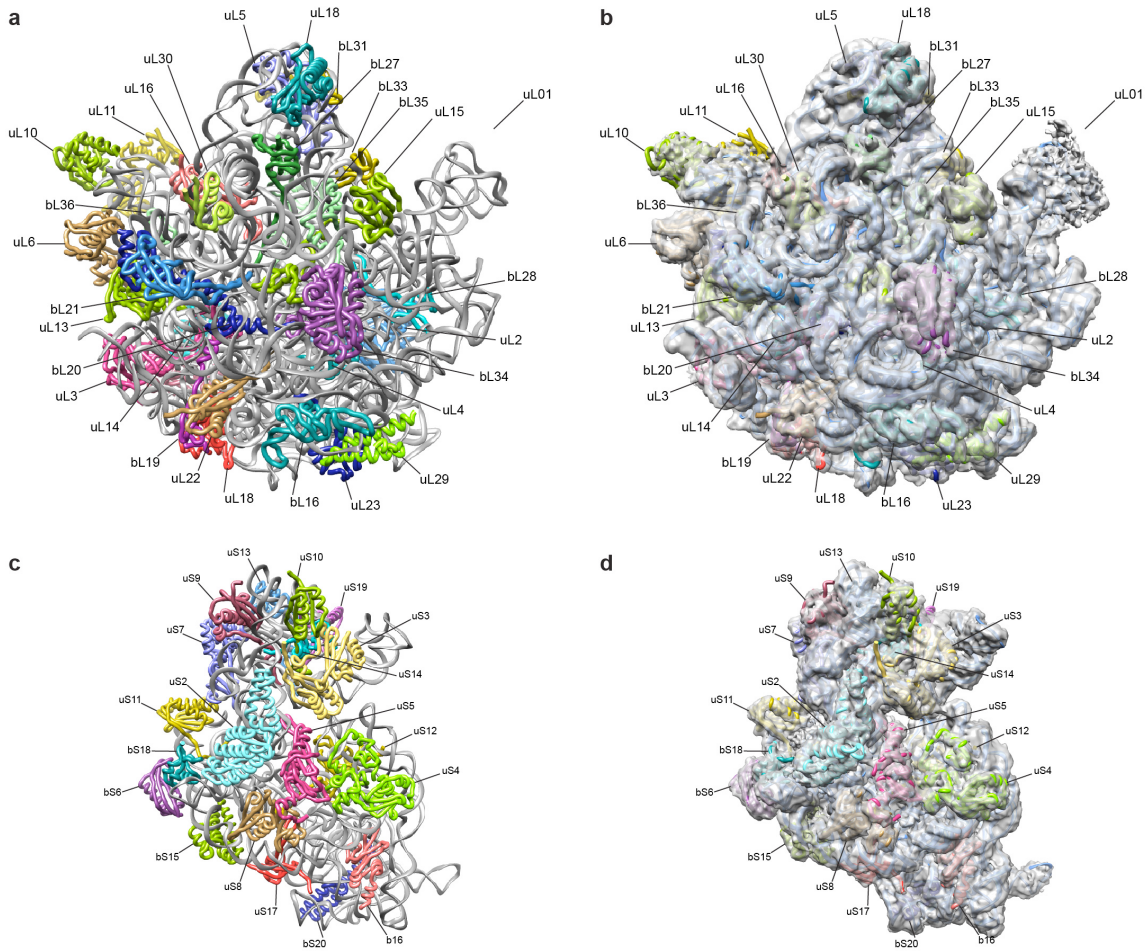
**Supplementary Figure 4: Differences in conformation of rRNA nucleotides between *B. subtilis*, *E. coli* and *T. thermophilus* ribosomes**

**a-d**, Comparison of selected regions where single rRNA nucleotide positions differ between *B. subtilis* (blue), *E. coli* (green, PDB4KIX/Y)<sup>3,4</sup> and *T. thermophilus* (yellow, PDB3I8H/I)<sup>5</sup> 70S ribosomes. The electron density for the MifM-SRC is shown as a grey mesh.



### Supplementary Figure 5: Validation of the molecular model of the *B. subtilis* 50S subunit

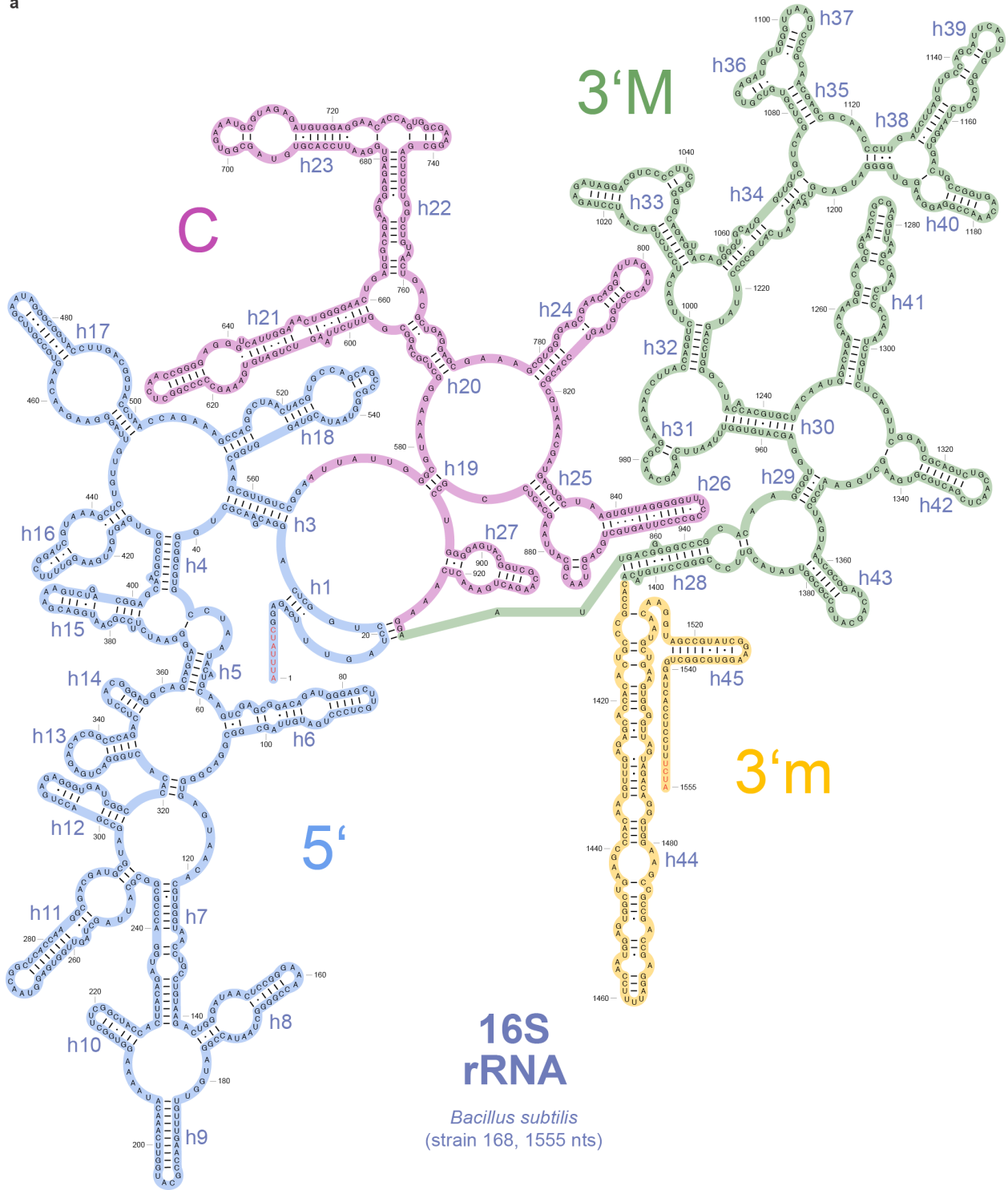
**a-c**, Overlay of the FSC of the cryo-EM map of the MifM-SRC (black line from fig. S2) compared to the FSC calculation between the molecular model and cryo-EM map (red line) for **a**, L22, **b**, L4, and **c**, the complete *B. subtilis* 50S subunit. **d-f**, Fit of the molecular models into the cryo-EM map (grey mesh) of the MifM-SRC for **d**, L22 (orange), **e**, L4 (cyan) and the **f**, 50S subunit.



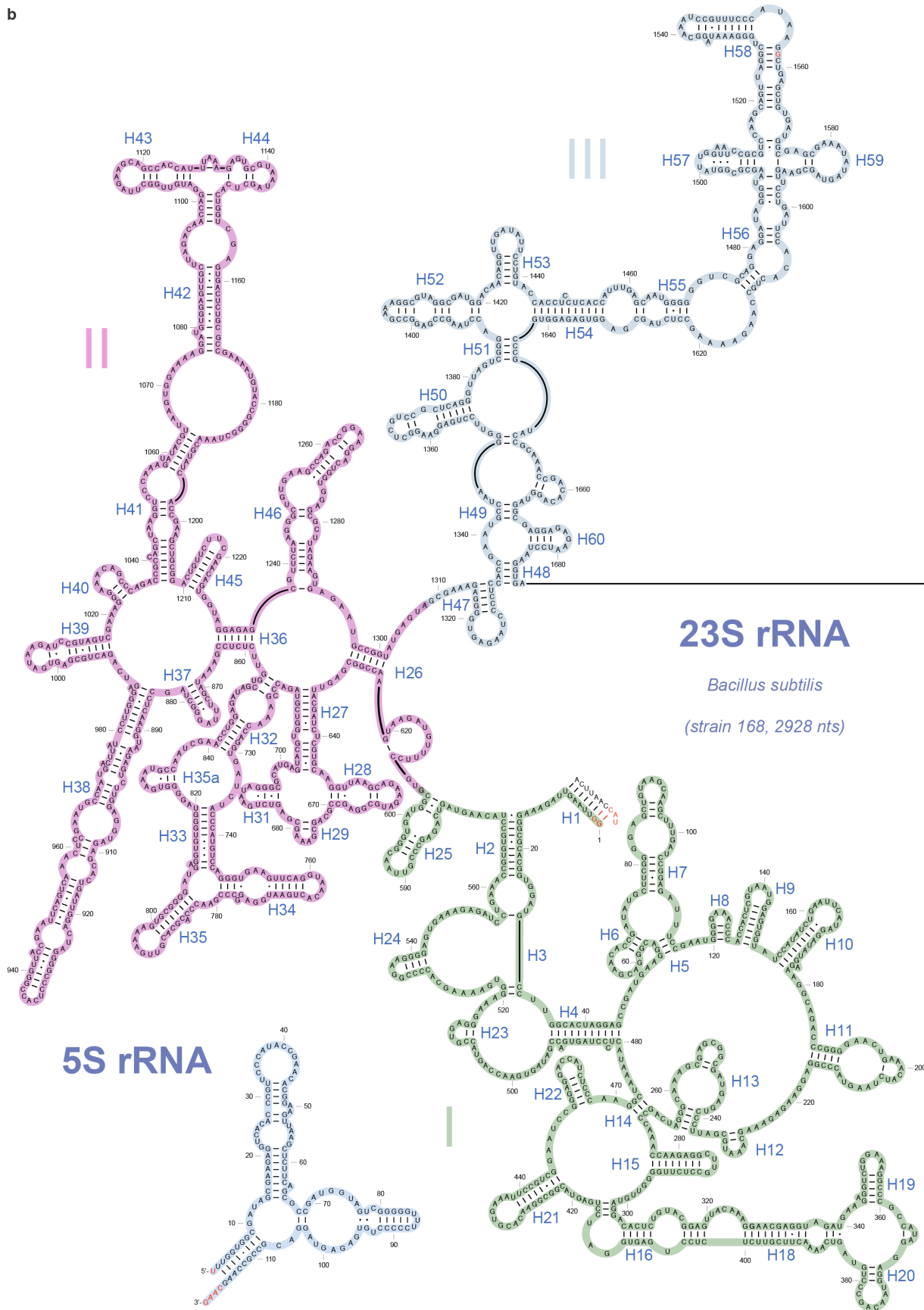
**Supplementary Figure 6: Molecular model for the *Bacillus subtilis* 70S ribosome**

**a-d**, Molecular model for the *B. subtilis* **a,b**, large 50S subunit and **c,d**, small 30S subunit. In **b**, and **d**, the molecular models are shown within the electron density (grey mesh) for the cryo-EM map of the MifM-SRC. Ribosomal proteins are coloured distinctly and labeled using the new nomenclature<sup>6</sup>.

a

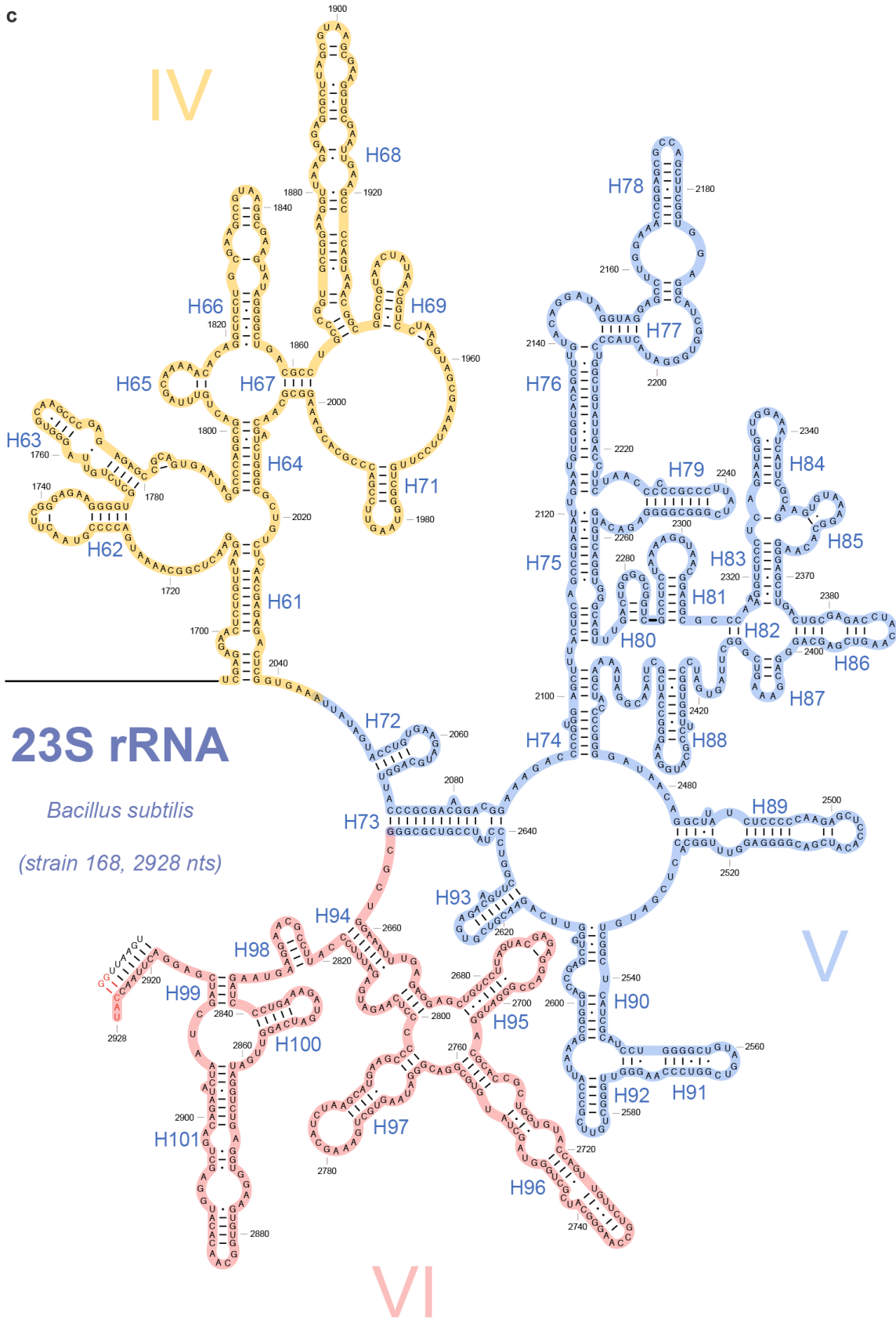


b



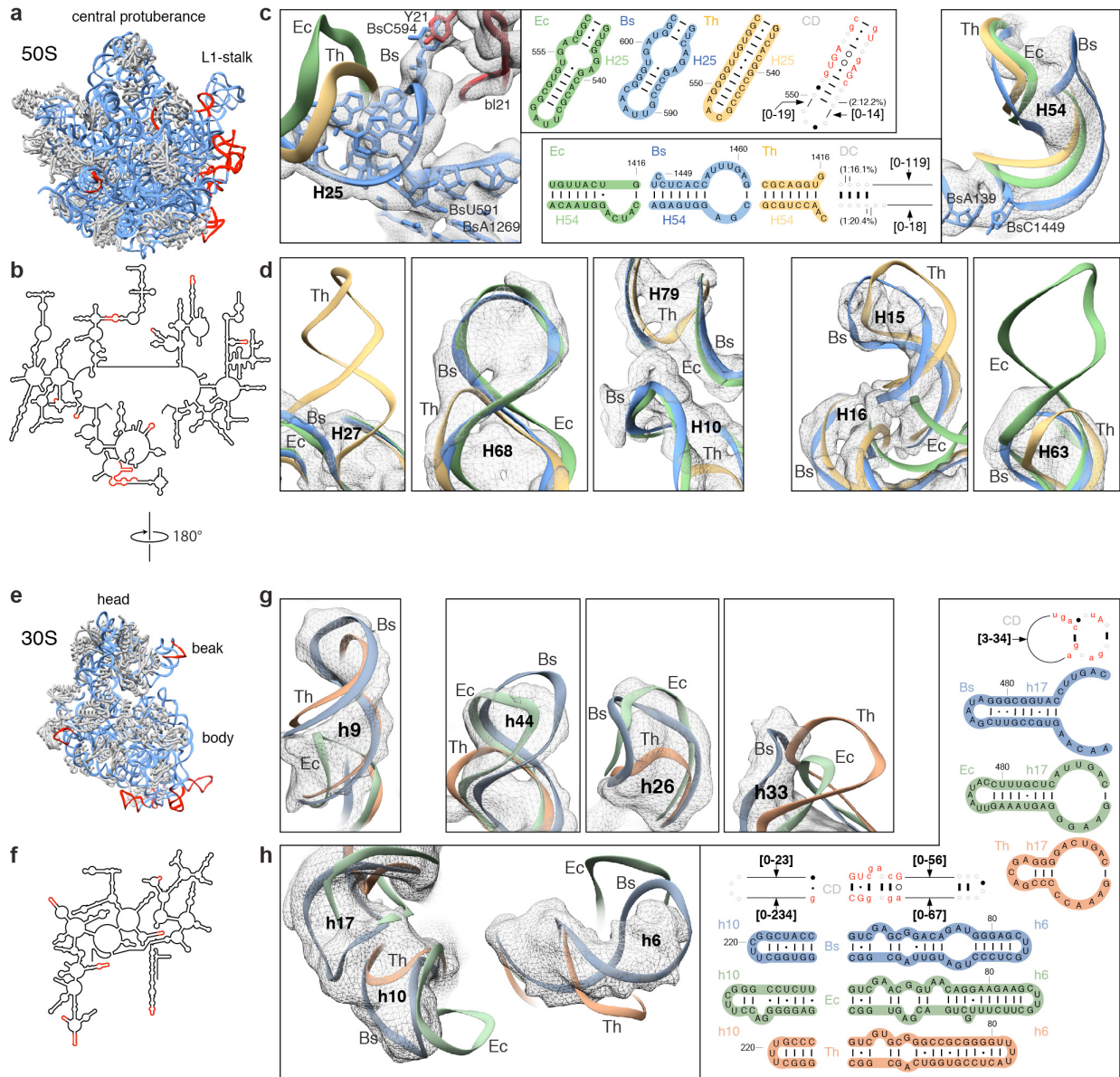


C



**Supplementary Figure 7: Secondary structures of the *B. subtilis* 5S, 16S and 23S rRNA.**

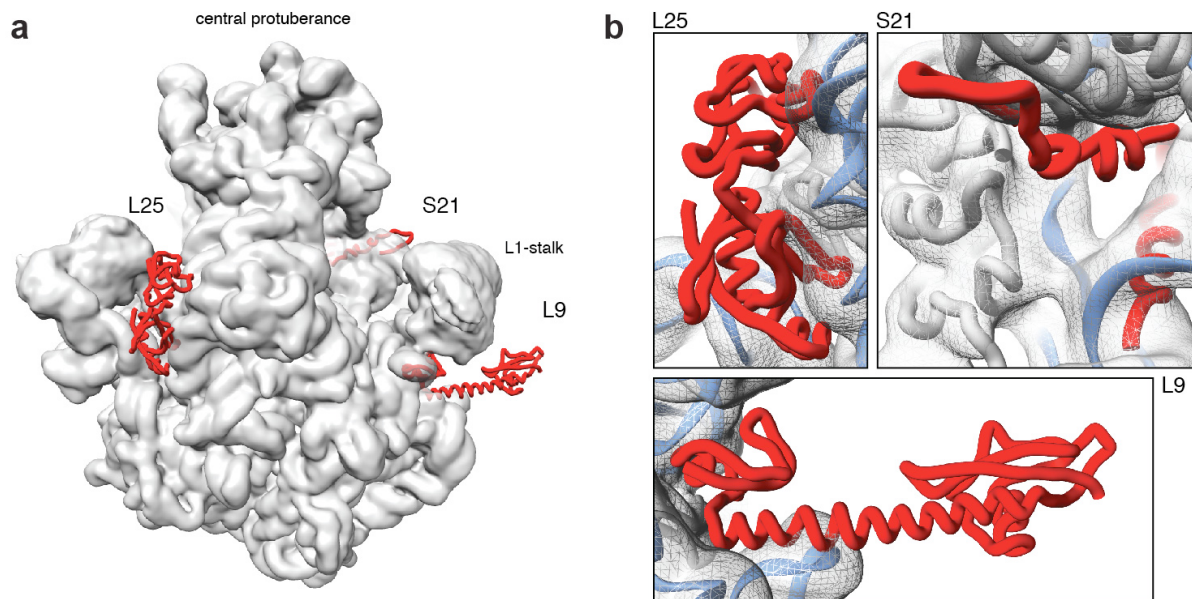
**a**, Secondary structure of the *B. subtilis* 16S rRNA coloured according to the 5' (blue), central (C, magenta), 3' major (3'M, green) and 3' minor (3'm, yellow) domains. **b-c**, Secondary structure of the *B. subtilis* 5S and 23S rRNA coloured differently for domains I-VI. The secondary structures were modified from that available at the comparative RNA web (CRW) site<sup>7</sup> based on the molecular model of the *B. subtilis* 16S rRNA from the cryo-EM structure of the MifM-SRC.



**Supplementary Figure 8: Comparison of *B. subtilis*, *E. coli* and *T. thermophilus* 70S ribosome structures.**

**a-d**, Regions of difference in 23S rRNA (red) mapped onto **a**, the *B. subtilis* 50S subunit (rRNA, blue and r-proteins, grey) and **b**, the secondary *B. subtilis* 23S rRNA. **c**, Panels show electron density (grey mesh) and rRNA models (blue) as well as secondary structures for regions of the *B. subtilis* 50S subunit that are distinct from the equivalent regions of the *E. coli* (green) and *T. thermophilus* (yellow) 50S subunit, namely H25 and H54. **d**, Panels show electron density (grey mesh) and rRNA models (blue) for regions of the *B. subtilis* 50S subunit that are either similar to *E. coli* (green, PDB4KIX/Y)<sup>3, 4</sup> and distinct from *T. thermophilus* (yellow, PDB318H/I)<sup>5</sup>, for example H27, H68 and H10/H79, or similar to *T. thermophilus* and distinct

from *E. coli*, namely, H16 and H63. **e-h**, Regions of difference in 16S rRNA (red) mapped onto **e**, the *B. subtilis* 30S subunit (rRNA, blue and r-proteins, grey) and **f**, the secondary *B. subtilis* 16S rRNA. **g**, Panels show electron density (grey mesh) and rRNA models (blue) for regions of the *B. subtilis* 50S subunit that are either similar to *E. coli* and distinct from *T. thermophilus*, for example h9 or similar to *T. thermophilus* and distinct from *E. coli*, namely, h44, h26 and h33. **h**, Panels show electron density (grey mesh) and 16S rRNA models (blue) as well as secondary structures for regions of the *B. subtilis* 30S subunit that are distinct from the equivalent regions of the *E. coli* (green) and *T. thermophilus* (yellow) 30S subunit, namely h10/h17 and h6.



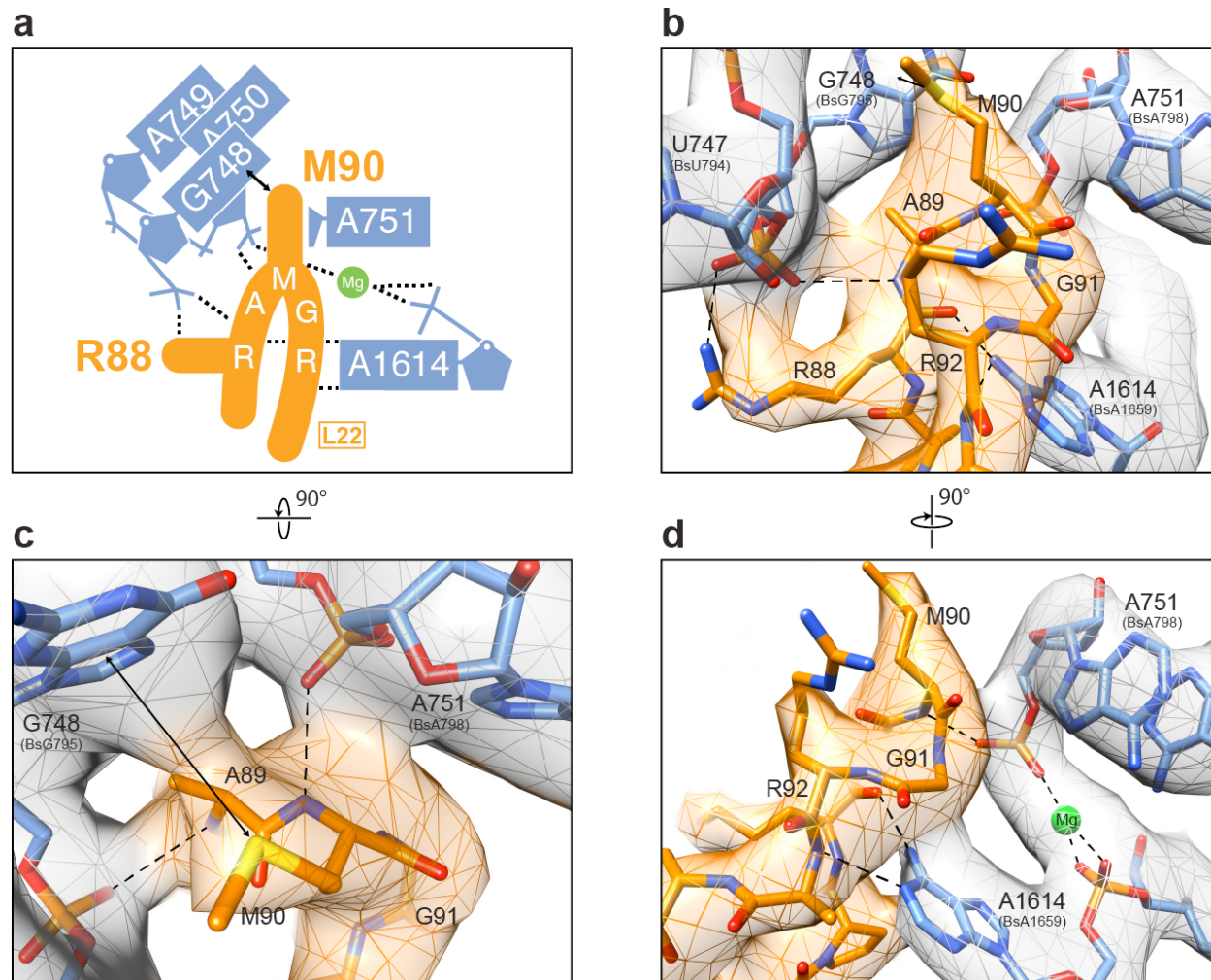
**Supplementary Figure 9: Differences in ribosomal proteins between *B. subtilis*, *E. coli* and *T. thermophilus* ribosomes.**

**a**, Overview and **b**, zoom of the cryo-EM map of the MifM-SRC (grey with map filtered to  $\sim 10$  Å), showing the absence of density for ribosomal proteins S21, L9 and L25. The binding positions of S21 and L9 were taken from the *E. coli* (PDB4KIX/Y)<sup>3,4</sup> and for L25 (red) from the *T. thermophilus* 70S ribosome (PDB3I8I)<sup>5</sup>.



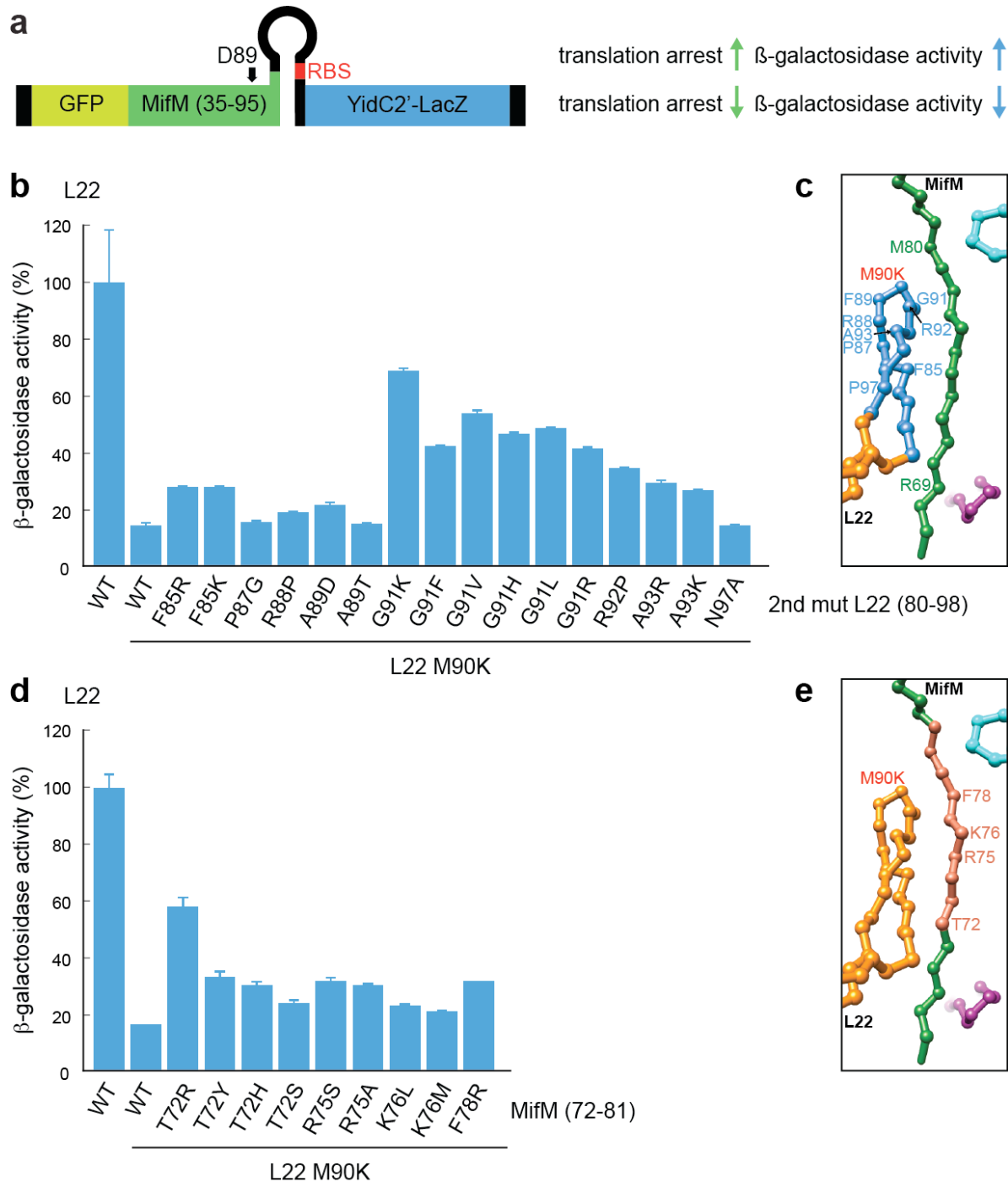
**Supplementary Figure 10: Sequence alignment of the tunnel lumen region of bacterial ribosomal protein L22.**

Boxshade representation of a Clustal W sequence alignment of selected bacterial ribosomal protein L22 sequences to highlight the sequence diversity at the position equivalent M90 in *B. subtilis* L22 (arrowed).



**Supplementary Figure 11: Network of interactions between the tip of the  $\beta$ -hairpin of L22 and H35 of the 23S rRNA.**

**a**, Schematic illustrating the network of interactions between the tip of the  $\beta$ -hairpin of L22 and the residues in H35 of the 23S rRNA. **b**, The backbone and the side chain of R88 of L22 are within hydrogen bonding distance to the phosphate-oxygens of G748. **c**, M90 of L22 can form sulphur-Pi interaction with the nucleobase of G748, whereas the backbone nitrogen of M90 is within hydrogen bonding distance to the phosphate-oxygen of A751. **d**, The backbone phosphate-oxygens of A751 and A1614 coordinate a  $Mg^{2+}$  ion, while the nucleobase of A1614 comes within hydrogen bonding distance of the backbone of A89 and R92 of L22.



**Supplementary Figure 12: Identification of second site mutations in MifM or L22 that restore stalling in the context of the L22-M90K mutation.** **a**, Schematic for the GFP-MifM-YidC2'-LacZ induction reporter used to monitor translational arrest via  $\beta$ -galactosidase activity in *B. subtilis* *in vivo*. In contrast to the GFP-MifM-LacZ reporter (see Figure 3g), translational arrest using the GFP-MifM-YidC2'-LacZ reporter leads to induction of  $\beta$ -galactosidase activity. **b**,  $\beta$ -galactosidase activity from the GFP-MifM-YidC2'-LacZ reporter using *B. subtilis* strains



bearing the *B. subtilis* L22-M90K mutation selected for second site suppressor mutations within L22 (residues 80-98) that restore stalling and induction of  $\beta$ -galactosidase activity. **c**, Overview of relative positions of MifM to tunnel lumen residues of L22 that rescue the L22-M90K mutation. **d**,  $\beta$ -galactosidase activity from the GFP-MifM-YidC2'-LacZ reporter when using *B. subtilis* strains bearing the *B. subtilis* L22-M90K mutation selected for second site suppressors within MifM (residues 72-81) that restore stalling and induction of  $\beta$ -galactosidase activity. **e**, Overview of relative positions of tunnel lumen residues of L22 to MifM residues that rescue the L22-M90K mutation. In **b** and **d**, the error bars indicate the standard deviation of three independent biological replicates.

**Supplementary Table 1: Strains**

Strains	genotype
BKG80	<i>amyE::rbsm1-gfp-mifM<sup>35-95</sup>-lacZΩcat, rpsSΩkanΩrplV(M90C)</i>
NAB263	<i>amyE::rbsm1-gfp-mifM<sup>35-95</sup>-lacZΩcat, rpsSΩkanΩrplV(M90V)</i>
NAB264	<i>amyE::rbsm1-gfp-mifM<sup>35-95</sup>-lacZΩcat, rpsSΩkanΩrplV(M90D)</i>
NAB265	<i>amyE::rbsm1-gfp-mifM<sup>35-95</sup>-lacZΩcat, rpsSΩkanΩrplV(M90F)</i>
NAB269	<i>amyE::rbsm1-gfp-mifM<sup>35-95</sup>-lacZΩcat, rpsSΩkanΩrplV(M90S)</i>
NAB270	<i>amyE::rbsm1-gfp-mifM<sup>35-95</sup>-lacZΩcat, rpsSΩkanΩrplV(M90N)</i>
NAB272	<i>amyE::rbsm1-gfp-mifM<sup>35-95</sup>-lacZΩcat, rpsSΩkanΩrplV(M90L)</i>
NAB275	<i>amyE::rbsm1-gfp-mifM<sup>35-95</sup>-lacZΩcat, rpsSΩkanΩrplV(M90I)</i>
NAB276	<i>amyE::rbsm1-gfp-mifM<sup>35-95</sup>-lacZΩcat, rpsSΩkanΩrplV(M90R)</i>
NAB277	<i>amyE::rbsm1-gfp-mifM<sup>35-95</sup>-lacZΩcat, rpsSΩkanΩrplV(M90E)</i>
NAB278	<i>amyE::rbsm1-gfp-mifM<sup>35-95</sup>-lacZΩcat, rpsSΩkanΩrplV(M90A)</i>
NAB279	<i>amyE::rbsm1-gfp-mifM<sup>35-95</sup>-lacZΩcat, rpsSΩkanΩrplV(M90T)</i>
NAB280	<i>amyE::rbsm1-gfp-mifM<sup>35-95</sup>-lacZΩcat, rpsSΩkanΩrplV(M90Q)</i>
NAB281	<i>amyE::rbsm1-gfp-mifM<sup>35-95</sup>-lacZΩcat, rpsSΩkanΩrplV(M90Y)</i>
NAB285	<i>amyE::rbsm1-gfp-mifM<sup>35-95</sup>-lacZΩcat, rpsSΩkanΩrplV(M90G)</i>
NAB286	<i>amyE::rbsm1-gfp-mifM<sup>35-95</sup>-lacZΩcat, rpsSΩkanΩrplV(M90P)</i>
NAB287	<i>amyE::rbsm1-gfp-mifM<sup>35-95</sup>-lacZΩcat, rpsSΩkanΩrplV(M90W)</i>
NAB291	<i>amyE::rbsm1-gfp-mifM<sup>35-95</sup>-lacZΩcat, rpsSΩkanΩrplV(M90H)</i>
SCB824	<i>amyE::gfp-mifM<sup>35-95</sup>-yidC2'-lacZΩcat</i>
SCB2592	<i>amyE::rbsm1-gfp-mifM<sup>35-95</sup>-lacZΩcat</i>
SCB2619	<i>rplWΩkanΩrplB</i>
SCB2634	<i>rplW(d65-69)Ωkan</i>
SCB2656	<i>rpsSΩkan</i>
SCB2939	<i>rplD(d63-67)Ωkan</i>
SCB2942	<i>rplD(d66-70)Ωkan</i>
SCB2912	<i>rpsSΩkanΩrplV(d82-86)</i>
SCB2917	<i>rpsSΩkanΩrplV(d86-90)</i>
SCB2920	<i>rpsSΩkanΩrplV(d91-95)</i>
SCB3348	<i>rpsSΩkanΩrplV(M90K)</i>
SCB2639	<i>amyE::rbsm1-gfp-mifM<sup>35-95</sup>-lacZΩcat, rplWΩkan</i>
SCB2639	<i>amyE::rbsm1-gfp-mifM<sup>35-95</sup>-lacZΩcat, rplWΩkan</i>
SCB2655	<i>amyE::rbsm1-gfp-mifM<sup>35-95</sup>-lacZΩcat, rplW(d65-69)Ωkan</i>
SCB2924	<i>amyE::gfp-mifM<sup>35-95</sup>-yidC2'-lacZΩcat, rpsSΩkanΩrplV</i>

SCB2956 *amyE::rbsm1-gfp-mifM<sup>35-95</sup>-lacZΩcat, rplD(d63-67)Ωkan*  
SCB2957 *amyE::rbsm1-gfp-mifM<sup>35-95</sup>-lacZΩcat, rplD(d66-70)Ωkan*  
SCB2958 *amyE::rbsm1-gfp-mifM<sup>35-95</sup>-lacZΩcat, rpsSΩkanΩrplV*  
SCB2959 *amyE::rbsm1-gfp-mifM<sup>35-95</sup>-lacZΩcat, rpsSΩkanΩrplV(d82-86)*  
SCB2960 *amyE::rbsm1-gfp-mifM<sup>35-95</sup>-lacZΩcat, rpsSΩkanΩrplV(d86-90)*  
SCB2961 *amyE::rbsm1-gfp-mifM<sup>35-95</sup>-lacZΩcat, rpsSΩkanΩrplV(d91-95)*  
SCB3247 *amyE::rbsm1-gfp-mifM<sup>35-95</sup>-lacZΩcat, rpsSΩkanΩrplV(M90K)*  
SCB3251 *amyE::rbsm1-gfp-mifM<sup>35-95</sup>-lacZΩcat, rpsSΩkanΩrplV(Ec-tip)*  
SCB3272 *amyE::rbsm1-gfp-mifM<sup>35-95</sup>-lacZΩcat, rpsSΩkanΩrplV(Ec-tip)-I85F/M86R*  
SCB3274 *amyE::rbsm1-gfp-mifM<sup>35-95</sup>-lacZΩcat, rpsSΩkanΩrplV(Ec-tip)-K90M*  
SCB3353 *amyE::gfp-mifM<sup>35-95</sup>-yidC2'-lacZΩcat, rpsSΩkanΩrplV(M90K)*  
SCB3353 *amyE::gfp-mifM<sup>35-95</sup>-yidC2'-lacZΩcat, rpsSΩkanΩrplV(M90K)*  
SCB3378 *amyE::gfp-mifM<sup>35-95</sup>-yidC2'-lacZΩcat, rplV94*  
SCB3523 *amyE::gfp-mifM(T72H)<sup>35-95</sup>-yidC2'-lacZΩcat, rpsSΩkanΩrplV(M90K)*  
SCB3524 *amyE::gfp-mifM(T72Y)<sup>35-95</sup>-yidC2'-lacZΩcat, rpsSΩkanΩrplV(M90K)*  
SCB3525 *amyE::gfp-mifM(T72S)<sup>35-95</sup>-yidC2'-lacZΩcat, rpsSΩkanΩrplV(M90K)*  
SCB3527 *amyE::gfp-mifM(R75S)<sup>35-95</sup>-yidC2'-lacZΩcat, rpsSΩkanΩrplV(M90K)*  
SCB3529 *amyE::gfp-mifM(K76M)<sup>35-95</sup>-yidC2'-lacZΩcat, rpsSΩkanΩrplV(M90K)*  
SCB3530 *amyE::gfp-mifM(K76L)<sup>35-95</sup>-yidC2'-lacZΩcat, rpsSΩkanΩrplV(M90K)*  
SCB3532 *amyE::gfp-mifM(F78R)<sup>35-95</sup>-yidC2'-lacZΩcat, rpsSΩkanΩrplV(M90K)*  
SCB3533 *amyE::gfp-mifM(R75A)<sup>35-95</sup>-yidC2'-lacZΩcat, rpsSΩkanΩrplV(M90K)*  
SCB3572 *amyE::gfp-mifM<sup>35-95</sup>-yidC2'-lacZΩcat, rpsSΩkanΩrplV(M90K/F85R)*  
SCB3573 *amyE::gfp-mifM<sup>35-95</sup>-yidC2'-lacZΩcat, rpsSΩkanΩrplV(M90K/F85K)*  
SCB3574 *amyE::gfp-mifM<sup>35-95</sup>-yidC2'-lacZΩcat, rpsSΩkanΩrplV(M90K/P87G)*  
SCB3575 *amyE::gfp-mifM<sup>35-95</sup>-yidC2'-lacZΩcat, rpsSΩkanΩrplV(M90K/R88P)*  
SCB3576 *amyE::gfp-mifM<sup>35-95</sup>-yidC2'-lacZΩcat, rpsSΩkanΩrplV(M90K/A89D)*  
SCB3577 *amyE::gfp-mifM<sup>35-95</sup>-yidC2'-lacZΩcat, rpsSΩkanΩrplV(M90K/A89T)*  
SCB3578 *amyE::gfp-mifM<sup>35-95</sup>-yidC2'-lacZΩcat, rpsSΩkanΩrplV(M90K/G91K)*  
SCB3579 *amyE::gfp-mifM<sup>35-95</sup>-yidC2'-lacZΩcat, rpsSΩkanΩrplV(M90K/G91L)*  
SCB3580 *amyE::gfp-mifM<sup>35-95</sup>-yidC2'-lacZΩcat, rpsSΩkanΩrplV(M90K/G91F)*  
SCB3581 *amyE::gfp-mifM<sup>35-95</sup>-yidC2'-lacZΩcat, rpsSΩkanΩrplV(M90K/G91H)*  
SCB3582 *amyE::gfp-mifM<sup>35-95</sup>-yidC2'-lacZΩcat, rpsSΩkanΩrplV(M90K/G91Y)*  
SCB3583 *amyE::gfp-mifM<sup>35-95</sup>-yidC2'-lacZΩcat, rpsSΩkanΩrplV(M90K/R92P)*  
SCB3584 *amyE::gfp-mifM<sup>35-95</sup>-yidC2'-lacZΩcat, rpsSΩkanΩrplV(M90K/A93R)*  
SCB3585 *amyE::gfp-mifM<sup>35-95</sup>-yidC2'-lacZΩcat, rpsSΩkanΩrplV(M90K/A93K)*  
SCB3586 *amyE::gfp-mifM<sup>35-95</sup>-yidC2'-lacZΩcat, rpsSΩkanΩrplV(M90K/G91R)*

SCB3605 *amyE::gfp-mifM<sup>35-95</sup>-yidC2'-lacZΩcat, rpsSΩkanΩrplV(M90K/N97A)*  
SCB3607 *amyE::gfp-mifM(T72R)<sup>35-95</sup>-yidC2'-lacZΩcat, rpsSΩkanΩrplV(M90K)*  
SCB3610 *amyE::gfp-mifM(T72R)<sup>35-95</sup>-yidC2'-lacZΩcat, rpsSΩkanΩrplV*

---

**Supplementary Table 2: Plasmids**

plasmid	genotype	Primer	Template	ref
pCH735	<i>amyE::mifM-lacZΩcat</i>			8
pCH835	<i>amyE::GFP-mifM<sup>35-95</sup>-yidC2'-lacZΩcat</i>			8
pCH913	<i>amyE::rbsm1-GFP-mifM<sup>35-95</sup>-yidC2'-lacZΩcat</i>			8
pCH1142	<i>spcRΩloxP-kanR-loxP</i>			9
pCH1517	<i>amyE::rbsm1-GFP-mifM<sup>35-95</sup>-lacZΩcat</i>			This study
pCH1570	<i>rplWΩkanΩrplB</i>			This study
pCH1584	<i>rplW(d65-69)ΩkanΩrplB</i>	SP1	pCH1570	This study
pCH1587	<i>rpsSΩkanΩrplV</i>			This study
pCH1590	<i>rpsSΩkanΩrplV(d82-86)</i>	SP2	pCH1587	This study
pCH1591	<i>rpsSΩkanΩrplV(d86-90)</i>	SP3	pCH1587	This study
pCH1592	<i>rpsSΩkanΩrplV(d91-95)</i>	SP4	pCH1587	This study
pCH1745	<i>spcRΩrplCD(d63-67)WΩkanΩrplB</i>	SP5	pCH1744	This study
pCH1746	<i>spcRΩrplCD(d66-70)WΩkanΩrplB</i>	SP6	pCH1744	This study
pCH1897	<i>rpsSΩkanΩrplV(M90K)</i>	SP7	pCH1587	This study
pCH1901	<i>rpsSΩkanΩrplV(Ec-tip)</i>			This study
pCH1904	<i>rpsSΩkanΩrplV(Ec-tip)-I85F-M86R</i>	SP8	pCH1901	This study
pCH1905	<i>rpsSΩkanΩrplV(Ec-tip)-K90M</i>	SP9	pCH1901	This study
pCH1958	<i>amyE::GFP-mifM(T72H)<sup>35-95</sup>-yidC2'-lacZΩcat</i>	SP10	pCH835	This study
pCH1959	<i>amyE::GFP-mifM(T72Y)<sup>35-95</sup>-yidC2'-lacZΩcat</i>	SP11	pCH835	This study
pCH1960	<i>amyE::GFP-mifM(T72S)<sup>35-95</sup>-yidC2'-lacZΩcat</i>	SP12	pCH835	This study
pCH1962	<i>amyE::GFP-mifM(R75S)<sup>35-95</sup>-yidC2'-lacZΩcat</i>	SP13	pCH835	This study
pCH1964	<i>amyE::GFP-mifM(K76M)<sup>35-95</sup>-yidC2'-lacZΩcat</i>	SP14	pCH835	This study
pCH1965	<i>amyE::GFP-mifM(K76L)<sup>35-95</sup>-yidC2'-lacZΩcat</i>	SP15	pCH835	This study
pCH1967	<i>amyE::GFP-mifM(F78R)<sup>35-95</sup>-yidC2'-lacZΩcat</i>	SP16	pCH835	This study
pCH1968	<i>amyE::GFP-mifM(R75A)<sup>35-95</sup>-yidC2'-lacZΩcat</i>	SP17	pCH835	This study
pCH1978	<i>amyE::GFP-mifM(T72R)<sup>35-95</sup>-yidC2'-lacZΩcat</i>	SP18	pCH835	This study
pEB71	<i>loxP-kanR-loxP</i>			10
pKG11	<i>rpsSΩkanΩrplV(M90C)</i>	SP19	pCH1587	This study
pNAR479	<i>rpsSΩkanΩrplV(M90V)</i>	SP20	pCH1587	This study
pNAR483	<i>rpsSΩkanΩrplV(M90D)</i>	SP20	pCH1587	This study
pNAR484	<i>rpsSΩkanΩrplV(M90F)</i>	SP20	pCH1587	This study
pNAR485	<i>rpsSΩkanΩrplV(M90S)</i>	SP20	pCH1587	This study

pNAR487	<i>rpsSΩkanΩrplV(M90N)</i>	SP20	pCH1587	This study
pNAR489	<i>rpsSΩkanΩrplV(M90L)</i>	SP20	pCH1587	This study
pNAR490	<i>rpsSΩkanΩrplV(M90I)</i>	SP20	pCH1587	This study
pNAR491	<i>rpsSΩkanΩrplV(M90R)</i>	SP20	pCH1587	This study
pNAR494	<i>rpsSΩkanΩrplV(M90E)</i>	SP20	pCH1587	This study
pNAR495	<i>rpsSΩkanΩrplV(M90A)</i>	SP20	pCH1587	This study
pNAR496	<i>rpsSΩkanΩrplV(M90T)</i>	SP20	pCH1587	This study
pNAR498	<i>rpsSΩkanΩrplV(M90Q)</i>	SP20	pCH1587	This study
pNAR499	<i>rpsSΩkanΩrplV(M90Y)</i>	SP21	pCH1587	This study
pNAR500	<i>rpsSΩkanΩrplV(M90G)</i>	SP22	pCH1587	This study
pNAR501	<i>rpsSΩkanΩrplV(M90P)</i>	SP23	pCH1587	This study
pNAR502	<i>rpsSΩkanΩrplV(M90W)</i>	SP24	pCH1587	This study
pNAR503	<i>rpsSΩkanΩrplV(M90H)</i>	SP25	pCH1587	This study

---

---

**Supplementary Table 3: Primers**

---

primer	sequence
MifMfor	5'- GAAATTAATACGACTCACTATAGGG -3'
MifMrev	5'- TTATTATTATTAGTCTTCCTCATCGTTCACAGG -3'
SP1	5'-ACTACAAAGGCAAATCAAAAAGTGGTATGACTAGCCGTCG-3'
SP2	5'-TCGTTGACGAAGGCCCTACGCCACGTGCTATGGGACGTGC-3'
SP3	5'-GCCCTACGTTAAAAAGATTCGGACGTGCGAGCCAAATCAA-3'
SP4	5'-GATTCCGCCCACGTGCTATGATCAACAAACGTACGAGCCA-3'
SP5	5'-AAGTACGCGGCGGAGGTCGTAAAGGTACTGGACGTGCCCG-3'
SP6	5'-GCGGAGGTTCGTAAACCATGGGGACGTGCCCGTCAAGGTTC-3'
SP7	5'-AGATTCCGCCCACGTGCTAAAGGACGTGCGAGCCAAATC-3'
SP8	5'-GGCCCTAGCATGAAGCGCTTCGCCCAGTGCAAAAGGTTCGT-3'
SP9	5'-CGCATTATGCCGCGTGCAATGGGTTCGTGCAGATCGCATC-3'
SP10	5'-ATATATCATCGCATTACACATTGGATACGTAAAGTCTTC-3'
SP11	5'-ATATATCATCGCATTACATATTGGATACGTAAAGTCTTC-3'
SP12	5'-ATATATCATCGCATTACAAGCTGGATACGTAAAGTCTTC-3'
SP13	5'-CGCATTACAACCTGGATAAGCAAAGTCTTCGCATGAAT-3'
SP14	5'-ATTACAACCTGGATACGTATGGTCTTCGCATGAATTCG-3'
SP15	5'-ATTACAACCTGGATACGTCTGGTCTTCGCATGAATTCG-3'
SP16	5'-ACTTGGATACGTAAAGTCCGCCGCATGAATTCGCCTGTG-3'
SP17	5'-CGCATTACAACCTGGATAGCGAAAGTCTTCGCATGAA-3'
SP18	5'-ATATATCATCGCATTACACGCTGGATACGTAAAGTCTTC-3'
SP19	5'-AGATTCCGCCCACGTGCTTGGCGACGTGCGAGCCAAATC-3'
SP20	5'-AGATTCCGCCCACGTGCTNNNGGACGTGCGAGCCAAATC-3'
SP21	5'-AGATTCCGCCCACGTGCTTATGGACGTGCGAGCCAAATC-3'
SP22	5'-AGATTCCGCCCACGTGCTGGCGGACGTGCGAGCCAAATC-3'
SP23	5'-AGATTCCGCCCACGTGCTCCGGGACGTGCGAGCCAAATC-3'
SP24	5'-AGATTCCGCCCACGTGCTTGGGGACGTGCGAGCCAAATC-3'
SP25	5'-AGATTCCGCCCACGTGCTCATGGACGTGCGAGCCAAATC-3'
SP26	5'-CAAGGAATGGTGCATGCAAGG-3'
SP27	5'-TCCATAGATCTTAAAAGAAGAGAACCGGCGTC-3'
SP28	5'-AATAAGAGCTCCAGGTTCTTGGTAAAGAAGGT-3'
SP29	5'-TCCTCTTTTCTACAGTATTTAGGATCCTTAGCGTCTTGTTTTTTGTC-3'
SP30	5'-TTTTACTGGATGAATTGTTTTAGCTCGAGTAATTAACAAGACGCTAAGAGAGGA-3'
SP31	5'-TTATTGCATGCTCTTCAGACACACCAGAAATT-3'

SP32 5'-TAAGGATCCTAAATACTGTAGAAAAGAGGA-3'  
SP33 5'-AATTACTCGAGCTAAAACAATTCATCCAGTAAAA-3'  
SP34 5'-GGGAAGAACAGTATGTTCGAGC-3'  
SP35 5'-CTGCACGACCTTTTGCACGCGGCATAATGCGCTTCATGCTAGGGCCTTCGTCAACGAATGC-3'  
SP36 5'-GCATTATGCCGCGTGCAAAAGGTCGTGCAGATCGCATCCTGAAACGTACGAGCCACATTACA-3'  
SP37 5'-TGTGGAATTGTGAGCGG-3'  
SP38 5'-GCATTTCGTTGACGAAGGCNNNACGTTAAAAAGATTCCGC-3'  
SP39 5'-TTCGTTGACGAAGGCCCTNNNTTAAAAAGATTCCGCCCA-3'  
SP40 5'-GTTGACGAAGGCCCTACGNNNAAAAGATTCCGCCACGT-3'  
SP41 5'-GACGAAGGCCCTACGTTANNNAGATTCCGCCACGTGCT-3'  
SP42 5'-GAAGGCCCTACGTTAAAAANNNTTCCGCCACGTGCTAAA-3'  
SP43 5'-GGCCCTACGTTAAAAAGANNNGCCACGTGCTAAAGGA-3'  
SP44 5'-CCTACGTTAAAAAGATTCNNNCCACGTGCTAAAGGACGT-3'  
SP45 5'-ACGTTAAAAAGATTCCGCNNNCGTGCTAAAGGACGTGCG-3'  
SP46 5'-TTAAAAAGATTCCGCCANNGCTAAAGGACGTGCGAGC-3'  
SP47 5'-AAAAGATTCCGCCACGTNNNAAAGGACGTGCGAGCCAA-3'  
SP48 5'-TTCCGCCACGTGCTAAANNNGTGCGAGCCAAATCAAC-3'  
SP49 5'-CGCCCACGTGCTAAAGGANNGCGAGCCAAATCAACAAA-3'  
SP50 5'-CCACGTGCTAAAGGACGTNNNAGCCAAATCAACAAACGT-3'  
SP51 5'-CGTGCTAAAGGACGTGCGNNNCAAATCAACAAACGTACG-3'  
SP52 5'-GCTAAAGGACGTGCGAGCNNNATCAACAAACGTACGAGC-3'  
SP53 5'-AAAGGACGTGCGAGCCAAANNNAACAAACGTACGAGCCAC-3'  
SP54 5'-GGACGTGCGAGCCAAATCNNNAAACGTACGAGCCACATT-3'  
SP55 5'-CGTGCGAGCCAAATCAACNNNCGTACGAGCCACATTACA-3'  
SP56 5'-ATATATCATCGCATTACANNNTGGATACGTAAAGTCTTC-3'  
SP57 5'-TATCATCGCATTACAACNNNATAACGTAAAGTCTTCCGC-3'  
SP58 5'-CATCGCATTACAACCTTGGNNNCGTAAAGTCTTCCGCATG-3'  
SP59 5'-CGCATTACAACCTTGGATANNNAAAGTCTTCCGCATGAAT-3'  
SP60 5'-ATTACAACCTTGGATACGTNNNGTCTTCCGCATGAATTCG-3'  
SP61 5'-ACAACCTTGGATACGTAAANNNTTCCGCATGAATTCGCCT-3'  
SP62 5'-ACTTGGATACGTAAAGTCNNNCGCATGAATTCGCCTGTG-3'  
SP63 5'-TGGATACGTAAAGTCTTCNNNATGAATTCGCCTGTGAAC-3'  
SP64 5'-ATACGTAAAGTCTTCCGCNNNAATTCGCCTGTGAACGAT-3'  
SP65 5'-CGTAAAGTCTTCCGCATGNNNTCGCCTGTGAACGATGAG-3'  
SP66 5'-TTTAGTCTAGATAATTATCTGTAAAAAGAAGG-3'  
SP67 5'-ATATTGCATGCACCACGAACGTAGGACGGAT-3'



SP68 5'-AATAAGGATCCCCAAAACGGTTCTACTGCTGG-3'  
SP69 5'-TTAAGCTTCAAAAATTTTCGAT-3'  
SP70 5'-GAAATCGAAATTTTGAAGCTTAAATACTGTAGAAAAGAGGAAGG-3'  
SP71 5'-CTAAAACAATTCATCCAGTAAAA-3'  
SP72 5'-AATTATCTAGACTAAAACAATTCATCCAGTAAAA-3'

---

'N' represents mixture of A, G, T and C.

## Supplementary References:

1. Scheres SH, Chen S. Prevention of overfitting in cryo-EM structure determination. *Nat Methods* **9**, 853-854 (2012).
2. Kucukelbir A, Sigworth FJ, Tagare HD. Quantifying the local resolution of cryo-EM density maps. *Nat Methods* **11**, 63-65 (2014).
3. Schuwirth B, *et al.* Structures of the bacterial ribosome at 3.5 Å resolution. *Science* **310**, 827-834 (2005).
4. Pulk A, Cate JH. Control of ribosomal subunit rotation by elongation factor G. *Science* **340**, 1235970 (2013).
5. Jenner LB, Demeshkina N, Yusupova G, Yusupov M. Structural aspects of messenger RNA reading frame maintenance by the ribosome. *Nat Struct Mol Biol* **17**, 555-560 (2010).
6. Ban N, *et al.* A new system for naming ribosomal proteins. *Curr Opin Struct Biol* **24**, 165-169 (2014).
7. Cannone JJ, *et al.* The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BioMed Central Bioinformatics* **3**, 2 (2002).
8. Chiba S, Lamsa A, Pogliano K. A ribosome-nascent chain sensor of membrane protein biogenesis in *Bacillus subtilis*. *EMBO J* **28**, 3461-3475 (2009).
9. Kumazaki K, *et al.* Structural basis of Sec-independent membrane protein insertion by YidC. *Nature* **509**, 516-520 (2014).
10. Aung S, *et al.* Dual localization pathways for the engulfment proteins during *Bacillus subtilis* sporulation. *Mol Microbiol* **65**, 1534-1546 (2007).

# **Structural basis for targeting and elongation arrest of *Bacillus* signal recognition particle**

Bertrand Beckert <sup>1</sup>, Alexej Kedrov <sup>1</sup>, Daniel Sohmen <sup>1</sup>, Georg Kempf <sup>2</sup>, Klemens Wild <sup>2</sup>, Irmgard Sinning <sup>2</sup>, Henning Stahlberg <sup>3</sup>, Daniel N. Wilson <sup>1</sup> and Roland Beckmann <sup>1,\*</sup>

<sup>1</sup> Gene Center and Center for integrated Protein Science Munich, Department of Biochemistry, Feodor-Lynen-Str. 25, University of Munich, 81377 Munich, Germany

<sup>2</sup> Heidelberg University Biochemistry Center (BZH), INF 328, D-69120 Heidelberg, Germany

<sup>3</sup> C-CINA, Biozentrum, University Basel, Mattenstrasse 26, CH-4058 Basel, Switzerland

\* Corresponding author: [beckmann@lmb.uni-muenchen.de](mailto:beckmann@lmb.uni-muenchen.de)

**Abstract:**

Signal recognition particle (SRP) recognizes signal sequences of nascent polypeptides and targets ribosome-nascent-chain complexes to membrane translocation sites. In eukaryotes, the translating ribosome is slowed-down by the Alu domain of SRP for efficient targeting. In prokaryotes, however, very little is known about structure and function of Alu domain containing SRPs. Here, we report a complete molecular model of SRP from the Gram-positive bacterium *Bacillus subtilis* based on cryo-electron microscopy. It consists only of two subunits, 6S RNA and SRP54 and facilitates elongation slow-down similar to its eukaryotic counterpart. Moreover, comparison with the ribosome-bound mammalian Alu domain shows that protein contacts with the small ribosomal subunit are substituted in bacteria by RNA-RNA interactions with the  $\alpha$ -sarcin-ricin loop and helices H43/H44 of the 23S rRNA. Our findings provide the functional and structural basis for co-translational targeting and RNA-driven elongation arrest in prokaryotes.

## Main Text:

The signal recognition particle (SRP) plays an essential role in co-translational targeting of secretory and membrane proteins mediated by the interaction with its membrane-bound cognate SRP receptor (SR)<sup>1,2</sup>. SRP is a ribonucleoprotein complex conserved in all three domains of life with a high diversity regarding size and composition<sup>3</sup>. In eukaryotes SRP is most complex and contains six proteins assembled on a 7SL RNA (**Fig. 1a**). It can be divided into two functional domains: the S domain with the SRP54 protein that recognizes N-terminal signal sequences as soon as they emerge from the ribosomal tunnel exit<sup>4</sup>, and the Alu domain that imposes an elongation arrest/slow-down by blocking the elongation factor entry site on the ribosome<sup>5-7</sup>. By arresting or delaying translation, it is believed that SRP prevents fast folding secretory proteins or aggregation-prone membrane proteins from being prematurely released from the ribosome before the ribosome-nascent chain complex (RNC) has correctly engaged with the translocation channel at the membrane<sup>1</sup>. The Alu domain of higher eukaryotes interacts with two Alu RNA binding proteins, SRP9 and SRP14 (**Fig. 1a**). The proteins stabilize the complex tertiary structure of the Alu RNA and contribute to ribosome binding by interaction with the small ribosomal subunit<sup>5,7</sup>. The Alu RNA itself establishes a direct contact with the large ribosomal subunit and thereby occludes the translation factor binding site<sup>7</sup>. Surprisingly, Gram-negative bacteria such as *E. coli* possess only a minimalistic SRP comprising a short RNA (4.5S) and the SRP54 subunit (Ffh) (**Fig. 1a**). This SRP lacks an Alu domain and, accordingly, is thought to be unable to promote elongation arrest. However, many prokaryotes such as archaea and Gram-positive bacteria possess an SRP with a large RNA (6S and 7S RNA, respectively; **Fig. 1a**) that also contains an Alu domain, usually carrying somewhat extended 5' and 3' regions compared to higher eukaryotes<sup>3</sup>. Yet, homologues of the SRP9/14 proteins have not been found in the genomes of archaea or bacteria so far. Crystal structures of the isolated RNA Alu domain of *B. subtilis* and *P. horikoshii* have recently been reported<sup>7,8</sup> and, as expected, show a high degree of similarity to each other (**Extended Data Fig. 1**) and to the eukaryotic one<sup>8,9</sup>. However, very little is known about the function of prokaryotic Alu containing SRPs in the context of the translating ribosome.

## RNC binding and elongation arrest

For the characterization of a prokaryotic, Alu domain-containing SRP we chose that of the Gram-positive bacterium *B. subtilis* that consists of a 6S RNA and an SRP54 protein (**Fig. 1a**). We reconstituted *B. subtilis* SRP (*BsSRP*) from *in vitro* transcribed RNA and purified recombinant SRP54 (**Extended Data Fig. 2a**). The histone-like DNA-binding protein HU1, that had previously been suggested to be part of the bacterial Alu domain<sup>10,11</sup> did not co-purify during *in vivo* pull-outs and could not be co-reconstituted (**Extended Data Fig. 2a**). We, therefore, conclude that HU1 lacks specific Alu RNA association or any function in SRP, and that the 6S RNA and the SRP54 subunit represent the complete set of *BsSRP* constituents.

Activity of the reconstituted SRP was first tested in pelleting assays for binding to either empty ribosomes or *B. subtilis* ribosome-nascent chain complexes (*BsRNCs*). The *BsRNCs* were generated using a cell-free translation system and carried a nascent peptide containing the FtsQ signal anchor sequence fused to the *B. subtilis* MifM stalling sequence (**Extended Data Fig. 2b**)<sup>12</sup>. While binding of the reconstituted *BsSRP* to empty 70S ribosomes was not detected, a specific and stable interaction with the signal sequence carrying RNCs was clearly observable (**Fig. 1b** and **Extended Data Fig. 2c**). Microscale thermophoresis (MST) was also used to further characterize the binding properties of the *BsSRP* to *BsRNCs*. To that end programmed and empty ribosomes were labeled with fluorescent dyes (**Extended Data Fig. 3a**) and different SRPs were titrated, so the complex assembly could be probed due to changes in ribosome thermophoretic. For the *BsSRP* we observed association constants in the low nanomolar range (5-15 nM) with *BsRNCs* carrying the signal anchor sequence of FtsQ (**Fig. 1c**). An even higher association constant was determined for an artificial more hydrophobic leucine-rich signal sequence (**Fig. 1c**), indicating that *BsSRP* has a preference for high hydrophobicity, as observed also for *E. coli* SRP<sup>13</sup>. In contrast, *BsSRP* exhibited a very low association rate for *BsRNCs* lacking a signal sequence (luciferase) and was too low to be determined for empty ribosomes under our conditions (**Extended Data Fig. 3b**). Validating these findings and in agreement with the literature<sup>1,14</sup>, similar binding properties and affinities (in the low nanomolar range) were observed using MST for *E. coli* SRP and *E. coli* RNCs carrying the FtsQ or the Foc signal sequences; also here the affinity to empty 70S ribosomes could not be determined (**Extended Data Fig. 3c**). Notably, *B. subtilis* and *E. coli* SRPs displayed essentially the same binding behavior when using *E. coli* RNCs carrying the Foc signal sequence (**Extended Data Fig. 3c**).

These data suggest that the overall binding properties of *Bs*SRP essentially resemble those of *E. coli* SRP due to a high degree of structural and functional conservation of both the SRP S-domain and the SRP binding site at the ribosomal tunnel exit. This is in agreement with the earlier finding that the *B. subtilis* and *E. coli* SRPs can indeed complement for each other *in vivo*<sup>15-18</sup>. However, unlike *E. coli* SRP, *Bs*SRP contains an Alu domain, therefore we tested whether it can induce elongation arrest similar to its eukaryotic counterparts. Using an *in vitro* cell free translation system the SRP substrate protein FtsQ was translated carrying either a mutated dysfunctional or a wild-type signal anchor sequence (**Extended Data Fig. 3d**). When translating the mutated FtsQ the presence of *Bs*SRP did not cause any detectable differences in translation efficiency, however, translation of the wild-type FtsQ in the presence of *Bs*SRP led to a delay in expression (**Fig. 1d**). Therefore, we conclude that similar to eukaryotic SRPs<sup>19,20</sup> the presence of the Alu domain in bacterial SRPs, such as *B. subtilis* SRP, confers elongation arrest/slow-down activity. By analogy, it is likely that the Alu domain containing SRPs of the *archaeal* kingdom are also endowed with an elongation arrest activity.

### **Structures of *Bacillus* and mammalian SRP**

For structural analysis, we reconstituted a *B. subtilis*. SRP-RNC consisting of purified stalled FtsQ-carrying *Bs*RNCs bound to reconstituted *Bs*SRP and subjected it to cryo-electron microscopy (cryo-EM) and single particle analysis. Cryo-EM data collected using a conventional CCD detector resulted in a reconstruction of  $\sim 7$  Å resolution that clearly revealed the programmed ribosome and additional “L”-shaped density for the SRP reaching from the ribosomal tunnel exit to the translation factor binding site (**Fig. 2a-c**). In an attempt to obtain the highest possible resolution, we also collected data sets on both K2 Summit and Falcon 2 direct electron detectors. As expected these reconstructions revealed details to below 4 Å in the conserved core of the ribosome, but unfortunately, however, the local resolution for SRP did not improve. In fact, the SRP density was only weakly represented in the highest resolution reconstruction (**Extended Data Fig. 4**), indicative of the flexible nature of the ribosome-bound SRP. Although the SRP54 interaction of the S domain confers the high affinity for the RNC, the SRP flexibility was more pronounced for the S domain than for the Alu domain. Further sub-sorting attempts failed, suggesting that there is an apparent lack of distinct conformational sub-states of SRP.

Nevertheless, we could use the 7-12 Å cryo-EM map to build the first complete molecular model of the ribosome-bound *BsSRP*. For the Alu domain a rigid-body docking of the recently reported crystal structure of the *B. subtilis* SRP<sup>8</sup> was possible with only very minor changes necessary; the S domain, including the SRP54 M domain, was first built based on its high degree of similarity to the known S domain of *E. coli* SRP<sup>21,22</sup> and then flexibly docked into the corresponding density (**Fig. 2d, e**). The linker part between the fixed Alu domain and S domain fragments was then built *de novo*<sup>23</sup>. The full 6S RNA model revealed the presence of a major kink (~90°), with a single hinge located between H5e and H5d. The presence of this hinge facilitates the bending of the RNA structure to orient the Alu domain for interaction at the intersubunit space of the ribosome.

To enable a detailed comparison with the eukaryotic SRP, we also improved the molecular model of mammalian SRP (*Canis familiaris*) in the functional context of a ribosome targeting complex<sup>7,24</sup> based on a new cryo-EM reconstruction (**Extended Data Fig. 5a**) and available crystal structures of the SRP68 bound to the S domain of 7S RNA<sup>25</sup> (**Fig. 2f, g**). Despite the fact that the density for the SRP68 remains weakly represented and the density for SRP72 appears to be not visible, indicating a clear flexibility of this region, it appears that the SRP68 binding restructures the distal end of helix H5f as suggested in<sup>25</sup>. While the 6S RNA seems to adopt an “L-shape”, the 7S RNA appears to adopt a more intricate structure mainly due to the presence of two hinges: A major kink between the distal end of H5e and H5d (Hinge 1), which forms a 90° turn similar to the 6S RNA structure, and a second kink of ~30° between the distal end of H5b and H5a (Hinge 2), which is required for the correct positioning of the Alu domain in the translation factor binding site as previously observed<sup>7</sup>.

### **SRP S-domain ribosome interaction**

A structural alignment of eukaryotic and bacterial SRPs with respect to the large ribosomal subunit was then performed (**Fig. 3a**). The overall positioning of the SRPs between tunnel exit and translation factor-binding site appeared very similar, however, there were some noticeable differences in both the Alu and the S domains. While the S domains of *BsSRP* and *E. coli* SRP superpose almost perfectly, the very distal end of the *E. coli* 4.5S RNA appeared to be more flexible than the corresponding region in the *BsSRP* connecting S and Alu domains (**Fig. 3a**). In contrast, the S domain of the mammalian SRP is bound in a position that is displaced by



10-15 Å from the ribosome compared to the bacterial SRP positions. This difference between bacterial and eukaryotic SRPs may be due to the eukaryotic-specific C-terminal extension of the SRP54 M domain that mediates the interaction between SRP and the ribosomal RNA helix 24 (H24) at the tunnel exit. The SRP54 NG domain that can also participate in ribosome interactions at the tunnel exit appeared highly flexible in the *BsSRP* SRP, and, similar to the *E. coli* SRP<sup>26</sup>, could only be seen when filtering the map to 15 Å resolution (**Fig. 3b**). This is different for the mammalian SRP54 in which the NG domain is rigidly bound to the 80S ribosome (**Extended Data Fig. 5b**), as observed previously<sup>7</sup>. At the given resolution of the *BsSRP* map, we can follow the nascent polypeptide chain from the PTC of the ribosome through the ribosomal tunnel into the M domain of SRP54 (**Fig. 2b, Extended Data Fig. 5c**). Here, a rod-like additional density, most likely representing the alpha-helical signal anchor sequence of FtsQ, is accommodated by the M domain in an open conformation (**Fig. 3c**). The overall position of the signal sequence agrees well with previously observed arrangements (**Fig. 3c, d**)<sup>7,24,27</sup>. Notably, however, in this *BsSRP* complex the signal sequence points towards and also interacts with ribosomal rRNA helix 59 (H59) via its N-terminal end (**Fig. 3e, f and Extended Data Fig. 5c**). An interaction of transmembrane (TM) domains of nascent chains with H59 has been observed before<sup>28</sup> and may contribute to the correct orientation of signal sequences or TM domains for transfer to the secYEG translocon or for adopting the correct orientation in the lipid bilayer. Apart from the SRP54-mediated interactions, we observed one additional RNA-RNA interaction between rRNA helix 100 (H100) of the ribosome and the 6S RNA of SRP (**Fig. 3f**). This interaction has also been observed for ribosome-bound *E. coli*<sup>26</sup> and mammalian SRP (<sup>7</sup> and this study) and may be of regulatory importance for NG domain GTPase activation<sup>25,29</sup>.

### **SRP Alu-domain ribosome interaction**

For a detailed understanding of the elongation arrest function of SRP in bacteria and mammalian cells, we analyzed the Alu domains within the ribosomal context (**Fig. 4a, b and Extended Data Fig 6a, b**). Interestingly, the docking of the mammalian Alu domain revealed an interaction mode purely driven by protein-RNA interactions (**Fig. 4a**): the tips of 7S RNA L3 and L4 loops directly contact the N-terminal part of the ribosomal protein uL11, providing the only detectible connection to the large ribosomal subunit (**Fig 4a and Extended Data Fig. 6a**). Contact between the 7S RNA and the small ribosomal subunit is established via the SRP9/14

protein dimer, where mainly the two SRP14  $\alpha$ -helices participate in contacts with h5 and h18 of the 18S rRNA (**Fig. 4a** and **Extended Data Fig. 6a**). Whereas the binding of the Alu domain to the ribosome competes directly with binding of elongation factors such as eEF2, the interaction mode of the SRP Alu domain appears to be different from that of canonical elongation factors: neither the tip of the rRNA helix 43 in the stalk base that was shown to interact with eEF2<sup>30</sup> nor the universally conserved  $\alpha$ -sarcin-ricin loop (SRL, H95) (**Fig. 4a**) are likely to participate in the stabilization of the mammalian Alu domain.

Similar to the mammalian Alu domain also the *Bs*SRP Alu domain was found to bind to the translation factor site and thereby competing with elongation factors (**Extended Data Fig. 7**), however, using an entirely different mode of interaction (**Fig. 4b** and **Extended Data Fig. 6b**): First, binding occurs essentially by RNA-RNA interactions rather than protein-RNA as observed for the mammalian SRP. Second, the *B. subtilis* Alu domain maintains a distance of at least 10 Å from the small ribosomal subunit. Thus, the absence of any SRP9/14-like proteins in the bacterial Alu domain results in the complete lack of interaction with the small ribosomal subunit (**Fig. 4b** and **Extended Data Fig. 6b**). The additional *B. subtilis* Alu RNA helix 1, which was found to point into the intersubunit space, is also too far away as to provide any contacts. However, this apparent lack of contact to the small subunit is compensated by a unique mode of tertiary RNA interaction between the 6S RNA and the large ribosomal subunit. Remarkably, the *B. subtilis* Alu RNA directly binds to the stalk base rRNA helices H43/H44 as well as to the rigidly positioned  $\alpha$ -sarcin-ricin loop (SRL, H95) (**Fig. 4b**). Here, the extended loop-loop pseudoknot t1 of SRP<sup>8</sup> appears to play an essential role in this docking by facilitating a continuous stacking from stalk base to SRL (**Fig. 4c, d**). The conformational changes observed in the stalk base, moving from an “open state” (unbound SRP) to a “half-closed state” (bound SRP) (**Extended Data Fig. 6c**) provide a perfect “landing pad” for the A24:U50 base pair of the 6S loop-loop pseudoknot t1. The A24:U50 base pair stacks effectively onto A1141 and A1113 of ribosomal rRNA helices H44 and H43, respectively (**Fig. 4c**). In this region, extra stability appears to be provided by interaction between uL11 proline 22 and C51 of the Alu RNA. Moreover, A44 from the Alu L4.1 lobe also stacks onto A2689 of the SRL tetraloop, thereby inducing a minor remodeling of the Alu L4.1 lobe. As a consequence, an induced base pairing appears to be possible between A44 and U29 providing in this way extra stability to the structure. Additional stability appears to be also provided by ribose interactions involving Alu A21 and A2689 of H95 as well as Alu G30

and G2690 of H95 (**Fig. 4c, d**). Finally, in this region the C-terminal extension of uL6 appears to stabilize this conformation by directly interacting with Alu C46.

Taken together, the bacterial Alu domain appears to have developed an distinctive way of interacting with the ribosome, relying mainly on RNA:RNA interactions. It exclusively interacts with the large ribosomal subunit essentially by establishing a connected RNA stacking system between the ribosomal stalk base and the  $\alpha$ -sarcin-ricin loop. The 3D structure of the 6S Alu RNA domain forms a unique tetraloop receptor allowing docking onto the SRL. Thus, the entire Alu domain acts as a ‘dock and lock’ system in which the translation factor binding site of the ribosome is locked temporarily in order to slow down the translation machinery for efficient targeting to a membrane translocation site.

## References and Notes:

- 1 Akopian, D., Shen, K., Zhang, X. & Shan, S. O. Signal recognition particle: an essential protein-targeting machine. *Annual review of biochemistry* **82**, 693-721, doi:10.1146/annurev-biochem-072711-164732 (2013).
- 2 Nyathi, Y., Wilkinson, B. M. & Pool, M. R. Co-translational targeting and translocation of proteins to the endoplasmic reticulum. *Bba-Mol Cell Res* **1833**, 2392-2402, doi:DOI 10.1016/j.bbamcr.2013.02.021 (2013).
- 3 Rosenblad, M. A., Larsen, N., Samuelsson, T. & Zwieb, C. Kinship in the SRP RNA family. *Rna Biol* **6**, 508-516 (2009).
- 4 Noriega, T. R., Chen, J., Walter, P. & Puglisi, J. D. Real-time observation of signal recognition particle binding to actively translating ribosomes. *eLife* **3**, doi:10.7554/eLife.04418 (2014).
- 5 Weichenrieder, O., Wild, K., Strub, K. & Cusack, S. Structure and assembly of the Alu domain of the mammalian signal recognition particle. *Nature* **408**, 167-173 (2000).
- 6 Weichenrieder, O. *et al.* Hierarchical assembly of the Alu domain of the mammalian signal recognition particle. *Rna* **7**, 731-740, doi:Doi 10.1017/S1355838201010160 (2001).
- 7 Halic, M. *et al.* Structure of the signal recognition particle interacting with the elongation-arrested ribosome. *Nature* **427**, 808-814, doi:10.1038/nature02342 (2004).
- 8 Kempf, G., Wild, K. & Sinning, I. Structure of the complete bacterial SRP Alu domain. *Nucleic acids research*, doi:10.1093/nar/gku883 (2014).
- 9 Bousset, L. *et al.* Crystal structure of a signal recognition particle Alu domain in the elongation arrest conformation. *Rna*, doi:10.1261/rna.047209.114 (2014).
- 10 Nakamura, K., Yahagi, S., Yamazaki, T. & Yamane, K. Bacillus subtilis histone-like protein, HBsu, is an integral component of a SRP-like particle that can bind the Alu domain of small cytoplasmic RNA. *Journal of Biological Chemistry* **274**, 13569-13576, doi:DOI 10.1074/jbc.274.19.13569 (1999).
- 11 Yamazaki, T., Yahagi, S., Nakamura, K. & Yamane, K. Depletion of Bacillus subtilis histone-like protein, HBsu, causes defective protein translocation and induces upregulation of small cytoplasmic RNA.

- Biochemical and biophysical research communications* **258**, 211-214, doi:DOI 10.1006/bbrc.1999.0615 (1999).
- 12 Chiba, S. & Ito, K. Multisite Ribosomal Stalling: A Unique Mode of Regulatory Nascent Chain Action Revealed for MifM. *Molecular cell* **47**, 863-872, doi:DOI 10.1016/j.molcel.2012.06.034 (2012).
- 13 Doud, S. K., Chou, M. M. & Kendall, D. A. Titration of Protein-Transport Activity by Incremental Changes in Signal Peptide Hydrophobicity. *Biochemistry* **32**, 1251-1256, doi:Doi 10.1021/Bi00056a008 (1993).
- 14 Holtkamp, W. *et al.* Dynamic switch of the signal recognition particle from scanning to targeting. *Nature structural & molecular biology* **19**, 1332+, doi:Doi 10.1038/Nsmb.2421 (2012).
- 15 Struck, J. C. R., Lempicki, R. A., Toschka, H. Y., Erdmann, V. A. & Fournier, M. J. Escherichia-Coli 4.5s Rna Gene-Function Can Be Complemented by Heterologous Bacterial Rna Genes. *Journal of bacteriology* **172**, 1284-1288 (1990).
- 16 Nakamura, K. *et al.* Small Cytoplasmic Rna (Scrna) Gene from Clostridium-Perfringens Can Replace the Gene for the Bacillus-Subtilis Scrna in Both Growth and Sporulation. *Microbiol-Uk* **141**, 2965-2975 (1995).
- 17 Nakamura, K., Imai, Y., Nakamura, A. & Yamane, K. Small Cytoplasmic Rna of Bacillus-Subtilis - Functional-Relationship with Human Signal Recognition Particle 7s Rna and Escherichia-Coli 4.5s Rna. *Journal of bacteriology* **174**, 2185-2192 (1992).
- 18 Nishiguchi, M., Honda, K., Amikura, R., Nakamura, K. & Yamane, K. Structural Requirements of Bacillus-Subtilis Small Cytoplasmic Rna for Cell-Growth, Sporulation, and Extracellular Enzyme-Production. *Journal of bacteriology* **176**, 157-165 (1994).
- 19 Walter, P. & Blobel, G. Translocation of Proteins across the Endoplasmic-Reticulum .3. Signal Recognition Protein (Srp) Causes Signal Sequence-Dependent and Site-Specific Arrest of Chain Elongation That Is Released by Microsomal-Membranes. *Journal of Cell Biology* **91**, 557-561, doi:Doi 10.1083/Jcb.91.2.557 (1981).
- 20 Mason, N., Ciufo, L. F. & Brown, J. D. Elongation arrest is a physiologically important function of signal recognition particle. *Embo Journal* **19**, 4164-4174, doi:DOI 10.1093/emboj/19.15.4164 (2000).
- 21 Ataide, S. F. *et al.* The crystal structure of the signal recognition particle in complex with its receptor. *Science* **331**, 881-886, doi:10.1126/science.1196473 (2011).
- 22 Batey, R. T., Rambo, R. P., Lucast, L., Rha, B. & Doudna, J. A. Crystal structure of the ribonucleoprotein core of the signal recognition particle. *Science* **287**, 1232-1239 (2000).
- 23 Jossinet, F., Ludwig, T. E. & Westhof, E. Assemble: an interactive graphical tool to analyze and build RNA architectures at the 2D and 3D levels. *Bioinformatics* **26**, 2057-2059, doi:10.1093/bioinformatics/btq321 (2010).
- 24 Halic, M. *et al.* Signal recognition particle receptor exposes the ribosomal translocon binding site. *Science* **312**, 745-747, doi:10.1126/science.1124864 (2006).
- 25 Grotwinkel, J. T., Wild, K., Segnitz, B. & Sinning, I. SRP RNA Remodeling by SRP68 Explains Its Role in Protein Translocation. *Science* **344**, 101-104, doi:DOI 10.1126/science.1249094 (2014).
- 26 Halic, M. *et al.* Following the signal sequence from ribosomal tunnel exit to signal recognition particle. *Nature* **444**, 507-511, doi:10.1038/nature05326 (2006).
- 27 Janda, C. Y. *et al.* Recognition of a signal peptide by the signal recognition particle. *Nature* **465**, 507-U139, doi:Doi 10.1038/Nature08870 (2010).
- 28 Bischoff, L., Wickles, S., Berninghausen, O., van der Sluis, E. O. & Beckmann, R. Visualization of a polytopic membrane protein during SecY-mediated membrane insertion. *Nature communications* **5**, 4103, doi:10.1038/ncomms5103 (2014).

- 29 Voigts-Hoffmann, F. *et al.* The Structural Basis of FtsY Recruitment and GTPase Activation by SRP RNA. *Molecular cell* **52**, 643-654, doi:DOI 10.1016/j.molcel.2013.10.005 (2013).
- 30 Anger, A. M. *et al.* Structures of the human and Drosophila 80S ribosome. *Nature* **497**, 80-85, doi:10.1038/nature12104 (2013).

### **Acknowledgments:**

We thank C. Ungewickell and O. Berninghausen for support with cryo-EM, E. van der Sluis and J. Musial for *E. coli* SRP purification, T. Becker for manual data collection. We thank Rishi Matadeen and Sacha DeCarlo for data collection at the NeCEN facility. R.B. is supported by the Deutsche Forschungsgemeinschaft (DFG) through grants SFB646, GRK1721, FOR1805, QBM, the Center for Integrated Protein Science and the European Research Council (Advanced Grant CRYOTRANSLATION). D.N.W. was supported by the Deutsche Forschungs-gemeinschaft through grants FOR1805, WI3285/3-1 and GRK1721. B.B. was supported by an EMBO Long Term Fellowship (ALTF 50-2011). Coordinates and structure factors are deposited in the RCSB protein data bank (PDB) with the code (XXX) (XXX) (XXX) and the EM data in the EMDataBank with the code (XXX).

**Fig. 1. Reconstitution, affinity and elongation arrest/slow-down activity of *Bacillus subtilis* SRP.**

(A) Schematic representation of the bacterial, archaeal and mammalian SRP. (B) Binding assays using *B. subtilis* RNCs with an excess of reconstituted *Bs*SRP. Supernatant (S) and pellet (P) fraction were analyzed by SDS-polyacrylamide gel electrophoresis and Coomassie blue staining. (C) *Bs*SRP binding affinity determination using microscale thermophoresis (MST). Thermophoretic mobility of labeled ribosomes was affected upon binding of SRP (top: raw MST trace, bottom left: measured MST response) allowing calculation of the affinity (bottom right). Bar diagram representing the association rate constants for different *Bs*SRP-RNCs (bottom right). (D) *Bs*SRP induces translation delay/slow-down of reporters containing a signal anchor transmembrane segment *in vitro*. Plot representing time-course translation experiments of full length product reporters encoding FtsQ wt or FtsQ mutant (mutated transmembrane segment) subjected to [<sup>35</sup>S] Met-labeled *in vitro* translation in absence or presence of *Bs*SRP.

**Fig. 2. Cryo-EM structures of SRP-RNC complexes.**

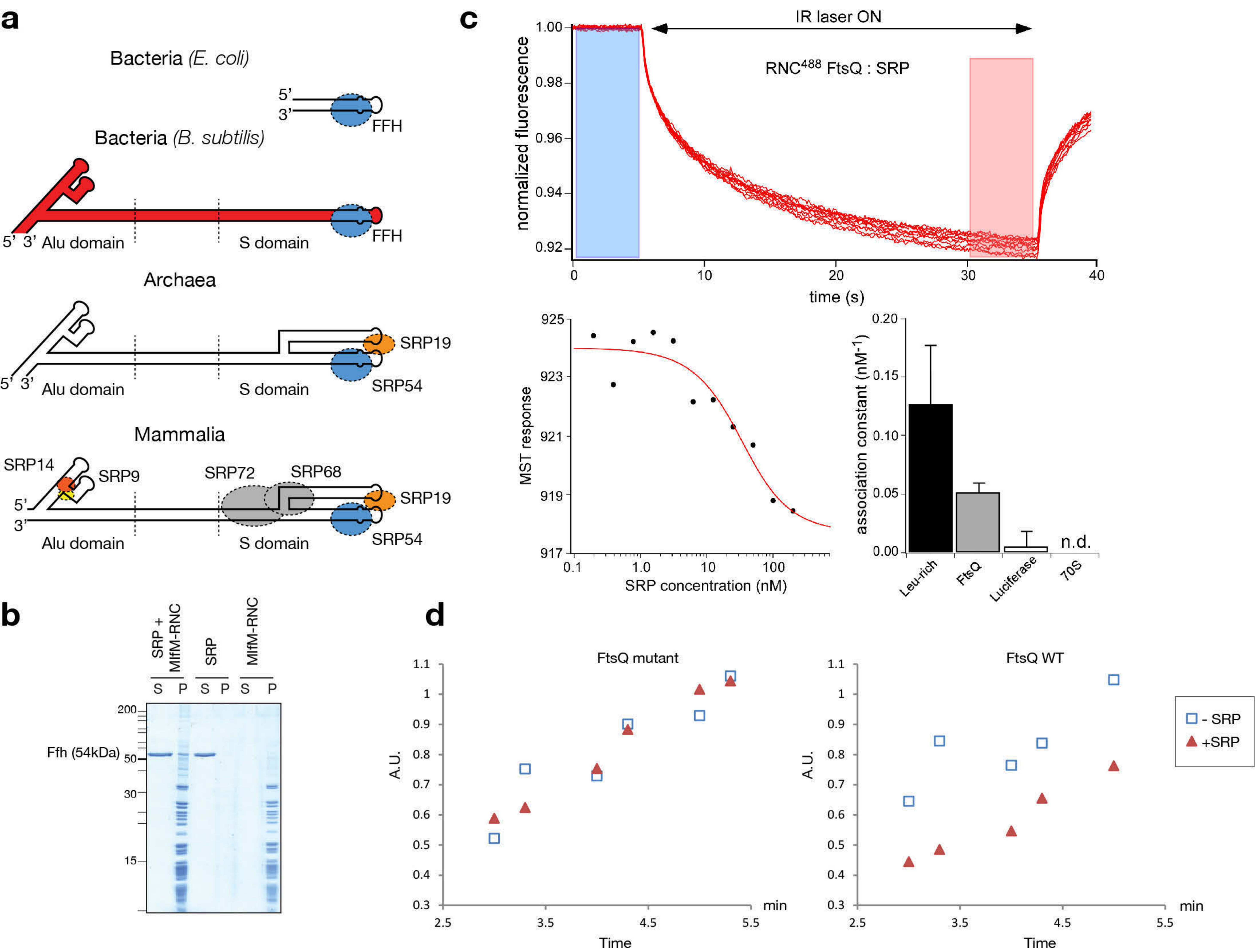
(A) Cryo-EM reconstruction of the *Bs*SRP-RNC. Small 30S subunit (yellow), large 50S subunit (grey), P-site tRNA (green), 6S RNA (red) and the density corresponding to Ffh M-domain (blue). (B) Transverse section of the *Bs*SRP-RNC, visualizing the nascent chain (green) that can be traced from the PTC through the entire ribosomal tunnel to the Ffh M-domain (blue). (C) Local resolution map of the *Bs*SRP-RNC. (D) *Bs*SRP model fitted into the density with corresponding secondary structure diagram. (E) Full *Bs*SRP model, including 6S RNA model (red) and Ffh M-domain (blue). The delocalized Ffh NG-domain is represented by dashed lines. (F). Mammalian SRP model fitted into the density with corresponding secondary structure diagram. (G) Full mammalian SRP Model, including the 7S RNA model (red), SRP9 (orange), SRP14 (purple), SRP54 (Blue), SRP68 (orange), SRP19 (yellow).

### **Fig. 3. Behavior of SRP54 and signal sequence binding**

**(A)** Relative positions of prokaryotic (*E. coli* 4.5A and *B. subtilis* 6S) and eukaryotic SRP (mammalian 7S) with respect to the large ribosomal subunit. **(B)** The Ffh NG domain remains delocalized and can only be visualized by filtering the *BsSRP* SRP-RNC cryo-EM map at 15 Å (red mesh). **(C, D)** Structural alignment of the M domain underlying the position of the nascent chain with respect to H4. All M domains adopt an open conformation where the signal sequence could dock. **(E and F)** Contacts between *BsSRP* and the tunnel exit environment. The Ffh M-domain is represented in blue, the 6S RNA in red, the signal sequence in green. The ribosomal RNA helices that interact with the M-domain (H24), with the signal sequence (H59) and with the 6S RNA (H100) are colored in grey. uL24 contacting the nascent chain is colored in pale yellow.

### **Fig. 4. SRPs Alu domain interaction with the ribosome.**

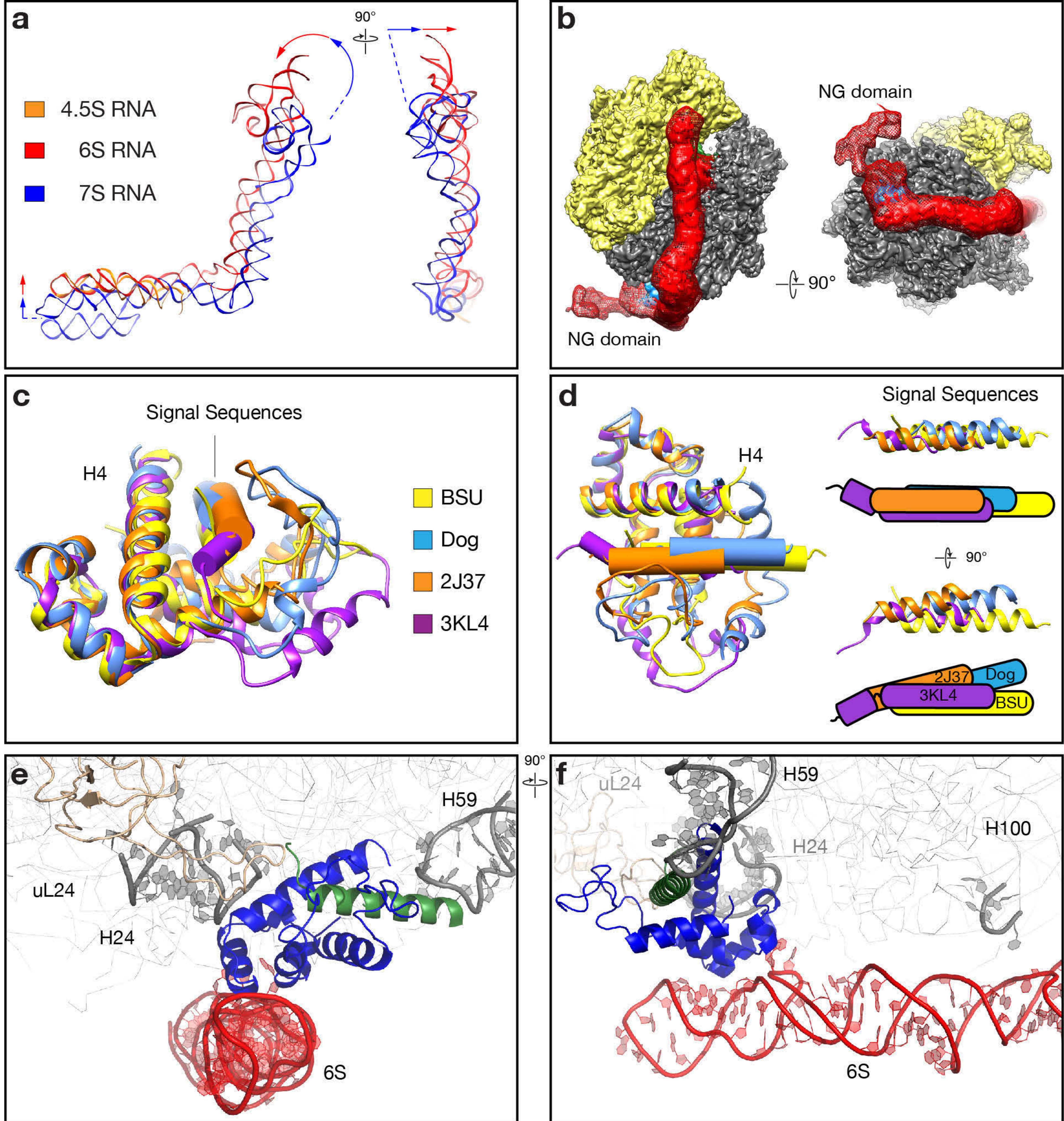
**(A)** Interaction of the mammalian SRP Alu domain with both ribosomal subunits at elongation factor binding site. **(B)** Interaction of the *BsSRP* Alu domain with the large ribosomal subunits. **(C)** Molecular model of *B. subtilis*. 6S Alu domain locking in by generating a continuous stacking between sarcin-ricin loop and helix H43/H44 of 23S rRNA. **(D)** Schematic representation of the molecular basis of the *BsSRP* Alu domain ribosome interaction.



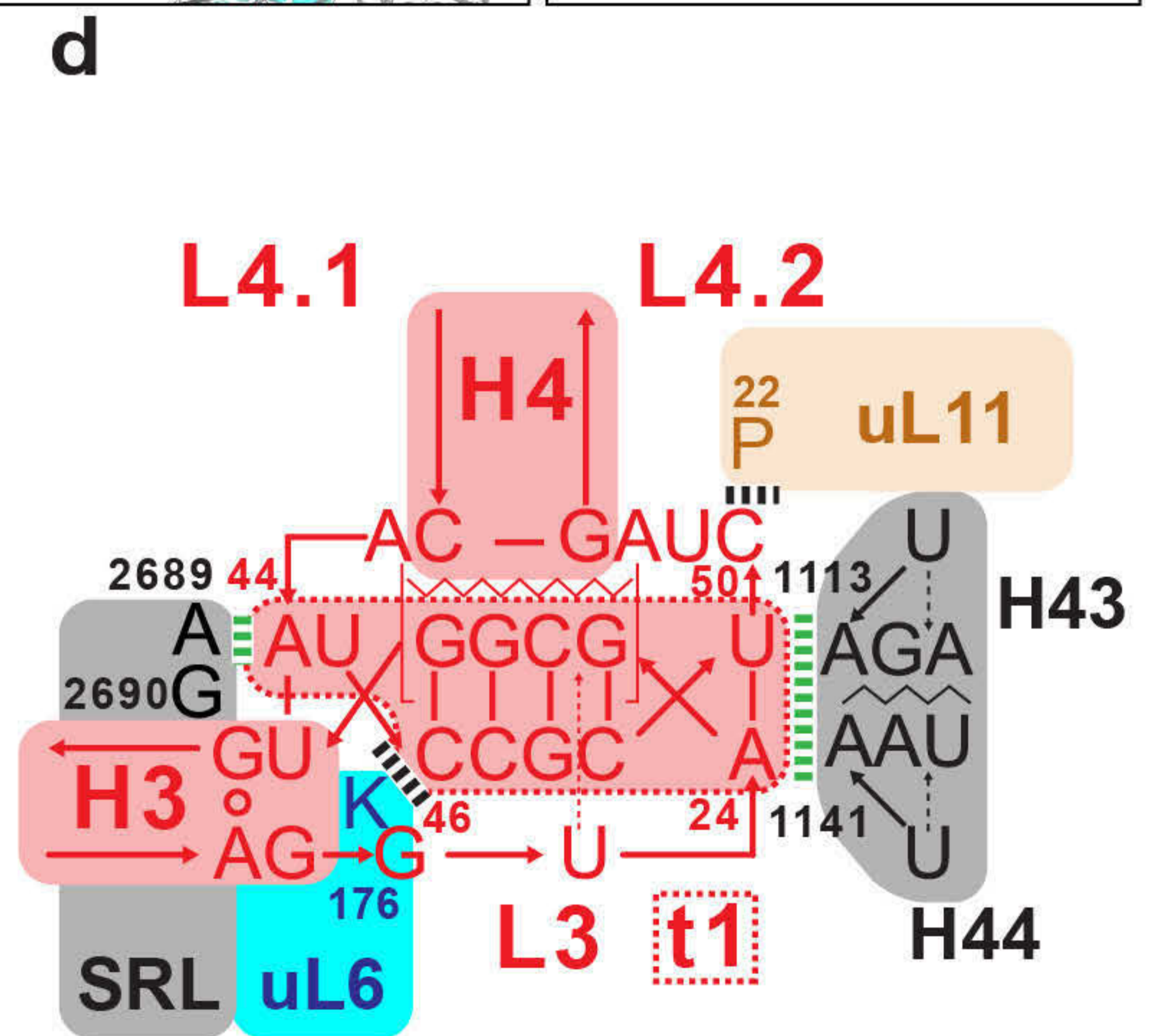
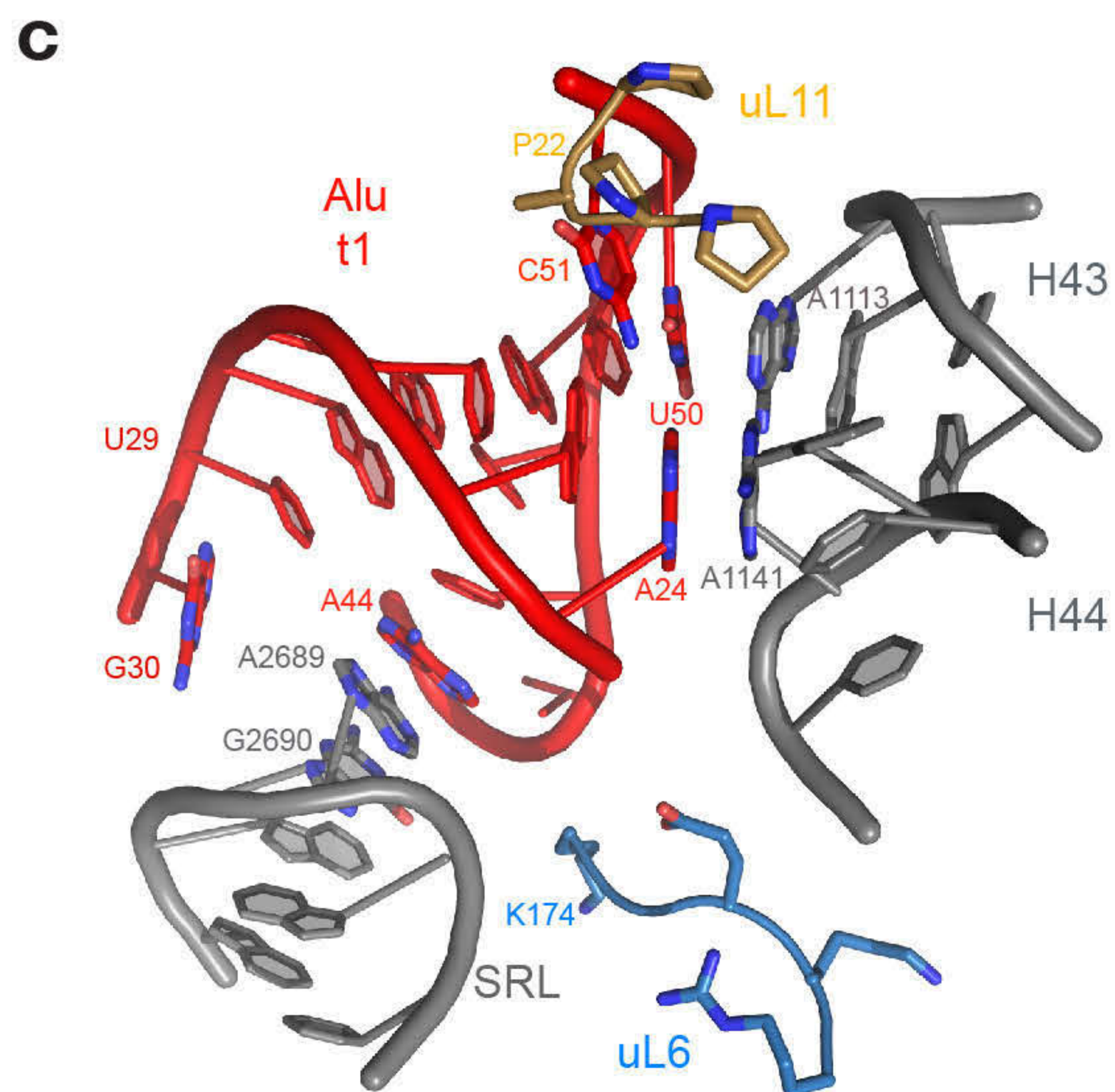
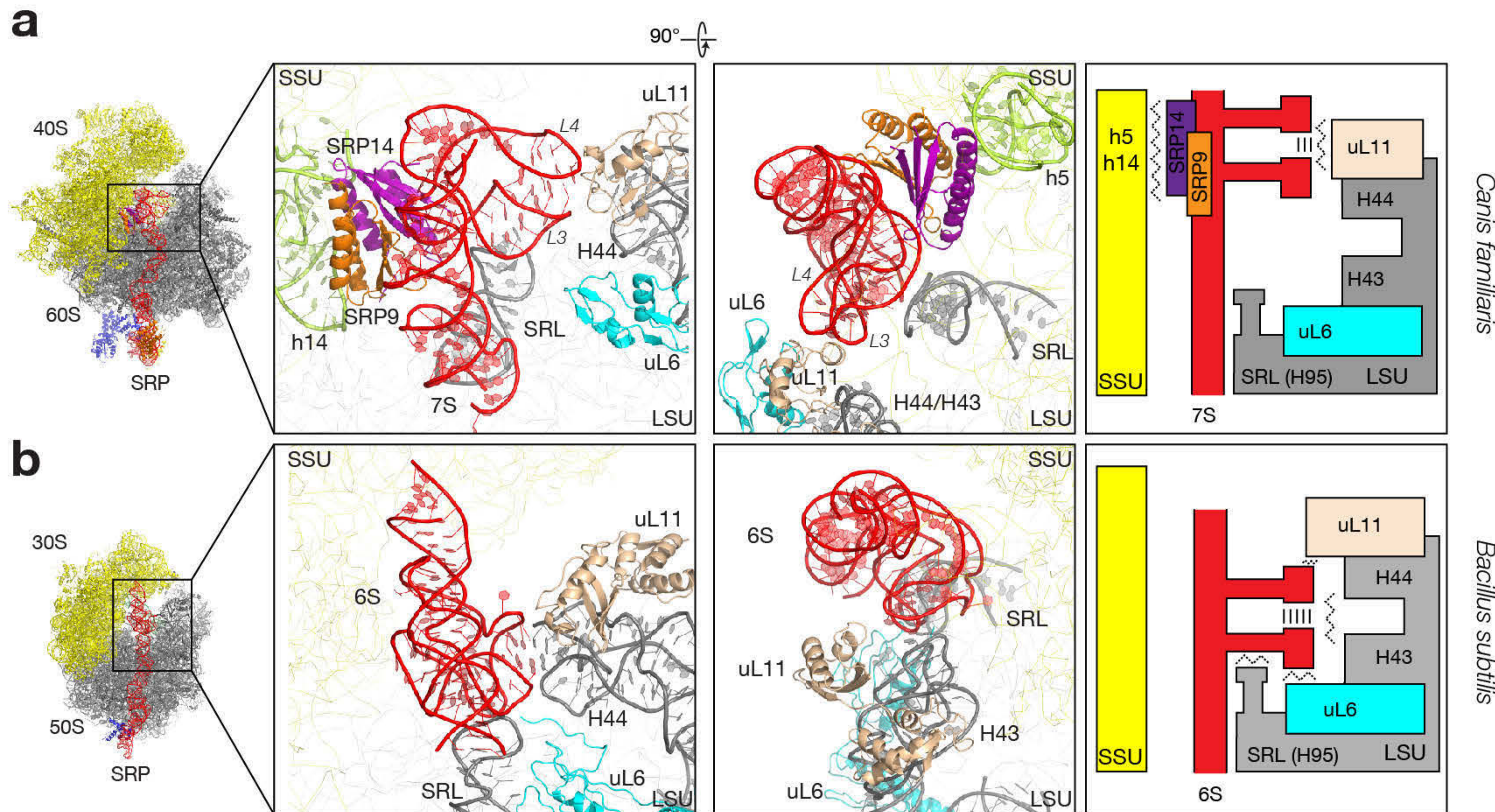
**Figure 1**







**Figure 3**



## Supplementary Materials

### **Structural basis for targeting and elongation arrest of *Bacillus* signal recognition particle**

Bertrand Beckert <sup>1</sup>, Alexej Kedrov <sup>1</sup>, Daniel Sohmen <sup>1</sup>, Georg Kempf <sup>2</sup>, Klemens Wild <sup>2</sup>, Irmgard Sinning <sup>2</sup>, Henning Stahlberg <sup>3</sup>, Daniel N. Wilson <sup>1</sup> and Roland Beckmann <sup>1,\*</sup>

#### **Affiliations:**

<sup>1</sup> Gene Center and Center for integrated Protein Science Munich, Department of Biochemistry, Feodor-Lynen-Str. 25, University of Munich, 81377 Munich, Germany

<sup>2</sup> Heidelberg University Biochemistry Center (BZH), INF 328, D-69120 Heidelberg, Germany

<sup>3</sup> C-CINA, Biozentrum, University Basel, Mattenstrasse 26, CH-4058 Basel, Switzerland

\* Corresponding author: [beckmann@lmb.uni-muenchen.de](mailto:beckmann@lmb.uni-muenchen.de)

### Purification of Ffh, HU1, 6S RNA and SRP reconstitution:

Ffh (ffh,BSU15980) and HU1 (HU1, BSU22790) were amplified from *Bacillus subtilis*, strain 168 genomic DNA using respectively oligonucleotides M18 (ATG CAT TTG AAG GAT TAG CCG ACC GAC TGC AGC AGA CG)/M19 (ATG GCA TTT GAA GGA TTA GCC GAC CGA CTG C), M16 (ATG AAC AAA ACA GAA CTT ATC AAT GCG GTT GCA GAA GC)/M17 (TTA TTT TCC GGC AAC TGC GTC TTT AAG CGC TTT ACC) and then cloned into pET46 vector. His<sub>6</sub>-ffh, His<sub>6</sub>-HU1 and HU1-His<sub>6</sub> protein were overexpress in *E. coli* BL21 at 28°C for 4h after induction with 1 mM IPTG. Cells were harvested and the pellet was suspended in F1B buffer (25 mM HEPES KOH pH 7.5, 500 mM KCl, 10 mM MgCl<sub>2</sub>, 10 mM imidazole, 10% glycerol). Lysis was achieved using a microfluidizer (Microfluidics) by 3 pass at 15.000 psi. The proteins were then purified from the supernatant by his-tag affinity purification using Ni-NTA agarose beads. The bound protein were washed with FwB buffer (25 mM HEPES KOH pH 7.5, 1 M KCl, 10 mM MgCl<sub>2</sub>, 25 mM imidazole, 10% glycerol) and then eluted with FeB buffer (25 mM HEPES KOH pH 7.5, 200 mM KOAc, 10 mM Mg(OAc)<sub>2</sub>, 150 mM imidazole, 10% glycerol).

6S RNA gene (scRNA, X06802) was amplified from *Bacillus subtilis* strain 168 genomic DNA, using oligonucleotides M14 (ATC GGA TCC CTA ATA CGA CTC ACT ATA GGC TTT GCC GTG CTA AGC GGG GAG GTA GCG GTG CCC TGT ACC TGC AAT CCG CTC TAG CAG GGC) /M15 (GC AAG CTT ACC GTG CAC CTT CTG TCG ATT CGC TAC CCT AAG CGT TAC TTA AGC AGT TAG CTC GGC CC) and then cloned into pUC19 -T7. The 6S RNA was produced by in vitro transcription using in-house prepared T7 polymerase and HindIII-linearized plasmid template. The RNA was then purified on 10% urea polyacrylamide gel, eluted from the gel and reconcentrated by ethanol precipitation followed size-exclusion chromatography (S200).

Prior to SRP complex formation, the 6S RNA was first heat-denatured and then refolded by slowly-cooled to 30°C in Cfb buffer (25 mM HEPES KOH pH 7.5, 150 mM KOAc, 10 mM Mg(OAc)<sub>2</sub>, 2% glycerol). SRP complex formation was performed by mixing folded 6S RNA with 10 fold excess of freshly purified proteins with an incubation of 20 min at 30°C. SRP

complex were then purified by size-exclusion chromatography (S200), aliquoted and stored at -80°C.

*Bacillus subtilis* strain 168, S12 translation extract preparation:

*Bacillus subtilis*, strain 168 cells were grown in an “INFORCE HT minifors” bench-top fermenter in 2x YPTG medium (16 g/L peptone, 10 g/L yeast extract, 5 g/l NaCl, 22 mM NaH<sub>2</sub>PO<sub>4</sub>, 40 mM Na<sub>2</sub>HPO<sub>4</sub>, 19.8 g/L glucose (sterile filtered)) feeding extra glucose (feed: 10) at 37°C, maintaining pH 7.0 and oxygen level (60%) to OD<sub>600</sub> 4.5. Cells were harvested at 5.000 g at 4°C for 15min and subsequently washed 3 times with cold Buffer A (10 mM Tris-acetate (pH 8.2), 14 mM Mg(OAc)<sub>2</sub>, 60 mM potassium glutamate, 1 mM dithiothreitol and 6 mM 2-mercaptoethanol). Cells were flash-frozen in liquid nitrogen and stored at -80°. About 15g of cells were thawed on ice and resuspended in 10ml of cold buffer B (buffer A without 2-mercaptoethanol) and broken open using a microfluidizer (Microfluidics M-110L) by 3 pass at 15.000 psi. The lysate was cleared subsequently via centrifugation at 12.000 g and 4°C for 10 min and incubated for 30 min at 37°C in a water bath. The cell extract was aliquoted, flash-frozen and stored at -80°C.

*In vitro* translation reaction and purification of unlabeled and 488 labeled MifM-stalled RNC:

1 ml of reaction mix containing 240 mM HEPES-KOH (pH 8.2), 60 mM glucose, 1.2 mM ATP, 0.85 mM CTP, GTP, and UTP, 2 mM DTT, 0.17 mg/mL *E. coli* total tRNA mixture (from strain MRE600 Sigma), 90 mM potassium glutamate, 80 mM of ammonium acetate, 8 mM magnesium acetate, 20 mM potassium phosphate dibasic (pH 7.2), 34 µg/mL L-5-formyl-5, 6, 7, 8tetrahydrofolic acid (folinic acid), 2.1 mM (each) amino acids mix, 2 mM Cysteine, 2 % (w/v) PEG 8000 was preheated to 30°C before the addition of 27 % (v/v) cell-extract (S12 extract). The whole reaction was incubated at 30°C for 2 min before the addition of MifM mRNA and further incubation at 30°C for 33 min.

The ribosomes were first isolated from the *in vitro* reaction mix by centrifugation through sucrose cushion (50 mM HEPES KOH pH 7.2, 250 mM KOAc, 25 mM Mg(OAc)<sub>2</sub>, 750 mM Sucrose, 0.1% DDM) for 180 min at 72000 g using TLA120.2 rotor. The pellet was then suspended in 250 buffer (50 mM HEPES KOH pH 7.2, 250 mM KOAc, 25 mM Mg(OAc)<sub>2</sub>, 500

mM Sucrose, 0.1% DDM) and the MifM-RNCs were isolated using TALON metal-chelate affinity resin (Clontech). The bound MifM-RNCs were washed with 500 buffer (50 mM HEPES KOH pH 7.2, 500 mM KOAc, 25 mM Mg(OAc)<sub>2</sub>, 500 mM Sucrose, 0.1% DDM) and eluted using 250i buffer (50 mM HEPES KOH pH 7.2, 250 mM KOAc, 25 mM Mg(OAc)<sub>2</sub>, 500 mM Sucrose, 250 mM imidazole, 0.1% DDM). The eluted MifM-stalled RNC were then loaded onto a linear sucrose gradient (5%-40% (w/v) sucrose in 250 buffer) for 18h at 43000 g in SW28 and the isolated RNC peak was then pelleted by centrifugation 3h at 139000 g Ti70.1 and the pellet was resuspended in RNC buffer (20 mM HEPES pH 7.2, 50mM KOAc, 6mM Mg(OAc)<sub>2</sub>, 250 mM sucrose, 0.05% DDM).

#### Binding assays of *Bs*RNC-SRP complexes and translation slowdown experiments:

Prior to the reconstitution of SRP-MifM-stalled RNC complexes, SRP was incubated in SRP buffer (25 mM HEPES KOH pH 7.5, 150 mM KOAc, 10 mM Mg(OAc)<sub>2</sub>, 1 mM DTT, 0.005 % Nikkol) at 32°C for 10 min. Then SRP-MifM-stalled RNC complexes were reconstituted by incubating 2 pmol RNCs with 10 pmol pre-incubated SRP for 20 min at 32°C. Binding was then tested by centrifugation through a sucrose cushion followed by SDS-PAGE analysis.

Independent in vitro translation triplicate experiments were performed using PURExpress system (NEB), according to the manufacturer's instruction. Briefly, reporter wt FtsQ or mutant FtsQ were cloned and translation was coupled to transcription using PCR templates containing a T7 promoter and supplemented with [35 S]-methionine (Perkin Elmer) plus/minus *B.s* SRP. Reactions were then incubated at 37°C, and 2.5 µl aliquots were removed at specific time points and stopped by placing them on ice. The translation reactions were then loaded on 18% SDS-PAGE and the products were detected by autoradiography using Amersham Hyperfilm MP film. The bands were quantified using ImageJ.

#### Microscale thermophoresis of ribosome:SRP complexes

Binding affinities of SRP to translating and empty ribosomes were measured using microscale thermophoresis (MST). MifM-stalled RNCs with exposed FtsQ, Leu rich, or luciferase nascent chain were prepared as described above. For the labelling of MifM-stalled

RNC and empty 70S ribosomes, after the optional *in vitro* translation reaction and the sucrose cushion step, the ribosomal pellet was resuspended in 250 buffer supplemented with 100  $\mu$ M AlexaFluor 488 C5-maleimide (Invitrogen/Life technologies). Ribosomes were incubated with the fluorophore for 30 min and further purified following the described protocol. Prior MST experiments ribosomes were diluted to 20 nM into 20 mM HEPES-KOH pH 7.2, 100 mM KOAc, 10 mM Mg(OAc)<sub>2</sub>, 2% glycerol, and 0.005% Nikkol. SRP composed of 6S RNA and the Ffh protein was titrated within the concentration range of 0.2 nM to 250 nM. Ribosome:SRP binding reactions were loaded into Premium coated capillaries (NanoTemper Technologies) and the thermophoretic response of fluorescently labeled ribosomes was measured using the Monolith .115 instrument (NanoTemper Technologies) at 40% and 60% infra-red (IR) laser power. Upon local, micrometer-sized heating with the IR laser a change in the fluorescence due to influx or efflux of labeled ribosomes was measured and then plotted versus the corresponding SRP concentration to form a binding trace. Binding constants were calculated as averages from independent duplicates or triplicates.

*E. coli* RNCs translating FtsQ was prepared *in vivo* and purified as previously described<sup>27,30</sup>. RNCs were fluorescently labeled with AlexaFluor 488-c5-maleimide during the metal chelating chromatography purification and used for MST measurements as described above. *E. coli* SRP was purified as previously described<sup>23</sup>.

#### Preparation the eukaryotic RNC-SRP complex:

Dog pancreas SRP, wheat germ DP90 RNCs and the complex reconstitution were prepared as described before<sup>6</sup>.

#### Cryo-EM specimen preparation and data collection:

For grid preparation 2.5  $\mu$ l *in vitro* reconstituted *B. subtilis* signal recognition particle (20 pmol/ $\mu$ l in SRP buffer: 25 mM HEPES, 100 mM KOAc, 10 mM MgOAc, 1 mM DTT, 2% Glycerol, pH 7.2 at 4°C) were diluted to a final volume of 44.2  $\mu$ l with buffer E (50 mM HEPES, 250 mM KOAc, 2 mM 2-mercaptoethanol, 0.06% Nikkol, pH 7.2 at 4°C) and activated by incubation at 30°C for 10 min. Subsequently, 5.8  $\mu$ l (9.6 pmol) of RNC was added and incubated



for 10 min at 30°C. Finally the prepared mix was diluted with 22.5 µl buffer F (50 mM HEPES, 250 mM KOAc, 2 mM 2-mercaptoethanol, pH 7.2 at 4°C) for optimal grid coverage. The prepared mix was then applied to 2 nm pre-coated Quantifoil R3/3 holey carbon supported grids and vitrified using a Vitrobot Mark IV (FEI Company).

For the 8k data collection, the micrographs were collected under low-dose conditions on a FEI TITAN KRIOS operating at 300 kV using a 8k x 8k TemCam-F416 CMOS camera and a final pixel size of 0.97 Å for the *B. s.* SRP and 1.035 Å for the WG DP90-SRP on the object scale. 20 910 for *B.s.* SRP, 3 627 for WG DP90-SRP were selected for single-particle analysis. Automated particle selection was performed using the program Signature<sup>31</sup>, resulting in 333 617 for *B.s.* SRP and 83 659 for WG DP90-SRP. Three dimensional reconstructions were then performed using the SPIDER software package<sup>32</sup>. Datasets were first cleaned from non-ribosomal particles and then sorted for the presence of a stoichiometric, homogenous P-site tRNA. The homogenous P-site tRNA particles were then sorted for the presence/absence of SRP bound close to the ribosomal exit tunnel. A final sub-dataset sorting contain 75 900 particles for *B.s.* SRP and 19 096 particles for WG DP90-SRP was refined to a final average resolution at FSC 0.5 of respectively 7.1 Å and 9 Å. To exclude potential over-fitting the data was processed using a frequency limited refinement protocol by truncating high frequencies (low-pass filter at 12 Å) during the whole refinement process<sup>33</sup>.

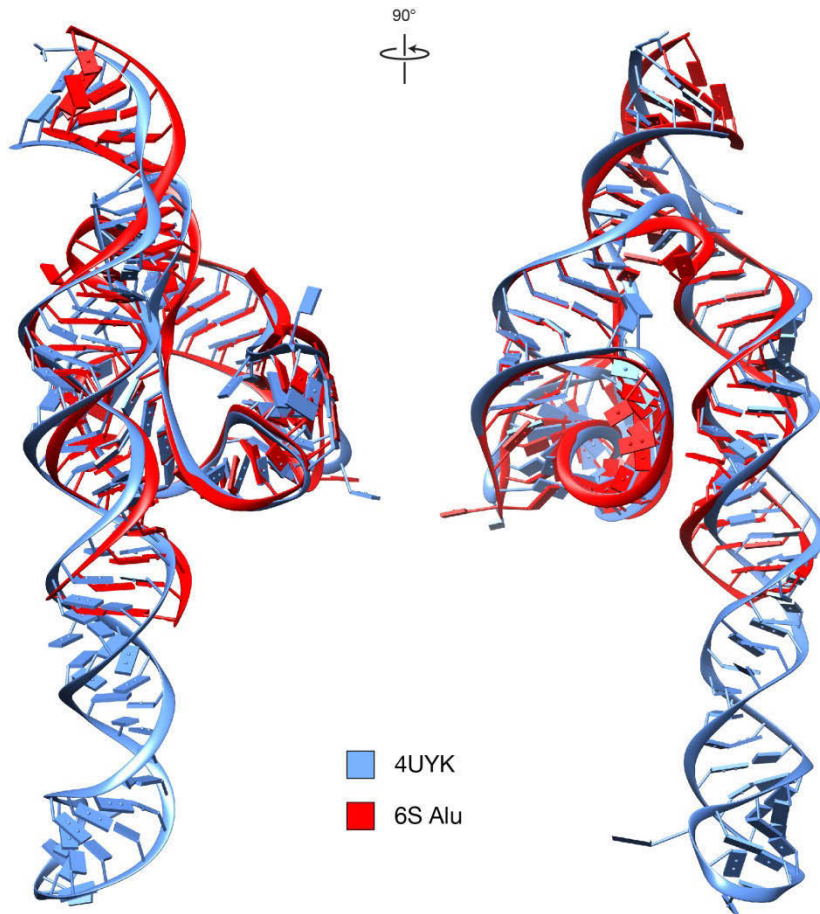
The first direct detector data collection was performed on a Titan Krios at 300 keV and magnification of 39 400 x at the plane of the K2 Summit direct detector camera (Gatan) resulting in an image pixel size of 1.27 Å per pixel on the object scale. Images were recorded using resolution counting mode following an establish protocol<sup>34</sup>. The dose rate on the camera was set was set to be 5e<sup>-</sup> counts per physical pixel per second. The total exposure time was 5 s, leading to a total accumulated dose of 20 electrons per Å<sup>2</sup> on the specimen. Each image was fractionated into 25 subframes, each with an accumulation time of 0.2 s per frame. All dose-fractionated cryo-EM images were recorded using a semi-automated acquisition program<sup>35</sup>. Images were recorded with a defocus in a range from 0.8 to 3.0 µm. 12000 micrographs were selected for single-particle analysis and the same procedure as previously describe above was used for three

dimensional reconstruction. 32 045 Particles were used for three dimensional reconstruction and the final volume after sorting contains 10 434 particles and reaches the resolution of 7.4 Å at 0.5 FSC.

The second direct detector data collection was performed at collected at NeCEN (Leiden, Netherlands) on a Titan Krios TEM (FEI Company) operated at 300 keV equipped with a Cs-corrector and a back-thinned FEI Falcon II direct electron detector. The camera was calibrated for a nominal magnification of 125,085x resulting and a pixelsize of 1.10 Å at the specimen. 8 s-1 frames were recorded in automatic mode with a dose of 4 e-/Å per frame at defocus values between 0.8 µm and 2.2 µm. The first and the last image were excluded from the dataset. The remaining frames were aligned using the .2389 micrographs were selected for single-particle analysis and the same procedure as previously describe above was used for three dimensional reconstruction. 236 436 Particles were used for three dimensional reconstruction and the final volume after sorting contains 78 812 particles and reaches the resolution of 5.2 Å at 0.5 FSC.

## References:

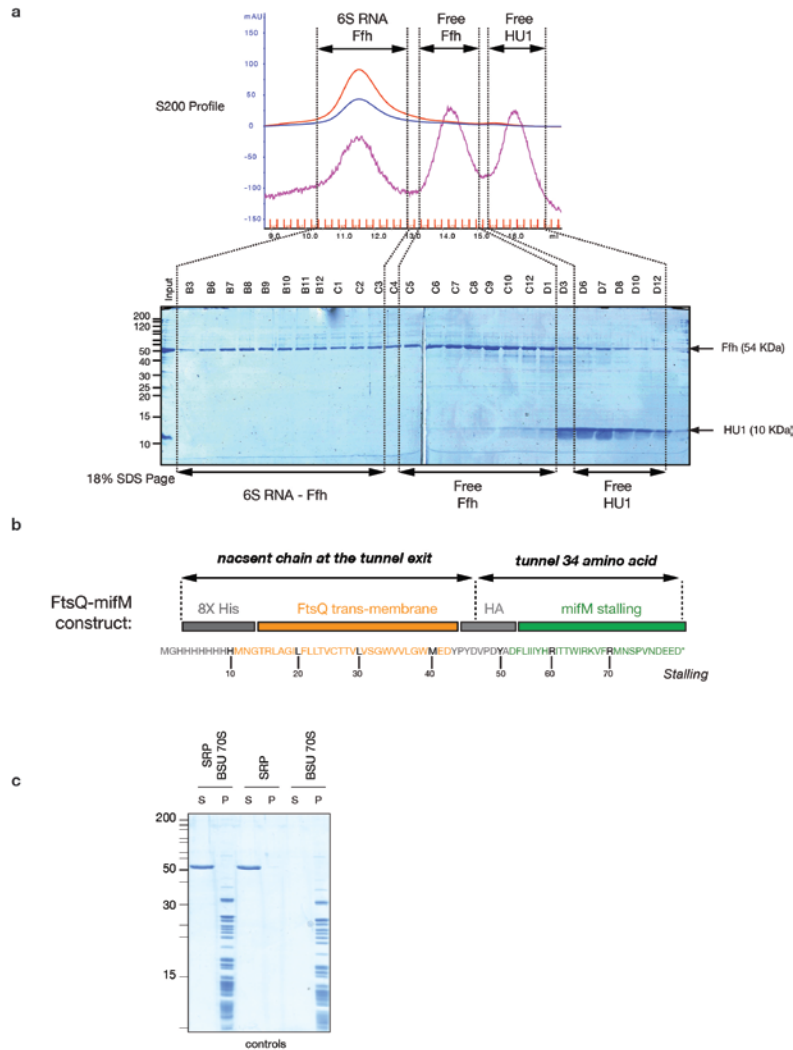
- 30 Wickles, S. *et al.* A structural model of the active ribosome-bound membrane protein insertase YidC. *eLife* **3**, e03035, doi:10.7554/eLife.03035 (2014).
- 31 Chen, J. Z. & Grigorieff, N. SIGNATURE: a single-particle selection system for molecular electron microscopy. *Journal of structural biology* **157**, 168-173, doi:10.1016/j.jsb.2006.06.001 (2007).
- 32 Frank, J. *et al.* SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *Journal of structural biology* **116**, 190-199, doi:10.1006/jsbi.1996.0030 (1996).
- 33 Scheres, S. H. & Chen, S. Prevention of overfitting in cryo-EM structure determination. *Nature methods* **9**, 853-854, doi:10.1038/nmeth.2115 (2012).
- 34 Li, X. *et al.* Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nature methods* **10**, 584-590, doi:10.1038/nmeth.2472 (2013).
- 35 Scherer, S. *et al.* 2dx\_automator: implementation of a semiautomatic high-throughput high-resolution cryo-electron crystallography pipeline. *Journal of structural biology* **186**, 302-307, doi:10.1016/j.jsb.2014.03.016 (2014).



Extended Data Figure 1

### Extended Data Figure 1: Structural alignment of RNA Alu domains

Comparison of the *B. subtilis* RNA Alu domain with the archaea *Pyrococcus horikoshii*. A structural superposition of the archaea *Pyrococcus horikoshii* Alu domain model (blue, PDB:4UYK), onto the *B. subtilis* RNA Alu domain (red) highlights the highly conserved tertiary fold of the Alu RNA core with an RMSD of 1.25 Å.



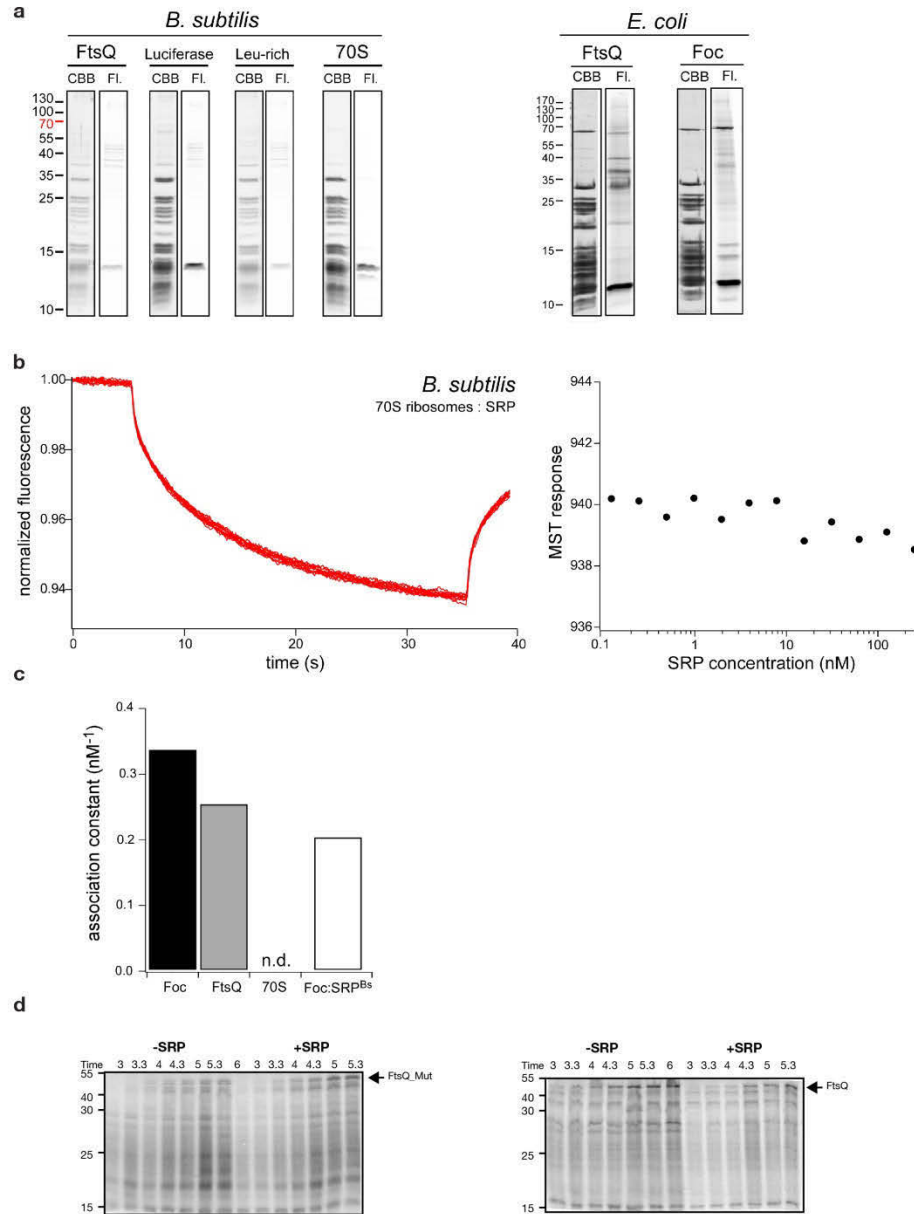
Extended Data Figure 2

**Extended Data Figure 2: *In vitro* *B. subtilis* SRP reconstitution, stalled *BsRNC* construct and binding assays.**

**(A)** Size exclusion chromatography diagram of the *in vitro* reconstitution of *BsSRP* and SDS-PAGE analysis of the fractions. The different components of *BsSRP* were either loaded separately or mixed together and subjected to size-exclusion chromatography. From the diagram profile and the SDS-PAGE analysis, Ffh and the 6S RNA co-elute while HU1 remains unbound.

**(B)** Construct used for the preparation of *B. subtilis* stalled RNC carrying FtsQ signal sequence.

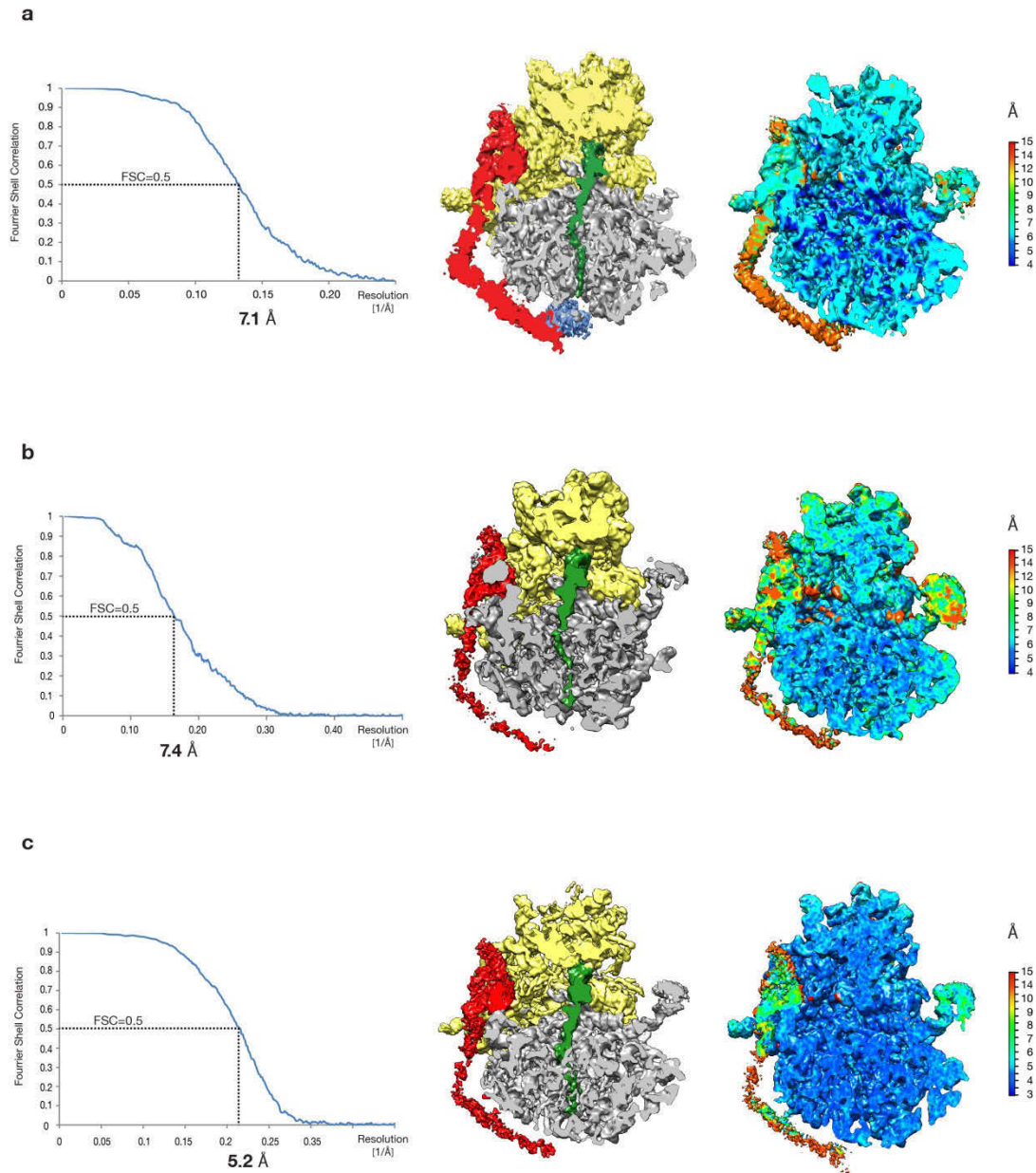
**(C)** Binding assays using purified 70S with an excess of reconstituted *BsSRP*. Supernatant (S) and pellet (P) fraction were analyzed by SDS-polyacrylamide gel electrophoresis and Coomassie blue staining.



Extended Data Figure 3

### Extended Data Figure 3: Microscale thermophoresis and translation assays

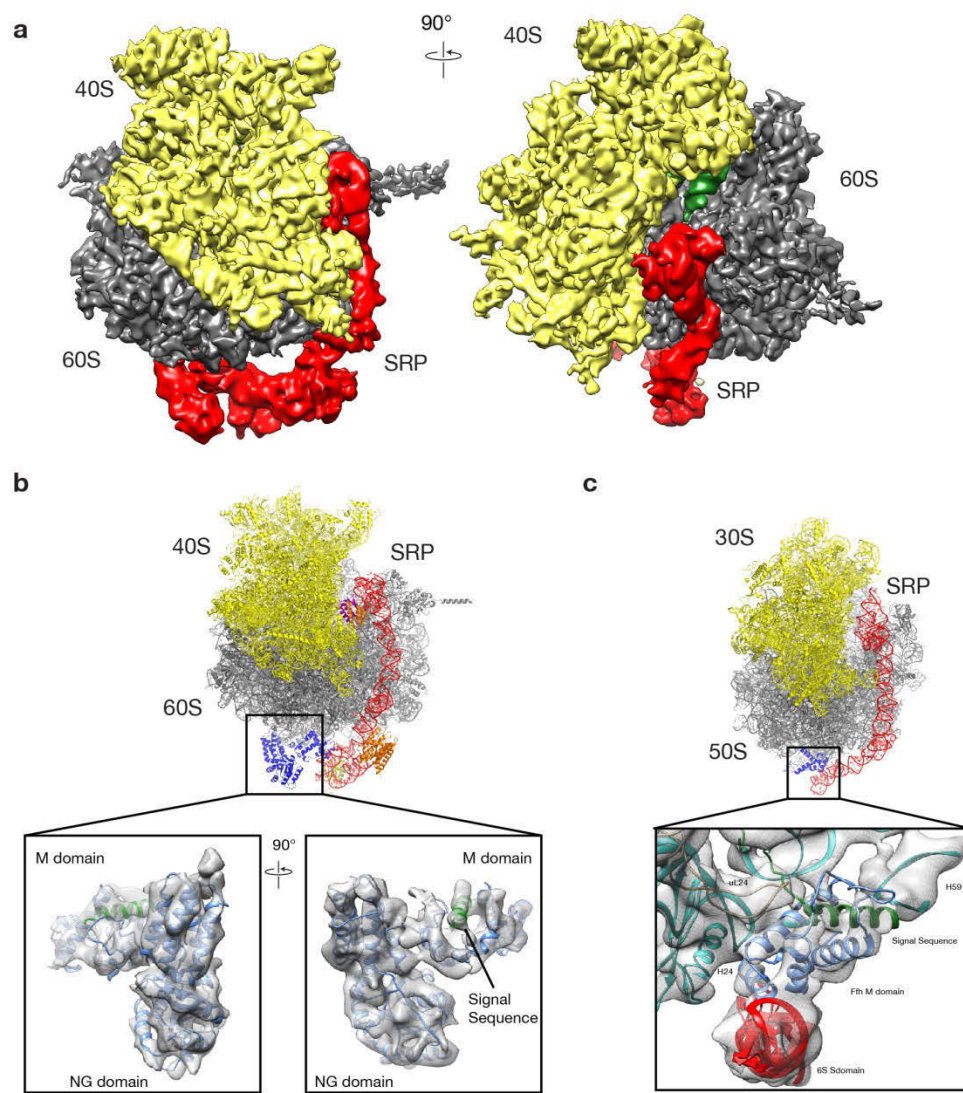
(A) Purified and labeled (AlexaFluor 488 C5-maleimide) RNCs after SDS-PAGE analysis. (B) Typical MST recording using puromycin-treated empty 70S ribosomes with SRP. (C) Bar diagram representing the association rate constants for different control experiments using *E. coli* SRP-RNC. (D) Time-course translation reaction. The formation of full length product reporters encoding FtsQ wt or FtsQ mutant were monitored by [<sup>35</sup>S] Met-labeled *in vitro* translation in absence or presence of *Bs*SRP.



**Extended Data Figure 4**

**Extended Data Figure 4: Cryo-EM reconstruction and resolution of *Bs*SRP-RNC complexes**

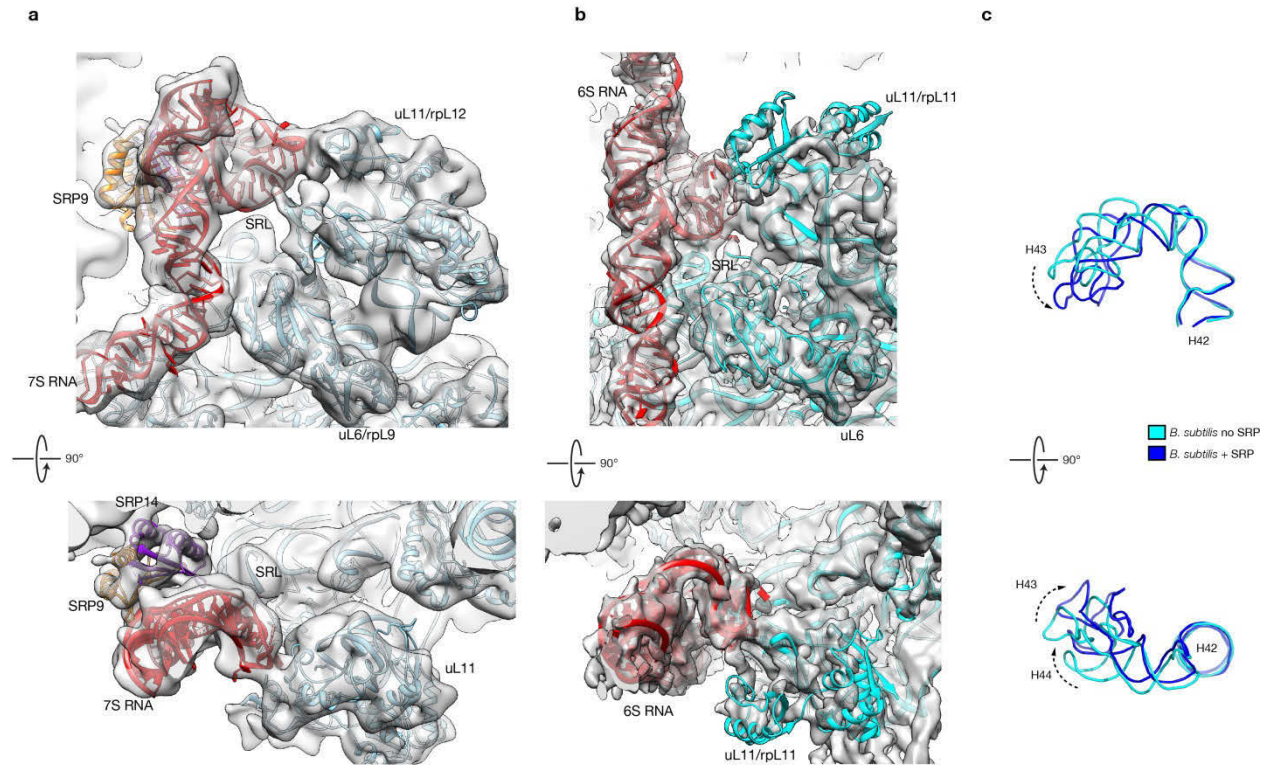
(A) Resolution curve, cryo-EM reconstruction and local resolution of *Bs*RNC-SRP complex using the 8 K CCD detector data. (B) Same as in (A) using K2 direct detector data. (C) same as in (A) using Falcon 2 direct detector data.



**Extended Data Figure 5**

**Extended Data Figure 5: Behavior of SRP54 and signal sequence binding.**

(A). Cryo EM of the mammalian SRP. Small 30S subunit (yellow), large 50S subunit (grey), P-site tRNA (green), SRP (red). (B) Isolated density of *C. familiaris* SRP54 at the ribosome exit tunnel. In contrast to the *B. subtilis* Ffh NG domain, the SRP54 NG domain adopts a more rigid conformation. The signal sequence could be fitted unambiguously into the mammalian SRP54 M domain. (C) Models of *Bs*SRP54 M domain and the signal sequence could be fitted into the density filtered at 8 Å.

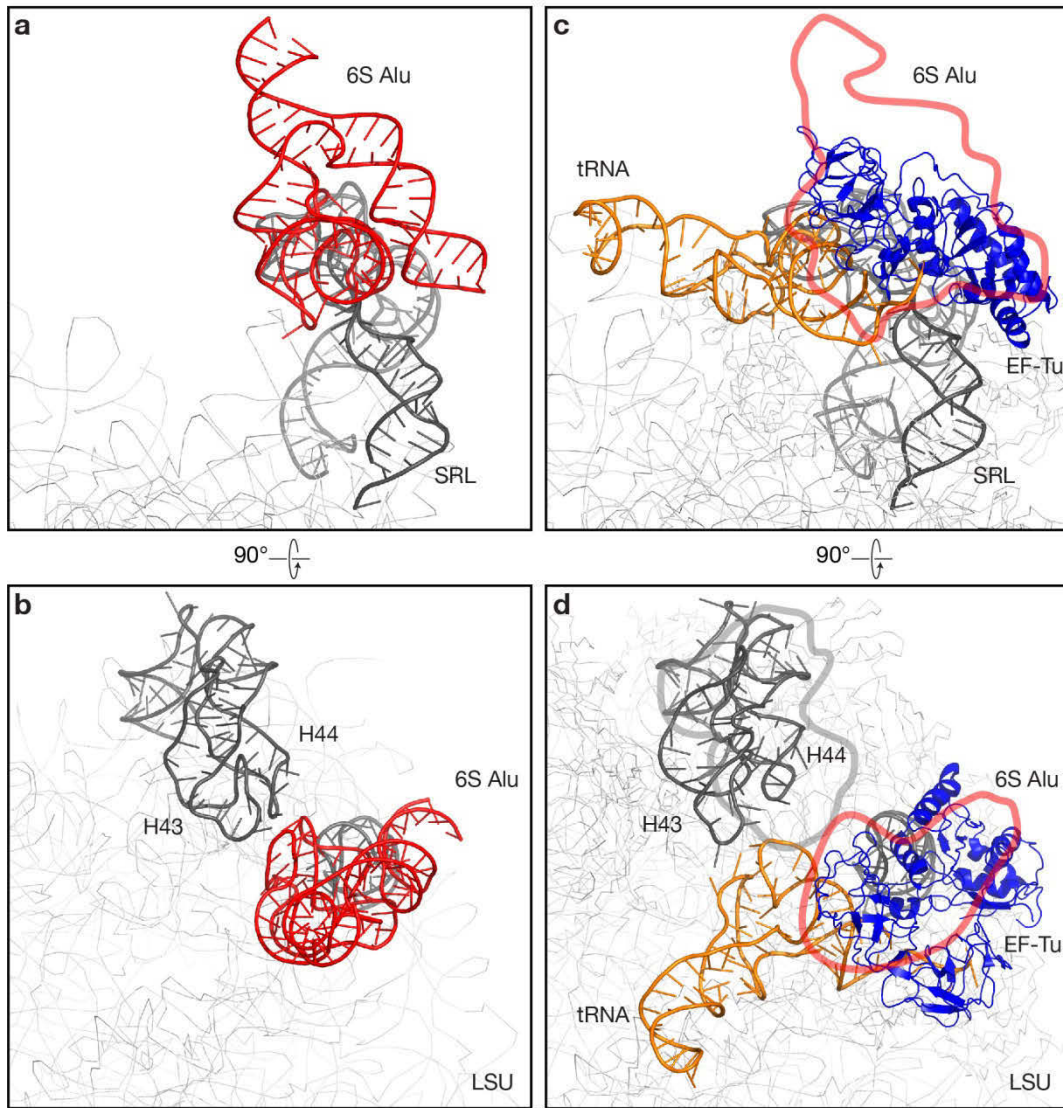


Extended Data Figure 6

### Extended Data Figure 6: Focusing on the Alu domain.

(A) Alu domain mode of interaction in the mammalian case. Density with the fitted model of the mammalian SRP model highlights the interaction mode of the Alu domain with the 80S ribosome by bridging the small and the large subunit at the A-site entry. SRP9/SRP14 interact with h5 and h15 of the 18S RNA. The 7S Alu domain interacts only with uL11. (B) Alu domain mode of interaction in the bacterial case. SRP density with the 6S model in an orientation similar to (A). In contrast to the mammalian SRP, *BsSRP* contacts only the LSU. Those interactions are mainly mediated by RNA:RNA interactions and a docking onto the sarcin-ricin loop. (C) Movement of H43-H44 upon binding of *BsSRP*. Binding of the SPR induces a structural change of H43-H44 from an “open” (cyan) to a “close” state (dark blue).





**Extended Data Figure 7**

**Extended Data Figure 7: Blocking of the translation factor binding site.**

(A and B) 6S Alu domain complex at the ribosome translation factor binding site. (C and D) EF-Tu-tRNA-GTP ternary complex at the ribosome translation factor binding site (pdb:2WRO). *BsSRP* Alu domain binds to the translation factor site (red contour) and thereby competes with elongation factors binding. Binding of the SPR induces a structural change of the stalk base (H43/H44 grey contour)