Biochemical and structural studies of the interactions between the bacterial actin homolog MreB and the Mur enzymes which are involved in the synthesis of the bacterial cell wall

Sandy Favini-Stabile

UNIVERSITÉ DE GRENOBLE

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Spécialité : **Biologie Structurale**

Arrêté ministériel : 7 août 2006

Présentée par

Sandy FAVINI-STABILE

Thèse dirigée par **Andréa DESSEN**

préparée au sein du **Laboratoire Pathogénie Bactérienne**
dans **l'École Doctorale Chimie Sciences du Vivant**

Études biochimiques et structurales des interactions entre la protéine MreB, homologue bactérien de l'actine, et les enzymes Murs impliquées dans le mécanisme de formation de la paroi des bactéries.

Thèse soutenue publiquement le **18 septembre 2013**,
devant le jury composé de :

Mme Paulette, CHARLIER
Directrice de recherche, Université de Liège, Rapporteur
Mr David ROPER
Associate Professor, University of Warwick, Rapporteur
Mr Jean-Pierre SIMORRE
Directeur de recherche, IBS, Grenoble, Président
Mme Andréa DESSEN
Directrice de recherche, IBS, Grenoble, Directrice de thèse

Abstract

Resistance to antibiotics is increasingly frequent, and therapy for patients infected by multi-resistant strains is more and more complicated and delicate, and sometimes even inefficient. Therefore, there is an urgent need for novel antibiotics to stave off the resurgent threat of bacterial epidemics.

The relatively well-known mechanism of bacterial cell wall formation remains a pathway of prime interest in the search for therapeutic targets. Strikingly, recent studies have suggested that the biosynthesis of its main component, peptidoglycan, would involve macromolecular protein-protein complexes. Particularly, Mur ligases were suggested to form a multipartite complex which would recruit the glycosyltransferase MurG and the bacterial actin homolog MreB as well. Interestingly, these enzymes are targetted by none of the antibiotics in clinical use.

The work carried out during this PhD on the *Thermotoga maritima* model showed by surface plasmon resonance and dot blot techniques that MurD, MurE, and MurF all recognize MurG and MreB, but not each other, whilst the two latter proteins interact. A crystallization screen allowed the determination of the crystallization fingerprints of single proteins and potential complexes, aiming for the crystal structure of one of the Mur complexes. Thanks to this screening, the structures of MurD, MurE, MurF were solved, suggesting that the conformational flexibility of their C-terminal domains could be involved in the formation and stability of complexes. In addition, one crystallization condition that could have isolated the MreB-MurF complex remains to be further investigated.

These results mark a further step in the characterization of the cytoplasmic peptidoglycan machinery, opening up towards novel therapeutic targets which would impair the integrity of the macromolecular complex.

Keywords: bacterial cell wall, peptidoglycan biosynthesis, Mur ligases, MurG, MreB, protein-protein interactions, surface plasmon resonance, X-ray crystallography, crystallization of protein-protein complexes, *Thermotoga maritima*, antibiotic targets.

Résumé

Les résistances aux antibiotiques sont de plus en plus fréquentes et la thérapie des patients infectés par des souches multi-résistantes devient très complexe et délicate, voire dans certains cas inefficace. Il devient donc urgent de trouver des antibiotiques innovants et ainsi repousser la menace renaissante d'épidémie.

Pour ce faire, la biosynthèse du peptidoglycane – l'un des composants majeurs de la paroi des bactéries, est une cible qui a amplement fait ses preuves dans la lutte contre les infections bactériennes et reste d'intérêt thérapeutique. Des études récentes ont en effet suggéré que ce processus impliquerait des complexes macromoléculaires dont l'intégrité pourrait être perturbée par de nouveaux antibiotiques. En particulier, il a été suggéré que les ligases Mur – qui participent à la synthèse de l'unité monomérique du peptidoglycane dans le cytoplasme, feraient partie d'un complexe multipartite recrutant probablement aussi la glycosyltransferase MurG et la protéine du cytosquelette MreB. En outre, ces enzymes ne sont à l'heure actuelle la cible d'aucun antibiotique médical, malgré leur intérêt thérapeutique largement reconnu.

Les travaux réalisés lors de cette thèse ont permis de montrer par résonance plasmonique de surface et par "dot blot", que les ligases MurD, MurE, MurF interagissent toutes avec MurG et MreB, ces deux dernières formant elles-mêmes un complexe. En revanche, aucune interaction n'a été détectée entre les ligases. Un criblage de conditions de cristallogenèse a été effectué afin de déterminer l'empreinte cristallogénique et cristallographique des protéines seules ainsi que des complexes potentiels dans le but de déterminer la structure atomique de l'un des complexes étudiés. Grâce à ce criblage, la structure par diffraction aux rayons X des trois ligases a pu être résolue, suggérant que la flexibilité conformationnelle de leur domaine C-terminal pourrait être importante dans les interactions protéiques. En outre, une condition donnant naissance à des cristaux pouvant contenir le complexe MreB-MurF a pu être identifiée et son étude cristallographique est en cours.

Ces résultats marquent les premiers pas dans la caractérisation de la machinerie cytoplasmique de la biosynthèse du peptidoglycane, ouvrant la porte à de nouvelles cibles thérapeutiques.

Mots-clés: paroi bactérienne, biosynthèse du peptidoglycane, Mur ligases, MurG, MreB, interactions protéine-protéine, résonance plasmonique de surface, cristallographie par diffraction aux rayons X, cristallisation de complexes protéiques, *Thermotoga maritima*, cibles antibiotiques.

Contents

Acknowledgements

My Ph.D. fellow at the Institute of Structural Biology has been a wonderful and enlightening experience. Clearly, the research project itself together with the scientific environment in Grenoble contributed to broaden my scientific and technical knowledge in a passionating field. But what made it an overwhelming adventure was all the gravitating tasks around as well, such as how to work in a group, give talks, set-up collaborations, write papers, stay focus on tens of things at once, open my eyes to all the knowledge and know-how available in the institute... and spend nights in front of scattering patterns until the birds start singing. I am indebted to many people for supporting my work, providing invaluable guidance, and making the time working on my Ph.D. an unforgettable experience.

First and foremost I would like to express my heartfelt gratitude to my advisor Andrea Dessen. Her enthusiasm was contagious and motivational to me and I admire her positive outlook. She was always supportive, inspirational, and encouraging in tough times, and has given me a great freedom in the way I wanted to tackle my research project. I hope that I could be as lively, generous, enthusiastic, and energetic as she is. I appreciate all her contributions of time, cooperation in allowing me to participate in associations, congresses, and all kinds of events, therefore making my Ph.D. experience productive, diverse, and stimulating. I am also thankful for the example she has provided as both a talented and hard-working researcher, and devoted mother as well. Above all, you made me feel a friend, which I appreciate from all my heart.

I am particularly indebted to Carlos, a knowledgeable expert in crystallography, with an amazing ability to manipulate 10 μm crystals. Carlos, known as *'the sample changer tamer'* of the *'fishing machine'*, who fueled my fondness for data collection and crystallography. Carlos is so fast in structure solution that I felt guilty when I had too few crystals to pass during our beam times and I could not collect enough data sets for him to play with. When Carlos is refining, he looks like a kid playing video games. It was a real pleasure to work with you and see how you could explain crystallography formulas with your hands. I wish you all my best for the research project you are starting.

I have been very privileged to get to know and to collaborate with Nicole Thielens, who taught me everything about Surface Plasmon Resonance in biology. Always ready to help, Nicole assisted me in data analysis and interpretation very much, and was extensively involved in the preparation of my first paper.

Special thanks go to Viviana Job, *'the cloning queen'*, who taught me, when I was a young Ph.D. student, great insights into molecular biology and protein purification, thus contributing to my good start. I am very grateful about her willingness to share her broad experience in protein purification. Your tremendous grasp of experimental issues had a great impact on me.

Other members of the Bacterial Pathogenicity Group have contributed immensely to my work through good advice and sometimes collaboration. I am particularly thankful to Tommaso Tosi, a great office mate who I would go to for answers to many questions. I am very much impressed by your technical know-how and deep understanding of cutting-edge methods. My deepest apologize for all the times I forgot to put the lab key back to the secret place. Special thanks go to Julia Nikolaidis, the self-taught specialist in membrane purification of the group, who shared with me some tips she picked up during her tens (hundreds?) of membrane purifications. As soon as I do not overuse your magic pipet, I am convinced it will become more productive in making membrane protein crystals. Best wishes to you as you finish up. Special thanks to Alex who helped me with bringing my very needed supplies from the other side of the world and contributed to the sportive atmosphere in the group, often talking about his weekly 30 km runs in the mountains.

Other past group members that I have had the pleasure to work with or alongside of are Karen Fulan-Discola (I was very much amazed by your ability to never give up, to never be discouraged by your very difficult projects); David Neves, the so friendly *'Junior Boss'*; Munan Shaik, the *'structure-killer'*, Steve Wong and his wife Jenn for sharing a few rock climbing afternoons with me, and Claire Gendrin for the very helpful discussions we had, her delicious cakes, and mountaineering trips as well; Alex Pflug for the enjoyable debates about life, Ph.D.s, politics, and so on. I am also indebted to the two undergraduate students I had the pleasure to work with, Emmanuel Hamelin and Florian Fabiani, who both worked very hard and learnt fast to become autonomous. Special thanks to Manu, who prepared a few nice crystals of MurF. It was a pleasure to welcome Damien Rondet for a voluntary internship. Added to the promising results he obtained, his childish joviality contributed to a great working atmosphere in the lab. I am glad to have worked along side the numerous summer and rotation students who have come through the lab and provided a pleasant and productive working atmosphere, especially Martina Hrast from Ljubljana who tested MurF activity.

I had also the pleasure to work in collaboration with Andrej Perdih from the Laboratory for Biocomputing and Bioinformatics of the National Institute of Chemistry in Slovenia with inhibitors of Mur ligases.

I would like to acknowledge our group's administrative assistant Chantal Robesson who helped me a lot in organizing my trips to workshops and symposiums and other very complicated administrative issues.

I thank the members of my PhD committee, who provided encouraging and constructive feedback. I very much appreciated Imre Berger enthusiasm and Jean-Pierre Simorre's pragmatism. All this work would not have been possible without the contribution of the platforms of the Partnership for Structural Biology and their responsible persons who made significant contribution to this work, including the MP3 platform and highly professional Michel Thepaut, the crystallization platform directed by the so friendly and helpful Delphine Blot, the HTX lab from EMBL which made more than 3,000 crystallization drops for my projects, the mass spectrometry platform and Luca Signor, the electron microscopy

platform with Daphna Fenel. People at the Institute of Structural Biology are genuinely nice and helpful and I am glad to have interacted with many people who were not part of the Dessen group, including André Zapun who provided me with his modified pETDuet constructs.

I gratefully acknowledge the IRTELIS fellowship from CEA that made my Ph.D. work possible.

My time in Grenoble was made enjoyable in large part due to the many friends and groups that became a part of my life. I am grateful for time spent with mountaineering partners, especially Fabrice Rastello and our memorable trips into the mountains, Matthieu Battini for rock climbing introduction, and Lionel Imbert for noon joggings and countless chocolate cups. Furthermore, I also want to thank all other friends who put up with me through the whole Ph.D. process and helped me with personal and family challenges, in particular my two 'filles au pair', Viktoriia Tovpeko and Lejla Ihracska, who took care of kids during beam time and endless experiments.

Lastly, I would like to thank my family for all their encouragement. My parents who raised me with a love of science and supported me in all my pursuits. Erwann who pretended listening to my experimental issues and complains about the lab. Though he met Andrea just once, he always had the same reply when I announced a good result about single proteins: "And what about complexes?"... My grand-mother who has always been showing interest in my research progress. Last but not least, I would like to thank my two children, Anna and Alan, who kindly let me sleep after synchrotron nights.

Preamble

The work of this Ph.D. was performed in the Bacterial Pathogenesis Group of the Institute of Structural Biology[1], under the supervision of Andréa Dessen and with the invaluable help of Carlos Contreras-Martel and Viviana Job.

This thesis is intended to a public who possesses a general scientific background. Indeed, no extensive knowledge in crystallography nor biology is required. However basics about proteins, DNA, transcription and translation processes, optics, and electromagnetic waves are needed.

Attention was made to carefully explain every technique and experimental methods used in this work, and chapters were written to facilitate any potential partial reading of this thesis. In small letters are mentioned some theoretical points which are not necessary for a first reading, but which can be of interest for those who want to go deeper into details. Appendices give even more technical details. A quick overview of the thesis can be obtained by reading text boxes which contain the main explanations, results, and conclusions.

The essence of the work presented here was the object of a publication in *Environmental microbiology*, entitled 'MreB and MurG as scaffolds for the cytoplasmic steps of peptidoglycan biosynthesis' and published in July 2013. Nevertheless, additional results are reported as well. Moreover, a review about resistance to antibiotics is being drafted in collaboration with Julia Nikolaidis and Andréa Dessen.

[1] 41 rue Jules Horowitz, 38027 Grenoble cedex 1, France.

Abbreviations

3D: three-dimensional
β-OG: n-octyl-beta-D-glucopyranoside
A_2pm: *meso*-diaminopimelic acid
AHT: Anhydrotetracycline
AMPPCP: β,γ-methyleneadenosine 5'-triphosphate
AMPPNP: Adenylyl-imidodiphosphate
A.U.: Absorbance units
B. subtilis: *Bacillus subtilis*
B. pertussis: *Bordetella pertussis*
C. crescentus: *Caulobacter crescentus*
CMC: Critical micelle concentration
CV: crystal violet
C-ter: C-terminal
DDM: n-Dodecyl β-D-Maltopyranoside
DNA: DeoxyriboNucleic Acid
dNTPs: DeoxyriboNucleotide Triphosphates
dsDNA: Double stranded DNA
DTT: Dithiothreitol
e.g.: *exempli gratia*
EDC: Ethyl(dimethylaminopropyl) carbodiimide
EDTA: Ethylenediaminetetraacetic acid
EGS: Ethylene glycol bis(sulfosuccinimidylsuccinate)
EM: Electron Microscopy
EMBL: European Molecular Biology Laboratory
E. coli: *Escherichia coli*
ESI: ElectroSpray Ionisation
FF: Fast-flow
FPLC: Fast protein liquid chromatography
FT: Flow Through
GlcNAc: *N*-acetylglucosamine
H. influenzae: *Haemophilus influenzae*
HEPES: 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HF: High-Fidelity
HP: High-performance
IBS: Institut de Biologie Structurale

IPTG: Isopropyl β-D-1-thiogalactopyranoside
IWZ: Inner wall zone
LB: Lysogeny broth
LMW: Low molecular weight protein marker from GE Healthcare
LPS: Lipopolysaccharide
LTA: Lipoteichoic acid
M9: Minimum medium
MALDI: Matrix-Assisted Laser Desorption
MES: 2-(N-morpholino)ethanesulfonic acid
MP3: Membrane Protein Purification Platform (IBS)
MS: Mass Spectrometry
M. tuberculosis: *Mycobacter tuberculosis*
N-ter: N-terminal
NaPi: Sodium Phosphate Buffer
NADPH: Nicotinamide adenine dinucleotide phosphate
NEB: New England Bio labs
NHS: N-Hydroxysuccinimide
OD: Optical density
O/N: Overnight
OWZ: Outer wall zone
PBS: Phosphate Buffered Saline
PBS-T: Phosphate Buffered Saline supplemented with 0.05% Tween 20
PBP: Penicillin Binding Protein
PCR: Polymerase Chain Reaction
PDB: Protein Data Bank
PG: Peptidoglycan
pI: Isoelectric Point
PSB: Partnership for Structural Biology
P. aeruginosa: *Pseudomonas aeruginosa*
res: residue
RFU: Relative Fluorescence Units
rpm: Revolutions per minute
RT: Room Temperature
S. agalactae: *Streptococcus agalactiae*
S. aureus: *Staphylococcus aureus*
SAD: Single-wavelength Anomalous Diffraction
SDS-PAGE: Sodium Dodecyl Sulfate PolyAcrylamide Gel Electrophoresis
SEC: Size Exclusion Chromatography
SeMet: Seleno-methioninylated
SOC: Super Optimal broth medium with Catabolite repression
T. maritima: *Thermotoga maritima*
ter: terminal
TIR: Total internal reflection
TIRF: Total internal reflection fluorescence microscopy

Tris: Tris(hydroxymethyl)aminomethane
TSA: Thermal Shift Assay
U: Unit
UDP: Uridine diphosphate
UDPGlcNAc: UDP-*N*-acetylglucosamine
UDPMurNAc: UDP-*N*-acetylmuramic acid
UMA: UDP-*N*-acetylmuramoyl-L-alanine
WB: Western Blot
Y. pestis: *Yersina pestis*

Part I

Introduction

Chapter 1

The bacterial cell wall

Many human illnesses, including pneumonia, tuberculosis, meningitis, and diverse skin and blood infections, are caused by bacterial pathogens. Unfortunately, current antibiotics are not able to fully control bacterial infections anymore, because of the emergence of an increasing number of multi-drug resistant strains, renewing the threat of pandemics. Therefore, scientists have to explore new ways for developing antibacterials [1].

The bacterial cell wall remains a target of prime interest. Particularly, disrupting the assembly of peptidoglycan, the major structural component of the bacterial cell wall [2] [3], has been successfully exploited in developing most antibiotics in clinical use as most enzymes involved in this biosynthetic pathway are essential, well-conserved, and do not have any mammalian homolog [4]. Strikingly, while most antibiotics target the late steps of peptidoglycan synthesis, the earlier, cytosolic, steps are still underexploited and are the purpose of the work presented here.

This introduction first gives the basics about the role and composition of peptidoglycan, then describes its biosynthesis pathway, and lastly provides a state-of-the-art description of the structural knowledge of proteins involved in this pathway.

1.1 Functions and constitution

In contrast to human cells, the cell membrane of bacteria is surrounded by a cell wall which provides rigidity, impermeability to some compounds, and strength to counteract internal osmotic pressure [5] (see Figure 1.1).

The composition of the cell wall is at the root of the very common classification of bacteria between Gram-positive and Gram-negative species (see Figure 1.2), according to the Gram staining protocol[1].

In Gram-positive bacteria, the cell wall is relatively thick (see Figure 1.2, left). Its major

[1]In this assay, cells are first heat fixed onto slides and incubated in a crystal violet (CV) aqueous solution, containing CV^+ and Cl^- ions. The dye is taken up in similar amounts by all bacteria. Cells are subsequently treated with an I_2-KI mixture containing iodine ions I^- which will form a complex with CV^+. This CV-I complex is captured within intact cells. Then, cells are washed briefly with 95% ethanol for destaining. Ethanol interacts with lipids and thus disrupts the outer membrane of Gram-negative cells. The peptidoglycan layer of Gram-negative cells is too thin to be impermeable to CV-I complexes, which will escape from the cells [7]. In contrast, the large CV–I complexes become trapped within Gram-positive cells due to the multilayered nature of its peptidoglycan [7].

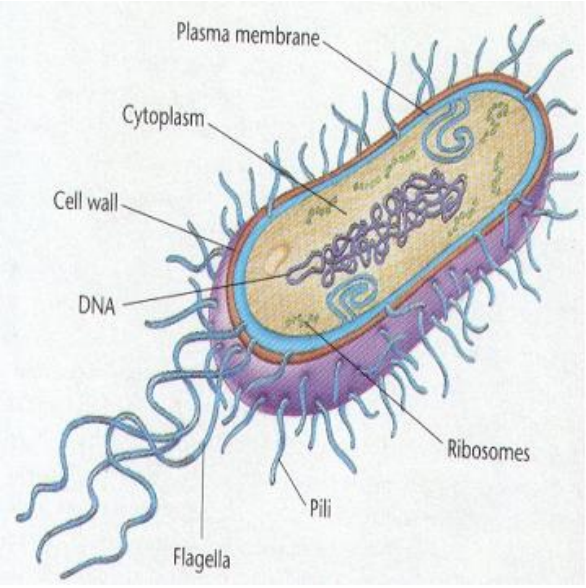


Figure 1.1: Scheme of a common bacterium [6].

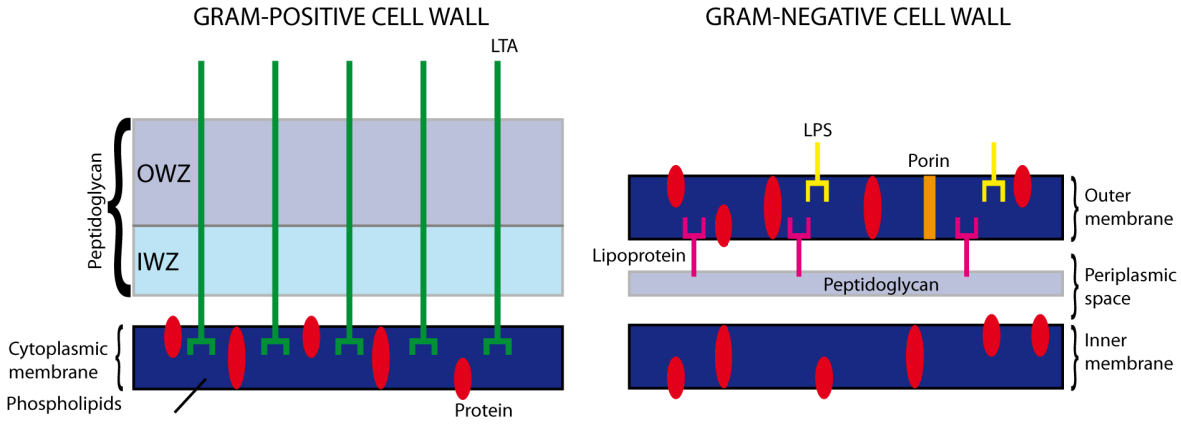


Figure 1.2: Composition of Gram-positive (Left) and Gram-negative (Right) bacterial cell walls. (Left) A Gram-positive bacterial cell wall is composed of a thick and multilayered peptidoglycan mesh outside of the cytoplasmic membrane. Hajipour and co-workers revealed the presence of two sublayers by cryo electron tomography technique: The inner wall zone (IWZ), and the outer wall zone (OWZ) which presents a higher electron density [8]. Teichoic acids are connected to and embedded in the peptidoglycan, and lipoteichoic acids (LTA) extend into the cytoplasmic membrane. (Right) A Gram-negative bacterial cell wall is composed of an outer membrane linked by lipoproteins to a thin peptidoglycan layer located within the periplasmic space that is formed between the outer and inner membranes. The outer membrane includes porins and lipopolysaccharide (LPS) molecules. Adapted from Hajipour *et al.*, 2012 [9].

component (from 50% to 90%) is peptidoglycan. It harbors accessory polymers such as lipoteichoic acid[2] and/or teichuronic acid[3] which are covalently linked to peptidoglycan, and does not contain any lipid, and often no protein. Interestingly, electron cryo-tomography revealed two layers: an inner wall zone (IWZ) of low-electron density, and an outer wall zone (OWZ) of high-electron density [10] [8].

[2]Lipoteichoic acid (LTA) is a phosphate-rich polymer whose structure varies between species and may contain long chains of ribitol or glycerol phosphate. LTA is anchored to the cell membrane via a glyceride. It acts as regulator of autolytic wall enzymes (muramidases) and possesses antigenic properties.

[3]Teichuronic acids are anionic, phosphate-rich polymers that play a role in the integrity of bacterial cell wall.

In Gram-negative bacteria, the cell envelope consists of a pair of membranes (inner and outer) with a thin, intermediate layer of peptidoglycan that could be multilayered, at least on some parts of the cell wall as reported by Vollmer *et al.* [11]. The outer membrane contains lipopolysaccharide (LPS)[4] as well as lipids and proteins.

1.2 Peptidoglycan

Thus, peptidoglycan is a major component of the bacterial cell wall [5]. A brief state-of-the-art overview of its structure and composition is presented here.

1.2.1 An essential structural element

Peptidoglycan contributes to the main features of the cell wall by providing both rigidity in order to maintain a defined cell shape and preserve cell integrity against the osmotic pressure, and the necessary flexibility to adapt bacteria to the different cell shapes relative to various stages of cell cycle. Moreover, it is directly involved in the processes of cell growth and cell division and serves as a harbor for anchoring other cell envelope components such as proteins and teichoic acids [11] [4].

Peptidoglycan is essential to the survival of most bacteria, and any inhibition of its biosynthesis or its specific degradation during cell growth results in cell lysis [11], as observed for many antibacterial treatments.

1.2.2 Chemical composition

Although the detailed chemical composition varies in different species, peptidoglycan is basically made of linear glycan strands interlinked by short peptides.

The glycan chains are composed of two alternating amino sugars, namely *N*-acetylglucosamine (Glc*N*Ac) and *N*-acetylmuramic acid (Mur*N*Ac) [12]. The D-lactoyl group of each Mur*N*Ac residue is substituted by a peptide stem whose composition depends on the species (see Section 1.2.4). Most often, the peptide is composed of L-Ala-γ-D-Glu-*meso*-A_2pm (or L-Lys)-D-Ala-D-Ala (A_2pm, 2,6-diaminopimelic acid) in nascent peptidoglycan, the last D-Ala residue being lost in the mature macromolecule.

A proportion of these peptides are cross-linked either directly or through a second short peptide generally between the carboxyl group of D-Ala at position 4 and the amino group of the diamino acid at position 3. It is this cross-link that gives rise to the three-dimensional peptidoglycan polymer [12], with specific mechanical features. Figure 1.3 depicts an overall scheme of peptidoglycan.

1.2.3 Three-dimensional structure

On one hand, the large size, structural heterogeneity and flexibility of the peptidoglycan network prevent it from crystallizing and hence the three-dimensional crystal structure is

[4]LPS is the major component of the outer membrane of Gram-negative bacteria, contributing greatly to the structural integrity of the bacteria, and protecting the membrane from certain kinds of chemical attack. LPS induces a strong immune response.

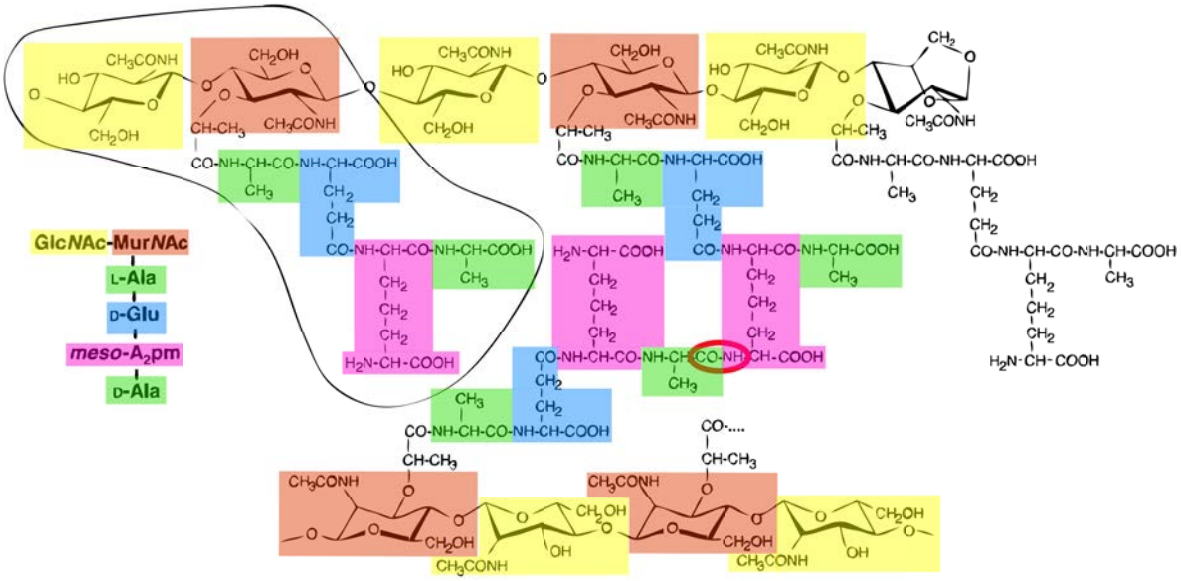


Figure 1.3: Structure of the peptidoglycan of *Escherichia coli*. The glycan strands consist of alternating, β-(1,4)-linked Glc*N*Ac and Mur*N*Ac residues, and are terminated by a 1,6-anhydroMur*N*Ac residue. The encircled part represents the basic disaccharide tetrapeptide subunit (monomer), which is also written with conventional abbreviations on the left-hand side. The middle part shows a cross-linked peptide, with the amide group connecting both peptide stems encircled in red. Adapted from Stenbak *et al.*, 2004 [13].

inaccessible [10]. On another hand, glycan strands and peptides are too small to be visualized by conventional electron microscopy [10]. These are major impediments to the knowledge of precise molecular organization of peptidoglycan, and thus three-dimensional structure remains elusive [10].

However, several models have been proposed, based on chemical and biophysical data. One of them assumes that the peptides protrude helically from twisted glycan strands, with three or four disaccharides per turn [10].

1.2.4 Variability between species

Peptidoglycans differ by the composition of the stem peptide, the nature and extent of peptide cross-linking, and the length of glycan strands [10]. Remarkably, these differences do not seem to be obviously related to the Gram-positive or Gram-negative characteristics of the species.

Length of glycan strands

Strikingly, there are Gram-positive species with a thick cell wall with either short (*S. aureus*) or long (*B. subtilis*) glycan strands. Similarly, there are Gram-negative species with either short (*Helicobacter pylori*) or long (*Proteus morganii*) glycan strands [11]. Therefore, it seems that there is no general correlation between the Gram-positive or Gram-negative nature of a species and the length of glycan strands in terms of 3D organization of peptidoglycan itself [11]. Moreover, it has been reported that glycan-strand lengths depend on strain, growth state, and growth conditions [10]. Figure 1.4 summarizes average dimensions of cell envelope components.

Cryotomography data [14] and analyses of glycan-strand lengths in *E. coli* [10] [15] strongly suggest that glycan strands must be parallel to the cell membrane within the peptidoglycan layer.

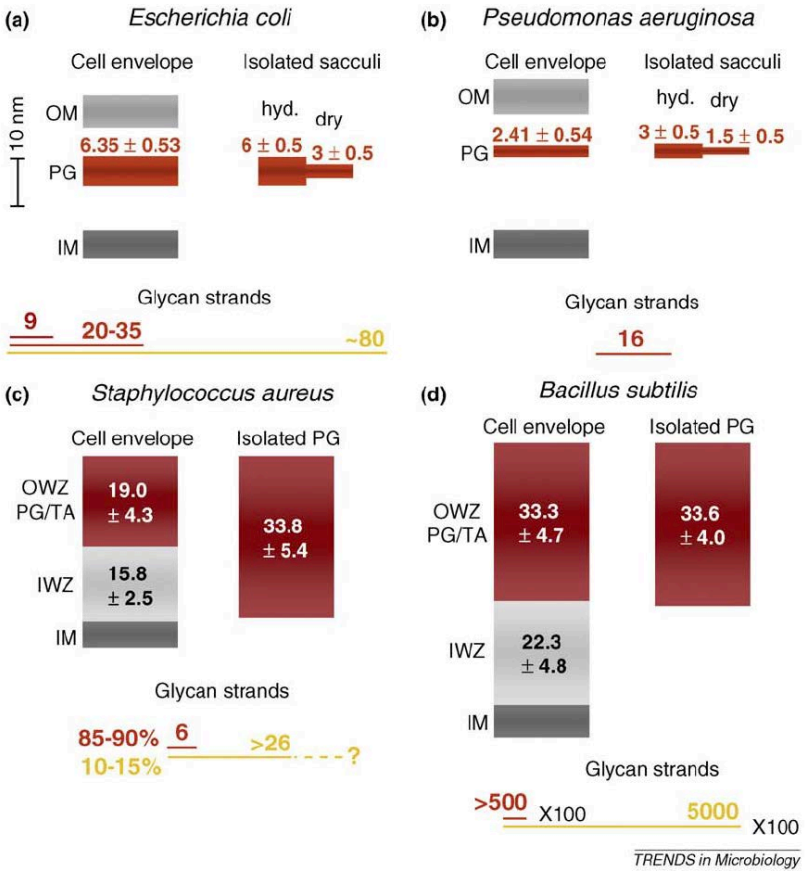


Figure 1.4: Dimensions of cell envelope layers, isolated peptidoglycan and glycan strands in Gram-negative (a, b) and Gram-positive (c, d) species. The thickness of inner membrane (IM), peptidoglycan layer (PG), outer membrane (OM), inner wall zone (IWZ) and outer wall zone (OWZ) (with peptidoglycan (PG) and wall teichoic acid (TA)) are drawn to scale. Numbers indicate thickness with standard deviation in nm, which were measured either by atomic force microscopy (isolated sacculi in (a) and (b)] or by cryo-TEM (isolated sacculi in (c) and (d), and cell envelope layers in all panels). The length of the glycan strands is indicated in nm and drawn to scale (length of a disaccharide unit: 1.03 nm). Dark red line, length of abundant glycan strand; red line, average length of glycan strands; and yellow line, longest glycan strands. The glycan strands in *B. subtilis* are 100-fold longer than the length indicated by the lines shown on the figure. hyd., hydrated. From Vollmer and Seligman, 2010 [10].

Peptide composition

There is a high diversity in the composition and sequence of the peptidoglycan peptides from different species [10]. The variations of the peptide stem can be divided into two categories:

- Those due to the specificity of the enzymes responsible for its biosynthesis;
- Those occuring at a later step of the biosynthesis, often at the level of Lipid II - the peptidoglycan monomer, before its incorporation into the 3D layer [11].

Particularly, peptidoglycan from the Gram-negative thermophilic bacterium *Thermotoga maritima* has been reported to be made of a L-lysine or D-lysine at the third position of the peptide chain, while most Gram-negative bacteria such as *E. coli* display a *meso*-

diaminopimelic acid (A_2pm) at this position [16] [17] [11]. Furthermore, the peptide composition may vary with growth conditions [10].

Peptide cross-linking

The most common peptide cross-linking is the 3-4 cross-linkage described above (Section 1.2.2). But other kinds of cross-links can be observed as well, such as the 3-3 cross-linkage, often seen in β-lactam-resistant strains [11]. Moreover, the size of the optional interpeptide bridge ranges from one to seven amino-acid residues, and its composition varies substantially between species [11].

Besides the diversity in the nature of cross-linking, there is a considerable variation in its degree. Indeed, while connections between only two peptides prevail in *E. coli* and *B. subtilis*, multimeric peptides with up to 20 connected peptides exist in *S. aureus* [11] [10]. Globally, it seems that peptidoglycans from Gram-positive bacteria are more cross-linked than Gram-negative peptidoglycans.

Peptidoglycan

The peptidoglycan layer is a specific component of bacteria responsible for the main mechanical features of the bacterial cell wall, and therefore essential to most species. Basically made of glycan strands cross-linked by short peptides, its composition displays a very high variability between species. However, the pathway of its biosynthesis is very well conserved within the bacterial world, and is addressed in the next chapter.

Chapter 2

Peptidoglycan biosynthesis

Biosynthesis of peptidoglycan is a multi-step process involving a series of enzymes and structural proteins (Figure 2.1), which can be described into two main steps. First, the monomer of peptidoglycan also known as Lipid II[1], is synthesized in the cytoplasm and incorporated in the membrane. Then, at the membrane level, Lipid II is flipped to the periplasm and subsequently transferred into the growing peptidoglycan layer where it is finally linked to the existing peptidoglycan network with the contribution of Penicillin Binding Proteins (PBPs) [5].

This chapter focuses on the synthesis of Lipid II, gives the basics of Lipid II polymerization and cross-linking into the peptidoglycan layer, and finally provides the current knowledge about the spatial organization of the peptidoglycan biosynthesis process which involves the bacterial actin homolog MreB.

2.1 Cytoplasmic steps: towards the synthesis of Lipid II

First of all, the MurA and MurB enzymes catalyze the synthesis of UDP-*N*-acetylmuramic acid (UDPMurNAc) from UDP-*N*-acetylglucosamine [18] [5]. Then, MurC, MurD, MurE, and MurF, which are ATP-dependent amino acid ligases, sequentially catalyze the addition steps of a short polypeptide chain to UDPMurNAc [5]. These first six steps occurring in the cytoplasm are followed by two steps taking place on the cytoplasmic side of the membrane where the final unit of peptidoglycan, known as Lipid II, is synthesized with the help of the transmembrane enzyme MraY and the glycosyltransferase MurG [5]. Based on Figure 2.1, this section describes into greater details the successive steps of Lipid II biosynthesis.

2.1.1 The very first precursor: UDP-*N*-acetylglucosamine

The basic precursor of peptidoglycan biosynthesis is a nucleotide sugar (see Figure 2.1) naturally synthesized through the hexosamine biosynthesis pathway from glucose, glucosamine, and uridine.[2]

[1] Lipid II is composed of a sugar moiety, a lipid, plus a peptide chain.

[2] The hexosamine pathway is used by a lot of glycosyltransferases to transfer N-acetylglucosamine residues to substrates in a high number of biological processes [19].

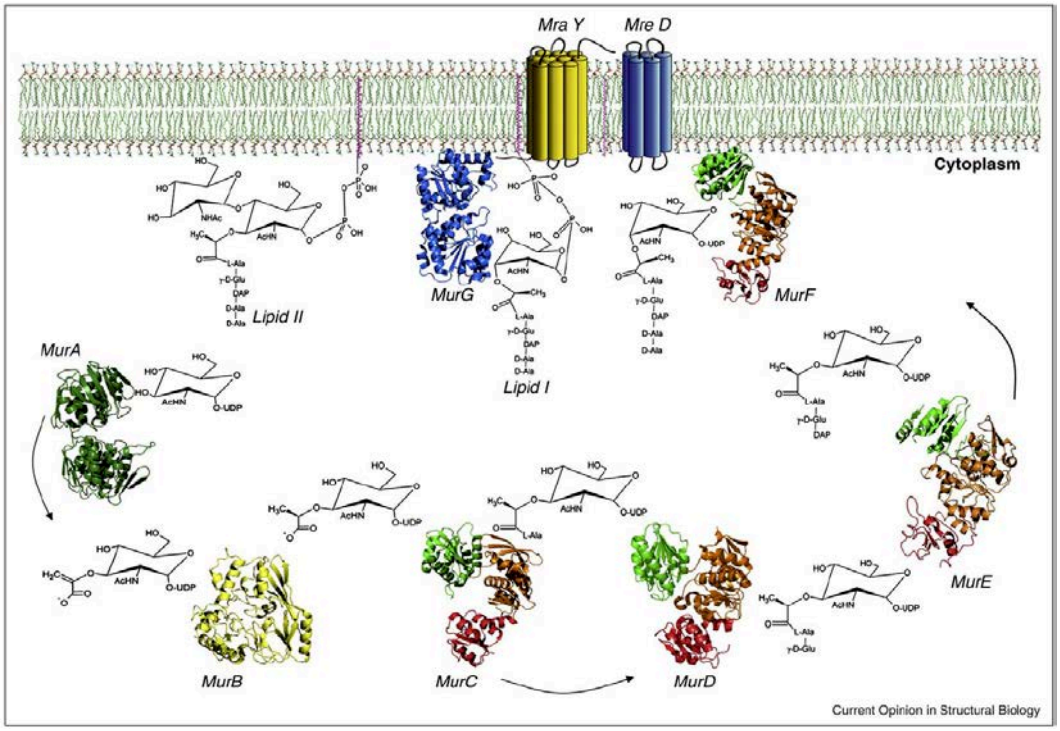


Figure 2.1: Schematic representation of the essential steps of peptidoglycan synthesis. The different domains in Mur ligases are shown in distinct colors (N-terminal, red; central domain, orange; C-terminal, green). Structures shown in the figure include MurA (1NAW), MurB (1MBT), MurC (1GQQ), MurD (2JFH), MurE (2WTZ), MurF (1GG4), and MurG (1F0K). The undecaprenyl phosphate carrier lipid is indicated in pink. From Mattei *et al.*, 2010 [18].

2.1.2 Synthesis of UDPMurNAc by MurA and MurB enzymes

First, MurA catalyzes the addition of a phosphoenolpyruvate to the hydroxyl at carbon 3 of the glucosamine ring of UDP-*N*-acetylglucosamine to form enolpyruvyl-UDP-*N*-acetylglucosamine[3]. Subsequently, MurB catalyzes the reduction of the pyruvyl moiety to a lactyl group by NADPH, giving UDP-*N*-acetylmuramic acid (UDPMurNAc)[4].

2.1.3 Addition of a short peptide by Mur ligases

A short polypeptide chain is then added to the sugar molecule UDPMurNAc through successive steps by four similar ATP-dependent amide bond ligases [5]. As described in the previous chapter, in most species the peptide is composed of the following residues: L-alanine, D-glutamate, *meso*-diaminopimelate, D-alanine and D-alanine [5].

- MurC is an UDP-*N*-acetylmuramoyl:L-alanine ligase
- MurD is an UDP-*N*-acetylmuramoyl-L-alanine:D-glutamate ligase
- MurE is an UDP-*N*-acetylmuramoyl-L-alanine-D-glutamate:*meso*-diaminopimelate ligase
- MurF is an UDP-*N*-acetylmuramoyl-L-alanine-D-glutamate-*meso*-diaminopimelate:D-alanyl-D-alanine ligase [5].

[3]Thus, MurA is an UDP-*N*-acetylglucosamine enolpyruvyl transferase.
[4]Thus MurB is a UDP-*N*-acetylpyruvylglucosamine reductase.

2.1.4 Final steps of Lipid II synthesis

The two last steps occur at the surface of the cytoplasmic membrane and are catalyzed by the transmembrane protein MraY and the membrane-bound glycosyltransferase MurG.

- The previously synthesized UDP-Mur*N*Ac-peptide is attached to a membrane-bound lipid carrier molecule, undecaprenol-phosphate[5], catalyzed by the integral membrane protein MraY [20], to give Lipid I.
- The glycosyltransferase MurG transfers *N*-acetylglucosamine (GlcNAc) to Lipid I, forming Lipid II, the monomer of the final peptidoglycan polymer [5].

2.2 Periplasmic steps: incorporation and polymerization

Lipid II is translocated across the membrane, most likely with the contribution of RodA/FtsW in order to be incorporated into the growing peptidoglycan layer by the formation of glycosidic linkages between the disaccharide units, cross-linking of the peptide tails, and cleavage from the lipid carrier, facilitated by transpeptidases, endopeptidases, and penicillin-binding proteins (PBPs), the latter being the main targets of antibiotics in clinical use [5] [21].

2.3 Spatial organization of peptidoglycan biosynthesis

As seen in the previous section, the pathway of peptidoglycan biosynthesis is very well known. Indeed, most steps correspond to a known protein which has been extensively studied for most cases. However, how all these steps are spatially organized within the cell remains unclear. Nevertheless, an increasing number of interactions between peptidoglycan biosynthesis actors have been described over the past decade, suggesting the existence of multipartite protein complexes. This section focuses on the clues supporting the peptidoglycan machinery hypothesis, and introduces the bacterial actin homolog MreB which is thought to play a key organizational role in the process.

2.3.1 MreB: a bacterial actin homolog with a strong morphogenetic role in rod-shaped bacteria

The *mreB* gene

In many organisms, *mreB* is part of the *mre* (murein cluster e) operon, one of the major operons involved in cell-shape determination in bacteria [22].

While most often Gram-negative species have only a single copy of *mreB*, Gram-positive organisms may have several *mreB*-like genes [23]. Such paralogues and orthologues could be specialized in different aspects of morphogenesis [23]. Interestingly, *mreB* genes are absent from most bacteria displaying spherical shapes [23].

[5]Undecaprenol-phosphate is a 55-carbon compound containing 11 isoprenoid units (five carbons).

Though *mreB* seems to be essential, the lethality or loss of its functionality can be suppressed either by overexpressing cell division proteins or reducing the rate of cell growth [24], or by growth in increased levels of Mg^{2+} [25] regarding *B. subtilis*.

The MreB protein: polymerization and membrane binding

Most *in vitro* work on MreB has been performed on the *Thermotoga maritima* protein, revealing that MreB presents two important and functional properties, namely polymerization and membrane binding.

Most of studied MreB homologs were found to be capable of polymerization. For instance, MreB from *T. maritima* forms long multilayered sheets of interwoven filaments in the presence of either ATP or GTP *in vitro* [26] [22]. The longitudinal repeats in MreB filaments resemble that of actin [22], although a few differences have been described between the superstructures of filaments, and between polymerization kinetics as well [22] [27] [28]. Recent biochemical and functional studies on *B. subtilis* [29], *Chlamydiaceae* [30], and *E. coli* [31] underlined the high diversity of MreB homologues. For instance, *B. subtilis* MreB does not require nucleotide to polymerize, and its kinetic behavior is much different from that of *T. maritima* [32]; while work from Bean and Amann [33] suggested that most *in vivo* polymerized MreB could be in an ADP-bound state.

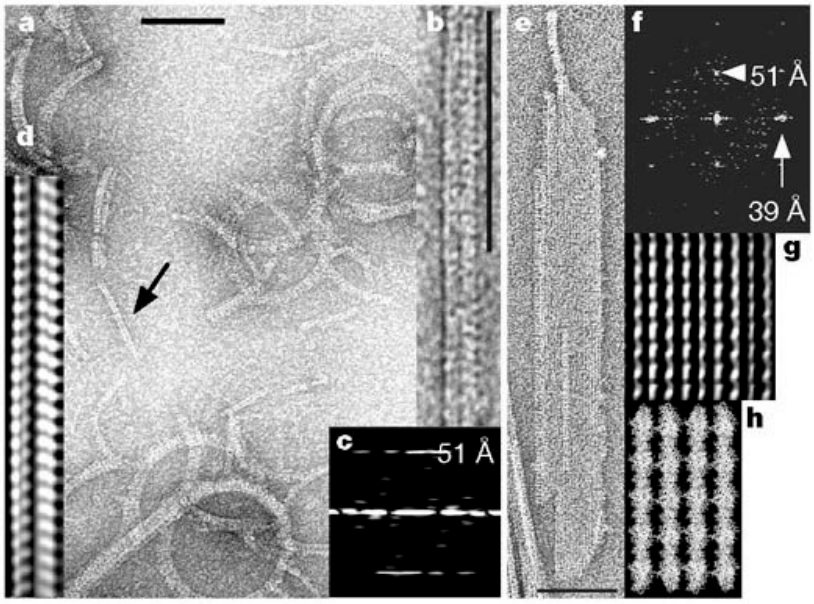


Figure 2.2: MreB polymers. Typical view of MreB filaments (a), double filaments (b, d: filtered image, c: diffraction image), and MreB sheets (e, g: filtered image, f: diffraction image) in different buffer conditions. In MreB sheets, the polymer is about 160 Åwide, which suggests four single protofilaments in total (each 40 Å). The longitudinal repeat is 51 Å, the lateral spacing is 39 Å. Scale bars, 100 nm. (h) The protofilaments found in the crystals of MreB fit well with the filtered image in (g). From van den Ent *et al.*, 2001 [22].

Salje *et al.* showed that MreB from both *T. maritima* and *E. coli* bind directly to cell membranes [34] and provided a structural analysis (see next Chapter). Interestingly, they revealed that membrane-binding activity in *E. coli* is essential for the function of MreB in cell shape determination [34]. Their work suggests that membrane binding of MreB would orient the protofilament along the membrane surface [34].

The role of MreB in cell-shape determination

The *in vivo* properties of MreB have been mainly characterized in *B. subtilis*, *Caulobacter crescentus*, and *E. coli*.

A number of depletion or functionality loss experiments have shown that MreB is required for cell shape in most non-spherical bacteria [35]. Figure 2.3 is an example of the aftermath of *mreB* depletion in *C. crescentus*: cells lose their regular shape and start lysing.

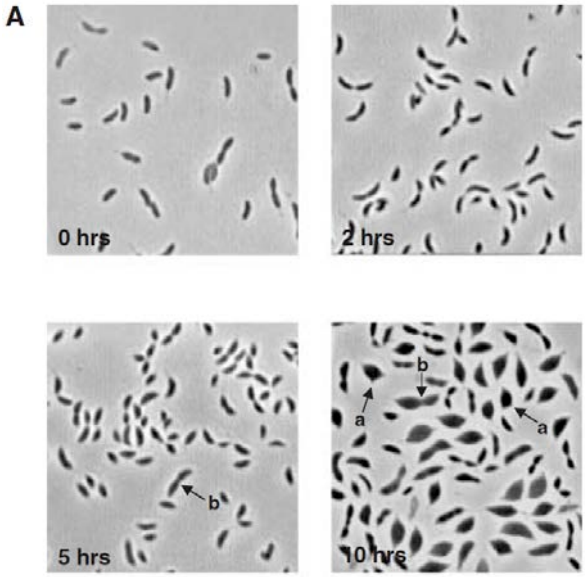


Figure 2.3: MreB depletion results in defects in cell shape. Cells of *C. crescentus* containing a xylose-inducible allele of *mreB* and a deletion in the wild-type copy of the *mreB* gene were grown in the presence of xylose, washed three times in inducer-free medium and resuspended in fresh medium lacking inducer. Phase-contrast images were obtained before washing (0 h) and after 2, 5 and 10 h of incubation without inducer, as indicated. Loss of MreB results in an abundance of lemon-shaped cells, some of which have membrane blebs indicating a loss of cell wall integrity (cells labelled 'a'). Less abundant is the presence of other cells with defects in cell division (cells labelled 'b'). These cells possess a constriction at the mid-cell, as would occur if they arrested at the pre-divisional stage. From Figge *et al.* [36].

In addition to depletion assays, MreB was shown by immunofluorescence to colocalize with RodZ. This bitopic inner membrane protein has been widely reported to be required for cell shape determination as well [37] [38], though a few studies suggested that it is not always the case [39]. The interaction was further confirmed by work of van den Ent *et al.* [35] in 2010, who solved the crystal structure of MreB in complex with the cytosolic part of RodZ and provided *in vitro* data supporting the evidence of a functional significance of the interaction [35] (see Chapter 3).

MreB and cell-cycle

The prevailing model for bacterial growth predicts two spatially specialized pathways that control growth in rod-shaped bacteria: one for elongation in which MreB is strongly involved, and the other one for division that would be MreB independent [40] [41].

In support, conventional microscopy experiments in *B. subtilis*, *E. coli*, *P. aeruginosa*, and *Rhodobacter* have characterized both helical (elongation phase), and medial (division phase) MreB distributions [42] [43] [44] [29] as illustrated by Figure 2.4. In addition, Daniel

and Errington showed that, in *B. subtilis*, synthesis of the cylindrical part of the cell wall occurs in a helical manner that is reminiscent of the helical structure observed for Mbl filaments (a MreB homolog) [23]. Therefore, localization of MreB is believed to be regulated during the cell cycle [36] [45], presenting a helix-like (or patches, see below) pattern during elongation and a medial distribution during division.

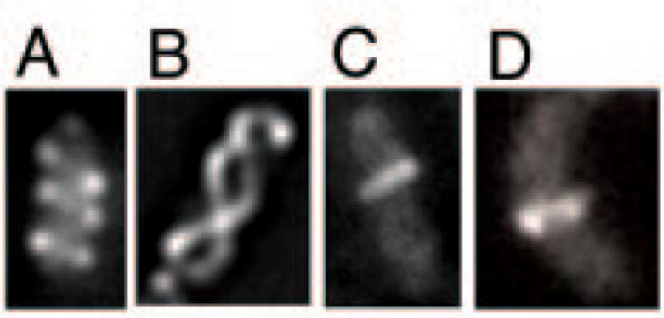


Figure 2.4: Cellular localization of MreB labeled with GFP or YFP. (A to C) *E. coli* MreB localizes into extended coils (A), intertwined double helices (B), and band-like structures (C). (D) *C. crescentus* MreB localizes into a band-like structure at the division site in pre-divisional cells [41].

MreB pattern during elongation

However, the exact pattern of MreB during elongation is still under controversy. Indeed, MreB was thought to build a large, rigid, membrane-associated, helical scaffold within the cell which was suggested by conventional and confocal microscopy assays (see Figure 2.4 A and B). By contrast, two recent reports suggested that in *B. subtilis* MreB polymers could be composed of short, dynamic filaments as illustrated in Figure 2.5. In addition, the helical MreB superstructure in *E. coli* was reported to be an artifact of the N-terminal YFP tag [46].

This theory is supported by two main studies:

- TIRF[6] experiments, in which Dominguez *et al.* could not see any MreB helix near or along the surface of the inner membrane during exponential growth of *B. subtilis* cells. Instead, the authors detected discrete MreB-patches that moved processively along peripheral tracks perpendicular to the cell axis and that colocalized with other morphogenetic factors [48].
- electron cryo-tomography studies of six different rod-shaped bacterial species at macromolecular resolution could not distinguish any long (> 80 nm), membrane-associated helical filaments encircling cells [49].

2.3.2 Role of MreB in peptidoglycan biosynthesis

The implication of MreB in peptidoglycan biosynthesis was first suggested by microscopy experiments revealing co-localization patterns with murein enzymes. These data have been completed by biochemical results which will be discussed in this section. In addition, it has

[6]TIRF (Total internal reflection fluorescence microscopy) is a high-resolution technique able to capture the surface of one side of a cell and useful for visualizing activities at the membrane

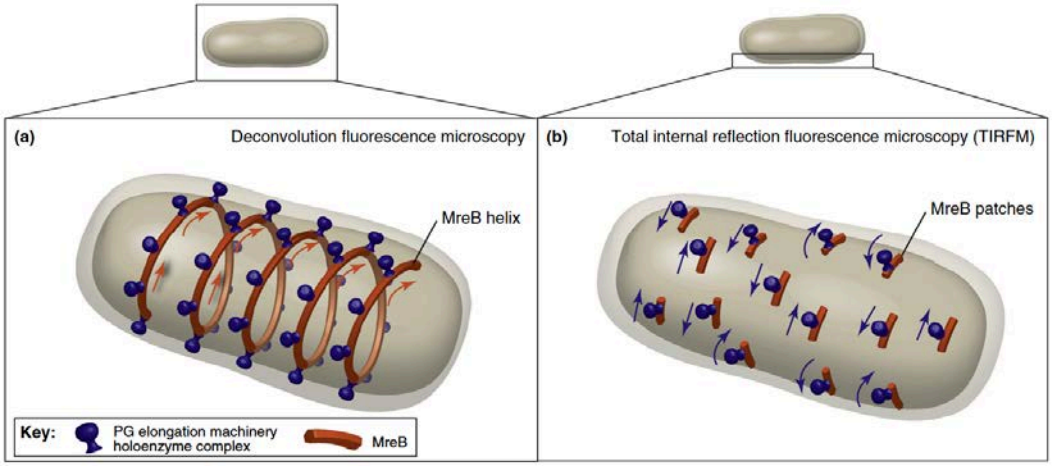


Figure 2.5: Schematic representation of MreB ultrastructure and movement seen by two different microscopic techniques. Both panels depict MreB (dark orange) coupled with the peptidoglycan (PG) elongation machinery (purple), which collectively represents cell wall synthetic enzymes and cell shape determining proteins, including, but not limited to, MreC, MreD, RodA, RodZ, PBPs and PG cytosolic synthetic enzymes. (a) Represents what is seen using deconvolution fluorescence microscopy, where a stack of images taken through the cell body depicts a helically structured MreB. (b) By contrast, TIRF microscopy found both MreB and a selection of several PG elongation proteins that move in short patches (b) as opposed to long helical filaments (a). Schematic representations are not drawn to scale [47].

been reported that inhibition of MreB with the small molecule A22 in *Caulobacter* leads to shortened cell wall glycan strands [50], indicating that MreB has a general influence on peptidoglycan assembly.

MreB as a bridge between the cytoplasmic and periplasmic steps

MreB, MreC, MreD, along with the cell wall assembly proteins PBP2 and RodA, all involved in late stages of peptidoglycan biosynthesis, have been proposed to form a complex [51] [48]. Therefore, the cytoplasmic MreB filaments may orchestrate peptidoglycan biosynthesis from the cytoplasm through its direct or indirect interactions with transmembrane or periplasmic proteins [52].

In addition, MreB was shown to interact with RodZ (see above). For instance, microscopy assays observed that RodZ was associated to MreB in space and time and marked future sites of peptidoglycan synthesis [37] in *Caulobacter crescentus*, suggesting an additional link through the bacterial cytoskeleton and future sites of peptidoglycan incorporation.

MreB could recruit cytoplasmic cell wall synthesizing proteins

White *et al.* found that in *C. crescentus* MreB cables are required for the organization of several other cytosolic murein biosynthetic enzymes such as MraY, MurB, MurC, MurE and MurF [53], each of these proteins adopting a subcellular pattern of localization comparable to MurG. These microscopy data strongly suggest the existence of cytoskeletal-dependent interactions among cytoplasmic actors of peptidoglycan biosynthesis [53] [54].

This has been supported by co-pelletting assays showing interactions between MreB and MurF [30], and bacterial two-hybrid system experiments, finding that MurG interacts with MreB in both *C. crescentus* [53] and *Chlamydia* [30].

White *et al.* proposed a model for peptidoglycan biosynthesis machinery, summarizing

known or suggested interactions among bacterial morphogenetic proteins and cell wall synthesizing enzymes (Figure 2.6). They propose that MreB would organize the cytoplasmic steps by interacting with a number of murein-synthesizing proteins such as Mur enzymes, MraY, and the RodZ and RodA morphogenetic proteins as well. On the periplasmic side, MreC woud position a peptidoglycan assembly complex consisting of PBPs and lytic enzymes (i.e. MltA, MipA). The integral membrane protein MreD would contribute to connection of both cytoplasmic and periplasmic complexes through interactions with MreB, MreD, and transmembrane proteins [53].

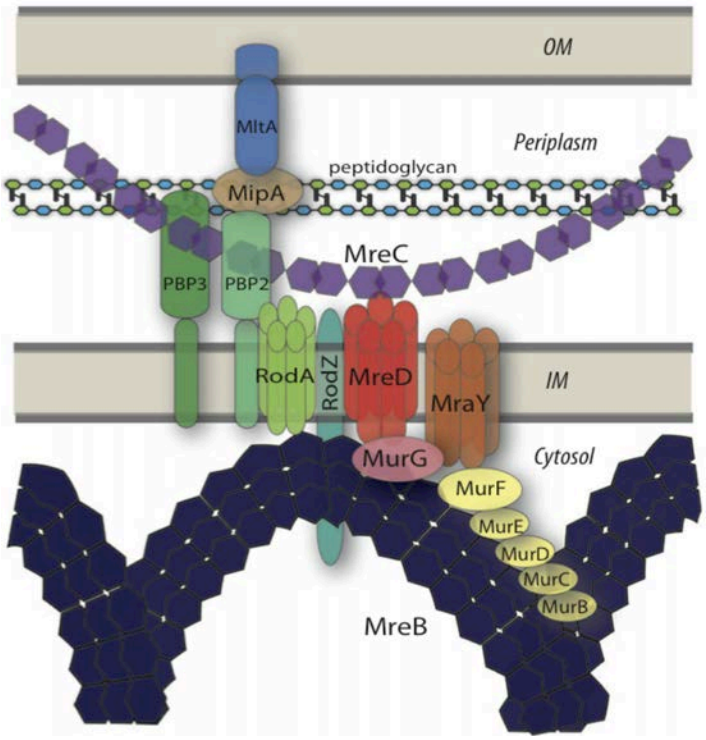


Figure 2.6: Model depicting the peptidoglycan biosynthesis machinery. Schematic view of interactions within peptidoglycan biosynthesis pathway, focusing on the crucial role of Mre proteins in spatially organizing other actors of peptidoglycan biosynthesis [53].

Exact role of MreB in peptidoglycan biosynthesis

All these data suggest the existence of an interaction between MreB polymers and the cell-wall synthetic machinery. However, the exact role of the bacterial actin homolog remains unclear, and two theories have been proposed to date.

The first model, based on the MreB helix mentioned above and helical incorporation of newly synthesized peptidoglycan monomers, postulates that helical MreB cables direct cell wall growth by appropriately positioning the peptidoglycan elongation machinery [55]. The second model, based on the recent studies revealing patches rather than a helical superstructure, proposes that MreB-patches, that contain sidewall elongation enzymes as well, would be effectively dragged along by peptidoglycan synthesis itself [55] [48] [56].

However, both models see MreB as an 'anchor', either to target peptidoglycan biosynthesis (first model) or to constraint the movement of the peptidoglycan synthesis machinery

(second model), ensuring the completion of peptidoglycan synthesis at a given site [55] and limiting diffusion of intermediates.

Nevertheless, direct evidence of interactions of MreB with peptidoglycan synthesis enzymes is still lacking, mainly because most of these enzymes are transmembrane proteins difficult to work with. But recent studies have suggested that MreB could also recruit cytoplasmic actors, therefore opening up the way towards the structural characterization of a cytoplasmic part of the peptidoglycan machinery.

Peptidoglycan Biosynthesis

The pathway of peptidoglycan biosynthesis involves a series of enzymes. Among them, the Mur ligases participate to the formation of UDPMurNAc-peptide in the cytoplasm. Then, at the membrane level, a lipid career and an additional disaccharide are added by MraY and MurG enzymes, forming Lipid II, the basic block of peptidoglycan polymers. The spatial organization of these steps has not been elucidated yet. However, the bacterial actin homolog MreB seems to act as an anchor at the membrane surface for a multipartite complex which would restrict the diffusion of enzymes and intermediates, resulting in an optimized biosynthesis.

Chapter 3

Structural background

Direct and molecular evidence of the existence of the cytoplasmic peptidoglycan synthesis machinery is still lacking. Notably, no structure of any complex has been solved to date between either of the morphogenetic or murein synthesizing proteins. However, single actors of peptidoglycan biosynthetic machinery have been thoroughly analyzed from a structural point of view, as summarized in this chapter for Mur ligases, MurG, and MreB.

3.1 Crystal structures of Mur ligases

The crystal structures of all four of these similar ATP-dependent amino-acid ligases are known in different species in apo form or in complex with substrate, product, ADP, or inhibitors (see Table 3.1). Together with kinetics studies, these data allow a good understanding of their substrate specificity and their catalytic mechanisms [5] in which conformational changes might be important for activity.

Since all four Mur ligases share structural similarities, the structural description [5] [57] presented here is based on MurC from *Haemophilus influenzae* (PDB code 1P3D), for which structures of each of the apo, substrate-bound, and product-bound forms are known. Numbering of amino-acid residues thus refers to *H. influenzae* MurC, unless mentioned. Any relevant similarity or difference seen in the other three enzymes will be mentioned.

Mur ligases consist of three α/β-sheet domains formed from contiguous segments in the amino acid sequence: the N-terminal domain, the central domain, and the C-terminal domain (Figure 3.1).

3.1.1 N-terminal domain

Description

In all Mur ligases, the N-terminal domain is about 100 amino acids in length. This domain shows the highest degree of structural and sequence diversity among the three domains of Mur ligases and is mainly responsible for binding the growing peptidoglycan precursor [5].

The N-terminal domain of MurC from *H. influenzae* contains a common Rossmann-type

Protein	Species	Form	Resolution	Publication	PDB
MurC	*T. maritima*	apo	2.3 Å	Spraggon *et al.*, 2004	1J6U
MurC	*E. coli*	Mg^{2+}	2.5 Å	Deva *et al.*, 2006	2F00
MurC	*H. influenzae*	apo	3.1 Å	Skarzynski *et al.*, -	1GQQ
MurC	*H. influenzae*	product.AMPPNP.Mn^{2+}	1.7 Å	Mol *et al.*, 2003	1P3D
MurC	*H. influenzae*	apo.AMPPCP.Mg^{2+}	1.8 Å	Hu *et al.*, 2003	1GQY
MurC	*H. influenzae*	substrate.Mn^{2+}	1.85 Å	Mol *et al.*, 2003	1P31
MurC	*Y. pestis*	apo.AMP	2.25 Å	Halavaty *et al.*, -	4HV4
MurD	*E. coli*	apo	2.4 Å	Bertrand *et al.*, 2000	1E0D
MurD	*E. coli*	substrate	1.9 Å	Bertrand *et al.*, 2000	1EEH
MurD	*E. coli*	substrate.Mg^{2+}	1.7 Å	Bertrand *et al.*, 1999	2UAG
MurD	*E. coli*	substrate.ADP	1.52 Å	Kotnik *et al.*, 2007	2JFG
MurD	*E. coli*	inhibitor.aa	2.2 Å	Humljan *et al.*, 2008	2VTE
MurD	*S. agalactiae*	substrate.ADP	1.5 Å	Stein *et al.*, -	3LK7
MurE	*M. tuberculosis*	substrate.Mg^{2+}	3.0 Å	Basavannacharya *et al.*, 2010	2WTZ
MurE	*E. coli*	product.Cl^{-}	2 Å	Gordon *et al.*, 2001	1E8C
MurF	*E. coli*	apo	2.3 Å	Yan *et al.*, 2000	1GG4
MurF	*S. pneumoniae*	inhibitor	2.8 Å	Longenecker *et al.*, 2005	2AM1

Table 3.1: Main structures of Mur ligases, in apo form or in complex with substrate, product, ATP analog, and/or amino-acid. Structures solved in complex with inhibitors are not mentioned here, unless no other structure is known for this enzyme. aa: incoming amino acid (D-ala for MurD); ADP: adenosine diphosphate; ANP: phosphoaminophosphonic acid-adenylate ester; AMP: adenosine monophosphate; AMPPCP: β,γ-methyleneadenosine 5'-triphosphate; AMPPNP: Adenylyl-imidodiphosphate; MurC substrate: UDP-*N*-acetylmuramic acid; MurD substrate: UDP-*N*-acetylmuramoyl-L-alanine; MurE substrate: UDP-*N*-acetylmuramoyl-L-alanine-D-glutamate; MurE product: UDP-*N*-acetylmuramoyl-L-alanine-D-glutamate-*meso*-diaminopimelate.

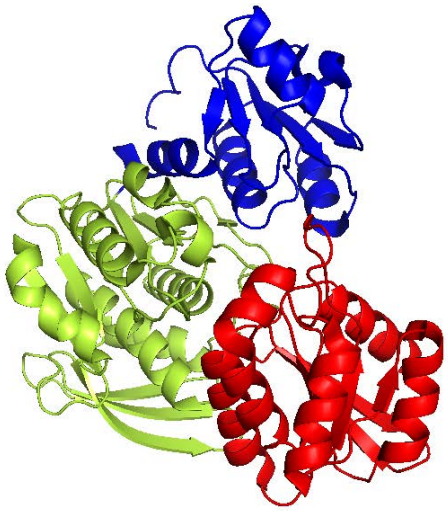


Figure 3.1: MurC from *H. influenzae*. Cartoon representation of MurC. Domain 1 (blue): residues 11-118; domain 2 (yellow): residues 119-324; domain 3 (red): residues 325-473. PDB entry: 1P3D.

α/β fold[1], consisting of a five-stranded parallel β-sheet flanked by four alternating α-helices [57]. This conserved topology is present in MurD as well. Indeed, the first two enzymes in the pathway share an essentially identical structure for domain 1 and both contain a variant of the GxGxxG fingerprint motif typical of dinucleotide binding domains [5] which is within a largely conserved region (Figure 3.2).

By contrast, although MurE and MurF also have an α/β fold for this domain, the topol-

[1]The Rossmann fold is a general nucleotide-binding structural motif. The most common structure with two repeats is composed of six parallel β-strands linked to two pairs of α-helices in the topological order $\beta-\alpha-\beta-\alpha-\beta$. Each Rossmann fold can bind one nucleotide [58].

ogy is much different and they display a mixed β-sheet [5] (Figure 3.3). In addition, no equivalent glycine-rich motif is found in the N-terminal domains of either of MurE and MurF enzymes.

```
          1         10        20        30        40        50        60
MurC_Ec   MNTQQLAKLRSIVPEMRRVRHIHFVGIGGAGMGGIAEVLANEGYQISGSDLAPNPVTQQL
MurC_Hi   .MKHSHEEIRKIIPEMRRVQQIHFIGIGGAGMSGIAEILLNEGYQISGSDIADGVVTQRL
MurC_Tm   ....................MKIHFVGIGGIGMSAVALHEFSNGNDVYGSNIETETERTAYL
```

Figure 3.2: Alignment of the 60 first residues of *E. coli* (Ec) MurC with *H. influenzae* (Hi) and *T. maritima* (Tm) MurCs. The consensus GxGxxG motif (black box) for dinucleotide binding domains is very well conserved among different species. In MurC, the motif is GIGGxGM.

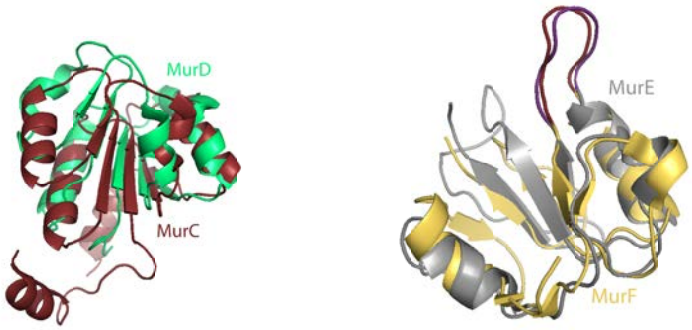


Figure 3.3: Domains 1 of Mur ligases from *E. coli* in apo forms. Left: superposition of the domains 1 of MurC (red) and MurD (green). Right: superposition of the domains 1 of MurE (grey) and MurF (yellow) from *E. coli*. In purple and red: the main loop involved in substrate binding. PDB codes: 2F00 (MurC), 1E0D (MurD), 18EC (MurE), 1GG4 (MurF).

Substrate binding by MurC and MurD.

The N-terminal domain binds the nucleotide moiety of the UDPMurNAc substrate, with a number of interactions [5] [57] (see Figure 3.4):

- The uracil ring is sandwiched between two hydrophobic loops (β2-α2 and β4-α4) which form a hydrophobic pocket where residues Ile 50 and Ile 87 pack the uridine.
- A conserved histidine residue (His 70) from the β3-α3 loop anchors the ring by hydrogen bonding.
- The glycine-rich loop (Gly 25 to Met 31) between β1 and α1 contacts the phosphate groups of the UDP. Together with a conserved serine residue (Ser84), this forms the diphosphate binding pocket.
- A conserved aspartate (Asp 49) from the C-terminus of strand β2 forms hydrogen bonds with the ribose hydroxyl groups.

Though the muramic acid moiety makes no interaction with the protein, the lactyl side-chain extends towards the catalytic centre and interacts with a Mg^{2+} ion [5].

The substrate binding site in MurD is very similar to that of MurC [5]. Nevertheless, MurD does not have the conserved histidine residue but instead uses a conserved threonine from its β2-α2 loop to anchor the uracil ring through a hydrogen bond [5] [59].

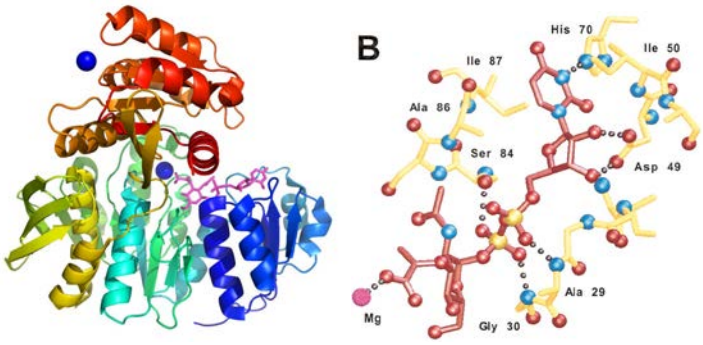


Figure 3.4: Interactions of *H. influenzae* MurC with its UDPMurNAc substrate. Left: overall view of MurC in complex with UDPMurNAc substrate (pink). Rainbow representation from N-terminus (blue) to C-terminus (red). Spheres: Mg^{2+}. Right: hydrogen-bonding interactions from domain 1 MurC residues (yellow) to the UDPMurNAc (red) are shown. From Mol *et al.*, 2003 [57].

Substrate binding by MurE and MurF

In contrast, MurE binds its substrate in a very different way from MurC and MurD [5] (see Figure 3.5). Indeed, a long loop between strand β2 and helix α2, extending towards domain 3, lies along the uridine group and makes hydrogen bonding interactions with the diphosphate moiety [5]. As in MurC, the uracil ring is tightly anchored, but the loop and residues involved are different (Tyr50 and β1-β2 loop) [5] (Figure 3.5).

While in MurC and MurD the substrate makes no interaction with domain 2, in MurE all of the other interactions with the UDP substrate are provided by the central domain [5].

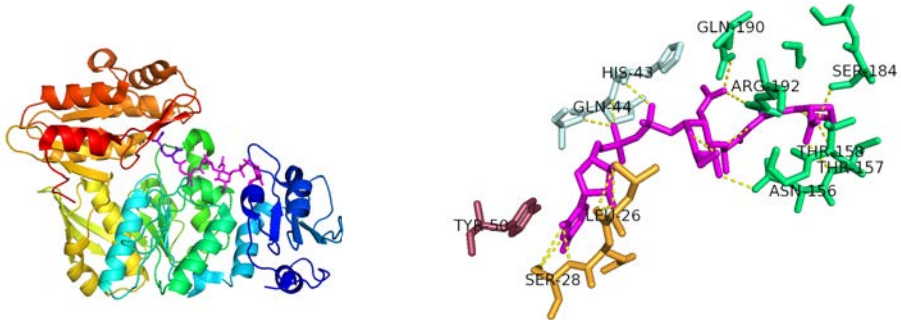


Figure 3.5: Interaction of *M. tuberculosis* MurE with its UDPMurNAc substrate. Hydrogen-bonding interactions from domain 1 MurE residues to the UDPMurNAc (pink) are shown. Blue: interacting residues from β2-α2 loop; yellow: interacting residues from β1-β2 loop; green: interacting residues from domain 2. PDB code: 2WTZ.

Although no structure of MurF in complex with its substrate is available yet, the superposition of domain 1 from apo MurF onto MurE (see Figure 3.3) shows that the same long loop is present and that the residues involved in UDP binding are essentially conserved, suggesting that a similar strategy is used for substrate binding by MurF [5] [1].

3.1.2 Central domain

Description

Domain 2 can be viewed as a structurally conserved "core" of the Mur ligases as illustrated by Figure 3.7 where each of the central domains of the four Mur ligases of *E. coli* are represented.

In MurC, this is the largest domain (extending from residues 119 to 324 in *H. influenzae* enzyme) and it presents a seven-stranded β-sheet surrounded by five α-helices; it is flanked by a smaller antiparallel, three-stranded β-sheet (Figure 3.6) [57]. This is a common fold seen in many ATP-binding proteins [60].

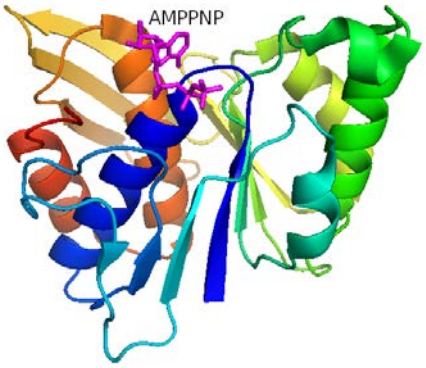


Figure 3.6: Central domain of MurC from *H. influenzae* in complex with AMPPNP. Cartoon representation from 1P3D entry of the PDB. N-terminus to C-terminus: Blue to red; Pink sticks: AMPPNP.

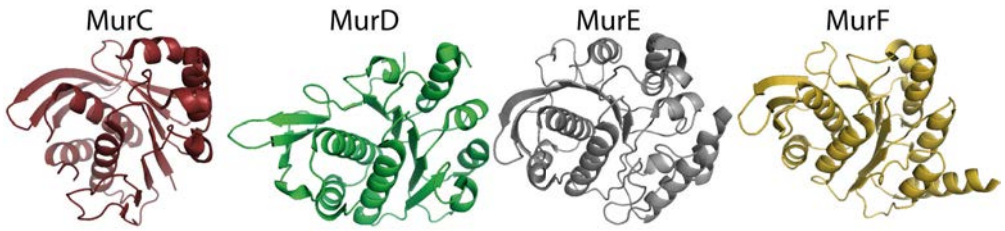


Figure 3.7: Comparison of domains 2 of MurC, MurD, MurE, and MurF from *E. coli*. Cartoon representation of domains 2 of Mur ligases. Red: MurC (2F00); Green: MurD (1E0D); Grey: MurE (1E8C); Yellow: MurF (1GG4).

ATP binding

As for all four ligases, the MurC ATP-binding site lies at the interface between the second and the third domains [57], with key interactions coming from the central domain, thereby often referred to as the ATP-binding domain. Domain 2 provides the following interactions for nucleotide binding:

- The triphosphate moiety is stabilized by hydrogen bonding interactions with the P-loop (residues 128 to 131), a glycine-rich loop which resembles the canonical mononucleotide

binding P-loop[2] [5].

- The adenine ring of AMPPNP inserts into a shallow pocket between the P-loop (mainly Gly 128) and the highly conserved His 291, forming hydrogen bonds involving the strictly conserved residue Asn 295 (see Figure 3.8) [57] [5].

In addition to direct interactions, Mg^{2+} ions were reported in Mur structures and seem required for ATP binding and/or catalysis, as in most ATPases [61]. In MurC and the other amide bond ligases, a conserved glutamate residue (Glu 173), adjacent to the P-loop, forms the basis of a first Mg^{2+} binding site [5], stabilizing the nucleotide interaction.

Domain 3 makes additional interactions with the nucleotide that will be described later.

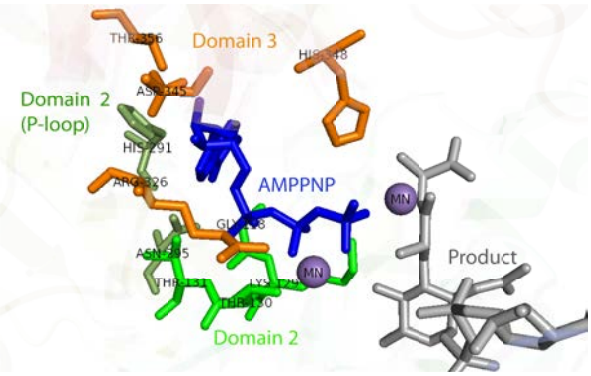


Figure 3.8: Interactions of MurC from *H. influenzae* with AMPPNP. Blue: AMPPNP; grey: UDP-*N*-acetylmuramoyl-L-alanine; Purple: Mn^{2+}; Orange: interacting residues from domain 3; Green: interacting residues from P-loop of domain 2; Light green: interacting residues from domain 2. PDB code 1P3D.

3.1.3 C-terminal domain

Description

The C-terminal domain, from residues 119 to 324 in *H. influenzae* MurC, consists of a six-stranded β-sheet with one antiparallel and five parallel β-strands, flanked by five α-helices. It contains a Rossmann dinucleotide-binding fold [57] (Figure 3.9). Although there are significant differences in the lengths and orientations of the α-helices and loops connecting the β-strands, this domain is quite well conserved among all of the Mur ligases [57].

Amino acid binding

The C-terminal domain is thought to bind the incoming amino acid in all the ligases [5]. Although the exact binding site for the amino acid has not yet been defined crystallographically for any of the ligases, it can be inferred from the structures of the product complexes [5], or from inhibitor-bound structures, especially from MurD [62].

In the structure of the product:MurC complex, the incoming alanine is positioned by two arginine residues (Arg 377 and Arg 380) of domain 3 that form hydrogen-bonds to the carboxylate atoms of the alanine (see Figure 3.10) [57]. Specificity for alanine comes from

[2]A P-loop (phosphate-binding loop), also known as Walker A motif is an ATP or GTP binding motif found in many nucleotide-binding proteins. It is a glycine-rich loop containing the pattern GXXXXGK(T/S) preceded by a β-strand and followed by an α-helix. It interacts with the phosphate groups of the nucleotide and with a magnesium ion, which coordinates the β- and γ-phosphates.

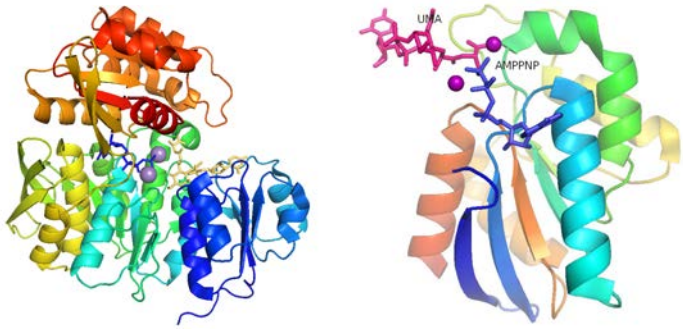


Figure 3.9: MurC from *H. influenzae* in complex with the product and AMPPNP. Left: Cartoon representation of MurC; from blue to red: N-terminus to C-terminus; stick representation in pink: UDP-*N*-acetylmuramoyl-L-alanine (product); stick representation in blue: AMPPNP; blue spheres: Mg^{2+}. Right: C-terminal domain, reoriented to better see the Rossmann-like fold; cartoon representation from blue to red: N-terminus to C-terminus; stick representation in pink: UDP-*N*-acetylmuramoyl-L-alanine (UMA); stick representation in blue: AMPPNP. Spheres: Mn^{2+} (PDB entry 1P3D).

the narrow and hydrophobic pocket between the side chains of His 348, His 376, Tyr 346, and Ala 459 [57].

ATP binding

The C-terminal domain furnishes additional contacts to the ribose sugar and α-phosphate group of the AMPPNP [57], most likely to orient and position the amino acid ligand with the growing peptidoglycan chain [57]. Indeed, the aspartate (Asp 345) and histidine (His 348) residues form hydrogen bonds to the ribose moiety and the γ-phosphate, respectively, and the conserved arginine (Arg 326) helps anchoring the α-phosphate group (see Figure 3.8) [5]. It is worthwhile mentioning that these residues are completely conserved in the four Mur ligases [5].

A second Mg^{2+} binding site has been identified in members of the Mur subfamily, situated in between the γ-phosphate of the ATP and the free carboxylate group of the substrate (see Figure 3.8) [5]. For instance, in the AMPPNP complex of *H. influenzae* MurC, this second cation binding site is formed by the side-chain of His 198, four water molecules and one of the γ-phosphate oxygen atoms [5].

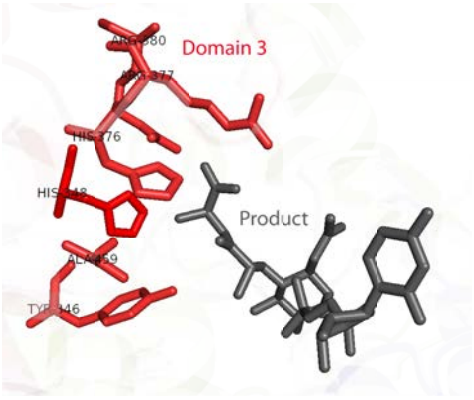


Figure 3.10: MurC interactions with the bound amino acid. Focus on the interaction of the L-Ala part of the product of MurC with the domain 3 of the ligase; grey: product; red: MurC interacting amino acid residues.

3.2 Conformational changes and other specificities of Mur ligases

3.2.1 Adaptation to a growing substrate

In the Mur ligase family, the individual reactions are identical. Indeed, for either of the ligases, a new amide bond is formed upon ATP activation. Moreover, each successive enzyme in the four-step pathway uses the product from the previous enzyme as substrate. That can explain why Mur ligases are related to each other from both a sequence and structural point of view.

Nevertheless, the elongating substrate imposes certain spatial restrictions on the positioning of its terminal carboxylate into the different enzymes. The Mur ligases have solved this problem first by orienting the substrate binding domain differently (between MurC and MurD), and then switching to a different substrate binding domain (for MurE and MurF). These adaptations allow the enzymes to retain a highly similar active site [5].

3.2.2 Mechanism and conformational change

All four ligases, which present structural similarities, have been suggested to share an identical reaction mechanism [1] which has been determined by several structural, kinetic, and simulation studies [59] [63].

ATP would be the first to bind [57] [64], and is believed to induce a conformational change from an open to a more closed state, resulting in activation of the enzyme [57]. Then, the substrate would enter the catalytic site where its C-terminal carboxylate would be phosphorylated by the γ-phosphate of ATP to form an acyl phosphate intermediate. This would be followed by nucleophilic attack by the amide group of the incoming amino acid resulting in the extension of the UDPMurNAc-peptide precursor, thus releasing an ADP and inorganic phosphate [65]. This mechanism suggests that structural flexibility and conformational changes are very important for the function of these enzymes.

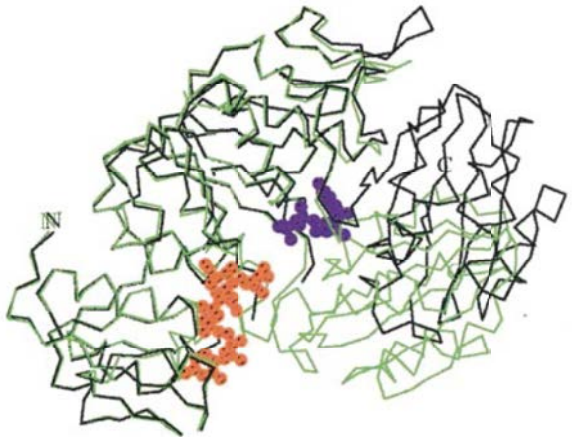


Figure 3.11: Superposition of the open-MurD in black and the more closed MurD.UMA.ADP.Mg2 complex in green. Representations show Cα atoms from the N-terminal and central domains. UMA and ADP from the MurD.UMA.ADP.Mg2 structure are shown in red and purple, respectively. From Bertrand *et al.*, 2000 [63].

3.2.3 Carbamoylation

In several Mur ligases structures, a carbamoylated lysine residue has been observed [5]. Carbamoylation is a post-translational modification of a lysine residue, which consists in the addition of a carbamoyl group to the amino-terminal of the lysine (Figure 3.12).

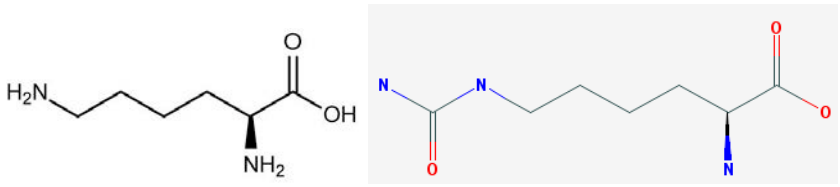


Figure 3.12: Lysine carbamoylation. Left: lysine; Right: carbamoylated lysine.

However, we still lack data to conclude about the conservation of carbamoylation of this lysine close to the active site. When present, the carbamoylated lysine residue seems to play a key role in stabilizing the second Mg^{2+} binding site (see Section 3.1.3) by interacting with one or two of the coordinating water molecules [5]. Moreover, mutagenesis and chemical rescue experiments suggested that this residue would be essential for the enzymatic function of MurD, MurE, and MurF [5] [66]. In MurC, where no lysine is present, this role is filled by a conserved glutamate residue (Glu 176) on the loop connecting β9 and β10 [5].

Crystal structures of Mur ligases

Mur ligases are composed of three structural domains which, individually, present a high similarity between the four ligases. The structures of the substrate and product complexes show that the ligases share a common active site which lies at the junction of the three structural domains and comprises specific binding pockets for the three substrates, UDPMurNAc, MgATP, and the incoming amino acid [5].
Lastly, crystal structures reveal different conformations, in an open or closed state believed to depend on both the nature and presence of ligands.

3.3 MurG

The crystal structures of the glycosyltransferase MurG (which couples a soluble donor sugar to the membrane anchored acceptor sugar, Lipid I, thus forming Lipid II) are known for *E. coli* and *Pseudomonas aeruginosa* in apo and/or substrate-bound forms (see Table 3.2), revealing key interactions involved in substrate binding and suggesting the membrane-association site of MurG.

3.3.1 Structure

E. coli MurG contains two α/β-sheet domains separated by a deep cleft [67] [68] (see Figure 3.13).

Species	Form	Resolution	Publication	PDB
E. coli	apo	1.9 Å	Ha *et al.*, 2000	1F0K
E. coli	donor substrate	2.5 Å	Hu *et al.*, 2003	1NLM
P. aeruginosa	donor substrate	2.23 Å	Brown *et al.*, 2012	3S2U

Table 3.2: MurG structures. Donor substrate = UDP-GlcNAc.

The first domain (residues 7-163 and 341-357 in *E. coli*) contains seven parallel β-strands and six α-helices [67]. The second domain (residues 164 to 340) contains six parallel β-strands and eight α-helices and presents a Rossmann-like fold [67]. Despite minimal sequence homology, both domains have high structural similarity [67].

The two domains are connected by an irregular bipartite helix (α-link) of the first domain linked to the first β-strand of the second domain [67].

There are three glycine rich loops in *E. coli* MurG that can be viewed as variants of the phosphate binding loops found in other dinucleotide binding proteins. Nevertheless, to date no structure in complex with nucleotide is available.

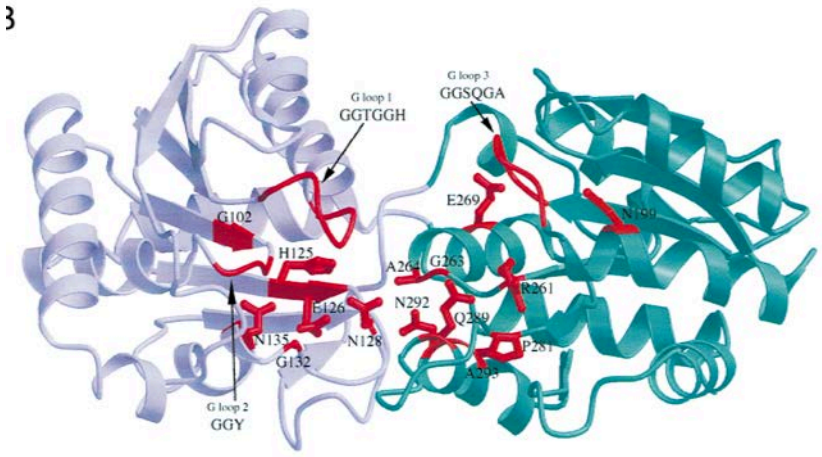


Figure 3.13: *E. coli* MurG. The Glycine-rich loops and other highly conserved residues (side chains are shown) are highlighted in red. From Ha *et al.*, 2000 [67].

3.3.2 Substrate binding

The two domains are similar in the presence and absence of substrate [68]. However, the relative orientation of the two domains is modified in the presence of UDP-GlcNAc mainly from a rigid body domain movement, so that MurG adopts a more closed conformation [68].

The UDP-GlcNAc moiety mostly contacts the second domain (see Figure 3.14, Right), mainly through a sequence motif made of prolines and glycines, several helices and loops located near the cleft between the two domains, and the GGS loop between β1 and α1, which undergoes conformational change upon substrate binding of the phosphate moiety [68]. This conformational change may be important for Lipid I binding, and may contribute directly to UDPGlcNAc binding.

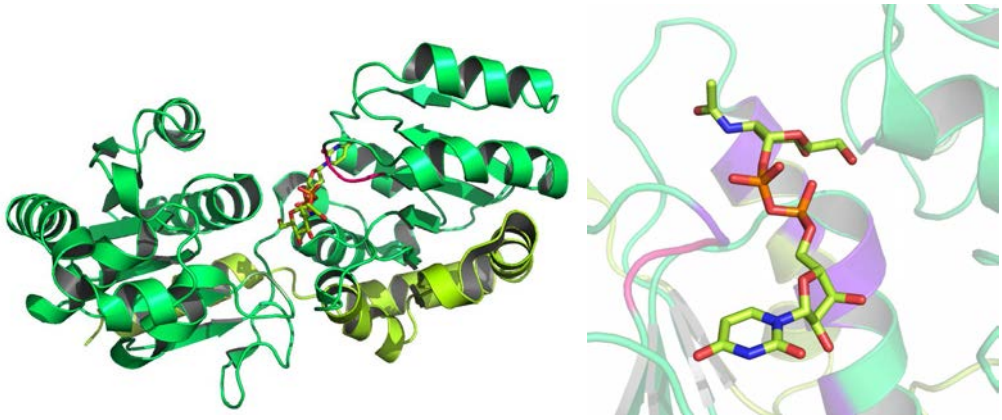


Figure 3.14: Structure of the MurG:UDP-GlcNAc complex. Left, cartoon representation showing the GGS motif displayed in pink and UDP-GlcNAc as sticks with C atoms in yellow, N atoms in blue, O atoms in red, and P atoms in orange. Light green indicates the 66 carboxy-terminal residues that have been suggested to play a role in membrane association [69]. Right, contacts between UDP-GlcNAc and MurG. Interacting amino acids 264, 265, 266, 269, 288, 289, 292, 192 (located in the invariant GGS loop), 128 (located in the invariant HEQN loop) are displayed in purple. From Hu *et al.*, 2003 [68].

3.3.3 Membrane-association site

E. coli MurG was shown to be associated with the cytoplasmic surface of bacterial membrane, which is in total accordance with its function of associating a soluble donor sugar to a lipid acceptor. Mohammadi *et al.* reported that the glycosyltransferase could be released from the inner membrane after treatment with detergents or trypsin digestion but not with salts, suggesting that the interaction of MurG with the cytoplasmic membrane is predominantly of hydrophobic nature [69]. Mohammadi and co-workers noticed that the carboxy-terminal 66 amino acid residues were essential for membrane localization, either because they interact with membrane components or because their absence affects the folding of MurG [69].

However, analysis of the *E. coli* MurG structure shows that there is a hydrophobic patch consisting of residues I75, L79, F82, W85, and W116 in the N-terminal domain, which is surrounded by basic residues K72, K140, K69, R80, R86, R89 [67] (see Figure 3.15). Ha *et al.* proposed that this is the membrane association site and that association involves both hydrophobic and electrostatic interactions with the negatively charged bacterial membrane. Therefore, the C-terminus of MurG could be rather important for proper folding as proposed by Mohammadi *et al.*.

MurG structure

MurG is a membrane-associated enzyme made of two similar domains, the active site lying within the cleft between the two domains. The region responsible for membrane binding is thought to be made of amino acids residues scattered within the N-terminal domain.

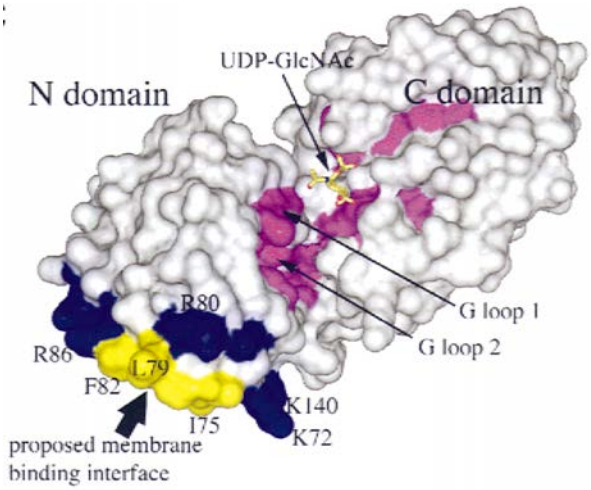


Figure 3.15: The surface of *E. coli* MurG. The G loops and other conserved residues in MurG are colored magenta. The proposed membrane binding interface is highlighted with hydrophobic residues in yellow and positively charged residues in blue. From Ha *et al.*, 2000 [67].

3.4 MreB

3.4.1 Crystal structure

The structure of MreB from *T. maritima* was solved in 2001 by Van den Ent and co-workers, showing the typical four-domain architecture of the actin family of proteins (see Figure 3.16). Thus, MreB consists of four subdomains IA, IB, IIA and IIB which correspond to subdomains 1, 2, 3 and 4 in actin. Despite weak amino-acid-sequence identity (about 15%), the three-dimensional structures of MreB and actin are remarkably similar [23].

The two larger subdomains (IA and IIA) have a common fold that comprises a five-stranded β-sheet surrounded by three α-helices, and are connected through helix H4. The smaller subdomains (IB and IIB) show exactly the same topology as actin.

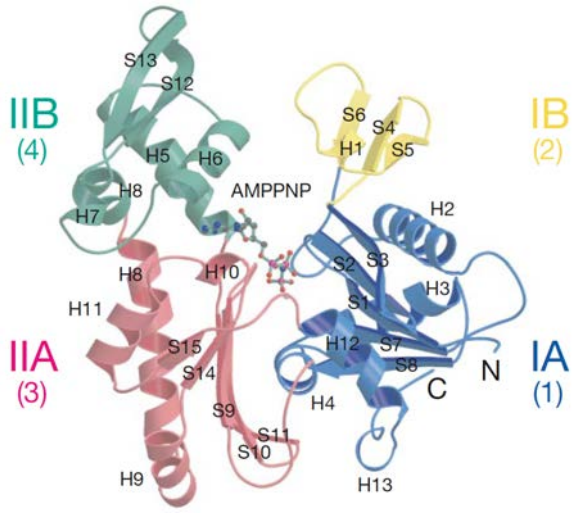


Figure 3.16: Structure of MreB in complex with AMPPNP. From van den Ent *et al.*, 2010 [35].

3.4.2 AMPPNP binding

Van den Ent *et al.* revealed that AMPPNP binds in a cleft between the main domains I and II of MreB (see Figure 3.17). Most of the residues involved in nucleotide binding appeared to be conserved or similar to that of the actin structure (PDB code 1YAG), with the exception of the residues located in subdomain IB. The authors proposed that this difference may be the result of a different nucleotide state of the two proteins. As in actin, a salt bridge can be seen between Glu 204 and Lys 49. The structures show a high-affinity divalent cation-binding site, with Mg^{2+} located near the base of the cleft between domains I and II as well.

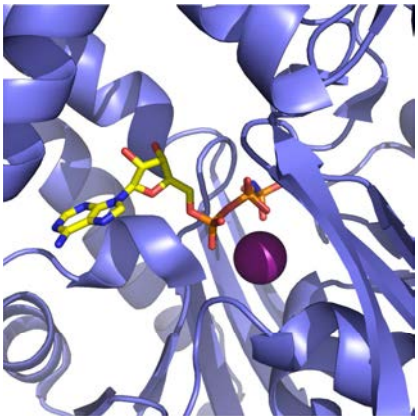


Figure 3.17: AMPPNP (sticks) and Mg^2+ (spheres) in the structure of MreB.

No conformational change was seen in crystal structures when comparing apo and nucleotide-bound forms of MreB.

3.4.3 Membrane-binding site

As mentioned in the previous chapter, MreB binds the cytoplasmic membranes through a membrane insertion loop in *T. maritima* MreB made of two hydrophobic residues Leu 93, Phe 94 (see Figure 3.19), and an N-terminal amphiphatic helix in *E. coli* MreB [34]. A schematic view of MreB membrane binding is presented Figure 3.18.

This feature of the bacterial cytoskeleton protein would make both MreB ends close to the membrane surface. That may explain the functionality issues which were previously reported for N- and C-terminal GFP fusions of MreB. Indeed, these fusion proteins could interfere with essential membrane binding [34].

3.4.4 MreB-RodZ complex

Figure 3.20 shows the crystal structure of MreB-RodZ, revealing a tiny surface of interaction (552 $Å^2$), in which not more than three amino acid residues would be involved.

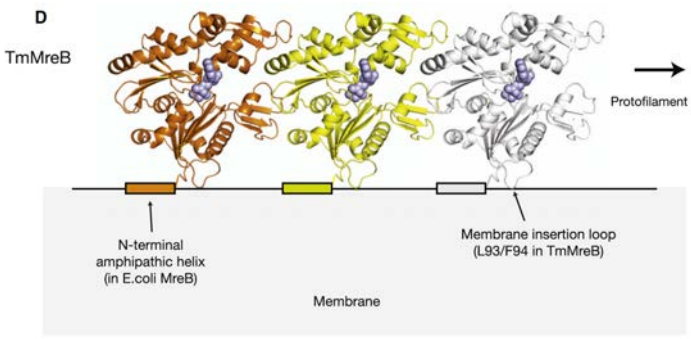


Figure 3.18: Model for the interaction of an MreB filament with the membrane. The membrane insertion loop required for *T. maritima* MreB is shown, as well as the N-terminal amphipathic helix from *E. coli* MreB. From Salje *et al.* [34].

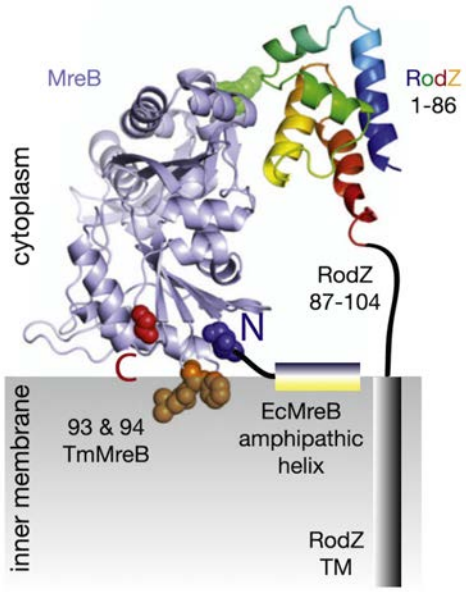


Figure 3.19: Schematic diagram showing the position of MreB on the membrane and indicating the residues involved in membrane binding. RodZ is represented (van den Ent *et. al*, 2010 [35]). The membrane insertion loop and *E. coli* MreB amphipathic helix are shown. The close positions of the N- and C- termini to the membrane may explain why N- or C-terminal fusion proteins were nonfunctional. From Salje *et al.*, 2011 [34].

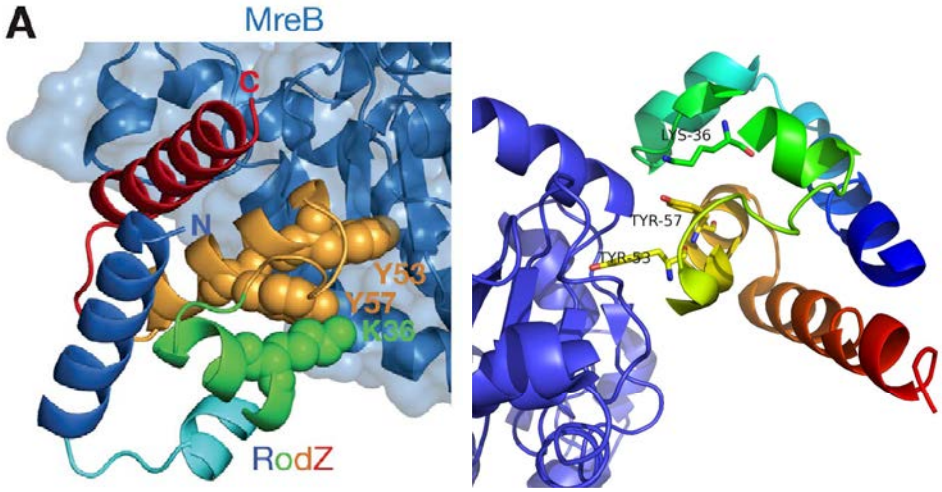


Figure 3.20: Two views of RodZ's interaction with MreB (blue). Residues Lys 36, Tyr 53 and Tyr 57, shown in spheres, are required for MreB interaction. From van den Ent *et al.*, 2010 [35].

Structure of MreB

MreB displays an actin-like fold in four sub-domains. It binds AMPPNP within the cleft between the two main domains, and binds to cell membranes. Strikingly, its interaction with RodZ involves very few amino acid residues.50

Chapter 4

Goals of the project

The goals of this work were first to detect and characterize the interactions between cytoplasmic actors of peptidoglycan biosynthesis, and second to isolate and crystallize a complex involving Mur enzymes, the ultimate objective being to solve the crystal structure such a complex.

Mur ligases, which are soluble proteins encoded within a unique operon in most bacterial strains according to the SMART [70] [71] software, seemed to be good candidates for such an investigation.

With the hope that such a Mur complex from a thermophilic organism would be more stable than its counterpart in *E. coli* or a pathogenic bacterium, *T. maritima* was chosen as model *in vitro*. This choice allowed to easily add MreB in our study, the bacterial actin homolog reported to play an organizational role in peptidoglycan synthesis. Indeed, MreB1 from *T. maritima* was the only MreB homolog reported to be tolerant to *in vitro* conditions of laboratories for crystallization purposes.

Besides, as MurG had been suggested to interact with MreB and Mur ligases in several publications [69] [53], it was added in the study as well.

Therefore, expression and purification protocols for Mur ligases and MurG from *T. maritima* had to be set up first. Indeed, except for MurC [72], none of the ligases had been previously purified, nor the glycosyltransferase. The investigation of Mur interactions was then assessed by both the single and polycistronic cloning possibilities, underlining aim for all experiments being the crystallization of a complex.

Bibliography

[1] Youwei Yan, Sanjeev Munshi, Barbara Leiting, Matt S. Anderson, John Chrzas, and Zhongguo Chen. Crystal structure of *Escherichia coli* UDPMur*N*Ac-tripeptide D-alanyl-D-alanine-adding enzyme (MurF) at 2.3 Å resolution. *J. Mol. Biol.*, 304(3):435 – 445, 2000.

[2] Ahmed El Zoeiby, François Sanschagrin, and Roger C. Levesque. Structure and function of the Mur enzymes: development of novel inhibitors. *Mol. Microbiol.*, 47(1):1–12, 2003.

[3] Walsh CT. Bugg TD. Intracellular steps of bacterial cell wall peptidoglycan biosynthesis: enzymology, antibiotics, and antibiotic resistance. *Nat. Prod. Rep.*, 9(3):199–215, 1992.

[4] Andrew L. Lovering, Susan S. Safadi, and Natalie C. J. Strynadka. Structural perspective of peptidoglycan biosynthesis and assembly. *Annu. Rev. Biochem.*, 81:451–478, 2012.

[5] Clyde A. Smith. Structure, function and dynamics in the Mur family of bacterial cell wall ligases. *J. Mol. Biol.*, 362(4):640–655, 2006.

[6] M.T Madigan, J.M. Martinko, and J. Parker. *Brock Biology of Microorganisms*. 2003 Brock Biology of Microorganisms. 10th ed. Pearson Education, Inc. Upper Saddle River, NJ.

[7] TJ Beveridge. Use of the Gram stain in microbiology. *Biotechnic & Histochemistry*, 76(3):111–118, 2001.

[8] Valrio R. Matias and Terry J. Beveridge. Native cell wall organization shown by cryo-Electronlectron microscopy confirms the existence of a periplasmic space in Staphylococcus aureus. *J. Bacteriol.*, 188(3):1011–1021, 2006.

[9] Mohammad J. Hajipour, Katharina M. Fromm, Ali Akbar Ashkarran, Dorleta Jimenez de Aberasturi, Idoia Ruiz de Larramendi, Teofilo Rojo, Vahid Serpooshan, Wolfgang J. Parak, and Morteza Mahmoudi. Antibacterial properties of nanoparticles. *Trends Biotechnol.*, 30:499–511, 2012.

[10] Waldemar Vollmer and Stephen J. Seligman. Architecture of peptidoglycan: more data and more models. *Trends Microbiol.*, 18(2):59 – 66, 2010.

[11] Waldemar Vollmer, Didier Blanot, and Miguel A. De Pedro. Peptidoglycan structure and architecture. *FEMS Microbiol. Rev.*, 32(2):149–167, 2008.

[12] Ahmed El Zoeiby, François Sanschagrin, Pierre C. Havugimana, Alain Garnier, and Roger C. Levesque. *In vitro* reconstruction of the biosynthetic pathway of peptidoglycan cytoplasmic precursor in *Pseudomonas aeruginosa*. *FEMS Microbiol. Lett.*, 201(2):229 – 235, 2001.

[13] Carolyn R. Stenbak, Ji-Hwan Ryu, François Leulier, Sebastien Pili-Floury, Claudine Parquet, Mireille Hervé, Catherine Chaput, Ivo G. Boneca, Won-Jae Lee, Bruno

Lemaitre, and Dominique Mengin-Lecreulx. Peptidoglycan molecular requirements allowing detection by the drosophila immune deficiency pathway. *J. Immunol.*, 173(12):7339–7348, 2004.

[14] Lu Gan, Songye Chen, and Grant J. Jensen. Molecular organization of Gram-negative peptidoglycan. *Proc Natl Acad Sci U S A.*, 105(48):18953–18957, 2008.

[15] Waldemar Vollmer and Joachim-Volker Höltje. The Architecture of the Murein (Peptidoglycan) in Gram-Negative Bacteria: Vertical Scaffold or Horizontal Layer(s)? *J. Bacteriol.*, 186(18):5978–5987, 2004.

[16] Audrey Boniface, Ahmed Bouhss, Dominique Mengin-Lecreulx, and Didier Blanot. The MurE synthetase from *Thermotoga maritima* is endowed with an unusual D-lysine adding activity. *J. Biol. Chem.*, 281(23):15680–15686, 2006.

[17] Audrey Boniface, Claudine Parquet, Michel Arthur, Dominique Mengin-Lecreulx, and Didier Blanot. The elucidation of the structure of *Thermotoga maritima* peptidoglycan reveals two novel types of cross-link. *J. Biol. Chem.*, 284(33):21856–21862, 2009.

[18] Matti Pierre-Jean, Neves David, and Dessen Andra. Bridging cell wall biosynthesis and bacterial morphogenesis. *Curr: Opin. Struc. Biol.*, 20(6):749 – 755, 2010.

[19] International Union of Biochemistry and Molecular Biology. UDP-N-acetylglucosamine biosynthesis, 2002. `http://www.chem.qmul.ac.uk/iubmb/enzyme/reaction/polysacc/UDPGlcN.html`.

[20] Ahmed Bouhss, Dominique Mengin-Lecreulx, Dominique Le Beller, and Jean Van Heijenoort. Topological analysis of the mray protein catalysing the first membrane step of peptidoglycan synthesis. *Molecular Microbiology*, 34(3):576–585, 1999.

[21] Athanasios Typas, Manuel Banzhaf, Carol A. Gross, and Waldemar Vollmer. From the regulation of peptidoglycan synthesis to bacterial growth and morphology. *Nat Rev Micro*, 10(2):123–136, February 2012.

[22] Fusinita van den Ent, Linda A. Amos, and Jan Lowe. Prokaryotic origin of the actin cytoskeleton. *Nature*, 413(6851):39–44, 2001.

[23] Rut Carballido-Lopez. Orchestrating bacterial cell morphogenesis. *Mol. Microbiol.*, 60(4):815–819, 2006.

[24] Felipe O. Bendezu and Piet A. J. de Boer. Conditional lethality, division defects, membrane involution, and endocytosis in *mre* and *mrd* shape mutants of *Escherichia coli*. *J. Bacteriol.*, 190(5):1792–1811, 2008.

[25] Alex Formstone and Jeffery Errington. A magnesium-dependent *mreB* null mutant: implications for the role of *mreB* in *Bacillus subtilis*. *Mol. Microbiol.*, 55(6):1646–1657, 2005.

[26] David Popp, Akihiro Narita, Kayo Maeda, Tetsuro Fujisawa, Umesh Ghoshdastider, Mitsusada Iwasa, Yuichiro Mada, and Robert C. Robinson. Filament structure, organization, and dynamics in MreB sheets. *J. Biol. Chem.*, 285(21):15858–15865, 2010.

[27] Osigwe Esue, Maria Cordero, Denis Wirtz, and Yiider Tseng. The assembly of MreB, a prokaryotic homolog of actin. *J. Biol. Chem.*, 280(4):2628–2635, 2005.

[28] Osigwe Esue, Denis Wirtz, and Yiider Tseng. GTPase activity, structure, and mechanical properties of filaments assembled from bacterial cytoskeleton protein MreB. *J. Bacteriol.*, 188(3):968–976, 2006.

[29] Laura J.F. Jones, Rut Carballido-Lopez, and Jeffery Errington. Control of cell shape in bacteria: helical, actin-like filaments in *Bacillus subtilis*. *Cell*, 104(6):913 – 922, 2001.

[30] Ahmed Gaballah, Anna Kloeckner, Christian Otten, Hans-Georg Sahl, and Beate Henrichfreise. Functional analysis of the cytoskeleton protein MreB from *Chlamydophila pneumoniae*. *PLoS ONE*, 6(10):e25129, 10 2011.

[31] Pearl Nurse and Kenneth J. Marians. Purification and characterization of *Escherichia coli* mreb protein. *J. Biol. Chem.*, 288(5):3469–3475, 2013.

[32] Joshua A. Mayer and Kurt J. Amann. Assembly properties of the *Bacillus subtilis* actin, MreB. *Cell Motil. Cytoskel.*, 66(2):109–118, 2009.

[33] Greg J. Bean and Kurt J. Amann. Polymerization properties of the *Thermotoga maritima* actin MreB: roles of temperature, nucleotides, and ions. *Biochem.*, 47(2):826–835, 2008.

[34] Jeanne Salje, Fusinita van den Ent, Piet de Boer, and Jan Löwe. Direct membrane binding by bacterial actin MreB. *Mol. Cell*, 43(3):478–487, 2011.

[35] Fusinita van den Ent, Christopher M Johnson, Logan Persons, Piet de Boer, and Jan Lowe. Bacterial actin MreB assembles in complex with cell shape protein RodZ. *EMBO J*, 29(6):1081–1090, 2010.

[36] Rainer M. Figge, Arun V. Divakaruni, and James W. Gober. MreB, the cell shape-determining bacterial actin homologue, co-ordinates cell wall morphogenesis in *Caulobacter crescentus*. *Mol. Microbiol.*, 51(5):1321–1332, 2004.

[37] S. Anisah Alyahya, Roger Alexander, Teresa Costa, Adriano O. Henriques, Thierry Emonet, and Christine Jacobs-Wagner. RodZ, a component of the bacterial core morphogenic apparatus. *Proc Natl Acad Sci U S A.*, 106(4):1239–1244, 2009.

[38] Felipe O Bendezu, Cynthia A Hale, Thomas G Bernhardt, and Piet A J de Boer. RodZ (YfgA) is required for proper assembly of the MreB actin cytoskeleton and cell shape in *E. coli*. *EMBO J*, 28(3):193–204, February 2009.

[39] Emma Tabe Eko Niba, Gaochi Li, Kenji Aoki, and Madoka Kitakawa. Characterization of *rodZ* mutants: RodZ is not absolutely required for the cell shape and motility. *FEMS Microbiology Letters*, 309(1):35–42, 2010.

[40] P Canepari M M Lleo and G Satta. Bacterial cell shape regulation: testing of additional predictions unique to the two-competing-sites model for peptidoglycan assembly and isolation of conditional rod-shaped mutants from some wild-type cocci. *J Bacteriol.*, 172(7):3758771, 1990.

[41] Yu-Ling Shih and Lawrence Rothfield. The bacterial cytoskeleton. *Microbiol. Mol. Biol. Rev.*, 70(3):729–754, 2006.

[42] Peter M. Slovak, George H. Wadhams, and Judith P. Armitage. Localization of MreB in *Rhodobacter sphaeroides* under conditions causing changes in cell shape and membrane structure. *J. Bacteriol.*, 187(1):54–64, 2005.

[43] Purva Vats, Yu-Ling Shih, and Lawrence Rothfield. Assembly of the MreB-associated cytoskeletal ring of *Escherichia coli*. *Mol. Microbiol.*, 72(1):170–182, 2009.

[44] Kimberly N. Cowles and Zemer Gitai. Surface association and the MreB cytoskeleton regulate pilus production, localization and function in *Pseudomonas aeruginosa*. *Mol. Microbiol.*, 76(6):1411–1426, 2010.

[45] Zemer Gitai, Natalie Dye, and Lucy Shapiro. An actin-like gene can determine cell polarity in bacteria. *Proc Natl Acad Sci U S A.*, 101(23):8643–8648, 2004.

[46] Matthew T. Swulius and Grant J. Jensen. The helical MreB cytoskeleton in *Escherichia coli* MC1000/pLE7 is an artifact of the N-terminal yellow fluorescent protein tag. *J Bacteriol.*, 194(23):6382–6386, 2012.

[47] Courtney L. White and James W. Gober. MreB: pilot or passenger of cell wall synthesis? *Trends in Microbiology*, 20(2):74 – 79, 2012.

[48] Julia Dominguez-Escobar, Arnaud Chastanet, Alvaro H. Crevenna, Vincent Fromion, Roland Wedlich-Sldner, and Rut Carballido-Lopez. Processive movement of MreB-associated cell wall biosynthetic complexes in bacteria. *Science*, 333 (6039):225–228, 2011.

[49] Matthew T. Swulius, Songye Chen, H. Jane Ding, Zhuo Li, Ariane Briegel, Martin Pilhofer, Elitza I. Tocheva, Suzanne R. Lybarger, Tanya L. Johnson, Maria Sandkvist, and Grant J. Jensen. Long helical filaments are not seen encircling cells in electron cryotomograms of rod-shaped bacteria. *Biochemical and Biophysical Research Communications*, 407(4):650 – 655, 2011.

[50] Constantin N. Takacs, Sebastian Poggio, Godefroid Charbon, Mathieu Pucheault, Waldemar Vollmer, and Christine Jacobs-Wagner. MreB drives de novo rod morphogenesis in *Caulobacter crescentus* via remodeling of the cell wall. *J. Bacteriol.*, 192(6):1671–1684, 2010.

[51] Thomas Kruse, Jette Bork-Jensen, and Kenn Gerdes. The morphogenetic MreBCD proteins of *Escherichia coli* form an essential membrane-bound complex. *Mol. Microbiol.*, 55(1):78–89, 2005.

[52] Joshua W. Shaevitz and Zemer Gitai. The structure and function of bacterial actin homologs. *Perspect. Biol.*, 2(9), 2010.

[53] Courtney L. White, Aleksandar Kitich, and James W. Gober. Positioning cell wall synthetic complexes by the bacterial morphogenetic proteins MreB and MreD. *Mol. Microbiol.*, 76(3):616–633, 2010.

[54] Arun V. Divakaruni, Cyril Baida, Courtney L. White, and James W. Gober. The cell shape proteins MreB and MreC control cell morphogenesis by positioning cell wall synthetic complexes. *Mol. Microbiol.*, 66(1):174–188, 2007.

[55] Andrew Jermy. Bacterial physiology: MreB takes a back seat. *Nat. Rev. Micro.*, 9(8):560–561, August 2011.

[56] Ethan C. Garner, Remi Bernard, Wenqin Wang, Xiaowei Zhuang, David Z. Rudner, and Tim Mitchison. Coupled, circumferential motions of the cell wall synthesis machinery and MreB filaments in *B. subtilis*. *Science*, 333(6039):222–225, 2011.

[57] Clifford D. Mol, Alexei Brooun, Douglas R. Dougan, Mark T. Hilgers, Leslie W. Tari, Robert A. Wijnands, Mark W. Knuth, Duncan E. McRee, and Ronald V. Swanson. Crystal structures of active fully assembled substrate- and product-bound complexes of UDP-N-acetylmuramic acid:L-alanine ligase (MurC) from *Haemophilus influenzae*. *J. Bacteriol.*, 185(14):4152–4162, 2003.

[58] C. Ramakrishnan, V.S. Dani, and T. Ramasarma. A conformational analysis of Walker motif A [GXXXXGKT (S)] in nucleotide-binding and other proteins. *Protein Eng.*, 15(10):783–798, 2002.

[59] Jay A. Bertrand, Genevi Auger, Lydie Martin, Eric Fanchon, Didier Blanot, Dominique Le Beller, Jean van Heijenoort, and Otto Dideberg. Determination of the MurD mechanism through crystallographic analysis of enzyme complexes. *J. Mol. Biol.*, 289(3):579–590, 1999.

[60] Konstantin A. Denessiouk and Mark S. Johnson. When fold is not important: A common structural framework for adenine and AMP binding in 12 unrelated protein families. *Proteins*, 38(3):310–326, 2000.

[61] Tulika Munshi, Antima Gupta, Dimitrios Evangelopoulos, Juan David Guzman, Simon Gibbons, Nicholas H. Keep, and Sanjib Bhakta. Characterisation of atp-dependent mur ligases involved in the biogenesis of cell wall peptidoglycan in <italic>mycobacterium tuberculosis</italic>. *PLoS ONE*, 8(3):e60143, 03 2013.

[62] Jan Humljan, Miha Kotnik, Carlos Contreras-Martel, Didier Blanot, Uros Urleb, Andrea Dessen, Tom Solmajer, and Stanislav Gobec. Novel naphthalene-N-sulfonyl-d-glutamic acid derivatives as inhibitors of MurD, a key peptidoglycan biosynthesis enzyme. *J Med Chem.*, 51(23):7486–7494, 2008.

[63] Jay A. Bertrand, Eric Fanchon, Lydie Martin, Laurent Chantalat, Genevi...ve Auger, Didier Blanot, Jean van Heijenoort, and Otto Dideberg. Open structures of MurD: domain movements and structural similarities with folylpolyglutamate synthetase. *J. Mol. Biol.*, 301(5):1257 – 1266, 2000.

[64] Andrej Perdih, Miha Kotnik, Milan Hodoscek, and Tom Solmajer. Targeted molecular dynamics simulation studies of binding and conformational changes in *E. coli* MurD. *Proteins: Structure, Function, and Bioinformatics*, 68(1):243–254, 2007.

[65] S Vaganay, ME Tanner, J van Heijenoort, and D. Blanot. Study of the reaction mechanism of the D-glutamic acid-adding enzyme from *Escherichia coli*. *Microb Drug Resist.*, 2(1):51–4., 1996.

[66] Sébastien Dementin, Ahmed Bouhss, Geneviève Auger, Claudine Parquet, Dominique Mengin-Lecreulx, Otto Dideberg, Jean van Heijenoort, and Didier Blanot. Evidence of a functional requirement for a carbamoylated lysine residue in MurD, MurE and MurF synthetases as established by chemical rescue experiments. *European Journal of Biochemistry*, 268(22):5800–5807, 2001.

[67] Sha Ha, Deborah Walker, Yigong Shi, and Suzanne Walker. The 1.9 Å crystal structure of *Escherichia coli* MurG, a membrane-associated glycosyltransferase involved in peptidoglycan biosynthesis. *Protein Sci.*, 9(6):1045–1052, 2000.

[68] Yanan Hu, Lan Chen, Sha Ha, Ben Gross, Brian Falcone, Deborah Walker, Maryam Mokhtarzadeh, and Suzanne Walker. Crystal structure of the MurG:UDP-GlcNAc complex reveals common structural principles of a superfamily of glycosyltransferases. *Proc Natl Acad Sci U S A.*, 100(3):845–849, 2003.

[69] Tamimount Mohammadi, Aneta Karczmarek, Muriel Crouvoisier, Ahmed Bouhss, Dominique Mengin-Lecreulx, and Tanneke Den Blaauwen. The essential peptidoglycan glycosyltransferase MurG forms a complex with proteins involved in lateral envelope growth as well as with proteins involved in cell division in *Escherichia coli*. *Mol. Microbiol.*, 65(4):1106–1121, 2007.

[70] Jörg Schultz, Frank Milpetz, Peer Bork, and Chris P. Ponting. SMART, a simple modular architecture research tool: Identification of signaling domains. *Proc Nat Aca Sci*, 95(11):5857–5864, 1998.

[71] Ivica Letunic, Tobias Doerks, and Peer Bork. SMART 7: recent updates to the protein domain annotation resource. *Nucleic Acids Research*, 40(D1):D302–D305, 2012.

[72] Glen Spraggon, Robert Schwarzenbacher, Andreas Kreusch, Christian C. Lee, Polat Abdubek, Eileen Ambing, Tanya Biorac, Linda S. Brinen, Jaume M. Canaves, Jamison Cambell, Hsiu-Ju Chiu, Xiaoping Dai, Ashley M. Deacon, Mike DiDonato, Marc-André Elsliger, Said Eshagi, Ross Floyd, Adam Godzik, Carina Grittini, Slawomir K. Grzechnik, Eric Hampton, Lukasz Jaroszewski, Cathy Karlak, Heath E. Klock, Eric Koesema, John S. Kovarik, Peter Kuhn, Inna Levin, Daniel McMullan, Timothy M. McPhillips, Mitchell D. Miller, Andrew Morse, Kin Moy, Jie Ouyang, Rebecca Page, Kevin Quijano,

Alyssa Robb, Raymond C. Stevens, Henry van den Bedem, Jeff Velasquez, Juli Vincent, Frank von Delft, Xianhong Wang, Bill West, Guenter Wolf, Qingping Xu, Keith O. Hodgson, John Wooley, Scott A. Lesley, and Ian A. Wilson. Crystal structure of an UDP-*N*-acetylmuramate-alanine ligase MurC (TM0231) from *Thermotoga maritima* at 2.3 Å resolution. *Proteins: Structure, Function, and Bioinformatics*, 55(4):1078–1081, 2004.

Part II

Materials and Methods

Chapter 5

Cloning

5.1 Cloning strategy

5.1.1 Single genes

Single genes from *T. maritima* were cloned in various expression vectors in order to be able to co-express several proteins. Genes were fused to different tags to permit pull-down experiments. Table 5.1 summarizes the vectors and the cloning sites that were used.

Gene	Expression Vector	Restriction sites	Resistance	Tag
murC	pASK-IBA3C-tev	BsaI / BsaI	Chloramphenicol	C-ter Strep, Tev cleavable
murD	pET15b	NdeI / XhoI	Ampicillin	N-ter His, Thrombin-cleavable
murD	pETDuetTev	NdeI / XhoI	Ampicillin	No tag
murE	pET30b	NdeI / XhoI	Kanamycin	C-ter His
murF	pET30b	NdeI / NotI	Kanamycin	C-ter His
murG	pASK-IBA3C-tev	BsaI / BsaI	Chloramphenicol	Strep, tev cleavable
mreB1	pET30b	NdeI / EcoRI	Kanamycin	C-ter His

Table 5.1: Cloning of single genes. Genes from *T. maritima* were cloned in the indicated vectors. Tags from indicated vectors were used, except for *mreB1* that was cloned as described in van den Ent *et al.* [1] with a two-residue linker between the C-terminus and the His tag. AflII, BamHI, BsaI, EcoRI-HF, NdeI, SpeI, XhoI were provided by NEB; BamHI, NcoI, XbaI were purchased from Fermentas. For more information about the vectors, see the corresponding section below. His tag is made of six histidine residues which are added at the end of the recombinant protein - either N- or C- terminus. Strep tag consists of the sequence WSHPQFEK added at the N- or C-terminus of the protein.

Detailed information about expression systems and vectors used is available in the Appendix.

5.1.2 Polycystronic forms

In order to investigate the potential role of the polycistronic forms in the formation of protein complexes, the genes *murE* and *murF* were cloned as a polycistronic form:

- Into the second site of a pETDuet vector modified by André Zapun from the Pneumococcus Group of IBS, referred to as pETDuetLIM1 (ampicillin resistance; Tev-cleavable N-ter His-tag (8 histidine residues); restriction sites: NcoI / BamHI)
- into the second site of the modified pETDuet vector containing *murD* (ampicillin resistance; Tev-cleavable N-ter His-tag on MurE; no tag on MurD; restriction sites: BamHI / AflII)

5.2 Cloning procedure

5.2.1 PCR

Genomic DNAs from *T. maritima* DSM3109 strain were used as templates for Polymerase Chain Reactions (PCR). PCR were performed according to protocols from polymerases suppliers.

50 μl reaction mixes contained the recommanded amount of buffers (GC rich buffer for Phusion Hot Start II polymerase) already supplemented with $MgCl_2$, 1 μl of genomic DNA, 2 μl of each of the sense and antisense primers at 10 pmol/μl, 1 μl of dNTPs at 10 mM, 1 μl of Pfu polymerase from Fermentas or 1 μl of Phusion Hot Start II polymerase from Finzymes.

PCR was performed in a thermocycler from Biometra® as described in Table 5.2.

Gene	**Polymerase**	**First step**	**Cycling: x3**	**Cycling: x27**	**End**
murC	Hot Start	98°C, 1 min	98°C, 30 s 53°C 30 s 72°C, 3 min	98°C, 30 s 61°C, 30s 72°C, 3 min	72 °C, 10 min
murD	Pfu	95°C, 1 min	95°C, 30 s 54°C 30 s 72°C, 5 min 30s	95°C, 30 s 59°C, 30 s 72°C, 5 min 30s	72 °C, 10 min
murE	Hot Start	98°C, 1 min	98°C, 30 s 53°C 30 s 72°C, 3 min	98°C, 30 s 61°C, 30 s 72°C, 3 min	72 °C, 10 min
murF	Hot Start	98°C, 1 min	98°C, 30 s 53°C 30 s 72°C, 3 min	98°C, 30 s 61°C, 30 s 72°C, 3 min	72 °C, 10 min
murE-murF	Pfu	95°C, 1 min	95°C, 30 s 54°C 30 s 72°C, 5 min 30 s	95°C, 30 s 59°C, 30s 72°C, 5 min 30 s	72 °C, 10 min
murG	Hot Start	98°C, 1 min	98°C, 30 s 53°C 30 s 72°C, 3 min	98°C, 30 s 61°C, 30 s 72°C, 3 min	72 °C, 10 min
mreB1	Pfu	98°C, 2 min	98°C, 30 s 61°C 30 s 72°C, 2 min	98°C, 30 s 65°C, 30 s 72°C, 2 min	72 °C, 10 min

Table 5.2: Polymerase Chain Reactions for amplifications of *murC*, *murD*, *murE*, *murF*, *murG*, *murE-murF*, and *mreB1* from *T. maritima*.

Amplifications of *murE* and *murF* for cloning into pETDuetLIM1 vector followed similar protocols.

5.2.2 Cloning into pCRBlunt

The cloning reaction was performed as recommanded by the Zero Blunt® TOPO® PCR Cloning Kit manual.

DNAs were chemically transformed into NEB5α or JM109 cells: 5 μl of DNA were added to 200 μl of home-made competent bacteria and incubated on ice for 30 min. A heat shock was done at 42°C for 35 s and bacteria were subsequently chilled on ice for 2-5 min. 800 μl of SOC or LB medium were added and bacteria were incubated for 1h at 37°C with 200 rpm

shaking without antibiotics. Cells were then spread onto LB-Agar Petri dishes supplemented with 40 μg/ml kanamycin.

For colonies screening, five to ten 15-ml tubes with 3 ml of LB supplemented with 40 μg/ml kanamycin were inoculated with isolated colonies and grown at 37°C overnight. DNA plasmids were purified with the NucleoSpin® Plasmid miniprep kit from Macherey-Nagel (1 purification column per 3 ml culture). After a 1h30-digestion with the corresponding restriction enzymes (see Table 5.1), 5 μl of digested constructs were loaded onto 1 % agarose gels. One positive clone per cloning was sent for sequencing to MWG or GATC, and 25%-glycerol stocks were frozen at -80°C for storing.

5.2.3 Insertion of a Tev-cleavage site in pASK-IBA3C vector

Primers were designed following the protocol from the QuikChange® Site-Directed Mutagenesis Kit from Stratagene, in order to add the nucleotide sequence encoding for the ENLYFQG Tev site between the cloning site and the Strep tag.

Principle of site-directed mutagenesis

A PCR is performed on the plasmid to be mutated with two oligonucleotide primers containing the desired mutation, each complementary to opposite strands of the vector. This generates a mutated plasmid.

While the parental template, which comes from a *dam* methylated *E. coli* strain such as NEB5α or JM109, can be digested by the DpnI endonuclease which is specific for methylated and hemimethylated DNA, the mutated vector cannot be digested by DpnI. Thereby, after DpnI digestion only the amplification product, hopefully with the desired mutation, remains.

Protocol

The PfuUltra® II Fusion DNA polymerase from Agilent was used for mutagenesis. The 50-μl reaction mix contained 5 μl of buffer, 34 ng of DNA template, 5 ng of each primer, 1 μl of dNTPs at 10 mM, 1 μl of polymerase (2.5 U). The reaction was performed in a thermocycler from Biometra® as follows: 30 s at 95°C; 18 cycles of 30 s at 95°C, 1 min at 55°C, 3 min 40 s at 68°C. 10 U of DpnI were added to the reaction product, then incubated at 37°C for 1h30. 10 μl of mutagenesis product were transformed into 200 μl of NEB5α cells as described earlier. Bacteria were plated in presence of 34 μg/ml of chloramphenicol. DNA plasmids from four colonies were purified with the NucleoSpin® Plasmid miniprep kit from Macherey-Nagel and sent for sequencing. Clone pASK-IBA3C-Tev4 was chosen and a 25 % glycerol stock was frozen at -80°C.

5.2.4 Cloning of single genes into expression vectors

The following vectors were employed:

- pET15b which allows the expression of proteins with a thrombin-cleavable His-tag,
- pET30b for expression of proteins fused to a C-terminal His-tag,

- a pETDuetTev for the expression of an untagged protein in the first cloning site and a Tev-cleavable N-terminal His-tagged protein in the second cloning site (pETDuet vector modified by Viviana Job)

(see FIGURE).

Cloning of *murC* and *murG* into pASK-IBA3C-Tev$_4$ vector

Previous to cloning into expression vectors, *murC* PCR product was purified by the NucleoSpin® Extract II kit from Macherey-Nagel, subsequently to electrophoresis on an agarose gel when several products of different sizes were present. A minipreparation of pASK-IBA3C-Tev$_4$ from 6 ml culture was done as described above (1 purification column per 6 ml culture).

Vector and PCR product were digested by incubating 24 μl of DNA with 3 μl of BsaI restriction enzyme from New England Biolabs (NEB) in 3 μl buffer 4 at 50°C for 1h. The digested products were purified by the NucleoSpin® Extract II kit from Macherey-Nagel after migration on a 1 % agarose gel, and concentrations were measured with a NanoVue device from GEHealthcare. Ligation was performed with T4 DNA ligase: 100 ng of vector were mixed to the insert with a 1 to 3 ratio in ligase buffer supplemented with 1 μl ATP, and incubated at 15°C overnight with 0.5 μl of ligase. 5 μl of the ligation product were transfomed into 50 μl of high-efficiency NEB5α competent cells as described previously.

DNA plasmids from colonies were purified. For each clone, 8.5 μl of DNA were digested by 0.5 μl BsaI for 1h at 50°C. Digested samples were loaded on a 1% agarose gel for electrophoresis. A positive clone was sent for sequencing to GATC, and a 25%-glycerol stock was frozen at -80°C.

A similar protocol was followed for *murG* cloning into pASK-IBA3C-Tev$_4$. In addition, subsequently to digestion with BsaI, the enzyme was inactivated at 65°C for 20 min. No electrophoresis was performed before purification of the digested PCR product.

Cloning of *mreB1*, *murD*, *murE*, *murF*, and *mreBbp*

pCRBlunt plasmids carrying the foreign genes were purified by minipreparation from 3 ml of culture as described above.

Then, the plasmids were sequentially over-digested by the two corresponding restriction enzymes for 2h30 at 37°C, each (see Table 5.1), with the following reaction mix: 24 μl of DNA, 3 μl of recommended buffer, 3 μl of enzyme. The two digestion steps were separated with a purification by the NucleoSpin® Extract II kit from Macherey-Nagel. After the second digestion, samples were loaded on a 1% agarose gel for DNA purification with the same kit.

Ligations were performed with T4 DNA ligase as described previously. Transformation of the ligation products into commercial NEB5α gave several colonies, which were screened as for pCRBlunt colonies. Positive clones were sent for sequencing to GATC.

Mutagenesis of the pET30b-*mreB1* construct to change the His tag to a Strep tag

The pET30b-*mreB1* expression vector was mutated to change the His tag to a Strep tag by site-directed mutagenesis (see mutagenesis of pASK-IBA3C for more details).

5.2.5 Cloning of *murD*, *murE*, and *murF* into pETDuetLIM1

murE was inserted into the second cloning site of pETDuetLIM1 (restriction sites NcoI / BamHI) with similar procedures, followed by insertion of *murF* (restriction sites NcoI / BamHI), resulting in the pETDuetLIM1-murEmurF$_2$ construct (partially sequenced), with similar procedures.

Likewise, *murD* was cloned into the first cloning site of pETDuetLIM1 as previously described for pETDuetTev. Then, *murE* and *murF* were inserted into the second site, resulting in the pETDuetLIM1-murD-murEmurF$_1$ construct (partially sequenced).

Chapter 6

Expression

6.1 Expression tests

For expression tests, expression vectors were transformed into some of the strains listed above and two colonies per strain were tested for expression. To do so, after a pre-culture at 37°C O/N, 3 ml of LB medium were inoculated with a 1:30 dilution from pre-cultures, and incubated at 37°C with 200 rpm shaking. When an optical density (OD) of 0.5 was reached:

- 200 μl of cell culture were harvested by centrifugation, resuspended in 50 μl 1X Laemmli buffer, heated at 100°C for 5 min, and kept on ice;
- After cooliing if required, 1 mM IPTG (pET vectors) or 400 ng/ml AHT (pASK vectors) were added to induce expression to the rest of the sample and cultures were incubated with shaking at the temperature to be tested.

After a 3h-induction or O/N induction, OD was measured and 200 μl of cell culture were harvested by centrifugation, resuspended in 50 μl 1X Laemmli buffer. 10 μl of non-induced cells and equivalent amounts of inducer-treated cells, according to OD measurements were loaded on 12.5 % to 15 % SDS-PAGE gel. 25%-glycerol stocks of positive colonies were frozen at -80°C.

6.2 Optimized expression protocols for native proteins

After a pre-culture at 37°C O/N from glycerol stocks selected following expression tests, growth was allowed as described in the previous section. When the appropriate OD was reached (see Table 6.1), flasks were cooled down for about 30 min in the cold room before induction. Expression was induced with 1 mM IPTG (pET vectors) or 400 ng/ml AHT (pASK vectors) at 20-25°C overnight, with 200 rpm agitation.

Protein	**Construct**	**Cells**	**OD before induction**
MurC	pASK-IBA3C-Tev4-murC$_1$	BL21Star^{TM}(DE3)	0.5
MurD	pET15b-murD-2$_1$ or pETDuetT5-murD4$_1$	BL21(DE3)	0.5
MurE	pET30b-murE$_3$	BL21(DE3)	0.9
MurF	pET30b-murF$_3$	BL21(DE3)	0.5
MurG	pASK-IBA3C-Tev4-murG$_1$	BL21(DE3)	0.9
MreB	pET30b-murF$_3$	C41(DE3)	0.5

Table 6.1: Induction protocols.

6.3 Over-expression of seleno-methionylated proteins

6.3.1 Principle

We chose the methionine biosynthesis pathway inhibition technique to produce seleno - methionylated (SeMet) derivatives. Most enzymes in this pathway are subject to regulation via feedback inhibition. Particularly, aspartokinases are inhibited in the presence of high concentration of isoleucine, lysine, and threonine [2].

The valuable advantage of this technique is that it does not require auxotrophic met- strains, nor any specific expression vector. Thus, no additional transformation is needed and the method is potentialy applicable to any expression strain.

6.3.2 Protocol

The same glycerol stocks used for expression of native proteins were used for the expression of the seleno-methionylated (Se-Met) derivatives, except for SeMet MurE, for which Rosetta strain was finally used instead of BL21(DE3).

After a first overnight pre-culture in 20 mL LB medium, cells were pelleted and washed in minimum medium M9+ before a second overnight pre-culture in 80 mL M9+. Cells were diluted in fresh M9+ medium to an OD of about 0.1 A.U. for growth at 37°C. When OD reached 0.4 A.U. (0.7 for MurE), amino acids were added as powders for inhibition of methionine synthesis (100 mg.L^{-1} of Lysine, Phenylalanine, Threonine, 50 mg.L^{-1} of Isoleucine, Leucine, Valine; all from Sigma), 200 mg.L^{-1} of thiamine (Sigma) was added, and 60 mg.L^{-1} of L(+)-seleno-methionine (from Acros Organics). After 10 min of growth with low shaking at room temperature, flasks were cooled down in cold room for 20 min and expression was induced at 25°C except for MurE expression for which expression was done at 20°C.

6.4 Co-expression

Although pET and pASK vectors contain both the ColE1 origin of replication, co-transformation and co-expression with these plasmids, which have different resistances, was assessed.

For co-transformation, two or three plasmids were added at the same time and the transformation protocol described earlier was followed. Cells were spread onto LB-Agar plates supplemented with appropriate antibiotics.

For co-expression tests, a procedure similar to the one used for expression tests was followed. For co-expression of both pASK and pET vectors, simultaneous or sequential inductions were assessed.

Chapter 7

Protein purification

7.1 Lysis and solubilization

7.1.1 Solubilization tests

For solubilization tests, several 30-ml cell cultures were grown at 37°C and induced at various induction temperatures (1 mM IPTG or 400 ng/μl AHT), optical densities, and times. Cells were subsequently lysed by sonication in different buffers with the small probe of the Digital Sonifier® from BRANSON (10% intensity), and finally centrifuged at 18,000 rpm for 45 min at 4°C. Equivalent amounts of cells before and after centrifugation were loaded to assess solubility. For MurC and MurG solubility tests, the lysis buffer was 25 mM NaPi, 0.2 M NaCl, supplemented with 5 CMC β-OG, or 2 or 5 CMC DDM, or 2 or 5 CMC Triton, or 2 CMC CYMAL-5, or 1% sarcosine, all with or without 5 mM DTT. For MurD solubility tests, the lysis buffer was 50 mM Tris pH 8.0, 0.15 M NaCl, 5% glycerol, supplemented with proteases inhibitors.

7.1.2 Lysis and solubilization for purifications

After an overnight induction, cultures were centrifuged at 4,000 rpm for 20 min at 4°C. Pellets were resuspended in buffer A (see Table 7.1 for composition) supplemented with proteases inhibitors by shaking bottles at 240 rpm at 16°C. Lysis was performed by sonication with the Digital Sonifier® from BRANSON: for 1 L of culture, 2 min sonication (2 sec pulses at 40% intensity separated by 10 s pauses).

Lysates were centrifuged at 18,000 rpm for 45 min at 4°C or ultracentrifuged at 40,000 rpm for 45 min at 4°C on Ti45 from Beckman (MreB), to harvest the soluble part.

7.2 Membrane extraction and solubilization for MurG purification

750 mL of BL21(DE3) cells carrying pASK-IBA-3C-Tev4-murG$_1$ were induced with 200 ng/ml AHT at 20°C O/N. Cells were pelleted and lysed in 40 ml buffer (25 mM NaPi pH 7.4, 0.3 M NaCl, 10% glycerol) by sonication as described earlier. The lysate was centrifuged at

10,000 rpm for 10 min at 4°C to spun down unbroken cells and large debris. The supernatant was ultracentrifuged at 40,000 rpm in Ti45 rotor for 1h at 4°C. The pellet was re-suspended and achieved to homogenization with a hand-driven Potter-Elvejem homogenizer in 50 mM Hepes pH 7.5, 0.3 M NaCl, 10% glycerol. 1% N-lauryl sarcosine and 1% DDM were added before 1h incubation on a rotating wheel at 4°C. The sample was ultracentrifuged at 40,000 rpm for 1h at 4°C. Sample was diluted twice before purification by Strep-trap affinity column (see below) in 50 mM Hepes pH 7.5, 0.3 M NaCl, 0.04% DDM.

7.3 Affinity chromatographies

7.3.1 Principle

Soluble fractions of cell lysates are loaded onto columns made of a chromatography matrix on which a specific ligand is coupled. The fast binding interaction allows proteins to bind the affinity media while the sample is passed through the column. Before elution with a chemical whose affinity for the ligand is higher than the protein-ligand affinity, the resin is washed to remove proteins that may have unspecifically bound the column.

His-trap^{TM}affinity chromatography

This immobilized metal affinity chromatography relies on the ability of histidine to bind Nickel. The metal is bound to a chromatography matrix by chelation. The histidine tag will bind Nickel with a micromolar affinity. His-tag fusion proteins can be easily desorbed from the resin with buffers containing high concentration of imidazole, since imidazole has a higher affinity for Nickel than the His tag. To regenerate the column, Nickel ions are removed by EDTA.

The 1 ml HisTrapHP column or the 1 ml HisTrapFF crude column (GE Healthcare) were used. Both columns are made of Sepharose, a matrix of highly cross-linked agarose beads, to which a chelating group has been immobilized.

Strep-trap affinity chromatography

This chromatography is based on the Strep-tag®-Strep-Tactin® system. The Strep-tag is a synthetic peptide (WSHPQFEK) which exhibits high affinity towards streptavidin[1]. Strep-Tactin® is a variant streptavidin with improved peptide-binding capacity [4].

When a Strep-tag fusion protein is applied to a Sepharose column with immobilized Strep-Tactin, it specifically binds to Strep-Tactin. Because of this high specificity, a short washing step allows to remove all other host proteins. Then, the purified Strep-tag fusion protein is eluted with a low concentration of *d*-desthiobiotin, which specifically competes for the biotin binding pocket of Strep-Tactin. To regenerate the column, *d*-desthiobiotin can be removed by NaOH.

[1]Streptavidin is a protein from the bacterium *Streptomyces avidinii*. Streptavidin homo-tetramers have an extraordinarily high affinity for biotin, one of the strongest non-covalent interactions known in nature [3].

7.3.2 Experimental procedures

Supernatants were loaded at 1 ml/min at 4°C onto affinity columns from GE. Washing and elution steps were done with an Äkta fast protein liquid chromatography (FPLC) device from Amersham, as described in Table 7.1 at a 0.5ml/min - 1 ml/min flow.

Protein	Basis Buffer (BB)	Lysis buffer	Column	Equilibration	Washing	Elution
MurC	25 mM NaPi pH 7.4, 150 mM NaCl	-	5 ml HP Strep-trap	BB	BB	2.5 mM
MurDhis	25 mM Hepes pH 7.4, 150 mM NaCl	25 mM	1 ml HP His-trap	25 mM	Gradient	500 mM
MurD	25 mM Hepes pH 7.4, 150 mM NaCl	25 mM	1 ml HP His-trap	25 mM	Gradient	500 mM
MurEhis	25 mM Hepes pH 7.4, 150 mM NaCl	25 mM	1 ml HP His-trap	25 mM	Gradient	500 mM
MurFhis	25 mM Hepes pH 7.4, 150 mM NaCl	25 mM	1 ml HP His-trap	25 mM	Gradient	500 mM
MurGstrep	25 mM Hepes pH 7.4, 150 mM NaCl, 0.04% DDM	0.5 % sarcosine	5 ml HP Strep-trap	BB	BB	2.5 mM
MreBhis	25 mM Hepes pH 7.4, 150 mM NaCl	35 mM; DNase, lysozyme	1 ml FF His-trap	35 mM	86 mM	500 mM
MreBstrep	25 mM Hepes pH 7.4, 150 mM NaCl	DNase, lysozyme	5 ml HP Strep-trap	BB	BB	2.5 mM

Table 7.1: Purifications. Basis Buffer indicates the buffer used for the whole purification, supplemented with the indicated reagents and the mentioned concentrations of imidazole (Lysis buffer, Equilibration buffer, Washing buffer and Elution buffer) or *d*-desthiobiotin. Affinity chromatography columns were purchased from GE Healthcare. Imidazole was from Euromedex, DNase from Sigma, lysozyme from Fluka, NaPi from Sigma, Hepes from Euromedex, N-lauryl sarcosine from Sigma, and all other detergents from Anatrace.

7.4 Heat purifications

As proteins from *T. maritima* are thermostable, heat purifications were tested.

For MurD heat purifications, cell were lysed in 50 mM Hepes pH 7.4, 0.1 M NaCl, 10% glycerol, 5 mM EDTA supplemented with proteases inhibitor, DNase, 2.4 μl/ml of $MgSO_4$ 1M and 2.4 μl/ml of $MnCl_2$ 1M. Lysates were heated at 90°C (MurD) for 45 min, and then centrifuged at 20,000 rpm for 30 min. Supernatants were concentrated with a Vivaspin20 30,000 MWCO from Sartorius Stedim Biotech before gel filtration chromatography (MurD, MurE, MurF).

For MurE and MurF tests of heat purifications, 50 ml of cells were induced at 37°C for 3h under 1 mM IPTG. Cultures were centrifuged and pellets were frozen at -80°C. Cell were resuspended and lysed by sonication (small probe) in 5 ml 25 mM Hepes pH 7.4, 0.15 M NaCl, 5% glycerol, 5 mM EDTA supplemented with proteases inhibitors and DNase. Lysates were heated at 78°C for 1 h, and then centrifuged at 13,500 rpm for 20 min. Supernatants were dialysed in 0.1 M NaCl, 25 mM Hepes pH 7.4, at 4°C overnight, then concentrated with a Vivaspin20 30,000 MWCO from Sartorius Stedim Biotech before gel filtration chromatography.

To test the feasibility of purifying MurG and MreB by heating, 50 mL of BL21(DE3) cells

carrying pASK-IBA3C-*murG* and C41(DE3) cells carrying pET30b-*mreB1* were induced at OD of 0.45 A.U. with 400 ng/mL AHT or 1 mM IPTG, respectively, at 20°C overnight. Cells were pelleted and resuspended in 50 mM Hepes pH 7.5, 5 mM EDTA, 2 mM DTT, 0.15 M NaCl, supplemented with DNase, protease inhibitors, 0.5 % sarkozyl (MurG) or 0.5 mg/ml lysozyme (MreB). Cells were either firstly heated at 70°C for 45 min and then sonicated, or first sonicated and then heated. Lysates were then centrifuged at 20,000 rmp for 30 min. An additional ultracentrifuge step for MreB was performed at 100,000 g for 45 min.

7.5 Size exclusion chromatographies

Principle

With this method, also known as gel filtration chromatography, molecules in solution are separated by their size by passing the sample through a resin made of pores of different sizes at a stationary flow. More precisely, it has been suggested that separation depends on the hydrodynamic volume[2]. A small protein can penetrate every small pore of the gel and will elute late, while a very large molecule which cannot penetrate the pore system and travels only in the interparticle volume will elute earlier.

Protocol

SuperDex200 or SuperDex70 gel filtration columns from GE Healthcare were used for purifications of MreB, MurC, MurD, MurE, and MurF ligases (running buffer: 25 mM Hepes pH 7.4, 150 mM NaCl, 1 mM EDTA), and Superose6 column was used for MurG purification (similar buffer supplemented with 0.04% DDM).

7.6 Tag cleavage

Best conditions for the cleavage of the His tag of MurD were screened with different amounts of Thrombin protease and different incubation times in 25 mM NaPi pH 7.4, 0.15 M NaCl. The concentration of MurD was 1 mg/ml. The best condition was overnight at room temperature with 5 units of protease per mg of MurD.

7.7 Purification of Se-Met derivatives

Identical purification protocols were followed for purification of Se-Met proteins. The only difference was that every buffer contained 2 mM dithiothreitol (DTT).

7.8 Storage

After purification, proteins were stored in gel filtration buffer at 4°C for one week maximum before use.

[2]The hydrodynamic volume is the sum of the time-average of the molecular volume and the volume of the solvent molecules associated with it.

Chapter 8

Protein characterization

8.1 SDS-PAGE and Western Blotting

Protein samples were analyzed by SDS-PAGE on 12.5 % - 15 % polyacrylamide gels. Gels were stained with Coomassie blue. The Low Molecular Weight standard from GE Healthcare was used to evaluate protein sizes.

For Western Blotting, after running a classical SDS-PAGE proteins were transferred to a nitrocellulose membrane for 1 hour at a constant voltage (300V) or with the iBlot Dry Blot system from Invitrogen. Following protein transfer, the membrane was blocked for one hour with shaking at room temperature or overnight at 4°CC in 3% BSA in PBS-T (PBS-Tween 0.05%). After three washing steps in PBS-T, the membranes were incubated with either the Streptactin HRP conjugate from IBA at 1:20000, or the SuperSignal West HisProbefrom Pierce Biotechnology at 1:2000 in PBS-T. Signal was revealed with the SIGMA FASTDAB tablets with Metal Enhancer according to the manufacturers instructions. The protein ladder Page RulerPlus Prestained from Thermo Scientific was used.

8.2 Protein concentration

Concentrations of purified proteins were evaluated with a NanoDrop 2000c spectrophotometer from ThermoScientific by measuring the absorbance at 280 nm and calculating the concentration with a molar extinction coefficient calculated from the protein sequences with the ProtParam [5] tool from Expasy.

8.3 Mass spectrometry

Quality of protein samples was controlled by mass spectrometry. This technique measures the mass-to-charge ratio m/z of ionized molecules in the gas-phase and allows the determination of the exact masses of the molecules within a protein sample, providing some information such as:

- possible degradation of the protein,
- presence and characterization of contaminants,
- chemical modifications of the protein, e.g. presence of disulfide bridges, oxydation, etc.

- percentage of incorporation of seleno-methionine.

Further explanation of the technique is available in Appendix B.

For mass spectrometry analyses, purified proteins at concentrations in the mg/ml range were analyzed on the Mass Spectrometry platform of IBS by Luca Signor. MALDI- and ESI-TOF techniques were used, with the following instruments:

- MALDI-TOF MS (Autoflex, Bruker Daltonics); mass range up to 500 kDa in linear mode and a resolution up to 20'000 in reflectron mode.
- LC ESI-TOF MS (6210, Agilent Technologies), with a mass range of 30-3000 Da and a resolution $>$ 20'000.

8.4 Thermal Shift Assay

Thermal shift assay monitors the thermal stability of proteins according to denaturation with temperature increase. This assay allows to investigate factors affecting this stability, such as buffer nature, NaCl concentration, or ligand binding.

8.4.1 Free energy of unfolding

The stability of a protein can be defined by its tendency to remain folded. If the protein unfolds in a reversible two-state manner the equilibrium thermodynamics models will apply. In the thermal shift assay, we consider the reaction:

$$F \leftrightarrow U$$

where F and U are the folded and unfolded states, respectively.

The folding stability of a protein is related to its Gibbs free energy of unfolding ΔG_u. Basically, at a given temperature, ΔG_u is the difference in Gibbs free energy between the unfolded U and the folded F states. Therefore, the larger and more positive ΔG_u, the more stable the protein is [6] [7].

ΔG_u depends on temperature:

$$\Delta G_u(T) = \Delta H_u(T) - T\Delta S_u(T)$$

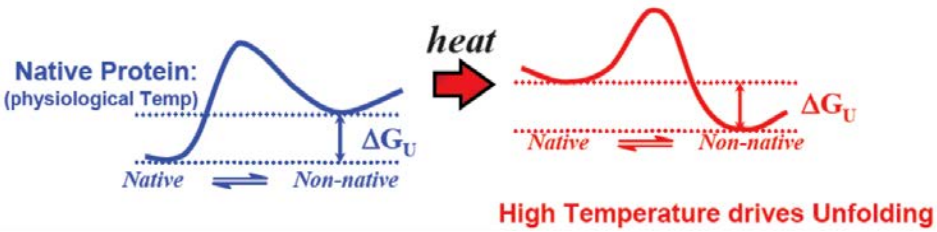


Figure 8.1: Gibbs free energy as function of the unfolding reaction showing that temperature inverses the way of spontaneous reaction. 'Native' refers to folded state and non-native to unfolded state. From Fluorescence-based thermal shift assays by Carrie Partch, Dept. of Chemistry and Biochemistry, UC Santa Cruz http://partch.chemistry.ucsc.edu/pdf/ThermalShiftAssay.pdf

Thus, as $\Delta S_u(T) > 0$ according to the second law of thermodynamics, when the temperature increases the free energy of unfolding decreases, meaning the protein is less stable.

8.4.2 Monitoring unfolding of proteins upon temperature increase

In the thermal shift assay, a fluorescent probe is added to the protein sample, and reveals the unfolded state of the protein. The fluorescence signal y is monitored as a function of temperature. It is the sum of the fluorescence of the folded proteins and the unfolded proteins:

$$y(T) = y_F^*.n_F(T) + y_U^*.n_F(T) \tag{8.1}$$

where $n_F(T)$ and $n_U(T)$ are the number of proteins in folded and unfolded proteins, respectively, at a temperature T and y_F^* and y_U^* the fluorescence of a protein molecule in the folded or unfolded state, respectively. Let us define y_F and y_U, the fluorescences of the sample when all the proteins are in folded state, unfolded state, respectively:

$$\mathrm{y}_F{=}\mathrm{y}(0){=}\mathrm{y}_F^* n_{tot} \tag{8.2}$$

$$\mathrm{y}_U{=}\lim_{T\to\infty} y(T) = y_U^* n_{tot} \tag{8.3}$$

where n_{tot} is the total number of protein molecules

Then, equation 8.1 becomes:

$$y(T) = y_F P_F(T) + y_U P_U(T) \tag{8.4}$$

where P_F P_U are the probabilities for a protein to be in the unfolded or folded state, respectively.

The equilibrium constant is defined as follow:

$$K = \frac{[U]}{[F]} = \frac{P_U}{P_F} = \frac{P_U}{1 - P_U} \tag{8.5}$$

Besides,

$$K = \mathrm{e}^{-\Delta G_u(T)/RT} \tag{8.6}$$

By combining equations 8.5, 8.4, and 8.6, one obtains:

$$\boxed{y(T) = y_F + \frac{y_U - y_F}{1+\mathrm{e}^{-\Delta G_u/RT}}} \tag{8.7}$$

8.4.3 Melting temperature

There is a specific temperature, T_m, named melting temperature, for which

$$\Delta G_u(T_m) = 0$$

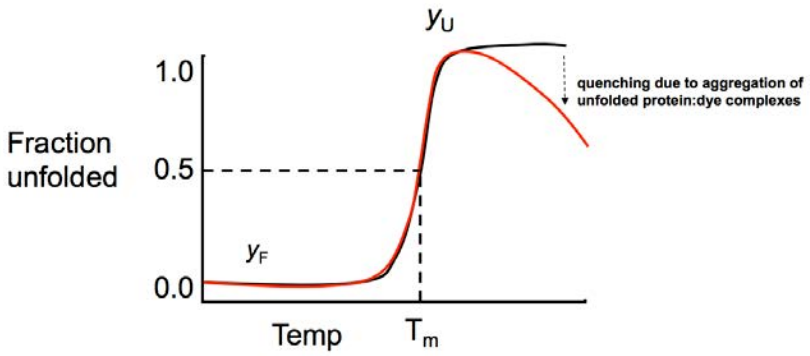


Figure 8.2: Typical experimental (red) and theoretical (black) curves for a thermal shift assay. From Fluorescence-based thermal shift assays by Carrie Partch, Dept. of Chemistry and Biochemistry, UC Santa Cruz http://partch.chemistry.ucsc.edu/pdf/ThermalShiftAssay.pdf

Thus,

$$\boxed{y(T_m) = \frac{y_F + y_U}{2}} \tag{8.8}$$

The stability of most proteins decreases with temperature. A given protein will be most stable at the conditions where T_m is the highest. Therefore, the thermal shift assay is a way to evaluate protein stability and determine the best buffer for protein stability [8].

8.4.4 Sypro® orange

The Sypro Orange probe, whose fluorescence is quenched in an aqueous environment, interacts with hydrophobic regions of proteins. When the protein is well folded, the hydrophobic parts are mostly inaccessible, so no fluorescence can be detected. As the temperature rises, the protein undergoes thermal unfolding and exposes its hydrophobic core region. Then the probe can bind the hydrophobic regions and becomes unquenched.

8.4.5 Protocol

Pure proteins were diluted in NaPi 25mM pH 7.4, 0.15 M NaCl at 0.1 mg/ml, 0.3 mg/ml, 1 mg/ml, and 3.2 mg/ml. 25 μl of each dilution were transfered to a 96-well plate, and 2 μl of Sypro Orange probe freshly diluted in water were added (final dilution in Sypro Orange 1:200). Fluorescence signal was measured with a BioRadiQ5 device and data were analyzed by Excel.

8.5 Electron Microscopy

8.5.1 Principle

An electron microscope illuminates a protein sample and produces a magnified image using an electron beam, by contrast to an optical microscope which uses a source of photons.

Magnification is limited by diffraction. Because electrons in electron microscopes have wavelengths in the range of 1 pm, an electron microscope has greater resolving power than a light microscope in which photons cannot have wavelengths shorter than 200 nm. There-

fore, electron microscopy is the right technique to get an insight into the ultrastructure of biological entities such as protein complexes, protein polymers, or even large proteins.

In a scanning transmission electron microscope, a beam of electrons is transmitted through a part of ('scanning') an ultra thin specimen such as a carbon-coated grid layer of protein in buffer. Electrons lose energy as they pass through the specimen, depending on the thickness, the electron density, and chemical properties. A signal is detected from the electrons transmitted through the specimen and is transformed to an optical image, then magnified and focused onto an imaging device (such as a CCD camera). This provides information about the properties of the specimen surface, such as its topography and composition.

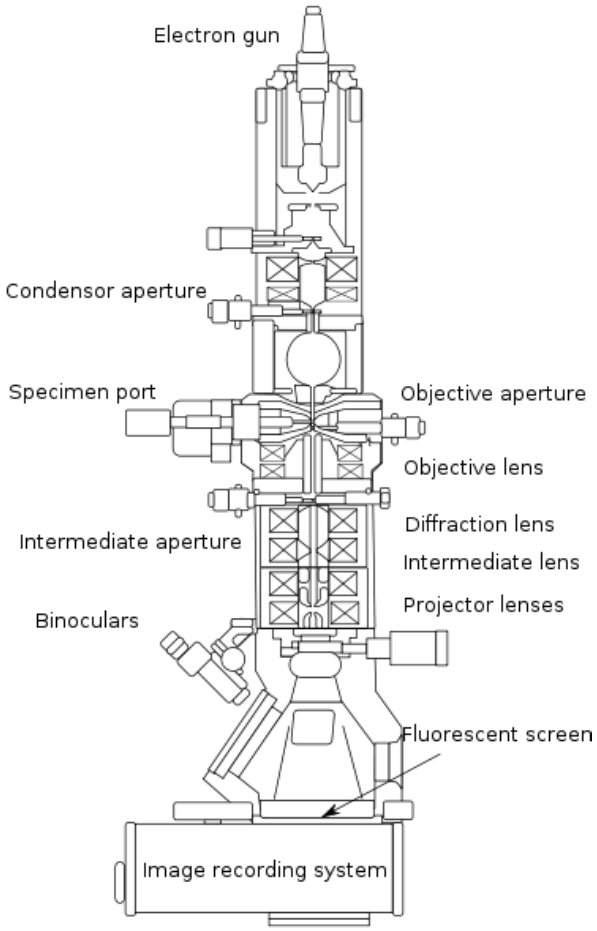


Figure 8.3: Diagram outlining the internal components of a basic TEM system. From Wikipedia.

8.5.2 Methodology

Electron microscopy (EM) studies were performed at the EM platform of the IBS on a Scanning Transmission Electron Microscope CM12 of 120kV from Philips with 2.6k x 4k CCD camera.

A grid was prepared by negative coloration on carbon-coated grids. For that purpose, a carbon film was evaporated onto a mica surface and the sample was applied to the surface of the carbon film attached to the mica. Then, the carbon film and the sample were picked

up onto a specimen grid. In order to increase contrast, negative staining with heavy atoms technique was used. Since heavy atoms fix onto the surface of absorbed proteins, the stain absorbs or scatters part of the electron beam. The contrast between regions with heavy atoms such as proteins and other regions is enhanced, and the proteins appear lighter than the rest of the sample. Two different stain solutions were used: aqueous Uranyl Acetate or Sodium Silico Tungstate (SST). The sample was examined in the electron microscope when completely dry.

8.6 N-terminal sequencing of MurG

N-terminal sequencing is very useful in the confirmation of the nature of proteins present in a sample after purification. N-terminal sequencing was performed on the N-sequencing platform of the Institute of Structural Biology by Jean-Pierre Andrieu.

8.6.1 Principle

Amino acid sequence determination based on Edman degradation was performed. With this technique, the amino-terminal residue is labeled and cleaved from the peptide without disrupting the peptide bonds between other amino acid residues. The derivative of the terminal amino acid is selectively extracted and treated for identification by chromatography or electrophoresis. Then, the procedure is repeated to identify the following amino acids in the protein.

8.6.2 Sample preparation and analysis

A SDS-PAGE was run with MurG proteins. The gel was then washed in 10 mM CAPS pH 11.0 (Sigma), 10 % methanol. A PVDF (polyvinylidene difluoride) membrane, previously soaked in methanol, is washed in the same buffer. Transfer at 300 mA for 1h (Transfer device from Bio-Rad) was performed from the gel to the membrane. The membrane was subsequently stained in Coomassie blue with 50 % methanol and 1 % acetic acid for 5 min, and then destained in 50 % methanol by 2 10-min washing steps. Protein lanes to be analyzed were cut for N-terminal sequencing by the N-sequencing platform of the IBS.

Analysis were performed using an Applied Biosystems gas-phase sequencer model 492 (s/n: 9510287J). Phenylthiohydantoin amino acid derivatives generated at each sequence cycle were identified on-line with an Applied Biosystems Model 140C HPLC system using the data analysis system for protein sequencing from Applied Biosystems Model 610A (software version 2.1). The procedures and reagents used were as recommended by the manufacturer.

Chapter 9

Protein crystallization

Protein crystal structures help the understanding of biological functions at the atomic level by providing a concrete and spatial representation of proteins. In addition to fundamental breakthroughs in biological science, crystal structures of proteins provide a basis for designing therapeutic molecules. To date, X-ray scattering is the most efficient way - if not still the only way, to get protein structures at a resolution below 3 Å which is needed for an atomic understanding of biomolecules. However, to obtain an X-ray scattering pattern from macromolecules with sufficient data, a highly ordered arrangement of protein units within the sample is needed.

9.1 Protein crystals

Crystals, which consist in three-dimensional repeating packings of molecules, satisfy this requirement. They are generated by a low precipitation process which allows molecules to arrange in a regular way, thus achieveing the lowest free energy state [9].

As any other crystals, protein crystals present flat planes on their surface which reflect the periodic packing[1] and the underlying symmetry.

Nevertheless, three main differences from small molecule crystals can be described.

- Due to the macromolecular size of proteins, samples are often more heterogeneous causing more crystallographic irregularities than small molecules crystals;
- The possibilities in terms of symmetry are limited by the chirality of proteins. For instance, mirrors are not allowed in protein crystals;
- Protein crystals contain a high amount of solvent - from 30 % to 70 %, which is mainly unordered within the crystalline lattice.

This high amount of solvent helps keeping crystallized proteins in an environment which is very close to their natural aqueous environment, supporting the hypothesis that while in crystalline form, proteins behave the same as in solution. This hypothesis is further supported with studies in which crystallized proteins present an enzymatic activity.

Producing protein crystals is often the rate-limiting step in protein structure determination as crystallization is a complex, multi-parameter process which is not fully understood yet. Therefore, it is mainly a trial-and-error procedure in which the protein is slowly pre-

[1] By contrast, amorphous solids such as glass are characterized by soft and round surfaces

cipitated from solution [9]. This chapter first presents simple mechanisms and models which explain the basics of crystallization. Then, the vapour diffusion method and factors influencing protein solubility and thus the formation of crystals are listed. Lastly, experimental procedures for the production of crystals of Mur ligases and MreB are exposed, and the screening procedure for crystallization conditions of a MreB-Mur crystal as well.

9.2 Supersaturation, a requirement for crystallization

9.2.1 Crystallization is a phase transition

When crystallizing, the protein goes from a soluble state to a particular, regularly ordered, solid state.

As a general rule, when reducing the solubility of a protein solution, an amorphous precipitate forms. However, under specific and unpredictable conditions, complementary patches can be present on the surfaces of protein molecules, causing specific attractive interactions. If these interactions are geometrically favorable, the crystallization process can start [10].

An intuitive approach suggests that the transition must be slow enough to allow the formation of such well-ordered arrangements.

9.2.2 Supersaturation

This can be achieved only if the solution enters an out-of-equilibrium zone called supersaturation state. In a supersaturated solution, the kinetics of the phase transition is so slow that the system lies in a metastable state characterized by an excess of soluble protein molecules regarding thermodynamic laws. Protein units can thus slowly arrange into crystals if conditions allow attractive forces to generate a geometrically favorable packing.

Definition of supersaturation

The degree of supersaturation is defined by the ratio between the activity a of the soluble protein and its corresponding theoretical activity a^* in the same conditions but under thermodynamic equilibrium:

$$S = \frac{a}{a^*}$$

$S = 1$ refers to the thermodynamic equilibrium, while $S > 1$ corresponds to a supersaturated solution. The latter is the metastable state required for crystallization.

Supersaturation diagram

This phenomenon can be depicted by the supersaturation diagram (Figure 9.1). There, the supersolubility curve separates two supersaturated zones:

- the *'nucleation zone'* where supersaturation leads to nucleation, the first step in crystallization,

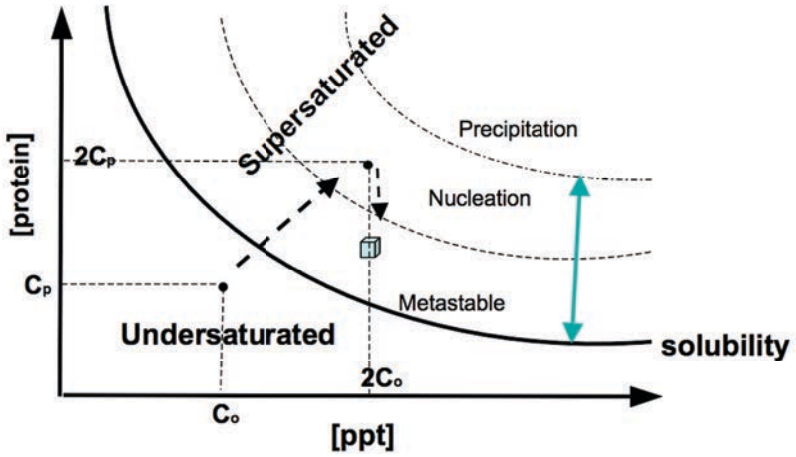


Figure 9.1: Solubility-supersolubility phase diagram of a protein as function of its concentration and a precipitating agent (ppt). $[C_o; C_p]$ corresponds to a condition where no crystal will appear since below the solubility curve, while $[2C_o; 2C_p]$ corresponds to a favorable condition for crystallization. Arrows depict the path of a protein solution during a crystallization experiment [11].

- and the *'metastable zone'* where the lower supersaturation allows crystal growth but is not sufficient for nucleation[2].

9.3 Nucleation

Nucleation, the starting point of crystallization, is the appearance of very small crystals called nuclei (of about one hundred molecules), which do not dissolve in solution.

There are different nucleation mechanisms:

- *primary nucleation*: crystallization occurs spontaneously from a clear solution, either by progressive formation of protein aggregates (*homogeneous nucleation*), or by nucleation onto a foreign surface present in the sample such as dust or the container surface (*heterogeneous nucleation*),
- *secondary nucleation*: nucleation is artificially induced by the presence of crystalline protein particles within the solution. This is commonly referred to as *'seeding'*.

9.3.1 Homogeneous nucleation

Model

In this model, nucleation is the result of spontaneous molecular interactions between protein molecules which lead to the formation of oligomers, and then little spherical aggregates with a primitive regular structure [12].

[2]These terms may be a little bit confusing as the nucleation zone corresponds to a metastable state as well.

Excess free enthalpy

The change in excess free enthalpy[3] ΔG of n molecules of molecular volume v from a soluble state to a spherical cluster with a radius r is given by:

$$\boxed{\Delta G = \tfrac{4}{3v}\pi r^3 kT \ln S + 4\pi r^2 \sigma} \tag{9.1}$$

where σ is the surface tension of the growing aggregate, k Boltzman's constant, and T the temperature. The first term corresponds to the volume contribution and the second term to the surface contribution (see Figure 9.2).

Model for homogeneous nucleation, from Garcia *et al.*, 2003 [12]

Indeed, let us consider the primary aggregation of protein molecules $\alpha \leftrightarrow \beta$ where α and β refer to the soluble and solid phases, respectively.
This modification is characterized by the change in excess free enthalpy ΔG

$$\Delta G = \Delta G_S + \Delta G_V$$

where ΔG_S is the surface contribution and ΔG_V the volume contribution.
An energy ΔG_S is needed to form an interface, due to surface tension σ of the novel interface. In the case of a spherical aggregate of radius r one has:

$$\Delta G_S = 4\pi r^2 \sigma$$

The volume energy $\Delta G_V < 0$ that is released when a spherical aggregate is created, is given by the difference in chemical potential of the n molecules in solution μ_α and in the solid phase μ_β by:

$$\Delta G_V = -n(\mu_\alpha - \mu_\beta)$$

The chemical potentials can be expressed as functions of the activities a_α and a_β of the molecules in the solution and in the solid phase, respectively:

$$\mu_\alpha = kT \ln(a_\alpha)$$

$$\mu_\beta = kT \ln(a_\beta)$$

With S the degree of supersaturation defined earlier, ΔG_V can be expressed by:

$$\Delta G_V = nkT \ln(S)$$

Therefore,

$$\Delta G_V = \frac{4}{3v}\pi r^3 kT \ln(S)$$

Hence,

$$\boxed{\Delta G = \tfrac{4}{3v}\pi r^3 kT \ln S + 4\pi r^2 \sigma} \tag{9.2}$$

Critical radius

Spontaneous nucleation can occur if $\frac{dG}{dr} < 0$,

$$\frac{dG}{dr} = 4\pi r^2 (2\sigma + r\Delta G_v) < 0$$

[3]The excess free enthalpy is defined by comparing the free enthalpy of a non-ideal binary system to a perfect solution

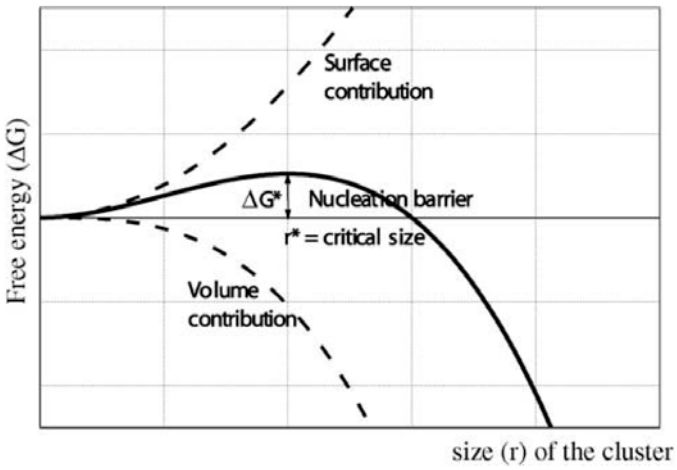


Figure 9.2: Excess free enthalpy diagram which shows the existence of a critical radius [11].

giving the critical radius for which $\frac{dG}{dr}(r^*) = 0$:

$$\boxed{r^* = -\frac{2v\sigma}{kT \ln S}} \tag{9.3}$$

The free energy needed to form a cluster of this critical radius r^* is thus:

$$\boxed{\Delta G^* = \frac{16\pi v^2 \sigma^3}{3(kT \ln S)^2}} \tag{9.4}$$

This reveals that:

- There is an energetic barrier ΔG^*, known as the nucleation barrier, that must be crossed to induce the formation of stable nuclei [12];
- The higher the supersaturation, the lower the activation barrier. Indeed, for $S = 1$ (equilibrium), $r^* \to \infty$, meaning no nucleation is possible. By contrast, at very high supersaturation levels, the size of the critical nucleus becomes smaller than the smallest structural unit, meaning the activation barrier disappears and amorphous phases form [12].

In another words, supersaturation is the driving force for nucleation and because of the existence of a critical radius, the solution must be supersaturated beyond a threshold value to be able to form crystals [12], explaining the existence of the 'metastable zone' on the phase diagram.

Interestingly, as illustrated by Figure 9.2, addition of new molecules to clusters larger than the critical radius releases available work [9]. That means that, once nuclei are big enough, crystallization is no longer limited by nucleation, but possibly by diffusion or interaction kinetics [12].

Nucleation rate and induction time

The nucleation barrier can be crossed thanks to fluctuations in the environment. Thereupon, nucleation is a probabilistic phenomenon [12].

As a result, the probability that an aggregate becomes a stable nucleus (nucleation frequency) can be estimated according to Boltzman's distribution law:

$$J = \kappa_0 \exp(-\frac{\Delta G^*}{kT})$$

$$\boxed{J = \kappa_0 \exp \frac{-16\pi v^2 \sigma^3}{3(kT)^3 [\ln S]^2}} \tag{9.5}$$

where the pre-exponential factor κ_0 is related to the kinetics of attachment of growth units to the forming cluster [12] and cannot be easily modeled or predicted as it is intimely linked to the properties of the protein in the given solution which at present are not well understood.

As illustrated by Figure 9.3, the nucleation rate J is very low at low supersaturation values and increases fast when a critical value is achieved [12].

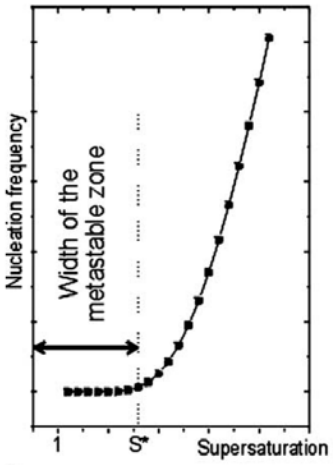


Figure 9.3: Nucleation rate as a function of supersaturation [12]

One can define the time elapsed between the creation of supersaturation and the formation of nuclei of critical size, known as *'induction time'* and inversely proportional to J [12].

Supersaturation diagram and nucleation

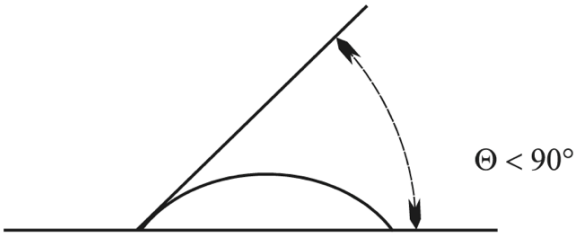
As indicated by Figure 9.3, J increases rapidly after a certain critical value of supersaturation S^* is achieved. This explains the existence in the supersolubility diagram of the metastable zone where there is a very low probability of nucleation [12]. Indeed, this region is delimited by the curve corresponding to $S = 1$ (saturation, thermodynamic nature) and the curve of kinetic nature corresponding to S_c (where J reaches significant values).

The value of S_c and therefore the location of the metastable limit depends not only on the protein and the solvent natures and concentrations, but also on the rate at which supersaturation is created in the crystallization experiment [12], which depends on even more factors.

9.3.2 Heterogeneous primary nucleation

Nevertheless, the absence of any foreign matter in the sample is very rare in practice. Samples often contain irregularities brought from another phase such as dust, contaminants, bubbles, or phase separation. The contact of the protein solution with the surfaces of these materials triggers aggregation onto these surfaces.

There, nascent nuclei present a flattened spherical shape instead of the spherical shape in the first model, resulting from intermolecular interactions between the protein aggregate and the surface. The flattened spherical shape is characterized by the angle θ between the surface of the cluster and the surface of the foreign matter as observed in the phenomenon of droplet wetting (see Figure 9.4).



(a) water droplet wetting

Figure 9.4: Droplet wetting. When a liquid droplet stands onto a solid surface, wetting occurs: the contact angle between the droplet surface and the solid surface reflect the intermolecular interactions between water and the molecules from the surface.

In this model, the activation energy is given by

$$\Delta G^*_{hetero} = \Phi(\theta).\Delta G^*_{homo}$$

where Φ depends on θ and $\Phi(\theta) < 1$.

Except for this difference, heterogeneous primary nucleation is very similar to homogeneous primary nucleation, and the only consequence of heterogeneous nucleation is a lower free energy barrier and thus a higher probability of nucleation.

9.3.3 Secondary nucleation

Nucleation is favored when very small crystals of protein are already present in solution as the nucleation barrier is decreased.

There are three different secondary nucleations:

- The presence of very small crystals in solution may favor the homogeneous primary nucleation of novel crystals by maintaining a high supersaturation level;
- Desaggregation of the very small crystals may serve as novel surfaces for nucleation;
- Introduced crystals may serve as seeds on which a new crystal will grow, based on the packing of the primary crystal.

Interestingly, secondary nucleation reduces - or abolishes in the last case, the induction time. The first two processes are involved in microseeding techniques, while the third one is applicable to macroseeding.

9.3.4 Conclusion

The introduction of these models gave the basics of nucleation, revealing the importance of supersaturation, and kinetics of the phase transition in crystallization. Particularly, the greater the supersaturation, the smaller and more numerous the nuclei. Though these models depend on a high number of parameters, they do not explain all crystallization data, and novel nucleation models are being developed.

9.4 Crystal growth

9.4.1 Growth process

In this second, faster, process compared to nucleation, the soluble proteins progressively add to the growing crystal in a regular way, continuing the crystal lattice started in crystal nucleation.

The attachment of additional molecules occurs preferably at small irregularities on the crystal surface, where the binding energy is higher compared to areas where the surface is flat.

Growth rate

The global growth rate is determined by the level of supersaturation when nucleation occurs. The higher the supersaturation, the higher the growth rate. However, if supersaturation is too high, new nuclei will form as first crystals are growing, thus limiting the growth rate of formed crystals.

If crystals grow fast, supply in protein units may limit the growth rate. This depends on diffusion and convection.

Mechanisms

Studies on lysozyme crystals confirmed that macromolecular crystals share the same growth mechanisms as small molecule crystals [13]:

- *The bidimensional nucleation*: bidimensional ordered nuclei form onto the surface of the primary nuclei, forming platforms.
- *The dislocation spiral growth*: a small dislocation in the crystal propagates as a spiral.
- *The tridimensional growth*: adsorption of tri-dimensional aggregates or small crystals at the surface of a larger crystal. This occurs at high supersaturation levels and often results in a high level of irregularities in the growing crystal.

Random addition of single units onto the surface of the growing crystal would be possible only at high saturation rates since the absence of any irregularities infers a higher activation energy. However, at such supersaturation levels, the three-dimensional growth is energetically favored and the probability of formation of irregularities is higher, explaining why such a crystallization mechanism is not seen in practice.

9.4.2 Growth arrest

The arrest of crystal growth can be due to:

- thermodynamic equilibrium: progressively forming crystals make the system reach the solubility curve where the concentration of soluble proteins equals the solubility. As many protein molecules pass from the soluble phase to the crystal phase as proteins detach from the crystal.
- poisoning by impurities or irregularities. An arrest in the ordered arrangement of the crystal lattice prevents the packing of additional proteins. To resume growth, the irregular surfaces must be dissolved first.

Crystal growth and crystal quality

For good X-ray diffraction data, protein crystals must be very well ordered to obtain a nice scattering pattern, and of a sufficient size to have a detectable signal.
Low supersaturation levels will permit only slow growth mechanisms and thus enables to reach a maximum degree of order in crystal structure. Besides, if the supersaturation level is too high, the nucleation stage will result in many nuclei. But, if too many nuclei are formed during the nucleation process, less protein will be available for growth and therefore the crystals obtained till thermodynamic equilibrium will be smaller.
Therefore, to obtain the best quality and size of crystals, the supersaturation must be reduced to a relatively low level [10].

9.5 Managing supersaturation

9.5.1 From solubility to supersaturation: the vapour diffusion method

To go from a soluble and stable solution to a supersaturated solution, protein solubility must be decreased slowly. In protein crystallization experiments, the most widely used technique is evaporation through the vapour diffusion method, the basis for the hanging-drop and sitting-drop assays [10].

Here is the protocol followed for crystallization studies of all hand-made assays of the present study. 1 μL of the protein solution was added to a 1 μL-drop of the crystallization liquor hanging on the underside of a microscope cover slip. The reservoir solution contained 1 mL of crystallization solution. At the starting point of such an experiment, the protein drop contains twice less precipitant than the reservoir (see Figure 9.5). Over time, water vapour diffuses from the drop to the reservoir toward thermodynamic equilibrium, increasing the concentration of both the protein and precipitant in the drop. Progressively, the protein enters in supersaturation state, the first requirement for crystallization.

Protein solubility, and thus the surpersaturation curve driving crystallization depends on a myriad of factors such as pH, temperature, solvants and ionic strengths, counter ions,

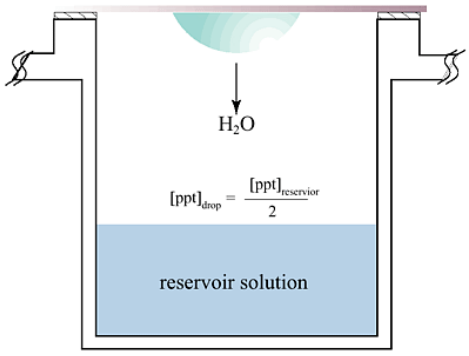


Figure 9.5: Hanging-drop technique. A well containing the reservoir solution is closed by a glass coverslip where a drop containing a 1:1 mix of reservoir solution with protein hangs.

additives, protein sequence, and potential contaminants. Here are listed the main factors and their major effects.

9.5.2 Effect of pH

At a solution pH that is different from the isoelectric point (pI)[4], the surface of the protein is predominantly negatively or positively charged, and proteins molecules are repelled. At the pI the net charge of the protein is zero, repulsive electrostatic forces are thus reduced, enhancing aggregation and precipitation processes.

Though generally the point of least solubility is the isolectric point, other minima can be seen, which may result from pH-dependent conformational changes and are usually unpredictable. In addition, at high ionic strength, the isoelectric point may be different due to counter-ions which may give rise to a net charge at the isoelectric point of the protein.

9.5.3 Effect of temperature

The effect of temperature on protein solubility is highly dependent on the protein sequence and composition of the buffer. Indeed, some proteins display an increased solubility with higher temperature, while other proteins will present a lower solubility with an increased temperature as illustrated Figure 9.6 [9] [14] [15].

Protein stability and sensitivity to protease contamination must be regarded as well when looking at consequences of temperature change.

9.5.4 Kosmotropic and chaotropic effects

Chaotropic and kosmotropic effects are thought to correspond to the ordering of water molecules. Under this perspective, a chaotropic ('chao-tropic' = order-breaking) effect refers to the apparently correlating property of decreasing the ordering of water due to a low interaction with water molecules. For instance, sodium chloride is a chaotrope. By contrast, a kosmotropic effect is associated with an increase in water structuring as it strongly interacts with it. The most used kosmotropes in protein crystallization are ammonium sulfate and

[4]pH of a solution at which the net primary charge of a protein becomes zero.

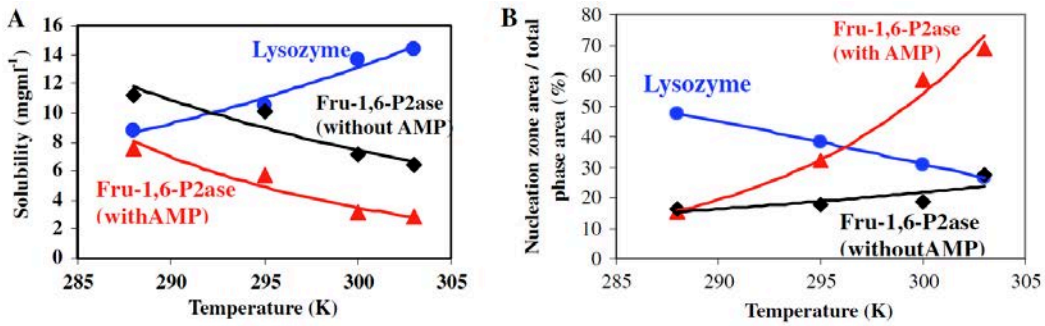


Figure 9.6: Effects of temperature on protein solubility (A) and nucleation (B). Three protein samples were studied: lysozyme, Fru-1,6-P2ase without or with AMP. The nucleation zone area / total phase area ratio was determined by experimentally [14].

PEGs. By consequence, when a chaotropic agent or a kosmotropic agent is added to the solution, a change in protein solubility is observed (see Figure 9.7). However, such properties may vary depending on the detailed composition of the protein sample, or even the temperature. Therefore, it is more accurate to talk about the kosmotropic or chaotropic effect of a reagent in the context of a specific system.

> Model for a Molecular Explanation of the Chaotropic and Kosmotropic Effects
>
> A chaotropic reagent breaks hydrogen bonds between water molecules, suppressing water ordering [16]. As a result, more water molecules are available for the protein hydration layer, with which the chaotropic reagent does not interact. Therefore, protein-protein interactions are minimized and the solubility is increased. In addition, a chaotropic solute preferentially interacts with protein macromolecules or salts rather than with water molecules [17], thus disrupting interactions mediated by hydrogen bonds, van der Waals forces, or hydrophobic effects, between or inside molecules. This causes protein denaturation.
>
> By constrast, kosmotropes are preferentially solubilized within the bulk of the solution by strongly interacting with water molecules (see Figure 9.7). Consequentially, they diminish the number of water molecules available for the hydration layer of proteins [18] and contribute to an increase in intramolecular protein interactions. However, if the effect is stronger, the resulting shrinking of the solvation layer will finally cause protein aggregation by exhibiting hydrophobic patches on the protein surface and increasing intermolecular protein interactions. Besides, a denatured protein exposes more hydrophobic patches than in folded state. More water molecules are thus needed for its solubilization. The free energy of denaturation will thus be higher in presence of a kosmotropic reagent as it contributes to reduce the amount of water molecules available for water-protein interactions. Therefore, kosmotropic reagents disfavour the denaturation process.

Nevertheless, the exact physical mechanism for the changes in water structure, and for the ability of such compounds to influence protein solubility in aqueous media is still under investigation [16].

9.5.5 Salting-in and salting-out effects

A graph showing protein solubility as function of ionic strength[5] is depicted Figure 9.8, revealing a bell-shaped curve which defines the *salting-in* and *salting-out* effects observed for

[5]The ionic strength is an important property of solutions, which affects the dissociation or the solubility of salts and proteins. It is a measure of the total concentration of ions in solution, and is defined by: $I = \frac{1}{2}\sum_{i=1}^{n} b_i.z_i^2$, where b_i is the molality of ion i, z its charge number, and n the total number of ions in solution.

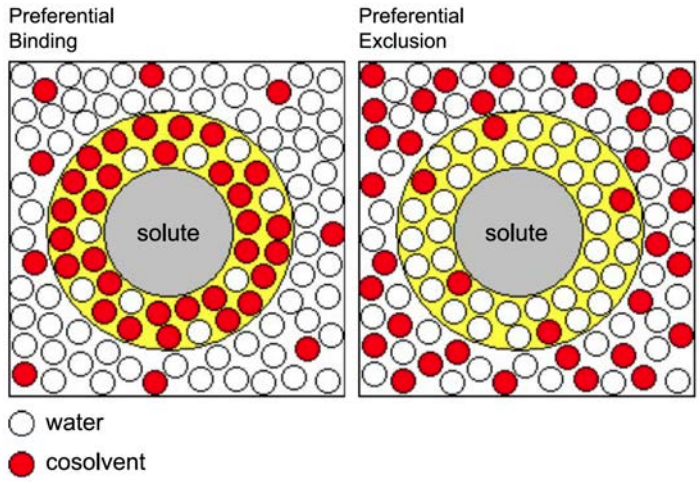


Figure 9.7: Chaotropic and kosmotropic effects. Schematic representation of preferential phenomena in a mixture of water and hydrophobic solute particles in the presence of chaotropic (left) and kosmotropic (right) cosolvents [16].

protein solutions. Both provide opportunities for the creation of supersaturated macromolecular solutions and crystal growth [19]. These effects are additive over all ions in solution though anions have larger effect than cations. An effort was made in this paragraph to establish a parallel with kosmotropic and chaotropic effects.

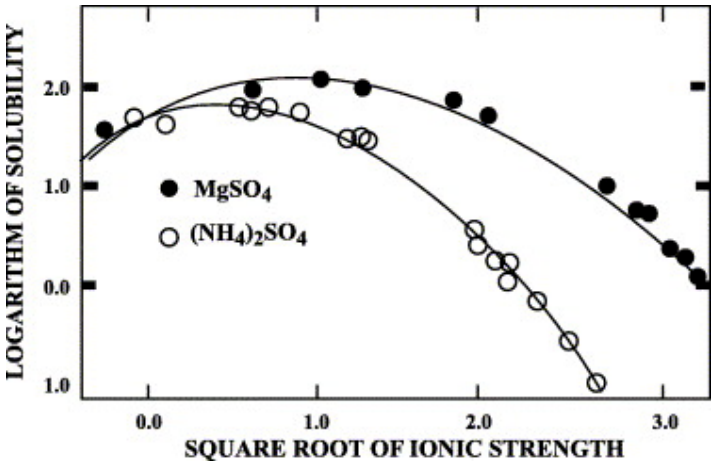


Figure 9.8: Salting-in and salting-out effects. The solubility of Enolase is shown here as a function of ionic strength produced by two different, common salts. The left (respectively, right) side of the bell-shaped curves is called *'salting-in'* region (respectively, *'salting-out'* region) [19].

Salting-in

On the left of the maximum of the curve is the *'salting-in'* zone where the dielectric constant[6] raises while the concentration in salt increases. The more polar and larger the solute, the higher the effect. This results in the charges on the protein surface interacting better with the environment, and therefore in a higher solubility.

However, the transition from salting-in to salting-out is not understood at the molecular level.

[6]The dielectric constant or relative static permittivity measures the solvent's ability to reduce the field strength of the electric field surrounding a charged particle immersed in it. It is a measure of its polarity.

Salting-out

On the right side of the curve, the *'salting-out'* zone, water molecules in the solvation layer are brought back into the bulk phase through interactions with the added salt, as described for kosmotropic reagents. This results in a reduction of the solvation layer around the protein, and therefore first in stabilization of native intramolecular protein interactions, and for stronger effects in the exhibition of hydrophobic patches on the protein surface, leading to protein aggregation.

Hofmeister series

Generally, ions follow the Hofmeister series, which rank them regarding their ability to 'salt out' or 'salt in' proteins (see Figure 9.9). As a general rule, attraction between proteins increases with ionic charge density.

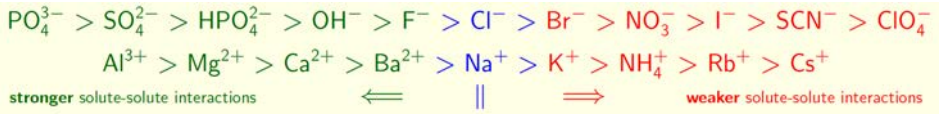


Figure 9.9: Hofmeister series [20].

Large, singly charged ions, with low charge density exhibit weaker interactions with water than water with itself and thus interfere little in the hydrogen bonding of the surrounding water. Therefore, they are chaotropes. They increase the solubility of nonpolar molecules ("salting in") and decrease the order in water; indeed, they weaken the hydrophobic effect.

By contrast, small or multiply-charged ions, with high charge density, are kosmotropes. They exhibit stronger interactions with water molecules than water with itself and therefore are capable of breaking water-water hydrogen bonds. Early members of the series increase solvent surface tension and decrease the solubility of nonpolar molecules ("salting out"); in effect, they strengthen the hydrophobic interaction.

Counter-ions

Solvent counter-ions interact with charged surface residues on the protein. In addition, as the protein is charged (for example negatively charged at a pH above its isoelectric point), a gradient of counter-ions is formed when moving away from the protein, compensated by an inversed gradient of co-ions. Such a counter-ions layer acts as a screen for ionic forces between protein units, resulting in a higher solubility. However, such counter-ions are rarely seen in protein crystal structures, suggesting that they are not well ordered in the crystalline lattice.

In addition, ions may also interact directly with proteins through polar interactions and may even bind them specifically[7]. This may modify the effect of counter-ions on protein conformation and solubility.

[7]For instance, Magnesium is often found in Mur ligases structures, as described in the Introduction Part.

9.5.6 Effect of organic solvents and polymers

Such as ions, by modulating the hydration layer around protein molecules, organic solvents alter protein solubility. Indeed, they lower the dielectric constant, thus increasing the coulombic attraction between unlike charges on the protein molecules and lowering the solubility [9].

However, organic solvents often denature proteins. Therefore, organic polymers such as PEGs (polyethyleneglycol) are very often preferred in protein crystallization assays. In addition, due to their polyhydroxy composition, they have a typical kosmotropic effect.

9.5.7 Effect of protein amino acid composition

The solubility of proteins depends on the distribution of hydrophilic and hydrophobic amino acid residues on the protein surface. On one hand, charged and polar surface residues interact with ionic groups in the solvent and increase solubility. On the other hand, the presence of hydrophobic patches on the surface of the protein is responsible for a lower solubility. In addition, repulsive or attractive forces may exist between proteins through permanent and induced dipoles, influencing protein solubility.

9.6 Crystals of protein-protein complexes

9.6.1 Still a new research field

The crystallization of protein-protein complexes remains a new field. Indeed, while all biocellular processes rely on protein-protein interactions, the number of structures of complexes deposited on the Protein Data Bank [21] is very small. A study of Sergei Radaev *et al.* published in 2006, reported only 659 unique dissociable protein-protein complexes[8]. Most of them are complexes between two proteins including a very small protein.

Yet, getting the structure of protein complexes allows not only a confirmation or discovery of an interaction but also a clear understanding of protein-protein interactions at the molecular level, which represents invaluable information for further research and development of new drugs.

9.6.2 Crystallization of protein-protein complexes

If crystallization of proteins is hard to control, crystallization of protein-protein complexes is even more arduous. Indeed, the crystallographer has to find a condition in which first the interaction is stable enough and second the crystallization is favored for the complex entities and disfavored for the single entities. Therefore, getting the crystal structure of a protein complex is long work and the more stable the complex the higher the chances to crystallize it. That explains why there are so few published structures of protein-protein complexes to date.

[8]including antibody-antigen complexes, cellular protein complexes, enzyme-inhibitor complexes, receptor-ligand complexes, signal transduction complexes, large multicomponent protein complexes such as ribosomes and other types of protein-protein complexes.

In addition, while the space of possible crystallization conditions of single proteins can be very broad - some proteins may crystallize at very high concentration of salts, precipitants, or extreme values of pH, for protein-protein complexes this space is much more limited as extreme pH values or salt concentrations may destabilize protein-protein interactions. Sergei Radaev *et al.* noticed that most published structures of protein complexes were found to crystallize within a less broad range of pH, and lower precipitant concentrations [22].

9.6.3 Protein-protein interactions and crystal packing

Moreover, once the structure is solved, protein-protein interactions due to crystal packing can be seen as well, and additional biochemical or *in vivo* work may be needed to determine which interactions are physiologically relevant, as van den Ent and co-workers did for the MreB-RodZ structure [23].

9.7 From theory to practice

When a protein is solubilized, molecules of solvent interact with surface amino acid residues thus forming a layer around the macromolecule and decreasing protein-protein attractive forces. The purpose of the crystallization reagents is to slowly reduce the hydration layer during the crystallization experiment to allow protein units to regularly arrange into a crystal lattice.

9.7.1 The paradox nucleation versus growth

While a high level of supersaturation will favor the initiation of crystallization (the nucleation process), a low level of supersaturation is needed for growth of good-quality crystals. However, the solution conditions that favor the nucleation are not always the same conditions that favor the subsequent growth. crystallization is then a matter of balance between favoring nucleation or growth, a delicate problem linked to the kinetics of evolution of the phase diagram during the experiment.

To crystallize a protein, the protein solution must be first pushed into the nucleation zone of the phase diagram (refer to Figure 9.1) where nucleation occurs and stay long enough (longer than the induction time, see section 9.3.1) to allow nuclei to appear. Nevertheless, maintaining the protein solution within the nucleation zone for too long will result in rapid appearance of too many crystal nuclei that will not grow much. Therefore, once a few nuclei are formed, the protein solution must leave the nucleation zone to stop the formation of additional nuclei, and stay in the metastable zone as long as possible where supersaturation allows crystals to grow.

9.7.2 Finding a crystallization condition: high-throughput screens

The two different processes of crystal nucleation and growth are often difficult to individually control because of the high number of parameters influencing protein solubility, and the

overall process of crystallization is even more ardeous to control because of its intrinsic duality.

Moreover, several factors are known to inhibit or thwart crystallization, such as temperature variations, vibrations and shocks, contaminants, conformational flexibility which may disturb the order in the growing crystal.

Because of the complexity of the crystallization process and the impossibility to control all the described parameters, to date, crystallization is still unpredictable at first. Therefore, the most common strategy to find suitable crystallization conditions is to screen the multi-parameter space influencing the solubility of the protein for the best supersaturation condition. Then, basic principles of crystallization are applied to optimize and scale up the experiment.

Having failed to crystallize a given macromolecule with any precipitant and additive, the crystallographer may first try to play with temperature and concentrations. A modification of the purification protocol of the target protein can help crystallization as well. However, if the protein still does not crystallize, a slightly different version of the molecule can be assessed, for instance after a limited proteolysis, or after removing flexible parts of the protein, or mutating a few residues. Indeed, even small changes in protein sequence can lead to large differences in solubility and protein-protein solubility, resulting in strong effects on crystallization behavior.

9.8 Experimental procedures

9.8.1 Sample preparation

After purification, pure proteins were concentrated with a Vivaspin centricon MWCO 30,000 to a concentration slightly above the desired one. The protein was then mixed to other reagents (see below) and diluted if needed in the purification buffer to achieve the final concentration, for a total volume of 100 μl. Samples were spun down for 5 min at 11,000 rpm at 4°C before setting of crystallization drops. 24-well crystallization plates from Hampton were prepared at room temperature and incubated at 20°C.

9.8.2 High-Throughput Screening

Screening assays for crystallization conditions were performed at the HTX lab of EMBL (Grenoble, France) with a nano-robot Cartesian PixSys 4200 crystallization robot (Genomic Solutions, U.K.) making 100 nl sitting-drops and using the six 96-well Greiner CrystalQuick plates (flat bottom, untreated):

- The Classics: a sparse matrix and ionic sampling screening strategies from Qiagen/Nextal
- Crystal Screen Lite PEG/Ion: a primary sparse matrix and primary or secondary, polymer, salt and pH matrix from Hampton Research
- MembFac Natrix: a primary sparse matrix for membrane proteins, samples with limited solubility and for nucleic acids protein/nucleic acid complexes from Hampton Research

- QuickScreen Grid screens "Ammonium Sulfate, Sodium Malonate - Sodium Formate": Primary or secondary, salt, polymer, organic and pH grid Home made, reagents from Hampton Research
- Grid screens PEG 6K, PEG/LiCl, MPD - Screen MME: Primary or secondary, salt, polymer, organic and pH grid - Home made with reagents from Hampton Research
- Index Screen: Primary, diverse reagent system crystallization screen for proteins, complexes, peptides, nucleic acids, water soluble small molecules from Hampton Research

The best conditions were reproduced and possibly refined to the following conditions, with the hanging-drop method as described earlier. Here below are the crystallization conditions and crystal shapes which gave rise to the best protein crystals (Table 9.1).

9.8.3 Optimized conditions for single proteins

Table

Protein	**Purification buffer**	**Ligand**	**Crystallization liquor**	**Growth time**	**Crystal shape**
SeMet-MurD 6-10 mg/ml	25 mM Hepes pH 7.4, 150 mM NaCl, 1 mM EDTA, 2 mM DTT	5 mM AMPPNP, 1 mM UMA	2 M Ammonium Phosphate	3 weeks	Plates, polycrystalline
SeMet-MurE 7-10 mg/ml	25 mM Hepes pH 7.4, 150 mM NaCl, 1 mM EDTA, 2 mM DTT	5 mM AMPPNP	5.5 M Sodium Formate, 0.1 M Sodium Acetate pH 5.5	5 days	Flowers of swords or fatty hexagonal sticks
SeMet-MurF 10-13 mg/ml	25 mM Hepes pH 7.4, 150 mM NaCl, 1 mM EDTA, 2 mM DTT	5 mM AMPPNP	0.1 M Sodium Acetate pH 5.5, 45-49% (v/v) MPD	1 week	Needles, polycrystalline
MurF 10-13 mg/ml	25 mM Hepes pH 7.4, 150 mM NaCl, 1 mM EDTA	5 mM AMPPNP	0.1 M Sodium Acetate pH 5.5, 45-49% (v/v) MPD	3 days	Needles, polycrystalline

Table 9.1: Crystallization conditions for single proteins. Crystals of Mur ligases were grown with the hanging-drop vapor diffusion technique at 20°C.

9.8.4 Towards the crystallization of a MreB-Mur ligase complex

For co-crystallization assays of Mur ligases with MreB, purified proteins were mixed with a 1:1, 1:2, or 2:1 molar ratios, then concentrated with a Vivaspin 500 μl 10,000 MWCO in order to reach a final concentration of MreB of 3 mg/ml (unless otherwise mentioned). 5 mM AMPPNP were added.

High-throughput screening

100 μl of samples prepared as described above were sent to the HTX lab (EMBL) for high-throughput screening.

Optimization

Interesting conditions from the HTX screen were reproduced, scaled-up, and optimized on the crystallization platform of IBS by Delphine Blot in 96-well Crystal Quick HTS plates (round bottom) with 100 μl reservoir and drops made of 1 μl protein solution and 1 μl crystallization liquor, using a Tecan Genesis RSP 100/8 robot. Finally, the best conditions were reproduced by hand (see the Results part for the composition of crystallization liquors).

SeMet crystals

SeMet MreB and MurE mixes were prepared as described above, but in presence of 2 mM DTT, before setting up crystallization plates.

Heavy atoms

For soaking assays with heavy atoms, salts of heavy atoms were dissolved at 500 mM in water or DMSO if not soluble in water, and then diluted at 10 mM in the crystallization liquor. 0.5 μl of these heavy atom solutions were added to the 2 μl drops of crystallization. A few hours later, 0.5 μl were added, and so on and so forth, until a concentration of 5 mM was achieved in the crystallization drop.

For co-crystallization assays with heavy atoms, protein samples were prepared with 5 mM of heavy atom but no heavy atom was added into the reservoir. Plates were set up as usual.

Chapter 10

Crystallography

This chapter first introduces properties of protein crystals and their preparation for X-rays scattering experiments. Then, basics of the theory of crystal scattering are exposed[1] and experimental procedures followed during this work are detailed before that basics of data processing techniques are explained, focussing on the methods that were used during this Ph.D. Lastly, the sequence of software used for structure solution are mentioned.

10.1 Protein crystals

This section addresses major definitions about protein crystals and basic symmetry properties.

10.1.1 Unit cell, lattice, and asymmetric unit

Unit cell

A crystal is basically a periodic arrangement of unit cells stacked in three dimensions. In other words, a unit cell contains all of the structural and symmetry information needed to generate the whole crystal with discrete translation operations.

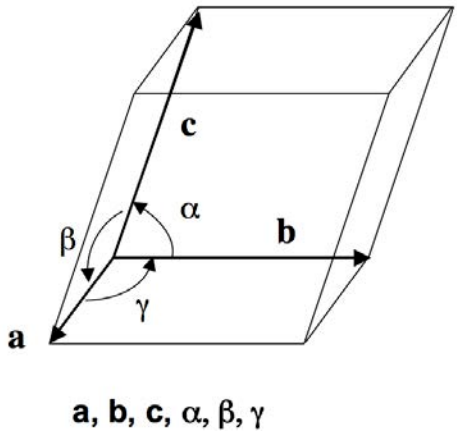


Figure 10.1: Unit cell [25].

[1]The reader is invited to look at the Crystallography Course of the University of Cambridge [24].

The unit cell can be described through a vector notation with the primary vectors **a**, **b**, and **c**, or by six parameters: the three lengths of the unit cell edges (a,b,c), and the angles between them (α, β, γ)[2]. By convention, the unit cell is chosen as the smallest unit of volume that can virtually reproduce the crystal by translations. The position of an atom j in the unit cell is given by its position vector $\mathbf{r_j} = x_j\mathbf{a} + y_j\mathbf{b} + z_j\mathbf{c}$ where x_j, y_j, and z_j are fractional coordinates of the crystal axes **a**, **b**, **c**.

Bravais lattice

The array of points which is generated by translations $\mathbf{t} = n\mathbf{a} + m\mathbf{b} + p\mathbf{c}$ where n, m, p are any integers, is called the Bravais lattice. By symmetry, each lattice point has the same environment as any other lattice point. Thus, a crystal can be viewed as a periodic arrangement of basic units repeated at each lattice point[3].

Asymmetric unit

Within the unit cell is the asymmetric unit, the smallest unit of volume that contains all of the structural information and that can reproduce the unit cell, and thus the whole crystal, by application of the symmetry operations given by the space group (both translations and rotations). Thereupon, each asymmetric unit contains the same environment as any other asymmetric unit. It is worthwhile mentioning that the choice of the asymmetric unit is not unique.

Files downloaded from crystallographic databases such as the Protein Data Bank (PDB) [21] generally contain coordinates for the atoms within the asymmetric unit.

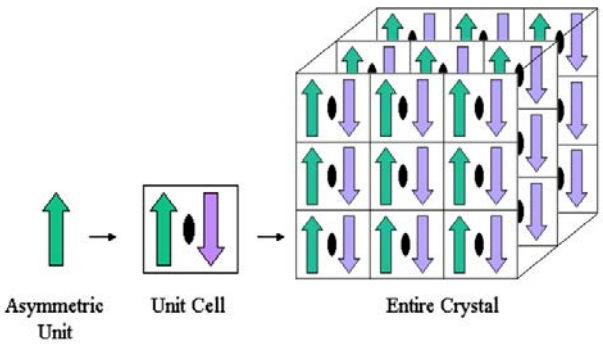


Figure 10.2: Asymmetric unit, unit cell, and lattice [21].

[2] α refers to the angle between the **b** and **c** vectors, β is the angle between **a** and **c**, and γ between **a** and **b**.

[3] Primitive unit cells contain only one lattice point, which is made up from the lattice points at each of the corners. Non-primitive unit cells contain additional lattice points, either on a face of the unit cell or within the unit cell, and so have more than one lattice point per unit cell.

10.1.2 Symmetry and space groups

As proteins are chiral molecules, mirror planes and inversion centers do not exist in protein crystals, and only rotations and screw axes are allowed.

The various possibilities in terms of crystal symmetry are described by the space groups. There are 65 chiral space groups that can describe protein crystals.

The Hermann-Mauguin notation is generally used to describe the symmetry of protein crystals. The first symbol - a letter, describes the centering of the Bravais lattice. The next three symbols describe the symmetries visible when projected along one of the high symmetry directions of the crystal: The screw axis is noted by a number with a subscript, n_m such that the angle of rotation is $360/n$ and the degree of translation is m lengths of the lattice vector. These four symbols contain all the symmetry information needed to reconstitute the complete crystal with the asymmetric unit. Here are a few examples of space groups:

- $P1$: no symmetry, except for the lattice translations along the **a**, **b**, and **c** axes;
- $P2_12_12_1$: three perpendicular two-fold screw axes, with a translation of one unit cell between each rotation;
- $P6_1$: six-fold screw axis in one direction, with a translation of one unit cell between each rotation.

Crystals and Symmetry

Symmetries of crystals can be divided into two types: Translations of the unit cell plus symmetry operations between asymmetric units within the unit cell. Translation operations of the unit cell define the crystal lattice. The space group of a crystal gives all the symmetry information needed to generate the complete crystal if the composition of the asymmetric unit is known.

10.2 X-ray scattering by protein crystals

10.2.1 Properties of X-rays and scattering by crystals

To produce significant diffraction, the spacing between the scatterers and the wavelength of the incident wave must be of the same order of magnitude. Atomic details of a protein such as covalent chemical bonds and atomic radius are about 1 Å in length. Therefore, to detect protein details through the diffraction technique, the incident electromagnetic wave must be in the range of 1 Å. This corresponds to X-rays[4].

However, the scattering signal from a single protein or from an aperiodic set of molecules would be too weak to allow the resolution of the structure at the atomic level. Fortunately, the periodic arrangement of identical units within a crystal allows the production of coherent interference of the beams scattered by every crystal unit, resulting in a strong signal that can be easily detected and then computed to solve protein structures.

[4]X-rays have wavelengths in the range of 0.01 nm to 10 nm.

10.2.2 The theory of X-ray scattering

In X-ray scattering experiments for protein crystallography, a beam strikes the protein crystal, producing a scattered wave which hits the detector, resulting in pattern of reflection spots.

After a brief introduction to electromagnetic waves, this section shows that X-ray scattering of protein crystals may be described as a summation of all the diffused waves, each scattered by one electron in the crystal [10]. Then, the relationship between diffraction data and electron density distribution in the crystal is given.

Electromagnetic waves

The incident wave is regarded as a monochromatic plane wave. It can be written as follows:

$$\Psi(\mathbf{x}, t) = \Psi_0 \exp\left[i(\omega t - \mathbf{k}.\mathbf{x}) + \Phi_0\right]$$

where Ψ_0 is the amplitude of the wave, $\omega = \frac{2\pi c}{\lambda}$ its angular frequency with λ its wavelength and c the light velocity, $\mathbf{k}$ the wave vector with $||\mathbf{k}|| = \frac{2\pi}{\lambda}$, Φ_0 its phase at the spatial origin, t the time, and $\mathbf{x}$ the position vector. Φ_0 will be set to 0.

Scattering by electrons

The electric field of the incident wave exerts a force on the electrons which makes them accelerate. In turn, the electrons emit radiation with a wave vector $\mathbf{k}'$: The wave is scattered [10] [26]. For X-ray crystallography experiments, the diffusion can be considered as elastic: the emitted wave has the same frequency as the incident wave, and $||\mathbf{k}|| = ||\mathbf{k}'||$.

Scattering by atoms

It can be modeled by a classical theory: the Thomson model. In this model, the diffusion cross section[5] of a scatterer is inversely proportional to its mass. Therefore, the contribution of nuclei to the scattered wave can be neglected. Besides, as the energy of X-rays is much greater than the energy of valence electrons, electrons can be regarded as free electrons to a good approximation[6] [10].

Under these hypotheses, every atom j in position $\mathbf{r_j}$ diffuses the incident wave, and the resulting wave in position $\mathbf{x}$ is given by:

$$\Psi_j^{atom}(\mathbf{x}, t) = \Psi_0 e^{i(\omega t + \phi(\mathbf{x}))}.f_j \tag{10.1}$$

where f_j is the atomic scattering factor, and $\phi(\mathbf{x})$ the phase shift of the scattered wave relative to the phase of the incident wave at its origin [27].

[5] The scattering cross section describes the likelihood of an electromagnetic wave to be scattered by a particle.

[6] However, this approximation cannot be assumed in anomalous scattering experiments. See Section 10.4.3.

Atomic Scattering Factor

The atomic scattering factor[a] is a measure of the amplitude of a wave scattered by an isolated atom.
For an atom with a spherical symmetry[b], the atomic scattering factor is defined by:

$$f(\mathbf{S}) = \int \rho(\mathbf{r}) e^{i\mathbf{S}.\mathbf{r}} \mathrm{d}^3\mathrm{r} \tag{10.2}$$

where $\mathbf{r}$ corresponds to polar coordinates around the center of mass of the atom, $\rho(\mathbf{r})$ is the electron density of the atom, and $\mathbf{S} = \mathbf{s} - \mathbf{s_0}$ with $\mathbf{s_0}$ and $\mathbf{s}$ the incident and scattered wave vectors by a system of two electrons, respectively.
Therefore, the atomic scattering factor is the Fourier transform of the electron density of the atom. It depends on the atomic number and incident angle of the beam.

[a] or atomic diffusion factor

[b] The hypothesis of a spherical distribution of electrons around the atom is good enough for most X-ray crystallography experiments. Particularly, it allows to consider the atomic scattering factor as real. Indeed, by centrosymmetry of the electron cloud, the imaginary part of every scattering vector will by compensated by its centro-symmetrical equivalent. However, when the wavelength becomes close to the absorption edge, this centrosymmetry of the atomic cloud is lost and the imaginary part of the atomic scattering factor must be considered (see Anomalous scattering section).

By incorporating the phase shift $\phi(\mathbf{x}, t) = \mathbf{K}.\mathbf{r_j} - \mathbf{k}'.\mathbf{x}$ with $\mathbf{K} = \mathbf{k}' - \mathbf{k}$ being the scattering vector, into equation 10.1, one obtains:

$$\Psi_j^{atom}(\mathbf{x}, t) = \Psi_0 e^{i(\omega t - \mathbf{k}'.\mathbf{x})} f_j . e^{i\mathbf{K}.\mathbf{r_j}} \tag{10.3}$$

Equation 10.3 shows that the scattering amplitude of X-rays increases with the atomic number.

Phase shift of the wave scattered by an atom
Indeed, the phase shift $\phi(\mathbf{x}, t)$ is made of two terms:

- The phase shift ϕ_1 of the incident wave at the position $\mathbf{r_j}$ of the atom j, relative to its origin: $\phi_1 = -\mathbf{k}.\mathbf{r_j}$;
- The phase shift ϕ_2 of the scattered wave between its origin $\mathbf{r_j}$ and position $\mathbf{x}$: $\phi_2 = -\mathbf{k}'.\mathbf{x} - \mathbf{r_j}$;

Scattering by a unit cell and structure factor

The kinematic theory of diffraction is considered here, meaning that the scattered waves do not scatter any more[7]. Under these conditions, the scattered wave for a unit cell (*u.c.*) l containing n atoms is the summation of the scattered waves of every atoms within the unit cell:

$$\Psi_l^{u.c.}(\mathbf{x}, t) = \sum_{j=1}^{n} \Psi_j^{atom}(\mathbf{x}, t) = \Psi_0 . e^{i(\omega t - \mathbf{k}'.\mathbf{x})} . \sum_{j=1}^{n} f_j . \exp(i\mathbf{K}.\mathbf{r_j})$$

with n the number of atoms in the unit cell.

One defines the structure factor[8] by[10]:

$$F(\mathbf{K}) = \sum_{j=1}^{n} f_j . \exp(i\mathbf{K}.\mathbf{r_j})$$

giving the scattered wave by a unit cell l:

$$\Psi_l^{u.c.}(\mathbf{x}, t) = \Psi_0 . e^{i(\omega t - \mathbf{k}'.\mathbf{x})} . F(\mathbf{K}) \tag{10.4}$$

[7] The kinematic theory is a good approximation for imperfect crystals as it is the case in real life where the coherence is not maximal.

[8] The structure factor depends on the protein structure in the unit cell, hence its name

Scattering by a crystal

A crystal is made of m unit cells. Likewise, for the unit cell l in $\mathbf{u_l}$ one has

$$\Psi_l^{u.c.}(\mathbf{x},t) = \Psi_0 e^{i(\omega t - \mathbf{k'}.\mathbf{x})}.F(\mathbf{K}).\exp(i\mathbf{K}.\mathbf{u_l})$$

Therefore, for the whole crystal the scattered wave in position $\mathbf{x}$ is given by[28]:

$$\boxed{\Psi^{crystal}(\mathbf{x},t) = \Psi_0 e^{i(\omega t - \mathbf{k'}.\mathbf{x})}.F(\mathbf{K}).S(\mathbf{K})} \tag{10.5}$$

where[9]

$$S(\mathbf{K}) = \sum_{l=1}^{m} \exp i\mathbf{K}.\mathbf{u_l}$$

10.2.3 Experimental procedures

Synchrotron radiation

When charged particles such as electrons are accelerated radially, an electromagnetic radiation is emmitted. When electrons are moving fast enough, they emit X-rays.

A synchrotron is basically made of four main components which are subsequently required for the production of powerful X-ray sources[29]:

- Linac: Here, electrons are produced, packed in 'bunches', and accelerated;
- Booster synchrotron: Electrons from the Linac are injected into the Booster synchrotron where they are further accelerated;
- Storage ring: A few times a day, the Storage ring is 'filled' with bunches from the Booster synchrotron. In this large ring, circulating electrons pass through magnets which modify their linear trajectory, and therefore produce X-rays with a wide and continuous spectrum;
- Beamlines: The X-ray beams emitted by the electrons are directed toward 'beamlines' where monochromators and filters select the proper wavelength and intensity for the experiment.

Cryofreezing and crystal mounting

A drawback of X-ray exposure is the inevitable formation of radical formation and resulting chemical reactions. This causes strong radiation damage of protein crystals if exposed for too long as during X-ray scattering experiments. In order to slow down the destructive process, the experiment is performed at a temperature of 100 K under a stream of cold nitrogen gas [10].

However, slowly freezing the crystals in an aqueous environment would result in ice formation in and around the crystal. As a result, crystal packing of protein units would be damaged as crystallized water expands into the protein crystal lattice. In addition, the presence of ice would cause a high background in scattering data. To prevent ice formation, crystals are transferred to a cryoprotectant solution, and rapidly cooled down within liquid

[9] This entity is named 'facteur de forme' in French as it depends on the crystal shape.

nitrogen. With such a procedure, the solvent freezes to a vitreous phase. Protein crystals are less damaged, though an optimization of cryo-solution may be required [10].

Then, crystals are mounted in a small loop made from a thin fiber of nylon or onto racket-like surface in Mylar[10] film [10], and stored in liquid nitrogen till X-ray scattering experiment.

As MurF crystallizes in MPD - a well-known cryoprotectant, MurF crystals were directly flash cooled in liquid nitrogen. MurD and MurE crystals were transferred into a cryoprotectant solution containing 75% of the crystallization liquor and 25% glycerol or 25 % MPD before cryocooling.

Data collection

Crystallographic data were collected on various beamlines at the European Synchrotron Radiation Facility (Grenoble, France) and Soleil (Saclay, France) synchrotrons. Detailed information about data collection is available in the article 'MreB and MurG as scaffolds for the cytoplasmic steps of peptidoglycan biosynthesis' published in *Environmental Microbiology* in 2013 and attached as Appendice in this thesis.

10.2.4 From intensities to electron density

Intensity

Intensity is the only quantity measured by detectors, as a function of the direction of the emitted wave. It corresponds to the power of a wave transferred per surface unit. For a monochromatic, spherical, wave such as the scattered waves, the intensity is proportional to the square of the amplitude of the scattered wave vector:

$$I(\mathbf{x}, t) \propto ||\Psi^{crystal}(\mathbf{x}, t)||^2$$

taking into account a corrective factor due the $\frac{1}{r^2}$ attenuation of the wave.

In an X-ray scattering experiment, the crystal of volume V_{cr} is rotated at an angular velocity ω and exposed with an incident beam of intensity I_0 and wavelength λ. The intensity can be viewed as a function of the scattering vector, and is basically proportional to the square of the amplitude of structure factor:

$$I(\mathbf{K}) \propto \frac{\lambda^3}{\omega.V^2}.V_{cr}.I_0.L.T.P.A.|F(\mathbf{K})|^2 \tag{10.6}$$

where V is the volume of the unit cell, L is the Lorentz factor which corrects effects of volume and depends on the geometry of the diffraction experiment, P is the polarization factor as the scattered wave is polarized [10], A is the absorption factor, and T is the temperature factor due to oscillation of atoms in the crystal about their average position (see Section 10.5.1).

[10]Mylar is a commercial name for biaxially-oriented polyethylene terephthalate, which presents a high transmission for X-rays.

Knowing the proportionality factors, an X-ray scattering experiment thus allows the measurement of the amplitudes of the structure factor $|F(\mathbf{K})|$.

Structure factor and electron density

The structure factor has been defined as:

$$F(\mathbf{K}) = \sum_{j=1}^{n} f_j . \exp(i\mathbf{K}.\mathbf{r_j})$$

with n the number of atoms j in the unit cell ($u.c.$).

A continuous expression can be used instead:

$$F(\mathbf{K}) = \int_{u.c.} \rho(\mathbf{r}) . \exp(i\mathbf{K}.\mathbf{r}) \mathrm{d}^3\mathrm{r}$$

with $\rho(\mathbf{r})$ the electron density at position $\mathbf{r}$ in the unit cell. Therefore, the structure factor is the Fourier transform of the electron density.

X-RAY SCATTERING

In an X-ray scattering experiment of a protein crystal, the incident electromagnetic wave interacts with electrons of the crystal, resulting in a wave scattered in different directions. The intensity of the scattered wave in a given direction is proportional to the square of the structure factor. This entity appears to be the Fourier transform of the electron density. Thus, diffraction can be regarded as a Fourier transformation.

10.3 Diffraction conditions and indexation

10.3.1 Laue conditions

A peak in a scattering pattern corresponds to a maximal intensity, i.e. a local maximum of the structure factor F as a function of the scattering vector $\mathbf{K}$. This occurs when diffused waves described by the diffusing vector $\mathbf{K}$ from every atoms are in phase, meaning $\mathbf{K} = \mathbf{H}$, with $\mathbf{H}$ such that:

$$\mathbf{H}.\mathbf{r_j} \equiv \mathbf{H}.\mathbf{r_l} \mod 2\pi \tag{10.7}$$

for any atom pair (i, j). The set of scattering vectors $\mathbf{H}$ following equation 10.7 are a subset of the diffusing vectors $\mathbf{K}$.

Let us consider $(\mathbf{e_1}, \mathbf{e_2}, \mathbf{e_3})$ the basis of the lattice. Thus, the position vector for an atom j is:

$$\mathbf{r_j} = x_j\mathbf{e_1} + y_j\mathbf{e_2} + z_j\mathbf{e_3} \tag{10.8}$$

where x_j, y_j, and z_j are positive numbers higher or equal to 1.

It can be proved that condition 10.7 infers that components of $\mathbf{K}$ in the lattice base are integers.

Let us define the base $(\mathbf{e_1^*}, \mathbf{e_2^*}, \mathbf{e_3^*})$ which defines the reciprocal space:

$$\mathbf{e_1^*} = \frac{\mathbf{e_2} \times \mathbf{e_3}}{V} \tag{10.9}$$

$$\mathbf{e_2^*} = \frac{\mathbf{e_3} \times \mathbf{e_1}}{V} \tag{10.10}$$

$$\mathbf{e_3^*} = \frac{\mathbf{e_1} \times \mathbf{e_2}}{V} \tag{10.11}$$

with V the lattice volume: $V = [\mathbf{e_1}, \mathbf{e_2}, \mathbf{e_3}]$. Interestingly, $\forall (i,j) \in [\![1;3]\!]^2$, $\mathbf{e_i}.\mathbf{e_j^*} = \delta_{ij}$.

Vectors $\mathbf{H}$ can be expressed in the reciprocal space:

$$\mathbf{H} = h\mathbf{e_1^*} + k\mathbf{e_2^*} + l\mathbf{e_3^*}, \tag{10.12}$$

$$\text{and} \tag{10.13}$$

$$\forall (i,j) \in [\![1;3]\!]^2, \mathbf{H}.\mathbf{e_i} = \mathbf{H}.\mathbf{e_i^*} \tag{10.14}$$

Therefore, one has diffraction spots if and only if components h, k, l of $\mathbf{H}$ in the reciprocal base are integers. This corresponds to the Laue diffraction condition[11].

10.3.2 Miller indices and indexation

The scattering vector $\mathbf{H}$ can thus be expressed in the reciprocal lattice with the integers h, k, l (see above). Thus, the structure factors and the electron density can be written as functions of (h, k, l) as well:

$$\rho(\mathbf{r}) = \frac{1}{V} \sum\sum\sum |F_{hkl}| \exp[i(\mathbf{H}.\mathbf{r}) + i\Phi_{hkl}] \tag{10.15}$$

Thereupon, the wave vectors give diffraction spots which correspond to scattering vectors[12] that can be indexed by Miller indices. In other words, each diffraction spot corresponds to a point in the reciprocal lattice and represents a wave with an amplitude and a relative phase.

The first step in data processing is indexing of reflection spots: Once the dimensions of the unit cell are determined, each peak is assigned to a position in reciprocal space by its Miller indices. From symmetry properties of the scattering pattern, the space group can be identified. Then, the data is integrated: intensity for each reflection is estimated in the 3D-space from the hundreds of diffraction images. This was performed using XDS software [30] for the work of this Ph.D.

[11] It can be shown that the Laue diffraction conditions are equivalent to the well-known Bragg conditions.
[12] Scattering vectors are diffusion vectors which follow the Laue conditions.

Scattering Pattern

When scattered waves are coherent, a signal can be detected as a diffraction spot. This is observed for scattered wave vectors which have integer components (Miller indices) in the reciprocal lattice. Thus, diffraction spots can be indexed by their Miller indices: this is the first step of data processing.

10.4 The phase problem

During a scattering experiment, photons are counted, giving intensities of diffraction spots. As seen in Section 10.2.4, this allows the measurement of the corresponding amplitudes of structure factors. However, the phases are not measured, and inverting the Fourier transform to derive the electron density is not possible. This is called the phase problem [24].

10.4.1 The Patterson function

Definition

The Patterson function is the inverse Fourier transform of the intensities [10] [31]. As scattered intensity by a crystal is a discrete function in the reciprocal space, a discrete notation can be used:

$$P(\mathbf{u}) = \sum_{\mathbf{H}} |F(\mathbf{H}).F^*(\mathbf{H})|e^{i(\mathbf{H}.\mathbf{u})} = \sum_{\mathbf{H}} |F(\mathbf{H})|^2 e^{i(\mathbf{H}.\mathbf{u})} \tag{10.16}$$

with $\mathbf{H}$ the diffraction vectors in the reciprocal space, $F(\mathbf{H})$ the structure factors, and $\mathbf{u}$ the position vector in the unit cell.

Besides, as seen above, structure factors are Fourier transforms of the electron density[10]:

$$F(\mathbf{H}) = \int_{\mathbf{r}} \rho(\mathbf{r}) e^{i\mathbf{H}.\mathbf{r}} \mathrm{d}\mathbf{r} \tag{10.17}$$

$$\text{and} \tag{10.18}$$

$$F^*(\mathbf{H}) = \int_{\mathbf{r}} \rho(\mathbf{r}) e^{-i\mathbf{H}.\mathbf{r}} \mathrm{d}\mathbf{r} = \int_{\mathbf{r}} \rho(-\mathbf{r}) e^{i\mathbf{H}.\mathbf{r}} \mathrm{d}\mathbf{r} \tag{10.19}$$

$$\tag{10.20}$$

According to the convolution theorem resulting from properties of Fourier functions, the Patterson function can be defined as the autocorrelation function of electron density as well:

$$P(\mathbf{u}) = \rho_{\mathbf{r}} * \rho_{-\mathbf{r}}(\mathbf{u}) = \int_{\mathbf{r}} \rho(\mathbf{r}).\rho(\mathbf{u}+\mathbf{r}) \mathrm{d}\mathbf{r} \tag{10.21}$$

In other words, the Patterson function can be regarded as the convolution of the structure $\mathbf{r} \to \rho(\mathbf{r})$ and its inverse $\mathbf{r} \to \rho(-\mathbf{r})$.

Properties

More concretely, the peaks in the Patterson function are the interatomic distance vectors weighted by the product of the number of electrons in the atoms concerned. A vector map of relative position vectors between each pair of atoms in the structure can be built: the Patterson map.

PATTERSON FUNCTION: FROM THE FORMULA TO THE INTERPRETATION OF THE PATTERSON MAP

Indeed, let us move within the crystal unit cell with a vector **u**. For every position of **u**, let us multiply the electron density at the beginning of **u** with the electron density at the end of **u** and take the integral of these values. Here we have the Patterson function [10]. The result will be nonzero only if nonzero electron density is present at both the beginning and the end of **u**. Therefore, the Patterson function is made of maxima which correspond to interatomic vectors in the real structure. The height of a peak is proportional to the product of the heights of the two peaks in the electron density map, i.e. it is determined by the atomic number of the atoms.

Use

Interestingly, as the Patterson function is defined by intensities only, it can be directly calculated from measured data.

For small molecules, phases can thus be derived from the Patterson map. However, for large molecules such as proteins, it is still impossible to directly derive the atomic positions from interatomic vectors of the Patterson map. Nonetheless, the Patterson map is widely used as a tool for solving the phase problem.

10.4.2 Molecular replacement

Molecular replacement is the most widely used technique to solve the phase problem. It relies on that phases of similar structures are correlated.

Principle

The first step in molecular replacement is to find a structure of a homolog which is expected to be relatively similar to the novel structure to be solved, or at least to parts of it [32]. Usually, 30 % sequence identity is a very minimum. The second step is to orient and localize the model within the unknown unit cell in order to have the best correlation between data computed from the oriented/translated model, and experimental data. To do so, most programs now use algorithms based on maximum likelihood, such as Phaser [33] that was used for solving structures of this work. However, molecular replacement was historically based on the Patterson function. For sake of simplicity, a brief explanation of this historical method is exposed in this section and allows an understanding of the essence of the molecular replacement method.

Once the similar protein is properly placed in the experimental unit cell, phases can be calculated from this model and then taken as first estimations of phases of the new structure.

The Patterson map as a basis for molecular replacement

An interesting feature can be derived from the definition of the Patterson map: two Patterson maps of structurally similar proteins should be very closely correlated if both proteins are in the same orientation and location. Therefore, the proper orientation of the model can be determined by maximizing the correlation between the experimental Patterson map of the unknown structure and Patterson maps obtained by different orientations and translations of the model.

To place one object in 3D space, 6 parameters are required: three rotation angles, and three translational parameters. The molecular replacement problem is thus a 6-dimensional problem if there is one molecule in the asymmetric unit [32]. However, it can be intuitively understood that intramolecular vectors, which depend only on the orientation of the molecule, will be in average shorter than intermolecular vectors in the Patterson map. Therefore, the model is first rotated using only the part of the Patterson map near the origin, to obtain a maximum fit between observed and calculated maps [32]. Then, the model is translated to find the correct position within the unit cell. In this way, molecular replacement solving is usually separated into two 3D problems: first, rotation and second, translation.

From the oriented model to phases of the new structure

Cartesian coordinates of the best orientations and locations of the structural model within the new unit cell are output in pdb files[13].

By artificially packing the model into the unit cell of the crystal to be analyzed, the structure factors F of the model in a similar crystal arrangement to the one of the unknown molecule can be calculated. Particularly, phases can be extracted, and then used to 'complete' experimental data for the unknown structure: An electron density map can be inferred from measured intensities and calculated phases according to the model.

This preliminary map of the new structure is then used to determine parts of the new structure which are then used to simulate a new set of phases. This process is reiterated untill the best correlation between calculated and experimental data is obtained (see Section 10.5).

Limitations

Though a very fast technique, molecular replacement presents a few important drawbacks:

- A very close model must be found. Particularly, it has been noted that molecular replacement often fails when differences in atomic positions between the model and the crystal structure are more than about 1 Å [32].
- In case of incomplete models (for instance, no structure of a homolog is known, or notable conformational changes are present, or a model is available for a part of the content of the unit cell only), the low correlation between a single copy of the model and the diffraction data may cause molecular replacement to fail.
- As phases play a major role in the structural information, structures solved by molecular replacement are often biased by the model used.

[13]PDB format is a standard file for representation of macromolecular structures.

In addition, despite several decades of experience with this method, the reasons for success or failure remain obscure [32].

10.4.3 Anomalous scattering

Normal diffusion and Friedel's law

Atomic electrons are shared between *'electron shells'* which can be imagined as orbits in which electrons would travel around the nucleus (Figure 10.3, Left). Each shell is composed of one or more subshells, which are themselves made of atomic orbitals. It is of note that atomic orbitals are centro-symmetric in the basal state (see Figure 10.3, Right) [34].

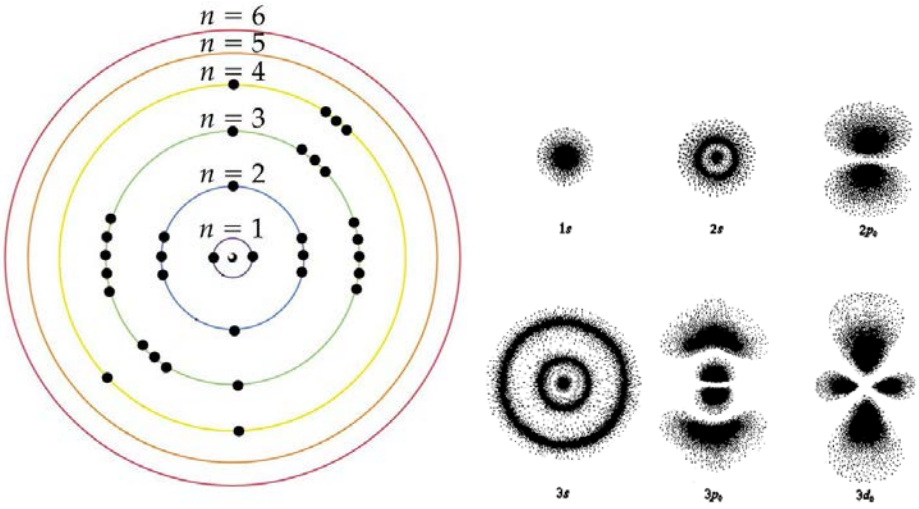


Figure 10.3: Schematic representation of selenium shells (left) and atomic orbitals (right) [34].

This centro-symmetry in the scattering atomic clouds is translated into the pattern of scattered X-ray intensities [34]. Indeed, for any space group and particularly for non centro-symmetric space groups such as protein space groups, X-ray diffraction data present centro-symmetry. This rule is known as the Friedel's law [35]: for any reflection (h, k, l), by defining $|F^+| = F(h, k, l)$ and $|F^-| = F(-h, -k, -l)$,

$$|F^+| = |F^-| \text{ and } \phi^+ = -\phi^- \tag{10.22}$$

Reflections (h, k, l) and $(-h, -k, -l)$ are called Friedel's pairs.

A mathematical view of Friedel's law

Friedel's law is in fact a property of Fourier transforms of real functions. Given a real function f, its Fourier transform:

$$\mathcal{F}(k) = \int f(x) e^{ikx} \mathrm{d}x$$

One can easily see that

$$\mathcal{F}(k) = \mathcal{F}^*(-k)$$

where $\mathcal{F}^*$ is the complex conjugate of $\mathcal{F}$.
In addition, the squared amplitude is centrosymmetric:

$$|\mathcal{F}(k)|^2 = |\mathcal{F}(-k)|^2$$

Likewise, as the structure factor is the Fourier transform of the electron density, its squared amplitude is centrosymmetric, hence Friedel's law.

Friedel's law is not restricted to Friedel's pairs and can be extended to Bijvoet pairs as well. Each member a and b of a Friedel's pair of reflections (a, b)[14] may have N and M true symmetry equivalents α_i and β_j, resp. (with i and j integers to identify the various symmetry equivalents)[15]. By symmetry:

$$\forall (i, j) \in [\![1; N]\!] \times [\![1; M]\!], \tag{10.23}$$

$$|F(\alpha_i)| = |F(a)| \text{ and } \phi(\alpha_i) = \phi(a) \tag{10.24}$$

$$|F(\beta_i)| = |F(b)| \text{ and } \phi(\beta_i) = \phi(b) \tag{10.25}$$

As Friedel's law states $|F(a)| = |F(b)|$ and $\phi(a) = -\phi(b)$, one has:

$$\forall (i, j) \in [\![1; N]\!] \times [\![1; M]\!], |F(\alpha_i)| = |F(\beta_j)| \text{ and } \phi(\alpha_i) = -\phi(\beta_j) \tag{10.26}$$

In other words, the pairs (α_i, β_j), known as Bijvoet pairs, follow Friedel's law as well[16].

Electron transition

Electron shells correspond to different energy levels of electrons in atoms (see Figure 10.4, top left). When excited by X-rays of a proper energy, core electrons can be excited to a higher level by absorbing a photon [36]. This phenomenon can be easily detected by plotting the absorption of an element as a function of the wavelength (Figure 10.4, bottom), revealing an *'absorption edge'*. The centro-symmetry of the electron cloud is disturbed (see Figure 10.4, top right), resulting in an anomalous scattering [34].

Atomic diffusion factor of an anomalous scatterer

As seen earlier, in the case of normal diffusion, the atomic diffusion factor f_0 for a spherical atom is real. By contrast, when an electronic transition occurs, the scattering atom looses its centro-symmetry and the imaginary part of the atomic scattering factor becomes nonzero.

[14] with $a = (h, k, l)$ and $b = (-h, -k, -l)$.
[15] For instance, 2-fold related reflections in $P2$ are true symmetry equivalents.
[16] A Friedel's pair is a Bijvoet pair, but a Bijvoet pair is not necessarily a Friedel's pair.

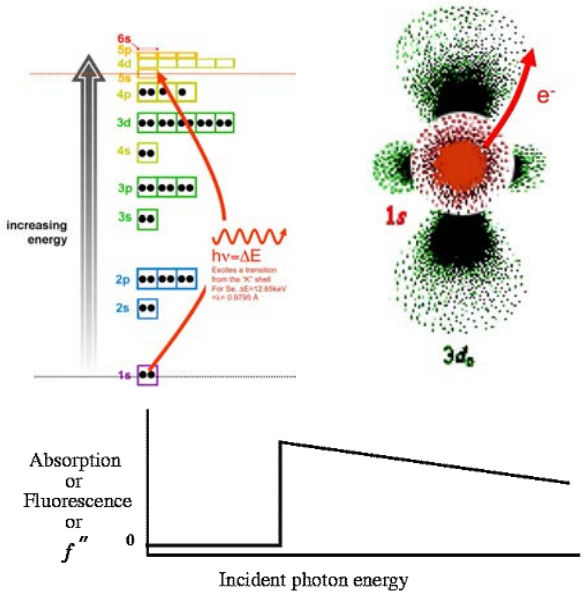


Figure 10.4: Electronic transition of core electrons of selenium excited by X-rays (top left) results in an absorption edge when plotting absorption versus energy [36] (bottom), and a loss of centro-symmetry in the atomic orbital [34] (top right).

Such an anomalous scatterer can be modeled by adding two corrective terms, based on an analogy of the atom to a forced oscillator under resonance conditions:

$$f^H(\omega, \mathbf{S}) = f_0^H(\mathbf{S}) + \Delta f'(\omega, \mathbf{S}) + i\Delta f''(\omega, \mathbf{S})$$

where $\Delta f'$ and $\Delta f''$ describe the change in intensity and phase of the diffusion factor from f_0, $\mathbf{S}$ is the diffusion vector, and $\omega = \frac{2\pi c}{\lambda}$ is the angular frequency of the wave[17] with c the light celerity.

Thereupon, the structure factor of a wave scattered by an anomalous scatterer H can be written as:

$$F^H(\mathbf{H}) = f^H(\omega).e^{i\mathbf{H}.\mathbf{r}} \tag{10.27}$$

$$\text{with} \tag{10.28}$$

$$f^H(\omega) = f_0^H + \Delta f'(\omega) + i\Delta f''(\omega) \tag{10.29}$$

with $\mathbf{H} = (h, k, l)$ the scattering vector in position $\mathbf{r} = (x, y, z)$.

A small component $\Delta f'$ of the scattered radiation, called the *dispersive component* [37] is π out of phase (real) with the normally scattered radiation given by f_0. By consequence, it always diminishes f_0 [34]. Another small component $\Delta f''$, called the anomalous component [37], is $\frac{\pi}{2}$ out of phase (imaginary) with the normally scattered radiation given by f_0 and modifies its phase. This correction of the atomic diffusion factor is illustrated on an Argand

[17] Since the absorption depends on the incident energy, the corrective terms in case of anomalous scattering depend on the wavelength.

diagram[18] Figure 10.5 and can be generalized to the expression of the wave scattered by the whole crystal.

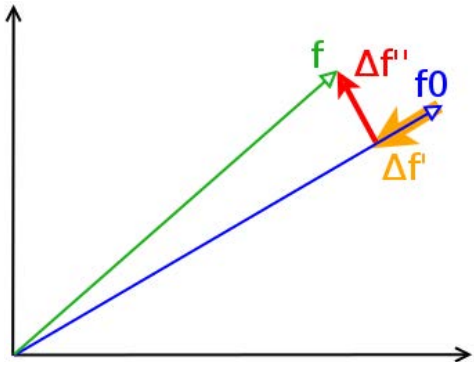


Figure 10.5: Argand diagram showing the contribution of the anomalous part of a scatterer in the atomic diffusion factor f^H.

Experimental determination of $\Delta f'$ and $\Delta f''$

The values of $\Delta f'$ and $\Delta f''$ can be determined experimentally by measuring the linear absorption coefficient μ[19] for the incident direction with the formulas [37]:

$$\Delta f'(\omega) = \frac{2}{\pi} \int_0^\infty \frac{\omega' \Delta f''(\omega')}{\omega^2 - \omega'^2} \mathrm{d}\omega' \tag{10.30}$$

$$\Delta f''(\omega) = \frac{mc\omega\omega}{4\pi N e^2} \mu\omega \tag{10.31}$$

where σ is the absorption cross section[20] of the atom, m the mass of the electron and e its charge.

Anomalous scattering infers breakdown of Friedel's law

When only anomalous scatterers are present, Friedel's law is not conserved for phases anymore, as illustrated Figure 10.6, left. When both non anomalous and anomalous scatterers are present, both the amplitude and phase relationships of the Friedel's law are broken (see Figure 10.6, right).

In protein crystals, carbon, nitrogen, oxygen, and hydrogen atoms do not contribute to anomalous scattering at X-ray wavelengths used for protein X-ray crystallography. However, if a heavy atom is present in the protein crystal and is well ordered, and if the wavelength is close to the absorption edge of the heavy atom, an anomalous signal is measured: intensities of Friedel's pairs (and Bijvoet's pairs as well) are not equivalent.

[18] Complex numbers such as structure factors can be conveniently represented on an Argand diagram where numbers are depicted as vectors in the plane using the x-axis as the real axis and y-axis as the imaginary axis.

[19] The linear absorption coefficient is defined by $I = I_0 e^{-\mu x}$ with I the intensity of the transmitted wave, I_0 the intensity of the incident wave, x the path length.

[20] The absorption cross section is a measure of the probability of an absorption process.

It is of note that the anomalous scattering phenomenon is roughly proportional to the number of electrons of the atom, as the amount of energy that is absorbed depends on the atomic number. Thus, the heavier the atom, the greater the anomalous signal.

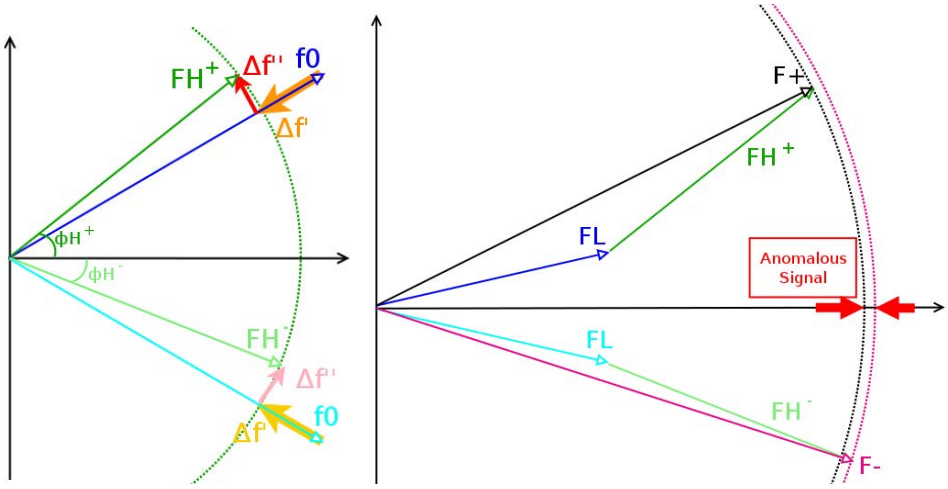


Figure 10.6: Argand diagrams showing the breakdown of Friedel's law in presence of anomalous scatterers [34]. Left: structure factors F_H^+ and F_H^- for reflections (h,k,l) and $(-h,-k,-l)$, respectively, by an anomalous scatterer H. Right: structure factors F^+ and F^- for reflections (h,k,l) and $(-h,-k,-l)$ by a protein made of both anomalous scatterers H and non-anomalous scatterers L [34].

Anomalous Patterson map: locating heavy atoms

As there are usually a limited number of anomalous scatterers (heavy atoms) per asymmetric unit in protein crystals, their position and phases can be determined by a Patterson method. A Patterson map is calculated with $\Delta|F_{ano}|^2$ called the anomalous difference:

$$\Delta|F_{ano}| = \frac{f_0 + \Delta f'}{2\Delta f''}\{|F^+| - |F^-|\}$$

It can be shown that

$$(\Delta|F_{ano}|)^2 = \frac{1}{2}|F^H|^2 - \frac{1}{2}|F^H|^2\cos\{2(\phi^+ - \phi^-)\}$$

Therefore, the anomalous Patterson map gives peaks corresponding to heavy atoms [10].

Phasing

Once the phases of anomalous scatterers are known, phases of the protein part of the structure factor can be derived.

A very convenient way to show this is the Harker construction. In this geometrical method shown Figure 10.7, the opposite of the structure factor of the heavy atoms F_0^H and its anomalous contributions $\Delta f'$ and $\Delta f''$ are drawn from the origin of a diagram. Then, two circles are drawn: one with its center corresponding to $+\Delta f''$, and a radius of F^+, and a second circle with its center corresponding to $-\Delta f''$ and a radius of F^-. The two circles intersect twice. One of these intersection indicates the structure factor of the protein F_L.

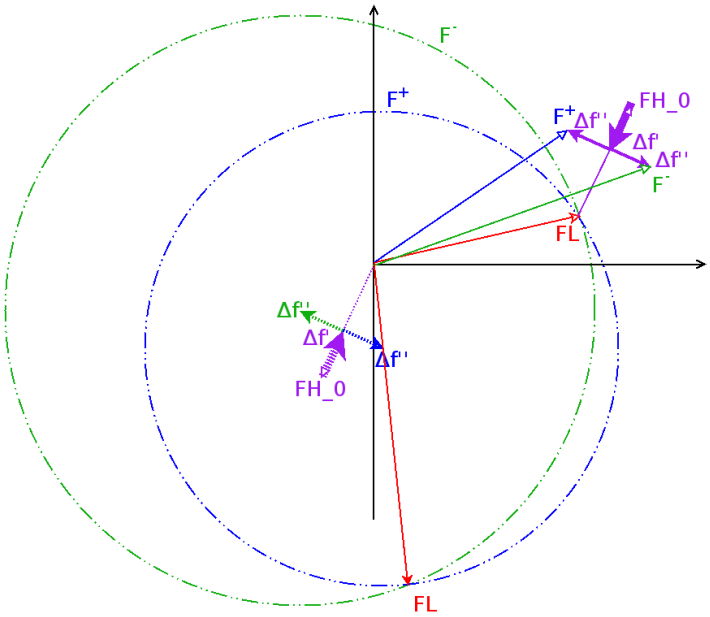


Figure 10.7: Harker construction for the determination of protein phases with anomalous signal.

Simple Mathematical View of the Relationship Between Anomalous Signal and Protein Phases.

For a protein P made of N light atoms and M heavy atoms of the same type H, the structure factor is [38]:

$$F(\mathbf{K}) = \sum_{j}^{N} f_j e^{i\mathbf{K}.\mathbf{r_j}} + \sum_{k}^{A} (f_{0k}^{H} + \Delta f_k' + i\Delta f_k'').e^{i\mathbf{K}.\mathbf{r_k}} \tag{10.32}$$

Let us define F^L, F_{0H}^H, $\Delta F'$, and $\Delta F''$ such that $F(\mathbf{H}) = F^L + F_{0H}^H + \Delta F' + i\Delta F''$, and $F_{Re} = F^L + F_{0H}^H + \Delta F'$. Interestingly, F_{PH} depends on the real parts of the atomic scattering factors only, and $F(\mathbf{H}) = F_{Re} + i\Delta F''$. The composition of $F(\mathbf{H})$ is depicted in the Argand diagram here below:

As usually proteins contain a few heavy atoms, one can assume $F_{Re} >> \Delta F''$, resulting in the following approximation:

$$F_{Re} \approx \frac{F^+ + F^-}{2} \tag{10.33}$$

with $F^+ = F(h,k,l)$ and $F^- = F(-h,-k,-l)$.
By defining the Bijvoet difference $\Delta_{ij} = F^+ - F^-$, the law of cosines for Figure 10.8 gives:

$$F_+^2 = F_{Re}^2 + \Delta F''^2 - 2F_{Re}\Delta F'' \cos\left(\Phi_{Re} - \Phi_H + \frac{\pi}{2}\right) \tag{10.34}$$

$$F_-^2 = F_{Re}^2 + \Delta F''^2 + 2F_{Re}\Delta F'' \cos\left(\Phi_{Re} - \Phi_H - \frac{\pi}{2}\right) \tag{10.35}$$

This can be written as follows:

$$F_+^2 = F_{Re}^2 + \Delta F''^2 + 2F_{Re}\Delta F'' \sin\left(\Phi_{Re} - \Phi_H\right) \tag{10.36}$$

$$F_-^2 = F_{Re}^2 + \Delta F''^2 - 2F_{Re}\Delta F'' \sin\left(\Phi_{Re} - \Phi_H\right) \tag{10.37}$$

Substraction of equations 10.36 gives [38]:

$$(F^+)^2 - (F^-)^2 = 4F_{Re}\Delta F'' \sin\left(\Phi_{Re} - \Phi_H\right) \tag{10.38}$$

$$\tag{10.39}$$

In other terms,

$$\boxed{\Delta_{ij} \approx 2\Delta F''_H \sin\left(\Phi_{Re} - \Phi_H\right)} \tag{10.40}$$

As the anomalous Patterson map allowed to obtain phases for anomalous atoms Φ_H, Equation 10.40 gives a first approximation of phases Φ_{PH} from which phases of the protein can be derived and used to calculate a first electron density map.

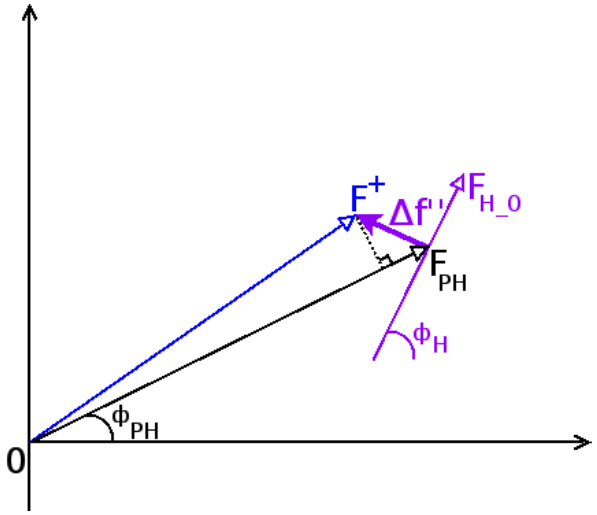


Figure 10.8: Bijvoet difference.

Experimental aspects

Structures of MurD, MurE, MurF, were solved by single anomalous diffraction (SAD) experiments, by acquisition of a scattering data set at the absorption edge of selenium by SeMet derivative crystals. Wavelengths for optimal contributions of $\Delta f'$ and $\Delta f''$ were experimentally determined on beamlines.

Phasing

During an X-ray experiment, only intensities are measured so the phase information is lost. To solve the structure, phases must be recovered. To do so, two different techniques were used during the work presented in this thesis. The molecular replacement method uses phases of a homologous structure as first estimates for phases of the new structure. Single Anomalous Diffraction relies on the energy absorption of heavy atoms at proper X-ray wavelengths, resulting in a non centro-symmetric (anomalous) scattering which can be observed in the difference in intensity levels of Friedel's pairs.

10.5 Refinement

10.5.1 Temperature factor and R-factors

Temperature factor

Atoms in crystals oscillate about their average position. The vibration may be isotropic (same vibration in all directions) or not[21]. These oscillations are estimated by the B-factor[22] which is defined as follows:

$$B_j = 8\pi^2 u_j^2$$

for every atom j in the structure, where u_j^2 is its mean square displacement, in the case of isotropic and harmonic vibration. X-ray crystallography allows an indirect measurement of the size of these oscillations.

As a consequence, the incident X-ray beam does not encounter identical atoms on exactly the same position in successive unit cells. In other words, the scattering power of the atoms is weakened, and the atomic scattering factor for atom j must be multiplied by a temperature-dependent factor:

$$T_j = exp[-B_j \frac{\sin^2 \theta}{\lambda^2}]$$

where θ is the scattering angle and B_j is the B-factor of atom j. As u_j increases, B_j increases and the contribution of the atom to the scattering is decreased.

Interestingly, B-factors can be viewed as indications of the degree of flexibility of parts of the structure: Atoms with low B-factors will indicate a well-ordered part while atoms with greater B-factors will belong to flexible parts. But B-factors may also indicate errors in model building. Indeed, B-factor of an incorrect atom in the model will tend to be higher than correctly built atoms in the neighbourhood.

R-factor

To estimate the accuracy of the model, the R-factor is used. It measures how well the simulated scattering pattern from the crystallographic model explains the experimental X-ray diffraction data. It is defined by:

$$R = \frac{\sum ||F_{obs}| - |F_{calc}||}{\sum |F_{obs}|}$$

where $|F_{obs}|$ are the structure factors computed from measured data, and $|F_{calc}|$ are the structure factors calculated by a Fourier transform of the model, and the sums extend over all the reflections measured and their calculated counterparts.

Basically, the lower the value the better the model. Typical values for R-factors of refined structures are in the range of 15 % to 25 %[23].

[21] Most often, an isotropic hypothesis is sufficient. However, at high resolution with sufficient data, anisotropic consideration of B-factors in the refinement process may be justified.

[22] Also known as Debye-Waller factor or temperature-factor.

[23] A totally random set of atoms would give an R-factor of about 63 %. By contrast, a perfect fit would result in an R-factor of 0.

10.5.2 Refinement process

Principle

When an initial set of phases has been determined (for instance by molecular replacement or SAD), the electron density map can be derived. However, the first phases are often not very accurate, resulting in an unclear and incomplete electron density map. Nevertheless, portions of the protein structure can be determined from this initial electron density map. This is commonly done with the software COOT [39]. These parts are then used to simulate a scattering in identical conditions to the experiment (space group, cell parameters...) by Fourier transform, and calculate a new set of improved phases which are subsequently reapplied to the measured amplitude to derive an improved electron density map. This step was often performed with Refmac software [40] in the work presented here. From this refined map, the model is corrected. This process, called refinement, is repeated until no further improvement can be obtained. This is estimated by calculating a statistical value which will be discussed later: R_{free}.

Parameters

Therefore, through refinement, the atomic model is adjusted to improve the agreement with the measured diffraction data. Parameters which are refined include atomic positions, but B-factors as well. However, for typical X-ray scattering data of proteins, there are only a few observations for each structure factor, which could result in poor statistics. As a consequence, additional parameters are considered in the refinement process, such as restraints on chemical bond lengths, angles, and distance contacts.

TLS (Translation/Libration/Screw refinement) allows further crystallographic refinement by considering additional parameters that describe the possible mean square displacements of rigid bodies of groups of atoms. It is based on TLS Motion Determination which analyzes a crystal structure for evidence of flexibility between partitions of the structure, based on B-factors [41] [42]. In other words, TLS refinement describes anisotropic displacements within portions of the protein. It is in between isotropic and full anisotropic refinement and can be used at medium resolution (2 Å). TLS refinement was used for refinement of MurD and MurF protein structures.

Maximum likelihood based refinement

The most commonly used method in refinement is based on maximum likelihood. The consistency between model and observation is measured by the probability that the observations should have been made, given a model.

R-free

So the refining strategy consists of fitting experimental data by a model as best as possible, through successive runs. However, such modeling carries the risk of overfitting the structure factor observations [43]. Indeed, there is no way to know whether a lower R-value is the result of an improvement of the protein structure after a refinement cycle or the aftermath

of mimicking the variations in intensities of the given X-ray data set, noise included. Such an over-fitted model may not explain a data set of an identical crystal which would contain a different noise, and results in a inaccurate, or even wrong model. Such overfitting generally occurs when a model has too many parameters relative to the number of observations. This is typically the case for crystallographic data, which are thus particularly susceptible to overfitting.

Techniques such as cross-validation are now used to indicate possible overfitting. The essence of cross-validation is to test how a model derived from the main part of a data set can explain an independent, complementary, randomly chosen, part of the data set [44].

Before refinement of crystallographic data, a portion of the data set (5 to 10 %), corresponding to the testing subset, is put aside and the refinement is only performed with the remaining data (working subset). The R_{free} factor is computed in the same way as the R-factor but with a summation on the testing subset only [45]. Therefore, R_{free} tests how well the model predicts experimental observations that have not been used to fit the model. In case of overfitting of the work set by the model, the model will likely not be able to explain the test subset. Indeed, if R_{work} decreases as a result of fitting to noise of the working set, R_{free} will not decrease as its noise was not present in the working subset, but will instead increase. Therefore, R_{free} is a good indicator of overfitting[24].

End of refinement process and structure validation

When R_{free} converges and no enhancement can be obtained by modifying the parameters of the model, and geometric restraints are respected, the refinement of the structure is considered as finished. At that point, the value of R_{free} must be in the range of 15 % - 25 %, depending on the resolution [46], and R_{free} should be greater than R_{work}, otherwise it indicates an overfitting of the noise of the test set.

Refinement Process and Structure Validation

Once estimates of phases for the new structure are found, they are refined to match the scattering data better by iterative cycles: this is the refinement process. To monitor how much the estimated phases explain the data, and limit the risk of overfitting the data throughout the process, a cross-validation method is used through the statistic value R_{free}. To validate the structure, the value of R_{free} must be low enough and general properties of geometry of proteins must be respected.

10.6 Data processing

Diffraction images were indexed and scaled with XDS [30]. Selenomethionine or zinc sites were identified and refined using autoSHARP [47]. The MurD, MurE and MurF structures

[24]Interestingly, Brünger [45] showed that the R_{free} value is correlated with the accuracy of phases.

were traced with ARP/ wARP 7.0.1 [48]. In the case of MurF, PHASER 2.5.1 [49] was used to perform molecular replacement with the model generated by the SAD experiment in order to phase data to 1.65 Å. For MreB, the same program was used for molecular replacement with the PDB entry 1JCF. The full structures were completed by cycles of manual model building with COOT 0.6.2 [39] and BUCANEER 1.5.2 [50]. Cycles of restrained refinement employing TLS [41] [42] were performed with REFMAC 5.7 [40] as implemented in the CCP4. Stereochemical verification was performed by PROCHECK [51] and secondary structure assignment by DSSP [52]. Figures were generated with PyMol (http:// www.pymol.org). The structures of MurD, ADP-MurE and ADP-MurF from *T. maritima* were deposited in the PDB database with accession numbers: 4BUC, 4BUB, and 3ZL8.

Chapter 11

Interactions

11.1 Dot-Blot assay

Purified proteins were loaded as dots from 1 μg to 7 μg onto a nitrocellulose membrane. After 15 min drying, the membranes were blocked in PBS-T (PBS-Tween 0.05%) supplemented with 3% Bovine Serum Albumin (BSA) for 1h at room temperature with shaking. After one washing step in PBS-T, the membranes were incubated overnight at 4°C in a PBS-T solution containing 0.1 mg.mL^{-1} of the protein to test the interaction with. After three washing steps in PBS-T, the membranes were incubated with either Streptactin HRP conjugate from IBA at 1:20000, or SuperSignal West HisProbefrom Pierce Biotechnology at 1:2000 in PBS-T. Signals were developed with the SIGMA FASTTM DAB tablets with Metal Enhancer. Negative controls: BSA, 7 μg per dot, or other proteins from the Bacterial Pathogenesis Group or from the Membrane and Pathogens group, >5 μg per dot.

11.2 Pull-down assay

General experimental conditions, such as columns and buffers used are mentioned in tables in the Results part. Here below are detailed the protocols for main pull-down assays.

11.2.1 MreB-Mur ligases

Co-expression of MreBhis and MurD

pETDuet-*murD* and pET30b-*mreBhis* were co-transformed in C41(DE3) cells. Induction was done at 37 °C for 3 h under 1 mM IPTG, at an OD of 0.5 A.U. 500 mL of cells were lysed by sonication in 50 mM Hepes pH 7.4, 5 % glycerol, 0.2 M NaCl in the presence of DNase, lysozyme, and proteases inhibitors. Ultracentrifuged sample was loaded on a 1 mL His-trap column. Washing buffer was similar to lysis buffer, and elution was performed by steps of imidazole concentration (maximum concentration: 500 mM). Samples were analyzed by SDS-PAGE.

Co-lysis of MreBstrep and MurEhis expressing cells

1 L of C41(DE3) cells containing pET30b*mreBstrep* were induced and mixed to 500 mL of BL21(DE3) cells expressing MurEhis from a pET30b vector. Cells were harvested and frozen at -80°C. Co-lysis was performed by sonication in 25 mM Hepes pH 7.4, 50 mM NaCl, supplemented with DNase, lysozyme and proteases inhibitors. After ultracentrifucation for 45 min at 100,000 g, sample was loaded on a 5 mL Strep-trap column and elution was performed with 2.5 mM d-desthiobiotin.

Co-lysis of MreBstrep and MurFhis expressing cells

A similar protocol was followed for this pull-down assay, though pellets were not frozen.

Buffer screening for co-lysis MreBhis, MurD, hisMurE, MurFhis

Expressions were induced from pETDuet-*murD-murEmurF* and pET30b-*mreBhis* in BL21(DE3) and C41(DE3) cells, respectively, at 0.6 A.U. of OD with 1 mM IPTG overnight at 22°C. Cultures were mixed in equivalent amounts of cells and pellets from 25 mL of cell cultures were frozen at -80°C. Cells were lysed by heat shock in 1.5 mL of different buffers: 25 mM MES pH 6 or 6.5, or Hepes pH 7 or 7.5, or Tris pH 8 or 8.5, 0.1 M NaCl, DNase, lysozyme, proteases inhibitors. After ultracentrifugation at 100,000 g for 1h, supernatants were loaded on 100 μL of His resin in a centrifuge column. Washing steps and elution steps were: 12.5 mM, 25 mM, 50 mM, 100 mM , 500 mM imidazole. Samples were analyzed by SDS-PAGE.

11.2.2 MreB-MurG

Several attempts to purify the MreB-MurG complex were performed. Here is described a general protocol for such an experiment.

BL21(DE3) cells expressing MurG-strep were prepared by induction with 400 ng/mL AHT at an OD of 0.5 A.U. overnight at 22°C. C41(DE3) cells expressing MreB-his were prepared by induction with 1 mM IPTG at an OD of 0.5 A.U. overnight at 22°C. Equivalent amounts of MurG- and MreB-expressing cells were mixed and harvested before lysis by sonication in 50 mM Hepes pH 7.4, 150 mM NaCl, 5%-10% glycerol, 0.1 % sarkozyl, DNase and proteases inhibitors. After ultracentrifugation at 100,000 g for 45 min, supernatant was loaded on a 5 mL Strep-trap column (GE Healthcare) previously equilibrated with a similar buffer (no sarkozyl). Elution was done with 2.5 mM d-desthiobiotin.

Depending on the experiment, elution fractions were analyzed by SDS-PAGE, Western Blotting, investigated for EGS cross-linking, assessed for the presence of MurG and MreB by dot blot, loaded on a His-trap column, or loaded on the Superose6 gel filtration column.

11.2.3 MurG-Mur ligases

Typical procedure

Most large-scale pull-downs between MurG-Mur ligases that were tried were performed as follows.

750 mL of BL21(DE3)-pASK-IBA3C-*murGstrep*, 750 mL of BL21(DE3)-pET30b-*murEhis*, 500 mL of BL21(DE3)-pETDuet-*murD* were prepared. MurD and MurE cell cultures were mixed. Pellets were frozen at -80°. MurG-expressing cells were lysed by sonication in 50 mM Tris pH 7.1, 0,05 mM NaCl, 1 mM DTT, 1 mM EDTA, 0.02 % DDM, 0.2 % sarkozyl supplemented with proteases inhibitors. Mix of MurD-expressing cells and MurE-expressing cells were lysed in a similar buffer without sarkozyl. Supernatants of MurG-expressing cells lysates were loaded on a 5 mL HP Strep-trap affinity column (GE) and column was washed. Then, supernatant of MurD- and MurE-expressing cells lysates were loaded onto the column. Elution was done with 2.5 mM d-desthiobiotin. Elution peak was loaded on Superose6 (p198), and then on a His-trap FF column. Elution was done with steps of imidazole.

Buffer screening in a small-scale experiment

After induction, equivalent amounts of each strain (expressing MurD, MurE, or MurG) were mixed and aliquoted in 50 mL and pelleted. Cells were lysed by sonication in 6 mL buffer: 50 mM Tris pH 7.1 or 8 or 9, 2 mM DTT, 1 mM EDTA, 0.02 % DDM, 0.05 M NaCl plus 0.25 % sarkozyl or 1 M NaCl, supplemented with proteases inhibitors. After centrifugation at 18,000 rpm for 45 min, supernatants were loaded onto 100 μL Strep resin. Washing buffers: 50 mM Tris pH 7.1 or 8 or 9, 0.05 M NaCl, 0.02 % DDM, 1 mM EDTA, 2 mM DTT. Elution was done with 2.5 mM d-desthiobiotin.

11.3 Size Exclusion Chromatographies

This technique allows the detection and separation of protein complexes from single proteins, depending on the hydroscopic volume of macromolecules. Nevertheless, complexes must be relatively stable to be detected by this method.

Purified proteins were mixed at the indicated ratios (see Results) at least 30 min before injection onto the gel filtration column. Different buffers (pH 6 - 6.5: MES, pH 7 - 7.5: Hepes, pH 8 - 8.5: Tris, 100 - 300 mM NaCl, 0 - 1 mM EDTA, 0 - 2 mM DTT, 0 - 0.04% DDM) were tried, and different columns (SuperDex200, SuperDex75, Superose6, all from GE) were tested as indicated in tables in the next Part. Here are given two detailed examples of such experiments.

11.3.1 MurG-MreB, SuperDex200

MurG-strep and MreB-his were purified as described earlier in 25 mM Hepes pH 7.4, 150 mM NaCl, 1 mM EDTA, 5 % glycerol, 0.04 % DDM. For lyses, proteases inhibitors were added, and 0.5 mg/mL lysozyme plus DNase (MreB) or 0.5% sarkozyl (MurG) as well. Purifications were made from 1 L of cell culture. After affinity chromatography, both proteins were further purified by size exclusion chromatography on a Superose6 (MurG) or a SuperDex75 (MreB) in 25 mM Hepes pH 7.4, 150 mM NaCl, 1 mM EDTA, supplemented with 0.04 % DDM for MurG. The purest elution fractions of MreB and MurG were directly mixed in a 1:1 ratio and incubated overnight at 4°C on a rotating wheel. For each run, about 50 μg were loaded

onto the gel filtration column. Running buffer: 25 mM Hepes pH 7.4, 150 mM NaCl, 1 mM EDTA, 5 % glycerol, 0.04 % DDM.

11.3.2 MurG-MreB, Superose6

MurG-strep and MreB-his were purified as described earlier in 50 mM Hepes pH 7.4, 5% glycerol, 10 mM DTT, 300 mM NaCl. For lyses, proteases inhibitors were added, and 0.5 mg/mL lysozyme plus DNase (MreB) or 0.5% sarkozyl (MurG) as well. Purifications were made from 800 mL of cell culture. After affinity chromatography, both proteins were further purified by size exclusion chromatography on a Superose6 in 50 mM Hepes pH 7.4, 5% glycerol, 10 mM DTT, 300 mM NaCl. The purest elution fractions of MreB and MurG were directly mixed in a 1:1 ratio and incubated overnight at 4°C on a rotating wheel. 500 μL of this mix were loaded on the Superose6 in 50 mM Hepes pH 7.4, 5% glycerol, 10 mM DTT, 300 mM NaCl.

11.4 Cross-linking assay

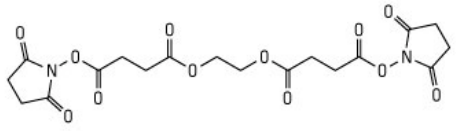
11.4.1 Principle

As most *in vivo* protein-protein interactions are transient, it may be very difficult to capture protein complexes through long methods such as pull-down or gel filtration assays. Cross-linking reagents provide the means for capturing protein complexes by covelently binding them together as they interact. The high reactivity of functional groups on crosslinkers reagent allows even transient interactions to be frozen in place.

Crosslinking reagents carry reactive moieties that bind to specific amino acid functional groups on target proteins.

11.4.2 EGS cross-linking assays

EGS is a crosslinker that contains amine-reactive NHS-ester ends around a 12-atom spacer arm.



EGS
Ethylene glycol bis(succinimidylsuccinate)
MW 456.36
Spacer Arm 16.1 Å

Figure 11.1: EGS formula.

A fresh solution of EGS at 100 mM in DMSO was prepared, and then four dilutions in protein buffer were made: 5 mM, 10 mM, 20 mM, and 50 mM. 5 μg of 'protein A' were mixed to 5 μg of 'protein B' in 5 different tubes and volumes were adjusted to 9 μL with buffer. Then, 1 μL of either of the EGS dilutions (or buffer as negative control) was added

and samples were incubated at room temperature for 1 h. Reaction was stopped by addition of Laemmli blue and heated before SDS-PAGE.

11.4.3 Glutaraldehyde cross-linking assays

Glutaraldehyde is an aggressive and indiscriminant crosslinking reagent, that can be used for a first interaction test. However, its products is often very heterogeneous.

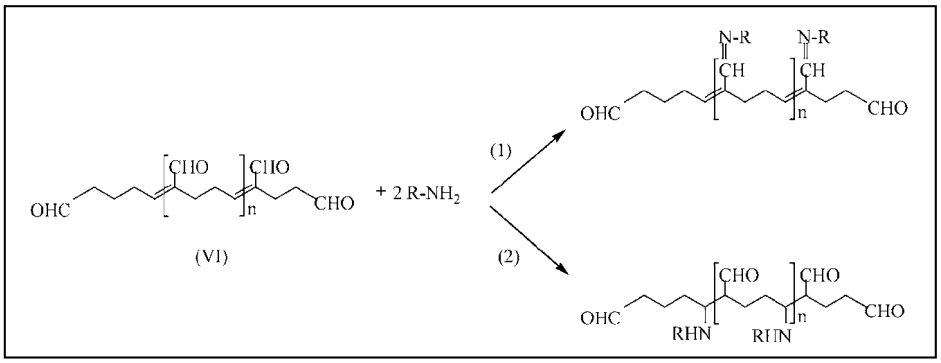


Figure 11.2: Glutaraldehyde. From Migneault *et al.*, 2004 [53].

Proteins (MurC, MurD, MurE, MurF) were diluted at 1 mg/mL in 25 mM NaPi pH 7.4, 150 mM NaCl, mixed at a 1:1 ratio in a total reaction volume of 60 μL. Cross-linking reagent was added at a final concentration of 0.025 %. 10 μL were taken every 10 min for SDS-PAGE analysis.

11.5 Native gels

For native gels, proteins were mixed at different ratios, with final concentrations around 1 mg/mL, incubated for 30 min at room temperature and then 5 μL of samples were loaded on native gels (no stacking).

Gels were prepared as for normal SDS-PAGE, but without SDS. For 1 L of 10x running buffer: Tris 30.3 g, glycine 144 g. 10 mL of loading buffer: 2.5 mL of Tris pH 6.8 1M, 4 mL glycerol, 10 mg bromophenol blue.

11.6 Surface Plasmon Resonance spectroscopy

Interactions between proteins can be monitored thanks to the optical phenomenon of Surface Plasmon Resonance (SPR). This section exposes the physical principle of SPR and its application to studies of biomolecules interactions. In addition, experimental procedures which were followed during SPR assays for assessing Mur interactions are described.

11.6.1 Principle of SPR

SPR is a phenomenon that occurs in thin conducting films at an interface between media of different refractive indices.

Two theoretical approaches are presented here. First, the SPR phenomenon is explained with optical and electromagnetic theories, introducing total internal reflection and surface plasmon resonance. Second, an optical approach is used to show the relationship between the reflectance (reflected intensity) and the incident angle, unveiling a minimal reflectance due to the SPR phenomenon. The resonance angle corresponding to this minimum is the entity that is measured in BIAcore devices.

Total Internal Reflection

When light travels from one medium to another, it follows the Snell's law (see Figure 11.3, left) $n_1 \sin\theta_1 = n_2 \sin\theta_2$, where θ_1 and θ_2 are the incident, resp. transmitted angles and n_1 and n_2 are the refractive indices in the first, resp. second, medium.

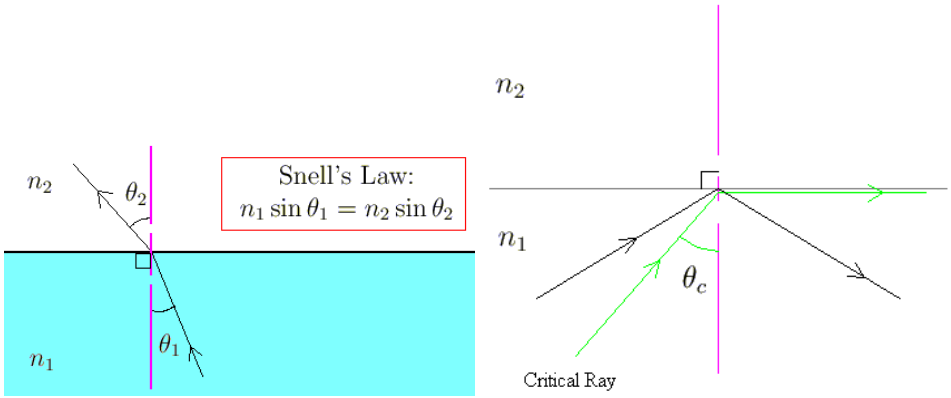


Figure 11.3: Left: Transmission of light from a medium with a refractive index n_1 to a medium with a refractive index n_2. Right: Critical angle for which internal reflection occurs. From University of British Columbia website [54].

There exists a critical incident angle θ_c for which all the light is totally reflected. In other words, for $\theta_1 = \theta_c$, $\theta_2 = \frac{\pi}{2}$, giving $\theta_c = \arcsin\frac{n_2}{n_1}$.

Therefore, at higher incident angles than θ_c, the phenomenon of total internal reflection (TIR) occurs (see Figure 11.3, right). This critical angle is only defined when n_2/n_1 is less than 1. For instance, TIR can be seen from glass ($n_1 \approx 1.5$) to water or protein buffer ($n_2 \approx 1.33$).

Evanescent wave

In terms of electromagnetic waves, let us consider a plane wave. The corresponding transmitted electromagnetic field $\overrightarrow{E_t}$ can be written as follows:

$$\overrightarrow{E_t}(\overrightarrow{r}, t) = \overrightarrow{E_0} \exp[i(\overrightarrow{k_t}.\overrightarrow{r} - \omega t)] \tag{11.1}$$

where $\overrightarrow{E_0}$ is the incident electromagnetic field, $\overrightarrow{k_t}$ the transmitted wave vector, $\overrightarrow{r}$ the position vector, ω the angular frequency, and t the time.

$\overrightarrow{k_t}$ and $\overrightarrow{r}$ have two components:

$$\overrightarrow{k_t} = k_x\overrightarrow{u_x} + k_z\overrightarrow{u_z} = k_t \sin\theta_t \overrightarrow{u_x} + k_t \cos\theta_t \overrightarrow{u_z} \tag{11.2}$$

$$\overrightarrow{r} = x\overrightarrow{u_x} + z\overrightarrow{u_z} \tag{11.3}$$

where $\overrightarrow{u_x}$ and $\overrightarrow{u_z}$ are the unit vectors of the wave plane, respectively, and $k_t = \frac{\omega n_2}{c}$, with c the light celerity.

In addition, since $n_2/n_1 < 1$, and Snell's law gives $\sin\theta_t = \frac{n_1}{n_2}\sin\theta_i$, $\sin\theta_t$ is higher than 1, meaning $\cos\theta_t$ is complex: $\cos\theta_t = i\sqrt{\sin^2\theta_t - 1}$.

After a few calculations, one obtains:

$$\boxed{\overrightarrow{E_t} = \overrightarrow{E_0} e^{-\kappa z} e^{i[kx-\omega t]}} \tag{11.4}$$

with $\kappa = \frac{\omega}{c}\sqrt{(n_1 \sin\theta_i) - n_2^2}$ and $k = \frac{\omega n_1}{c}\sin\theta_i$.

This wave, characterized by its propagation in the x direction and its exponential attenuation in the z direction in second medium is known as the *evanescent wave*.

Plasmons and resonance

If the surface of the glass prism is coated with a thin film of a noble metal[1] (e.g. gold), the evanescent wave interacts with free electrons in the metal layer.

Indeed, according to Maxwell's equations, surface plasmons can propagate along a metallic surface and have a spectrum of frequencies related to the wave-vector by a dispersion relation[2]:

$$k_x = k_0\sqrt{\frac{\epsilon_1\epsilon(\omega)}{\epsilon_1 + \epsilon(\omega)}} \tag{11.5}$$

where ϵ_1 is the dielectric constant of glass, $\epsilon(\omega)$ the dielectric constant of the metal, and $\frac{\omega}{c} = k_0$, the wave vector in void at frequency ω, with c the light celerity. Figure 11.4 represents the dispersion relation for surface plasmon.

Besides, the wave vector k_x of the evanescent wave formed by the incident light traveling through the prism is described by:

$$k_x = \frac{\omega}{c}\sqrt{\epsilon_1}\sin\theta_1 \tag{11.6}$$

Since the wave is evanescent, its z component is imaginary, so it can be represented by a line in Figure 11.4. The two curves intersect at the resonance, occurring at k_{xR} which corresponds to the resonance incident angle θ_R defined by the relationship:

$$\boxed{\mathrm{k}_x = k_0\sqrt{\frac{\epsilon_1\epsilon(\omega)}{\epsilon_1+\epsilon(\omega)}} = k_0\sqrt{\epsilon_1}\sin\theta_R} \tag{11.7}$$

[1] Noble metals are metals that are resistant to corrosion and oxidation in atmosphere.

[2] detailed calculation is available in *Le champ proche optique - Théorie et applications*, edited by Daniel Courjon and Claudine Bainier [55].

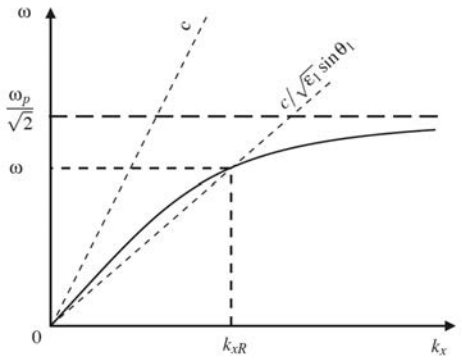


Figure 11.4: Left: Dispersion relation of the surface plasmon mode in metal. Coupling of a plasmon with an evanescent wave, in condition of total reflexion with a prism of $\epsilon_1 > 1$. ω_p is the plasmon angular frequency in a Drude model, $k_0 = \omega/c$. From *Le champ proche optique - Theorie et applications*, Daniel Courjon et Claudine Bainier [55].

Fresnel coefficients

A frequently used theory for explaining the experimental surface plasmon resonance phenomenon is based on the Fresnel coefficients for a double interface prism-metal $(p - m)$ and metal-sensing layer $(m - s)$. In this model, each material is treated as a homogeneous continuum, described by a frequency-dependent relative permittivity $\epsilon(\omega)$.

Let us consider the Kretschmann configuration used in BIAcore devices, in which a metal film coats the glass prism (see Figure 11.5). The light illuminates the prism with an incident angle θ, and an evanescent wave penetrates through the metal film and then the sample before vanishing. As seen previously, the plasmons within the metal film are excited and 'resonate' with the evanescent wave if the incident angle is close to the resonance angle θ_R.

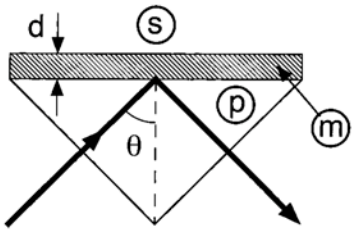


Figure 11.5: Kretschmann configuration of a prism-metal-sensing layer interface. The glass prism, the metallic film and the sensing layer are labeled p, m, and s, respectively. θ is the incident angle, and d is the thickness of the metallic film. From Kurihara and Suzuki [56].

When the p-polarized light is incident at an angle θ on this three-layer system, the reflectance R of the light is given by the three-layer Fresnel equations as follows:

$$R = \left| \frac{r_{pm} + r_{ms} \exp(2ik_{mz}d)}{1 + r_{pm}r_{ms} \exp(2ik_{mz}d)} \right|^2 \tag{11.8}$$

with

$$r_{pm} = \frac{k_{pz}\epsilon_m - k_{mz}\epsilon_p}{k_{pz}\epsilon_m + k_{mz}\epsilon_p} \tag{11.9}$$

$$r_{ms} = \frac{k_{mz}\epsilon_s - k_{sz}\epsilon_m}{k_{mz}\epsilon_s + k_{sz}\epsilon_m} \tag{11.10}$$

and

$$k_{jz} = (\epsilon_j \frac{\omega^2}{c^2} - k_x^2)^{1/2} \quad \text{for} j = p, m, s \tag{11.11}$$

$$k_x = \sqrt{\epsilon_p}\frac{\omega}{c}\sin\theta \tag{11.12}$$

where r_{pm} and r_{ms} are the amplitude reflectance given by Fresnel formulas of p-polarization form prism-metal and metal-sensing layer interfaces, respectively; ϵ_j and k_{jz} are the dielectric constant and the wave-vector component perpendicular to the interface in medium j; k_x is the component of the incident wave-vector parallel to the interface; d is the thickness of the metallic film; ω is the angular frequency of the incident light and c is the velocity of light [56].

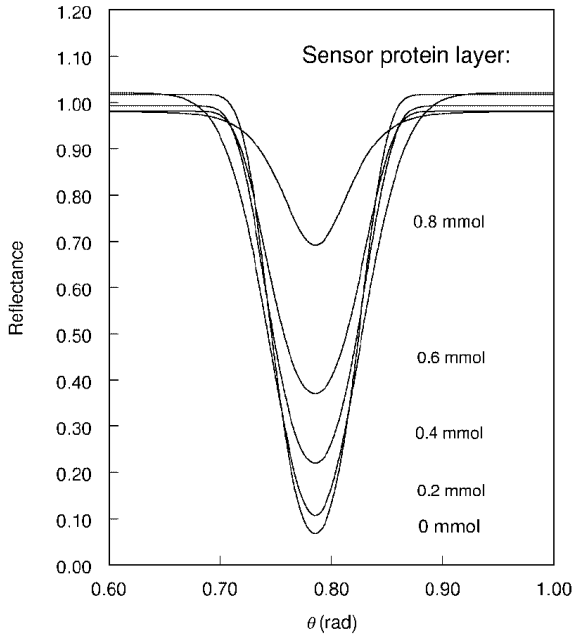


Figure 11.6: Evolution of SPR spectra simulated for increasing ligand protein concentrations. Simulations were performed using the Fresnel three layer model for a light beam at 600 nm striking the metal and sensing layers through a glass prism. From Englebienne *et al.*, 2003 [57].

According to equations 11.8 to 11.11, the reflectance R varies as a function of the dielectric constant ϵ_s of the sensing layer.

The application of these equations together with mathematical models for dielectric constants (see Kurihara *et al.*, 2002 [56] for more details) allows to simulate the SPR spectra displayed in Figure 11.6.

SPR as a measurement of the refractive index

As seen in the previous paragraph, there exists an angle θ_{SPR} at which the reflectance is minimum. It can be shown that the change in SPR angle θ_{SPR} of reflected light at a given

wavelength is directly related to both the change in refractive index n at the surface[3], and the change in thickness d of the layer, according to the relationship [57]:

$$\boxed{\Delta\theta_{SPR} = c_1\Delta n + c_2\Delta d} \tag{11.13}$$

In addition, the Lorentz-Lorentz relationship shows that the change in refractive index Δn is linked to the change in thickness Δd [57] [58]:

$$\Delta n = -\frac{1}{6n}(n^2+2)^2\left(\frac{n^2-1}{n^2+2} - \frac{n_\omega^2-1}{n_\omega^2+2V}\frac{V_p}{V}\right)\frac{\Delta d}{d} \tag{11.14}$$

where n is the refractive index of the protein, n_ω the refractive index of water, V_p the volume of the protein, and V the volume of the protein layer ($V = V_p + V_w$).

SPR PHENOMENON

When a beam of incoming light passes from a material with a high refractive index (e.g. glass) into material with a low refractive index (e.g. water) with an incident angle greater than the critical angle, the light is completely reflected (total internal reflection). Under such conditions, an evanescent wave propagates in the second medium, exponentially decreasing in the direction perpendicular to the incident plane. When a thin metal film is added at the interface glass-water, at an appropriate wavelength and incident angle, the evanescent wave interacts with free electrons of the metal layer and resonate. At the resonance angle, the reflectance is minimal. This angle depends on the refractive index of the (first layers of the) second medium.

11.6.2 From resonance angle to interaction assay

Measurement of the resonance angle

In BIAcore systems, the media are (see Figure 11.7):

- the glass of the sensor chip,
- a thin layer of gold[4] on the sensor chip surface,
- the matrix coated on the gold surface in which the ligand is immobilized.

Monochromatic light with a broad range of incident angles is directed towards the glass face of the sensor chip. A flow with the analyte protein is passed on the matrix side, and the reflected lights reach the detector array at different points [57] [56]. The detector continuously records the position of reduced light intensity corresponding to resonance and calculates the SPR angle [57].

[3]It is of note that the change in refractive index can be detected only if it occurs within the effective penetration depth of the evanescent wave (about 150 nm) from the surface.

[4]Gold is used in Biacore sensor chips because it combines favorable SPR characteristics with a strong resonance effect, with stability and a high level of inertness in biomolecular interaction contexts (gold is very resistant to oxidation and other atmospheric contaminants but is compatible with many chemical modification systems).

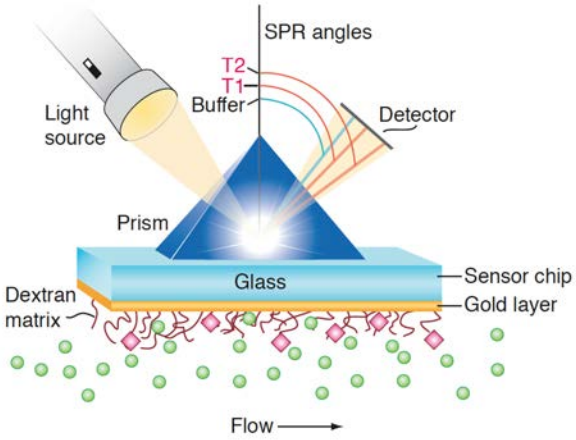


Figure 11.7: SPR optical unit and a sensor chip detecting the analyte protein (green) in the flow solution, which pass by the ligand molecules (pink) linked to the dextran matrix coated onto the gold film. The SPR angle defines the position of the reduced-intensity beam. Time points T1 and T2 correspond to two SPR angles, which shift as the protein binds to the immobilized ligand over time [59].

An increase in resonance angle corresponds to a binding event

As seen previously (see Equation 11.13), this angle strongly depends on the refractive index in the sample medium.

As proteins have a higher refractive index than buffer [60], any change in protein layer thickness onto the gold film (binding event or large conformational change) will result in a refractive index change that will modify the resonance angle measured by the instrument. Interestingly, this change is linear[5], allowing in-real time monitoring of interactions.

The SPR angle change is reported in BIAcore device as resonance units (RU), a response of 10^3 RU corresponding to an angle change of 0.1 degree [57] [59]. Empirical measurements have shown that the binding of 1 ng/mm^2 of protein to the sensor surface leads to a response of about 1000 RU [57] [59]. Since the matrix where the ligand is bound is about 100 nm thick, this represents an increase in protein concentration within the matrix of 10 μg/mL.

It is worthwhile mentioning that as light does not penetrate the sample, interactions can be followed in colored, opaque, or cloudy samples. Moreover, in SPR assays, no labels are required and detection is instantaneous by contrast to most biochemical methods.

11.6.3 Experimental aspects

Immobilization of the ligand protein onto the sensor surface

For most applications, a carboxymethylated dextran matrix linket to a gold layer enables molecules to be immobilized to the sensor surface through covalent amine binding [57]. The most common sensor surface used is the CM5 sensor chip from GE Healthcare.

Ligand immobilization can be divided into three steps:

- activation of the carboxymethyl groups of the dextran matrix with a mix of 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC) and N-hydroxysuccinimide (NHS), thus

[5]However, for other biomolecules, interactions may have a refractive index increment different and some calibration may be necessary

creating a highly reactive succinimide ester,

- coupling: injection of the ligand in a buffer with a pH lower than the protein's pI[6],
- blocking: the remaining activated carboxymethyl groups are 'blocked' by injecting 1 M ethanol-amine - HCl at pH 8.5. The high concentration of ethanolamine contributes to elute any non-covalent bound material as well.

The surface is then ready to be used for binding experiments.

Since the dextran matrix is flexible, it allows relatively free movement of attached ligands within the surface layer and provides a hydrophilic environment for interactions. However, as proteins usually contain several available amine groups, the site of immobilization and therefore the orientation of the immobilized protein cannot be determined with amine coupling. This unspecific binding direction may interfere with protein flexibility and potential conformational changes that may be required for interaction.

Data acquisition

When sample is passed over the sensor surface, the 'sensorgram' shows an increasing response when molecules interact, as illustrated Figure 11.8. Basically, the response is proportional to the mass of bound material, and remains constant if the interaction reaches equilibrium or saturation. When sample is replaced by buffer, the response decreases as the interaction partners dissociate.

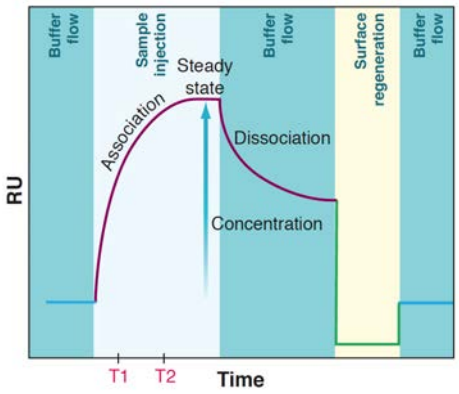


Figure 11.8: Typical sensorgram of an interaction detected by a BIAcore device. Time points T1 and T2 correspond to the two SPR angles depicted Figure 11.7, which shift as protein molecules bind to the immobilized ligand over time. From Wilson, 2002 [59].

From these data, complete profiles of binding and dissociation are generated in real time, and specificity, affinity, kinetic behavior can be determined and characterized.

Regeneration

Most often, some protein remains bound to the immobilized ligand and a regeneration step is needed. For this purpose, a solution which will disrupt the interaction is injected. As the ligand is covalently coupled, it will remain attached to the dextran surface. However, the regeneration solution must be chosen as it will not cause any damage to the ligand.

[6]The dextran layer has a net negative charge which assists proteins to be electrostatically attracted to the dextran when they are positively charged. Succinimide ester groups in the dextran matrix react with the primary amine and other nucleophilic groups on proteins to form covalent bonds.

11.6.4 Experimental procedures

Immobilization

A standard protocol for amine coupling was used with the amine coupling kit from GE-Healthcare. For that purpose, flow cells of CM5 sensor chips (GE Healthcare) were activated on a BIAcore 3000 machine with 60 μL of 0.2 M EDC and 0.05 M NHS at 5 μL.min^{-1}. Then, MreB, MurE, MurF, or MurG at a concentration of 20 μg.mL^{-1} (MreB, MurE, MurF) and 5 μg.mL^{-1} respectively, were injected over one of the activated flow cells in 10 mM Sodium Acetate pH 5.0 (MreB, MurE, MurF) or 10 mM HEPES pH 7.0 (MurG) until a coupling level of about 2,000 resonance units (RUs) was reached. A flow cell with no protein immobilized (buffer instead) was used as a negative control.

Data acquisition

Real-time monitoring of the interactions was performed on a BIAcore 3000 machine (GE Healthcare). Samples diluted in running buffer (10 mM HEPES, 150 mM NaCl, pH 7.4) supplemented with 0.005% v/v polysorbate 20 (or 0.04% DDM for MurG), were injected over the different surfaces at a flow rate of 20 μL.min^{-1}. The formed complexes were washed in running buffer and surfaces were regenerated with 1 M NaCl, followed by 1 mM EDTA if required.

Data analyses

Binding curves were analyzed with BIAevaluation software (GE Healthcare Bio-Sciences AB, Uppsala, Sweden), and reproduced in Excel (Microsoft).

Curves were fitted with the two-state (conformational change) model from BIAcore which describes a 1:1 binding followed by a conformational change in the complex. In addition, it is assumed that the conformationally changed complex can only dissociate through the reverse of the conformational change. Therefore, the model is:

$$A + B = AB = AB^*$$

It is worthwhile mentioning that this very simple model does not take into account any conformational change in any of free ligands.

Bibliography

[1] Fusinita van den Ent, Linda A. Amos, and Jan Lowe. Prokaryotic origin of the actin cytoskeleton. *Nature*, 413(6851):39–44, 2001.

[2] S. Doublié. Preparation of selenomethionyl proteins for phase determination. *Methods Enzymol.*, 276:523–30, 1997.

[3] Jason DeChancie and K.N. Houk. The origins of femtomolar proteinigand binding: hydrogen bond cooperativity and desolvation energetics in the biotinStrept)avidin binding site. *J Am Chem Soc.*, 129(17):5419429, 2007.

[4] Thomas GM Schmidt and Arne Skerra. The Strep-tag system for one-step purification and high-affinity detection or capturing of proteins. *Nature Protocols*, 2007.

[5] E. Gasteiger, C. Hoogland, Duvaud S. Gattiker, A and, M.R. Wilkins, R.D. Appel, and Bairoch A. Protein identification and analysis tools on the ExPASy server. *The Proteomics Protocols Handbook, Humana Press*, pages 571–607, 2005.

[6] Danuta Maksel. Thermal shift assay: monitoring solution-dependent changes in protein stability. `http://thermofluor.org/resources/Maksel_Protein-Stability.pdf`.

[7] Carrie Partch. Fluorescence-based thermal shift assays. dept. of chemistry and biochemistry, uc santa cruz. `http://partch.chemistry.ucsc.edu/pdf/ThermalShiftAssay.pdf`.

[8] Ulrika B. Ericsson, B. Martin Hallberg, George T. DeTitta, Niek Dekker, and Pär Nordlund. Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Analytical Biochemistry*, 357(2):289 – 298, 2006.

[9] T.L. Blundell and L.N. Johnson. *Protein Crystallography.* Academic Press, 1976.

[10] Jan Drenth. *Principles of Protein X-Ray Crystallography.* Springer, 2007.

[11] Hong Ling. Protein crystallography, 2006. The University of Western Ontario, `instruct.uwo.ca/biochemistry/523b/Ling1.pdf`.

[12] Juan Manuel Garc-Ruiz. Nucleation of protein crystals. *Journal of Structural Biology*, 142:221, 2003.

[13] Claude Sauter and Richard Giegé. La cristallogenèse des macromolécules biologiques. *Regard sur la biochimie*, 3:21–31, 2001.

[14] Mazumdar M Zhou M Xu GJ Lin SX. Zhu DW, Garneau A. Attempts to rationalize protein crystallization using relative crystallizability. *J Struct Biol.*, 154(3):297–302, 2006.

[15] G.K. Christopher, A.G. Phipps, and R.J. Gray. Temperature-dependent solubility of selected proteins. *Journal of Crystal Growth*, 191:820–826, 1998.

[16] Susanne Moelbert, B. Normand, and Paolo De Los Rios. Kosmotropes and chaotropes: modelling preferential exclusion, binding and aggregate stability. *Biophysical Chemistry*, 112(1):45 – 57, 2004.

[17] P. E. Mason, G. W. Neilson, C. E. Dempsey, A. C. Barnes, and J. M. Cruickshank. The hydration structure of guanidinium and thiocyanate ions: Implications for protein stability in aqueous solution. *Proc. Natl. Acad. Sci. U.S.A.*, 100:4557–4561, 2003.

[18] I. Yu and M. Nagaoka. Slowdown of water diffusion around protein in aqueous solution with ectoine. *Chemical physics letters*, 388:316–321, 2004.

[19] Alexander McPherson. Introduction to protein crystallization. *Methods*, 34(3):254 – 265, 2004.

[20] Ronen Zangi. Can salting-in/salting-out ions be classified as chaotropes/kosmotropes? *The Journal of Physical Chemistry B*, 114(1):643–650, 2010. PMID: 19994836.

[21] F.C.Bernstein, T.F.Koetzle, G.J.Williams, E.E.Meyer Jr., M.D.Brice, J.R.Rodgers, O.Kennard, T.Shimanouchi, and M.Tasumi. The Protein Data Bank: A computer-based archival file for macromolecular structures. *J. of. Mol. Biol.*, 112:535–542, 1977.

[22] Sergei Radaev, Sean Li, and Peter D. Sun. A survey of proteinrotein complex crystallizations. *Acta Crystallogr D Biol Crystallogr.*, 2006.

[23] Fusinita van den Ent, Christopher M Johnson, Logan Persons, Piet de Boer, and Jan Lowe. Bacterial actin MreB assembles in complex with cell shape protein RodZ. *EMBO J*, 29(6):1081–1090, 2010.

[24] Randy J Read. Macromolecular crystallography course of CIMR, University of Cambridge and the MRC-LMB, 1999-2000.

[25] Phil Jeffrey. X-ray data collection course. Princeton University, 2006. `http://xray0.princeton.edu/~phil/Facility/Guides/XrayDataCollection.html`.

[26] Wikipedia. Thomson scattering, Consulted in 2013. `http://en.wikipedia.org/wiki/Thomson_scattering`.

[27] Wikipedia. Atomic form factor, Consulted in 2013. `http://en.wikipedia.org/wiki/Atomic_form_factor`.

[28] Techno-Science.net. Théorie de la diffraction sur un cristal, Consulted in 2013. `http://www.techno-science.net/?onglet=glossaire\&definition=3708`.

[29] European Radiation Synchrotron Facility, Consulted in 2013. `http://www.esrf.eu/about/synchrotron-science/synchrotron`.

[30] Wolfgang Kabsch. *XDS. Acta Crystallographica Section D*, 66(2):125–132, Feb 2010.

[31] A.L. Patterson. Fourier series method for the determination of the components of interatomic distances in crystals. *Physical Review*, 46:372–376, 1934.

[32] Jr. Charles W. Carter and Robert M. Sweet, editors. *Methods in Enzymology.* Academic Press, 1997.

[33] Airlie J. McCoy, Ralf W. Grosse-Kunstleve, Paul D. Adams, Martyn D. Winn, Laurent C. Storoni, and Randy J. Read. Phaser crystallographic software. *J. Appl. Cryst.*, 40:65874, 2007.

[34] Michael R. Sawaya. Making sense of a MAD experiment, 2006. `http://www.doe-mbi.ucla.edu/M230B/`.

[35] The International Union of Crystallography. Online dictionary of crystallography, 2006. `reference.iucr.org/`.

[36] Ethan A Merritt. X-ray anomalous scattering, 2012. `http://skuld.bmsc.washington.edu/scatter/AS_index.html`.

[37] Manfred S. Weiss. Introduction to anomalous scattering and its use in macromolecular crystallography, 2007. `http://www.esrf.eu/events/conferences/embo2007/weiss_AnomScatt_2007.pdf` EMBL.

[38] Evans Gwyndaf. *Anomalous diffraction using synchrotron radiation at optimal X-ray energies:application to protein crystallography.* PhD thesis, University of Warwick, 1994.

[39] P. Emsley and K. Cowtan. Coot: model-building tools for molecular graphics. *Acta Cryst.*, D60:2126–2132, 2004.

[40] G.N. Murshudov, A.A.Vagin, and E.J.Dodson. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Cryst.*, D53:240–255, 1997.

[41] Jay Painter and Ethan A. Merritt. Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Cryst.*, D62:43950, 2006.

[42] Jay Painter and Ethan A. Merritt. TLSMD web server for the generation of multi-group TLS models. *J. Appl. Cryst.*, 39:10911, 2006.

[43] Rensselaer Polytechnic Institute. Protein structure determination, 2012. `http://www.bioinfo.rpi.edu/bystrc/courses/bcbp4870/lecture12.pdf`.

[44] Andrew Moore. Cross-validation for detecting and preventing overfitting. `http://www.autonlab.org/tutorials/overfit10.pdf` School of Computer Science, Carnegie Mellon University.

[45] Axel T. Brünger. Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature*, 355:472 – 475, 1992.

[46] G.J. Kleywegt and T.A. Jones. Model-building and refinement practice. *Methods in Enzymology*, 277:208–230, 1997.

[47] Clemens Vonrhein, Eric Blanc, Pietro Roversi, and Gérard Bricogne. Automated structure solution with autosharp. In Sylvie Doublié, editor, *Macromolecular Crystallography Protocols*, volume 364 of *Methods in Molecular Biology*^{TM}, pages 215–230. Humana Press, 2007.

[48] Serge X. Cohen, Marouane Ben Jelloul, Fei Long, Alexei Vagin, Puck Knipscheer, Joyce Lebbink, Titia K. Sixma, Victor S. Lamzin, Garib N. Murshudov, and Anastassis Perrakis. *ARP/wARP* and molecular replacement: the next generation. *Acta Crystallographica Section D*, 64(1):49–60, Jan 2008.

[49] Airlie J. McCoy, Ralf W. Grosse-Kunstleve, Laurent C. Storoni, and Randy J. Read. Likelihood-enhanced fast translation functions. *Acta Crystallographica Section D*, 61(4):458–464, Apr 2005.

[50] Kevin Cowtan. Fitting molecular fragments into electron density. *Acta Crystallographica Section D*, 64(1):83–89, Jan 2008.

[51] R. A. Laskowski, M. W. MacArthur, D. S. Moss, and J. M. Thornton. *PROCHECK*: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography*, 26(2):283–291, Apr 1993.

[52] Sander C. Kabsch W. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers.*, 22(12):2577–637, 1983.

[53] Isabelle Migneault, Catherine Dartiguenave, Michel J. Bertrand, and Karen C. Waldron. Glutaraldehyde: behavior in aqueous solution, reaction with proteins, and application to enzyme crosslinking. *BioTechniques*, 37:790–802, 2004.

[54] University of British Columbia. `http://www.math.ubc.ca/cass/courses/m309-01a/chu/Fundamentals/snell.htm`.

[55] Daniel Courjon and Claudine Bainier, editors. *Le champ proche optique - Theorie et applications*. Collection technique et scientifique des télécommunications, Springer-Verlag France et France Télécom R&D, 2001.

[56] Kazuyoshi Kurihara and Koji Suzuki. Theoretical understanding of an absorption-based surface plasmon resonance sensor based on Kretchmann's theory. *Anal. Chem.*, 74:696–701, 2002.

[57] Patrick Englebienne, Anne Van Hoonacker, and Michel Verhas. Surface plasmon resonance: principles, methods and applications in biomedical sciences. *Spectroscopy*, 17:255–273, 2003.

[58] S. Boussaad, J. Pean, and N. J. Tao. High-resolution multiwavelength surface plasmon resonance spectroscopy for probing conformational and electronic changes in redox proteins. *Anal. Chem.*, 72:222–226, 2000.

[59] David W. Wilson. Analyzing biomolecular interactions. *Science*, 295:2103–2105, 2002.

[60] Janos Vörös. The density and refractive index of adsorbing protein layers. *Biop Phys J.*, 87:553–561, 2004.

Part III

Results

Chapter 12

Purification and characterization of single proteins

A requirement for crystallization of protein complexes is the production of large amounts of highly pure protein complexes or proteins to be mixed. In order to know the *in vitro* behaviour of each protein better and perform *in vitro* interaction studies at the molecular level, protocols of purifications were set up for each of MurC, MurD, MurE, MurF, MurG, and MreB proteins from *T. maritima*. This chapter aims at reporting the optimization of purification protocols, the quality of the samples finally obtained, and specificities of each proteins that may be important for interaction and crystallization assays. Brief information about experimental procedures is given to facilitate the reading. However, for further details the reader is invited to refer to the Materials and Methods part.

12.1 MurC

MurC catalyzes the first step in the addition of a short peptide to the sugar precursor of peptidoglycan. MurC from *T. maritima* is an UDP-*N*-acetylmuramoyl:L-alanine ligase.

12.1.1 Cloning, protein sequence, and predictions

The *murC* gene from *T. maritima* was cloned into a modified pASK-IBA3C vector, for expression with a Tev-cleavable C-terminal Strep tag.

The resulting protein is made of 476 amino acid residues, has a molecular weight of 54 kDa, and a predicted isoelectric point of 5.7. It presents a theoretical molar extinction coefficient of 46,300 $M^{-1}.cm^{-1}$ corresponding to an absorbance of 0.86 A.U. for 1 mg/ml.

12.1.2 Expression and solubility tests

Expression

pASK-IBA3C-*murC* was transformed into BL21(DE3) and a glycerol stock was prepared. For expression tests, 400 ng/ml anhydrotetracyclin (AHT) were added when OD (optical density) reached 0.5 A.U. for induction at 37°C for 2h30. Induction was clearly detected by SDS-PAGE, though at a moderate level.

Solubility

A first solubility test was performed, revealing that MurC expressed in BL21(DE3) was not soluble, either of the three induction temperatures tested (see Figure 12.1, Left).

Solubility of MurC was then assessed in other bacterial strains: C41(DE3), Origami(DE3), STAR(DE3). Expression levels were similar to those in BL21(DE3) cells, though Origami cells grew very slowly upon induction. MurC was found to be more soluble when expressed in STAR(DE3) cells (see Figure 12.1, Right).

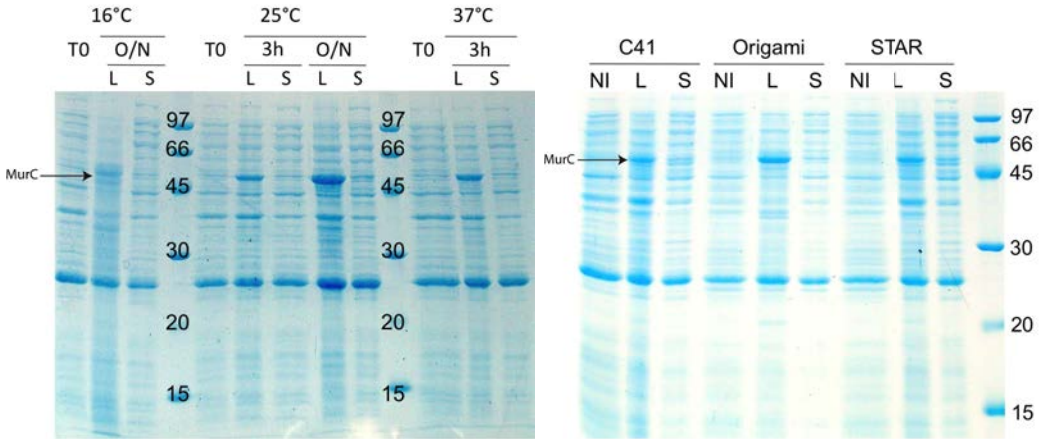


Figure 12.1: Left: SDS-PAGE analysis of MurC solubility for different induction conditions in BL21(DE3). Right: SDS-PAGE analysis for MurC solubility tests in different strains. The solubility of MurC was assessed for different induction temperatures and times in BL21(DE3) cells with 400 ng/ml AHT induction. For strains tests, induction was performed at 37°C for 3 h under 400 ng/ml AHT. T0: before induction; L: lysate after induction; S: soluble part (supernatant) after induction. Equivalent amounts of cells were loaded.

Thus, the STAR(DE3) strain was chosen for further optimization tests. The best induction condition was found to be similar to that in BL21(DE3) cells, that is to say 25°C overnight (see Figure 12.1). MurC solubility did not seem sensitive to inducer concentration, though concentration of AHT below 100 ng/mL remains to be tested.

12.1.3 Purification

MurC purification, as all purifications presented in this work consisted of an affinity chromatography followed by a size exclusion chromatography step.

Figure 12.2 shows typical results obtained for MurC purifications from 1 L of STAR(DE3) cell culture. Despite partial solubility from cell lysates, sufficient amounts of protein could be obtained for interaction assays. A high purity could be achieved though a contaminant of high molecular weight was present, most likely a chaperone from *E. coli*'s cells. However, the protein tended to aggregate. Therefore, we later focused on the other three ligases that appeared to behave better *in vitro*.

12.2 MurD

MurD catalyzes the second step in the addition of the peptide to the sugar precursor of peptidoglycan. MurD is an UDP-*N*-acetylmuramoyl-L-alanine:D-glutamate ligase.

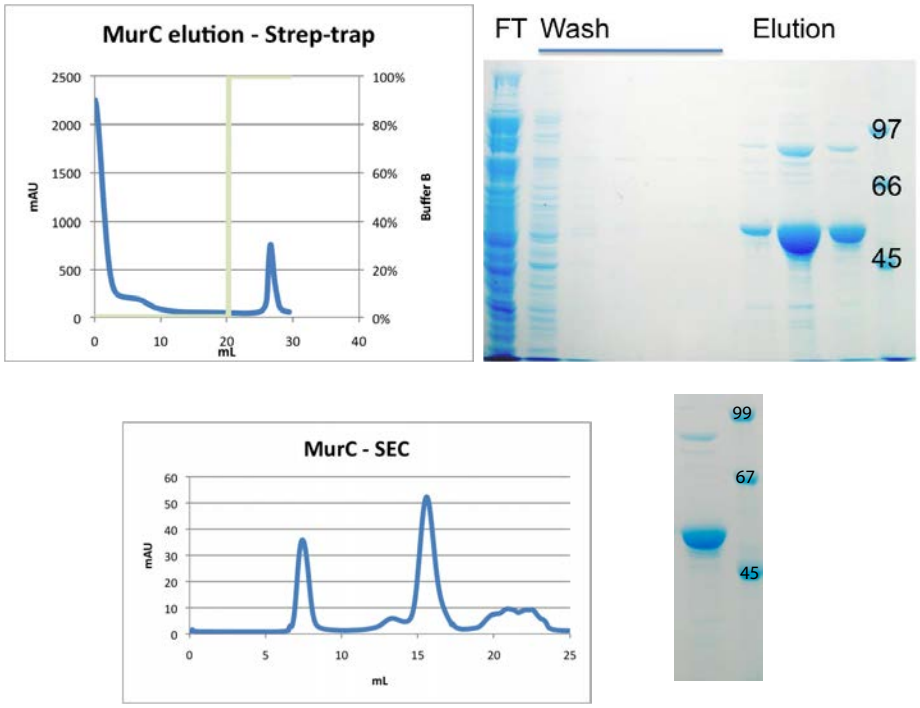


Figure 12.2: MurC purification: affinity chromatography (top) and subsequent size exclusion chromatography (bottom). A 5 ml Strep-trap HP column and a SuperDex200 column from GE Healthcare were used on an Äkta device at room temperature. Left: corresponding chromatograms with absorbance at 280 nm in blue and elution buffer concentration in green. The gel filtration chromatogram reveals the presence of aggregates (column void volume : 8 mL). Right: corresponding SDS-PAGE analyses, showing the persistence of a high-molecular weight contaminant, even after gel filtration. FT: flow through. Lysis buffer: 25 mM NaPi pH 7.4, 10 % glycerol, 0.15 M NaCl, protease inhibitors. Washing buffer: 25 mM NaPi pH 7.4, 5 % glycerol, 0.15 M NaCl. Elution buffer: 25 mM NaPi pH 7.4, 5 % glycerol, 0.15 M NaCl, 2.5 mM d-desthiobiotin.

12.2.1 Cloning, protein sequence, and predictions

The *murD* gene from *T. maritima* was cloned into a modified pET15b vector, for expression with a Thrombin-cleavable N-terminal His tag, and in a pETDuet vector, for expression as natural form.

The resulting his-tagged protein is made of 450 amino acid residues (430 for the untagged protein), has a molecular weight of 51.3 kDa (49.1 kDa for the untagged protein), and a predicted isoelectric point of 6.0. It contains two cysteines and presents a theoretical extinction coefficient of 44,935 $M^{-1}.cm^{-1}$ corresponding to an absorbance 0.88 A.U. for 1 mg/ml.

12.2.2 Expression and solubility tests

Expression of MurD was assessed in BL21(DE3), RIL, and Rosetta2 cells, after 2h30 of induction at 37°C. As shown in Figure 12.3, expression was high in all strains. For further experiments, we chose the BL21(DE3) strain.

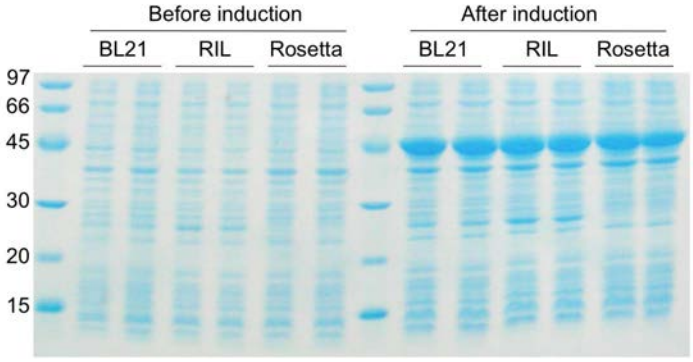


Figure 12.3: Expression tests of MurD: SDS-PAGE analysis for three different *E. coli* strains. The two lanes for each strain correspond to two different colonies. Samples were loaded onto a 12.5 % SDS-PAGE in equivalent amounts of cells.

MurD solubility was assessed for three different induction temperatures. Figure 12.4 shows that MurD was very soluble and that the best induction condition was 25°C overnight.

No additional expression and solubility tests were performed for his-MurD expression since these conditions worked very well.

12.2.3 Purification of native his-tagged MurD

Figure 12.5 shows typical results obtained for MurD purifications from 500 mL to 1 L of cell culture, yielding high amounts of pure protein.

In size exclusion chromatography, MurD eluted in a sharp peak as a monomer, and 10 to 20 % of the protein eluted as a dimer, depending mainly on protein concentration. For crystallization assays, only fractions from the monomer peak were pooled.

In order to be able to test Mur interactions through cross-linking experiments, Tris buffer used for solubility tests was replaced by Sodium phosphate buffer. Indeed, Tris buffer contains a free amine group which may interfere with cross-linking experiments. For crys-

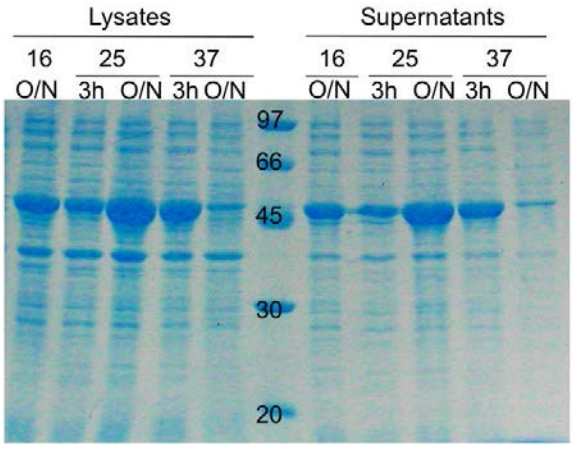


Figure 12.4: Solubility tests of MurD for different induction temperatures and times in BL21(DE3) cells. Induction was done at 16°C overnight, 25°C for 3 h or overnight, or 37°C for 3 h or overnight. Samples were loaded onto a 12.5 % SDS-PAGE in equivalent amounts of cells.

tallization trials, Hepes buffer was used to avoid formation of phosphate crystals. MurD did not seem sensitive to the nature of the buffer.

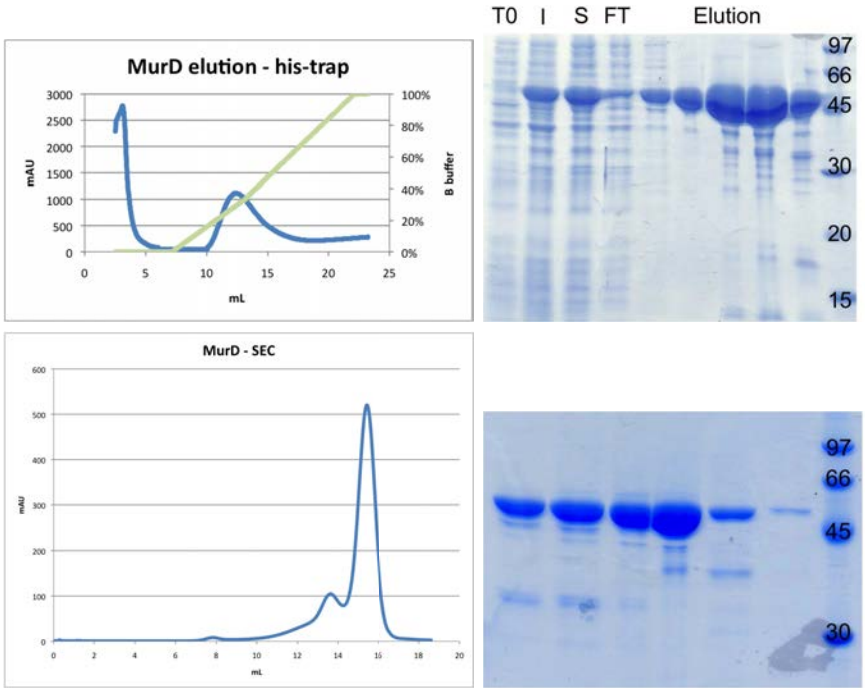


Figure 12.5: MurD purification: affinity (top) and size exclusion (bottom) chromatographies. A 1 ml His-trap FF column and a SuperDex200 column from GE Healthcare were used on an Äkta device at room temperature. Left: corresponding chromatograms with absorbance at 280 nm (mAU) in blue and elution buffer concentration in green. Right: corresponding SDS-PAGE analyses. T0: before induction, I: after induction, FT: flow through, S: supernatant. Lysis buffer: 25 mM NaPi pH 7.4, 5% glycerol, 0.15 M NaCl, protease inhibitors. Washing buffer: 25 mM NaPi pH 7.4, 5% glycerol, 0.15 M NaCl, 25 mM imidazole. Elution buffer: 25 mM NaPi pH 7.4, 5% glycerol, 0.15 M NaCl, 500 mM imidazole. Gel filtration buffer: 25 mM NaPi pH 7.4, 5% glycerol, 0.15 M NaCl.

12.2.4 Purification of untagged MurD

Purification by heating

As the studied proteins come from a thermophilic organism, a high temperature could be required for the detection of protein-protein interactions. That is the reason why heat purifications were tried.

Untagged MurD, expressed from pETDuet-*murD* vector was purified by heating. Sample was then concentrated and loaded onto a size exclusion chromatography, indicating a high amount of DNA contamination. Treatment with PEI[1] did not result in a sufficient reduction of DNA contamination as detected by measuring the A_{260}/A_{280} ratio. Therefore, affinity purification was finally preferred to heat purification.

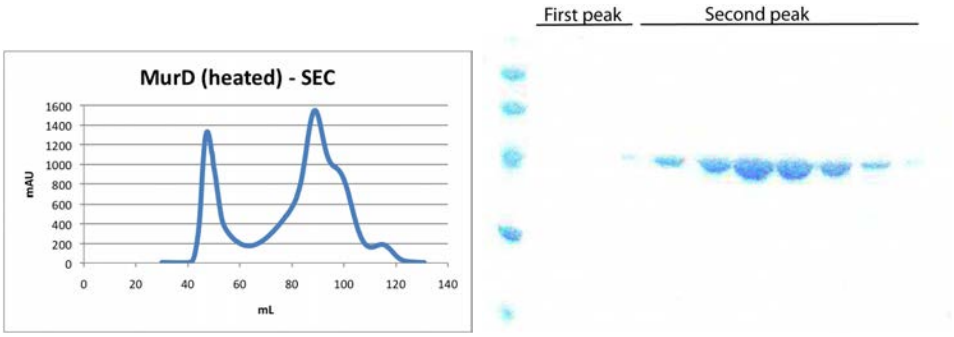


Figure 12.6: Purification of untagged MurD by heating: left, size exclusion chromatography (absorbance at 280 nm); right, SDS-PAGE analysis of SEC fractions showing that the peak at the void volume of the column does not contain any protein. Lysis buffer: 50 mM Hepes pH 7.4, 0.1 M NaCl, 10% glycerol, 5 mM EDTA supplemented with protease inhibitor, DNase, 2.4 μl/ml of $MgSO_4$ 1M and 2.4 μl/ml of $MnCl_2$ 1M. Supernatants were concentrated with a Vivaspin20 30,000 MWCO from Sartorius Stedim Biotech before gel filtration chromatography (SuperDex200). Gel filtration buffer: 25 mM Hepes pH 7.4, 0.15 M NaCl.

Tag cleavage of his-tagged MurD

In order to assess MurD crystallization with no tag and no DNA contamination, the His tag was cleaved with Thrombin from the his-tagged MurD enzyme, using 5 units of protease per mg of ligase.

12.2.5 Purification of seleno-methionylated his-MurD

As molecular replacement did not work for native MurD with homologs (see next part), a SeMet derivative was prepared from 1 L culture in M9 medium as for its native counterpart, giving similar results.

12.2.6 Characterization

General considerations about stability

MurD from *T. maritima* is very stable. No degradation was detected and no precipitation was seen at work concentrations (15 mg/ml maximum). After several months at 4°C, a

[1]Polyethylenimine is a highly positively charged polymer. At neutral pH PEI interacts strongly with nucleic acids, which are negatively charged, causing their precipitation and thus removal from the sample.

small part of the sample was aggregated according to gel filtration assay, but no degradation was visible either on gel filtration or SDS-PAGE. However, for all crystallization and surface plasmon resonance assays, the protein was freshly prepared (one-week-old maximum).

Nevertheless, our attempts to dialyze MurD failed, either because the concentration was too high or because MurD interacted with the dialysis membrane, resulting in protein precipitation.

Thermal Shift Assay

In order to check folding and stability, a TSA was performed on his-MurD protein in Sodium phosphate buffer, resulting in a very high melting temperature. Data, which are limited by the 100°C maximum, did not allow to determine the melting temperature precisely. However, it is clearly higher than 90°C.

A TSA performed by HTX lab gave similar results for untagged MurD in Hepes buffer, confirming that MurD does not present a strong sensitivity to buffer nature.

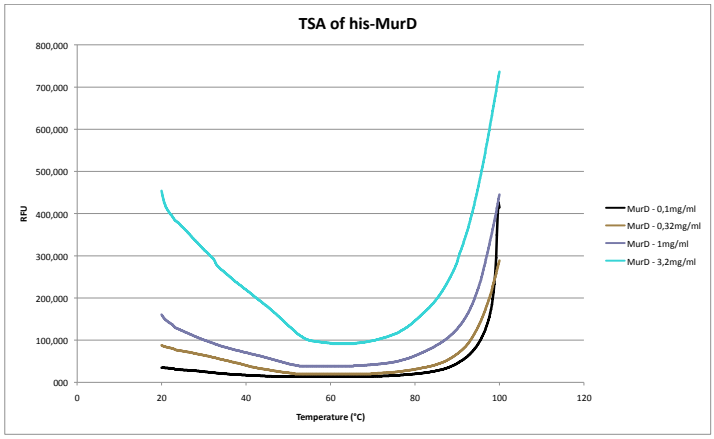


Figure 12.7: Thermal shift assay of his-MurD. Assay performed at the IBS, on a BioRadiQ5 device, with the SYPRO orange probe in 25 mM NaPi pH 7.4, 0.15 M NaCl, at four different concentrations.

Mass Spectrometry

Mass Spectrometry analysis gave a measured mass in total agreement with the theoretical mass of his-MurD, sometimes with loss of the N-terminal methionine. A disulfide bridge could be seen for samples with no DTT.

12.3 MurE

MurE catalyzes the third step in the sequential addition of the peptide to the sugar precursor of peptidoglycan. MurE from *T. maritima* is an UDP-*N*-acetylmuramoyl-L-alanine-D-glutamate:L-lysine ligase.

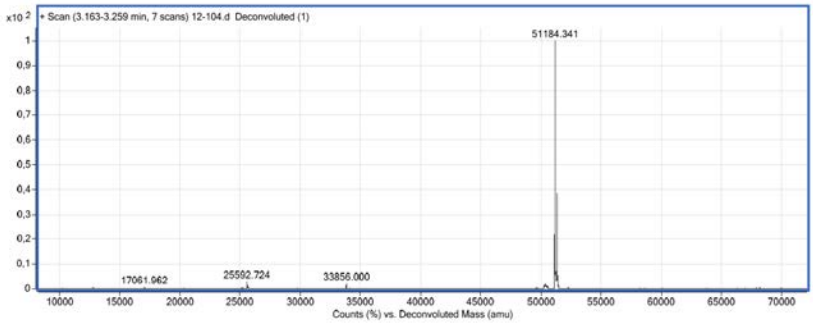


Figure 12.8: Mass spectrometry analysis of his-MurD.
Deconvoluted spectrum from a ESI-TOF MS analysis.

12.3.1 Cloning, protein sequence, and predictions

murE gene from *T. maritima* was cloned in a pET30b vector, for expression with a C-terminal His tag.

The resulting protein is made of 498 amino acid residues, has a molecular weight of 55.8 kDa, and a predicted isoelectric point of 6.0. It contains one cysteine and presents a theoretical extinction coefficient of 30,370 $M^{-1}.cm^{-1}$ corresponding to an absorbance of 0.54 A.U. for 1 mg/ml.

12.3.2 Purification of native MurE

As for MurD, BL21(DE3) bacteria were used to express MurE upon induction conditions of 1 mM IPTG overnight at 20-25°C.

Purifications were done from 1 L to 2 L of cell culture, with an affinity chromatography step followed by a size exclusion chromatography step, yielding about 15 mg of pure protein per liter of culture. The yield was lower than other Mur ligases because MurE was slightly toxic to the cells. Indeed, upon induction, BL21(DE3) cells grew very slowly. Therefore, for all subsequent purifications, cells were induced at an OD of 0.9-1 A.U., in order to increase the yield. This toxicity was described in a paper from Boniface *et al.*, 2006 [1].

A typical purification is depicted Figure 12.9. As for MurD, MurE eluted mainly in monomeric form, and the dimeric form could be detected on size exclusion chromatography as well.

As MurE was not sensitive to purification buffer, Hepes was eventually used to facilitate crystallization and interaction assays.

12.3.3 Heat purification of MurE

A purification at small scale (50 ml of culture) of MurE by heating was tried. The sample looked slightly less pure and much less homogeneous than with affinity purifications. Thereupon, no additional heat purification was performed for MurE.

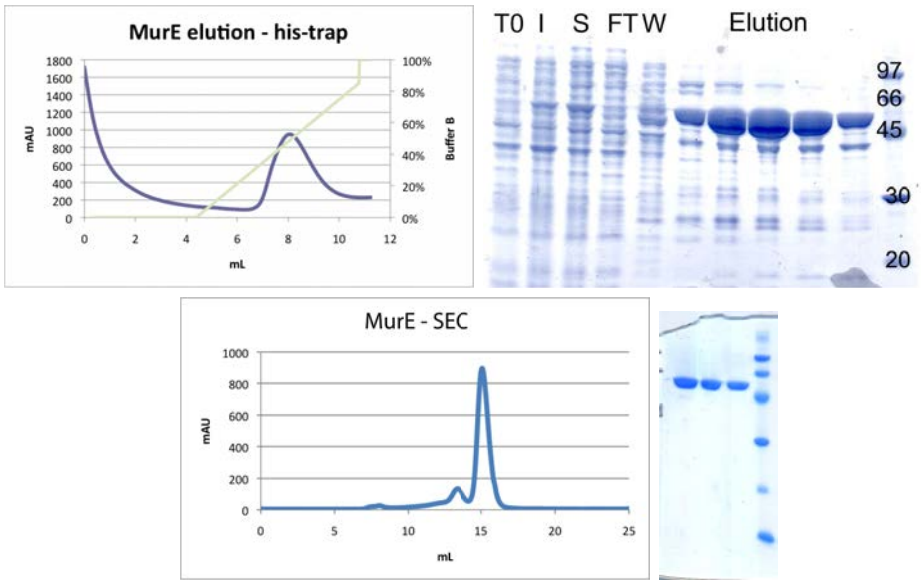


Figure 12.9: MurE purification: affinity (top) and size exclusion (bottom) chromatographies.
A 1 ml His-trap FF column and a SuperDex200 column from GE Healthcare were used on an Äkta device at room temperature. Left: corresponding chromatograms with absorbance at 280 nm (mAU) in blue and elution buffer concentration in green. Right: corresponding SDS-PAGE analyses. T0: before induction, I: after induction, FT: flow through, S: supernatant, W: wash. Lysis buffer: 25 mM NaPi pH 7.4, 5% glycerol, 0.15 M NaCl, protease inhibitors. Washing buffer: 25 mM NaPi pH 7.4, 5% glycerol, 0.15 M NaCl, 25 mM imidazole. Elution buffer: 25 mM NaPi pH 7.4, 5% glycerol, 0.15 M NaCl, 500 mM imidazole. Gel filtration buffer: 25 mM NaPi pH 7.4, 5% glycerol, 0.15 M NaCl.

12.3.4 Purification of seleno-methionylated MurE

Because molecular replacement failed (see next chapter), the seleno-methionylated derivative was prepared. Se-Met MurE-his was first purified from 2 L culture in M9 medium as for its native counterpart, giving similar results.

However, mass spectrometry analysis revealed that the protein sample was heterogeneous, presenting -27 Da sequential shifts as illustrated in Figure 12.10, top. It is of note that no crystals could be obtained from this batch. BL21(DE3) cells may have incorporated wrong amino acids into *T. maritima* MurE, because of the presence of rare codons for *E. coli* in the nucleotide sequence. Using Rosetta(DE3) cells which are optimized for rare codons, and performing a slow induction at 20°C a more homogeneous protein sample could be obtained (see Figure 12.10, bottom).

12.3.5 Characterization

MurE from *T. maritima* is very stable. No degradation was detected and precipitation was seen at very high concentrations (above 25 mg/ml). As for MurD, after several months at 4°C, a small part of the sample was aggregated according to a gel filtration assay, but no degradation was visible either on gel filtration or SDS-PAGE. Despite of this apparent high stability, the protein was freshly prepared for all crystallization and surface plasmon resonance assays (one-week-old maximum).

However, attempts to dialyze MurE failed, either because the concentration was too high or because MurE interacted with the dialysis membrane, resulting in protein precipitation.

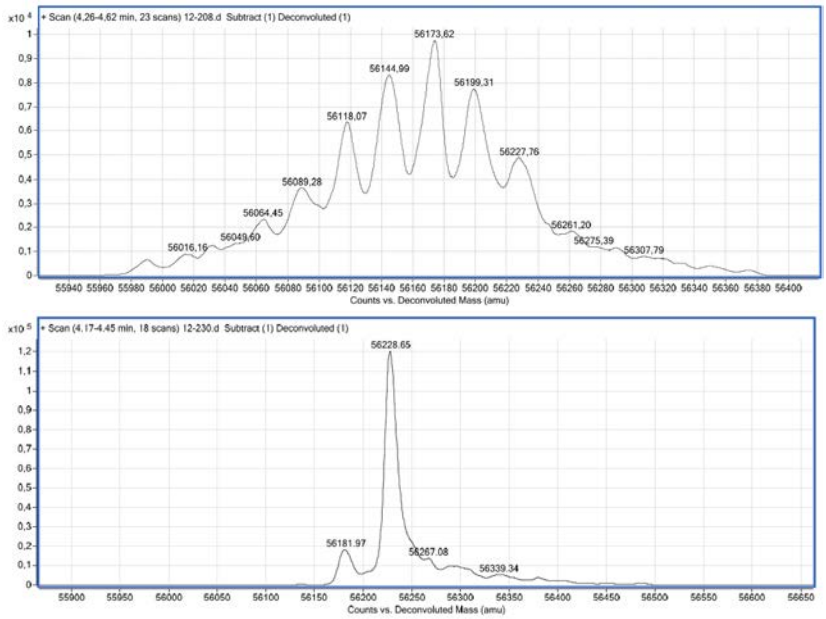


Figure 12.10: Mass spectrometry analysis of seleno-methionylated MurE-his coming from BL21(DE3) cells (left) and Rosetta(DE3) cells (right). Deconvoluted spectrum from ESI-TOF MS analyses performed on IBS MS platform showing -27 Da shifts in SeMet-MurE coming from BL21(DE3) cells and a more homogeneous sample when MurE was expressed in Rosetta(DE3) cells at low induction temperature.

Thermal Shift Assay

In order to check folding and stability, a TSA was performed on MurE-his in Sodium phosphate buffer, resulting in a melting temperature higher than 80°C.

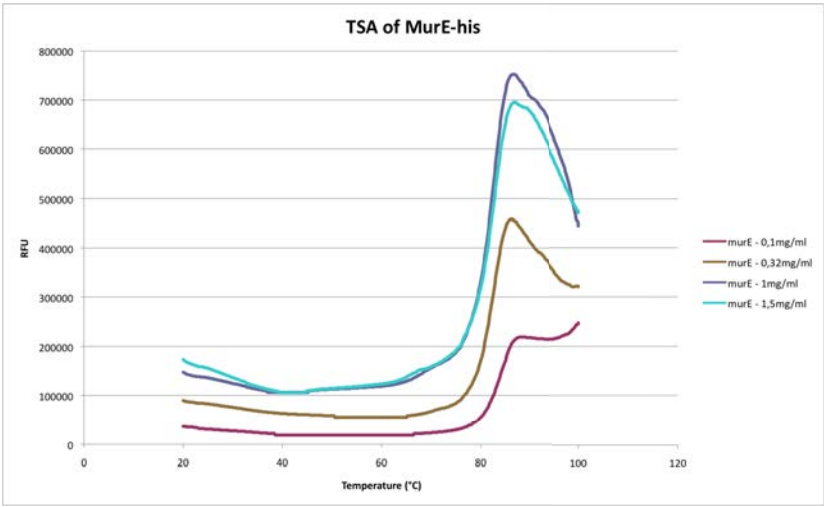


Figure 12.11: Thermal shift assay of MurE-his. Left: assay performed at the IBS, on a BioRadiQ5 device, with SYPRO orange probe in 25 mM NaPi pH 7.4, 0.15 M NaCl, at four different concentrations.

Mass Spectrometry

Sample of native MurE was analyzed by mass spectrometry and lead to the detection of a macromolecule of a redueced measured mass, suggesting that the N-terminal methionine was cleaved.

12.4 MurF

MurF catalyzes the last step in the sequential addition of the peptide to the sugar precursor of peptidoglycan. MurF from *T. maritima* is an UDP-*N*-acetylmuramoyl-L-alanine-D-glutamate-L-lysine:D-alanyl-D-alanine ligase [2].

12.4.1 Cloning, protein sequence, and predictions

murF gene from *T. maritima* was cloned in a pET30b vector, for expression with a C-terminal His tag.

The resulting protein contains 437 amino acid residues, has a molecular weight of 49.2 kDa, and a predicted isoelectric point of 6.9. It contains two cysteines and presents a theoretical extinction coefficient of 37,025 $M^{-1}.cm^{-1}$ corresponding to an absorbance at 280 nm of 0.75 A.U. for 1 mg/ml.

12.4.2 Purification of native MurF

BL21(DE3) bacteria were used to induce MurF expression at an OD of 0.5 A.U. with 1 mM IPTG overnight at 20-25°C. Purifications were done from 500 mL of cell culture, easily yielding 60 mg of highly pure enzyme since MurF was very highly over-expressed. As for MurD and MurE ligases, a part of MurF appeared as a dimeric form in size exclusion chromatography. Hepes was preferred for all subsequent purifications.

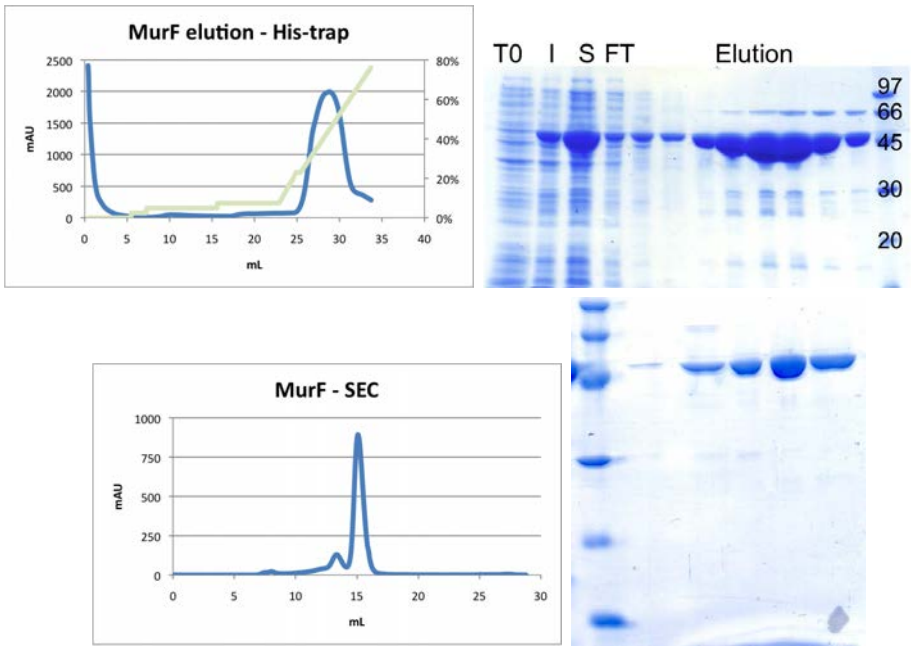


Figure 12.12: MurF purification: affinity (top) and size exclusion (bottom) chromatographies. A 1 ml His-trap FF column and a SuperDex200 column from GE Healthcare were used on an Äkta device at room temperature. Left: corresponding chromatograms with absorbance at 280 nm (mAU) in blue and elution buffer concentration in green. Right: corresponding SDS-PAGE analyses. T0: before induction, I: after induction, FT: flow through, S: supernatant. Lysis buffer: 25 mM NaPi pH 7.4, 5% glycerol, 0.15 M NaCl, protease inhibitors. Washing buffer: 25 mM NaPi pH 7.4, 5% glycerol, 0.15 M NaCl, 25 mM imidazole. Elution buffer: 25 mM NaPi pH 7.4, 5% glycerol, 0.15 M NaCl, 500 mM imidazole. Gel filtration buffer: 25 mM NaPi pH 7.4, 5% glycerol, 0.15 M NaCl.

12.4.3 Heat purification of MurF

A purification at small scale (50 ml of culture) of MurF by heating was tried as well. The sample looked less pure and much less homogeneous than with affinity purifications, and a significant part of the sample was aggregated. Thus, no additional heat purification was performed for MurF.

12.4.4 Characterization

MurF from *T. maritima* is very stable. Indeed, after several months at 4°C, though a small part of the sample was aggregated according to a gel filtration assay, no degradation was detected either by gel filtration or SDS-PAGE.

However, a degradation was detected after two months at 4°C by mass spectrometry in a sample without EDTA (see below) and precipitation was seen at very high concentrations (above 40 mg/ml). For instance, MurF was seen to precipitate just after elution when the concentration was too high, probably due to the very high expression level of this enzyme in BL21(DE3) cells. In addition, as for MurD and MurE, attempts to dialyze MurF failed either because the concentration was too high or because MurF interacted with the dialysis membrane, resulting in protein precipitation.

Fresh protein (less than one-week-old) was prepared for all crystallization and surface plasmon resonance assays.

Thermal Shift Assay

In order to check folding and stability, a TSA was performed on MurF-his in Sodium phosphate buffer, resulting in a melting temperature higher than 80°C. This result was confirmed by HTX lab.

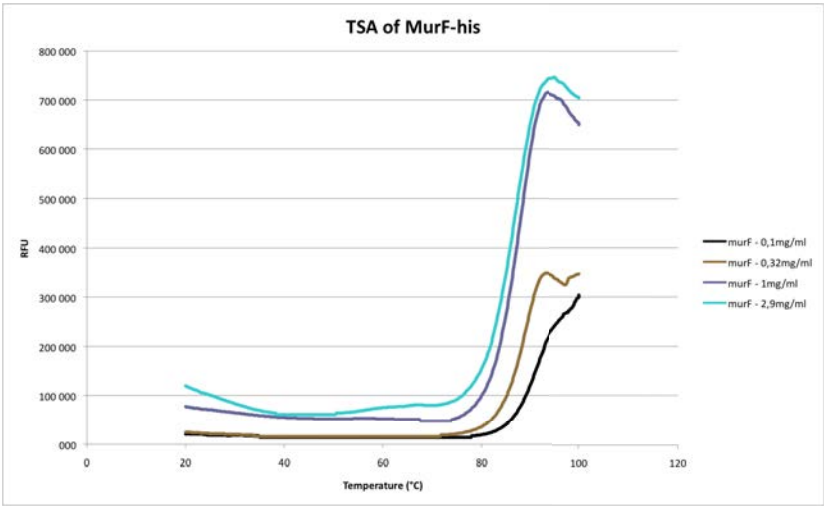


Figure 12.13: Thermal shift assay of MurF-his. Left: assay performed at the IBS, on a BioRadiQ5 device, with SYPRO orange probe in 25 mM NaPi pH 7.4, 0.15 M NaCl at four different concentrations.

Mass Spectrometry

Sample of native MurF was analyzed by mass spectrometry, leading to a measured mass in agreement with the theoretical molecular weight. However, for some batches, heterogeneity (-27 Da species) was detected as for seleno-methionine MurE, though at a lower extent. In addition, after two months in a buffer without EDTA, the protein was degraded (see Figure 12.14).

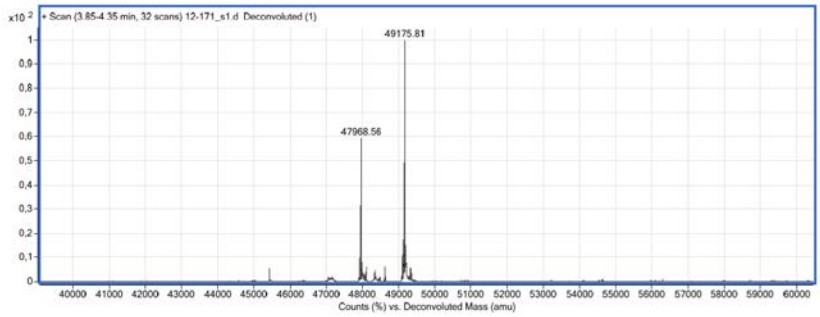


Figure 12.14: Mass spectrometry analysis of MurF-his. Deconvoluted spectrum from a ESI-TOF MS analysis of a two-month-old MurF sample, showing a degraded form in addition to the full-length protein.

12.4.5 Purification of seleno-methionylated MurF

Molecular replacement having failed (see next chapter), a Se-Met derivative of MurF-his was prepared in the same way as its native counterpart, giving similar results. Se-Met incorporation was checked by mass spectrometry.

12.5 MurG

Once MraY has added a lipid carrier to the UDP-Mur*N*Ac-peptide previously synthesized by MurF, MurG transfers *N*-acetylglucosamine (GlcNAc) to Lipid I to form Lipid II, which is the monomer of the final peptidoglycan polymer [2].

12.5.1 Cloning, protein sequence, and predictions

murG gene from *T. maritima* was cloned in a modified pASK-IBA3C vector, for expression with a Tev-cleavable C-terminal Strep tag.

The resulting protein is made of 358 amino acid residues, has a molecular weight of 39.9 kDa, and a predicted isoelectric point of 8.6. It contains one cysteine and presents a theoretical extinction coefficient of 49,390 $M^{-1}.cm^{-1}$ corresponding to an absorbance of 1.24 A.U. for 1 mg/ml.

12.5.2 Expression and solubility tests

pASK-IBA3C-*murG* was transformed into BL21(DE3) bacteria for expression and solubility tests. Though over-expression was high, the protein was not soluble as illustrated in Figure 12.15.

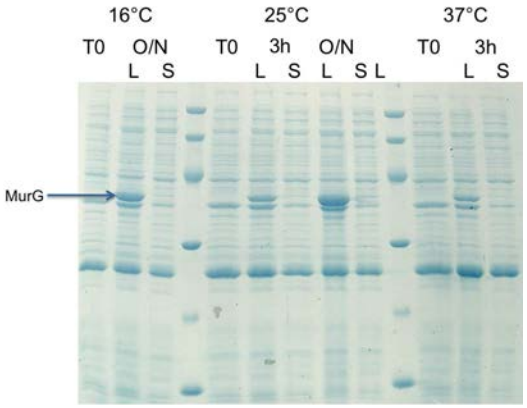


Figure 12.15: SDS-PAGE analysis of MurG solubility tests in BL21(DE3). Solubility of MurG was assessed for different induction temperatures and induction durations in BL21(DE3) cells with 400 ng/ml AHT induction. T0: before induction; L: lysate after induction; S: soluble part (supernatant) after induction. Equivalent amounts of cells were loaded.

Consequently, other bacterial strains were assessed for MurG expression and solubility. MurG was not expressed in neither Origami nor C41(DE3) strains while a good expression level was obtained with STAR cells. However, solubility was not higher in STAR cells, in contrast to MurC.

As *E. coli* has been reported to attach to cell membranes, a detergent screen was performed to solubilize *T. maritima* MurG enzyme from BL21(DE3) cell lysates. The only detergent that was able to solubilize a major part of the expressed MurG appeared to be N-lauroyl sarcosine and was used for all subsequent purifications of MurG.

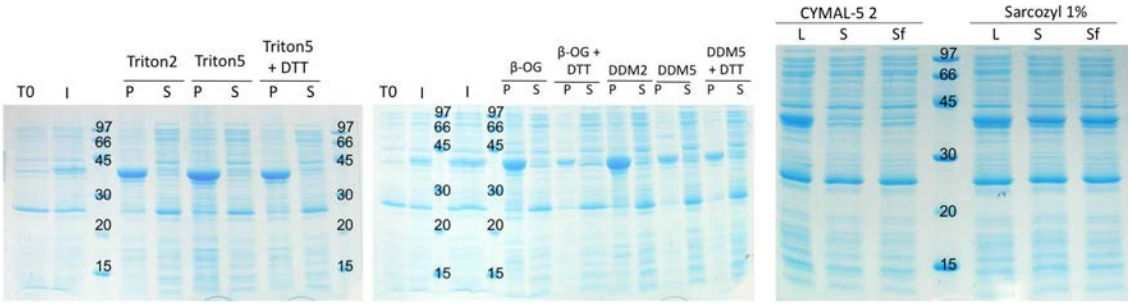


Figure 12.16: MurG solubility tests in detergents. MurG was induced in STAR cells with 400 ng/ml AHT at 37°C for 3 h. Cells were lysed in the presence of different detergents by sonication. The indicated number after detergent name indicates the concentration in cmc units. When indicated, DTT was added at 5 mM. T0: before induction; I: after induction; L: lysate ; P: pellet; S and Sf: soluble parts from higher centrifugation speed (supernatant). Equivalent amounts of cells were loaded.

12.5.3 Purification

Classical protocol

1 L of BL21(DE3) cells carrying pASK-IBA3C-*murG* were induced at an OD of 1 A.U. with 400 ng/ml AHT overnight at 20-25°C. Purifications were done from 1 L of cell culture. Cells grew slowly upon MurG expression, with an OD after overnight induction of about 1.3 A.U.

The result of a typical affinity chromatography experiment is shown Figure 12.17. A notable amount of MurG was found in the flow through. This could have been due to the

presence of MurG oligomers (see below) which could have hidden the Strep tag and prevented it from binding to the affinity column.

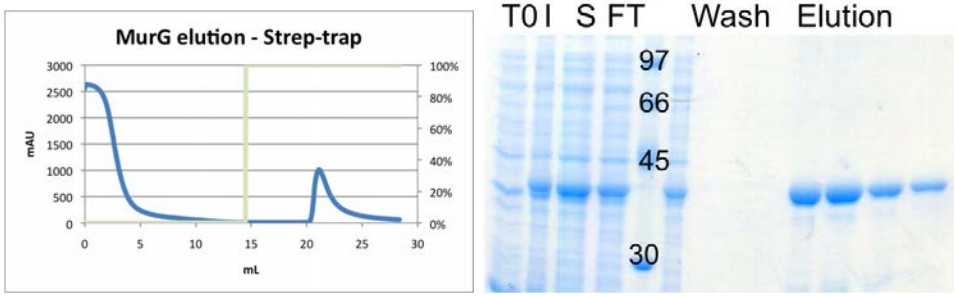


Figure 12.17: MurG purification: affinity chromatography. Purification on a 5 ml Strep-trap HP column from GE Healthcare, on a Äkta device at room temperature. Lysis buffer: 50 mM Hepes pH 7.4, 5% glycerol, 0.3 M NaCl, 10 mM DTT, protease inhibitors. Washing buffer: 50 mM Hepes pH 7.4, 5% glycerol, 0.3 M NaCl, 10 mM DTT. Elution buffer: 50 mM Hepes pH 7.4, 5% glycerol, 0.3 M NaCl, 10 mM DTT, 2.5 mM d-desthiobiotine. Left: absorbance at 280 nm (mAU) in blue, and % of elution buffer in green. Right: SDS-PAGE analysis of affinity chromatography. T0: before induction, I: after induction, S: supernatant, FT: flow-through.

The purest fractions were pooled and concentrated before further purification by size exclusion chromatography. A SuperDex200 column was tried first. However, MurG appeared to be aggregated and aggregation was increasing over the time as described in Figure 12.18. Therefore, DDM was added to purification buffers to prevent this phenomenon. Nevertheless, MurG appeared to interact with the column, resulting in an increase in column pressure and loss of part of the sample. The protein could be detached by washing the column in NaOH 0.5 M. Another gel filtration column was thus used to get rid of this problem, namely the Superose6. The difference between these two gel filtration columns is the absence of dextran matrix in the Superose6 column.

After optimization of purification buffers and gel filtration column, a protocol was finally set up for MurG purification, yielding high quantities of very pure protein. MurG eluted as monomeric and oligomeric forms, depending mostly on its concentration, and time between affinity and gel filtration chromatographies.

Since DTT did not enhance purification results, it was not added in subsequent purifications of the glycosyltransferase.

Membrane extraction protocol

A purification of MurG by membrane extraction was assessed. This experiment allowed a purification of good quantities of relatively pure protein, much as the initial affinity purification protocol, supporting the hypothesis that MurG from *T. maritima* is attached to cell membranes. However, as such a protocol is much more time consuming, normal purifications were preferred.

Strep tag cleavage

Strep tag could be cleaved easily by the Tev protease. However, because of the relatively low stability of the enzyme, for most assays the tag was not cleaved in order to reduce the number of purification steps.

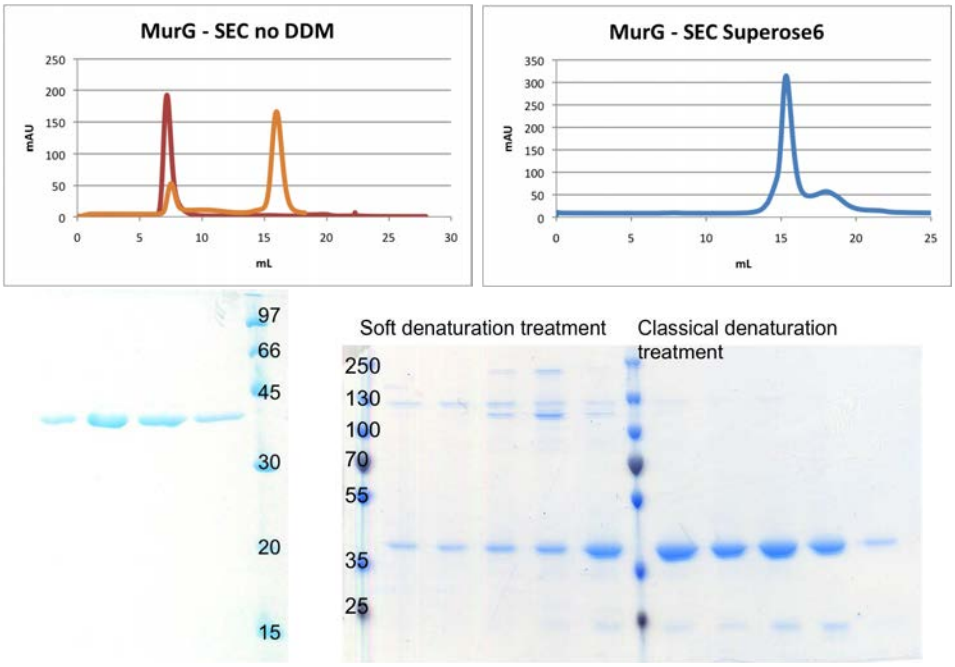


Figure 12.18: MurG purification: size exclusion chromatographies. Top left: two serial runs on a SuperDex200 gel filtration column revealing progressive aggregation of MurG in the absence of detergent. Orange: first run; Red: second run. Top right: SEC on Superose6 column in the presence of 0.04 % DDM. No more aggregation and two oligomeric states are detected, most likely the monomer (elution volume of 18 ml) and a tri- or tetramer (elution volume of 15 ml).

Heat purification

A purification of MurG by heating was assessed as for the MurE enzyme. Though sarkozyl was added as for other purifications, more than 90 % of the protein remained in the pellet, suggesting a precipitation of MurG-containing entities. Therefore, no additional trials were performed to heat-purify MurG.

12.5.4 Characterization

Stability

MurG tended to precipitate at relatively low concentration (2-3 mg/ml). In addition, some degradation could be seen by SDS-PAGE, resulting into two or three bands with a difference of about 1-5 kDa as compared to the full-length protein, even in the presence of 1 mM EDTA. This could be seen directly after affinity chromatography, and the effect increased over the time.

Recent purification trials suggested that MurG could be more stable in Tris buffer, though the purity achieved seemed to be lower.

Mass Spectrometry

Mass Spectrometry analyses on purified MurG detected a protein at the expected mass. However, for one batch of protein, an additional mass of 38,374.50 Da could be detected that could correspond to the first 345 residues of MurG. This could correspond to one of the degradation bands often seen by SDS-PAGE.

Moreover, a similar heterogeneity to that observed for seleno-methionylated MurE could be detected, with sequential -27.6 Da differences between species. It seems that the induction temperature is linked to the heterogeneity of the sample.

Electron Microscopy

Electron Microscopy assays showed that MurG oligomerizes in small ring-like particles (see Figure 12.19). Though the heterogeneity of the sample did not allow to determine any model, it seems that particles are about 3 to 5 entities. Optimization of purification buffers is needed to achieve higher homogeneity.

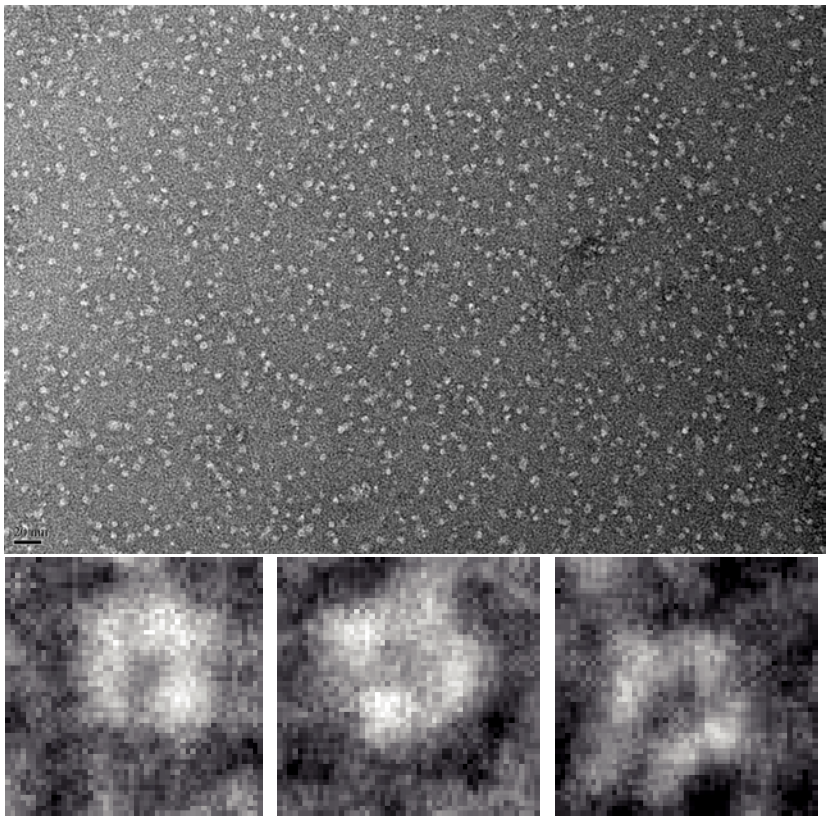


Figure 12.19: Electron Microscopy analysis of a MurG sample. MurG forms ring-like particles. Sodium silico tungstate was used for negative staining (best stain).

12.6 MreB

12.6.1 Cloning, protein sequence, and predictions

The *mreB1* gene from *T. maritima* was cloned into a modified pET30b vector, for expression with a C-terminal His tag, as previously reported in van den Ent *et al.*, 2001 [3].

The resulting protein is made of 344 amino acid residues, has a molecular weight of 36.7 kDa, and a predicted isoelectric point of 6.0. It presents a theoretical extinction coefficient of 11,460 $M^{-1}.cm^{-1}$ corresponding to an absorbance of 0.31 A.U. for 1 mg/ml.

12.6.2 Purification of native MreB

C41(DE3) cells were used for MreB over-expression as described in van den Ent *et al.* [3], though generally an overnight induction at 25°C was preferred.

Purification by affinity and size exclusion chromatography

Purifications were done from 2 L of cell culture after ultracentrifugation of lysates as MreB polymerizes within the cells, by an affinity step followed by a size exclusion chromatography as for other proteins. This protocol gave 10 mg of highly pure and homogeneous protein. A typical purification is depicted Figure 12.20. It is of note that a high concentration of imidazole (86 mM) was required in washing buffer to obtain a good purity. MreB eluted as a monomer in size exclusion chromatography assays.

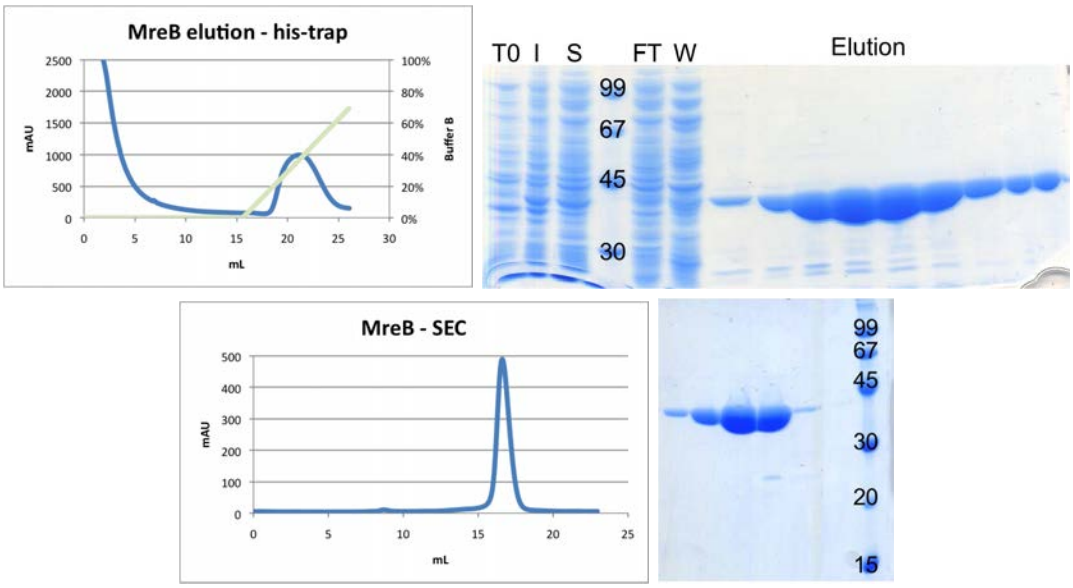


Figure 12.20: MreB purification: affinity (top) and size exclusion (bottom) chromatographies. A 1 ml His-trap FF column and a SuperDex200 column from GE Healthcare were used on an Äkta device at room temperature. Left: corresponding chromatograms with absorbance at 280 nm (mAU) in blue and elution buffer concentration in green. Right: corresponding SDS-PAGE analyses. T0: before induction, I: after induction, S: supernatant, FT: flow through, W: wash. Lysis buffer: 25 mM Hepes pH 7.4, 0.15 M NaCl, DNase, lysozyme, protease inhibitors. Washing buffer: 25 mM Hepes pH 7.4, 0.15 M NaCl, 35 mM imidazole. Elution buffer: 25 mM Hepes pH 7.4, 0.15 M NaCl, 500 mM imidazole. Gel filtration buffer: 25 mM Hepes pH 7.4, 0.15 M NaCl, 1 mM EDTA.

Heat purification

A purification of MreB by heating was assessed as for other studied proteins. Though the purity seemed good, most of the protein was in the pellet, suggesting polymerization or aggregation of MreB could have occurred during the heating step. Therefore, the above mentioned classical purification scheme was preferred.

12.6.3 Purification of seleno-methionylated MreB

In order to phase potential crystals of the MreB-MurE complex (see Chapter 14), Se-Met MreB was prepared. As the SeMet derivative could not be expressed in C41(DE3) cells,

pET30b-*mreB* plasmid was transformed into BL21(DE3) cells. Upon induction, cells did not grow but expressed the protein of interest. 1 L of culture yielded a few mg of protein.

However, mass spectrometry analysis revealed that seleno-methionine had not been incorporated at 100 % in this batch. Indeed, three species were detected with 7, 8, or 9 seleno-methionines over the 9 methionines of the protein sequence.

12.6.4 Characterization

Stability

MreB tended to precipitate at a relatively low concentration (3 mg/ml). In addition, a notable degradation could be seen on SDS-PAGE when no EDTA was added in the gel filtration buffer. Attempts to dialyze the protein resulted in precipitation.

However, at low concentration and in the presence of EDTA, the protein could be easily stored in stable condition for one week. Strikingly, at 2-3 mg/ml, the MreB solution often appeared white. Nevertheless, this was not a precipitate since centrifugation did not result in pelleting. This might have been the consequence of a spatial ordering of MreB molecules in solution. This property did not prevent the protein from crystallizing.

Mass Spectrometry

Mass Spectrometry analyses on purified MreB detected a macromolecule at the expected mass.

Thermal Shift Assay

TSA performed by the HTX lab resulted in a melting temperature of 52°C. This seemed low for a thermostable protein. However, MreB crystallized very well so no buffer optimization was performed regarding its stability.

In vitro behaviour of Mur ligases, MurG, and MreB

MurD, MurE, MurF from *T. maritima* can be easily purified in high amounts for crystallization purposes and present a high stability. The glycosyltransferase MurG can be obtained in high quantity as well but its stability and homogeneity remain to be enhanced. The bacterial actin homolog MreB can be purified in notable amount and high purity.

12.7 Automation of purifications

After protocols were established, the purification of his-MurD, MurE-his, MurF-his, MreB-his were subsequently automated on the MP3 platform at the IBS. Monomeric ligases could be easily obtained in similar purity and yield. Some optimization was required to get high

amounts of pure MreB, as a high amount of imidazole was required to remove some contaminants. Much more protein than needed was prepared in order to select the purest fractions.

Chapter 13

Crystal structures of MurD, MurE, MurF, and MreB from *T. maritima*

To expand the structural knowledge of the cytoplasmic steps of peptidoglycan biosynthesis, structures of the ligases MurD, MurE, and MurF from *T. maritima* were solved at 2.17 Å, 2.9 Å, and 1.65 Å resolution, respectively. In addition, a structure of MreB from *T. maritima* at 1.44 Å resolution was obtained. Here are reported the crystallization processes for each protein, the X-ray scattering of their crystals, and their atomic structures.

13.1 Crystallization

13.1.1 MurD

High-throughput screening

A crystallization screen in nanodrops (HTX lab, EMBL) for his-MurD mixed with both AMPPNP and UMA (UDP-*N*-acetylmuramoyl-L-alanine) substrate resulted in very few detected conditions, mostly needles or micro-crystals. Ligands were added because they are required for crystallization of the *E. coli* homolog. The best condition was 0.1 M Tris pH 8.5 plus 2.0 M ammonium phosphate, giving rise to nice sticks in 2 months (see Figure 13.1, A).

Other crystallization conditions were found as well, where small needles or sea urchins grew:

- 0.2 M ammonium sulfate, 15% PEG8000
- 0.2 M magnesium chloride, 0.1 M Hepes pH 7.5, 25% PEG3350
- 1 M ammonium sulfate

The high-throughput screens on MurD enzyme without tag did not give better leads, be it from the heat purified enzyme or from the Thrombin-treated enzyme.

Reproducibility and optimization

His-MurD crystals could be manually reproduced. However, high variability in nucleation time and crystal quality was observed. In particular, the nice sticks that had been obtained

in the HTX screen could be reproduced once only, with the PEG8000 condition and after a 3-month growth. Unfortunately, they were too small for X-ray scattering experiments.

Actually, the easiest condition for MurD crystallization was the ammonium phosphate condition. Nevertheless, when hand-made, MurD crystals grew most often as polycrystalline plates. Pictures of MurD crystals are shown Figure 13.1.

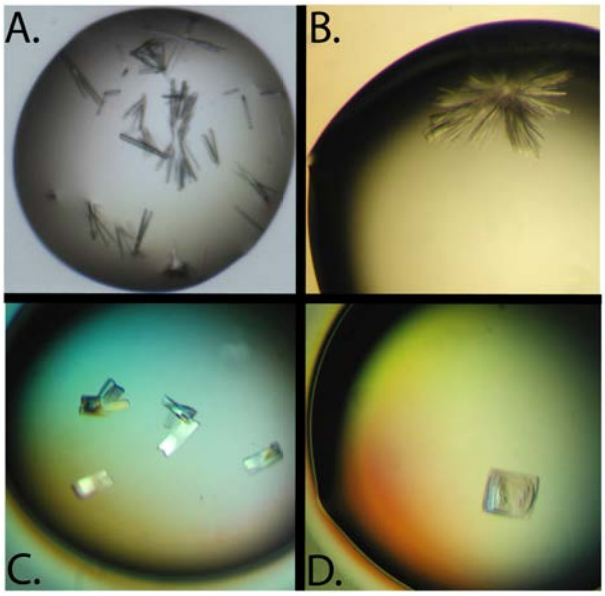


Figure 13.1: MurD crystals. A, best hit from the HTX screen. his-MurD at 11 mg/ml with 5 mM AMPPNP and 1 mM UMA in 0.1 M Tris pH 8.5, 2.0 M ammonium phosphate, 60 days growth, hard to reproduce. B, his-MurD at 11 mg/ml with 5 mM AMPPNP and 1 mM UMA in 0.1 M Tris pH 8.0, 2.0 M ammonium phosphate, 6-week growth, difficult to reproduce. C and D: his-MurD (C), SeMet his-MurD (D) at 10 mg/ml with 5 mM AMPPNP and 1 mM UMA in 2.0 M ammonium phosphate, six- (C) or four- (D) week growth. C and D show typical polycrystalline plates obtained for MurD crystals.

Diffraction and crystallization of SeMet MurD

MurD crystals in the shape of polycrystalline plates diffracted well, up to 2 Å resolution. However, the crystals were often too polycrystalline and obtaining good diffraction data required to select a part of crystal which was the least polycrystalline when flash cooling, properly orient the crystal under the X-ray beam, and reduce the beam size in order to expose only a small section of the crystal, thus avoiding the exposure of several crystal lattices. This technique lead to the collection of a data set to 2.17 Å resolution.

As molecular replacement did not work, SeMet MurD was crystallized in the same condition, producing similar shapes (Figure 13.1, C and D).

13.1.2 MurE

High-throughput screening

MurE-his crystallized very well in a large number of conditions. The best ones are illustrated in Table 13.1.2. Interestingly, many different crystal shapes were seen.

Protein sample	Crystallization liquor	Crystal shape	Growth time
10 mg/ml to 15 mg/ml MurE-his, 5 mM AMPPNP	1.6 M to 2.0 M ammonium sulfate, pH 7.0, 8.0, 8.5, or 9.0 in 0.1 M Hepes, Tris, or Bicine	Large parallelepiped	3 days
10 mg/ml to 15 mg/ml MurE-his, 5 mM AMPPNP	0.2 M NaSCN, 20% PEG3350, pH 6.9	Polycrystalline	3 days
10 mg/ml MurE-his, 5 mM AMPPNP	0.2 M LiSO4, 15% PEG4000, 0.1 M Tris pH 8.5	Needles	3 days
15 mg/ml MurE-his, 5 mM AMPPNP	0.1 M LiCl or 0.1 M $(NH_4)_2PO_4$, 10-12% PEG6000, 0.1 M Tris pH 8.0-8.5	Sticks	3 days
10 mg/ml MurE-his, 5 mM AMPPNP	0.1 M LiCl or 0.1 M $(NH_4)_2PO_4$, 5% PEG6000, 0.1 M citric acid pH 5.0	Sticks	3 days

Table 13.1: Best crystallization conditions obtained from the HTX screen for MurE.

Reproducibility and optimization

Conditions with ammonium sulfate, PEG3350, and PEG6000 could be easily reproduced by hand and no optimization was required as the crystals manually obtained were better than the crystals from nanodrops. Interestingly, the PEG3350 condition lead to various crystal shapes as illustrated Figure 13.2.

Figure 13.2: MurE crystals. A to D, MurE-his at 9 mg/ml with 5 mM AMPPNP crystallized in a crystallization liquor made of 0.1 M Hepes pH 7.0, 12-18% PEG3350, 0.2-0.3 M NaSCN. Hand-made drops of 1 μl protein with 1 μl crystallization solution.

Diffraction

Crystals from the ammonium sulfate condition were tested for X-ray diffraction. Unfortunately, no diffraction pattern could be detected. Thus, crystals from PEG3350 were tested as well. Most of them diffracted poorly, but some gave data sets at 2.8 Å - 3 Å resolution.

Crystallization of SeMet MurE

As molecular replacement with known structures of homologs did not succeed, a SeMet derivative was prepared and crystallized in PEG3350 in order to phase data from X-ray diffraction by MurE crystals. However, as the SeMet crystals needed more time than the native ones to grow, a sample was sent to the HTX lab to screen potential additional crystallization conditions for the SeMet derivative.

SeMet MurE was then found to crystallize in even more conditions than its native counterpart. Nice crystals were obtained in ammonium sulfate as for the native form. In addition, single and large crystals in PEG6000 and PEG8000 were obtained, and Sodium formate was shown to be a good precipitant for the SeMet protein as well. Flower-like crystals were obtained in PEG3350 drops, and hexagonal crystals grew in a Sodium formate condition (see Figure 13.3). Tens of SeMet crystals were screened. Finally, a crystal from the Sodium formate condition diffracted at 3.47 Å resolution, and a complete data set was collected.

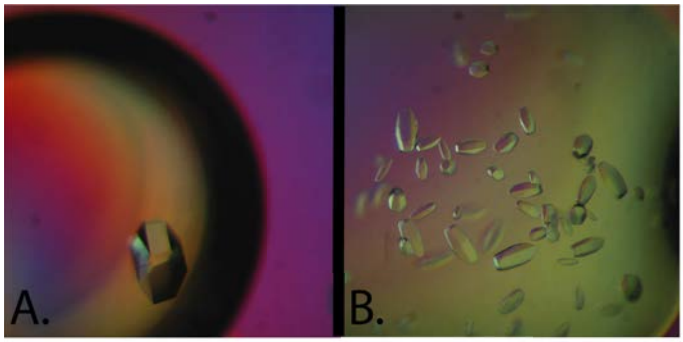


Figure 13.3: SeMet MurE crystals. Hand-made drops of 1 μl SeMet-MurE at 10 mg/ml with 5 mM AMPPNP plus 1 μl crystallization solution. A, 1 M Sodium formate, 0.1 M Sodium acetate pH 5.5. B, 0.5 M Sodium formate, 0.1 M Sodium acetate pH 5.0.

13.1.3 MurF

Crystallization

The high-throughput screen for MurF-his resulted in a single condition, containing 2-methyl-pentanediol (MPD), and forming tiny needles. Crystallization could be easily performed in hand-made drops but lead to sea urchin-like crystals or very small needles. Optimization through modification of crystallization temperature, pH, MPD concentration, additives, protein concentration, was not successful. In addition, crystal quality seemed to be batch-dependent, and finally the best way to obtain good crystals was to prepare several identical crystallization plates, in which some thicker and less polycrystalline needles could grow. Interestingly, a higher reproducibility was observed once the purification of MurF was automated with the MP3 platform at the IBS.

Diffraction and crystallization of SeMet MurF

Manually reproduced crystals diffracted X-rays up to 1.65 Å resolution. Despite significant effort, no molecular replacement solution could be obtained using the known structures of

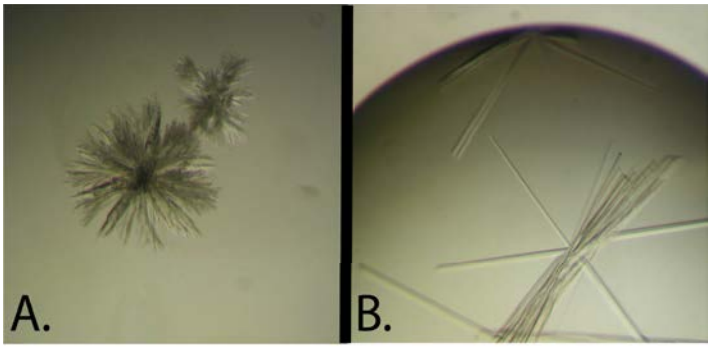


Figure 13.4: SeMet MurF crystals. Hand-made drops of 1 μl MurF at 10 mg/ml with 1 μl crystallization solution, in presence of 5 mM AMPPNP. Crystallization liquor: 0.1 M MES pH 5.5, MPD 48 % v/v.

MurF homologues from *E. coli* [4] or *Streptococcus pneumoniae* [5]. Therefore, a SeMet derivative was prepared and crystallized under similar conditions (see Figure 13.4 for crystal pictures). A data set was then collected at the Selenium edge at 2.4 Å resolution.

13.1.4 MreB

In order to acquire an extensive knowledge of the crystallization behaviour of MreB, the bacterial actin homolog was sent for high-throughput screening to the HTX lab, resulting in a high number of crystallization conditions. The table depicted Figure 13.5 summarizes the conditions where notable crystals were obtained. Conditions which gave rise to micro-crystals, very polycrystalline clusters, or spherulites are not reported here.

MreB crystals usually diffracted well (below 2 Å). In particular, a data set at 1.44 Å resolution was collected at the ESRF.

Crystallization of Mur ligases and MreB from *T. maritima*

Despite the high structural similarity previously reported among the Mur ligase family, the crystallization of three of the Mur ligases from *T. maritima* highly differs in terms of crystallization liquor, growth, and crystal quality for X-ray scattering experiments. Nice crystals can be easily obtained for MurE, while MurD and MurF often crystallize into tiny or polycrystalline crystals. Strikingly, MurE crystals diffract X-rays much less than MurD and MurF. As MurD crystals need several weeks to grow, MurF, which crystallizes in a few days, appears to be the best candidate among *T. maritima*'s ligases for co-crystallization assays with inhibitors. Lastly, the crystallization space of MreB from *T. maritima* is extraordinarily broad, giving rise to a wide variety of crystal shapes and making MreB a good candidate for protein crystallization studies.

Concentration	Buffer			Main precipitant		Salt		Second precipitant / salt		Crystal shape
(mg/ml)	Conc.	Name	pH	Name	Conc.	Name	Conc.	Name	Conc.	
10	0,05	sodium cacodylate	6,5	1,6-hexanediol	10%w/v	KCl	0,2	Magnesium chloride	0,005	Sticks
5 and 10	0,1	tri-sodium citrate dihydrate	6,5	Ethylene imine polymer	2%v/v	NaCl	0,1			Hexagones
10	0,1	HEPES	7,5	Isopropanol	10%	tri-Sodium citrate dihydrate	0,2			Large, single
5	0,1	tri-sodium citrate dihydrate	5,6	Isopropanol	10%			PEG 4000	10%	Single
10	0,1	tri-sodium citrate dihydrate	5,6	Isopropanol	10%			PEG 4000	10%	Triangles
5	0.05	sodium cacodylate	6,5	Isopropanol	10%	Ammonium acetate	0,1			Clusters / cluster needles
10	0,1	tri-sodium citrate dihydrate	5,6	Isopropanol	10%v/v	tri-sodium citrate dihydrate	0,1			Large, single
5	0.05	sodium cacodylate	6	Isopropanol	15%v/v	NaCl	0,1	Magnesium chloride	0,025	Cluster of needles
5				K/Na tartrate	0,4					Large, single
5	0,1	tris	8	Lithium chloride	1					Small crystals
5	0,1	bicine	9	Lithium chloride	1					Hexagones
10	0,1	tris	8,5	Lithium sulfate monohydrate	0,2					Large, single hexagones
5	0,1	HEPES	7,5	Magnesium formate	0,5					Hexagones
5	0,1	tris hydrochloride	8,5	Magnesium sulfate hydrate	0,4	Potassium sodium tartrate tetrahydrate	0,1			Large, single hexagones
10			5	Malonate	1,5					Large, single
5	0,1	MES	6	MPD	10%v/v					Clusters
5	0,1	tris	8	MPD	10%v/v					Clusters
5	0,1	Bicine	9	MPD	10%v/v					Clusters
10	0,1	ADA	6,5	MPD	12%					Cluster of needles
10	0.05	HEPES	7	MPD	15%v/v	KCl	0,1			Cluster of needles
5	0,1	MES	6	MPD	20%v/v					Cluster of needles
5	0,1	HEPES	7	MPD	20%v/v					Cluster of sticks
5	0,1	tris	8,5	NaCl	3					Hexagones
5	0,1	tris	8,5	NaCl	0,1					Clusters
10	0,05	MES	6	NaCl	0,6	Ammonium acetate	0,1	Magnesium sulfate	0,005	Needles
5	0,1	HEPES	7,5	PEG 10000	8%w/v			Ethylene glycol	8%v/v	Cluster of needles
5			6,9	PEG 3350	20%	Sodium thiocyanate	0,2			Cluster of needles
5			7	PEG 3350	20%	potassium thiocyanate	0,2			Cluster of sticks
10			7,9	PEG 3350		di-ammonium hydrogen phosphate	0,2			Large, single hexagones
10	0,1	Tris	8,5	PEG 400	15%	tri-Sodium citrate dihydrate	0,1			Large, single
5	0,05	Tris	7,5	PEG 4000	10%	KCl	0,2	Magnesium chloride	0,05	Cluster of needles
10	0,1	tri-sodium citrate dihydrate	5,6	PEG 4000	12%w/v	NaCl	0,1			Cluster of needles
5	0,1	ADA	6,5	PEG 4000	12%w/v	Lithium sulfate monohydrate	0,1	Isopropanol	2%v/v	Cluster of needles
10	0,1	HEPES	7	PEG 6000	10%w/v					Needles
10	0,1	Tris	8	PEG 6000	10%w/v	Lithium chloride	1			Small hexagones
5	0,1	Bicine	9	PEG 6000	10%w/v	Lithium chloride	1			Small hexagones
5	0,1	Citric Acid	5	PEG 6000	5%w/v					Cluster
5	0,1	HEPES	7	PEG 6000	5%w/v					Needles
5	0,1	Tris	8	PEG MME 5000	5%					Clusters
5	0,1	Bicine	9	PEG MME 5000	5%					Clusters
5	0,1	Tris	8	PEG MME 5000	10%					Clusters
5	0,1	Bicine	9	PEG MME 5000	10%					Clusters
5	0,1	Citric Acid	5	PEG MME 5000	20%					Small nice crystals
5	0,1	HEPES	7	PEG MME 5000	10%w/v	Tacsimate pH 7.0	5%w/v			Clusters
5	0,1	Tris	8,5	PEG MME 5000		Potassium sodium tartrate	0,8			Micro hexagones
5	0,1	Tris	8,5	PEG400	5%v/v	tri-sodium citrate dihydrate	0,1			Large, single
10	0,1	tri-sodium citrate dihydrate	5,6	PEG4000	12%w/v	NaCl	0,1			Cluster of needles
10	0,1	tri-sodium citrate dihydrate	5,6	PEG4000	12%w/v	Lithium sulfate monohydrate	0,1			Large, single
10	0,1	tri-sodium citrate dihydrate	5,6	PEG4000	15%w/v	Ammonium acetate	0,2			Cluster of needles
5			7,9	PEG8000	20%	di-ammonium hydrogen phosphate	0,2			Large, single hexagones
5				PEG8000	30%w/v	Ammonium sulfate	0,2			Cluster of needles
5	0,1	Tris	8,5	PEG8000	8%w/v					Cluster of needles
10	0,05	HEPES	7,5	Pentaerythritol propoxylate (5/4 PO/OH)	30%v/v	KCl	0,2			Small needles
5	0,1	Tris	8,5	Polyvinylpyrrolidone K15	20%	Cobalt chloride	0,1			Large sticks
10	0,1	HEPES	7,5	potassium sodium tartrate tetrahydrate	0,1	Lithium sulfate monohydrate	0,1			Large sticks
10	0,1	HEPES	7,5	Potassium sodium tartrate tetrahydrate	0,4					Single crystal
10	0,1	Imidazole	6,5	Sodium acetate	1					Cluster of needles
5	0,1	Tris	8,5	Sodium acetate trihydrate	0,1					Clusters
10	0,1	Bicine	9	Sodium formate	1,6					Small hexagones
10	0,8	MES	6	Sodium formate	0,8					Needles
5	0,1	Bicine	9	Sodium formate	1,6					Clusters
5 and 10	0,1	Sodium acetate	4,6	Sodium formate	2					Cluster of needles
5	0,1	Tris	8	Sodium formate	2,4					Hexagone
10	0,1	Bicine	9	Sodium formate	2,4					Hexagones
10	0,1	Bicine	9	Sodium formate	3,2					Hexagones
5	0,1	Bicine	9	Sodium formate						Clusters
5	0,1	Citric Acid	4	Sodium formate						Clusters
10	0,1	Citric Acid	5	Sodium formate pH5						Crystals (cubic)
5	0,8	Tris	8	Sodium formate pH8	0,8					Clusters
5	1	Malonate	5							Small crystals
10	1,5	Malonate	5							Crystals (cubic)

Figure 13.5: Crystallization conditions for MreB from *T. maritima*. Results obtained from a high-throughput screen from the HTX lab.

13.2 Structure solution

Six crystal structures for MurD, MurE, MurF, and MreB were solved with the invaluable help of Carlos Contreras-Martel. Three of them were published in *Environmental Microbiology* in July 2013 in the article 'MreB and MurG as scaffolds for the cytoplasmic steps of peptidoglycan biosynthesis' [6].

13.2.1 Data processing

Structures of the ligases MurD, MurE, MurF, the two latter in complex with ADP, were solved by single-wavelength anomalous diffraction (SAD) experiments on SeMet derivatives. Resulting models for MurE and MurF were used as templates for molecular replacement on higher resolution data sets.

Additionally, a structure of apo MurF at 1.9 Å resolution in another space group was solved by molecular replacement using the previously solved structure. This structure is under refinement. A second structure of MurE at 2.7 Å resolution in complex with ADP was obtained during co-crystallization trials with MreB (see next chapter), in a $P2_12_12_1$ space group by SAD at the zinc edge. This structure is under refinement as well.

Co-crystallization assays (see next chapter) lead to the collection of three data sets in $P3_121$ and $C2$ space groups for MreB. The highest resolution data set has been processed, resulting in a structure of MreB at 1.44 Å resolution, thus providing more details to the structure at 2.1 Å published by van den Ent *et al.* in 2001.

Table 13.2.1 summarizes the main statistics about data collection, phasing, and refinement of these six structures. More detailed information is available in the Appendix C.

Crystal	**MurD**	**MurE** $P6_1$	**MurE** $P2_12_12_1$	**MurF** $P2_12_12_1$	**MurF** P_1	**MreB**
Ligands	-	ADP, Mg^{2+}	ADP, Mg^{2+}	ADP	-	ADP
Beamline	Proxima	ID29 (SAD), ID14eh4	BM30a (SAD), ID14eh4	BM30a (SAD), ID29	ID14eh4	ID14eh1
Resolution (Å)	2.17 Å	2.9 Å	2.7 Å	1.65 Å	2.07 Å	1.44 Å
Space group	$P2_1$	$P6_1$	$P2_12_12_1$	$P2_12_12_1$	$P1$	$P3_121$
Molecules per asymmetric unit	2	2	2	1	4	1
Cell parameters	56.2 135.9 67.3 90 98.2 90	74.4 74.4 441.8	53 108 140	48.6 49.8 178.0	67.4 84.1 104.9 90.7 93.5 113.6	51.4 51.4 291.7
Rsym (last shell)	7.8 % (44.1%)	5.6 % (100.3 %)		4.1 % (47.8 %)	7 % (42.2 %)	
Mean B factor data / model (Å^2)	38 / 45	114 / 140	refinement	25 / 29	in progress	in progress
R_{work}/R_{free}	17 % / 21 %	24 % / 29 %	refinement	16 % / 22 %	17 % / 23 %	18 % / 21 %
Allowed residues in Ramachandran plot	99.7 %	100 %	refinement	100 %	in progress	in progress

Table 13.2: Data collections and structure solution statistics. Statistics are indicated for the highest resolution data sets.

13.2.2 Crystal structures of Mur ligases

Mur ligases from different bacteria display considerable structural similarity, and basically consist of a small N-terminal domain that recognizes the growing peptidoglycan building

block, a central domain that binds the nucleotide, and a C-terminal domain that binds the incoming amino acid [2].

MurF

MurF from *T. maritima* (see Figure 13.6, A) presents a typical Mur ligase fold which is composed of three mixed α/β-sheet domains that, individually, are highly reminiscent of those from its *E. coli* and *S. pneumoniae* counterparts [4] [5] despite the relatively low sequence identity (27%, 29%, respectively).

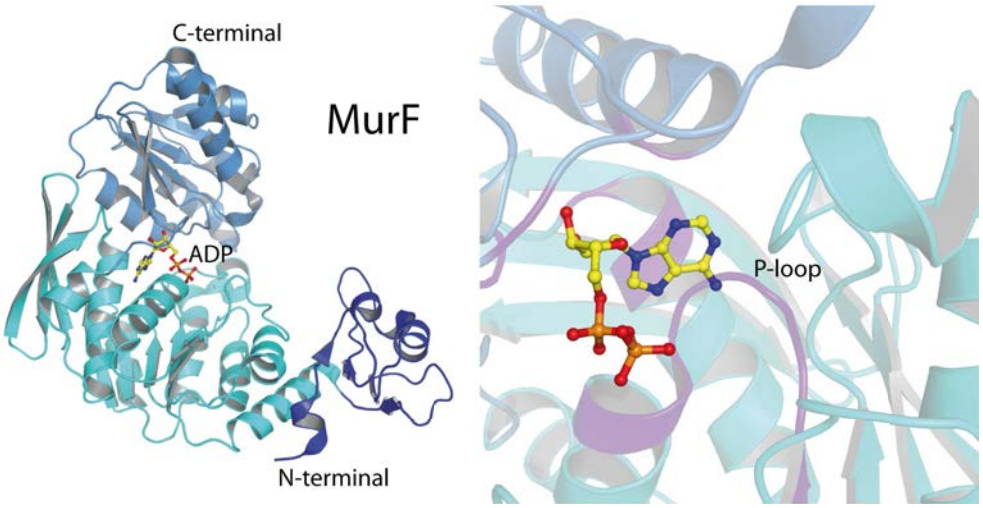


Figure 13.6: Structure and ADP binding site of MurF from *T. maritima*. MurF in interaction with ADP (left) and zoom onto the ADP-binding site (right). The three different shades of blue correspond to the three open α/β-sheet domains of MurF. ADP-interacting residues are highlighted in purple, including the P-loop. ADP is depicted as sticks, with C atoms in yellow, N atoms in blue, O atoms in red, and P atoms in orange. PDB entry: 3ZL8.

The N-terminal domain of MurF from *T. maritima* (res 1–63) consists of a small α/β-fold that lacks the 20 first residues that are normally present in other Mur ligases. This is a unique feature of MurF – it lacks the characteristic Rossmann nucleotide binding fold seen in the N-terminal domains of other Mur ligases [7]. The central domain (res 64–286) adopts a mononucleotide-binding fold seen in other ATP-binding enzymes, and consists of a central, 7-stranded β-sheet interwoven by 6 α-helices, backed up by a smaller, 3-stranded antiparallel β-sheet. Lastly, the C-terminal domain (res 287–427) adopts a dinucleotide (Rossmann)-binding fold, and is connected to the central domain by a short linker.

The crystal structure obtained in the $P2_12_12_1$ space group shows MurF in complex with ADP. The nucleotide binds within a cleft made by both the second and third domains employing a set of highly conserved residues, including the well-characterized P-loop (Figure 13.6, right). Strikingly, in addition to the common set of amino acids that provide side chains for interactions with ADP in all structures solved to date, in the structure of *T. maritima* MurF, the ribose forms hydrogen bonds with two additional residues, Asn 258 and Lys 405. In addition, Asn 310, which interacts with the phosphate group in MurF, has no counterpart in *E. coli* MurD. The higher number of residues involved in ADP binding may be linked to the thermophilic origin of the protein.

Strikingly, the structure of apo MurF obtained from the $P1$ space group data set presents exactly the same conformation as the ADP-MurF complex in $P2_12_12_1$.

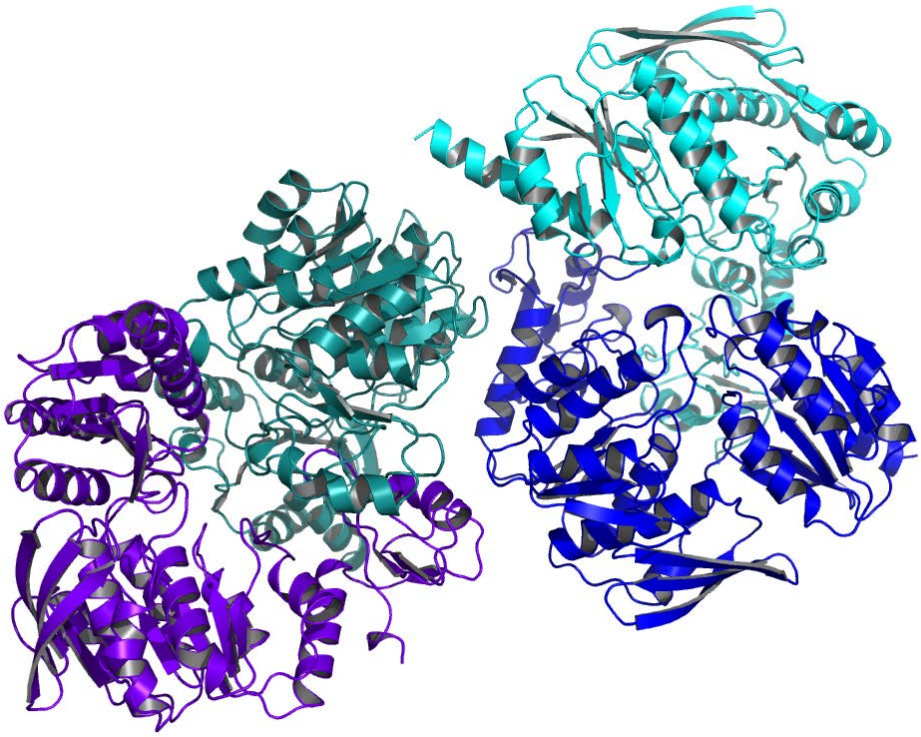


Figure 13.7: Structure of apo MurF in a $P1$ space group. The four different colours indicate the four chains within the asymmetric unit.

MurE

The structure of *T. maritima* MurE also reveals a three-domain molecule, in which the C-terminus is closed 'over' ADP, stabilizing it within the central domain (see Figure 13.8).

Notably, the structure of MurE reported here is highly reminiscent of that of the enzymes from *E. coli* and *Mycobacterium tuberculosis*, both crystallized in the presence of the reaction product UDPMurNAc-L-Ala-D-Glu-*meso*-diaminopimelate [8] [9] (Figure 13.9). Superposition of the three structures reveals root-meansquare deviation (rms) values between 1.6 and 2.2 Å over 347 Cα atoms.

To our knowledge, this is the first structure solved for an UDP-*N*-acetylmuramoyl-L-alanine-D-glutamate:L/D-lysine ligase, thus providing a structural basis for the understanding of substrate specificity of MurE ligases.

The second structure of MurE that was solved, in $P2_12_12_1$ space group, shows a slight shift of the C-terminal domain when comparing with the structure in $P6_1$. This subtle conformational change can be observed when superimposing the electron density maps.

MurD

Figure 13.10 shows the structure of *T. maritima* MurD at 2.17 Å. Although MurD was crystallized in the presence of AMPPNP and UMA, no ligands could be identified in the

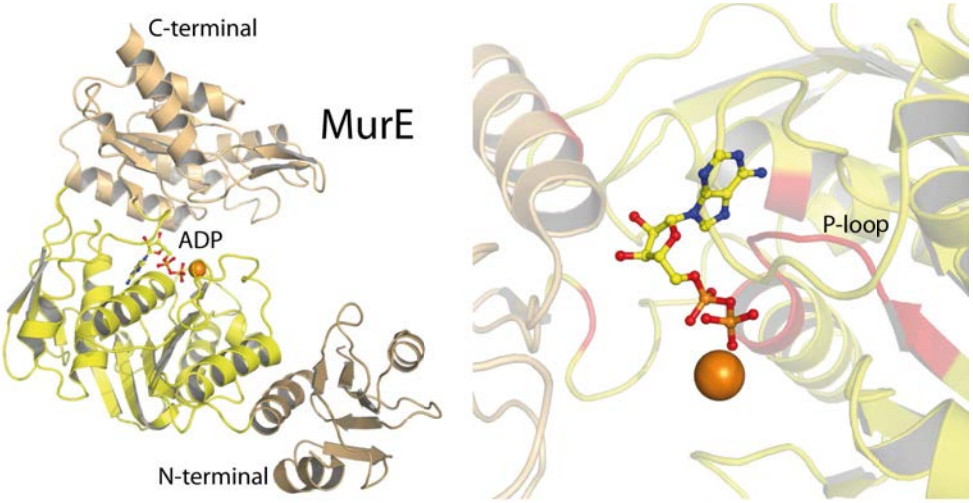


Figure 13.8: Structure and ADP binding site of MurE from *T. maritima*. MurE in interaction with ADP (left) and zoom onto the ADP-binding site (right). The three different shades of yellow correspond to the three open α/β-sheet domains of MurE. ADP-interacting residues are highlighted in red, including the P-loop. ADP is depicted as sticks, with C atoms in yellow, N atoms in blue, O atoms in red, and P atoms in orange. Orange sphere: Mg^{2+}. PDB entry: 4BUB.

Figure 13.9: Superposition of the structures of MurE from *T. maritima* (orange, in complex with ADP) and MurE from *E. coli* (1E8C; cyan) in complex with the product of the reaction (sticks). Note that the conformation of the C-terminal domain in both molecules is comparable, despite the fact that only the *E. coli* enzyme was crystallized in the presence of the reaction product.

structure, potentially due to the high concentration of phosphate in the crystallization solution. Thus the structure presented here is in apo form.

The enzyme displays the characteristic three-domain organization of its *E. coli* and *Staphylococcus aureus* counterparts [10] [11] [12]. Mapping of the crystallographic temperature (B) factors onto the structure model, as shown in Figure 13.10, where dark blue indicates stable regions (low B factors) and lighter colours are indicative of flexibility (high B factors), shows that the C-terminal domain of this Mur enzyme is also its most conformationally flexible feature. This domain makes no contacts with its central region, and loop 308–315 that links the two domains is highly flexible and difficult to trace in the electron density map. Notably, this feature generates two different conformations of the domain in the same asymmetric unit, which are indicated in Figure 13.10, where the two molecules are superposed, and the distinct conformation of the C-terminal domain is clearly visible.

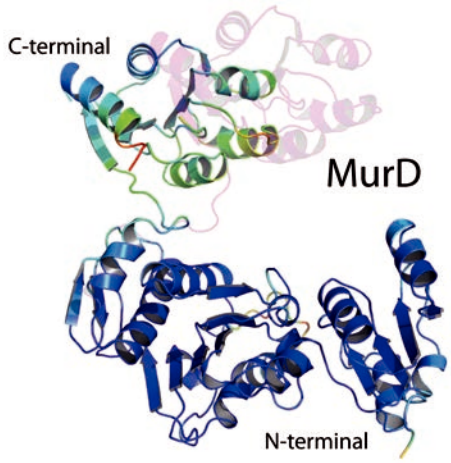


Figure 13.10: Structure of MurD from *T. maritima*. Superpositions of cartoon representations of chain A (colours correspond to B-factors; low B-factors: blue, high B-factors: red) and chain B (transparent, magenta). PDB entry: 4BUC.

Notably, the superposition of the MurD structures from *E. coli* and *T. maritima* reveals that the N-terminal and central domains are highly similar, while the C-terminal domain is in a completely different position, being located clearly much farther from the active site than the C-terminus in the ligand-bound structure from *E. coli* (Figure 13.11).

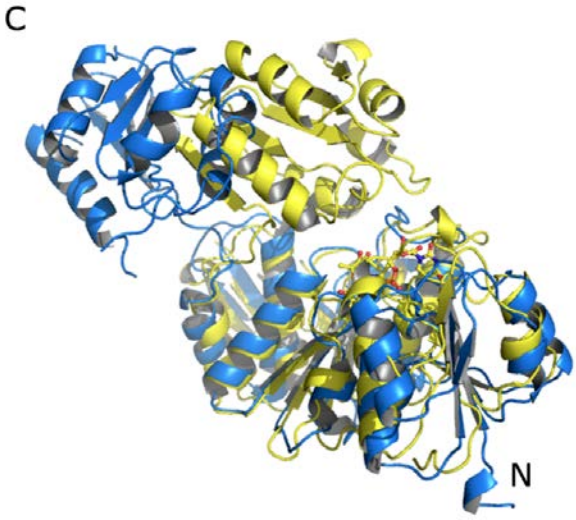


Figure 13.11: Superposition of the structures of MurD from *T. maritima* (blue) and from *E. coli* (yellow), the latter having been crystallized in the presence of UMA (1UAG). Note that the N-terminal and central domains superpose well, whilst the C-terminal domain displays considerable conformational flexibility.

Ligand binding and confomational flexibility

Mur enzymes have been reported to crystallize in 'open' and 'closed' conformations, depending on the presence and identity of the bound ligand. These conformational changes involve rotations of the N- and C-terminal domains either 'towards' the central domain (closed form) or 'away' from it (open form) [2] [13] [12].

A comparison of the ADP-liganded MurF structure to the open apo *E. coli* MurF structure, and an inhibitor-bound, closed *S. pneumoniae* MurF structure [5], reveals that the conformation of MurF from *T. maritima* in complex with ADP is in fact intermediary between that of the two other structures (Figure 13.12). The C-terminal domain of MurF acts as a 'claw' that locks onto the ADP molecule, helping to stabilize it within a deep cleft. It is of note that the inhibitor-bound MurF structure from *S. pneumoniae* represents the most compact Mur structure published to date. However, the second structure of MurF that was solved during the work of this thesis shows an identical conformation for an apo form of MurF from *T. maritima*, suggesting that nucleotide binding does not result in a major conformational change. However, the phosphates which are present in this structure could mimic the nucleotide and therefore infer an 'ADP-conformation' of the ligase.

When comparing the conformations of MurE enzymes from *E. coli*, *M. tuberculosis*, and *T. maritima*, the C-terminal domains of all three molecules are positioned similarly, despite the fact that one is in complex with the nucleotide, and the others with the reaction product. Thus, in the case of MurE, ligand binding does not necessarily dictate the positioning of the C-terminus, suggesting that the preferential conformations of this enzyme may not be ligand-dependent. Moreover, it is of note that the previously solved structure of apo MurC displayed a relatively closed conformation (13.14).

Added to the high flexibility of the C-terminal domain of MurD from *T. maritima*, these observations support the idea that there is increased flexibility of the apo forms of the enzymes in contrast to the substrate-bound forms, but no clear ligand-related conformation.

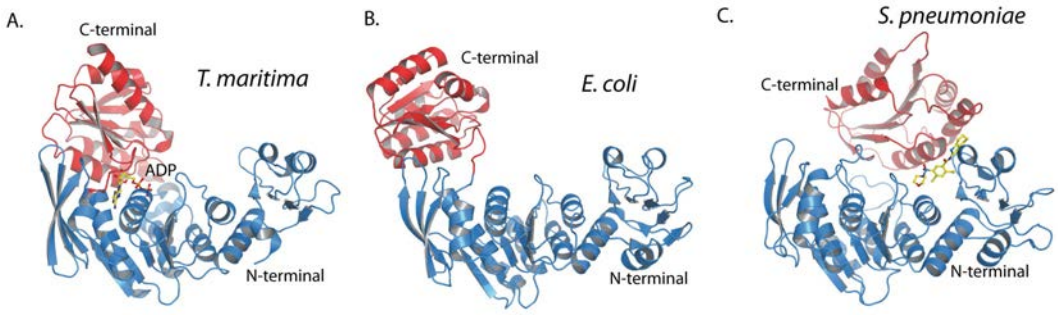


Figure 13.12: Flexibility of the C-terminal domain in the known structures of MurF. A, *T. maritima* MurF in complex with ADP (apo form of the same enzyme presents an identical structure). B, apo form of *E. coli* MurF (1GG4). C, *S. pneumoniae* MurF in complex with an inhibitor (2AM1). C-terminal domains are represented in red. ADP and inhibitor are depicted as sticks.

13.2.3 Crystal structure of MreB at 1.44 Å resolution

The structure of MreB in $P3_121$ was obtained at 1.44 Å resolution and is under refinement. It provides more details to the previously known structure at 2.1 Å.

Figure 13.13: Crystal structure of MreB at 1.44 Å resolution.

Crystal structures of Mur ligases from *T. maritima*

These crystal structures contribute to complete the set of Mur ligases structures of *T. maritima*, as MurC has already been published. They present high structural similarity to their homologs, though sequence identity is relatively low (below 30 %). MurD structure underlines the very high flexibility of the C-terminal domain towards N-terminal and central domains and displays an open conformation. By contrast, MurE and MurF structures, both in complex with ADP, present a relatively closed conformation. Interestingly, the second structure of MurF, in apo form, presents the same conformation as the ADP-bound form; And the second structure of MurE, also in ADP-bound form, shows a slight conformational change by comparing to the first MurE structure. These results suggest that there is no ligand-related conformation for Mur ligases, but instead a reduced flexibility of the nucleotide-bound forms. Lastly, the higher resolution obtained for the structure of MreB could be useful for the design of compounds targeting the bacterial actin homolog.

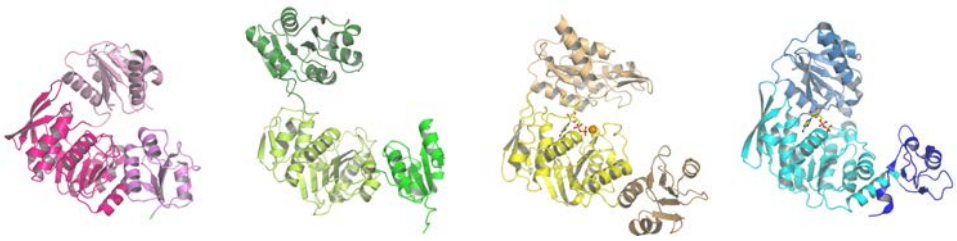


Figure 13.14: Crystal structures of Mur ligases from *T. maritima*. PDB codes: MurC (1JCU), MurD (4BUC), MurE (4BUB), MurF (3ZL8).

Chapter 14

Studies of interactions

The purification, characterization, and crystallization described in the previous chapters provided a deep knowledge of the *in vitro* behaviour of Mur ligases, MurG, and MreB from *T. maritima*, thus giving a sound basis for the study of the interaction network of the cytoplasmic actors of peptidoglycan biosynthesis.

14.1 Biochemical studies of Mur interactions

14.1.1 Set-up of co-expression protocols

Expression from pETDuet vectors

The two pETDuet constructions lead to high amounts of MurD - cloned in the first site, and much less MurE - cloned in the second site. Regarding the pETDuet-*murD-murEmurF* construct, expression of MurF could not be detected by Coomassie gel. As controlling the ratio of potential partners might be required to detect the formation of a complex, this method of expression was not extensively used.

Co-transformations

BL21(DE3) cells were transformed with two or three plasmids in order to try co-expression of several recombinant proteins.

However, for most cases it was very difficult to control the ratio of expression. While induction of pETDuet (first cloning site) co-transformed with pET30b resulted in similar amounts of proteins (both IPTG-induced vectors), the co-expression of an IPTG-induced together with an AHT-induced protein was very difficult to control. Indeed, IPTG-induced proteins were much more expressed than AHT-induced proteins. By adding AHT first and IPTG two hours later, and lowering the concentration of IPTG, higher amounts of AHT-induced proteins could be obtained. However, it was difficult to control the expression levels and with this method the expression was not simultaneous but consecutive.

Because of these difficulties, a co-lysis protocol was preferred for assays involving both AHT and IPTG-induced proteins, as such a strategy allowed a better control of protein quantities.

14.1.2 Pull-down assays

Pull-down assays between Mur ligases

No interaction between Mur ligases could be detected by different pull-down assays (see Table 14.1) involving MurC, MurD, and MurE, even when ligases were co-expressed in a same bacterial strain.

Proteins	**Buffer**	**Result**
MurCstrep and hisMurD	pH 7.4, 0.2 M NaCl, 10% glycerol	No interaction
MurCstrep, MurD	pH 7.4, 0.15 M NaCl	No interaction, MurC aggregated
MurD, MurEhis	pH 7.4, 0.2 M NaCl, 10% glycerol	No interaction

Table 14.1: Co-lysis and co-purification trials for Mur complexes. BL21(DE3) cells expressing single or several (*italicized*) proteins were (co)-lysed by sonication and loaded onto affinity columns (5 mL Strep-trap column, 1 mL His-trap column) or mixed to 100 μl His resin in batch mode. No imidazole was added in 'buffer A'. Elution was performed on an Äkta device or by gravity with 2.5 mM d-desthiobiotin or steps of imidazole concentration.

Though in some experiments, a very tiny band that could correspond to the interacting protein was seen, it is hard to determine whether this band was the result of a specific interaction or was just a contaminant. Anyway, the amount of this potential complex was far too small to consider purifying it for crystallization trials. Figure 14.1 shows a typical result of a pull-down assay, through the example of MurD - MurEhis.

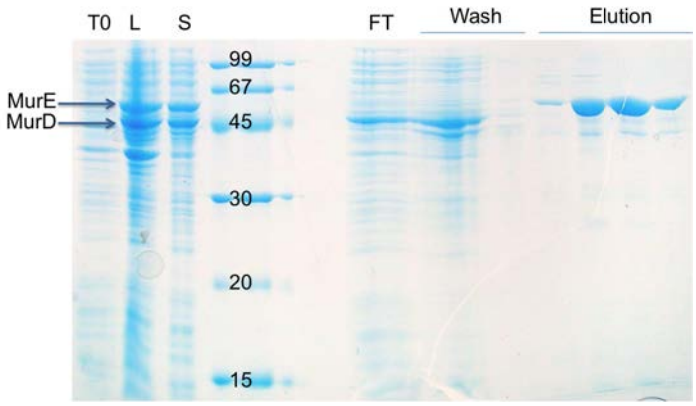


Figure 14.1: SDS-PAGE analysis of a pull-down trial between MurD and MurE-his showing no interaction between the two ligases. BL21(DE3) co-expressing MurD (pETDuet) and MurE-his (pET30b) were lysed in 25 mM NaPi, pH 7.4, 0.2 M NaCl, 10% glycerol. After centrifugation, lysates were loaded onto a His-trap column. T0, before induction; L, lysate; S, supernatant. MurD, 49 kDa; MurE, 56 kDa.

MreB-Mur ligase pull-downs

As MreB has been reported to play an organizational role in peptidoglycan biosynthesis [14], interactions between Mur ligases and the bacterial actin homolog were assessed as well. Four combinations of pairwise pulldown assays were tried (see Table 14.2) but none of them could unveil a clear interaction. In particular, the pH screen showed that MreB-his could not be detected in the elution fractions of the His-trap affinity purification step when in presence of MurD, MurE-his, and MurF at pH 6.5, 7.0, and 7.5, while MreB-his was clearly detected at pH 8.0 and 8.5. By contrast, bands at the sizes of MurD or MurF were detected in the

elution fractions of the samples for a pH below 7, but not for a pH above 7. Additional experiments are required to determine whether or not this pH-dependent behaviour is the result of interactions between Mur ligases and MreB.

Proteins	**Buffer**	**Result**
MreBhis, MurD	pH 7.4, 0.2 M NaCl, 5% glycerol	No interaction
MreBstrep, MurEhis	pH 7.4, 0.075 M NaCl, 2 mM DTT	No interaction
MreBstrep, MurFhis	pH 7.4, 0.05 M NaCl, 2 mM DTT	No interaction
MurD, hisMurE, MurFhis, MreBhis	pH 6 to 8.5, 0.1 M NaCl, 25 mM imidazole	No interaction detected, unclear results for some pH

Table 14.2: Co-lysis and co-purification trials for Mur complexes. BL21(DE3) cells expressing single or several (*italicized*) proteins were (co)-lysed by sonication and loaded onto affinity columns (5 mL Strep-trap column, 1 mL His-trap column) or mixed to 100 μl His resin in batch mode. No imidazole was added in 'buffer A', unless mentioned. Elution was performed on an Äkta device or by gravity with 2.5 mM d-desthiobiotin or steps of imidazole concentration.

MurG-MreB pull-downs

Thereafter, interaction between MurG-strep and MreB-his was assessed (see Table 14.3) since it has been suggested in several publications in other organisms [14] [15] [16]. For these experiments, the first step was a Strep-trap affinity column, as there is less non-specific binding on a Strep-trap resin compared to a His-trap resin, and the second step was a His-trap affinity column.

Proteins	**Buffer**	**Result**
MreBhis, MurGstrep	pH 7.4 without DDM, 0.15-0.3 M NaCl 5-10% glycerol	unclear results
MreBhis, MurGstrep	pH 6.5 to 8.5, with or without DDM 0.04 %, 0.15 M NaCl	Double band when no DDM (MurG degradation?)
MreBhis, MurGstrep	pH 7.5 without DDM, 0.15 M NaCl	MurG aggregated and degraded, no MreB

Table 14.3: Co-lysis and co-purification trials for Mur complexes. BL21(DE3) cells expressing single proteins were co-lysed by sonication and loaded onto a 5 mL Strep-trap column or mixed to 100 μl Strep resin in batch mode (first line). Elution peak (second line) was then loaded onto a 1 mL His-trap column. No imidazole was added in 'buffer A'. Elution was performed on an Äkta device or by gravity with 2.5 mM d-desthiobiotin or steps of imidazole concentration.

A first trial gave an interesting double band after double affinity chromatography purification in the presence of DDM, according to Western Blotting analysis: A band stained with Streptactin HRP probe suggested the presence of MurG, and a band stained with anti-his antibody suggested that MreB had been co-eluted. However, this result could not be easily reproduced. Moreover, when the sample was loaded onto a size exclusion chromatography, no macromolecule of a higher molecular weight than the usual MurG oligomer could be clearly detected.

Then, a screen for optimization of buffers was performed, with a range of pH and the presence or absence of detergent. This analysis suggested that DDM could interfere with the formation or stability of a Mur complex. The experiment could be reproduced at a higher scale without detergent, giving the results reported in Figure 14.2.

However, in absence of DDM, MurG appeared to partially aggregate as shown by gel filtration assays (see Chapter 12).

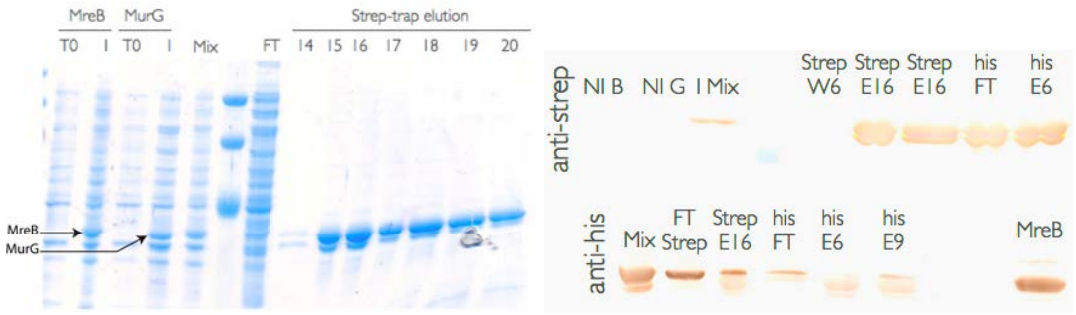


Figure 14.2: SDS-PAGE and Western Blotting analyses of MreB-MurG pull-down. C41(DE3) and BL21(DE3) cells expressing MreB-his and MurG-strep, respectively, were mixed and lysed by sonication in 50 mM Hepes pH 7.4, 10 % glycerol, 300 mM NaCl. After centrifugation, supernatants were loaded onto a Strep-trap column and bound proteins were eluted with d-desthiobiotin. Elution fractions were loaded onto a His-trap column and eluted with a gradient of imidazole. Left, SDS-PAGE analysis of the Strep-trap elution, revealing a double band which could correspond to MurG-MreB complex. T0, before induction; I, induced; Mix, mix of cell strains before lysis; FT, flow-through. Right, Western Blotting analysis. NI, non-induced; B, MreB-induced; G, MurG-induced; Mix: mix of MurG- and MreB-expressing cells before lysis; Strep refers to the Strep affinity column step and his to the His affinity column step. W refers to a washing fraction; E refers to an elution fraction; FT refers to the flow through. This WB analysis indicates that both MurG (39.9 kDa) and MreB (37.5 kDa) were present in the elution fraction E16 of the Strep-trap elution, and the elution fraction E6 of the His-trap column (tiny band for MreB).

Moreover, other attempts to reproduce and refine experimental conditions gave very varying results: Several times no protein bound the His-trap column though a high elution peak had been obtained after the first Strep-trap affinity column. In addition, a dot blot experiment performed on Strep-trap elution fractions could not detect MreB, nor mass spectrometry analysis. Lastly, parallel purifications of single MurG and MreB showed a degradation of both proteins in the buffer used for pull-down assays.

Because of all these difficulties, no further work was done regarding MurG-MreB pull-downs. Though the complex might have been trapped in some experiments, the difficulties to reproduce the results, the instability of MurG and its tendency to oligomerize, render the experiment very hard to pursue, limiting the chances of a successful isolation of the MurG-MreB complex for crystallization purpose. Optimizing the stability and homogeneity of MurG, adding EDTA to avoid MreB degradation, and using other techniques to detect the complex could be useful in the purification of MreB-MurG complex.

MurG-Mur ligase pull-downs

MurG-Mur ligase interactions were assessed by pull-down assays as well, and were the most promising co-purification experiments. Indeed, several potential complex-related bands were obtained for different pull-downs.

A pull-down was tried between MurG-strep, MurD, and MurE-his. In the light of previous experiments, Tris buffer was chosen as MurG seemed more stable in this buffer, the concentration of DDM was lowered to 0.02 % DDM because in MurG-MreB pull-downs, DDM seemed to interfere with the formation of a complex, and a low pH was chosen to get closer to the most promising co-crystallization trials (see below). The elution fractions contained many contaminants, most likely because of the low pH used in this assay which reduces the performance of the Strep-trap resin. However, when loaded onto a Superose6,

the elution peak was very similar to the elution profile of MurG in the presence of a low concentration of DDM so no complex could be detected by gel filtration chromatography assay. Strep-trap elution fractions were then loaded onto a His-trap column and a SDS-PAGE analysis of this affinity step is depicted Figure 14.3. A Western Blot analysis showed that one of the band was recognized by the anti-his antibody, another band was recognized by the anti-MurD antibody, and the two lower bands most likely corresponded to MurG and its degradation product, suggesting that a first step towards the purification of a Mur complex had been accomplished (not shown).

Proteins	**Buffer**	**Result**
MurD, MurE-his, MurFhis, MurGstrep	pH 7.8, 0.2 M NaCl, 5% glycerol	No interaction
MurGstrep, MurD, MurEhis	Tris pH 7.1, 0.05 M NaCl, 0.02 - 0.04 % DDM, 1 mM EDTA, 1 mM DTT	Several bands eluted from the Strep-trap colum, but MurG non-specifically binds the His trap column

Table 14.4: Co-lysis and co-purification trials for MurG-Mur ligase complexes. BL21(DE3) cells expressing MurG-strep were lysed, and the soluble part was loaded onto a Strep-trap column. After washing, lysates of MurD and MurE-his were loaded onto the Strep-trap column where MurG was attached. Elution fractions were then loaded onto a His-trap column.

However, a negative control where pure MurG was loaded onto a His-trap column revealed that the glycosyltransferase tended to interact non-specifically with this column. Therefore, the His-trap column (Ni Sepharose High Performance, from GE) cannot be used for investigating the formation of a MurG-Mur ligase complex. In addition, an experiment in batch mode was performed to optimize the co-purification buffer; however, no complex could be detected in this trial. Other His resins have to be assessed to find one to which MurG does not bind.

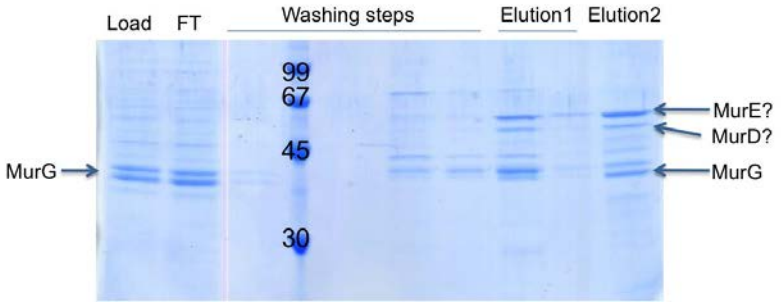


Figure 14.3: SDS-PAGE analysis of a MurG-strep, MurD, MurE-his pull-down experiment showing a double band that may correspond to MurG-strep and its degradation product (39.9 kDa), a band that may correspond to MurE-his (56 kDa), and a band that may correspond to MurD (50 kDa). SDS-PAGE analysis shows representative fractions of the second, His-trap, affinity step. Load: sample loaded onto the column; FT, flow-through; elution1 = 100 mM imidazole; elution 2 = 500 mM imidazole.

14.1.3 Gel filtration assays

Mur ligases and MurG

Size exclusion chromatographies (see Table 14.5), after mixing pairs of purified ligases were assessed, but no interaction could be detected.

Protein	Buffer	Column	Result
hisMurD, MurFhis	pH 7.4, 0.15 M NaCl, DTT	SuperDex200	No interaction
MurD, MurEhis	pH 7.4, 0.2 M NaCl, 10% glycerol	SuperDex75	No interaction

Table 14.5: Size exclusion chromatography trials for mixes of Mur ligases. Purified proteins were mixed at least 30 min before the interaction tests, in different ratios. See Materials and methods for more details.

Size exclusion chromatographies involving MreB and/or MurG

The MurG-MreB pair was assessed as well, in the conditions reported in Table 14.6. As illustrated by the chromatograms Figure 14.4, no interaction was detected.

Proteins	Buffer	Column	Result
MurGstrep, MreBhis	pH 7.4, 0.15 M NaCl, EDTA, DDM, glycerol	SuperDex200	No interaction
MurGstrep, MreBhis	pH 7.4, 0.15 M NaCl, EDTA, DDM, glycerol	SuperDex75	No interaction
MreBhis, MurEhis	pH 6.0, 0.15 M NaCl, AMPPNP	SuperDex200	No interaction
MurGstrep, MreBhis	pH 7.4, 0.3 M NaCl, DDM	Superose6	No interaction

Table 14.6: Gel filtration trials for the MreB/MurG pair.

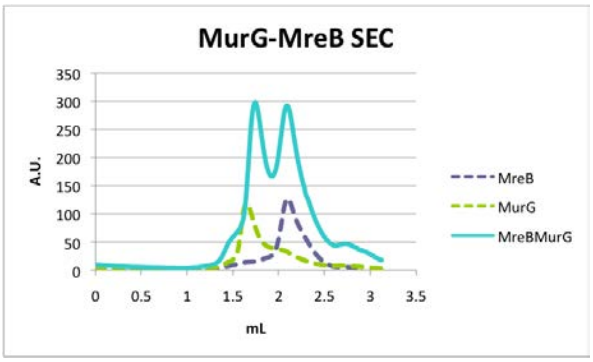


Figure 14.4: Size exclusion chromatrophy for assessing the MurG-MreB interaction. The elution profile of the mix of proteins corresponds to the addition of the elution profiles of the single proteins. Note: for a clearer view, curves are not drawn to scale.

A size exclusion chromatography was performed on a mix of MurF and MurG, but the formation of a complex could not be detected either.

14.1.4 Cross-linking and native gels

Cross-linking assays and native gels to analyze mixes of ligases could not reveal any interaction (see Table 14.7). A typical SDS-PAGE of a cross-linking assay is represented Figure 14.5.

Protein	Crosslinking agent	Result
MurCstrep, hisMurD, MurEhis, MurFhis by pairs	Glutaraldehyde, ratio 1:1	No interaction
MurCstrep, hisMurD, MurEhis, MurFhis by pairs	EGS, ratio 1:1	No interaction
hisMurD, MurEhis, MurFhis by pairs or all together	native gel, ratio 1:1 to 1:4	No interaction

Table 14.7: Crosslinking and native gels trials for Mur complexes.

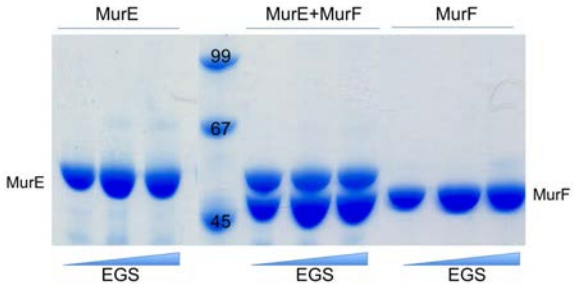


Figure 14.5: SDS-PAGE analysis of cross-linking assay between MurE and MurF with EGS. Cross-linked products are shown for increasing concentrations of EGS. Controls with single proteins are shown as well. Note that for MurE, a band around 70 kDa starts appearing for the highest EGS concentration.

Cross-linking by EGS for MreB-MurG (ratio 1:2), MreB-MurF (ratio 1:3), and MurG-MurF (ratio 1:3) pairs was assessed without success. Regarding native gel trials with MreB, the actin homolog did not enter the gels.

14.1.5 Dot-blot assays suggested that the Mur interaction network could be based on MurG and MreB

As pull-down assays between Mur ligases were all negative, Mur interactions were further assessed with MurG and MreB by a simpler technique, namely dot blot. In this assay, protein A is dotted on a nitrocellulose membrane. After drying and washing, the membrane is incubated in a solution of protein B. The presence of protein B on a protein A dot is developed as a regular Western Blot.

When MurG was immobilized on a membrane, an interaction could be detected with MurE, MurF, and MreB (Figure 14.6) and the signal was proportional to the amount of protein loaded onto the membrane.

Likewise, when the ligases and MreB were immobilized and the membrane incubated with MurG, interactions could be detected between MurG and MreB, MurE, MurF in a dose-dependent way (Figure 14.7). By contrast, though a signal could be detected for the MurG/MurD assay, it was not dose-dependent.

No significant signal was detected for negative controls[12] (Figure 14.7), suggesting that these interactions between pure proteins were specific. In addition, MurG-strep was not detected by the anti-his antibody solution and neither his-tagged Mur ligases nor MreB-his was detected by the streptactin-HRP probe (not shown).

A similar assay showed the interaction between MreB-strep and MurF-his, as illustrated

[1]LMO corresponds to the protein InlK, a member of the internalin family specific to *Listeria monocytogenes*. The protein was provided by David Neves.

[2]PulS is the Pullulanase secretion protein from *Klebsiella oxytoca*, and was kindly provided by Tommaso Tosi.

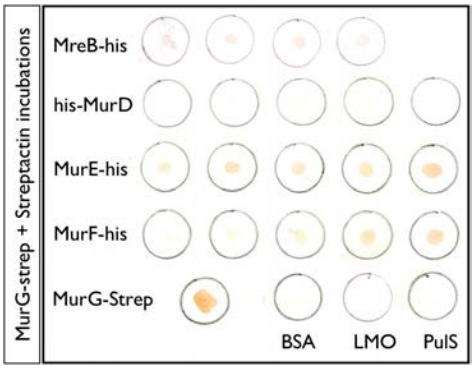


Figure 14.6: Dot blot on purified proteins showing direct interaction between MurG and immobilized MreB, MurE, and MurF. The nitrocellulose membrane was dotted with increasing amounts of his-tagged MreB, MurD, MurE, MurF, Strep-tagged MurG as positive control, and three negative controls (BSA; LMO; PulS, as indicated). Membranes were incubated with 0.1 mg.mL^{-1} of MurG, and subsequently incubated with streptactin-HRP for development.

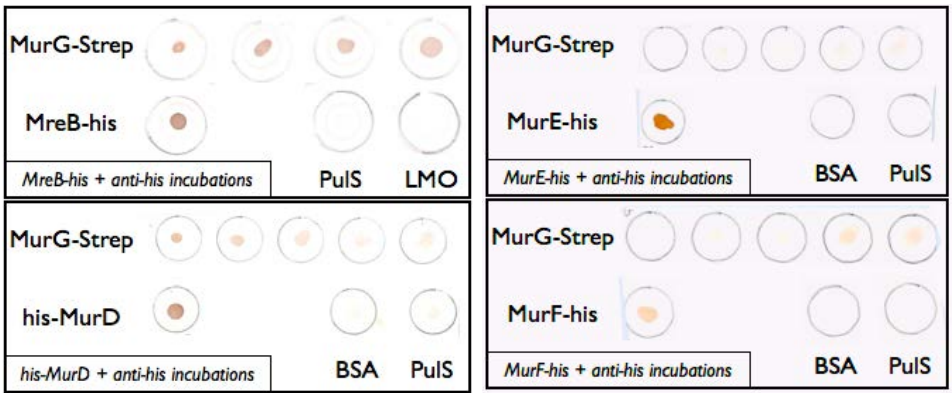


Figure 14.7: Dot blot on purified proteins showing direct interaction between immobilized MurG and MreB, MurE, and MurF. The nitrocellulose membranes were dotted with increasing amounts of Strep-tagged MurG. Membranes were incubated with 0.1 mg.mL^{-1} of either MreB, MurD, MurE, or MurF as indicated, subsequently incubated with anti-his HRP antibody for development. His-tagged proteins were loaded as positive controls, and negative controls (BSA, PulS) are represented as well.

in Figure 14.8.

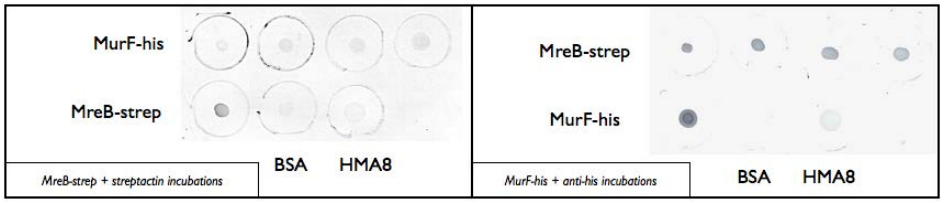


Figure 14.8: Dot blot on purified proteins showing direct interaction between MurF-his and MreB-strep. The nitrocellulose membranes were dotted with increasing amounts of his-tagged MurF (left) or Strep-tagged MreB (right). Positive and negative controls (BSA, and HMA8, a chloroplast protein.) were loaded as well. Membranes were incubated with 0.1 mg.mL^{-1} of either MreB-strep (left), of MurF (right) as indicated, subsequently incubated with anti-his HRP antibody or streptactin HRP probe for development, as mentioned. Note that a signal was detected for the negative control with HMA8.

14.2 Surface Plasmon Resonance assays

In order to confirm the positive results from the dot blot experiments, Surface Plasmon Resonance spectroscopy was performed for further testing Mur interactions.

14.2.1 Deciphering the interaction network of Mur ligases

Increasing concentrations of MurD, MurE, and MurF, ranging from 0.2 μM to 2.4 μM were injected over a CM5 sensor chip onto which MreB had been immobilized (Figure 14.10, left). The response was clearly dose-dependent, indicating that each of the three ligases, MurD included, readily bound to the bacterial actin homolog. Similarly, MurG was covalently linked to a CM5 gold chip and assessed for interaction with MurD, MurE, MurF, as analytes in increasing concentrations from 0.25 μM to 8 μM (Figure 14.10, right). Resulting curves confirmed that the glycosyltransferase is an interaction partner of MurD, MurE and MurF. The same experiment performed using immobilized MurE, confirmed its interaction with MurG and MreB (not shown).

The kinetic analysis of MreB biosensorgrams revealed K_D constants in the range of 7–30 nM and χ^2 values around 1.0 (see Figure 14.9). It is of note that these values were obtained with a two-state reaction model where a conformational change is considered in addition to a classical 1:1 binding mode.

Immobilized protein	Analyte	ka1 ($M^{-1}s^{-1}$)	kd1 (s^{-1})	ka2 (s^{-1})	kd2 (s^{-1})	K_D	χ_2
MreB	MurD	$4.8\ 10^4$	$7.0\ 10^{-3}$	$8.7\ 10^{-3}$	$7.7\ 10^{-4}$	$1.2\ 10^{-8}$	1.05
MreB	MurE	$8.0\ 10^4$	$9.1\ 10^{-3}$	$8.9\ 10^{-3}$	$5.8\ 10^{-4}$	$7.4\ 10^{-9}$	0.54
MreB	MurF	$3\ 10^4$	$5.9\ 10^{-3}$	$9.5\ 10^{-3}$	$1.7\ 10^{-3}$	$2.9\ 10^{-8}$	0.63

Figure 14.9: Kinetic constants for the interactions of Mur ligases with MurG and MreB. Binding was measured as described in material and methods. The association (ka1, ka2) and dissociation (kd1, kd2) rate constants were determined by global fitting of the data using a two-state binding model. The dissociation constants K_D were determined from the (kd1/ka1)*kd2/(ka2+kd2) ratio.

MurG-MurD and MurG-MurE binding curves, despite clearly indicating that there is interaction between the two protein pairs, could not be fitted using any of the models available in the BIA Evaluation software package, indicating that binding may not follow a simple kinetic model.

By contrast, no signal could be detected when MurD or MurE were tested over a MurF-immobilized surface (not shown), confirming our previous observations that Mur ligases do not interact with each other.

14.2.2 Interaction between MurG and MreB

MurG and MreB were found to interact with each other as well, thus indicating that MurG and MreB could be the main actors of the Mur interaction network. Indeed, when MreB was injected over a MurG surface, a dose-dependent signal was detected (see Figure 14.11).

The interaction was assessed when MreB was immobilized as well. However, MurG stuck to the surface which, then, could not be regenerated.

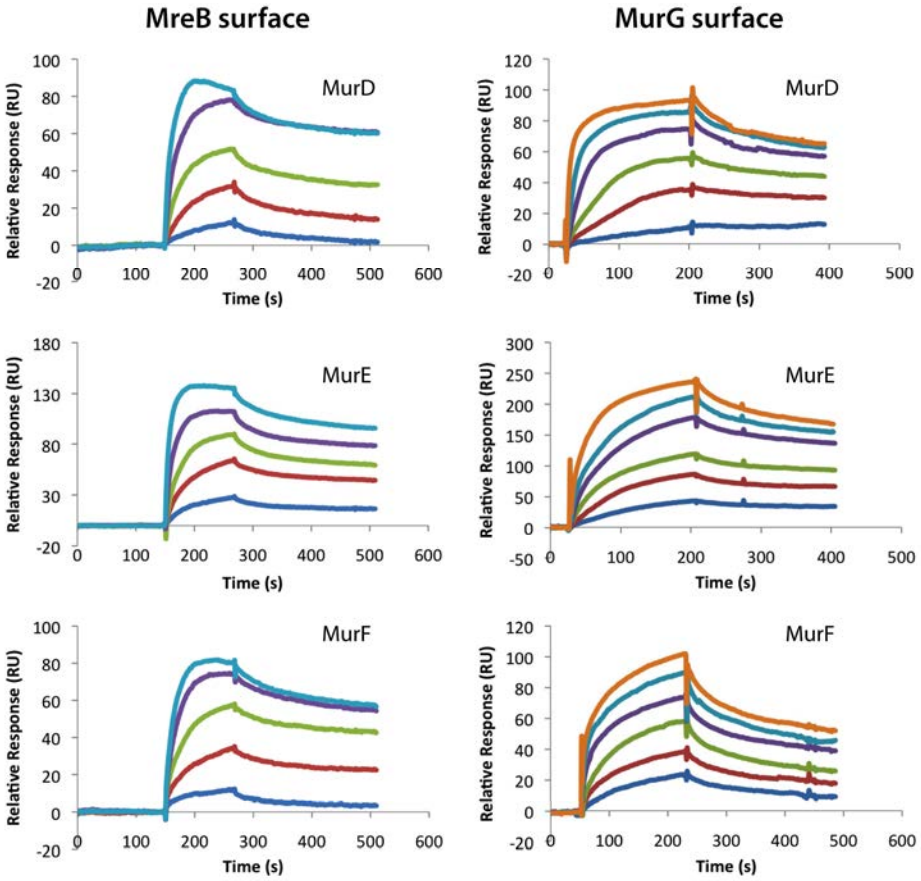


Figure 14.10: Surface plasmon resonance assays showing that MurG and MreB directly interact with MurD, MurE, and MurF. MurG was immobilized onto a CM5 sensor chip and MreB, MurD, MurE and MurF were tested as analytes in varying concentrations (from bottom to top): 0.25 μM, 0.5 μM, 1 μM, 2 μM, 4 μM, 8 μM. MreB was immobilized onto a CM5 sensor chip, and Mur ligases were injected as analytes. From bottom to top: 0.2 μM, 0.4 μM, 0.8 μM, 1.6 μM, 2.4 μM.

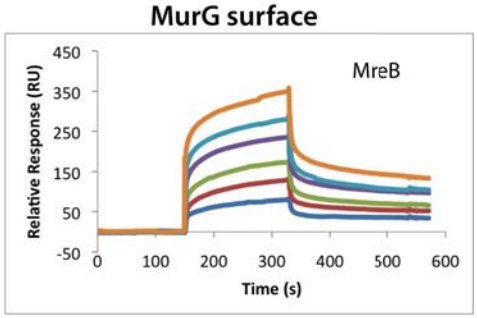


Figure 14.11: Surface plasmon resonance assays showing that MreB interacts with MurG. MurG was immobilized onto a CM5 sensor chip and MreB was tested as analyte in varying concentrations (from bottom to top): 0.25 μM, 0.5 μM, 1 μM, 2 μM, 4 μM, 8 μM.

Notably, the MurG-MreB sensorgram indicates that no saturation point is reached at high MreB concentrations, suggesting that the cytoskeletal protein oligomerizes onto the MurG surface as previously suggested by Gaballah and colleagues [15].

14.2.3 Self-interactions of MurG and MreB

In addition to these interactions, MurG and MreB were both found to self-interact, presenting typical curves in which no plateau was reached, supporting an oligomerization or polymerization (data not shown).

14.2.4 Negative conrols

To confirm the specificity of the interactions detected by SPR, BSA and his-tagged proteins with no relationship with peptidoglycan biosynthesis were injected onto MreB and MurG surfaces in similar conditions. No signal could be detected. Indeed, protein controls 1 and 2 did not display any binding to neither MurG nor MreB surfaces, whereas MurE and MurF clearly interacted with both immobilized proteins (see Figure 14.12). This confirms the specificity of the interactions.

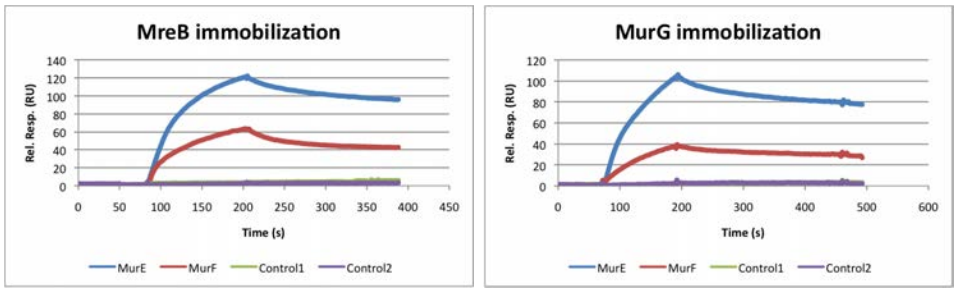


Figure 14.12: SPR control experiments performed with two distinct his-tagged bacterial proteins whose function is not related to peptidoglycan biosynthesis. MurE, MurF and control proteins 1 and 2 were injected at a concentration of 600 nM over immobilized MurG and MreB.

The Mur interaction network

Dot blot assays and Surface Plasmon Spectroscopy indicated that MurD, MurE and MurF all recognize MurG and MreB, but not each other, whilst the two latter proteins interact. Optimization is still required to isolate a Mur complex in sufficient purity and quantity for crystallization purposes.

14.3 Towards the crystal structure of a MreB-Mur complex

As purification conditions for the formation of a stable Mur complex could not be found, an alternative approach was tried: co-crystallization. As MreB behaved better than MurG *in vitro*, priority was given to the study of potential complexes between MreB and Mur ligases. Thus, single proteins were purified and mixed by pairs before crystallization screening.

The strategy was the following:

- High-throughput screening[3]: about 600 crystallization conditions per complex and per ratio, in nanodrops. In parallel, identical screenings were performed on single proteins as controls;
- Analysis of the results to identify conditions which seemed specific to complexes;
- Optimization and control of conditions for potential crystals of complexes[4]: in the same well, one drop for the protein mix, and two drops with each of the two single proteins as controls;
- Characterization of the most interesting crystals for X-ray scattering on beam lines.

14.3.1 Crystallization trials of MreB:Mur complexes

1:1 and 2:1 ratios

Since MreB forms polymers in the cell, an intuitive approach suggests that several units of MreB could be necessary for a stable interaction with Mur partners. For this reason, 1:1 and 2:1 ratios for each of the MreB:MurD, MreB:MurE, MreB:MurF mixes were assessed first.

As MreB alone crystallizes very well in a myriad of conditions, an extensive data analysis was needed to try to identify the conditions that seemed specific to complexes. An optimization with single proteins as controls in the same crystallization wells, allowed to further check specificity of crystallization conditions for protein mixes. For this higher-scale screening, 6 96-well plates were prepared by the crystallization platform of the IBS, with one drop with MreB alone, one with the Mur ligase alone, and one with the MreB-Mur mix. Figure 14.13 shows an example of such a well.

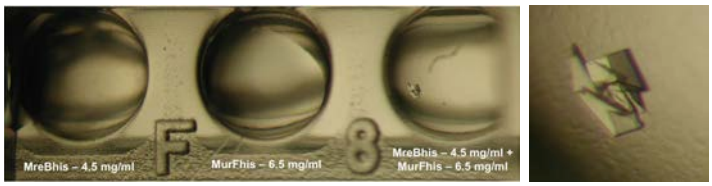


Figure 14.13: One well of the crystallization plate set up by the crystallization platform of the IBS (Left) and zoom onto a potential MreB-MurF crystal (Right). Left drop: MreBhis 4.5 mg/ml (clear), Middle drop: MurFhis 6.5 mg/ml (clear), Right drop: MreBhis 4.5 mg/ml + MurFhis 6.5 mg/ml (crystal). Crystallization condition: 0.05 Sodium cacodylate pH 6.5, 0.2 M ammonium acetate, 20 % iso-propanol; 20 °C.

Most interesting crystallization conditions could be reproduced. From them, conditions in which one of the single proteins crystallized in the same way as the mix of proteins were excluded, and just a few interesting conditions remained, for which crystals were tested for X-ray diffraction. Unfortunately, either diffraction was too poor or crystals presented the same cell parameters and space group as MreB.

1:2 ratio

MreB crystallizes at about 4 mg/ml, whereas Mur proteins crystallize around 10 mg/ml. Thus, when mixing proteins at a 1:1 ratio, MreB is very close to supersaturation, while Mur

[3]performed at the HTX lab at the European Molecular Biology Laboratory (EMBL) of the Partnership for Structural Biology (PSB).

[4]performed at the crystallization platform of the IBS in 2 μl-drops by Delphine Blot.

proteins are still very soluble. This may not be optimal for stabilizing a MreB-Mur complex. Thus, a second high-throughput screen was performed, with a 1:2 ratio, in which both single proteins are close to supersaturation state.

Much less crystallization conditions were obtained, most likely because of the relatively high concentration in total protein (about 15 mg/ml), leading to a high percentage of precipitated drops. However, this screen shed light on a few particularly interesting crystallization conditions. Most were for MreB-MurE, and a few for the two other complexes. Optimization on the IBS crystallization platform suggested that MreB-MurE was the best candidate for co-crystallization assays, as MreB-MurD and MreB-MurF conditions appeared to be unspecific for protein mixes or could not be reproduced. Therefore, focus was put on MreB-MurE mix and further optimization was performed as listed Table 14.8, except for one crystal from a 'MreB-MurF' drop (see next paragraph).

Condition	**MreBhis**	**MurEhis**	**1MreBhis:2MurEhis**
PEG 6000 7%-12%, 0.1 M MES pH 6-6.5	Light precipitate	Clear or flower	Toblerone-like form
PEG 6000 7%-11%, pH 8-9, 1 M LiCl	Small hexagons	Clear	Hexagons clusters and sticks
PEG 8000 7%-10%, 0.025-0.075 M KH_2PO_3	Clear	Flowers	Chromosome-like crystals and sticks
PEG 4000 3%-7% pH 6-6.5	Clear	Flowers	Toblerone-like form
PEG 20000 12%-13%, 0.1 M MES pH 6-6.5	Light precipitate	Light precipitate	Sticks and chromosome-like
PEGMME 5000 5%, 0.1 M MES pH 6	Light precipitate	Sticks at pH 5.0 and 7% PEG	Toblerone-like form
PEG 1500 10%-20%	Light precipitate	Precipitate	Precipitate
PEG 4000 12%-17%	Light precipitate	Precipitate	Precipitate

Table 14.8: Interesting co-crystallization conditions for MreBhis-MurEhis complex unveiled by the HTX platform, and tried to be reproduced on the crystallization platform of the IBS.

Further knowledge on MurE crystallization suggested that flower-like crystals were in fact MurE crystals. However, Toblerone-like crystals had never been seen in single protein drops and seemed highly specific to MreB-MurE mixes. Thus, a focus was put on this crystal shape which could be very easily reproduced in high amounts and appeared to be relatively stable.

14.3.2 X-ray scattering

Toblerone-like crystals were screened for X-ray scattering on various beam lines and with various cryoprotectants, directly from robot plates or from hand-made plates. The quality of the diffraction pattern obtained from these crystals was highly variable: In average, 1 crystal out of 25 diffracted around 3 Å resolution, while the rest diffracted to a resolution lower than 8 Å. This variability could not be explained, and might be due to uncontrollable parameters of crystallization. Strikingly, the diffraction pattern for these toblerone-like crystal was different from the previously solved structures for the single proteins, and data sets were collected.

Other crystals from mixes of MreB with MurE were tested. Some of them diffracted well, but presented the scattering features of MreB alone: Cell parameters and space groups

corresponded to known structures of MreB.

In parallel to MreB-MurE optimization, an interesting lead for MreB-MurF from the HTX screen appeared six weeks after the plate set-up in a precipitated drop (see Figure 14.14), but has never been reproduced yet. This potential crystal of MreB-MurF showed a nice diffraction pattern as well, which was different from any known diffraction pattern for single proteins. Table 14.9 summarizes the two best crystallization conditions identified for potential MreB-MurE and MreB-MurF complexes.

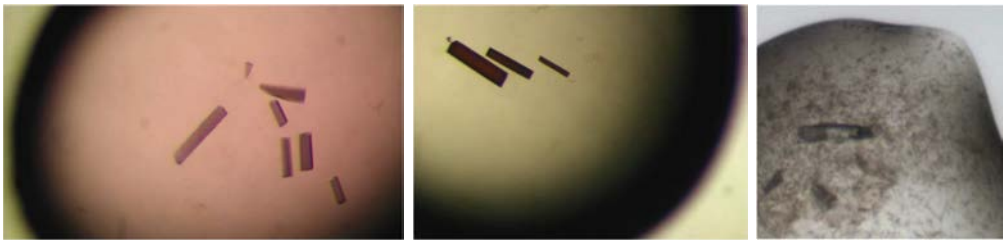


Figure 14.14: Crystals from two hand-made MreB-MurE mix drops (left and center), and crystal from a MreB-MurF mix nanodrop (right). Crystallization condition for MreB-MurE: 0.1 M MES pH 6 - 6.5, 0.5 % to 5 % PEG 4000 or 6000, 1 day to 1 week growth - soaked with Ruthenium Red (Left) or Osmium (Center). Crystallization condition for MreB-MurF: : 0.1 M MES pH 6.0 5 % PEG6000, six-week growth; 20°C.

Protein	**Crystallization liquor**	**Growth time**	**Crystal shape**
MreB-MurE ratio 1:2	0.1 M MES pH 6.0, 1% PEG 4000	1 week	Toblerone-like
MreB-MurF 1:1 to 1:2 ratio	0.1 M MES pH 6.0, 5% PEG 6000	2 months	Irregular stick

Table 14.9: Most promising crystallization conditions for MreB-MurE and MreB-MurF complexes). Proteins were purified in 25 mM Hepes pH 7.4, 150 mM NaCl, 1 mM EDTA, mixed, and 5 mM AMPPNP were added. Crystals of Mur ligases were grown with the hanging-drop vapor diffusion technique at 20°C. Note: As MreB started to precipitate during the set up of the MreB-MurF screen, accurate ratio for this experiment is unknown.

14.3.3 Indexation

A complete data set could be collected for 'MreB-MurE'crystals of this best condition, while a partial data set was collected for the 'MreB-MurF' crystal (see Table 14.10). Interestingly, space groups and cell parameters were different than for single proteins.

Protein	**'MreB-MurE'**	**'MreB-MurF'**
Resolution (Å)	2.7	2.7
Space group	$P2_12_12_1$	$P2_12_12_1$
Cell parameters (Å)	53 108 140	47 129 159
Crystal mosaicity (°)	0.163	0.354
Completeness	91 %	50 %
R-factor	2.9 %	

Table 14.10: Data sets of potential MreB-MurE and MreB-MurF crystals. Scattering data were collected on ID14eh4 (MreB-MurE) and ID23eh1 (MreB-MurF) at the ESRF.

Data regarding MreB-MurF data set could not be further processed as the completeness was too low.

14.3.4 Molecular replacement

Molecular replacement trials using Phaser and the structure of MurE from *T. maritima* solved previously during this thesis only identified poor phases, while no solution could be obtained for MreB.

14.3.5 Crystallization of SeMet MreB-MurE mix

As reported in the Chapter 12, SeMet derivatives for both proteins were prepared. However, only MurE-like flower crystals grew, confirming our previous observation that when selenomethionylated, MurE preferentially crystallizes into this flower-shape. Thus, SeMet MreB was mixed to native MurE, giving rise to very polycrystalline crystals. Unfortunately, the latter did not diffract properly for a X-ray scattering experiment.

14.3.6 Soaking with heavy atoms

In order to obtain better phases for 'MreB-MurE' crystals, soaking assays and co-crystallization with heavy atoms were performed. Crystals could grow easily even in presence of heavy atoms, and were relatively resistant to soaking. After flash cooling in a cryoprotectant solution containing 10 mM of heavy atom, crystals were screened for X-ray scattering at the ESRF. Table 14.11 summarizes the different heavy atoms which were tested, and the corresponding data sets that could be collected.

Salt	**Solubility**	**CoX / soaking**	**Crystal appearance**	**Collection**
Ruthenium Red	ok	CoX + soaking	ok	2 data sets
KCl_6Os IV	ok	CoX + soaking	ok	no
Thimerosal (Hg)	ok	CoX + soaking	ok	2 data sets
Ethyl HgCl	poor	soaking	ok	-
Gadolinium III acetate	no	-	-	-
Samarium Nitrate	no	-	-	-
Sodium Tungstate	ok	soaking	dissolved	-
Sodium Selenite	ok	CoX + soaking	ok	no
Potassium Iodide	ok	CoX	ok	no
Cesium Chloride	ok	soaking	ok	no
Barium Chloride	ok	CoX + soaking	ok	1 data set

Table 14.11: Soaking assays and co-crystallization of 'MreB-MurE' crystals with heavy atoms. 'Solubility' refers to the solubility in the crystallization liquor. Heavy atoms were used at a final concentration in the crystallization drop of 5 mM. CoX = co-crystallization. Collection was performed at the ESRF on various beamlines.

Though some anomalous signal was detected for the different data sets that were collected, phasing was not successful.

Finally, soaking assays with 5 mM of zinc sulfate was performed. About 60 crystals were screened to collect four data sets. One of them allowed to phase the data, confirming that MurE was the only protein contained in these crystals.

Towards the crystal structure of a MreB-Mur complex

As isolating a Mur complex through biochemical assays was particularly arduous, high-throughput crystallization screens were performed in order to co-crystallize a pair MreB-Mur ligase. These assays unveiled an interesting condition which was highly specific for MreB-MurE mixes. However, after testing hundreds of crystals for X-ray scattering, a SAD experiment at the zinc edge allowed to conclude that MreB had not been co-crystallized with MurE. Another condition, regarding the MreB-MurF pair remains to be further tested.

Chapter 15

Discussion and future perspectives

A summary of the results obtained, and perspectives on the purification of a Mur complex are proposed in this part. Taking into account the knowledge acquired through experiments, further assays and novel research projects are suggested as well.

15.1 Single proteins

This section focuses on the *in vitro* and crystallogenic behaviour of Mur ligases, MurG, and MreB from *T. maritima*.

15.1.1 Mur ligases

This work set up purification protocols to get high amounts of pure Mur ligases from *T. maritima* by two steps: an affinity chromatography, followed by a size exclusion chromatography. This allowed to complete the structural knowledge of Mur ligases from *T. maritima*, by adding to the known structure of MurC, a crystal structure of MurD, two structures of MurE in complex with ADP, one structure of apo MurF and one of MurF in complex with ADP. Results, specificities of purification of each enzyme, and global analysis of the structures are discussed in this section.

Though MurD was seen to be very stable and easily heat-purified, with a melting temperature extremely high according to TSA, it did not crystallize so easily. Actually, crystals needed several weeks to grow, were often very polycrystalline, small, and a high variability was observed depending on protein batches and crystallization plates. Nevertheless, the structure of the enzyme in apo form could be solved at 2.2 Å resolution in a $P2_12_12_1$, with two molecules per asymmetric unit. Interestingly, the two molecules present two different conformations where the C-terminal domain is more or less away from the rest of the molecule, supporting the very high flexibility of this domain in Mur enzymes.

By contrast, MurE crystallized very well in a high number of crystallization conditions in many different crystal shapes but most crystals diffracted poorly. However, this allowed to solve the structure of MurE in complex with ADP at 2.9 Å resolution in a $P6_1$ space group, and a second structure in a $P2_12_12_1$ at 2.7 Å resolution was obtained as well. These data give a molecular basis for the understanding of the specificity of the enzyme for L-lysine, as

described by Boniface *et al.*, 2006 [1]. However, assays of crystallization of MurE in complex with its product failed, so did co-crystallization trials with inhibitor compounds kindly provided by Andrej Perdih (Laboratory for Biocomputing and Bioinformatics of the National Institute of Chemistry in Slovenia) or Martina Hrast (University of Ljubljana, Faculty of Pharmacy, Slovenia). This suggests that, when complexed with one of these ligands, MurE could undergo a significant conformational change, therefore exposing hydrophobic patches on its surface, resulting in lower solubility. It is of note that the activity of the recombinant protein was suggested by the toxicity of its over-expression in *E. coli* cells as the enzyme from *T. maritima* incorporates a 'wrong' amino-acid residue for *E. coli* [1].

MurF could be purified in very high amounts, and its activity was checked by Martina Hrast. This ligase most often crystallized in tiny and polycrystalline needle crystals, difficult to optimize and batch dependent. However, the variability could be significantly reduced with automated purification. MurF crystals grew fast and diffracted well, making this ligase the best candidate for compounds screening among the Mur ligases from *T. maritima*. Crystallization and diffraction assays resulted in two crystal structures of MurF: one at 1.65 Å resolution in a $P2_12_12_1$ space group in complex with ADP, and a second one at 2.1 Å resolution in apo form in a $P1$ space group. Interestingly, an identical, relatively closed, conformation is seen in the two structures, supporting the idea that nucleotide binding does not infer a strong conformational change in Mur ligases. However, as phosphates are present in the nucleotide pocket of the second structure, the phosphate moiety of the nucleotide could be sufficient for conformational closure.

Although co-crystallization assays with a compound did not give the crystal structure of a complex yet, other types of compounds, and lower concentration of phosphate in the crystallization liquor have to be tried.

The purification of MurC, for which a crystal structure was already available [17], needs to be optimized for interaction assays. Indeed, under our purification conditions, MurC was not very soluble in cell lysates, tended to aggregate, and remained bound to a high-molecular weight protein after the two-step purification. However, the published crystal structure reported a N-terminal His-tag fusion, while the protein purified in this work was a C-terminal Strep-tag fusion. This suggests that MurC from *T. maritima* could be less stable when expressed with a C-terminal Strep tag.

In addition to these specificities of each Mur ligase, MurD, MurE, and MurF were all found to dimerize partially *in vitro*. However, this homo-oligomerization has never been reported to play a key role in functionality, as assessed by the work of Jin *et al.* [18]. In addition, -27 Da shifts in mass spectrometry analyses were seen several times. This can be due to incorporation of wrong amino acid residues by *E. coli*, as *mur* genes from *T. maritina* contain rare codons for *E. coli*.

Regarding Mur structures, despite minimal sequence homology (arount 30 %), Mur ligases have high structural similarity with their homologs. However, there were too many little differences with homologs for molecular replacement to work. Thereupon, Se-Met derivatives were required for phasing. Thanks to the high flexibility of the C-terminal domain, the boundary between central and C-terminal domains could be easily determined. By contrast, the delimitation between the N-terminal and central domains is much more difficult to place.

With Mur ligases from *E. coli*, the structures of *T. maritima* Mur ligases is the second complete set of structures of the Mur ligase family, providing an additional basis for the study of the adaptation of the Mur enzymes to the growing substrate of peptidoglycan. As the concentration of ATP in the cell is in the millimolar range, while

15.1.2 MurG

While Mur ligases were relatively easy to handle and purify, *in vitro* behaviour of MurG appeared to be much more difficult to control.

Particularly, MurG could not be solubilized with any detergent but N-lauroyl sarcosine[1]. This reagent is used in purification of recombinant proteins with two main purposes:

- for specific solubilization of the cytoplasmic membrane in *E. coli* [19];
- and for the recovery of proteins aggregated into inclusion bodies. Indeed, sarcosyl has the capability of solubilizing inclusion bodies or protein aggregates [20] [21], and is known to be a protecting osmolyte[2] [3] [22]. Particularly, purifications of proteins from inclusion bodies thanks to sarkosyl have been reported to produce functional, folded proteins [23] [24] [25].

Size exclusion chromatography profiles, oligomerization, and effective interactions of MurG suggested that the protein was well folded. However, MurG aggregation in absence of detergent *in vitro*, and possible expression as inclusion bodies *in vivo* support the hypothesis of the presence of a hydrophobic region that, *in vivo*, would be responsible for membrane binding [26] [16], as it has been suggested for *E. coli*'s homolog [26] [27].

MurG was found to interact with SuperDex columns, but not with Superose6. Binding of MurG on the CM5 sensorchip made of dextran was very fast as well. These observations suggest that MurG interacts with dextran. This has been reported for glucosyltransferase from *Streptococcus sobrinus* [28]. However, no dextran-binding domain could be identified in MurG sequence.

MurG often appeared as oligomeric form in size exclusion chromatographies and electron microscopy experiments. Interestingly, high-molecular bands could be seen on SDS-PAGE when the sample was not heated and under non-reducing conditions. However, the composition of these lanes must be confirmed by N-terminal sequencing. The oligomeric state was seen to be very sensitive to experimental conditions and was time-dependent. An optimization of MurG buffer in order to further control its oligomerization and stability could allow the acquisition of better electron microscopy data, and open up crystallographic possibilities. Several ways for optimization of MurG solubility can be proposed:

- Expression in C41(DE3) or C43(DE3) cells which have been selected for their resistance to expression of membrane proteins and present higher amounts of membranes than usual *E. coli* cells. Indeed, by providing more membrane material to MurG, its overexpression may result in a more stable protein.
- Strongly reducing the amount of inducer and the temperature induction in order to prevent aggregation in the cell. However, the yield might be reduced.

[1] N-Dodecanoyl-N-methylglycine (sarkosyl), is an anionic surfactant. It is made of a hydrophobic 14-carbon chain (lauroyl) and a hydrophilic carboxylate.

[2] An osmolyte is defined by its ability to move the equilibrium unfolded - folded towards the folded state.

[3] For instance, sarkosyl is known to counteract the denaturing effect of urea [22].

- Set up techniques to monitor MurG stability and folding. As TSA with traditional Sypro probe cannot be used because of the presence of detergent and hydrophobic patches on the protein, alternative techniques must be found to optimize MurG buffers. Circular dichroism studies may be of paramount interest in the optimization of MurG purification conditions.
- Change the construct for a N-terminal tag. The carboxy 66-terminal amino acid residues of *E. coli* MurG has been suggested to be involved in proper folding of the enzyme [26]. The presence of a C-terminal tag may interfere with the proposed folding function of the C terminus of MurG.

15.1.3 MreB

In order to investigate the potential role of MreB in formation of a Mur complex, the bacterial actin homolog was added to the study. Thus, a purification protocol was set up and a crystal structure solved by molecular replacement.

While BL21(DE3) cells stop growing under MreB expression, C41(DE3) grew normally and enable a good overexpression of recombinant MreB. As C41(DE3) cells bear more membrane material than original BL21(DE3), this is in line with a study of Salje *et al.* [29] which reported that MreB binds cell membranes.

However, as MreB polymerizes in the cells, a notable part was lost during the ultracentrifugation step, as polymerized form. Therefore, the yield of MreB purification was lower than for Mur ligases, but enough protein for crystallization trials could be easily produced. Strikingly, TSA of MreB resulted in a relatively low melting temperature, indicating that either MreB was not very stable in purification buffer, or that thermostability of proteins could depend on their interaction with partners as well. The propensity of MreB solution to become slightly white may indicate a stacking in solution though gel filtration assays detected the monomeric form only.

In spite of the apparently low stability of MreB from *T. maritima* in solution, the bacterial actin homolog crystallized very well in many conditions, making MreB a good candidate for crystallization studies. A structure at 1.4 Å resolution could be solved in a $P32_12_1$ space group, providing further structural information to the previous published structure at 2.1 Å resolution [3]. Crystallographic screens combined with assays of inhibition of polymerization and *in vivo* experiments on other bacterial strains, should be fruitful in finding a good MreB inhibitor.

15.2 Towards the crystallization of a Mur complex

Dot blot and SPR assays showed that MurD, MurE, and MurF all recognize MurG and MreB, but not each other, whilst the two latter proteins interact. This suggests that MurG and MreB would act as a scaffold for the cytoplasmic steps of peptidoglycan biosynthesis, and therefore highlights the need to determine *in vitro* conditions in which MurG would be more stable, with the aim at crystallizing a Mur complex.

15.2.1 Detection

Dot blot assay appeared to be an easy, sensitive, and specific technique for investigation of protein-protein interactions. Additional similar experiments can be easily performed to confirm the interactions between MreB-strep and his-MurD, or MurE-his.

SPR assays revealed clear, specific, dose-dependent interactions. However, for all Mur assays, the maximum analyte binding capacity was low compared with the amount of immobilized protein. This could be explained by different quaternary states adopted by these proteins, and by the conformational flexibility of Mur ligases as well. In addition, immobilization on CM5 chips may interfere with a proper interaction, as covalent binding to such chips is not oriented. SPR technique may be used to identify which domain of Mur ligases is involved in binding, and whether the presence of ligands (AMPPNP, growing peptidoglycan unit, or the incoming amino acid) in the solution of the analyte could help stabilizing the interactions. Native gels were problematic to set up because MreB did not enter the gels and MurG was a basic protein.

The formation of complexes could not be detected by cross-linking experiments. However, only two cross-linking agents were tried - EGS and glutaraldehyde. As Mur ligases tend to dimerize, MurG to form homo-oligomers, and MreB to polymerize or at least 'stack' in absence of ATP, using hetero-bifunctional cross-linkers[4] could be helpful as they avoid self-conjugation.

15.2.2 From detection to crystallization of a Mur complex

Biochemical isolation

Nevertheless, the purpose of this project was to solve the structure of a Mur complex. Therefore, emphasis was put on finding experimental conditions for purification of a stable complex. Despite the clear signal and the reproducibility of the interactions detected by SPR, purifying a Mur complex appeared to be very difficult as the interaction could not be clearly detected by SDS-PAGE and best results from co-purification assays were often hard to reproduce.

Reproducibility issues for MurG-Mur ligases pull-down assays were most likely due to the high sensitivity of the glycosyltransferase to experimental conditions and its propensity to interact with purification columns. To increase the chances of isolating such a complex, purification conditions for MurG must be further optimized to control its oligomerization and enhance its stability, and a resin with which MurG does not interact at all has to be found.

To increase the odds of catching a MreB-Mur complex in a purification experiment, co-expressing a MreB mutant which does not polymerize with other proteins could be a good strategy. Indeed, if co-expressing native MreB together with Mur proteins, Mur enzymes may prefer to bind MreB polymers that are present in the cells rather than MreB monomers. This

[4]Hetero-bifunctional cross-linkers have two distinct reactive groups, thereby avoiding the formation of dimers and polymers. For instance, one protein reacts with the amine-specific end of a reagent while the other protein is treated with a sulfhydryl-addition reagent to create sulfhydryl groups. Then, the two proteins are mixed to allow the sulfhydryl-reactive groups of the first protein to conjugate with the sulfhydryl groups of the second protein.

would result in loss of complex entities during the ultracentrifugation step of the purification. Besides, purification of MreB polymers is not suitable for crystallographic experiments which require a high homogeneity in protein sample.

Reproducibility of MreB-MurG pull-down experiments was problematic as well. Here again, a better control of the glycosyltransferase is required. Moreover, a more sensitive detection of the complex must be set up, such as 2D gels or N-terminal sequencing of both proteins, as MurG and MreB migrate the same on SDS-PAGE. However, as both MreB and MurG interact with either of the Mur ligases, purification of a Mur complex based on this 'scaffold' may be or paramount interest.

Co-expression and co-lysis experiments were preferred to pull-downs on pure proteins in order to keep co-factors, products, substrate from the cells. Indeed, such ligands could potentially be required for the formation of a stable complex, for instance by limiting the conformational flexibility. The presence of tags may interfere with interactions stability as well, and a screen for expression constructs may be useful in the search for experimental conditions for the purification of a Mur complex.

Isolation by co-crystallization

High-throughput screening of co-crystallization conditions allows to test a very broad range of conditions, enhancing the chances of finding the one which stabilizes a protein complex. Such an experiment for MreB-Mur ligases pairs could identify a few conditions in which single proteins did not crystallize, but mixes of MreB with MurE, or MurF, did crystallize. Nevertheless, the best condition for co-crystallization trials of MreB with MurE finally resulted in crystals containing only MurE, as revealed by a SAD experiment at the zinc edge. The best condition for the MreB/MurF pair has still to be reproduced, in order to collect a complete data set.

This work, in addition to providing novel structures of Mur ligases, marks a further step in the characterization of a cytoplasmic peptidoglycan multi-partite complex. Such a machinery would limit the diffusion of peptidoglycan precursor intermediates and channel peptidoglycan building blocks towards the inner membrane. Though pair complexes formation could be clearly detected by SPR, purifying them appeared particularly arduous. The structures of MurD, MurE, MurF that were solved during the work of this PhD underlined the high flexibility of these enzyme, suggesting that conformational changes could be important for the formation and stabilization of a Mur complex. Besides, polymers of MreB could be required for the formation of a stable Mur complex. One one hand, this would complicate crystallization trials. On the other hand, this would open up new possibilities of electron microscopy experiments. For this PhD, *T. maritima* was chosen as model to study the Mur interaction network, because of its thermophilic aspect and the good behaviour of MreB from this species *in vitro*. However, another bacterial model may be more suitable for purification of the complex. Particularly, neither MreC nor MreD homologs were identified in *T. maritima*, suggesting that another organization of peptidoglycan biosynthesis may be followed in this organism. Indeed, MreC and MreD have been reported to play a major role

in the periplasmic organization of peptidoglycan biosynthesis [16].

More generally, this study supports that biology must not be seen only as pathways, but rather as multi-partite machineries, opening up new possibilities of drugs which would target not only active sites, but interaction surfaces as well.

Bibliography

[1] Audrey Boniface, Ahmed Bouhss, Dominique Mengin-Lecreulx, and Didier Blanot. The MurE synthetase from *Thermotoga maritima* is endowed with an unusual D-lysine adding activity. *J. Biol. Chem.*, 281(23):15680–15686, 2006.

[2] Clyde A. Smith. Structure, function and dynamics in the Mur family of bacterial cell wall ligases. *J. Mol. Biol.*, 362(4):640–655, 2006.

[3] Fusinita van den Ent, Linda A. Amos, and Jan Lowe. Prokaryotic origin of the actin cytoskeleton. *Nature*, 413(6851):39–44, 2001.

[4] Youwei Yan, Sanjeev Munshi, Barbara Leiting, Matt S. Anderson, John Chrzas, and Zhongguo Chen. Crystal structure of *Escherichia coli* UDPMur*N*Ac-tripeptide D-alanyl-D-alanine-adding enzyme (MurF) at 2.3 Å resolution. *J. Mol. Biol.*, 304(3):435 – 445, 2000.

[5] Kenton L. Longenecker, Geoffrey F. Stamper, Philip J. Hajduk, Elizabeth H. Fry, Clarissa G. Jakob, John E. Harlan, Rohinton Edalji, Diane M. Bartley, Karl A. Walter, Larry R. Solomon, Thomas F. Holzman, Yu Gui Gu, Claude G. Lerner, Bruce A. Beutel, and Vincent S. Stoll. Structure of MurF from *Streptococcus pneumoniae* co-crystallized with a small molecule inhibitor exhibits interdomain closure. *Protein Science*, 14(12):3039–3047, 2005.

[6] Sandy Favini-Stabile, Carlos Contreras-Martel, Nicole Thielens, and Andréa Dessen. MreB and MurG as scaffolds for the cytoplasmic steps of peptidoglycan biosynthesis. *Environmental Microbiology*, pages 1–11, 2013.

[7] Andrew L. Lovering, Susan S. Safadi, and Natalie C. J. Strynadka. Structural perspective of peptidoglycan biosynthesis and assembly. *Annu. Rev. Biochem.*, 81:451–478, 2012.

[8] Elspeth Gordon, Bernard Flouret, Laurent Chantalat, Jean van Heijenoort, Dominique Mengin-Lecreulx, and Otto Dideberg. Crystal Structure of UDP-N-acetylmuramoyl-L-alanyl-D-glutamate:meso-Diaminopimelate Ligase from *Escherichia Coli*. *J. Biol. Chem.*, 276(14):10999–11006, 2001.

[9] Chandrakala Basavannacharya, Giles Robertson, Tulika Munshi, Nicholas H. Keep, and Sanjib Bhakta. ATP-dependent MurE ligase in *Mycobacterium tuberculosis*: Biochemical and structural characterisation . *Tuberculosis*, 90(1):16 – 24, 2010.

[10] Jay A. Bertrand, Genevieve Auger, Eric Fanchon, Lydie Martin, Didier Blanot, Jean van Heijenoort, and Otto Dideberg. Crystal structure of UDP-N-acetylmuramoyl-L-alanine:D-glutamate ligase from *Escherichia coli*. *EMBO J*, 16(12):3416–3425, June 1997.

[11] Jay A. Bertrand, Genevi Auger, Lydie Martin, Eric Fanchon, Didier Blanot, Dominique Le Beller, Jean van Heijenoort, and Otto Dideberg. Determination of the

MurD mechanism through crystallographic analysis of enzyme complexes. *J. Mol. Biol.*, 289(3):579–590, 1999.

[12] Jay A. Bertrand, Eric Fanchon, Lydie Martin, Laurent Chantalat, Genevi...ve Auger, Didier Blanot, Jean van Heijenoort, and Otto Dideberg. Open structures of MurD: domain movements and structural similarities with folylpolyglutamate synthetase. *J. Mol. Biol.*, 301(5):1257 – 1266, 2000.

[13] Andrej Perdih, Miha Kotnik, Milan Hodoscek, and Tom Solmajer. Targeted molecular dynamics simulation studies of binding and conformational changes in *E. coli* MurD. *Proteins: Structure, Function, and Bioinformatics*, 68(1):243–254, 2007.

[14] Courtney L. White, Aleksandar Kitich, and James W. Gober. Positioning cell wall synthetic complexes by the bacterial morphogenetic proteins MreB and MreD. *Mol. Microbiol.*, 76(3):616–633, 2010.

[15] Ahmed Gaballah, Anna Kloeckner, Christian Otten, Hans-Georg Sahl, and Beate Henrichfreise. Functional analysis of the cytoskeleton protein MreB from *Chlamydophila pneumoniae*. *PLoS ONE*, 6(10):e25129, 10 2011.

[16] Tamimount Mohammadi, Aneta Karczmarek, Muriel Crouvoisier, Ahmed Bouhss, Dominique Mengin-Lecreulx, and Tanneke Den Blaauwen. The essential peptidoglycan glycosyltransferase MurG forms a complex with proteins involved in lateral envelope growth as well as with proteins involved in cell division in *Escherichia coli*. *Mol. Microbiol.*, 65(4):1106–1121, 2007.

[17] Glen Spraggon, Robert Schwarzenbacher, Andreas Kreusch, Christian C. Lee, Polat Abdubek, Eileen Ambing, Tanya Biorac, Linda S. Brinen, Jaume M. Canaves, Jamison Cambell, Hsiu-Ju Chiu, Xiaoping Dai, Ashley M. Deacon, Mike DiDonato, Marc-André Elsliger, Said Eshagi, Ross Floyd, Adam Godzik, Carina Grittini, Slawomir K. Grzechnik, Eric Hampton, Lukasz Jaroszewski, Cathy Karlak, Heath E. Klock, Eric Koesema, John S. Kovarik, Peter Kuhn, Inna Levin, Daniel McMullan, Timothy M. McPhillips, Mitchell D. Miller, Andrew Morse, Kin Moy, Jie Ouyang, Rebecca Page, Kevin Quijano, Alyssa Robb, Raymond C. Stevens, Henry van den Bedem, Jeff Velasquez, Juli Vincent, Frank von Delft, Xianhong Wang, Bill West, Guenter Wolf, Qingping Xu, Keith O. Hodgson, John Wooley, Scott A. Lesley, and Ian A. Wilson. Crystal structure of an UDP-*N*-acetylmuramate-alanine ligase MurC (TM0231) from *Thermotoga maritima* at 2.3 Å resolution. *Proteins: Structure, Function, and Bioinformatics*, 55(4):1078–1081, 2004.

[18] Haiyong Jin, John J. Emanuele, Robert Fairman, James G. Robertson, Mark E. Hail, Hsu-Tso Ho, Paul J. Falk, and Joseph J. Villafranca. Structural studies of *Escherichia coli* udp-n-acetylmuramate:l-alanine ligase. *Biochemistry*, 35(5):1423–1431, 1996.

[19] Camille Filip, Gail Fletcher, Judith L. Wulff, and C. F. Earhart. Solubilization of the cytoplasmic membrane of *Escherichia coli* by the ionic detergent sodium-lauryl sarcosinate. *J Bacteriol.*, 115(3):71722, 1973.

[20] SM Singh and AK. Panda. Solubilization and refolding of bacterial inclusion body proteins. *J Biosci Bioeng.*, 99(4):303–10, 2005.

[21] H Tao, W Liu, BN Simmons, HK Harris, TC Cox, and MA Massiah. Purifying natively folded proteins from inclusion bodies using sarkosyl, Triton X-100, and CHAPS. *Biotechniques.*, 48(1):61–4, 2010.

[22] N Kumar and N Kishore. Structure and effect of sarcosine on water and urea by using molecular dynamics simulations: Implications in protein stabilization. *Biophys Chem.*, 171:9–15, 2013.

[23] XA Yang, XY Dong, Y Li, YD Wang, and Chen WF. Purification and refolding of a novel cancer/testis antigen BJ-HCC-2 expressed in the inclusion bodies of *Escherichia coli*. *Protein Expr Purif.*, 33(2):332–8, 2004.

[24] VG Francis, MA Majeed, and SN. Gummadi. Recovery of functionally active recombinant human phospholipid scramblase 1 from inclusion bodies using N-lauroyl sarcosine. *J Ind Microbiol Biotechnol.*, 39(7):1041–8, 2012.

[25] S Frankel, R Sohn, and L. Leinwand. The use of sarkosyl in generating soluble protein after bacterial expression. *Proc Natl Acad Sci U S A.*, 15;88(4):1192–6, 1991.

[26] Sha Ha, Deborah Walker, Yigong Shi, and Suzanne Walker. The 1.9 Å crystal structure of *Escherichia coli* MurG, a membrane-associated glycosyltransferase involved in peptidoglycan biosynthesis. *Protein Sci.*, 9(6):1045–1052, 2000.

[27] Yanan Hu, Lan Chen, Sha Ha, Ben Gross, Brian Falcone, Deborah Walker, Maryam Mokhtarzadeh, and Suzanne Walker. Crystal structure of the MurG:UDP-GlcNAc complex reveals common structural principles of a superfamily of glycosyltransferases. *Proc Natl Acad Sci U S A.*, 100(3):845–849, 2003.

[28] K Kaseda, H Yokota, Y Ishii, T Yanagida, T Inoue, K Fukui, and Kodama T. Single-molecule imaging of interaction between dextran and glucosyltransferase from *Streptococcus sobrinus*. *J Bacteriol.*, 182(4):1162–6, 2000.

[29] Jeanne Salje, Fusinita van den Ent, Piet de Boer, and Jan Löwe. Direct membrane binding by bacterial actin MreB. *Mol. Cell*, 43(3):478–487, 2011.

Appendices

Appendix A

Vectors and strains

A.1 Vectors

A.1.1 pCRBlunt

The pCRBlunt (pCR™-Blunt II-TOPO®) vector was used as means of storing the amplified genes.

The Zero Blunt TOPO® PCR Cloning technique from Invitrogen relies on Topoisomerase I from *Vaccinia* virus. This enzyme binds the plasmid vector at specific sites, cleaves the phosphodiester backbone in one strand, and forms a covalent bond between the 3' phosphate of the cleaved strand and a tyrosyl residue. Thus, the plasmid vector pCRBlunt is linearized with *Vaccinia* virus DNA topoisomerase I covalently bound to the 3' end of each DNA strand. Interestingly, the cleavage/binding reaction can be reversed by attack of the phospho-tyrosyl bond by the 5' hydroxyl of the original cleaved strand, or by a blunt PCR product, resulting in a pCR-Blunt vector containing the gene of interest.

Moreover, the vector contains the lethal *E. coli ccdB* gene [1] fused to the C-terminal side of the *lacZ*α fragment of the β-galactosidase enzyme[1]. In empty vectors, the resulting LacZ-Ccdb fusion protein retains both the CcdB killer activity and the ability to α-complement the truncated LacZ. The cells therefore show no β-galactosidase activity and are not viable upon plating. Ligation of a blunt-end PCR product breaks expression of the *lacZ-ccdB* gene fusion, permitting growth of only positive recombinants upon transformation, and restoring β-galactosidase activity.

The pCRBlunt vector holds the kanamycine resistance gene.

A.1.2 The pET system

We chose pET vectors as expression systems for *T. maritima mreB1*, *murD*, *murE*, and *murF*.

These plasmids rely on the pET System from Novagen, where the target genes are cloned after the T7 promoter DNA sequence from T7 bacteriophage, the lac operator DNA sequence, and the ribosome binding site (see FIGURE).

[1] LacZα and LacZω are the two peptides making up the β-galactosidase enzyme. Neither of these peptide is active by itself. Thus, β-galactosidase activity requires the presence of both peptides which spontaneously reassemble into a functional enzyme [2].

pET vectors also contain the *lacI* gene from the lac operon that codes for the lac repressor protein LacI. When bound to the operator region of the lac operon, LacI blocks RNA polymerase from binding, and so prevents transcription of the target gene [3].

When Isopropyl β-D-1-thiogalactopyranoside (IPTG)[2] is added to the culture medium, it binds to the lac repressor LacI causing an allosteric change in its shape which presents it from binding its operator, thereby allowing the transcription of the foreign gene. This effect is referred to as induction, because it induces expression of the target gene [3].

Howerer, the *E. coli* RNA polymerase can not recognize the T7 promoter in front of the target gene. Thus, to be expressed, pET vectors must be inserted into *E. coli* strains that have been modified to carry a gene coding for T7 RNA polymerase. In order to limit expression leakage when no inducer is present, T7 DNA polymerase expression is controlled by a modified lac operon system (referred to as DE3; see below). Indeed, instead of a T7 promoter sequence in front of the lac operator sequence, there is a lac promoter sequence that native *E. coli* RNA polymerase is able to bind.

Thereupon, the basal state of expression of the foreign gene is off, and the effect of adding the inducer is to abolish the repression by LacI, allowing both the transcription and translation of the T7 RNA polymerase and the subsequent transcription of the target gene by the T7 DNA polymerase [4] [3].

Both pET15b (N-terminal His-tag, ampicillin resistance) and pET30b (C-terminal His-tag, kanamycin resistance) were used in this work.

A.1.3 Modified pETDuet vectors

In pETDuet vectors, there are two separate cloning regions, both under T7 promoters. This provides an easy way for the co-expression of two proteins for detection and purification of protein complexes. Originally, this vector allows the expression of an untagged protein from the first cloning site, and a His-tagged protein from the second cloning site.

pETDuet vector with a Tev-cleavage site

Viviana Job (Bacterial Pathogenesis Group, IBS) added a Tev-cleavage site in order to allow removal of the His-tag from the protein expressed in the second cloning site, resulting in the pETDuetTev construct.

pETDuet vector with a Tev-cleavable C-terminal Strep-tag

André Zapun (Pneumococcus Group, IBS) modified the pETDuet vector to allow the expression of an untagged protein in the first site, and a Tev-cleavable Strep-tagged protein at the C-terminus from the second cloning site, resulting in the pETDuetLIM2.

[2]IPTG resembles allolactose which is responsible for triggering transcription of the lac operon

Sequential cloning in modified pETDuet vectors modified by André Zapun

André Zapun modified the pETDuet vectors for the expression of an untagged protein in the first cloning site and a Tev-cleavable N-terminal His-tagged (8 histidines) protein in the second cloning site. Interestingly, this construct allows sequential cloning of polycistronic genes into the second site, resulting in the pETDuetLIM1.

Then, if a *gene1* is amplified with primers such that the PCR product is NcoI-*gene1*-SpeI-BamHI, its insertion into NcoI/BamHI restriction sites results in a construct with the NcoI-*gene1*-SpeI-BamHI sequence. Then, PCR product XbaI-*gene2*-SpeIBamHI digested with XbaI and BamHI can be ligated into the SpeI / BamHI sites as XbaI and SpeI sites can be complemented, resulting into NcoI-*gene1*-XbaI-*gene2*-SpeIBamHI sequence. More genes can be sequentially added following this procedure.

A.1.4 pASK-IBA3C vector

We used the pASK-IBA3C expression vector from IBA for cloning of each of *murC* and *murG* genes. This vector bears the chloramphenicol resistance gene, opening up the possibility to co-transform it with pET vectors (ampicillin or kanamycin resistances).

The induction system of pASK-IBA vectors uses the tet repressor to turn on or off gene expression upon the presence of anhydrotetracyclin[3] [5]. There, the target gene is cloned after the tet operator (tetO) DNA sequence. The *tet* repressor is encoded on pASK-IBA plasmids as well, under the control of an independent promoter and then is constitutively expressed as TetR protein. The interaction between the repressor protein TetR and the tet operator (tetO) DNA sequence, represses the activity of the tet promoter, and thereby prevents transcription of the foreign gene. Transcription is turned on when anhydrotetracycline - a non-antibacterial homolog of tetracyclin, binds to TetR and causes a conformational change that prevents TetR from binding the operator. The activity of the promoter is thus restored and the target gene is transcribed. Thereby, the introduction of anhydrotetracyclin (AHT) to the system initiates the transcription of the genetic product [6].

Interestingly, this system does not require any specific *E. coli* strain or extra plasmid to work.

Previously to *murC* and *murG* cloning, pASK-IBA-3C vector has been modified to allow cleavage of the Strep tag by the Tev protease (see below).

A.2 Expression strains

A.2.1 The BL21(DE3) strain

Expression of pET-cloned genes requires specific strains which carry the T7 polymerase gene such as BL21(DE3), the most commonly used strain for expression of proteins encoded within such vectors (originally developed by Brookhaven National Laboratories). This *E. coli* B strain [7] [8] was engineered with the following genotype:

[3] In tetracyclin resistant bacteria, the tet promoter triggers the expression of TetA, the protein that pumps tetracycline antibiotic out of the cell and its respressor, TetR.

- F^-: does not carry the F-plasmid[4] so BL21(DE3) bacteria cannot transfer genes to another bacterium
- ompT: deficient in the Outer Membrane Protease OmpT[5], due to a mutation in the *ompT* gene. As a result, proteolysis of expressed and secreted proteins is reduced.
- gal: cannot use galactose as a carbon source.
- dcm: mutation in the *dcm* gene which deletes the capability of *E. coli* to methylate the second cytosine in CCWGG DNA sequences.
- lon: the *lon* gene encoding for the Lon serine protease has been deleted. This enzyme degrades misfolded proteins, but also some normal proteins which have a transient function into bacteria. Therefore, *lon* is regarded as a house-keeping gene. Nevertheless, it may degrade over-expressed proteins.
- $\text{hsdS}_B(r_B^-\text{-}m_B^-)$: mutations in the *S* gene from the restriction-modification system located at the *hsd* locus. Normal *E. coli* B strains methylate their DNA at B sites (TGA(N)8TGCT), and DNA that is not methylated at these sites is degraded by a restriction enzyme. With this mutation, BL21 cannot methylate, nor digest unmethylated DNA at B sites. Thereby, BL21(DE3) cells cannot degrade DNA plasmids.
- λ(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5]): BL21(DE3) cells have the T7 RNA polymerase gene from the λ phage under the control of the lacUV5 promoter (a mutant lac promoter that is stronger than wild type lac promoter); this comes from the λDE3 phage genome.

A.2.2 Other expression strains

For expression of the genes of interest, we used the following strains:

- C41(DE3) cells are derived from BL21(DE3) cells, by selection upon over-expression of a membrane protein. Thereby, these cells are more resistant to over-expression of membrane or membrane-bound proteins [11].
- BL21(DE3)pLysS have been transformed with the chloramphenicol-resistant pLysS plasmid that encodes T7 phage lysozyme, a natural inhibitor of T7 polymerase. Therefore, the basal activity of the polymerase is strongly reduced in non-induced cells, allowing relatively toxic genes to be established in BL21pLys cells.
- BL21-CodonPlus(DE3)-RIL cells contain extra copies of the *argU*, *ileY*, and *leuW* tRNA genes in a plasmid bearing chloramphenicol resistance. These genes encode tRNAs that recognize the arginine codons AGA and AGG, the isoleucine codon AUA, and the leucine codon CUA, respectively. All these codons are rare codons in *E. coli* and are the most frequently responsible for translation restriction of heterologous proteins.
- Rosetta(DE3) are derived from BL21(DE3). They contain the pLysSRARE plamid (chloramphenicol resistant) which, in addition to the pLysS plasmid, carries additional

[4]The F-plasmid is an episome (a plasmid that can integrate itself into the bacterial chromosome by homologous recombination) which allows genes to be transferred from one bacterium carrying the factor to another bacterium lacking the factor by conjugation [9].

[5]Most bacteria have proteases on their surfaces in order to degrade extracellular proteins that can be a source of amino acids, such as OmpT. OmpT is the one that causes the most trouble in expression of recombinant proteins [10].

tRNA genes to enhance the transcription of the *E. coli* rare codons AGG, AGA, AUA, CUA, CCC, and GGA.

- BL21 Star™(DE3) are derived from BL21(DE3) and carry a mutated *rne* gene which encodes a truncated RNase E enzyme that lacks the ability to degrade mRNA. This results in an increase in mRNA stability.
- Origami(DE3) are K-12 derivatives that have mutations in both the thioredoxin reductase (trxB) and glutathione reductase (gor) genes, which greatly enhance disulfide bond formation in the cytoplasm.

Appendix B

Mass spectrometry

This technique measures the mass-to-charge ratio m/z of ionized molecules in the gas-phase and allows the determination of the exact masses of the molecules within a protein sample.

B.1 Ionization techniques: MALDI and ESI

Ionization of proteins can be obtained by either MALDI (Matrix-Assisted Laser Desorption/Ionisation) or ESI (ElectroSpray Ionisation) techniques. These are soft ionization techniques which allow the production of large ions in the gas phase with no fragmentation for protein samples. A short introduction to these techniques is given here, based on a review of Shibdas Banerjee and Shyamalava Mazumdar published in 2012 [12], a review of Hanno Steen and Matthias Mann published in 2004 [13], and two books: one written by Gary Siuzdak [14], and the second one written by Edmond de Hoffmann and Vincent Stroobant [15].

B.1.1 MALDI

In the MALDI technique, laser pulses are directed onto a dry protein sample co-crystallized with a solid matrix constituted of small organic molecules.

The irradiation by the laser beam induces rapid heating of the matrix components, thus causing localized sublimation of the matrix, and entraining intact protein molecules in the gas phase, mainly as mono-protomated ions.

Ionization reactions may occur at any time during this process, though the exact mechanism of ions production is still under controversy. Nevertheless, gas-phase proton transfer in the expanding plasma between photo-ionized matrix molecules and protein molecules is the most widely accepted ion formation mechanism.

Therefore, the laser pulses accomplish both vaporization and ionization of the sample. Moreover, the matrix overcomes the propensity of macromolecules to fragment when ionized by absorbing most of the incident energy.

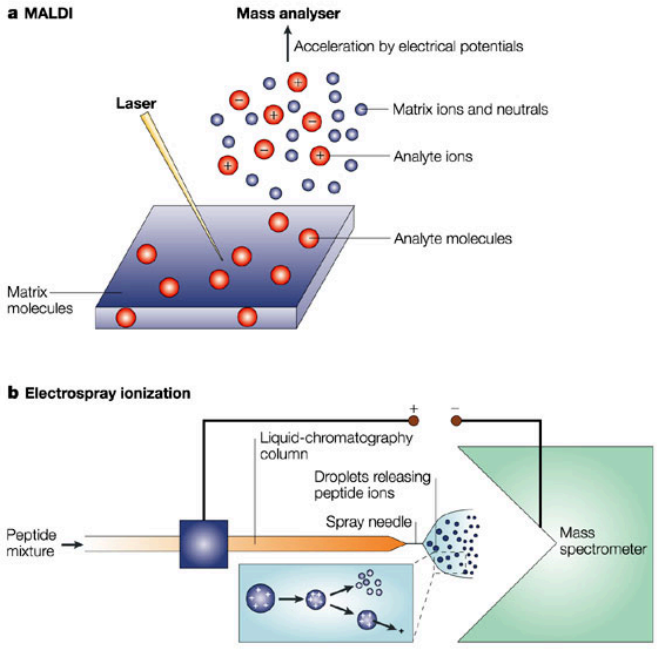


Figure B.1: Ionization techniques. a: MALDI and b: ESI. From Steen and Mann [13].

B.1.2 ESI

In the ESI method, the liquid containing the analyte is dispersed by electrospray into a fine aerosol, in order to produce charged proteins in the gas phase. The electrospray is generated by applying a strong electric field to the liquid protein sample passing through a capillary tube with a weak flux. As a result, the analyte is dispersed as multiply charged ions.

The mechanism of electrospray ionization can be described into three main steps (see Figure B.2), as described here below.

Production of charged droplets.
The protein solution is slowly pumped through the capillary. The high-voltage field applied to the capillary originates electrochemical reactions of solvent molecules, resulting in an electron flow.

Progressively, charges accumulate at the drop-like end of the protein solution coming out from the capillary tip, causing an elongation of the drop. When the surface tension is broken because of Coulomb forces, the shape of the drop changes to a 'Taylor cone'. This zone of high turbulence gives rise to a desintegration of the drop, forming smaller drops driven away from eath other by Coulombic repulsion. Therefore, the electrospray is formed.

Desintegration of the charged droplets into very small and highly charged droplets.

The solvent contained in the formed droplets evaporates, making them shrink till the point,

known as Rayleigh limit[1], where the surface tension can no longer sustain the Coulomb force of repulsion. Therefore, Coulomb fission occurs: the parent droplet disintegrates into much smaller offspring droplets.

The new droplets undergo desolvation too, repeating the process. A cascade of Coulomb fissions, yields increasingly small and charged droplets.

Gas-phase ion formation

The mechanism by which the gas-phase is formed from micro-droplets remains unclear. Nevertheless, it seems that large molecules such as proteins follow the Charge Residue Model.

In this model, the series of solvent evaporation and Coulomb fission results in an extremely small charged droplet which contains only one analyte molecule. The gas-phase protein ions would form after the remaining solvent molecules evaporate.

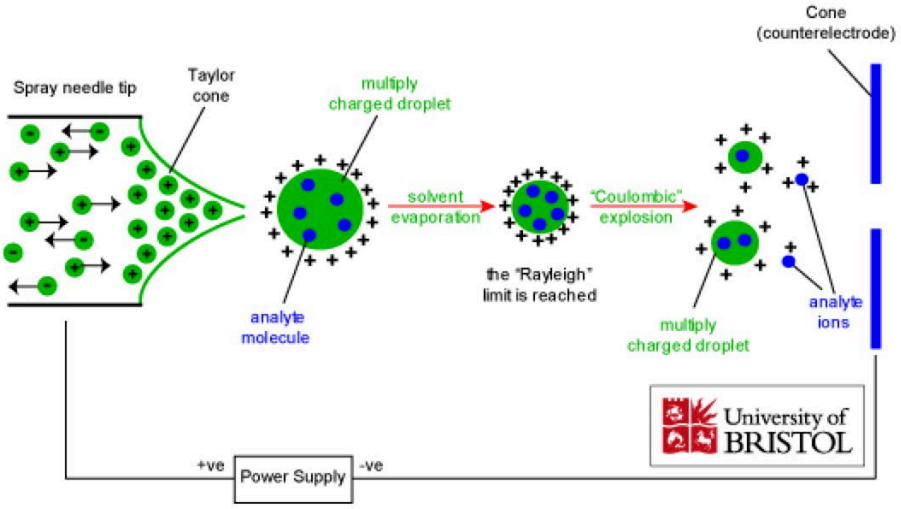


Figure B.2: A schematic of the ESI technique. From http://www.chm.bris.ac.uk/ms/theory/esi-ionisation.html

In contrast to MALDI, ESI may produce multiply charged ions from macromolecules, extending the mass range of the analyzer. As for MALDI, the ESI technique minimizes sample fragmentation during ionization since very little residual energy is retained by the analyte.

B.2 Time-of-flight (TOF) mass analyzer

Once the molecules in the sample are ionized, they are directed by electrostatic lenses from the ionization source into the mass analyzer.

There, ions are first accelerated by an electric field. They acquire a potential energy, E_p:

$$E_\mathrm{p} = zU$$

[1]Breakdown of the droplets can occur before the limit given by the Rayleigh equation is reached, because the droplets are deformed by mechanical vibrations, thus reducing the repulsion necessary to breakdown the droplets

where z is the charge of the particle, and U is the electric potential difference According to the theorem of conservation of mechanical energy,

$$E_{\mathrm{p}} + E_{\mathrm{k}} = cste$$

where E_{k} is the kinetic energy.

The time-of-flight mass analyzer consists in a chamber under high vacuum that contains no electric fields. The ions drift through the analyzer with the kinetic energy obtained from the potential energy of the electric field.

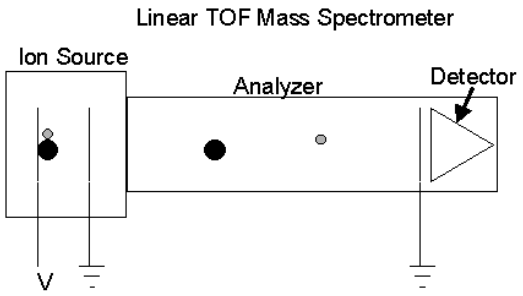


Figure B.3: Basic linear TOF mass spectrometer displaying two ions traversing the analyzer.
From NIH / NCRR Mass Spectrometry Resource Washington University in St. Louis

Therefore:

$$zU = \frac{1}{2}mv^2 \tag{B.1}$$

where m is the mass of the particle, and v its velocity [15].

If all ions obtained the same kinetic energy (same charge), the ions of smaller m/z (lower mass) will have higher velocity than ions of bigger m/z (higher mass). This results in separation in space of ions according to their m/z ratio as they flight along the analyzer (see Figure B.3). Thus, if one knows the distance of the flight of the ion, d, and one measures the time of the flight of the ion t, one can determine the m/z ratio.

Indeed:

$$v = \frac{d}{t}$$

So from B.1:

$$zU = \frac{1}{2}m\left(\frac{d}{t}\right)^2$$

giving

$$\boxed{t = \frac{d}{\sqrt{2U}}\sqrt{\frac{m}{z}}} \tag{B.2}$$

Therefore, the time of flight of the ion varies with the square root of its mass-to-charge ratio m/z.

A detector is positioned at the end of the analyzer and measures the arrival time of ions.

This allows to plot the abundance versus time, thus showing the arrival time distribution of the ions. Each of the peaks in the plot can be associated with different m/z values (see Figure B.4). Thus, the exact masses of molecules can be determined.

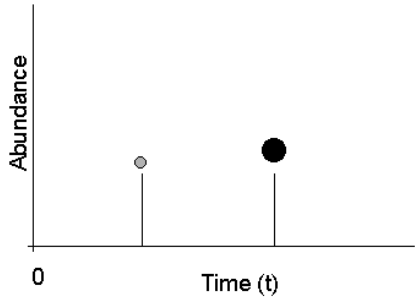


Figure B.4: Plot of abundance versus time for ions detected in TOF mass spectrometer. From NIH / NCRR Mass Spectrometry Resource Washington University in St. Louis

Appendix C

Publication

environmental microbiology

Environmental Microbiology (2013) doi:10.1111/1462-2920.12171

MreB and MurG as scaffolds for the cytoplasmic steps of peptidoglycan biosynthesis

Sandy Favini-Stabile,[1,2,3] Carlos Contreras-Martel,[1,2,3] Nicole Thielens[1,2,3] and Andréa Dessen[1,2,3,4]*

[1]*Institut de Biologie Structurale (IBS), Université Grenoble I, Grenoble, France.*

[2]*Commissariat à l'Energie Atomique (CEA), Grenoble, France.*

[3]*Centre National de la Recherche Scientifique (CNRS), Grenoble, France.*

[4]*Brazilian National Laboratory for Biosciences (LNBio), CNPEM, Campinas, São Paulo, Brazil.*

Summary

Peptidoglycan is a major determinant of cell shape in bacteria, and its biosynthesis involves the concerted action of cytoplasmic, membrane-associated and periplasmic enzymes. Within the cytoplasm, Mur enzymes catalyse the first steps leading to peptidoglycan precursor biosynthesis, and have been suggested as being part of a multicomponent complex that could also involve the transglycosylase MurG and the cytoskeletal protein MreB. In order to initialize the characterization of a potential Mur interaction network, we purified MurD, MurE, MurF, MurG and MreB from *Thermotoga maritima* and characterized their interactions using membrane blotting and surface plasmon resonance. MurD, MurE and MurF all recognize MurG and MreB, but not each other, while the two latter proteins interact. In addition, we solved the crystal structures of MurD, MurE and MurF, which indicate that their C-termini display high conformational flexibilities. The differences in Mur conformations could be important parameters for the stability of an intracytoplasmic murein biosynthesis complex.

Introduction

The bacterial cell wall is a complex structure that plays key roles in the bacterial life cycle, protecting bacteria from osmotic stress and ensuring cell shape. One of its main components is peptidoglycan, a mesh-like polymer formed by polymerized *N*-acetyl-glucosamine (GlcNAc) and *N*-acetyl-muramic acid (MurNAc) chains and cross-linked by short peptides (Vollmer *et al*., 2008). The peptidoglycan biosynthetic pathway has been of prime interest for the development of antibiotics for decades, but to date, only a limited number of proteins have proven to be useful targets towards the generation of antimicrobials. Nevertheless, because of the essential nature of many of its enzymes and its absence in mammals, the peptidoglycan formation process continues to be of keen interest for the development of novel antibacterial drugs and a further understanding of its functional, organizational, and structural intricacies could provide leading information towards this objective.

Peptidoglycan biosynthesis is a multistep process that occurs in different cellular compartments (cytoplasm, inner membrane and periplasm), and involves a number of mostly sequential enzymes. Cytoplasmic steps consist in the synthesis of the monomer building block of peptidoglycan, lipid II, from UDP-GlcNAc through the sequential action of Mur enzymes (Bouhss *et al*., 2008). Subsequently, lipid II is translocated by enzymes of the shape, elongation, division and sporulation family to the periplasmic side of the membrane (Mohammadi *et al*., 2011), where it is incorporated into the growing peptidoglycan layer through transglycosylation and transpeptidation reactions catalysed by Penicillin-Binding Proteins (Matteï *et al*., 2010; Lovering *et al*., 2012). While most current antibiotics target the periplasmic stage of peptidoglycan biosynthesis, the cytoplasmic steps, and in particular Mur enzymes, are still underexploited despite the fact that most of them are essential and specific to eubacteria. Although a number of inhibitors of Mur enzymes have been identified, few have presented promising antibacterial activity (Silver, 2006; Kotnik *et al*., 2007; Barreteau *et al*., 2008; Humljan *et al*., 2008; Chopra, 2013), necessitating a more precise knowledge of the mechanism of action of Mur enzymes within the peptidoglycan synthesis machinery.

Among the cytoplasmic actors of peptidoglycan biosynthesis, the MurC, MurD, MurE and MurF enzymes are adenosine triphosphate (ATP)-dependent ligases that catalyse the sequential addition of amino acids to the UDP-MurNAc precursor in order to generate a

Received 9 February, 2013; accepted 27 May, 2013. *For correspondence. E-mail andrea.dessen@ibs.fr; Tel. 00 (+33) 4 38 78 95 90; Fax 00 (+33) 4 38 78 54 94.

short peptide which will be cross-linked within the peptidoglycan layer (Barreteau *et al*., 2008). Following the synthesis of UDP-MurNAc by the MurA and MurB enzymes from UDP-GlcNAc, the MurC ligase adds an L-alanine to form UDP-*N*-acetylmuramoyl-L-alanine (UMA). Subsequently, MurD adds a D-glutamate moiety to UMA, producing UDP-*N*-acetylmuramoyl-L-alanine-D-glutamate (Bertrand *et al*., 1997), which is then used as substrate by MurE for the binding of a diamino acid, i.e. *meso*-diaminopimelate [or L-lysine, depending on the species (Patin *et al*., 2010)]. The last enzyme in this ligation pathway, MurF, catalyses the addition of a D-ala-D-ala dipeptide to the UDP-Mur*N*Ac-tripeptide to form UDP-Mur*N*Ac-pentapeptide (Mengin-Lecreulx *et al*., 1996; Patin *et al*., 2010). Following these cytosolic steps, two enzymes located on the cytoplasmic side of the bacterial membrane, MraY and MurG, successively catalyse the transfer of the phospho-MurNAc-pentapeptide group of UDP-MurNAc-pentapeptide to a lipid carrier, generating lipid I, and then of a GlcNAc group to lipid I, giving rise to lipid II (Bouhss *et al*., 2004; 2008; Mohammadi *et al*., 2007).

A number of actors involved in peptidoglycan biosynthesis have been shown to interact with each other. Among them, MreB has been widely reported to play a key role in the spatial organization of the overall process in rod-shaped bacteria. This actin-like protein involved in cell-shape maintenance is thought to restrain the diffusion of peptidoglycan components within the cell membrane in order to achieve more efficient incorporation of newly synthesized peptidoglycan building blocks (den Blaauwen *et al*., 2008; White and Gober, 2012). MreB was hypothesized in different organisms to interact with proteins involved in the membrane-associated and periplasmic stages of peptidoglycan biosynthesis, such as MreC (Kruse *et al*., 2005; Divakaruni *et al*., 2007; White *et al*., 2010), MraY, MurG (Mohammadi *et al*., 2007; White *et al*., 2010) and RodZ (van den Ent *et al*., 2010). In contrast, little is known about the role of this cytoskeletal protein in the cytoplasmic steps. However, it is of interest that in *Caulobacter crescentus*, MreB was shown to be required for the proper spatial positioning of several cytosolic murein biosynthetic enzymes including MurB, MurC, MurE and MurF, all of which adopted a subcellular pattern of localization similar to MurG (White *et al*., 2010). These observations hint to the existence of a cytoskeletal-dependent cytoplasmic complex of murein enzymes (Silver, 2006; Matteï *et al*., 2010; White *et al*., 2010; Lovering *et al*., 2012). To date, however, few studies have tackled the question of spatial organization of the cytoplasmic steps of peptidoglycan biosynthesis.

In order to investigate potential interactions that could support the existence of such a complex, we purified and characterized the ligases MurD, MurE, MurF and the glycosyltransferase MurG, as well as MreB, from the Gram-negative bacterium *Thermotoga maritima*. Surface plasmon resonance (SPR) spectroscopy and dot blot interaction experiments provided evidence that MurD, MurE and MurF all recognize and bind to MreB and MurG, but not to each other, while MreB and MurG were shown to interact. In addition, we solved the crystal structures of MurD, MurE and MurF, the latter two in complex with ADP. The structures reveal that the C-terminal domains of the molecules may adopt different orientations, and support the hypothesis that Mur ligases display distinct conformations upon ligand recognition. This fact could play an important role in MreB and MurG recognition and binding. These results provide an initial framework for understanding the structural requirements of a cytosolic complex of murein-synthesizing enzymes involving the cytoskeletal protein MreB.

Results

Mur enzymes interact with MreB

MurD, MurE, MurF and MreB from *T. maritima* were all overexpressed in *Escherichia coli* cells as His-tagged proteins, whereas MurG was fused to a Strep tag. We chose to study enzymes from *T. maritima*, a rod-shaped, thermophilic organism, in order to increase our chances of characterizing a cytoplasmic murein-synthesizing complex that could potentially be more stable than one studied from a non-thermophilic bacterium. All proteins were purified by affinity and size exclusion chromatographies. While MurD, MurE and MurF were highly soluble, MreB solubility was more limited and MurG solubilization required the employment of N-lauroyl-sarcosine as well as the addition of 0.04% N-dodecyl-β-D-maltopyranoside (DDM) in all buffers in order to prevent aggregation.

We initially attempted to identify which were the preferential interaction partners within a Mur network. At first, the ability of purified Mur ligases to interact with each other was tested using pull-down affinity assays and gel filtration chromatography, but no interactions could be detected using these techniques for any combination of MurD, MurE or MurF. Thus, interactions were investigated using dot blot assays. In this test, increasing quantities of His-tagged MurD, MurE, MurF and MreB were spotted onto a nitrocellulose membrane and subsequently incubated with Strep-tagged MurG (Fig. 1A); the membrane was then treated with horseradish peroxidase (HRP) streptactin. The detection of signal following increasing amounts of loaded enzymes indicated that MurG readily bound both MurE and MurF. A weak signal could be detected for the MurG/MreB interaction. In the reverse experiment, where MurG-spotted membranes were incubated with MreB, MurD,

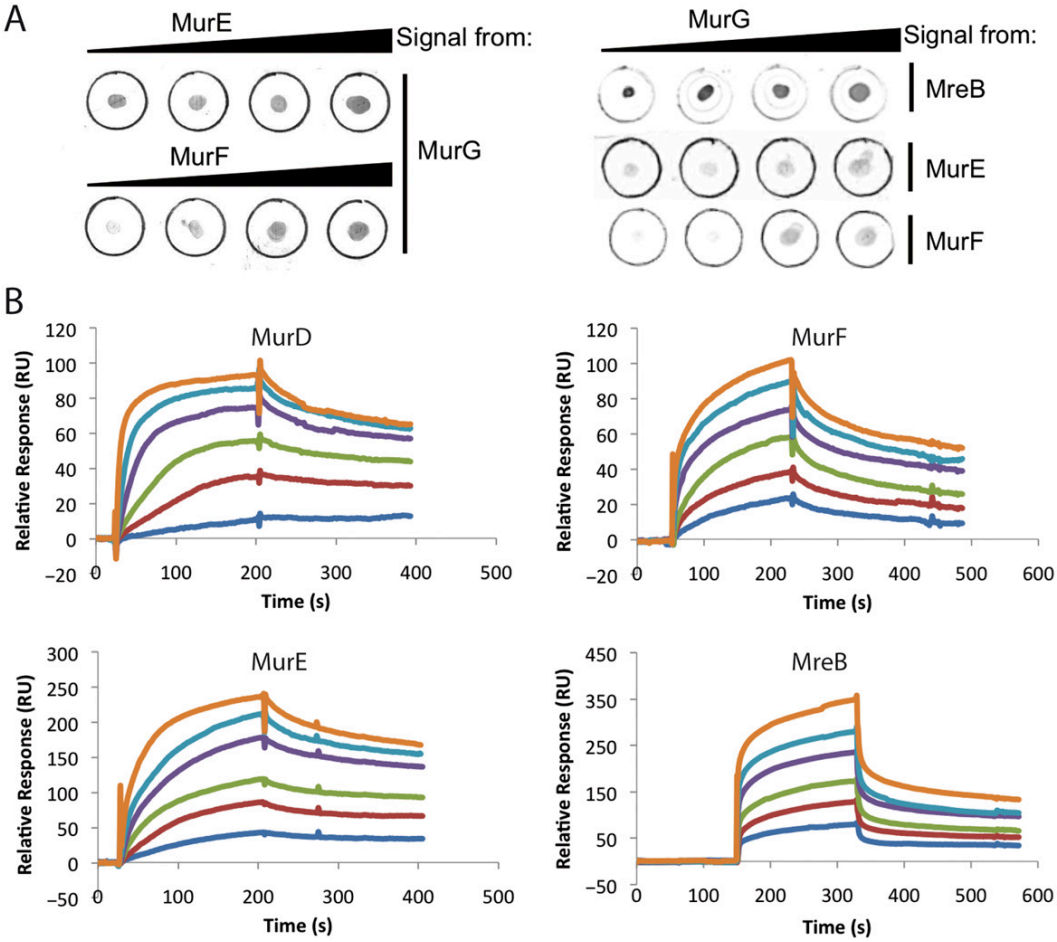


Fig. 1. MurG directly interacts with MurD, MurE, MurF and MreB.
A. Dot blot on purified proteins showing direct interaction between MurG and MreB, MurE, MurF. The nitrocellulose membranes were dotted with increasing amounts (see the bar) of His-tagged MreB, MurE and MurF or Strep-tagged MurG, as indicated. Membranes were incubated with 0.1 mg ml^{-1} of MurG, MreB, MurE or MurF as indicated on the right of each frame, subsequently incubated with HRP anti-his antibody or streptactin HRP for development.
B. Surface plasmon resonance assays of the interactions between MurG and MurD-F ligases, and between MurG and MreB. MurG was immobilized onto a CM5 sensor chip and MreB, MurD, MurE and MurF were tested as analytes in varying concentrations (from bottom to top): 0.25 μM, 0.5 μM, 1 μM, 2 μM, 4 μM, 8 μM.

MurE or MurF solutions, a clear signal appeared for the MurG/MreB, MurG/MurE and MurG/MurF pairs (Fig. 1A). No signal was detected for MurG/MurD in either of the experiments. A negative control (BSA) displayed no signal in either one of the experiments, providing initial evidence that MurG from *T. maritima* could interact directly with MreB, MurE and MurF.

To further characterize these interactions, we used SPR spectroscopy. Increasing concentrations of MurD, MurE, MurF and MreB, ranging from 0.25 μM to 8 μM, were injected over a CM5 sensor chip onto which MurG had been immobilized (Fig. 1B). The response was clearly dose-dependent, indicating that each of the three ligases, MurD included, as well as MreB, readily bound to the glycosyltransferase. Similarly, MreB was covalently linked to a CM5 gold chip and assessed for interaction with MurD, MurE, MurF, as analytes in increasing concentrations from 0.25 to 2.4 μM (Fig. 2). Resulting curves confirmed that MreB is an interaction partner of MurD, MurE and MurF. The same experiment performed using immobilized MurE, confirmed its interaction with MurG and MreB (not shown). It is of note that employment of BSA or other His tagged proteins as controls did not display an interaction signal with either the MurG or MreB surfaces (Fig. S1), indicating that these interactions are specific. No signal could be detected when MurD or MurE were tested over a MurF-immobilized surface (not shown), confirming our previous observations that Mur ligases do not

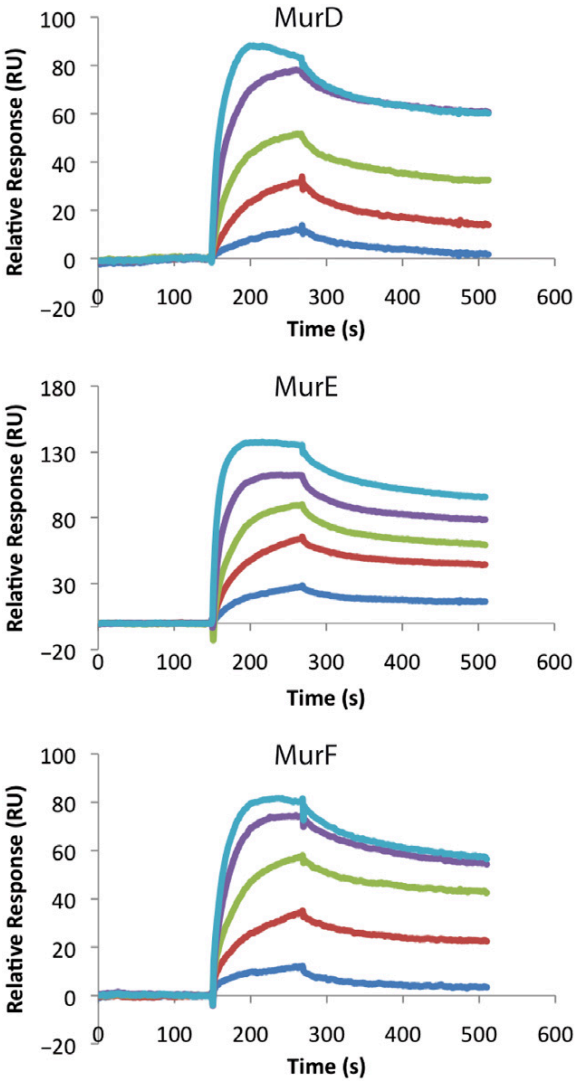


Fig. 2. SPR assays showing direct interaction between MreB and MurD, MurE and MurF. MreB was immobilized onto a CM5 sensor chip, and Mur ligases were injected as analytes. From bottom to top: 0.2 μM, 0.4 μM, 0.8 μM, 1.6 μM, 2.4 μM.

interact with each other, and that MurG and MreB are the main actors of the interaction network.

The kinetic analysis of MreB biosensorgrams revealed K_D constants in the range of 7–30 nM and χ^2 values around 1.0 (see Table 1). It is of note that these values were obtained with a two-state reaction model where a conformational change is considered in addition to a classical 1:1 binding mode. MurG-MurD and MurG-MurE binding curves, despite clearly indicating that there is interaction between the two protein pairs, could not be fitted using any of the models available in the BIA Evaluation software package (GE Healthcare Bio-Sciences AB, Uppsala, Sweden), indicating that binding may not follow a simple kinetic model. The MurG-MreB sensorgram indicates that no saturation point is reached at high MreB concentrations, suggesting that the cytoskeletal protein oligomerizes onto the MurG surface at high concentrations, as suggested by Gaballah and colleagues (2011).

Crystal structures of MurD, MurE and MurF from T. maritima

Our data thus indicated that MurD, MurE and MurF do not interact directly with each other, but can recognize both MreB and MurG. In order to analyse if there are structural features of the three Mur ligases that could shed light on these affinity profiles, we crystallized MurD, MurE and MurF, and solved all three crystal structures.

Initially, crystallization conditions for MurF were screened at the High Throughput Crystallization Laboratory [Partnership for Structural Biology (PSB), Grenoble]. Manually reproduced crystals diffracted X-rays to 1.65 Å resolution at the European Synchrotron Radiation Facility (ESRF) in Grenoble and were in space group $P2_12_12_1$, with 1 molecule per asymmetric unit. Despite significant effort, no molecular replacement solution could be obtained using the known structures of MurF homologues from *E. coli* (Yan *et al.*, 2000) or *Streptococcus pneumoniae* (Longenecker *et al.*, 2005). A selenomethionylated (SeMet) derivative, prepared and crystallized under similar conditions, allowed us to solve the structure of MurF to a resolution of 2.4 Å by performing a single wavelength anomalous dispersion (SAD) experiment on the selenium edge. The model was subsequently used to phase the higher resolution native data set by molecular replacement. Data collection, phasing and refinement statistics are included in Table S1.

Mur ligases from different bacteria display considerable structural similarity, and basically consist of a small

Table 1. Kinetic constants for the interactions of Mur ligases with MreB. Binding was measured as described in *Experimental procedures*. The association (ka1, ka2) and dissociation (kd1, kd2) rate constants were determined by global fitting of the data using a two-state binding model (BIAEvaluation, Biacore). The dissociation constants K_D were determined from the (kd1/ka1)*kd2/(ka2 + kd2) ratio.

Immobilized protein	Analyte	ka1 (M^{-1} s^{-1})	kd1 (s^{-1})	ka2 (s^{-1})	kd2 (s^{-1})	K_D	χ_2
MreB	MurD	$4.8\ 10^4$	$7.0\ 10^{-3}$	$8.7\ 10^{-3}$	$7.7\ 10^{-4}$	$1.2\ 10^{-8}$	1.05
MreB	MurE	$8.0\ 10^4$	$9.1\ 10^{-3}$	$8.9\ 10^{-3}$	$5.8\ 10^{-4}$	$7.4\ 10^{-9}$	0.54
MreB	MurF	$3.0\ 10^4$	$5.9\ 10^{-3}$	$9.5\ 10^{-3}$	$1.7\ 10^{-3}$	$2.9\ 10^{-8}$	0.63

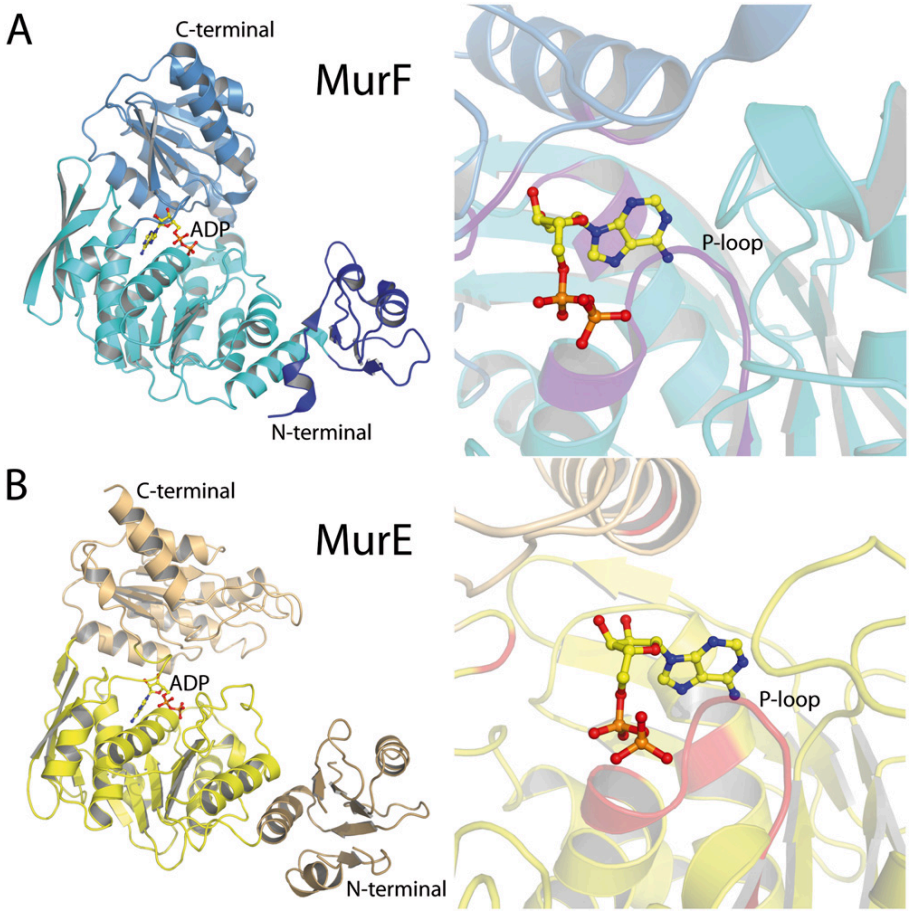


Fig. 3. Structures and ADP binding sites of MurE and MurF from *T. maritima*.
A. MurF in interaction with ADP (left) and zoom onto the ADP-binding site (right). The three different shades of blue correspond to the three open α/β-sheet domains of MurF. ADP-interacting residues are highlighted in purple, including the P-loop.
B. MurE in interaction with ADP (left) and zoom onto the ADP-binding site (right). The three different shades of yellow correspond to the three domains of MurE. ADP-interacting residues are highlighted in red, including the P-loop. ADP is depicted as sticks, with C atoms in yellow, N atoms in blue, O atoms in red, and P atoms in orange.

N-terminal domain that recognizes the growing peptidoglycan building block, a central domain that binds the nucleotide, and a C-terminal domain that binds the incoming amino acid (Smith, 2006). This is the case for MurF (Fig. 3A), our highest resolution structure, which is composed of three mixed α/β-sheet domains that, individually, are highly reminiscent of those from its *E. coli* and *S. pneumoniae* counterparts (Yan *et al*., 2000; Longenecker *et al*., 2005) despite the relatively low sequence identity (27%, 29%). The N-terminal domain of MurF from *T. maritima* (res 1–63) consists of a small α/β-fold that lacks the 20 first residues that are normally present in other MurF ligases. This is a unique feature of MurF – it lacks the characteristic Rossmann nucleotide-binding fold seen in the N-terminal domains of other Mur ligases (Lovering *et al*., 2012). The central domain (res 64–286) adopts a mononucleotide-binding fold seen in other ATP-binding enzymes, and consists of a central, 7-stranded β-sheet interwoven by 6 α-helices, backed up by a smaller, 3-stranded antiparallel β-sheet. This domain carries the characteristic P-loop, which plays a key role in ADP binding (Fig. 3A, right). Lastly, the C-terminal domain (res 287–427) adopts a dinucleotide (Rossmann)-binding fold, and is connected to the central domain by a short linker.

Mur enzymes have been reported to crystallize in 'open' and 'closed' conformations, depending on the identity of the bound ligand and involving rotations of the N- and C-terminal domains either 'towards' the central domain (closed form) or 'away' from it (open form) (Smith, 2006). A comparison of our ADP-liganded MurF structure to the open apo *E. coli* MurF structure, and an inhibitor-bound,

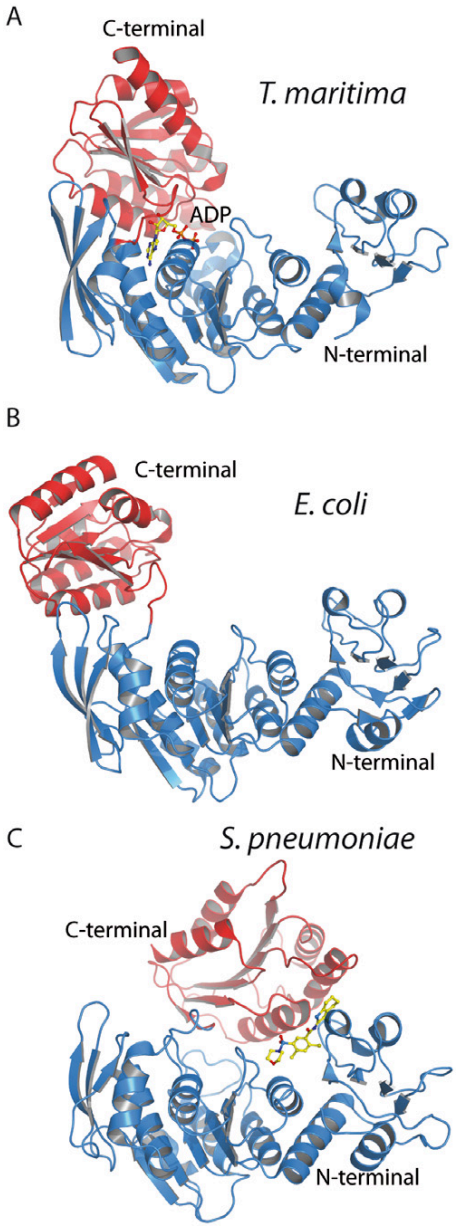


Fig. 4. Flexibility of the C-terminal domain in the known structures of MurF.
A. *T. maritima* MurF in complex with ADP.
B. Apo form of *E. coli* MurF (1GG4).
C. *S. pneumoniae* MurF in complex with an inhibitor (2AM1).
C-terminal domains are represented in red. ADP and inhibitor are depicted as sticks.

closed *S. pneumoniae* MurF structure (Longenecker *et al.*, 2005), reveals that the conformation of MurF from *T. maritima* in complex with ADP is in fact intermediary between that of the two other structures (Fig. 4). The C-terminal domain of MurF acts as a 'claw' that locks onto the ADP molecule, helping to stabilize it within a deep cleft. It is of note that the inhibitor-bound MurF structure from *S. pneumoniae* represents the most compact Mur structure published to date.

Insight into the conformational flexibility of Mur enzymes was further obtained when we solved the crystal structure of MurE from *T. maritima* to 2.9 Å by performing SAD experiments at the selenium edge at the ESRF followed by molecular replacement onto a higher resolution data set (Table S2). The structure of *T. maritima* MurE also reveals a three-domain molecule, in which the C-terminus is closed 'over' ADP, stabilizing it within the central domain (Fig. 3B). Notably, the structure of MurE reported here is highly reminiscent of that of the enzymes from *E. coli* and *Mycobacterium tuberculosis*, both crystallized in the presence of the reaction product UDP-MurNAc-L-Ala-D-Glu-*meso*-diaminopimelate (Gordon *et al.*, 2001; Basavannacharya *et al.*, 2010) (Fig. S2). Superposition of the three structures reveals root-mean-square deviation values between 1.6 and 2.2 Å over 347 Cα atoms. Notably, the C-terminal domains of all three molecules are positioned similarly, despite the fact that one is in complex with the nucleotide, and the others with the reaction product. Thus, in the case of MurE, ligand binding does not necessarily dictate the positioning of the C-terminus, suggesting that the preferential conformations of this enzyme may not be ligand-dependent.

In the structures of MurE and MurF from *T. maritima*, ADP binds within a cleft made by both the second and third domains employing a set of highly conserved residues, including a well-characterized P-loop (Figs 3 and S3). Strikingly, in addition to the common set of amino acids that provide side chains for interactions with ADP in all structures solved to date, in the structure of *T. maritima* MurF, the ribose forms hydrogen bonds with two additional residues, Asn 258 and Lys 405 (Fig. S3). In addition, Asn 310, which interacts with the phosphate group in MurF, has no counterpart in *E. coli* MurD. The higher number of residues involved in ADP binding may be linked to the thermophilic origin of the protein.

Figure 5 displays the structure of *T. maritima* MurD, which was solved to 2.17 Å by employing SAD techniques at the selenium edge at the SOLEIL synchrotron (Table S3). Although MurD was crystallized in the presence of AMP-PNP and UMA, no ligands could be identified in the structure, potentially due to the high concentration of phosphate in the crystallization solution. The enzyme displays the characteristic three-domain organization of its *E. coli* and *Staphylococcus aureus* counterparts (Bertrand *et al.*, 1997; 1999; 2000). Mapping of the crystallographic temperature (B) factors onto the structure model, as shown in Fig. 5, where dark blue indicates stable regions (low B factors) and lighter colours are indicative of flexibility (high B factors), shows that the

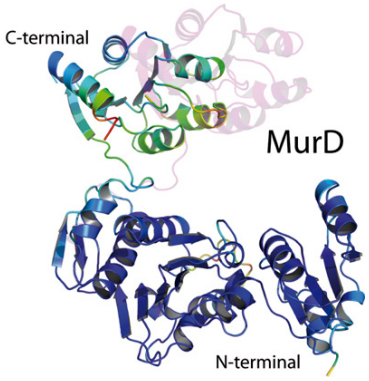


Fig. 5. Structure of MurD from *T. maritima*. Cartoon representation of chain A (colours correspond to B-factors; low B-factors: blue, high B-factors: red) and chain B (transparent, magenta).

C-terminal domain of this Mur enzyme is also its most conformationally flexible feature. This domain makes no contacts with its central region, and loop 308–315 that links the two domains is highly flexible and difficult to trace in the electron density map. Notably, this feature generates two different conformations of the domain in the same asymmetric unit, which are indicated in Fig. 5, where the two molecules were superposed, and the distinct conformation of the C-terminal domain is clearly visible. Nevertheless, the superposition of the MurD structures from *E. coli* and *T. maritima* reveals that the N-terminal and central domains are highly similar, while the C-terminal domain is in a completely different position, being located clearly much farther from the active site than the C-terminus in the ligand-bound structure (Fig. S3). These observations support the idea that there is increased flexibility of the apo forms of the enzymes in contrast to the substrate-bound forms, a role mostly played by the C-terminal domain of Mur ligases.

Discussion

The peptidoglycan biosynthetic steps that occur within the bacterial cytoplasm are crucial for the generation of a pool of precursors that will be essential both for the cell division and cell wall elongation processes. Mur ligases, which catalyse the consecutive reactions leading to formation of the stem peptide on UDP-MurNAc, have been suggested as being integral components of a cytoplasmic complex of murein-synthesizing enzymes whose function would be to 'channel' peptidoglycan building blocks towards the inner leaflet of the membrane while being spatially coordinated by the cytoskeletal protein MreB (White *et al*., 2010). The potential existence of such a complex is supported by evidence indicating that MreB interacts with membrane-associated MurG (Mohammadi *et al*., 2007; White *et al*., 2010; Gaballah *et al*., 2011) which, in addition to all Mur ligases, displays MreB-dependent localization within the cytoplasm (White *et al*., 2010). Interestingly, MreB was also shown to interact with MurF in a copelleting assay (Gaballah *et al*., 2011).

The participation of Mur ligases within a multi-component complex is an attractive hypothesis from different points of view. In the Mur ligase-catalysed reactions, the substrate of one reaction is the product of the previous one (Barreteau *et al*., 2008), and the spatial limitation would facilitate transfer of the product of one reaction directly onto the active site of the next Mur, thus potentially restricting diffusion of peptidoglycan building blocks throughout the cytoplasm (Lovering *et al*., 2012). In addition, Mur ligases have been the targets of drug development studies for decades, but to date, no compounds with promising antibacterial activity have been identified, even in the presence of membrane permeabilizers (Silver, 2006; Barreteau *et al*., 2008; Lovering *et al*., 2012; Chopra, 2013). This could be due to the fact that Mur ligase active sites are made inaccessible to diffusion of molecules through their involvement in a multi-enzyme complex. In order to shed light on the hypothesis of the potential existence of this complex, we set out to characterize interactions between Mur ligases D, E and F, and also MurG and MreB.

Because we were not able to detect direct interactions between MurD, MurE and MurF, we hypothesized that multipartite complex formation could require 'scaffolds', and tested if MurG and MreB could fulfil such a role, initially using dot blot assays. These results identified interactions between MurG and MurE, MurF, and MreB, but no signal could be detected for MurD interactions using this assay. Subsequent SPR experiments performed coating either MurG or MreB on a sensor chip showed dose-dependent curves for the interactions of the two proteins with all Mur ligases, including MurD. Despite the fact that all sensorgrams were clearly dose-dependent and reproducible (Figs 1 and 2), it is worth mentioning that often only a fraction the immobilized protein – be it MurG or MreB – interacted with the analyte, and best fits could only be obtained with a 'two conformation state' model from the BIAevaluation software. This diminished binding capacity could be potentially explained by different quaternary states adopted by MurG and MreB. MurG, as many other glycosyltransferases, has been reported to oligomerize *in vitro* (Ha *et al*., 1999), an observation that we confirmed during our own purifications and Biacore experiments (data not shown). MreB has also been suggested to undergo conformational changes upon nucleotide binding (Dye *et al*., 2011), and a number of studies have shown that its eukaryotic homologue actin can display remarkable conformational flexibility, in both monomeric and polymerized states, in response to interactions with partners (Hild *et al*., 2010). Thus, the

behaviours of both MurG and MreB themselves could potentially affect differential recognition of interacting surface on Mur ligases.

The crystal structures of all three Mur ligases, D, E and F, reveal that their three domains are highly reminiscent of those of other, previously characterized homologues. However, these structures highlight the fact that Mur ligases undergo considerable conformation changes, with the C-terminal domains being more or less free to adopt different conformations in relation to the other domains of the proteins; the C-terminus of MurD, for example, displays extreme flexibility, not making a single interaction with the central domain and being present in two very different conformations even within the same asymmetric unit. The fact that the C-terminal domain of MurD from *T. maritima* is not more stably associated to the rest of the protein could explain why its interaction with MreB/MurG could not be detected using a dot blot technique, which is less sensitive than SPR experiments.

Despite the fact that our structures reveal that the C-terminal domains of Mur ligases could be 'free' to test different conformations in three-dimensional (3D) space, it has been reported that catalysis by Murs requires that the enzymes adopt a 'closed' conformation, where the substrate binds to the N-terminal domain, the incoming amino acid to the C-terminal region, and ATP, the first ligand to bind, within the catalytic cleft between the C-terminal and central domains (Perdih *et al*., 2007). Our structures, analysed in the light of the structures of Mur ligases from other species, suggest that neither nucleotide nor substrate binding is sufficient to infer substantial closure; MurE and MurF, bound to ADP, present relatively open conformations. Thus, we postulate that rather than inducing significant conformational modifications upon binding, the incoming ligands simply limit the flexibility of the C-terminal domain.

The conformational sampling of 3D space by the C-terminal domains of Mur ligases in the absence of ligands, together with our SPR data indicating that both MreB and MurG are capable of interacting with Mur ligases under these conditions, suggest that within the bacterial cytoplasm, the two latter proteins could serve as 'scaffolds' to maintain Mur ligases assembled in a multicomponent entity even in their 'open', inactive forms. Because the concentration of ATP in the cytoplasm is in the millimolar range, it is probable that all Mur ligases exist within the complex in forms that display a 'conformationally stable' C-terminal domain. Further experimentation will be required to define if a complex between MreB/MurG and Mur ligases is further stabilized by substrate/nucleotide binding. The data presented here, combined with structural and cellular localization studies of Murs from other bacterial species, provide an initial framework towards the understanding of the interactions between a biosynthetic machinery that channels precursors towards the inner leaflet of the membrane.

Experimental procedures

Cloning and expression of MreB, MurD, MurE, MurF and MurG

The *mreB1* gene from *T. maritima* (DSM 3109) was cloned into a modified pET30b vector as described in van den Ent and colleagues (2001). The *murD* gene was cloned into pET15b. The *murE*, *murF* genes were cloned into pET30b and *murG* into a modified pASK-IBA3C vector that allows the expression of proteins with a Tev-cleavable C-terminal Strep tag. MreB, MurD, MurE, MurF and MurG were expressed in *E. coli* BL21(DE3) or C41(DE3) (MreB) overnight at 25°C and induced with either 1 mM isopropyl β-D-thiogalactopyranoside (IPTG) or 400 ng/ml anhydrotetracycline (MurG).

Purification of MurD, MurE and MurF from T. maritima

Cells were lysed by sonication in buffer A (25 mM HEPES pH 7.4, 150 mM NaCl, 25 mM imidazole) supplemented with protease inhibitors. After a centrifugation step at 18 000 r.p.m. for 45 min at 4°C, the soluble fractions were loaded onto a 1 ml HP His-trap affinity column from GE. Elution was performed with buffer B (25 mM HEPES pH 7.4, 150 mM NaCl, 500 mM imidazole), followed by size exclusion chromatography (Superdex200, GE) in 25 mM HEPES pH 7.4, 150 mM NaCl, 1 mM EDTA.

Purification of MurG and MreB from T. maritima

A similar purification protocol was followed for MurG on a 5 ml HP Strep-trap column from GE with 2.5 mM d-desthiobiotin added into the elution buffer instead of imidazole. All buffers were supplemented with 0.04% DDM, lysis was performed in the presence of 0.5% N-lauryl sarcosine (Sigma), and size exclusion chromatography was performed on a Superose6 column (GE Healthcare Bio-Sciences AB). MreB purification was similar to the MurD, MurE and MurF purification protocols. Nevertheless, buffer A contained 35 mM imidazole, and the lysis buffer was supplemented with lysozyme and DNase. The centrifugation step was replaced by an ultracentrifugation step at 100 000 g for 1 h at 4°C and a washing step at 86 mM imidazole was added.

Expression and purification of SeMet Mur ligases

The same cells and expression vectors were used for the expression of the SeMet ligases, with the exception of MurE, which was expressed in *E. coli* Rosetta cells. After

a first overnight preculture in Luria–Bertani (LB) medium, a second overnight pre-culture in M9 was performed. Cells were diluted in fresh M9 medium until an OD_{600nm} of 0.1 AU for growth at 37°C. When the OD_{600nm} reached 0.4 AU (0.7 AU for MurE), amino acids were added as powders for inhibition of methionine synthesis (100 mg l^{-1} of lysine, phenylalanine, threonine, 50 mg l^{-1} of isoleucine, leucine, valine), and subsequently 200 mg l^{-1} of thiamine, and 60 mg l^{-1} of SeMet were also added. After 30 min of growth with low shaking at room temperature, expression was induced as for the native proteins. Purifications were performed as described earlier, but in the presence of 2 mM DTT.

Dot Blot assay

One microgram to 7 μg of purified proteins were loaded as dots onto a nitrocellulose membrane. After 15 min of drying, the membranes were blocked for 1 h at room temperature in PBS-T (PBS-Tween 0.05%) supplemented with 3% BSA (bovine serum albumin; Euromedex). After one washing step in PBS-T, the membranes were incubated overnight at 4°C in a PBS-T solution with 0.1 mg ml^{-1} of the potential partner protein. After three washing steps in PBS-T, the membranes were incubated with either the Streptactin HRP conjugate from IBA at 1:20 000, or the SuperSignal West HisProbe from Pierce Biotechnology at 1:2000 in PBS-T. Membranes were developed with the SIGMA FASTTM DAB tablets with metal enhancer). Negative control: BSA (Euromedex), 7 μg per dot.

SPR spectroscopy

Real-time monitoring of the interactions was performed on a BIAcore 3000 device (GE Healthcare Bio-Sciences AB). For that purpose, flow cells of CM5 sensor chips (GE Healthcare) were activated with 50 μl of 0.2 M N-ethyl-N'-(dimethylaminopropyl)-carbodiimide and 0.05 M N-hydroxysuccimide at 5 μl min^{-1}. Subsequently, MreB and MurG, at a concentration of 10 μg ml^{-1} and 20 μg ml^{-1} respectively, were injected over one of the activated flow cells in 10 mM sodium acetate pH 5.0 or 10 mM HEPES pH 7.0, respectively, until a coupling level of approximately 2000 resonance units was reached. A flow cell with no immobilized protein (containing only buffer) was used as a blank. Ethanolamine-HCl 1 M pH 8.5 was injected to saturate the surface at 5 μl min^{-1} for 8 min. For binding experiments, samples diluted in running buffer (25 mM HEPES, 150 mM NaCl, pH 7.4) supplemented with either 0.005% v/v polysorbate 20 or 0.04% DDM were injected over the different surfaces at a flow rate of 20 μl min^{-1}. The surfaces were washed in running buffer and regenerated with pulse injections of 10 μl of 1 M NaCl, followed by 1 mM EDTA, if required. Negative controls included BSA as well as two His-tagged proteins involved in pilus assembly in *S. pneumoniae*. Binding curves were analysed with the BIAevaluation software 3.2 (GE Healthcare). The specific binding signals were obtained by subtracting both the background signal recorded over the blank surface and the buffer blank (double referencing).

Crystallization and data collection

Crystallization screens were performed at the HTX lab (PSB), leading to preliminary hits, which underwent manual optimizations. Crystals were grown at 20°C using the hanging-drop method, by mixing protein and reservoir solutions in equal volumes of 1 μl. Concentration of protein samples were: MurD (10 mg ml^{-1}), MurE (7 mg ml^{-1}) and MurF (10 mg ml^{-1}). Reservoirs for the different experiments contained: MurD, 2 M ammonium phosphate (Sigma); MurE, 0.1 M sodium acetate (Sigma) pH 5.0, 0.5 M sodium formate (Fisher Scientific); MurF, 0.1 M sodium acetate pH 5.5, 45–49% (v/v) 2-methyl-2,4-pentanediol (MPD) (Sigma). Crystals were mounted onto loops and flash cooled by direct immersion into liquid nitrogen (MurF) or by immersion in the crystallization mother liquor supplemented with 25% glycerol (MurD and MurE). SAD data for MurD were collected on the PROXIMA beamline at SOLEIL (Saclay, France). For MurE and MurF, SAD data were collected on the ID29 and BM30 beamlines at the ESRF respectively. All SAD data were collected at the experimentally determined wavelength of the selenium peak. High resolution data sets were collected on ID14-4 (MurE) and ID29 (MurF).

SAD phasing, model building and refinement

Diffraction images were indexed and scaled with XDS (Kabsch, 1993). Selenomethionine sites were identified and refined using autoSHARP (Vornhein *et al.*, 2007). The MurD, MurE and MurF structures were traced with ARP/wARP 7.0.1 (Cohen *et al.*, 2008). In the case of MurF, PHASER 2.5.1 (Storoni *et al.*, 2004) was used to perform molecular replacement with the model generated by the SAD experiment in order to phase data to 1.65 Å. The full structures were completed by cycles of manual model building with COOT 0.6.2 (Emsley and Cowtan, 2004) and BUCANEER 1.5.2 (Cowtan, 2008). Cycles of restrained refinement employing TLS (Painter and Merritt, 2006) were performed with REFMAC 5.7 (Murshudov *et al.*, 1997) as implemented in the CCP4. Stereochemical verification was performed by PROCHECK (Laskowski *et al.*, 1993) and secondary structure assignment by DSSP (Kabsch and Sander, 1983). The detailed representation of ADP binding pocket was made by Ligplot (Wallace

et al., 1995). Figures were generated with PyMol (http://www.pymol.org). The structures of MurD, MurE and MurF from *T. maritima* were deposited in the PDB database with accession numbers: 4BUC, 4BUB, and 3ZL8.

Acknowledgements

This work used the platforms of the Grenoble Instruct Center (ISBG; UMS 3518 CNRS-CEA-UJF-EMBL) with support from FRISBI (ANR-10-INSB-05-02) and GRAL (ANR-10-LABX-49-01) within the Grenoble PSB. The authors wish to thank J. Marquez and the HTX Lab team (PSB, Grenoble) for access to and help with high throughput crystallization, the ESRF and SOLEIL for access to beam lines, Isabelle Bally (IBS) for assistance and access to the Biacore facility, Michel Thépaut (IBS) from the Membrane Protein Purification platform for help with purifications of MreB, MurD, MurE, and MurF, and Luca Signor for assistance and access to the Mass Spectrometry platform. S. F.-S. was supported by an IRTELIS fellowship from the CEA. Authors declare no competing interests.

References

Barreteau, H., Kovac, A., Boniface, A., Sova, M., Gobec, S., and Blanot, D. (2008) Cytoplasmic steps of peptidoglycan biosynthesis. *FEMS Microbiol Rev* **32:** 168–207.

Basavannacharya, C., Moody, P.R., Munshi, T., Cronin, N., Keep, N.H., and Bhakta, S. (2010) Essential residues for the enzyme activity of ATP-dependent MurE ligase from Mycobacterium tuberculosis. *Protein Cell* **1:** 1011–1022.

Bertrand, J.A., Auger, G., Fanchon, E., Martin, L., Blanot, D., van Heijenoort, J., and Dideberg, O. (1997) Crystal structure of UDP-N-acetylmuramoyl-L-alanine:D-glutamate ligase from *Escherichia coli*. *EMBO J* **16:** 3416–3425.

Bertrand, J.A., Auger, G., Martin, L., Fanchon, E., Blanot, D., Le Beller, D., *et al*. (1999) Determination of the MurD mechanism through crystallographic analysis of enzyme complexes. *J Mol Biol* **289:** 579–590.

Bertrand, J.A., Fanchon, E., Martin, L., Chantalat, L., Auger, G., Blanot, D., *et al*. (2000) 'Open' structures of MurD: domain movements and structural similarities with folylglutamate synthetase. *J Mol Biol* **301:** 1257–1266.

den Blaauwen, T., de Pedro, M.A., Nguyen-Distèche, M., and Ayala, J.A. (2008) Morphogenesis of rod-shaped sacculi. *FEMS Microbiol Rev* **32:** 321–344.

Bouhss, A., Crouvoisier, M., Blanot, D., and Mengin-Lecreulx, D. (2004) Purification and characterization of the bacterial MraY translocase catalyzing the first membrane step of peptidoglycan biosynthesis. *J Biol Chem* **279:** 29974–29980.

Bouhss, A., Trunkfield, A.E., Bugg, T.D.H., and Mengin-Lecreulx, D. (2008) The biosynthesis of peptidoglycan lipid-linked intermediates. *FEMS Microbiol Rev* **32:** 208–233.

Chopra, I. (2013) Discovery of antibacterial drugs in the 21st century. *J Antimicrob Chemother* **68:** 496–505.

Cohen, S.X., Ben Jelloul, M., Long, F., Vagin, A., Knipscheer, P., Lebbink, J., *et al*. (2008) ARP/wARP and molecular replacement: the next generation. *Acta Crystallogr D Biol Crystallogr* **64:** 49–60.

Cowtan, K. (2008) Fitting molecular fragments into electron density. *Acta Crystallogr D Biol Crystallogr* **64:** 83–89.

Divakaruni, A.V., Baida, C., White, C.L., and Gober, J.W. (2007) The cell shape proteins MreB and MreC control cell morphogenesis by positioning cell wall synthetic complexes. *Mol Microbiol* **66:** 174–188.

Dye, N.A., Pincus, Z., Fisher, I.C., Shapiro, L., and Theriot, J.A. (2011) Mutations in the nucleotide binding pocket of MreB can alter cell curvature and polar morphology in Caulobacter. *Mol Microbiol* **81:** 368–394.

Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60:** 2126–2132.

van den Ent, F., Amos, L.A., and Löwe, J. (2001) Prokaryotic origin of the actin cytoskeleton. *Nature* **413:** 39–44.

van den Ent, F., Johnson, C.M., Persons, L., de Boer, P., and Löwe, J. (2010) Bacterial actin MreB assembles in complex with cell shape protein RodZ. *EMBO J* **29:** 1081–1090.

Gaballah, A., Kloeckner, A., Otten, C., Sahl, H.G., and Henrichfreise, B. (2011) Functional analysis of the cytoskeleton protein MreB from *Chlamydophila pneumoniae*. *PLoS ONE* **6:** e25129.

Gordon, E., Flouret, B., Chantalat, L., van Heijenoort, J., Mengin-Lecreulx, D., and Dideberg, O. (2001) Crystal structure of UDP-N-acetylmuramoyl-L-alanyl-D-glutamate: meso-diaminopimelate ligase from *Escherichia coli*. *J Biol Chem* **276:** 10999–11006.

Ha, S., Chang, E., Lo, M.-C., Men, H., Park, P., Ge, M., and Walker, S. (1999) The kinetic characterization of *Escherichia coli* MurG using synthetic substrate analogues. *J Am Chem Soc* **37:** 8415–8426.

Hild, G., Bugyi, B., and Nyitrai, M. (2010) Conformational dynamics of actin: effectors and implications for biological function. *Cytoskeleton (Hoboken)* **67:** 609–629.

Humljan, J., Kotnik, M., Contreras-Martel, C., Blanot, D., Urleb, U., Dessen, A., *et al*. (2008) Novel napthalene-N-sulfonyl-d-glutamic acid derivatives as inhibitors of MurD, a key peptidoglycan biosynthesis enzyme. *J Med Chem* **51:** 7486–7494.

Kabsch, W. (1993) Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J Appl Crystallogr* **26:** 795–800.

Kabsch, W., and Sander, C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22:** 2577–2637.

Kotnik, M., Humljan, J., Contreras-Martel, C., Oblak, M., Kristan, K., Herve, M., *et al*. (2007) Structural and functional characterization of enantiomeric glutamic acid derivatives as potential inhibitors of MurD ligase. *J Mol Biol* **370:** 107–115.

Kruse, T., Bork-Jensen, J., and Gerdes, K. (2005) The morphogenetic MreBCD proteins of *Escherichia coli* form an essential membrane-bound complex. *Mol Microbiol* **55:** 78–89.

Laskowski, R.A., MacArthur, M.W., Moss, D.S., and Thornton, J.M. (1993) PROCHECK: a program to check the stereo chemical quality of protein structures. *J Appl Crystallogr* **26:** 283–291.

Longenecker, K.L., Stamper, G.F., Hajduk, P.J., Fry, E.H., Jakob, C.G., Harlan, J.E., *et al.* (2005) Structure of MurF from *Streptococcus pneumoniae* co-crystallized with a small molecule inhibitor exhibits interdomain closure. *Protein Sci* **14:** 3039–3047.

Lovering, A.L., Safadi, S.S., and Strynadka, N.C.J. (2012) Structural perspective of peptidoglycan biosynthesis and assembly. *Annu Rev Biochem* **81:** 451–478.

Matteï, P.-J., Neves, D., and Dessen, A. (2010) Bridging cell wall biosynthesis and bacterial morphogenesis. *Curr Opin Struct Biol* **20:** 749–766.

Mengin-Lecreulx, D., van Heijenoort, J., and Park, J.T. (1996) Identification of the mpl gene encoding UDP-N-acetylmuramate: L-alanyl-g-D glutamylmeso- diaminopimelate ligase in *Escherichia coli* and its role in recycling of cell wall peptidoglycan. *J Bacteriol* **178:** 5347–5352.

Mohammadi, T., Karczmarek, A., Crouvoisier, M., Bouhss, A., Mengin-Lecreulx, D., and den Blaauwen, T. (2007) The essential peptidoglycan glycosyltransferase MurG forms a complex with proteins involved in lateral envelope growth as well as with proteins involved in cell division in *Escherichia coli*. *Mol Microbiol* **65:** 1106–1121.

Mohammadi, T., van Dam, V., Sijbrandi, R., Vernet, T., Zapun, A., Bouhss, A., *et al.* (2011) Identification of FtsW as a transporter of lipid-linked cell wall precursors across the membrane. *EMBO J* **30:** 1425–1432.

Murshudov, G., Vagin, A., and Dodson, E. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53:** 240–255.

Painter, J., and Merritt, E.A. (2006) Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr D Biol Crystallogr* **62:** 439–450.

Patin, D., Boniface, A., Kovac, A., Hervé, M., Dementin, S., Barreteau, H., *et al.* (2010) Purification and biochemical characterization of Mur ligases rom *Staphylococcus aureus*. *Biochimie* **92:** 1793–1800.

Perdih, A., Kotnik, M., Hodoscek, M., and Solmajer, T. (2007) Targeted molecular dynamics simulation studies of binding and conformational changes in *E. coli* MurD. *Proteins* **68:** 243–254.

Silver, L.L. (2006) Does the cell wall of bacteria remain a viable source of targets for novel antibiotics? *Biochem Pharmacol* **71:** 996–1005.

Smith, C.A. (2006) Structure, function and dynmics in the *mur* family of bacterial cell wall ligases. *J Mol Biol* **362:** 640–655.

Storoni, L., McCoy, A., and Read, R. (2004) Likelihood-enhanced fast rotation functions. *Acta Crystallogr D Biol Crystallogr* **57:** 1373–1382.

Vollmer, W., Blanot, D., and de Pedro, M.A. (2008) Peptidoglycan structure and architecture. *FEMS Microbiol Rev* **32:** 149–167.

Vornhein, C., Blanc, E., Roversi, P., and Bricogne, G. (2007) Automated structure solution using autoSHARP. *Methods Mol Biol* **364:** 215–230.

Wallace, A.C., Laskowski, R.A., and Thornton, J.M. (1995) LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng* **8:** 127–134.

White, C.L., and Gober, J.W. (2012) MreB: pilot or passanger of cell wall synthesis? *Trends Microbiol* **20:** 74–79.

White, C.L., Kitich, A., and Gober, J.W. (2010) Positioning cell wall synthetic complexes by the bacterial morphogenetic proteins MreB and MreD. *Mol Microbiol* **76:** 616–633.

Yan, Y., Munshi, S., Leiting, B., Anderson, M.S., Chrzas, J., and Chen, Z. (2000) Crystal structure of *Escherichia coli* UDPMurNAc-tripeptide d-alanyl-d-alanine-adding enzyme (MurF) at 2.3 Å resolution. *J Mol Biol* **304:** 435–445.

Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Fig. S1. SPR control experiments performed with two distinct His-tagged bacterial proteins whose function is not related to peptidoglycan biosynthesis. MurE, MurF and control proteins 1 and 2 were injected at a concentration of 600 nM over immobilized MurG and MreB as described in *Experimental procedures*. Controls 1 and 2 did not display any binding to neither MurG nor MreB whereas MurE and MurF clearly interacted with both immobilized proteins.

Fig. S2. Superposition of the structures of MurE from *T. maritima* (orange, in complex with ADP) and *E. coli* (1E8C; cyan), with the product of the reaction, present only in the structure of MurE from *E. coli,* shown as sticks. Note that the conformation of the C-terminal domain in both molecules is comparable, despite the fact that only the *E. coli* enzyme was crystallized in the presence of the reaction product.

Fig. S3. Superposition of the structures of MurD from *T. maritima* (blue) and from *E. coli* (yellow), the latter having been crystallized in the presence of UMA (1UAG). Note that the N-terminal and central domains superpose well, while the C-terminal domain displays considerable conformational flexibility.

Table S1. MurF data collection, phasing, molecular replacement and structure refinement statistics.

Table S2. MurE data collection, phasing, molecular replacement and structure refinement statistics.

Table S3. MurD data collection, phasing, molecular replacement and structure refinement statistics.

Bibliography

[1] Martine Couturier, El Mustapha Bahassi, and Laurence Van Melderen. Bacterial death by DNA gyrase poisoning. *Trends in Microbiology*, 6(7):269 – 275, 1998.

[2] Brian W. Matthews. The structure of *E. coli* β-galactosidase. *Comptes Rendus Biologies*, 328(6):549 – 556, 2005.

[3] Mitchell Lewis. The lac repressor. *Comptes Rendus Biologies*, 328(6):521 – 548, 2005.

[4] F.William Studier and Barbara A. Moffatt. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.*, 189(1):113 – 130, 1986.

[5] Barbara Ulmasov, John Capone, and William Folk. Regulated expression of plant tRNA genes by the prokaryotic tet and lac repressors. *Plant Molecular Biology*, 35(4):417–424, 1997.

[6] A. Skerra. Use of the tetracycline promoter for the tightly regulated production of a murine antibody fragment in *Escherichia coli*. *Gene*, 151(1-2):131–5, 1994.

[7] Evelyn M. Witkin. Inherited differences in sensitivity to radiation in *Escherichia Coli*. *Proc Natl Acad Sci U S A.*, 32(3):598, 1946.

[8] W.B. Wood. Host specificity of DNA produced by *Escherichia coli*: bacterial mutations affecting the restriction and modification of DNA. *J. Mol. Biol.*, 16:118–133, 1966.

[9] Philip M. Silverman. Towards a structural biology of bacterial conjugation. *Mol. Microbiol.*, 23(3):423–429, 1997.

[10] J Grodberg and J J Dunn. *ompT* encodes the *Escherichia coli* outer membrane protease that cleaves T7 RNA polymerase during purification. *J Bacteriol.*, 170(3):1245253, 1988.

[11] Bruno Miroux and John E. Walker. Over-production of proteins in *Escherichia coli*: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J. Mol. Biol.*, 260(3):289 – 298, 1996.

[12] Shibdas Banerjee and Shyamalava Mazumdar. Electrospray ionization mass spectrometry: A technique to access the information beyond the molecular weight of the analyte. *Int J Anal Chem.*, page 282574, 2012.

[13] Hanno Steen & Matthias Mann. The abc's (and xyz's) of peptide sequencing. *Nat. Rev. Mol. Cell Bio.*, 5:699–711, 2004.

[14] Gary Siuzdak. *Mass Spectrometry for Biotechnology*. Elsevier Science (USA), 1996.

[15] Edmond de Hoffmann and Vincent Stroobant. *Mass Spectrometry - Principles and Applications. Second Edition*. John Wiley & Sons, Inc., 1999.