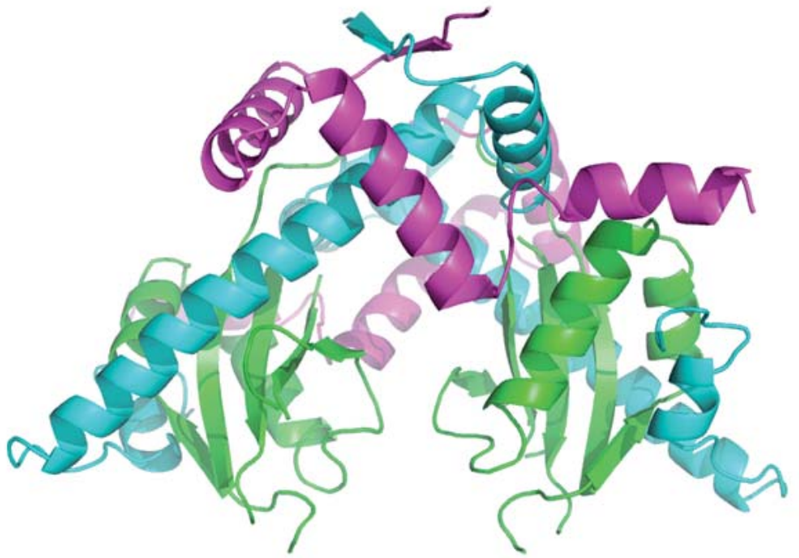
Universitat Autònoma de Barcelona

Structural Insights into Natural Transformation and Toxin-Antitoxin Systems

Biochemical and structural characterization of *Neisseria gonorrhoeae* ComE, ComA and DprA and *Streptococcus pneumoniae* RelBE2



Diana Martínez Llinàs
PhD thesis

IRB Barcelona - IBMB CSIC

2012

Structural Insights into Natural Transformation and Toxin-Antitoxin Systems Diana Martínez Llinàs



Universitat Autònoma de Barcelona

Structural Insights into Natural Transformation and Toxin-Antitoxin Systems

Biochemical and structural characterization of *Neisseria gonorrhoeae* ComE, ComA and DprA and *Streptococcus pneumoniae* RelBE2

PhD Thesis

Diana Martínez Llinàs

Barcelona, 2012





Universitat Autònoma de Barcelona

Thesis written by Diana Martínez Llinàs, PhD candidate enrolled in the program *Structure and Function of Proteins* at the Universitat Autònoma de Barcelona.

This work was carried out at the *Structural Biology of Protein & Nucleic Acid Complexes and Molecular Machines Group* at the Institute for Research in Biomedicine (IRB-Barcelona) and Molecular Biology Institute of Barcelona (IBMB-CSIC), under the supervision of Prof. Miquel Coll Capella.

PhD candidate	PhD supervisor	University advisor
Diana Martínez Llinàs	**Prof. Miquel Coll Capella**	**Dr. Xavier Daura Ribera**
IRB Barcelona	IRB Barcelona	Universitat Autònoma
IBMB-CSIC	IBMB-CSIC	de Barcelona

Barcelona, 2012

A sa meva família:
ma mare, mon pare,
sa padrina, na Jade
i en Francesco





"The time may come when penicillin can be bought by anyone in the shops. Then there is the danger that the ignorant man may easily underdose himself and by exposing his microbes to non-lethal quantities of the drug make them resistant."

(Alexander Fleming, Nobel lecture extract, 1945)

Acknowledgements

Behind any PhD project there are more minds, hearts and hands than those of the PhD candidate. I would like to thank some of those people who made this all possible.

First, I would like to thank all the members of the "Structural biology of protein & nucleic acid complexes and molecular machines" group. I owe my deepest gratitude to Prof. Miquel Coll who gave me the opportunity to take part in this exciting adventure and offered me support, advice and the appropriate environment and resources to carry out this PhD thesis. I am also indebted to Albert, Robert J, and Carme for their support and insightful advice in the design and performance of most experiments; and to Rosa, who crystallized the RelBE complex and gave me the chance to embark on the second project of this PhD thesis. I am extremely grateful to the lab technicians and research assistants Esther F, Nayibe, Maïlys and Leonor for their assistance and support; and to the visiting students Bàrbara, Bernat and Robert O for sharing many moments of joy and sorrow with me. Thanks to Radek, Zuzzanna, Juliana, Roeland, Fabio, Carlo, Marta, Raquel, Dani, Nereida, Pablo, Tomislav, Seb, Lionel, Anna, Maria, Esther P, Sol, Fernando, Cristina and Fran for their ideas, support, coffees, lunches, talks, and all the activities that made the lab a better working place.

Second, I am indebted to our collaborators in Sweden and in Madrid. From my Swedish adventure, I would like to thank Prof. Pär Nordlund, who host me at his lab, made me feel part of his team and introduced me to the field of membrane proteins, and to Marina, Esben, Christian and Pelle for their help in the design and performance of many experiments at Karolinska Institute. Thanks as well to Rosaria, Madhan, TC, Rebecca, Kerstin, Christine, Maria, Marie, Mikaela, Cedric, Madde, Fatma, and to all people I met at Jägargatan for making my stay in Stockholm one of the most enriching experiences in my life. From Madrid, I am extremely grateful to Prof. Manuel Espinosa and to Inma for

the purification and biochemical characterization of RelBE2Spn toxin-antitoxin complex.

Third, I would like to express my gratitude to other members from the IRB Barcelona, IBMB-CSIC and PCB facilities. I am extremely thankful to Jenny for her assistance in all purification experiments and to Sonia, Joan and Alícia for their unconditional support in crystallization trials and for countless lunches and chats throughout these years. Thanks to Prof. Miquel Pons and to Jesús for their support in the performance and interpretation of NMR experiments. Besides, I would like to thank the IRB direction and the members of the IRB PhD student community for giving me the chance to take part in exciting activities such as the organization of a PhD Student Symposium and the founding of the IRB PhD student council. Thanks as well to the members of my Thesis Advisory Committee: Prof. Ignasi Fita, Prof. Xavier Daura and Prof. Ferran Azorín for their insightful advice throughout these years.

Fourth, I wish to thank the entities and funding agencies that provided financial support to this work. Research projects were financed by the former Spanish Ministry of Science and Education and by the former Ministry of Science and Innovation (BFU2005-06758/BMC: 2005-2008 and BFU2008-02372/BMC: 2009-2011). My PhD scholarships were funded by the Spanish Research Council CSIC (JAE/I3P PhD scholarships program) and by the IRB Barcelona (IRB PhD scholarships program).

Finally, I would like to dedicate this work to my family because without their love, support, and faith in me, none of this would have been possible. Thanks to my mother who has always been my role model and source of inspiration. Thanks to my sister for being my guardian angel and best friend. Thanks to my father who inspired me as a child to dream the impossible and to pursue my dreams. Thanks to my grandmother who taught me how to be strong and hold my head high even in difficult times. Thanks to Francesco for being my soul mate and adventure companion.

Preface

"It was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the epoch of belief, it was the epoch of incredulity, it was the season of Light, it was the season of Darkness, it was the spring of hope, it was the winter of despair, we had everything before us, we had nothing before us, we were all going direct to Heaven, we were all going direct the other way - in short, the period was so far like the present period, that some of its noisiest authorities insisted on its being received, for good or for evil, in the superlative degree of comparison only."

(Charles Dickens, A Tale of Two Cities)

This is the written report of an interesting journey full of contrasts and surprises that started in September 2006 and brought me to spend four years at IRB Barcelona/IBMB-CSIC and almost one year at Karolinska Institute (Sweden), to meet many people and to learn something about bacterial genetics and structural biology.

This is a tale of two different projects that started at different time points and were carried out independently of one another. The first project consisted of the biochemical and structural characterization of *Neisseria gonorrhoeae* ComE, ComA and DprA, three proteins involved in natural transformation and the second consisted of the structural characterization of the *Streptococcus pneumoniae* RelBE2Spn toxin antitoxin system.

Although these two projects were born in different contexts, I believe they share important conceptual and methodological aspects. From a conceptual point of view, both projects are devoted to the understanding of molecular mechanisms potentially related to the acquisition and maintenance of antibiotic resistance in bacteria. From a methodological point of view, both projects use techniques for the expression, purification and biochemical and structural characterization of proteins. In an effort to highlight the resemblances and common aspects between both

projects, I have decided to structure the main body of the thesis in four chapters (introduction, material and methods, results and discussion and conclusions). But there could be alternative reading schemes. By citing the protagonist Morelli from Cortazar's book Hopscotch/Rayuela "*You can read my book any way you want to*", I invite you to explore the index of contents to find the order that works better for you.

I hope you will enjoy reading this PhD thesis and that by reading between the lines you will get a grasp of the meaning of this adventure.

Abstract

The incidence of gonococcal and pneumococcal infections remains high in developing countries and is increasing in many parts of the world. The need not only for treatment of the individual but also for control of the diseases at a community level is acute but the selection of appropriate treatments has become a complicated issue by the ability of *Neisseria gonorrhoeae* and *Streptococcus pneumoniae* to develop resistance to antibiotics. This PhD thesis investigates mechanisms related to horizontal gene transfer and growth arrest in *N. gonorrhoeae* and *S. pneumoniae*. The first part of the thesis is devoted to the study of natural transformation in *N. gonorrhoeae* and describes the procedures that have lead to the expression, purification and biochemical characterization of ComE, ComA and DprA, three proteins involved in the uptake and processing of environmental DNA. The second part of the thesis describes the structural characterization by X-ray crystallography of *S. pneumoniae* RelBE2, a chromosomally-encoded type II toxin-antitoxin system that has been linked to translation moderation and growth arrest under starvation conditions, and proposes a model for its regulation.

Keywords

Natural transformation, Toxin-antitoxin systems, Antibiotic resistance, RelBE, ComE, ComA, DprA, *Neisseria gonorrhoeae*, *Streptococcus pneumoniae*, X-ray crystallography, Structural biology





Table of contents

List of figures

List of tables

List of abbreviations

Å	Angstrom
AFM	Atomic force microscopy
ATP	Adenosine triphosphate
bp	Base pairs
BSA	Bovine serum albumin
CDC	Centers for Disease Control & Prevention
cDNA	Complementary DNA
CHAPS	3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate
CoFi	Colony-Filtration
C-terminal	Carboxy-terminal
CV	Column volumes
Dcw	Division cell wall
ddH_2O	double-distilled water
DDM	n-Dodecyl-β-maltoside
DLS	Dynamic light scattering
DMS	Dymethyl suberimidate
DMSO	Dimethylsulphoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
dsDNA	Double stranded DNA
DTBP	dimethyl 3,3'-dithiobispropionimidate
DTT	Dithiothreitol
DUS	DNA uptake sequence
EB	Elution buffer
EDTA	Etylendiaminetetraacetate
EK	Enterokinase
EMSA	Electrophoretic Mobility Shift Assay
FC	Fos-choline
Fig	Figure
FF	Fast flow
FL	Full length
g	Gram
g	Gravity acceleration constant
GF	Gel filtration
GST	Glutathione-S-Transferase
h	Hour
His-tag	Poly(6) Histidine-tag

HGT	Horizontal gene transfer
HP	High performance
HPLC	High performance/pressure liquid chromatography
HRP	Horseradish peroxidase
IMAC	Immobilised Methal ion Affinity Chromatography
IP	Isoelectric point
IPTG	Isopropyl-D-beta-galactopyranoside
Kan	Kanamycin
Kb	Kilobase
KDa	Kilodalton
l	litre
L-Arg	L-arginine
LB	Luria Bertani broth
LDS	Lithium dodecyl sulfate
LGT	Lateral gene transfer
LIC	Ligation-independent cloning
LOS	lipooligosaccharides
µ	micro (10^{-6})
M	molar
m	milli (10^{-3})
MAD	Multiple-wavelength anomalous diffraction
MBP	Maltose binding protein
MES	2-(N-morpholino) ethane sulphonic acid
MPD	2-Methyl-2,4-pentanediol
min	Minute
mRNA	Messenger RNA
MW	Molecular weight
n	nano (10^{-9})
Na-P	Sodium phosphate
NHS	National Health Service
NMR	Nuclear magnetic resonance
NC-membrane	Nitrocellulose membrane
NCS	Non-crystallography symmetry
N-terminal	Amino-terminal
ODx	Optical density at x nm
ORF	Open reading frame
PAGE	Polyacrylamide gel
PBS	Phosphate buffer saline
PBPs	Penicillin-binding proteins
PCR	Polymerase chain reaction

PDB	Protein Data Bank
PI	Protease inhibitors
Pol	Polymerase
PVDF	Polyvinylidene fluoride
ppm	Parts per million
rDNA	Ribosomal DNA
rmsd	Root mean square deviation
rpm	Revolutions per minute
RHH	Ribbon-helix-helix
RNA	Ribonucleic Acid
RT	Room temperature
s	Seconds
SANS	Small-angle neutron scattering
SAXS	Small-angle X-ray scattering
SDS	Sodium dodecyl sulphate
SEC	Size exclusion chromatography
SeMet	Seleno-L-methione
SLS	Static light scattering
ssDNA	Single stranded DNA
tRNA	Transfer RNA
T4SS	Type IV secretion system
TA	Toxin-antitoxin
TAE	Tris Acetate EDTA buffer
Taq	*Thermus aquaticus*
TB	Terrific broth
TBST	Tris-buffered saline tween-20
TCEP	Tris chloride ethanol phosphine
TEMED	N,N,N,N,-tetramethylendiamine
TEV	Tobacco Edge Virus
Tfp	Type IV pili
Tris	Tris(hydroxxymethyl)aminomethane
TM	Transmembrane
U	Unit
UV	Ultraviolet
V	volt
vol	volumes
[v/v]	volume per volume
[w/v]	weight per volume



Chapter 1
Introduction





1.1 The complicated relationship between human beings and bacteria

The possibility to fight against bacterial infections represents one of the major successes of modern medicine. This section presents a bit of history of microbiology and antibiotic development, introduces the problem of antibiotic resistance, describes the genera *Neisseria* and *Streptococcus* and addresses relevant aspects regarding the prophylaxis and treatment of gonococcal, meningococcal and streptococcal diseases. Besides the references provided throughout the text, descriptions of both genera can be found in the books "Brock Biology of microorganisms" (Brock *et al.*, 2006) and "Prescott's Microbiology" (Prescott *et al.*, 2002). More information regarding epidemiology, treatment and profilaxis of gonococcal, meningococcal and streptococcal diseases can be extracted from the CDC webpage [Web1] and from the WHO webpage [Web2].

1.1.1 Microorganisms and disease

The discovery of microorganisms has been attributed to Van Leeuwenhoek, a Dutch merchant without any scientific background who was curious enough to observe water under a rudimentary microscope. His drawings, dated late 1670s, were perceived by the scientific community as something interesting but rather irrelevant. Van Leeuwenhoek microscopes pictures and writings can be found in a report written on occasion of Van Leeuwenhoek tercentenary (Porter, 1976).

It took two centuries to work out that microorganisms, bacteria among them, could be transmitted among people causing diseases. The germ theory of disease that linked microorganisms and human disease was developed by several scientists over the 1860's and 1970's and culminated with the establishment of Koch's postulates in 1890. An interest-

ing analysis about the origin of Koch's postulates is debated in a paper written by Carter (1985).

However, although several microorganisms were found to cause diseases, doctors had no tools to fight against them. The situation changed when in 1928 Alexander Fleming, a Scottish doctor who was investigating *Staphylococcus*, incidentally got his cultures contaminated by mold. The mold, later identified as *Penicillum notatum*, was preventing bacterial growth. Fleming found that the active compound, penicillin, was effective against a wide range of bacteria. Growing mold and isolating penicillin was a cumbersome process, so the use of penicillin could not be extended until mid 1940's, when Florey, Chain and Heatley proved the efficiency of penicillin to treat animals and developed methods for its large-scale production. A review about the discovery and early development of penicillin can be found in Ligon (2004).

1.1.2 The rise and fall of the antibiotics empire

The discovery of penicillin marked the beginning of a new era in medicine. Many compounds with bacteriostatic and bactericidal activities were discovered upon the analysis of antagonistic behaviors between microorganisms (Waksman and Woodruf, 1942). The term antibiotic was coined by Waksman to refer to those substances produced by microorganisms that had the ability to kill bacteria (Kresge *et al.*, 2004).

Following the discovery of penicillin, the isolation of tyrothricin (mixture of gramicidin and tyrocidine) in 1939, ofactinomycin in 1940, streptothricin in 1942, streptomycin in 1943 and neomycin in 1949, opened the gold era of discovery of naturally occurring antibiotics. The discovery of new antibiotics began to grow exponentially. Originally produced by fungi, many antibiotics were soon produced synthetically. New techniques to modify natural compounds, to modify natural biosynthetic pathways or to produce recombinant antibiotics were intro-

duced as well. A good review about antibiotic formulation history can be found in Borders (2007).

Although at the beginning it seemed all bacterial infections could be conquered by the development of new antibiotics, many bacterial strains were found to be resistant to old and new antibiotics, and drug resistance, far from being controlled, was on the rise (Appelbaum, 1992).

Infectious disease mortality in developed countries decreased through most of the 20th century, in part thanks to the use of antimicrobial agents. However, the abuse and misuse of antibiotics has lead to the spread of multi-resistant strains that are responsible for many deaths. As a consequence, a change in epidemiological trends has been observed in many countries, and deadly infectious diseases are gaining ground. In this scenario, finding the clues to antibiotic resistance emergence and spread has gained importance.

1.1.3 Insights into the biochemical basis of antibiotic efficiency and inefficiency

Understanding the multilayered mechanisms by which antibiotics kill bacteria is clue to understand why and how bacteria become resistant to antibiotics. This section does not aim at analyzing in detail the phenomenon of antibiotic action and resistance, but just at giving an overview to the topic by introducing a bit of the biochemistry behind the processes.

1.1.3.1 The many mechanisms of action of antibiotics

In the 40's, it was already clear that antibiotics differed in their chemical structure, that antibiotic efficiency was strain and dose dependent and that there were various ways of action (Waksman and Woodruff, 1942). However, the biochemical details behind antibiotic action were largely unknown.

Nowadays we know that most antibiotics interfere with bacterial growth by targeting essential cellular processes such as DNA replication, RNA synthesis, and cell wall synthesis, but the modes of action of many antibiotics and molecular details behind their action are not known yet.

The list below classifies antibiotics as per their chemical structure and points out important members of each family and plausible mechanisms of action [Web3]:

- **Aminoglycosides** such as streptomycin, neomycin, kanamycin, gentamicin, tobramycin, and amikacin. These antibiotics inhibit protein biosynthesis through binding to the ribosome.
- **Ansamacrolides or ansamycins** such as rifampin. These antibiotics inhibit RNA polymerase.
- **β-lactams** is the most important group from a therapeutical point of view. It includes penicillins, cephalosporins, carbapenems, monobactams, nocardicins, and clavulanic acid. β-lactams inhibit bacterial cell wall synthesis. β-lactams bind to penicillin binding proteins (PBP's). β-lactams carboxypeptidase or transpeptidase activities and in some cases stimulatate the activity of autolysins.
- **Chloramphenicol** inhibits protein synthesis by binding to the 50s ribosomal subunit and blocking the peptidyltransferase reaction.
- **Glycopeptides** such as vancomycin, avoparcin, and teicoplanin. They inhibit bacterial cell-wall biosynthesis by specifically binding to D-alanyl-D-alanine units of the lipid-boud precursor found in the cell walls.
- **Licosamides** such as lincomycin and celesticetins are antibiotics with carbohydrate-like structures that interfere with protein biosynthesis through binding to the 50s subunit of the ribosome.
- **Lipopeptides** such as daptomycin which disrupts bacterial cell-wall membranes.
- **Macrolides** inhibit protein synthesis by binding to the 50s subunit.

- **Polyethers** interact with bacterial cell membranes and allow cations to pass through the membranes causing cell death.
- **Tetracyclines** inhibit protein synthesis by binding to the 30S subunit of the ribosome.
- **Synthetic antibacterial agents** not derived from microbial products or based on structures of microbial products. This group includes fluoroquinolones which prevent bacteria growth by inhibiting DNA gyrase and oxazolidinones that interfere with mRNA translation.

1.1.3.2 Strategies for resistance development

In order to deal with antibiotics, bacteria have developed many strategies. Some of them imply the modification of the target enzyme or the metabolic pathway (modification of drug target or molecular bypass). Others avoid the entrance, promote the exit or chemically modify the antibiotic (permeability barriers, efflux pumps of chemical modification). These mechanisms are summarized below with some representative examples.

- **Modification of drug target:** *i.e.* streptomycin resistance due to mutations in rDNA genes (rpsL) (Agarwal *et al.*, 2011) or β-lactams resistance due to changes in PBPs (Zapun *et al.*, 2008).
- **Molecular bypass:** *i.e* trimethoprim resistance due to overproduction of DHF (dihydrofolate) reductase (Maskell *et al.*, 2001).
- **Permeability barriers:** *i.e* outer membrane in gram negative bacteria (Bolla *et al.*, 2011).
- **Efflux pumps:** *i.e. Streptococcus pneumoniae* MefA (Bley *et al.*, 2011), *Staphylococcus aureus* NorA (Zhang and Ma, 2010) and *Pseudomonas aeruginosa* Mex (Kumar and Schweizer, 2004).
- **Chemical modification:** *i.e.* aminoglycoside-modifying enzymes (Ramirez and Tolmasky, 2010), β-lactamases (Bush, 2010), chloramphenicol acetyltransferases (Schwarz *et al.*, 2004).

Although the individual analysis of these systems has provided invaluable information to understand the biochemistry behind antibiotic resistance and to modify existing drugs, evidence points towards common mechanisms of drug-induced cell death and antibiotic resistance. In this context, antibiotic cell-death and antibiotic resistance, rather than being caused by a punctual mechanism, would be the result of a complex cell response to drug-stress. As a consequence, the problem of antibiotic resistance should be approached by taking into consideration the complexity of stress responses in individual bacteria and in bacteria populations.

1.1.3.3 Mechanisms for resistance development

The following subsections review in more detail some processes by which bacteria acquire variability putting emphasis on the development and spread of antibiotic resistance. First subsection discusses the importance of spontaneous and induced mutations in the development of antibiotic resistance. Second subsection introduces horizontal gene transfer as a means of antibiotic resistance spread. Third subsection points out phase variation and antigenic variation as two mechanisms important for the rapid generation of variation hence fast development of antibiotic resistance in some pathogenic organisms. Fourth subsection analyzes non-inheritable resistance mechanisms such as drug indifference, biofilm formation and other persistence phenomena. Last subsection discusses antibiotic resistance from a holistic perspective and the need of new therapeutical approaches to fight against bacteria and defeat or at least reduce the number of antibiotic resistant strains.

Gaining antibiotic resistance by spontaneous and induced mutations

De novo mutation is a powerful mechanism for the direct modification of drug targets and drug uptake systems (Miller *et al.*, 2002) and for the derepression of drug efflux systems (Grkovic *et al.*, 2002).

Mutation can be seen as a spontaneous phenomenon (spontaneous mutation rate) that may get enhanced under some conditions (induced mutagenesis). In 1943 Max Delbrück and Salvador Luria shown that in bacteria, random genetic mutations could arise at a certain rate in the absence of selection. Since Luria-Delbrück experiments, many efforts have been devoted to determine mutation rates in bacterial populations (Saunders *et al.*, 2003). Mutation rates seem to be determined by a balance between the deleterious consequences of many mutations, and the metabolic costs of further reducing mutation rates during replication (Denamur and Matic, 2006).

Assuming that most spontaneous mutations have negative effects, variants with increased mutation rates would be expected to have reduced fitness. But in some situations, *e.g.* when facing the dangers and possibilities of a continuously changeable environment, increased mutation rates might confer a selective advantage. Then, mutator strains (strains that permanently exhibit higher mutation rates than their counterparts) would be favoured.

Mutator phenotypes have been linked not only to the development of multiple-antimicrobial resistance (Macià *et al.*, 2005), but also to enhanced pathogenicity (Meyers *et al.*, 2003; Oliver and Mena, 2010). Not surprisingly, hypermutation phenotypes have been explained by deficiencies in DNA repair (García-Ortiz *et al.*, 2011), or in error avoidance mechanisms such as alterations in the function of the mismatch repair system, a highly conserved system that detects and repairs mispairs and short insertions or deletions produced during replication (Jolivet-Gougeon *et al.*, 2011). Genomic analysis has revealed that some microorganisms characterized by high mutation rates such as *Mycobacterium tuberculosis* lack of some components essential for DNA repair (Dos Vultos *et al.*, 2001).

Bacterial exposure to stressful conditions can trigger the activation of rescue mechanisms such as the SOS response, a global response to DNA damage characterized by cell cycle arrest, DNA repair and mutagenesis induction. Interestingly, antibiotics that interfere with DNA replication can trigger the activation of the SOS response (Maiques *et al.*, 2006). The activation of the SOS response not only alleviates DNA damage (Miller *et al.*, 2004), but also promotes the activation of error-prone polymerases that introduce variability in bacterial populations (Michel, 2005)

Spreading antibiotic resistance by horizontal gene transfer

While punctual mutations can confer resistance through facile modifications of target enzymes, horizontal gene transfer - *i.e.* transfer of DNA between organisms that are not parent and offspring - guarantees the spread of efflux pumps and metabolic pathways for antibiotic degradation. Indeed many systems that confer antibiotic resistance are encoded in genetic mobile elements that can be spread by horizontal gene transfer (Partridge, 2011).

Although the extensive use of antibiotics in the last 50 years has probably accelerated the spread of antibiotic resistance among human pathogens, most resistance elements have been circulating in bacterial populations for millennia.

Horizontal gene transfer is such a common event that it makes sense to consider every resistance gene or even every potential precursor both in pathogenic and non-pathogenic bacteria as a potential threat. In this context the term resistome is used to refer to the collection of all the antibiotic resistance genes and its precursors (Wright, 2007).

But the link between antibiotics and horizontal gene transfer is two-sided. On the one hand, horizontal gene transfer promotes the spread of antibiotic resistance genes. On the other hand, exposure to antibiotics

can activate horizontal gene transfer through the activation of the SOS response (Beaber *et al.*, 2004; Hastings *et al.*, 2004).

Evading transitory dangers by phase variation and antigenic variation

The term phase variation is commonly used to refer to an inheritable reversible switch between two states (on/off) that result in variation in the level of expression of one or more proteins between individual cells from a clonal population. After cell division, most cells will retain the expression phase of the parents, but a small number will switch expression phase. Phase variation is an efficient method to deal with changing environments and to evade the immune system (van der Woude, 2011).

Phase variation can be achieved through epigenetic modifications such as methylation that do not alter DNA sequence (van der Woude and Henderson, 2008) and through genetic modifications that rely on slipped strand mispairing (Torres-Cruz and van der Woude, 2003) or specific-site recombination events (van der Woude, 2011). Recombination events that lead to temporary DNA rearrangements are at the basis of *Salmonella typhimurium* flagellar phase variation (Bonifield and Hughes, 2003) and *Neisseria meningitidis* Opa and LOS phase variation (Hill and Davies, 2009).

The term antigenic variation has been used in various ways in the literature. This section refers to the process that consists of the expression of one or multiple forms of protein variants in bacterial populations (Vink *et al.*, 2011).

Antigenic variation has been well studied in the pathogens *Borrelia burgdorferi* and *Neisseria gonorrhoeae*. *B. burgdorferi* VslE undergoes antigenic variation through random, segmental recombination between a copy of the coding gene and multiple silent loci in a RecA independent manner (Lin, *et al.*, 2009). *N. gonorrhoeae* PilE undergoes homologous recombination with silent PilS copies in a process that requires the ac-

tion of RecA, RecX, RecJ, RecO, RecQ, RecR, RecG and RuvABC (Hill and Davies, 2009).

Extensive DNA rearrangements found in hyper-recombinant strains are also related with antigenic variation and pathogenesis. In a similar manner to that of hypermutant phenotypes, hyperrecombinant phenotypes are believed to be related with the emergence of antibiotic resistance in *S. pneumoniae* (Hanage *et al.*, 2009).

Non-inherited antibiotic resistance

The genetic modifications explained above cannot fully explain the resistance phenomenon. "Non-inherited antibiotic resistance" embraces a number of mechanisms by which bacterial populations that are genetically homogeneous and susceptible to antibiotics can become temporary resistant to their action (Levin and Rozen, 2006).

Most antibiotics only attack exposed and rapidly growing cells. "Non-inherited antibiotic resistance" phenomena are based on the fact that in a bacterial culture not all bacteria are equally exposed or actively dividing.

Persistence is a trait exhibited by a subpopulation of bacteria, characterized by slow growth and the ability to survive antibiotic treatment. Persistence phenomena were described for the first time in 1944 when Bigger observed that growing bacteria cultures could not be completely killed by antibiotic treatment and a small fraction of cells "persisted" (Kussell and Kishoni, 2005). The activation of persistent seems to be a tightly regulated mechanism to overcome transitory environmental stresses. Persistence has a direct benefit to cells because it allows survival during catastrophes, but it can also provide an indirect benefit to siblings, because it can reduce competition for limiting resources (Gardner *et al.*, 2007).

Some cells can become indifferent to antibiotic treatments or to other environmental stresses by entering a dormancy state or by stalling their cell growth. Entrance in a dormancy state can result in the formation of recognizable structures such as endospores (Errington, 2003), but not all dormant cells appear different under the microscope. The controlled synthesis of proteins that block or delay essential biological processes might trigger the entrance in a "less active" state in many organisms (Yamaguchi and Inouye, 2011). This is thought to be the *raison d'être* of some toxin-antitoxin systems.

Another mechanism to fight against environmental threats such as the exposure to antibiotics is biofilm formation. Biofilms are complex aggregates of cells that attach to surfaces and live within matrices of polymeric exudates that protect the cells from adverse conditions (Hall-Stoodley *et al.*, 2004). Biofilm matrix provides mechanical stability, mediates adhesion to surfaces, forms networks that interconnect and immobilize the cells and acts as an external digestive system (Flemming and Wingender, 2010). Biofilm formation has been associated with quorum-sensing-regulated mechanisms (Nadell *et al.*, 2008) and increased level of mutations (Macià *et al.*, 2011). Interestingly, biofilms have been found to promote horizontal gene transfer (Sørensen *et al.*, 2005).

Biofilms can grow onto many surfaces, but those that adhere to medical devices or thrive onto human tissues are very important in the clinical setting. Microorganisms such as *Enterococcus faecalis, Staphylococcus aureus, Staphylococcus epidermidis, Streptococcus viridians, Escherichia coli, Klebsiella pneumoniae, Proteus mirabilis,* and *Pseudomonas aeruginosa* can form biofilms onto catheters or other devices and cause severe infections if they gain access to the patient bloodstream (Donlan, 2001). Besides, biofilm formation within the human body is a common feature of chronic lung infections (Yang L. *et al.*, 2008), caries (Sbordone, 2003) and other pathologies [Web4].

1.1.4 New ways to fight against resistant bacteria

As mentioned before, rather than being caused by a single mechanism, antibiotic death and antibiotic resistance are the result of complex cell response to drug-stresses. All the mechanisms presented above do not act alone, but are part of an intricate network and share components with other stress-related responses.

SOS response, recombination and growth regulation play central roles in the development of antibiotic resistance. A better understanding of bacterial responses to stressful conditions would lead to the development of future generations of antibiotics designed to block processes that are essential for the establishment antibiotic resistance.

Finding new drugs toxic to bacteria but not to eukaryotic cells or to symbiotic bacteria is not an easy task. Broad spectrum antibiotics kill both pathogenic and commensal bacteria, and cause short-term imbalances in the natural flora of the gut, skin or any other treated area. But the most concerning issue is that continuous antibiotic treatments can promote the stabilization of drug-resistant commensal strains. Those strains, potentially able share resistance genes with pathogenic species through horizontal gene transfer mechanisms, can stress the risk of appearance of multiresistant strains (Jernberg *et al.*, 2010).

A possible way to cope with these difficulties is to gain insight into mechanisms and pathways unique or particularly important in pathogenic bacteria. Besides, exploring therapeutic approaches to prevent horizontal gene transfer and to treat persistent infections seems to be of great interest to interfere with antibiotic resistance spread and to reassure a longer lifespan to present and future antibiotics.

1.1.5 Neisseria sp. and Streptococcus sp.

This section briefly describes the genera *Neisseria* and *Streptococcus* and addresses relevant aspects regarding the prophylaxis and treatment of gonococcal, meningococcal and streptococcal diseases.

1.1.5.1 *Neisseria sp.*

Neisseria is a genus of gram negative bacteria that belongs to the Proteobacteria group. *Neisseria sp.* are kidney bean shaped cells that usually grow in pairs (diplococcus). The genus takes the name from Albert Neisser, the bacteriologist who discovered *N. gonorrhoeae* in 1879 in samples from patients with gonorrhoea (Ligon, 2005).

The genus is very relevant a clinical point of view because it includes two important human pathogens: *N. gonorrhoeae* and *N. meningitidis* and at least eight other species such as *N. lactamica* and *N. cinerea* which are considered commensals but can cause disease in immune-compromised patients (Morse, 1996).

N. meningitidis (also called meningococcus) exists as part of the normal flora of human nasopharinx, but can cause severe forms of meningitis in children and teenagers [Web5]. *N. meningitidis* can also reach the blood stream and cause septicaemia. *N. meningitidis* is easily spreadable and can cause epidemics. Epidemiology of meningococcal infections in Spain has been analyzed by Gil-Prieto *et al.* (2011).

Antibiotic treatment is available (Zalmanovici *et al.*, 2011) and prophylactic treatment is recommended for people who have been close to infected individuals to avoid infection spread. There are vaccines against *N. meningitis* types A, C, W135 and Y [Web6]. There are also vaccines used to control type B in epidemic settings but vaccines against type

B meningitis conferring long-lasting protection against type B are still unavailable (Yogev and Tan, 2011). In most cases, although vaccination can protect against invasive meningococcal disease, it does not prevent carriage, which is the primary way of transmission (see Meningitis Vaccine - Meningitis Research Foundation at [Web7].

N. gonorrhoeae (also called gonococcus) normally infects genital mucosae and causes inflammation and purulent discharges. In some cases infections are asymptomatic, but asymptomatic infections can be dangerous as well because, being untreated, they can lead to pelvic inflammatory disease and infertility in females (Walker and Sweet, 2011). *N. gonorrhoeae* can also cause conjunctivitis in newborns (Woods, 2005) and disseminated infections that can lead to death (Dal Conte *et al.*, 2006).

The characterization of *N. gonorrhoeae* strains is useful for epidemiological purposes, but also to provide effective treatment options. In particular, the surveillance of resistance to antibiotics is important to adapt treatments and decrease the transmission of resistant strains (Monfort *et al.*, 2009). Combinations of typing techniques are routinely used to provide a better characterization of circulating strains (Unemo *et al.*, 2009). Unfortunately, the development of an effective vaccine against *N. gonorrhoeae* is not yet in sight (Barh *et al.*, 2010).

Although *N. gonorrhoeae* and *N. meningitidis* infections can be treated with appropriate antibiotics, the rapid development of antibiotic resistance has obliged to change treatment guidelines several times and has raised concerns in the medical community. A wise use of antibiotics together with a tight surveillance on new cases of antibiotic-resistance is essential to guarantee the usefulness of antibiotics (Stefanelli, 2011). Understanding the underlying molecular basis of the high variability, high heterogeneity and low immunogenicity that characterize *Neisseria sp.* is of great interest to provide new treatment options. Interfering

with processes such as horizontal gene transfer, antigenic variation and phase variation would be a definitive goal in the fight against antibiotic resistance.

1.1.5.2 *Streptococcus sp.*

The word Streptococcus was coined by Theodor Billroth in 1868 to refer to organisms forming chain-like structures in the pus of human wounds. The word Streptococcus comes from the combination of two words: streptos (στρεπτος) - meaning chain formed of links - and kokkos (κοκκος) - meaning berry (Wilson, 1987).

At the time streptococci were observed for the first time the link between microorganisms and disease was not completely established. Microorganisms forming chain-like structures were found in a variety of pathological conditions including surgical infections, erysipelas, strep throat, tuberculosis and a number of fevers, but also in healthy individuals. Therefore streptococci were thought to belong to a number of different species. The discovery and early history of streptococci is excellently reviewed by Wilson (1987).

The genus *Streptoccocus* includes important pathogens such as S. *pyogenes, S. agalactie, S. pneumoniae, S. viridians* and *S. mutans* but also commensal species of human mouth, skin, intestine and upper respiratory tract. Related species are used in biotechnological applications such as wine making, milk fermentation, cheese ripening and yogurt production [Web8].

S. pyogenes can cause uncomplicated respiratory infections such as pharyngitis tonsillitis and otitis, but also skin infections, and severe internal infections that can lead to systemic infections and can cause death and dangerous sequels (McArthur *et al.*, 2011). *S. pneumoniae* is a normal colonizer of human respiratory tract but in immunocompromised indi-

viduals it can cause pneumonia, meningitis and in rare cases systemic infections (Mook-Kanamori *et al.*, 2011). *S. agalactie* can cause miscarriage in pregnant women and meningitis in the newborn, *S. viridans* is an important cause of endocarditis and *S. mutans* causes dental caries (Jevitz Patterson, 1996).

Streptococcal infections are usually treated with penicillin and derivatives (Jevitz Patterson, 1996). However, the misuse and abuse of antibiotics is threatening their usefulness. *Streptococcus* species are becoming resistant to commonly used antibiotics, and treatment failure and recurrence are becoming a concerning issue among clinicians.

There are vaccines against *S. pneumoniae*, but their efficiency is still far from being optimal (Mundany *et al.*, 2003). In the case of *S. pyogenes* and *S. agalactie*, there are only vaccine candidates in clinical trials (Steer *et al.*, 2009; Heath, 2011). The low immunogenicity and immunological cross-reactivity with human components, together with the diversity of strains have complicated the development of efficient vaccines (Nossal, 2011).

A better understanding of *Streptococcus* transmission, virulence, resistance and persistence mechanisms will open the door to new therapeutic approaches.

1.2 Swapping genetic material by horizontal gene transfer

Horizontal gene transfer (HGT), also called Lateral gene transfer (LGT), is defined as the process by which an organism can take up genetic material from another organism without being its offspring. The name was coined in contrast with vertical gene transfer (*i.e.* transmission of genetic information to offspring). This section describes the molecular biology behind horizontal gene transfer putting particular focus on natural transformation in *N. gonorrhoeae*. Good introductions to horizontal gene transfer can be found in *Molecular Genetics of Bacteria* (Dale and Park, 2004) *Horizontal Gene transfer 2nd edition* (edited by Syvanen and Kado, 2002).

1.2.1 Re-shaping the tree of life

In the 1970's, with the development of phylogenetic taxonomy of 16S ribosomal RNA, Woese opened the door to the study of bacterial phylogenies, but also to the depiction of a general scheme for the classification of all living organisms. According to this scheme, all living organisms belonged to one of the three interrelated primary kingdoms representing the three aboriginal lines of descent: eubacteria, archaebacteria and urkaryotes, the later represented now in the cytoplasmatic component of eukaryotic cells (Woese and Fox, 1977).

Over the last two decades, the analysis of complete genomes has revealed that history of genomes cannot be fully explained through vertical inheritance (*i.e.* gene transfer from parents to offspring). Horizontal gene transfer events (*i.e.* gene transfer from organisms non-related through parental ties) would explain why phylogenetically distant organisms can easily share non-essential genes and pathways with distant organisms occupying similar ecological niches. But large extents of horizontal gene

transfer may even lead to question the very concept of species, and the unit of selection in evolutionary theories (Andam and Gogarten, 2011). The evidence and abundance of horizontal gene transfer events among distant organisms, raises doubts about the usefulness of representations that merely represent vertical inheritance patterns. Some authors suggest that the relationships between species would be better represented by nets or other graphs (Kunin *et al.*, 2005), but others argue that tree-like representations are appropriate because they believe horizontal gene transfer events have minor effects in shaping the evolution of life when compared to vertical inheritance (Daubin *et al.*, 2003; Ciccarelli *et al.*, 2006).

It seems horizontal gene transfer events drove early evolution and still play an important role in the evolution of prokaryotic communities, but the global extent and the biological meaning of horizontal gene transfer are not clear yet.

The fate of incoming DNA is not clear. Most authors support the idea that HGT results in the acquisition of new capabilities such as new metabolic pathways (Iwasaki and Takagi, 2009), virulence factors (Faguy, 2003) or antibiotic resistance (Barlow, 2009), but others suggest that incoming DNA is used to repair deficient copies of genes (Hamilton and Dillard, 2006) or even as a food source for the recipient cell (Palchevskiy and Finkel, 2006).

Horizontal gene transfer is often seen as an unidirectional process beneficial for the recipient cell that has an important energy cost and can even cause cell death of the donor cell. Given its "altruistic nature", horizontal gene transfer has profound implications in the theory of evolution.

Indeed, in a classical organism-centered view of evolution, it is difficult to find the benefits of such an altruistic behavior. But if focus is put on

the survival of the fittest gene or on the survival of the fittest population or community, horizontal gene transfer events do not contradict this evolutionary principle, but rather reinforce it.

In gene-centered views of evolution, as the one largely defended in "*The selfish gene*" by which evolution is guided for the benefit of the genes and organisms are vessels designed by the genes to carry the genetic information from generation to generation (Dawkins, 1989), any strategy to promote gene transfer and survival would be justifiable. In population or community-centered views of evolution, sharing genetic information between members belonging to the same community would be justified even if it did not represent any benefit for the organism carrying the genes because it would improve the overall fitness of the population or the community. Interestingly, population-based antibiotic resistance mechanisms have been found in bacteria (Lee *et al.*, 2010). In any case a satisfactory answer for the evolutionary meaning of horizontal gene transfer is still to come.

1.2.2 Modes of horizontal gene transfer

Although horizontal gene transfer events have been described in all three kingdoms, they are best studied in prokaryotic organisms (Panoff and Chuiton, 2004). There are three ways by which prokaryotes can get DNA from unrelated organisms: transduction, conjugation and transformation.

Transduction is a process by which DNA is moved from one organism to another by a virus. When a bacteriophage infects a cell it may integrate its DNA into the bacterial chromosome and get it replicated and translated with the cellular machinery. New viral particles would ideally include a copy of all the bacteriophage genes, but DNA packaging within the viral particle is not perfect. Therefore viral particles can take up DNA from the bacteria, carry it to another bacteria and get it inte-

grated into the new chromosome (Kokjohn *et al.*, 1989). Transduction by means of virus-like particles might be of great importance in aquatic ecosystems (Chiura *et al.*, 2000; Jiang and Paul, 1998). The process is schematized in Fig. 1.1.

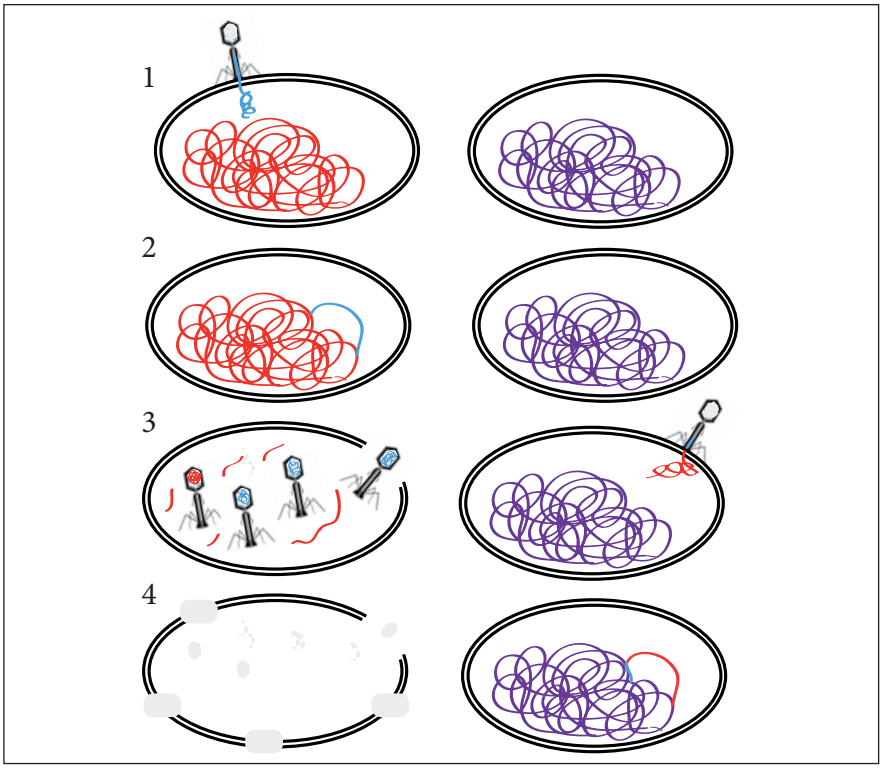


Fig. 1.1 Schematic representation of transduction. A virus infects the donor cell (1). The donor cell integrates viral DNA within its genome (2). The virus forms new viral particles that assemble capsids around viral DNA but also around bacterial DNA. A virus containing DNA from the donor cell infects a host cell (3). The host cell integrates viral DNA within its genome (4).

Conjugation is a process by which DNA is moved from one organism to another through direct cell-to-cell contact or through the establishment of a conjugation bridge. The best studied conjugative mechanism is the transfer of the F-plasmid in *E. coli*. This process has been reviewed by Llosa *et al.* (2002). The process is schematized in Fig. 1.2.

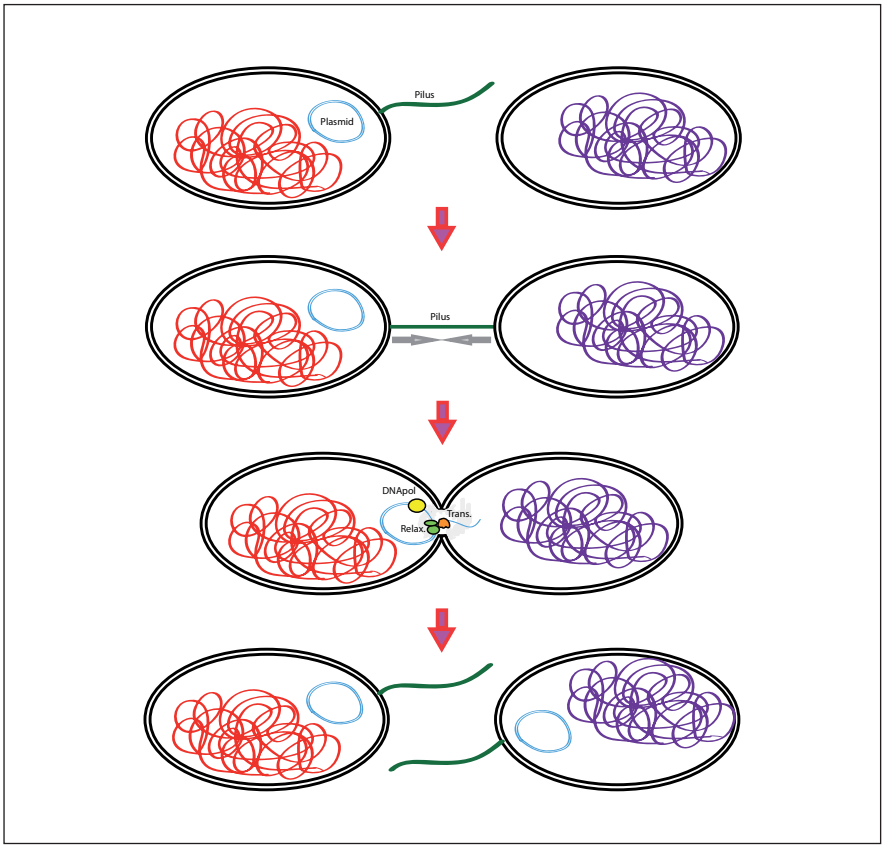


Fig. 1.2 Schematic representation of conjugation. A piliated donor cell containing a plasmid pilus gets closer to a host cell (1). The donor cell makes contact with the host cell through the pilus (2). The donor cell introduces DNA into the host cell through a DNA rolling-circle replication system linked to a type IV secretion system (T4SS) (3). Both cells become donor cells (4).

Transformation does not require direct contact between cells or between cells and viruses but consists of the acquisition of naked DNA from the environment (Chen *et al.*, 2005). Bacteria that are capable of being transformed, either naturally or artificially, are called competent. Transformation is explained in detail in the next section. The process is schematized in Fig. 1.3.

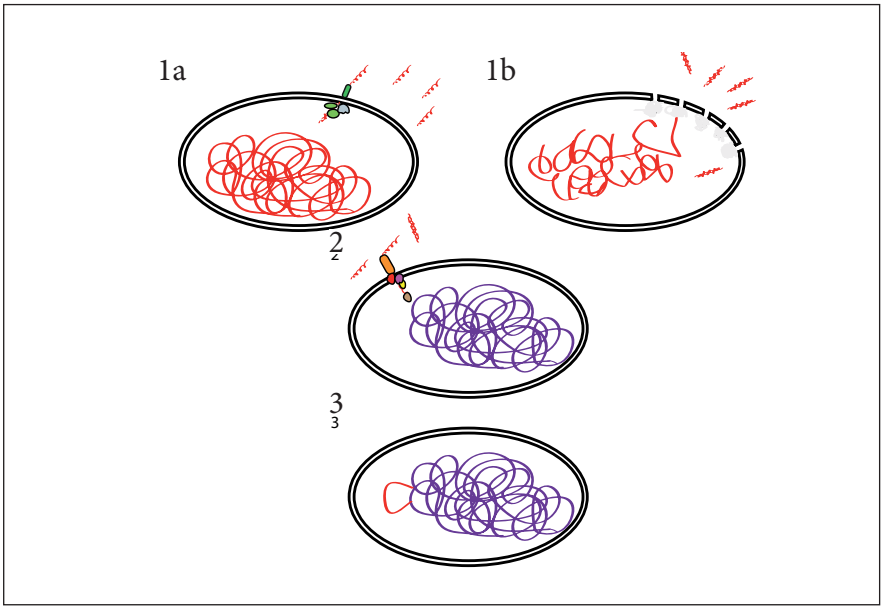


Fig. 1.3 Schematic representation of transformation. A donor cell donates DNA through a type IV secretion system (T4SS) or any other active secretion mechanism (1a) or through lysis (1b). The host cell takes DNA from the environment in a process later known as DNA uptake (2). The host cell integrates incoming DNA within the genome in a process later referred to as DNA processing (3).

1.2.3 Natural transformation

1.2.3.1 On the discovery of natural transformation

The term "transformation" was coined by Fred Griffiths in 1928. By experimenting with vaccine candidates against pneumonia in mice, Griffiths realized that some *S. pneumoniae* strains that were not dangerous could become dangerous after being exposed to dead or alive dangerous strains. Griffiths called that process "transformation" because a "transforming principle" had transformed non-dangerous R form (rough form) pneumococci into dangerous S form (smooth form) pneu-

mococci (Griffiths, 1928). The experiments were repeated many times, but the process remained obscure and unexplainable for more than a decade.

The nature of the transforming principle was discovered by Avery, McCarty and MacLeod in 1944. A brilliant experiment that consisted of the purification of *S. pneumoniae* extracts to test their ability in transforming R cells into S cells proved that DNA was the chemical responsible for the conversion of type of pneumococcus (Avery *et al.*, 1944). This experiment was also used as a proof for the definition of DNA as the biomolecule responsible for carrying genetic information and marked the beginning of molecular genetics.

Natural transformation was soon found in other organisms. Nowadays we know that more than 60 bacterial species distributed among distant taxonomic groups are susceptible to undergo natural transformation, but the number of species might be higher (Johnsborg *et al.*, 2007).

In most organisms transformation is a tightly regulated process because competence (*i.e.* the ability to take up naked DNA from the environment) is only developed under very specific environmental or physiological conditions (Chen and Dubnau, 2004). *N. gonorrhoeae* is an exception to the rule because it is constitutively competent and can take DNA coming from neighbouring gonococci without depending on any apparent environmental or physiological factor (Hamilton and Dillard, 2006).

1.2.3.2 Natural transformation in *N. gonorrhoeae*

N. gonorrhoeae and other species belonging to the genus *Neisseria* are well known for their variability. *N. gonorrhoeae* populations are panmictic, meaning that recombination events are so frequent that there is random association between loci (Smith *et al.*, 1993).

Apart from being a major source of variation, natural transformation can have additional roles. Incoming DNA might be used as a source of food (Palchevskiy and Finkel, 2006), but also as template for DNA repair (Hamilton and Dillard, 2006). Introducing variation to avoid variation might sound paradox, but in some cases, quoting Lampedusa's The Leopard, "*If we want things to stay as they are, things will have to change*". The combination of transformation and recombination events might provide an effective mechanism to purge the genome from deleterious mutations.

Understanding the molecular details of natural transformation in *N. gonorrhoeae* will provide more clues about its meaning and functions. Next subsections describe the molecular machinery involved in DNA donation, DNA uptake and DNA processing in *N. gonorrhoeae* transformation.

DNA donation

A peculiarity of *N. gonorrhoeae* transformation is that DNA coming from neighboring gonococci is preferred over DNA coming from other organisms. This preference is mainly achieved by taking preferentially DNA containing a specific sequence called DUS (5'-GCCGTCTGAA-3') spread over the genome (Hamilton and Dillard, 2006). This mechanism is shared with other members belonging to the *Neisseria* genus that share DUS with *N. gonorrhoeae* but also with *Haemophilus influenzae* and *Actinobacillus actinomycetemcomitans* that preferentially take DNA containing a sequence called UUS (5'-AAGTGCGGT-3') (Davidsen *et al.*, 2003). Transformation with DNA coming from other species is also possible but seems to be restricted to a few examples. In particular, it has been documented genetic exchange between *N. gonorrhoeae* and *H. influenzae* (Kroll *et al.*, 1998).

N. gonorrhoeae donates DNA at least by two mechanisms: cell death/lysis and active DNA secretion (Hamilton and Dillard, 2006). The first

mechanism for DNA donation involves cell death and lysis. *S. pneumoniae* competent cells have an active killing mechanism known as "fratricide" that is used to release DNA from non-competent pneumococci (Claverys *et al.*, 2007). No similar mechanisms have been described in *N. gonorrhoeae*, but *N. gonorrhoeae* can produce autolysins that promote its own cell lysis during stationary phase or under unfavorable conditions (Dillard and Seifert, 1997). The second mechanism is DNA secretion. *N. gonorrhoeae* has the ability to secrete large amounts of DNA during the log phase without significant cell death. Secreted DNA is an important component of gonoccocal biofilms, but is also efficient for natural transformation (Salgado-Pabon *et al.*, 2010).

The molecular details of DNA secretion are largely unknown. The analysis of *N. gonorrhoeae* DNA sequences has revealed that many pathogenic isolates contain the gonococcal genetic island (GGI), a mobile genetic element that encodes a putative type IV secretion system (T4SS) that would transport DNA in a similar manner to that used in conjugation (Hamilton *et al.*, 2005).

The system for DNA secretion is formed by putative partitioning proteins that bring the DNA to the secretion machinery (ParA and ParB), putative relaxases and helicases (TraI and Yea) that nick and unwind the DNA, a putative coupling protein that couples the relaxosome to the DNA transport apparatus (TraD), a number of periplasmic proteins that make breaks in the peptidoglycan layer for the insertion of the DNA transport machinery (Yag, TraW, TraF, TraH, and AtlA), a putative ATPase (TraC) and a number of proteins that make a transmembrane channel that would span both membranes (TraD, TraB, TraK, TraV, and TraN) (Ramsey *et al.*, 2011). The machinery might also allow the secretion of companion DNA-binding proteins that would protect DNA from eventual degradation or would even aid in the uptake process (Hamilton *et al.*, 2005).

With this information coming from sequence analysis it is tempting to think about a common origin for both means of horizontal gene transfer. Further biochemical and structural studies will eventually provide interesting information about function and evolution of DNA transport machineries in bacteria, and might relate them to other secretion machineries.

DNA uptake

The mechanisms by which DNA is taken up into the cell during transformation are poorly understood. Fig. 1.4 shows a schematic diagram of DNA uptake in *N. gonorrhoeae*.

Most studies suggest that in *N. gonorrhoeae* the best substrate for DNA transformation is dsDNA which is processed during its uptake and converted into ssDNA that enters the cytoplasm (Hamilton and Dillard, 2006). However, *N. gonorrhoeae* can take ssDNA as well, albeit with less efficience (Miao and Guild, 1970). Due to the low efficiency of ssDNA uptake with respect to dsDNA uptake, dsDNA is assumed to be the main substrate for DNA transformation, but this might change if further studies confirm that ssDNA secretion by T4SS systems is the main source of DNA release (Ramsey *et al.*, 2011).

As mentioned before, DNA uptake is highly enhanced by the presence of DUS sequences in incoming DNA (Elkins *et al.*, 1991). DNA is thought to bind in a specific manner to a receptor located in the outer membrane, but the protein or proteins that would provide such specificity are unknown. The role of the DUS may be variable between strains and influence multiple steps during transformation (Duffin and Seifert, 2010).

Among the proteins required for efficient transformation, type IV pili are the best studied components (Aas *et al.*, 2002). Type IV pili (Tfp) are surface-exposed helical fibers that are involved in many functions

related to *Neisseria spp.* pathogenesis (Craig *et al.*, 2006). In addition to their role in horizontal gene transfer, they are also important for cellular adhesion (Plant and Jonsson, 2006) and provide a mechanism to fight against the host immunitary system due to the fact that the main component of the pilus, the pilin PilE, undergoes extensive antigenic variation (Criss *et al.*, 2009).

Pilus assembly and disassembly are dynamic processes that require the concerted action of many proteins. Pilin subunits are inserted into the inner membrane and are processed by the pre-lipin peptidase PilD. Additional major and minor pilin subunits are assembled to the nascent fiber and are arranged in three helices that twist around each other and form positively charged grooves that have the potentiality to bind DNA or negatively charged surfaces in a non-specific manner (Craig *et al.*, 2006). PilF and PilT are hexameric ATPases that interact with auxiliary proteins and provide energy for the protrusion and retraction of the pilus (Maier *et al.*, 2004). The pilus passes the outer membrane through PilQ, a member of a secretin superfamily present in type II and type III secretion systems (Collins *et al.*, 2011). PilQ interacts with the pilus and with other proteins such as the pilot protein PilP (Balasingham *et al.*, 2007), PilW (Szeto *et al.*, 2011), the adhesins PilC1 and PilC2 (Jain *et al.*, 2011), and the lipoprotein ComL which interacts with PilQ and DNA (Benam *et al.*, 2011). PilG is another component of the pilus system. PilG is integrated in the inner membrane and may play a role in binding to other pilus biogenesis proteins in the cytoplasm and the periplasm and provide a link between both compartments (Collins *et al.*, 2007). The expression of components of *N. gonorrhoeae* Dsb system is required for PilQ folding and consequently for efficient DNA binding and uptake (Sinha *et al.*, 2008).

Some authors support that the assembly of pili fibers is not necessary for efficient transformation and have proposed that DNA uptake takes

place through a competence pseudopilus formed by PilE, PilD, PilF, PilG, PilQ, PilP and PilC (Chen and Dubnau, 2004).

DNA transfer throughout the periplasm is aided by auxiliary proteins such as ComE (Chen and Gotschlich, 2001). Interestingly, ComE is very similar to the C-terminal end of ComEA from *Bacillus subtilis*, a DNA receptor located that plays an important role in DNA attachment to the microorganism surface (Provvedi and Dubnau, 1999). *N. gonorrhoeae* and *N. meningitidis* genomes contain four identical copies of the gene *comE*. The deletion of a single copy of *comE* does not have any effect on transformation, but the deletion of two or more copy leads to a gradual decrease in transformation efficiency. ComE has unspecific DNA binding activity and a putative periplasmic signal peptide which is coherent with a putative role in DNA transport across the periplasm (Chen and Gotschlich, 2001).

Proteins putatively involved in cell division such as Tpc and Dca are also needed for efficient transformation. Tpc seems to play a role in murein hydrolysis (Fussenegger *et al.*, 1996) whereas Dca is encoded by a gene located in *N. gonorrhoeae dcw* cluster (Division Cell Wall cluster) (Snyder *et al.*, 2001).

Incoming DNA enters the cytosol through a transmembrane protein known as ComA (Fascius and Meyer, 1993). Little is known about *N. gonorrhoeae* ComA. Bioinformatic and biochemical assays performed with the homologous protein ComEC from *B. subtilis* suggested that this protein oligomerizes through its C-terminal end and spans the internal membrane through seven transmembrane helices (Draskovic and Dubnau, 2005). The activity of ComEC is regulated by ComFA, a membrane protein that collocalizes at the cellular poles with ComGA, Ssb and DprA/Smf (Kramer *et al.*, 2007). The activity of ComA might be regulated by other proteins located in the membrane and in the cytosol, but there are no studies in this respect.

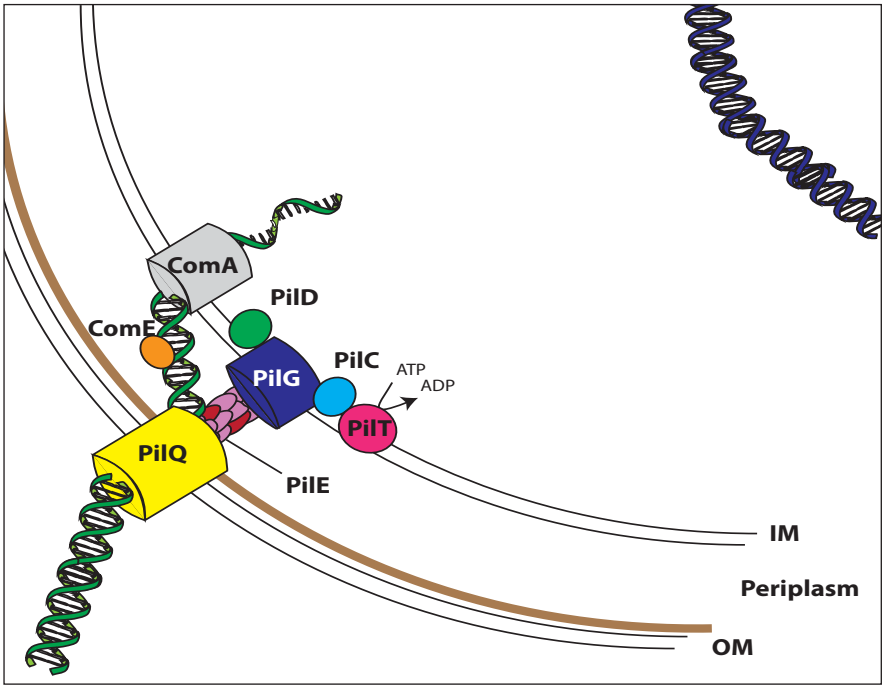


Fig. 1.4 DNA uptake of transforming DNA in *N. gonorrhoeae*. Pilins are assembled to form a competence pseudopilus in a process dependent on PilC, PilD, PilT and PilG. Then, dsDNA is channelled across the outer membrane through PilQ. In the periplasm, dsDNA binds to ComE ComE. A strand gets degraded and the other enters the cytoplasm through ComA.

DNA processing

Once in the cytosol, incoming DNA needs to be protected from cellular nucleases or restriction modification systems and integrated in the bacterial chromosome. This complex mechanism can be referred as a whole as DNA processing.

DNA processing has been best studied in *S. pneumoniae* and *B. subtilis.* Initial studies performed with *S. pneumoniae* (Fox, 1960) and *B. subtilis* (Venema *et al.* 1965) revealed that transforming DNA introduced into

competent cells underwent an "eclipse phase" characterized by transient loss of donor-marker transforming activity. Since ssDNA was a poor substrate for DNA uptake this behavior was attributed to the conversion from dsDNA to ssDNA. During the eclipse phase ssDNA was embedded in a nucleoprotein complex that conferred protection against cellular nucleases (Morrison and Mannarelli, 1979).

Recent investigations have provided more information about the nature of the nucleoprotein protein complex and have linked its formation to the integration of incoming DNA into the chromosome. DNA integration would require the concerted action of DprA and RecA, two proteins that are essential for transformation (Fig. 1.5). DprA and RecA would bind to ssDNA and would drive the formation of nucleofilaments that would catalyze the homology-dependent formation of joint molecules (Mortièr-Barriere, 2007).

The high similarity between *N. gonorrhoeae dprA* and its homologous counterparts from *S. pneumoniae* and *B. subtilis* suggests similar mechanisms for DNA processing and recombination are shared across species.

It is interesting to point out that *dprA* homologues are found in many genomes, even in genomes of organisms that do not manifest natural transformation (Mortièr-Barriere, 2007). The presence of *dprA* in non-transformable species suggests that DprA might play additional roles in other processes or that transformation is a more widespread phenomenon than one might think. In any case, the debate about the role of bacterial competence genes in non-competent bacteria is wider (Claverys and Martin, 2003).

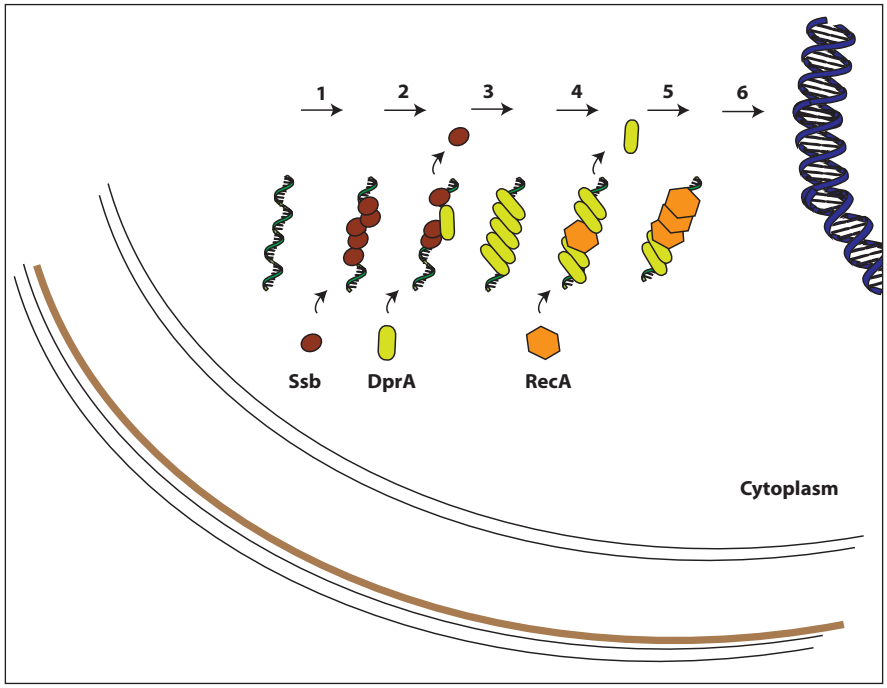


Fig. 1.5 DNA processing of transforming DNA in *N. gonorrhoeae*. Ssb covers incoming ssDNA (1), DprA displaces Ssb (2) and covers ssDNA (3). DprA recruits RecA (4) and RecA-DprA cover ssDNA forming nucleofilaments (5). The resulting nucleoprotein filament then searches out homologous sequences of DNA on the host chromosome (6).





1.3 Regulation of growth and death by toxin-antitoxin systems

Toxin-antitoxin systems have been introduced in section 1.1 in the context of persistance mechanisms. This section describes the molecular biology behind toxin-antitoxin systems. Toxin toxicity mechanisms and antitoxin regulatory mechanisms are discussed from a structural perspective. Similarities and differences among homologous systems are highlighted to point towards comprehensive models of toxin-antitoxin action. The end of the section is focused on RelBE systems and introduces the *S. pneumoniae* RelBE2 system.

1.3.1 An overview of toxin-antitoxin systems

Toxin-antitoxin (TA) systems are sets of closely related genes that encode a stable poisonous protein that can harm its host cell and a labile antidote that prevents the toxic action of the toxin (Van Melderen and Saavedra, 2009).

TA systems were first discovered in plasmids (Ogura and Higara, 1983; Gerdes *et al.*, 1985) and were called addiction modules because cells became addict to antitoxin production and therefore to the whole module (Engelberg-Kulka and Glaser, 1999). Being encoded in low-copy plasmids, it was hypothesized that TA systems could ensure the transmission of non-essential plasmids through post-segregational killing. By this mechanism, daughter cells that had inherited the plasmid could grow normally whereas daughter cells that had lost the plasmid would have lost the ability to produce antitoxin, toxin molecules would be released and cells would die (Gerdes el al., 1986).

However, TA modules were soon found in bacterial chromosomes (Tyndall *et al.*, 1997; Gronlund and Gerdes, 1999). High throughput sequencing of bacterial strains revealed that they were widespread. Rather

than being anecdotal, they were very abundant, especially in free-living organisms and pathogens (Pandey and Gerdes, 2005). The function of chromosomally encoded TA systems is not clear and still open to debate.

Chromosomal TA systems might be selfish entities; acquired through horizontal gene transfer and maintained within the genome thanks to their addictive properties (Mochizuki *et al.*, 2006). But TA systems might be more than selfish entities and confer selective advantage to their carriers (Van Melderen and Saavedra de Bast, 2009). Actually, TA systems might be involved in the stabilization of mobile genetic elements in the bacterial chromosome (Pandey and Gerdes, 2005), cell growth modulation (Christensen *et al.*, 2001), programmed cell death (Engelberg-Kulka *et al.*, 2006), persistence mechanisms (Maisonneuve *et al.*, 2011), development regulation (Nariya and Inouy, 2008) or even systems to escape from other addiction modules (Saavedra de Bast *et al.*, 2008) or from viral DNA (Villarreal, 2011). In my opinion, TA systems are diverse and diversity can be exploited to provide a rich set of physiological responses.

Toxin molecules are always protein molecules but the antidotes can be proteins or RNA molecules. Depending on the nature of the antidote and the inhibitory mechanism, toxin-antitoxin systems have been classified in three groups: type I, type II and type III (Van Melderen, 2010).

- **Type I systems:** antitoxin molecules are small non-coding RNAs complementary to the toxin mRNAs thus preventing their translation. Some examples are *hok/sok*, *fst/RNAII*, *tisB/lstR*, and *ldrD/rdID*. Type I toxin-antitoxin systems have been reviewed by Fozo *et al.* (2008).
- **Type II systems:** antitoxin molecules are labile proteins that prevent the toxic action of the toxins by tightly binding to it. Regulation can be achieved through the specific binding of the antitoxin or the toxin-antitoxin complex to a common promoter region located upstream of both genes (Fig. 1.6). Some examples are RelBE, MazEF,

ParDE and Kis/Kid. Predicted and experimentally detected toxin-antitoxin systems are integrated in the TADB database (Shao *et al.*, 2011). Type II TA systems are reviewed in Leplae *et al.* (2011).

- **Type III systems:** antitoxin molecules are non-coding RNAs that prevent the action of the toxin by specifically binding to them. Type III systems are largely unknown, with only one representative known so far, the ToxIN system, an abortive infection system encoded in a cryptic plasmid from the bacteria *Erwinia carotovora* (Fineran *et al.*, 2009).

In TA systems the toxic component is a stable protein that targets an essential cellular process. Therefore it is not surprising that the artificial overexpression of the toxin generally causes growth arrest or cell death (Yamaguchi and Inouye, 2011). According to their targets, toxins can be distributed in seven groups:

- **DNA gyrase poisons** that target the DNA gyrase and generate double-strand breaks such as CcdB (De Jonge *et al.*, 2010) and ParE (Jiang *et al.*, 2002).
- **Ribosomal-dependant mRNases** that cut mRNA in the translating ribosome such as RelE (Neubauer *et al.*, 2009) and HigB (Hurley and Woychik, 2009).
- **Ribosomal-independant mRNases** that cut mRNA without being associated to the ribosome such as MazF (Zhang *et al.*, 2005).
- **RNases** that target tRNAs and induce their cleavage such as VapC (Winther and Gerdes, 2011).
- **Phosphorilases** that target EF-Tu and inhibit translation such as HipA (Schumacher *et al.*, 2009).
- **Doc toxins** that target the translating ribosome and inhibit elongation (Liu *et al.*, 2008).
- **Proteins of unknown targets** such as the Zeta toxin that can inhibit replication, transcription and translation through an unknown mechanism (Lioy *et al.*, 2006).

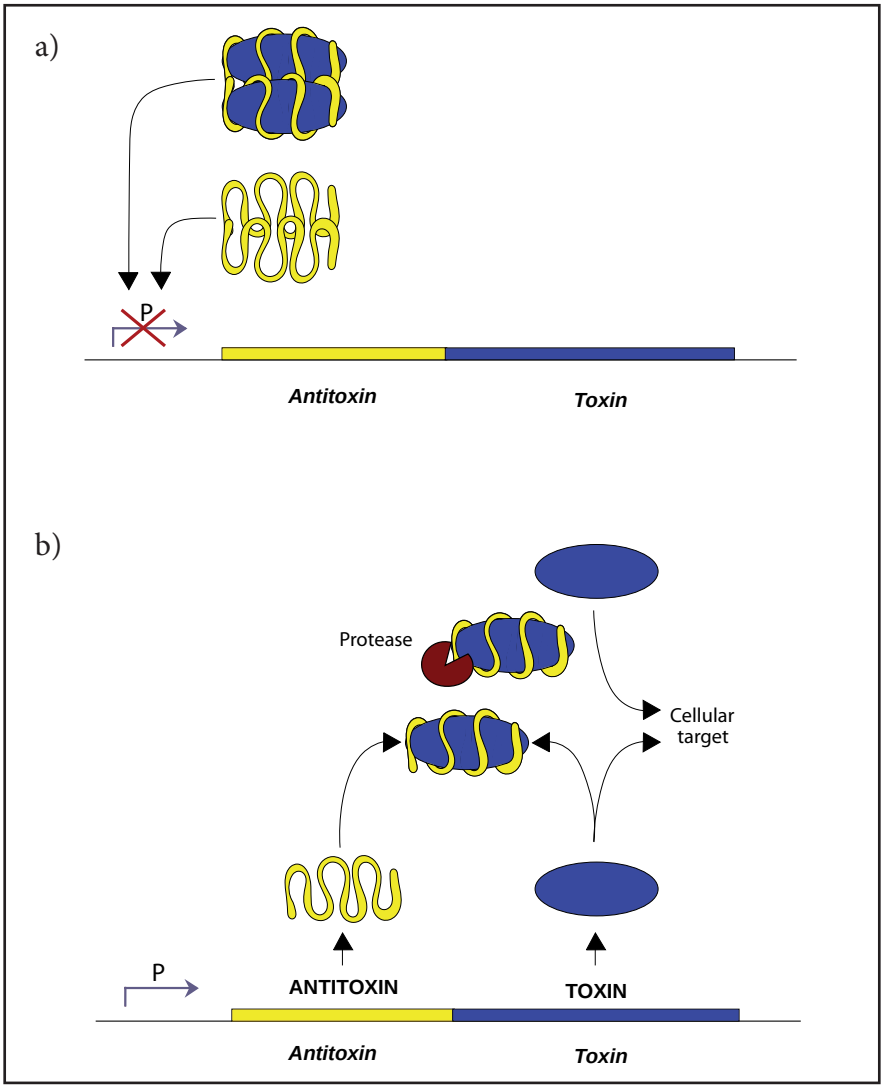


Fig. 1.6 Regulation of a typical type II TA system. Toxin and antitoxin genes are encoded in an operon regulated under a single promoter. a) In a steady state, either the toxin/antitoxin complex or the antitoxin/antitoxin complex bind to the promoter region and repress the synthesis of both components. b) If specific proteases are activated, the antitoxin gets degraded. Transcription gets de-repressed and the cell produces both components. Being freed from its partner antitoxin, toxin can target its cellular target.

1.3.2 Structural biology of TA systems

The structural characterization of TA systems has provided useful information to understand the molecular details behind toxin and antitoxin action. This section summarizes the state of the art of structural biology of TA systems.

The structure of type I systems will not be reviewed in this section. Just as a comment, the structural rearrangements that allow translation and antisense RNA binding in type I systems have been reviewed by Gerdes *et al.* (1997).

The structural characterization of type II and type III TA systems has posed technical difficulties due to the toxicity of the toxin component and to the intrinsic lability of the antitoxin component. These problems have been tackled by combining coexpression and directed mutagenesis techniques, and by combining X-ray crystallography with other structural determination techniques such as NMR spectroscopy. Table 4.1 summarizes toxin-antitoxin systems structurally characterized by X-ray or by NMR by September 2011.

Toxin-antitoxin systems have been best studied in *E. coli* (reviewed in Yamaguchi and Inouye, 2011). Structural information we have so far comes mostly from the structural characterization of *E. coli* toxin-antitoxin systems. Given the technical difficulties mentioned above, it is not surprising that most information comes from toxin-antitoxin complexes, which are less toxic and more stable than toxins or antitoxins without their counterpart.

From a functional point of view, it has been particularly insightful the information coming from the structural characterization of toxins in complex with their cellular target or antitoxins which are in complex with DNA operator regions.

Table 1.1 Toxin-antitoxin systems

Name	Description	PDB
Escherichia coli		
MqsA/MqsR	Antitoxin C-terminal domain	3FMY (X-ray)
	Antitoxin N-terminal domain	3GA8 (X-ray)
	Antitoxin full-length	3GN5 (X-ray)
	Antitoxin with promoter DNA	3O9X (X-ray)
	Toxin-antitoxin complex	3HI2 (X-ray)
RelB/RelE	Antitoxin DBD domain	2K29 (NMR)
	Toxin	2KC9 (X-ray)
	Toxin with antitoxin fragment	2KC8 (NMR)
	Toxin complex with 70S ribosome	3KIU, 3KIW, 3KIX, 3KIY (X-ray)
CcdA/CcdB	Antitoxin	2ADL, 2ADN
	Antitoxin with promoter DNA	2H3A, 2H3C
	Toxin-antitoxin complex	3G7Z, 3HPW
	Toxin complex with gyrase	3KU8, 1X75 (X-ray)
Kis/Kid	Toxin	1M1F (X-ray)
	Toxin with a an RNA substrate	2C06 (NMR)
YefM/YoeB	Toxin	2A6R, 2A6S (X-ray)
	Toxin-antitoxin complex	2A6Q (X-ray)
ParD/ParE	Antitoxin	2AN7 (NMR)
YeeV/YeeU	Toxin	2H28 (X-ray)
HigA/HigB	Antitoxin	2ICT (X-ray)
MazE/MazF	Toxin-antitoxin complex	1UB4
Mycobacterium tuberculosis		
RelB/RelE	Toxin-antitoxin complex	3G5O (X-ray)
YefM/YoeB	Antitoxin	3CTO, 3D55 (X-ray)
RelJ/RelK	Toxin-antitoxin complex	3OEI (X-ray)
VapBC (Rv0301/ Rv0300)	Toxin-antitoxin complex	3H87 (X-ray)
VapBC-5	Toxin-antitoxin complex	3DBO (X-ray)
Neisseria gonorrhoeae		
FitA/FitB	Toxin-antitoxin complex	2H1C (X-ray)
	Toxin-antitoxin complex with DNA	2H1O, 2BSQ (X-ray)

Table 1.1 Toxin-antitoxin systems

Name	Description	PDB
Phage P1		
PhD-Doc	Toxin-antitoxin complex	3K33, 3KH2 (X-ray)
	Phd truncation	3HS2 (X-ray)
Thermus thermophilus		
RelB/RelE	Toxin RelE	2KHE (NMR)
Pyrococcus horikoshii		
RelB/RelE	Toxin-antitoxin complex	1WMI (X-ray)
Cytophaga hutchinsonii		
CHU_2935	Antitoxin CHU_2935	3OMT (X-ray)
Shigella flexneri		
Q83JN9	Antitoxin Q83JN9	2INW (X-ray)
Erwinia carotovora		
ToxIN	Toxin-antitoxin (RNA) complex	2XDD (X-ray)
Pyrobaculum aerophilum		
VapBC	Toxin PAE0151 (VapC)	2FE1 (X-ray)
Methanococcus janaschii		
RelB/RelE	Toxin-antitoxin complex	3BPQ (X-ray)
Caulobacter crescentus		
ParD/ParE	ParDE complex	3KXE (X-ray)
Syntrophomonas wolfei		
SWOL_0700	Putative toxin	3K6Q (X-ray)
Streptococcus pneumoniae		
PezA/PezT	Toxin-antitoxin complex	2P5T (X-ray)
Streptococcus pyogenes		
Epsilon-Zeta	Toxin-antitoxin complex	3Q8X (X-ray)
Bacillus subtilis		
SpoIISA/SpoIISB	Toxin-antitoxin complex	3O6Q (X-ray)
Vibrio fischeri		
CcdB/CcdA	Toxin CcdB	2KMT (NMR)

The structural characterization of toxins in complex with their cellular targets is a difficult task because it involves working with labile substrates or with large macromolecules. Some examples of toxins characterized in complex with their cellular target are *E. coli* RelE in complex with the ribosome subunit 70S (Neubauer *et al.*, 2009), *E. coli* CcdB in complex with a gyrase domain (Dao-Thi *et al.*, 2005), and *E. coli* Kid in complex with RNA substrate (Kamphuis *et al.*, 2006). Some examples of antitoxins or toxin-antitoxin complexes with DNA operator regions are *E. coli* MqsA (Brown *et al.*, 2011), *E. coli* CcdA (Madl *et al.*, 2006) and *N. gonorrhoeae* FitAB (Mattison *et al.*, 2006).

1.3.2.1 Structural insights into toxicity mechanisms

The structural characterization of toxins has revealed that despite their low sequence identity, most of them share a common fold characterized by a central β-sheet made of 4-6 β–strands surrounded by 3-4 α-helices (Fig. 4.7).

The structural characterization of toxins in complex with their substrates has provided information about the residues involved in gyrase poisoning in the case of CcdB (Dao-Thi *et al.*, 2005) and about the catalytic mechanisms behind RelE (Neubauer *et al.*, 2009) and Kid function (Kamphuis *et al.*, 2006).

It is interesting to point out that homologues from distant organisms keep conserved structures, but also that toxins carrying out different functions are structurally similar (Fig. 4.7). Similar folds are used to carry out gyrase poisoning (CcdB), mRNA cleavage (RelE, MqsR, YoeB, Kid) or tRNA cleavage (VapC). RNA cleaveage mechanisms are similar to those of other RNases (Neubauer *et al.*, 2009; Fineran *et al.*, 2009).

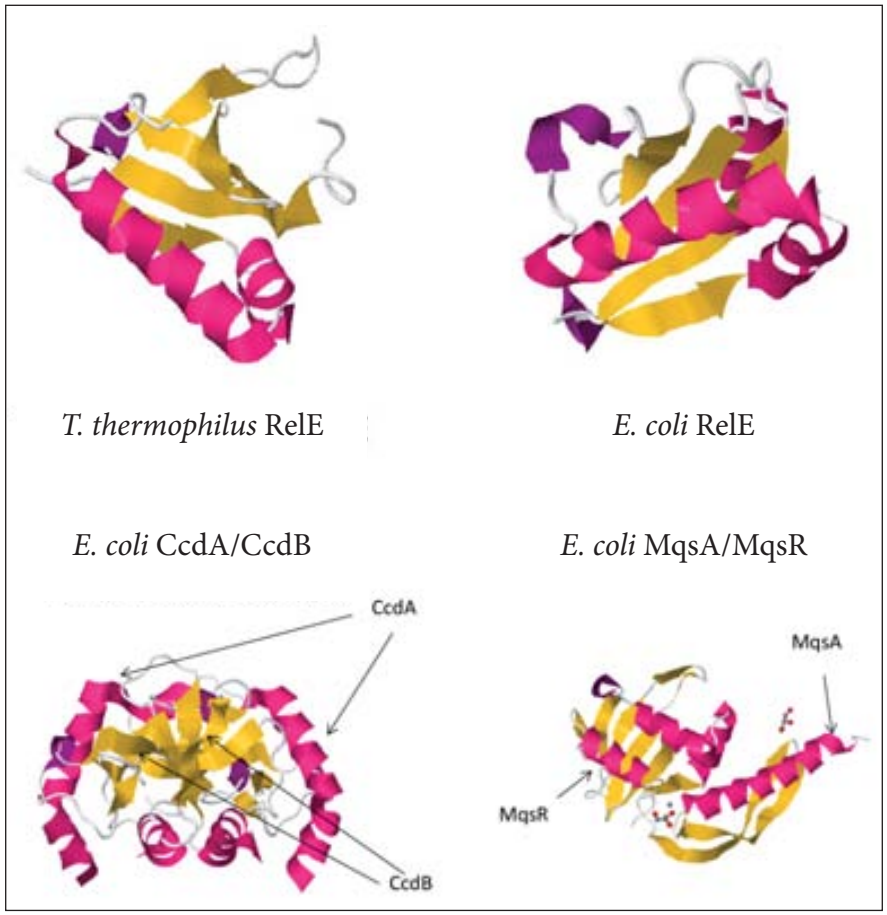


Fig 4.7 Toxin fold homologies a)RelE toxins from two distant homologues share the same fold. b) Toxins with different functions share the same fold.

1.3.2.2 Structural insights into antitoxin inhibitory mechanisms

Most antitoxins belonging to type II systems are mainly α–helical proteins that inhibit the toxin by wrapping around it (Fig. 4.8). Contrary to toxins, antitoxins show a range of different 3D structures and conformations even within closely related homologues. The interaction between toxin and antitoxin molecules is generally driven by electrostatic

interactions and hydrogen bond formation (Yamaguchi *et al.*, 2011). Since toxin activity requires the interaction with negatively charged substrates such as DNA and RNA, most toxins contain many aminoacids with positively charged side chains. Not surprisingly, these aminoacids interact as well with negatively charged surfaces of antitoxin molecules.

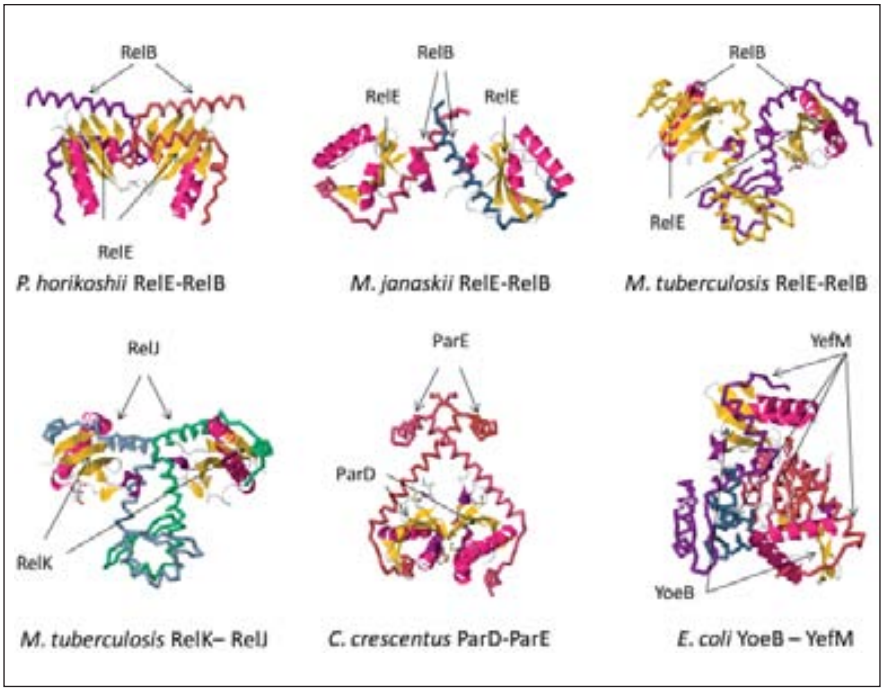


Fig 4.8 Antitoxins are mainly α-helical structures that wrap around the toxin in several ways.

The most frequent oligomeric states in toxin-antitoxin crystals are quaternary complexes in which one antitoxin binds to one toxin, and one complex toxin-antitoxin binds to another complex toxin-antitoxin through the interaction between the antitoxin molecules. But in a few cases, such as in the YefM-YoeB pair, two antitoxin molecules bind to one toxin molecule, and the trimer binds to another trimer through

the interaction between the antitoxin molecules (Kamada and Hanaoka, 2005).

Since the structures and the inhibitory mechanisms are conserved, some antitoxins seem to be interchangeable. In addition to inhibit their own partner, antitoxins can actually bind to other toxins and prevent their toxicity. Antitoxin exchangeability might provide a mechanism to counteract addiction and defense against invader toxins (Arbing *et al.*, 2010) or exert complex cross-regulation effects (Yang *et al.*, 2010).

Besides regulating toxicity by binding to their partner toxin, type II antitoxins can repress the synthesis of the toxin-antitoxin pair by specifically binding to an operator region located upstream of both genes (Yamaguchi and Inouye, 2011). In addition, some antitoxins might play a role in the regulation of other genes. This is the case of MqsA that has been found to repress rpoS, reducing thus the concentration of the internal messenger 3,5-cyclic diguanylic acid, an important mediator of the environmental stress response (Xiaoxue *et al.*, 2011) and to repress the expression the toxin CspD.

The structural characterization of some toxin-antitoxin systems in complex with DNA has revealed that DNA binding is generally carried out through helix-turn-helix and ribbon-helix-helix DNA binding motifs that are common to other repressor proteins such as the Cro repressor (PDB:2OR1) and the Arc repressor (PDB:1PAR) respectively.

In proteins belonging to the MqsA family, the interaction occurs through both ends of the protein and requires the oligomerization of two antitoxin molecules and the relative rotation between the domains to accommodate and bend the DNA. The interaction is stabilized through unspecific contacts between the DNA phosphate backbone and positive residues located at the N-terminal end (Brown *et al.*, 2009).

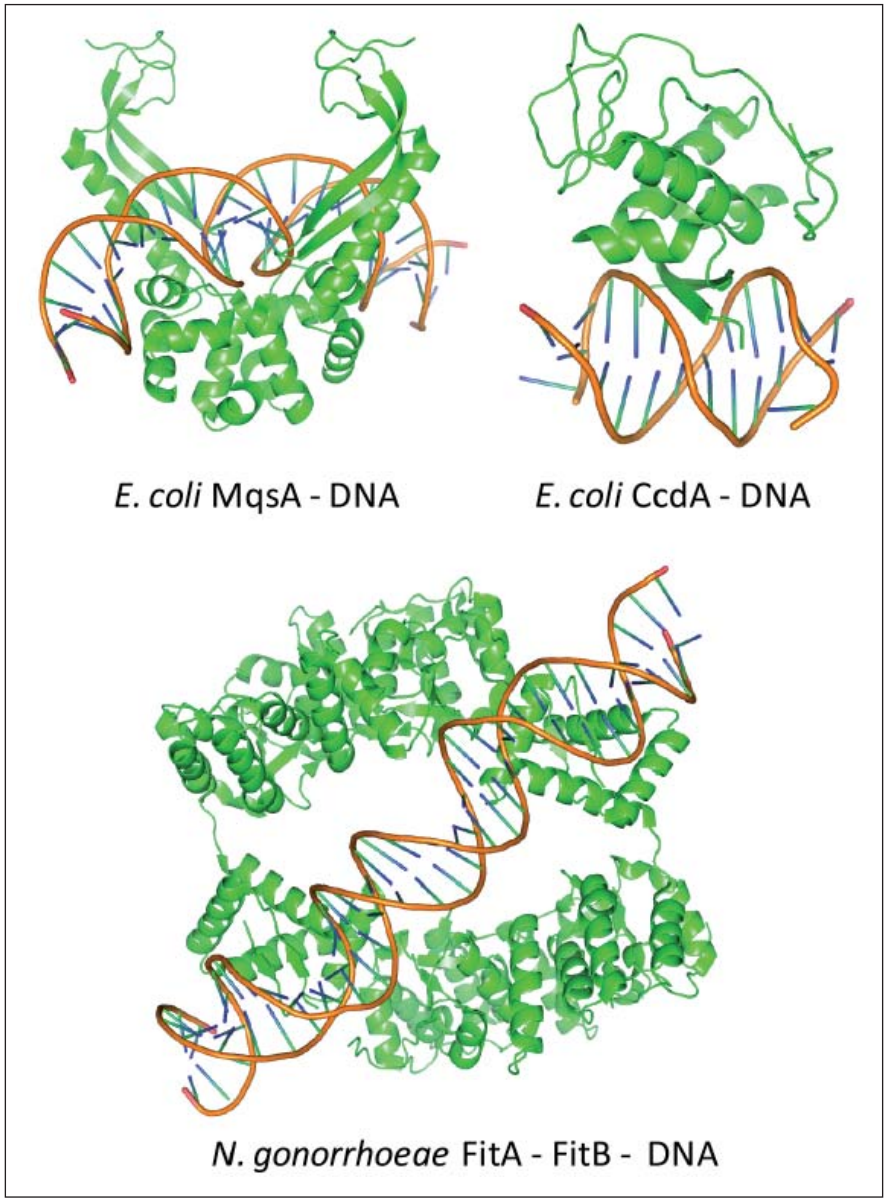


Fig. 4.9. Antitoxins bind to DNA through helix-turn-helix motifs (MqsA) and ribbon-helix-helix (CcdA, FitA).

Ribbon-helix-helix DNA binding motifs are found in many antitoxins including RelB (Overgaard *et al.*, 2009), YefM (Bailey and Hayes, 2009), ParE (Dalton KM and Crosson S., 2009), CcdA (Madl *et al.*, 2006) and FitA (Mattison *et al.*, 2006) (Fig. 4.9). In all cases, the DNA binding motif is formed by terminal residues located in two adjacent antitoxin molecules and the specific interaction occurs through apolar residues located in two anti-parallel β–strands that enter the major groove and positive residues in the helices that interact with the DNA phosphate backbone. Although two antitoxin molecules alone can bind DNA, DNA binding is more efficient when the antitoxins form complexes with their partner toxins, suggesting that the antitoxin may act as a co-repressor (Yamaguchi and Inouye, 2011).

1.3.3 The RelBE family

The *relBE* locus was first discovered in *E. coli* in 1998. *RelBE* was described as a 2-component system in which *relE* encoded a cytotoxin whereas *relB* encoded an antitoxin that prevented the toxic action of the *relE*-encoded toxin (Gotfredsen and Gerdes, 1998).

Members of the RelBE family were soon found to be widespread among bacteria and archea (Grønlund and Gerdes, 1999) and RelE was linked to a complex physiological response to aminoacid starvation that relied on the degradation of RelB by the Lon protease (Christensen *et al.*, 2001). RelE toxins were lately described to cleave mRNA in response to nutritional stress (Gerdes *et al.*, 2005).

The basis for mRNA recognition and cleavage by *E. coli* RelE was elucidated by combining data from the mutagenesis analysis and structural characterization of *Pyrococcus horikoshii* RelBE (Takagi *et al.*, 2005) and from the structural characterization of *E. coli* RelE in complex with *Thermus thermophilus* 70s ribosome (Neubauer *et al.*, 2009).

Site-directed mutagenesis in *P. horikoshii* RelE suggested that Arg85, Arg40, Leu48, Arg58 and Arg65 were involved in RelE toxicity (Takagi *et al.*, 2005). The structural characterization of RelE in complex with a ribosome confirmed that aminoacids important for RelE toxicity occupied equivalent positions in *E. coli* and *P. horikoshii.* In addition RelE was found to occupy the A site on the 30s subunit of the ribosome and cleave mRNA through a conserved mechanism that involved mRNA reorientation and activation for 2'-OH-induced hydrolysis (Neubauer *et al.*, 2009).

The structural characterization of several RelBE complexes has provided clues to understand the inhibitory mechanism of RelB by direct interaction with RelE toxic molecules (Takagi *et al.*, 2005), but the mechanisms behind the regulation of *relBE* transcription are still mysterious.

RelE activity depends on the transcription levels of the *relBE* operon and those levels are tightly autoregulated by RelB and RelE. RelB regulates transcription by binding to the promoter region and there is the hypothesis that, depending on the RelB:RelE ratio, the toxin RelE can act either as a co-repressor or as a de-repressor of transcription. On a steady state, RelB and RelE form tight complexes that bind to the *relBE* operator region, repressing thus its expression. Upon certain conditions, the expression of specific proteases such as the Lon protease alters the balance between RelB and RelE, the operator region is released and RelE molecules can exert their action.

The *relBE* operator region (*relO*) has been characterized in *E. coli* and more recently in *S. pneumoniae.* In *E. coli*, the *relBE* operator (*relO*) consists of two operator sites of 6 bp inverted repeats located within each leg of a larger 26 bp inverted repeat. RelB-RelE heterotrimeric complexes bind cooperatively to two operator sites in *relO* (Overgaard *et al.*, 2008). Mutational and footprinting analyses show that two RelB dimers bind to two consecutive operator sites on the same face of the

DNA helix and that the spacing between each half-site is essential for successful repression (Overgaard *et al.*, 2009). The mechanism is similar in *S. pneumoniae* (Moreno-Cordoba *et al.*, 2012).

Structural analysis has provided hints regarding the mode of interaction of RelB with DNA, which is dependant on the formation of a ribbon-helix-helix motif at N-terminal (Overgaard *et al.*, 2009), but the role of RelE as a co-repressor still remains elusive. RelE forms tight complexes with RelB but unfortunately, the tetrameric complexes found in RelBE crystals from *Pyrococcus horikoshii* (Takagi *et al.*, 2005), *Methanococcus janaschii* (Frankuski and Saenger, 2009) and *Mycobacterium tuberculosis* (PDB 3GO) are not compatible with the abovementioned co-repression model that would require the formation of two heterotrimers in order to bind to two consecutive operator sites in *relO*.

1.3.3.1 *S. pneumoniae* TA systems - RelBE2

S. pneumoniae genome encodes at least eight putative TA systems: *relBE1, relBE2, yefMyoeB, higAB, phd/doc, pezAT, tasAB,* and *hicAB.* Only three of them -*yefM-yoeB, pezAT* and *relBE2*- are genuine TA systems (Nieto *et al.*, 2010).

S.pneumoniae YefM-YoeB function and regulation have been investigated by Nieto *et al.* (2007) and Chan *et al.* (2011) respectively. *S.pneumoniae* PezAT function and regulation have been investigated by Khoo *et al.* (2007), and Mutschler *et al.* (2010, 2011).

S. pneumoniae RelBE2 function has been investigated by Nieto *et al.* (2007, 2010) and Moreno-Córdoba *et al.* (2012). RelBE2 appears to be a *bona fide* TA system that resembles *E. coli* RelBE related TA systems. Exposure of *E. coli* cells to *S. pneumoniae* RelE toxin results in cell growth arrest, which can be rescued by induction of *S. pneumoniae* RelB antitoxin if exposure is short, but not if it is too long. The activation of *S. pneumoniae* RelE subsequent to protein synthesis inhibition

by amino acid starvation or antibiotic exposure leads to a slower rate of cell growth when compared to mutant strains lacking of functional RelBE2 system suggesting that under unfavourable conditions RelBE2 contributes to modulate the survival response through stasis (Nieto *et al.*, 2010).

Experimental data obtained by footprinting analysis and ultracentrifugation assays suggest that the regulatory mechanism for *S. pneumoniae* repression is similar to that of *E. coli relBE* (Moreno-Córdoba *et al.,* 2012). Footprinting assays show that *S. pneumoniae* RelBE2 would bind to a 30 bp DNA sequence located upstream of *relBE2* start codon (Moreno-Córdoba *et al.,* 2012). This information suggests that two RelB dimers bind to two consecutive operator sites located on the same face of the DNA helix. Ultracentrifugation assays of RelBE2 alone or in complex with the operator region delimited by footprinting assays support that RelBE2 in solution is in a balance between heterotrimers and heterohexamers (being the latter the most abundant species), and that the addition of DNA stabilizes the formation of heterohexamers (Moreno-Córdoba *et al.*, 2012).



Chapter 2
Objectives





Objectives

The objective of this thesis is twofold. On the one hand, it aims at characterizing ComE, ComA and DprA, three proteins involved in natural transformation in *Neisseria gonorrhoeae*. On the other hand, it aims at determining the X-ray crystallographic structure of *Streptococcus pneumoniae* RelBE2 toxin-antitoxin system and at providing insights into its function and regulation.

In order to characterize *Neisseria gonorrhoeae* ComE, ComA and DprA, we set the following intermediate objectives:

- To clone full-length genes and truncations of *comE, comA* and *dprA* into suitable expression vectors.
- To perform small-scale expression and purification experiments under different conditions so as to define the best clones for each target.
- To optimize expression and purification protocols for large-scale production of the best clones.
- To perform crystallization trials and biochemical assays with the best clones.
- To provide a functional interpretation of the results.

In order to characterize *Streptococcus pneumoniae* RelBE2, we set the following intermediate objectives:

- To obtain diffraction quality RelBE2 crystals.
- To collect and process datasets from native and Se-Met labelled protein crystals.
- To determine the X-ray crystallographic structure of RelBE2.
- To provide models for RelBE2 DNA binding and regulation.





Chapter 3
Materials & Methods





3.1 An overview of methods

Structural biology is a branch of molecular biology dedicated to the elucidation of macromolecular structures and functions. Structural biology is an interdisciplinary and collaborative science that involves a range of complementary methods such as X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy, cryo-electron microscopy, single molecule biophysics techniques (*e.g.* fluorescence detection, atomic force microscopy (AFM), and optical and magnetic trapping), and other scattering techniques (*e.g.* static light scattering (SLS), dynamic light scattering (DLS), small-angle X-ray scattering (SAXS), and small-angle neutron scattering (SANS)). Among these techniques, X-ray crystallography is by far the most used. Indeed, most structures included in the PDB archive have been determined by X-ray crystallography.

Good introductions to the field of X-ray crystallography for readers with no prior knowledge of the field or its mathematical basis are "Crystallography Made Crystal Clear, Third Edition: A Guide for Users of Macromolecular Models" (Rhodes, 2006) and "Outline of Crystallography for Biologists" (Blow, 2002). A good reading to understand better crystallization procedures is "Protein Crystallization, Second Edition" (Bergfors, 2009).

The first protein structure was determined by John Kendrew and coworkers in Cambridge (Kendrew *et al.*, 1958). This was the structure of myoglobin, a protein that could be isolated in large amounts from the skeletal muscle of sperm whale and formed excellent crystals. By 1970, solved protein structures included myoglobin, hemoglobin, lysozyme, ribonuclease, carboxypeptidase, α-chymotrypsin, papain, and subtilisin, (Meyer, 1997), all of them proteins that could be isolated from natural sources in large amounts.

With the advent of recombinant technologies, structural biology made several steps forward. Nowadays most structural studies are performed with recombinant protein, and isolation from source is reserved for those cases in which recombinant protein expression is not feasible or does not provide satisfactory results. The possibility to obtain large amounts of protein regardless their relative abundance in cells or tissues has opened the door to studies on relevant targets that were not abundant in natural sources, and on targets from pathogens and humans. Likewise, the possibility to simplify expression and purification procedures has opened the door to automation.

Protein X-ray crystallography projects are characterized by a number of defined steps aimed at obtaining pure and highly concentrated protein samples (protein expression and purification), obtaining protein crystals (protein crystallization), placing the crystals to an intense beam of X-rays (crystal diffraction), and analyzing X-ray diffraction patterns so as to determine the position of electrons in the protein (phasing and refinement).

Fig. 3.1 shows a typical hanging drop vapor diffusion experiment (a) and an example of a phase diagram for protein crystallization (b). In vapor diffusion experiments, as the vapor pressure is lower for the reservoir solution, there is a net transfer of water through the vapor phase from the protein droplet to the reservoir. As water vaporizes from the drop and transfers to the reservoir, protein and precipitant concentration increase within the droplet. Crystallization will occur between the nucleation and the metastable zone.

Over the last years, the automation of molecular biology procedures, crystallization trials and X-ray data collection together with the development of user-friendly software for structural determination have led to an exponential growth in the number of structures. This growth has come by the hand of large scale efforts to provide a biological meaning

to complete genomes or protein families, and in particular by the hand of structural genomics consortia.

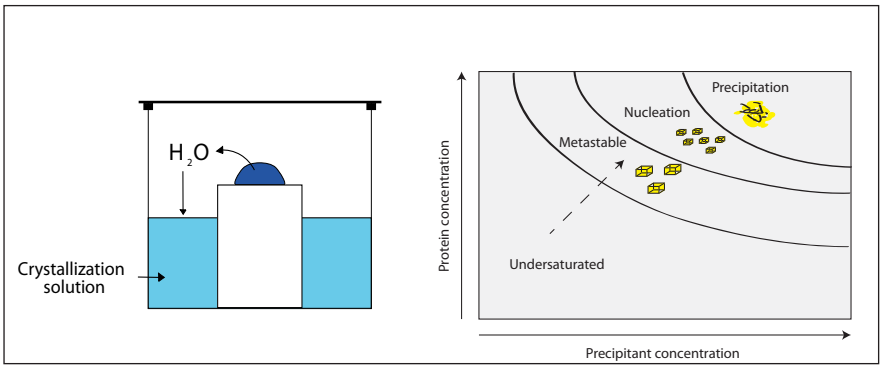


Fig. 3.1 Sitting drop setup (a) and phase diagram (b) for protein crystallization experiments. (a) A droplet of protein solution mixed with a buffer containing precipitant agents equilibrates with a larger reservoir containing a similar buffer. (b) Throughout a vapor diffusion experiment, protein and precipitant concentration increase within the droplet. Ideally, diffusion equilibrium will be achieved in the metastable zone.

But structural genomics groups have been accused of picking the low-hanging fruits, rather than the juiciest ones. Besides, in the era of high-throughput processes, the need for high-rate production of protein crystals and structures has undermined the primary goal of structural biology. Many structures are being determined without any functional interpretation and tricky targets fall of the path regardless their biological interest. In this scenario, smaller groups can take the crumbs left over by structural genomics consortia and try to provide a biological meaning to their results or take the risk and go for the high fruits, namely membrane proteins, eukaryotic targets and macromolecular complexes.

Indeed, protein purification and crystallization are still the major bottlenecks in X-ray crystallography projects. In order to crystallize, protein samples need to be pure and homogeneous. The most common reasons why a protein does not crystallize or crystallizes poorly are related to

its heterogeneity or propensity to aggregate. Aggregation and heterogeneity can be often detected at a purification level because they might prevent from obtaining soluble protein or result in multiple bands on SDS-PAGE gels. But it is important to keep in mind that not all protein heterogeneity can be detected during purification. In apparently homogeneous samples, mixed populations of proteins may exist due to the intrinsic flexibility of protein regions, to the existence of different folding states or to the lack of long-term stability.

In some cases, it is possible to play with expression, purification and storage conditions in order to infer the optimal conditions for protein stability. In other cases, the lack of stability can be related to the lack of an essential partner or substrate; but the missing element is often unknown and, even in those cases in which the element is known, the techniques for efficient co-expression or co-purification still pose many difficulties. Insolubility and instability problems are often a black box.

A common strategy to deal with instability consists of the removal of regions prone to aggregation or to proteolysis either at a genetic level or at a protein level. Sequence analysis can provide hints about the location of regions that can promote heterogeneity or aggregation, but the information provided by sequence analysis is not always accurate or sufficient and problems arise upon expression, purification or storage of protein samples. In those cases where heterogeneity is manifested by the appearance of degradation products, protein degradation patterns can provide interesting information for new rounds of cloning because they can point at regions that encode for a stable fragments of the protein. Alternatively, the blind design of a large number of constructs covering the problematic areas and the choice of alternative homologues are sensible options to rescue problematic targets.

Another problematic issue is that protein crystallization is still very much an empirical science with a low rate of success. With the automa-

tion and miniaturization of processes, crystallization has become more and more a random trial and error method with all the advantages and drawbacks that this implies. In any case, it is worth to bear in mind that commercially available crystallization screens are not exhaustive screens for crystal growing conditions but also that crystallization is a stochastic process in time and space guided by supersaturation and nucleation. This means that neither crystallography trials can be exhaustive, nor that results for a single condition are always reproducible.

Some of the problems detailed in this section motivate the decisions taken in the two projects that constitute this thesis.





3.2 General materials and methods

This section lists the materials used throughout this thesis and explains some methods that are widely used in biochemistry and structural biology labs, and are common to both projects. This section is organized in four subsections. Subsection 3.2.1 details DNA materials and methods, subsection 3.2.2 details microbiological materials and methods and subsection 3.2.3 details materials and methods for the purification and characterization of proteins.

3.2.1 DNA materials and methods

3.2.1.1 Vectors

Table 3.1 specifies the features of the vectors used throughout the thesis.

Table 3.1 Vectors

Vector	Length	Antibiotic	Promoter	N-term	Protease	C-term	Reference
pPROEX HTa	4750	Amp	pTrc/lacI	His (x6)	TEV	--	Gibco
pET22b	5493	Amp	T7/lacI	pelB	Signal	His (x6)	Novagen
pET28a	5369	Kan	T7/lacI	His (x6)	Thromb.	His (x6)	Novagen
pBAT4	4416	Amp	T7	--	--	--	Novagen
pETM40	6467	Kan	T7/lacI	MBP	TEV	--	Novagen
pGEX 4T-2TEV	4900	Amp	T7/lacI	GST	TEV	--	GE Health.
pBADM 40	5196	Amp	BAD	MBP	TEV	--	Invitrogen
pBADM 60	5570	Amp	BAD	NusA His (x6)	TEV	--	Invitrogen
pET46 EK-LIC	5200	Amp	T7/lacI	His (x6)	EK	--	Novagen
pET45	5200	Amp	T7/lacI	--	--	--	SGC

3.2.1.2 Plasmid purification and linear DNA purification

Plasmid purification from individual samples was performed with the Qiagen Miniprep kit (Qiagen) on a table-top centrifuge following the manufacturer's guidelines. Purifications from 96-well plates were performed with the Millipore Montáge miniprep kit and a vacuum manifold following manufacturer's instructions (Millipore).

Linear DNA fragment purification from individual samples (PCR products and cleaved vectors) was performed either with PCR and Gel Purification Kits (Qiagen), or with Illustra GFX PCR DNA and Gel Band Purification Kits (GE Healthcare) following the manufacturer's guidelines. Purifications from 96-well plates (extension products from the cycle sequencing reactions) were purified in 45µm poor size MAHVN4550 96 well filter plates (Millipore) loaded with Sephadex G-50 (GE Healthcare).

3.2.1.3 DNA electrophoresis

DNA electrophoreses were carried out with Bio-Rad electrophoresis tanks and power sources applying currents that ranged between 80V-100V. Stained DNA gels were visualized with a BioRad Fluor-S Multimager or with other UV light/blue light sources. The reagents used for gel preparation, gel running, and staining are detailed below.

- Agarose gels: 0.7-1% [w/v] agarose (Sigma) in 1X TAE, Tris-acetate-EDTA (TAE) buffer (stock solution 50X): 2M Tris, 1M acetic acid, 50mM EDTA, pH: 8.4; 6X DNA Loading dye (Fermentas).
- Molecular weight markers: λ DNA DNA/EcoRI + HindIII Marker, PhiX174 DNA/HinfI Marker, GeneRuler 1Kb DNA ladder (Fermentas).
- Staining solutions: SYBER green and SYBER safe (Invitrogen) (stock solution 10000X in DMSO, working concentration 1X diluted in TAE).

3.2.1.4 Determination of DNA concentration and purity

DNA concentration was routinely measured by Quant-iT dsDNA assays in a Qubit Fluorometer following manufacturer's instructions.

3.2.1.5 DNA precipitation

DNA precipitation was performed by sodium acetate/ethanol precipitation. A solution containing 3M sodium acetate (pH5.2) was added to the sample containing DNA to the final concentration of 0.3M. 2 volumes of ice-cold 100% ethanol were added to the sample, that was kept overnight at -20ºC. Samples were centrifuged at 14000rpm on a table top centrifuge for 15min and supernatant was removed. 500µl ice-cold 70% ethanol was added to the sample, and centrifugation was repeated. Supernatant was removed and tubes were left open to air-dry the pellets. DNA pellets were dissolved in water or in EB buffer (Qiagen).

3.2.1.6 DNA restriction digests

All restriction digests were performed either with New England Biolabs or Fermentas restriction enzymes (NcoI, HindIII, NdeI, EcoRI, SmaI, XhoI, PstI, NotI and AatII) at 37ºC for 1h in the digest buffers proposed by the manufacturers. Sample mix components are detailed in table 3.2.

Table 3.2 Sample mix preparation for restriction digests

Reagent	Cleavage test (x1)	Preparative digest (x1)
50µM plasmid	5 µl	40 µl
10X digest buffer	1 µl	5 µl
100X BSA*	0.1 µl	0.5 µl
Enzyme 1	0.25 µl	1 µl
Enzyme 2*	0.25 µl	1 µl
ddH_2O	Up to 10 µl	Up to 50 µl
Total volume	10 µl	50 µl

* If necessary

3.2.1.7 Colony PCR

Colony PCR with Taq polymerase (Stratagene) was used as a routine method to assess the success of cloning. Colonies were picked with a sterile pipette tip, and put into PCR eppendorf tubes containing the sample mix and conditions detailed in table 3.3. Primers are detailed in table 3.4.

Table 3.3 Sample mix preparation for colony PCR

Reagent	Sample mix (x1)
10X Taq polymerase buffer	2 µl
25 mM MgCl2	0.8 µl
10 mM dNTPs	0.4 µl
10 µM Forward primer	1 µl
10 µM Reverse primer	0.4 µl
Taq polymerase	0.24 µl
ddH_2O	Up to 20 µl
Total volume	20 µl
Cycles (n)	**Time / temperature**
1 cycle	5:00 min at 95ºC
30 cycles	1:00 min at 95ºC 1:30 min at 55ºC* 1:00 min at 72ºC
	5:00 min at 72ºC
Hold	∞ at 4ºC

3.2.1.8 Sequencing

Plasmids resulting from cloning procedures were sequenced either at Macrogen (Korea) or at Eurofins MWG Operon (Germany). Deletion libraries were sequenced at the MBB (Sweden). Details of protocol for in-house sequencing are specified in section 3.3.2.Primers are detailed in table 3.4.

Table 3.4 PCR primers for colony PCR and sequencing	
T7-promoter based vectors	
T7_promoter	TAATACGACTCACTATAGGG
T7_terminator	TATGCTAGTTATTGCTCAG
pGFPCR vectors	
pGFPCR_5	GCCTGCAGGTCGACTCTA
pGFPCR_3	CACTGGAGTTGTCCCAAT
pBAD vectors	
pBAD_forward	ATGCCATAGCATTTTTATCC
pBAD_reverse	GATTTAATCTGTATCAGG
ComA internal primer	
ComA_600	GTGGCTGGCGAAGCGG

3.2.2 Microbiological materials and methods

3.2.2.1 Growth media

Common chemical reagents for media and buffer preparation were purchased from Merck, Flucka, Sigma, Pronadisa and Difco. Common media formulations and reagents are detailed below:

- Luria-Bertani (LB) medium: 1% [w/v] bacto-tryptone, 0.5% [w/v] yeast extract, 1% NaCl.
- Terrific broth (TB) medium: 1.2% [w/v] bacto-tryptone, 2.4% [w/v] yeast extract, 0.4% [v/v] glycerol.
- LB agar plates: 1% [w/v] bacto-tryptone, 0.5% [w/v] yeast extract, 1% NaCl, 1.5% [w/v] agar.
- Antibiotics: ampicillin (stock solution: 100mg/ml in ddH_2O, working concentration: 100µg/ml); carbenicillin (stock solution: 100mg/ml in ddH_2O, working concentration: 100µg/ml); kanamycin (stock solution: 50mg/ml in ddH_2O, working concentration: 50µg/ml); chloramphenicol (stock solution: 35mg/ml in 100% [v/v] ethanol, working concentration: 35µg/ml).
- IPTG (stock solution: 1M).

3.2.2.2 Bacterial strains

Bacterial strains are detailed in Tables 3.5 and 3.6.

Table 3.5 Cloning strains

Cell strain	Genotype	Remarks
DH5a	F- endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG Φ80dlacZΔM15 Δ(lacZYA-argF)U169, hsdR17(rK-mK+), λ–	Cloning strain. Novagen
XL1-Blue	endA1 gyrA96(nalR) thi-1 recA1 relA1 lac glnV44 F'[::Tn10 proAB+ lacIq Δ(lacZ)M15] hsdR17(rK- mK+)	Cloning strain. Stratagene
MACH1	ΔrecA1398 endA1 tonA Φ80ΔlacM15 ΔlacX74 hsdR(rK- mK+)	Cloning strain. Invitrogen

Table 3.6 Expression strains

Cell strain	Genotype	Remarks
BL21 (DE3)	F– ompT gal dcm lon hsdSB(rB- mB-) λ(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5])	Expression strain with DE3, a λ prophage carrying the T7 RNA polymerase gene and lacIq. New England Biolabs
BL21 (0)	F- dcm ompT hsdS(rB- mB-) gal [malB+]K-12(λS)	Expression strain without DE3 for plasmids without T7 promoter. Stratagene
Rosetta (DE3)	F- ompT hsdSB(RB- mB-) gal dcm λ(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5]) pLysSRARE (CamR)	Expression strain with DE3, and pLysSRARE. pLysSRARE contains tRNA genes argU, argW, ileX, glyT, leuW, proL, metT, thrT, tyrU, and thru. Novagen
Rosetta (0)	F- ompT hsdSB(RB- mB-) gal dcm pLysSRARE (CamR)	Expression strain without DE3 but with pLysSRARE. Novagen
C41(DE3)	F– ompT gal dcm hsdSB(rB- mB-) λ(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5])	Expression strain with genetic mutations for conferring tolerance to toxic proteins. Strain widely used for the expression of membrane proteins. Lucigen corporation

3.2.2.3 Growth and storage of bacterial strains

Bacterial cells were commonly grown on LB and LB agar at 37ºC unless stated otherwise. Antibiotics were supplied to media for positive selection of clones. All media and instruments were sterilized by autoclaving or by filtration with 0.22 μm filter, and all manipulations were performed under sterile conditions.

For long-term storage, cells were mixed with glycerol 10-20% [v/v] and were rapidly flash-frozen in liquid nitrogen. Glycerol stocks were kept at -80 ºC.

3.2.2.4 Transformation of bacterial strains

Most transformations were performed with *E. coli* cells competent obtained by the $CaCl_2$ method and competence was induced by heat-shock treatment at 42 ºC for 45 s.

Deletion libraries and colony filtration blot experiments were performed with Z-competent cells (Zymoresearch) following the manufacturers' instructions.

3.2.3 Protein materials and methods

Common chemical reagents for media and buffer preparation were purchased from Merck, Flucka, Sigma, Pierce, Anatrace, Panreac, Invitrogen, Roche, Biorad and Fermentas.

3.2.3.1 Protein Electrophoresis

Protein electrophoreses were carried out using Bio-Rad electrophoresis tanks and power sources applying currents that ranged between 150V-250V. Stained gels were vacuum-dried, and scanned. Reagents for gel preparation, running and staining are detailed below:

- NuPAGE SDS-PAGE Gel system (Invitrogen): NuPAGE Bis-Tris Protein gels 4-12%, NuPAGE MES-SDS running buffer (20X) and NuPAGE LDS sample prep buffer.
- Novex Tris-Glycine Gel System (Invitrogen): Novex® Tris-Glycine gels 4-12%, Tris-Glycine Running Buffer (10X) and Tris-Glycine Native Sample Buffer (2X).
- SDS-PAGE gels: 10, 12.5, and 15% SDS PAGE-gels, 10X running buffer (0.25M Tris, 2M Glycine, 0.5% [w/v] SDS), 5X sample buffer.
- EMSA-PAGE gels: 15% SDS-PAGE gel, Tris-Acetate-EDTA (TAE) buffer, 6X DNA loading dye (Fermentas).
- Molecular weight markers: Seeblue and Mark12 (Invitrogen) for denatured proteins and MWND500 (Sigma) for native proteins.
- Staining solutions: Coomassie stain (1X): Coomassie Brilliant blue R-250, 40% [v/v] ethanol, 10% [v/v] acetic acid and Coomassie destain (10X): 10% [v/v]acetic acid for protein staining and SYBER gold stain (Invitrogen) (stock solution 10000X in DMSO, working concentration 1X diluted in TAE) for DNA staining.

3.2.3.2 Western blots and dot blots

Western blots and dot blots were performed either by using the Trans-Blot® SD Semi-Dry Electrophoretic Transfer Cell (BioRad) or the iBlot dry blotting system (Invitrogen) following the manufacturers' recommendations. Transfer was carried out with the following materials:

- Western blots with the Semi-Dry Electrophoretic Transfer Cell: Transfer buffer (200 ml methanol, 3g Tris, 3g glycine, 3ml 10% SDS, up to 1l H_2O), Whatmann 3MM paper (GE Healthcare), Nitrocellulose or PVDF membranes (GE Healthcare).
- iBlot Dry Blotting System (Invitrogen): Nitrocellulose transfer stacks and blotting pads.

In both cases following protein transfer, membranes were blocked O/N with Blotto/Tween PBS solution, incubated with HisProbe-HRP for 1h

at room temperature, and washed thrice with Tween PBS solution (3 washes 10 min each) or alternatively incubated for 1h at room temperature with anti-His primary antibody, washed thrice with Tween PBS (3 washes 10 min each), incubated for 1h at room temperature with anti-rabbit secondary antibody and washed again with Tween PBS (3 washes 10 min each).

- Blocking/washing solutions: Blotto/Tween-PBS (5.0g non-fat dry milk in 100ml 1XTween-PBS), Washing/Tween-PBS 1X Tween-PBS (10X solution: 0.1M PBS, 0.5% Tween 20, pH 7.4).

Protein detection was performed with SuperSignal West Dura HisProbe Kit (Pierce) or with Sigma Fast BCIP-NBT following the manufacturers' guidelines.

- Probes and detection reagents: HisProbe-HRP + SuperSignal West Dura HisProbe Kit (Pierce) or Anti-His primary antibody (1:1000 dilution in Tween/PBS) + Anti-rabbit secondary antibody (1:10000 dilution in Tween/PBS) + Sigma Fast BCIP-NBT (Sigma).

3.2.3.3 Chromatography columns and media

All purification columns and media were purchased from GE Healthcare. Protein purification protocols were carried out using the following materials:

- HisTrap HP columns (1ml and 5ml)
- HiTrap Heparin HP columns (1ml)
- GSTrap HP columns (1ml)
- Superdex 75 columns (16/60 and 26/60)
- Superdex 200 columns (10/300, 16/60 and 26/60)
- HiPrep Sephacryl S-200 HR
- HiTrap desalting column (5ml)
- Empty Disposable PD-10 columns filled with Ni-NTA Superflow (Qiagen)

Small-scale purifications were performed with MSDVN filter plates (Millipore) filled with Ni-NTA Superflow (Qiagen).

3.2.3.4 Protein concentration and dyalisis

Protein concentration was carried out with Vivaspin devices (GE Healthcare) of different sizes (0.5 ml, 2ml, 6ml and 20 ml) with different cut-off values (3000, 6000, 30000 MWCO).

Small volume (<2ml) dialysis was performed with Slide-A-Lyzer Dialysis Cassettes (Pierce). Larger volumes were dialyzed with dialysis membranes closed at both ends with clamps (Spectrum).

Protein concentration was measured either on a Nanodrop 2000 device (Thermoscientific) or by the Bradford method using Bradford reagent (Biorad) and BSA (Thermoscientific) as a standard following the manufacturers' guidelines.

Protein purity was assessed by analyzing the ration 260/280 on a Nanodrop 2000 device (Thermoscientific), by SDS-PAGE gels, by analyzing the gel filtration chromatograms, and by DLS.

3.2.3.5 Crystallization

Reagents and crystallization materials were purchased from Hampton Research and Molecular Dimensions. All crystallization experiments were performed at the Parc Científic de Barcelona Crystallography Platform or at the Karolinska Institute Crystallography Unit.

All crystallization experiments were performed using the hanging drop vapor diffusion method.

Initial screens were performed in 96 well MRC plates. Crystallization screenings were prepared with Tecan robots (Tecan, Switzerland). Crystallization drops were handled by Cartesian (Cartesian Technologies,

US) and Crystal Phoenix robots (Artrobbins, US), both providing microscale liquid handling for high-throughput screening of crystallization conditions. The best conditions were optimized on 24-well Linbro plates. Particular conditions and settings are indicated for every experiment in following sections.





3.3 Materials and methods: Natural transformation

This section describes the steps for cloning, expression, purification, biochemical characterization and crystallization trials of three proteins involved in natural transformation.

Section 3.3.1 describes the methodological aspects of a classical multi-construct approach based on bioinformatic predictions from construct design to expression and purification strategies for the best clones. Section 3.3.2 detail the protocols for the generation of exhaustive libraries with N-terminal, and C-terminal deletions (Erase-a-base libraries), the methods for expression screens at a colony level (Colony Filtration blot) and in small liquid cultures. Section 3.3.3 describes protocols for medium and large scale purification of the best clones.

3.3.1 Bioinformatics approach

3.3.1.1 Construct design

ComE, ComA and DprA initial construct design was based on the definition of domain borders by bioinformatics tools. The position of transmembrane helices was predicted by the tools: DAS (Cserzo *et al.*, 1997), TSEG (Kihara *et al.*, 1998), HMMTOP (Tusnády and Simon., 2001), TMHMM (Krogh *et al.*, 2001), SOSUI (Hirokawa *et al.*, 1998), TOPPRED (Claros and von Heijne, 1994), SPLIT (Juretic *et al.*, 2002), and TM-PRED (Hofmann and Stoffel, 1993). Disorder, globularity and folding were predicted with FoldIndex (Prilusky *et al.*, 2005), DisoPRED (Ward *et al.*, 2004), and GlobPlot (Linding *et al.*, 2003). Secondary structure elements were predicted with the tools PsiPRED (Jones, 1999), and DomPRED (Marsden *et al.*, 2002). Protein domains were predicted with SMART (Letunic *et al.*, 2012). Molecular weight, isoelectric point and extinction coefficient were calculated using tools available through the ExPASy server [Web9]. Regions predicted to be hy-

drophobic or disordered were systematically removed. Regions encoding domains predicted to be soluble were amplified from genomic DNA, and inserted in vectors with different fusion tags and promoters.

3.3.1.2 PCR and cloning

The genes *comE, comA* and *dprA* were amplified from *Neisseria gonorrhoeae* genomic DNA (ATCC 700825). Primers used to clone full-length and predicted soluble regions are shown in Table 3.7. Fig. 3.2 shows the pairs of primers used for the amplification of each fragment (*e.g. comE (B)* PCRs were performed with F_E02 and R_E02 primers). Restriction sites were introduced depending on the vector cloning site (*e.g. comE (B)* cloning into pPROEX HTa was performed with F_NcoI_E02 and R_HindIII_ E02 primers).

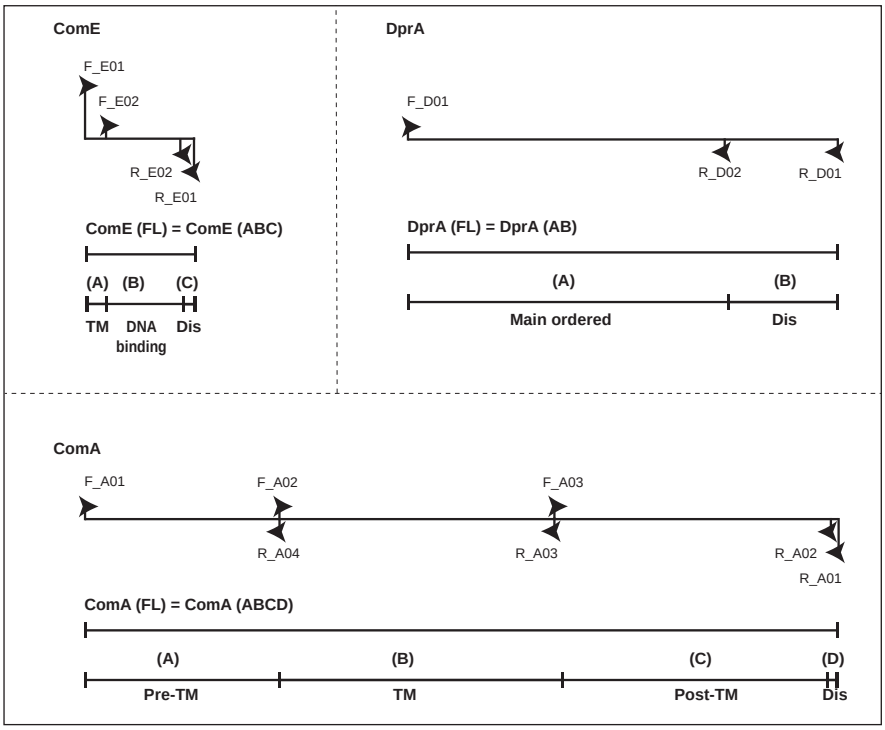


Fig. 3.2 Primers for the amplification of full-length proteins and soluble regions. For further information regarding the delimitation of soluble regions view Fig. 3.1

Table 3.7 PCR primers for cloning (bioinformatics approach)

ComE primers	
F_NdeI-E01	CAACCCTTAAAGGAAAACATATGAAAAAAATGTTTGTAT
F_NcoI-E01	TTAAAGGAAAAACCATGGAAAAAATGTTTGTATTGTT
F_NdeI-E02	GGCGGTAAACATCAATCATATGTCGCAGCAGGAGC
F_NCoI-E02	CGGTAAACATCAATGCCATGGCGCAGCAGGAGCT
R_EcoRI-E01	TCAGCAGTATGCATATGAATTCCTCTATTTTTTAACCGC
R_HindIII-E01	CAGCAGTATGCATATTAAGCTTTCTATTTTTTAACCGCA
R_EcoRI-E02	CGGGGCCTTTTGGAATTCGCGCCTAGACGGAAGCCTGG
R_HindIII-E02	CGGGGCCTTTTGGAAGCTTCGCCTAGACGGAAGCCTGG
ComA primers	
F_NdeI-A01	GCGTTTGCCGGTCATATGCTGTGCGTGTTG
F_NcoI-A01	GCGTTTGCCGGTCCCATGGTGTGCGTGTTGGCG
F_NdeI-A02	GGCCGTTGGGGCATATGCATTTGGTCAGTATCT
F_NcoI-A02	GGCCGTTGGGGCCCATGGATTTGGTCAGTATCTCG
F_NdeI-A03	GGTTTGTATCCTACCCATATGGAAGCAGTACCTGAAAA
F_NcoI-A03	GGTTTGTATCCTACCCCATGGAAGCAGTACCTGAAAA
R_EcoRI-A01	TTCCGGCACCAGCGAATTCGTTCACTCAAACGG
R_HindIII-A01	GGCACCAGCCTTCGAATTCACTCAAACGGTTT
R_EcoRI-A02	CGGTTTTTTCTGCCAAGAATTCTTATATTATCTCAAACGTTGAGC
R_HindIII-A02	GGTTTTTTCTGCCAATAAGCTTTATATTATCTCAAACGTTGAGC
R_EcoRI-A03	ACCGCAGCTTCGTTTGAATTCACCTATTCAGGACGGTAGGA
R_HindIII-A03	CCGCAGCTTCGTTTTAAGCTTCCTATTCAGGACGGTAGGA
R_EcoRI-A04	CCGAGATACTGACCAGAATTCTCTACCCCAACGGCCG
R_HindIII-A04	GAGATACTGACCAAAAGCTTCTACCCCAACGGCCG
DprA primers	
F_NdeI-D01	CGTAAGGGGATGGATCATATGACAGAGGACGAA
F_NcoI-D01	GTAAGGGGATGGATTCCATGGCAGAGGACGAACGTT
R_EcoRI-D01	TTCGAGTTCCTTAATATAGAATTCGTTCAAGTTCGGAT
R_HindIII-D01	CGAGTTCCTTAATATAAAAAGCTTTCAAGTTCGGATACG
R_EcoRI-D02	TATAGAATATGACGTTGAATTCGTTTATTGCAATAGCCCCG
R_HindIII-D02	TAGAATATGACGAAGCAAAGCTTTATTGCAATAGCCCCG

Polymerase Chain Reaction (PCR) with high fidelity polymerases (Vent DNA polymerase (New England Biolabs) and Expand DNA polymerase (Boehringer) was used as a method to amplify DNA for cloning purposes.

PCR reactions were carried out with the protocols reported in Tables 3.8 and 3.9. PCR results were analyzed on 1% agarose gels stained with SyberSafe (Invitrogen).

Table 3.8 PCR reaction with Vent polymerase

Reagent	Sample mix (x1)
Genomic DNA (50ng/µl)	1 µl
10X Thermopol buffer	5 µl
2 mM dNTPs	5 µl
10 µM Forward primer	1 µl
10 µM Reverse primer	1 µl
Vent polymerase (2u/µl)	0.5 µl
ddH_2O	Up to 50 µl
Total volume	50 µl

Number of cycles	Time and temperature
1 cycle	5:00 min at 95ºC
30 cycles	1:00 min at 95ºC 1:00 min at 60ºC 2:00 min at 72ºC
	10:00 min at 72ºC
Hold	∞ at 4ºC

Table 3.9 PCR reaction with Expand polymerase

Reagent	Sample mix (x1)
Plasmid DNA (25ng/µl)	2.5 µl
10X Expand buffer	5 µl
2 mM dNTPs	5 µl
10 µM Forward primer	1 µl
10 µM Reverse primer	1 µl
Expand polymerase (3.5u/µl)	0.75 µl
ddH_2O	Up to 50 µl
Total volume	50 µl

Number of cycles	Time and temperature
1 cycle	5:00 min at 94ºC
30 cycles	0:30 min at 95ºC 1:00 min at 60ºC 2:00 min at 72ºC
	10:00 min at 72ºC
Hold	∞ at 4ºC

Full-length *comE* and *comA* PCR products were initially cloned into the vector pGFPCR, and subsequently cloned into expression vectors. Cloning into pGFPCR was performed in a single reaction that comprised both the restriction and the ligation steps (table 3.10). The reaction mix was incubated for 20h at 20ºC, and SmaI was inactivated at 65ºC for 20min. Ligation products were used to transform DH5α or BL21 chemically competent cells. Fragment insertion was assessed by UV fluorescence of positive colonies, plasmid digestion by flanking XbaI digestion sites, and sequencing.

Table 3.10 Sample mix preparation for cloning into pGFPCR

Reagent	Sample mix (x1)
pGFPCR vector	1.5 µl
SmaI	0.5 µl
PCR product	5 µl
T4 DNA Ligase (NEB)	1 µl
T4 buffer 10X	1 µl
ddH_2O	Up to 10 µl
Total volume	10 µl

DprA was directly cloned into the expression vector pPROEXHTa because it contained an internal restriction site for SmaI. *ComA, comE* and *dprA* truncations were amplified from plasmid DNA and recloned into the vectors pPROEXHTa, pET22b, pET28a, and pBAT4. In addition, *comA* and *dprA* truncations were cloned into vectors providing fusion tags (pETM40, pGEX-4T-2TEV, and pBAD M60).

DprA was later cloned into pET 45 so as to introduce a C-terminal tag. PCR was performed with the primers detailed in table 3.4 and the protocol detailed in table 3.11 for PCR with the enzyme Phusion Hot-start.

Table 3.11 PCR primers for cloning with C-terminal His-tag

DprA primers	
F_DprA	GACTCCATGGCAGAGGACGAAC
R_DprA	GATCAAGCTTTCAGTGATGGTGGTGGTGATGAGTTCGGATACGCTGGTATC
R_DprA_TEV	GATCAAGCTTTCAGTGATGGTGGTGGTGATGGAAAACCTG-TATTTTCAGGGTAGTTCGGATACGCTGGTATC

PCR products and vectors were double-digested with the restriction pairs NcoI-HindIII or NdeI-EcoRI (New England Biolabs). Digested PCR products were purified from solution or from agarose gel bands with the Illustra Kit (GE Healthcare). PCR products and vectors were ligated with T4 DNA ligase (Roche).

Ligation products were used to transform XL-Blue or DH5a chemically competent cells via the heat shock protocol. Several colonies were picked from each plate, and plasmids were purified and sequenced.

3.3.1.3 ComE, ComA and DprA expression trials

Expression trials were performed in BL21(DE3) and Rosetta *E. coli* cells. Expression trials were performed in 250ml flasks containing 50 ml LB media and in 2l flasks containing 600 ml LB media. Several induction conditions were tried modifying induction temperature (20ºC or 37ºC), induction time (20h or 3h), inductor type (IPTG or arabinose), and inductor concentration (1mM or 0.1mM), and OD550 at the time of induction.

Cells were harvested at 4000g, and cellular pellets were resuspended in native buffer (30mM Tris-HCl pH 7.5, 250 mM). All solutions were sonicated on ice with a probe-tip sonicator, and centrifuged at 20000g. Soluble and insoluble fractions were analyzed on SDS-PAGE gels stained with Coomassie blue.

3.3.1.4 ComE purification trials

Three strategies were compared to define a protocol for ComE(B) purification:

1. Ni-NTA affinity purification + gel filtration
2. Ni-NTA affinity purification + desalting + histidine cleavage + gel filtration
3. Heparine affinity purification + gel filtration

Ni-NTA affinity purification and gel filtration

ComE(B) cell pellets coming from large scale cultures were resuspended in lysis buffer (30mM Tris HCl pH 7.5, 30mM imidazole, 250 mM NaCl protease inhibitors (1 tablet/50ml), and 100 U/ml DNAaseA). Supernatant was purified in two steps: Ni-NTA affinity chromatography and size exclusion chromatography.

Affinity chromatography was carried out with 1ml and 5ml Histrap FF columns in an ÄKTA Purifier system (GE Healthcare). Columns were washed with 40 CV binding buffer (30mM TrisHCl pH 7.5, 250mM NaCl, 30mM imidazole). Samples were loaded with binding buffer and eluted with elution buffer (30mM TrisHCl pH 7.5, 250mM NaCl, 500mM imidazole) in a 20CV linear gradient from 30 to 500 mM imidazole.

Eluted fractions from the affinity column were concentrated and loaded on a gel filtration column equilibrated in gel filtration buffer. Tris HCl pH 7.5 buffers containing different NaCl concentrations (0.25, 0.5 and 0.75 NaCl) were tested. Initial purification trials were performed with Superdex 75 10/300. Purification was scaled up using Superdex 75 16/60 and Superdex 75 26/60 changing the volume of injected sample and flow rates accordingly.

Ni-NTA affinity purification, desalting, histidine cleavage, and gel filtration
ComE(B) cell pellets coming from large scale cultures were resuspended in lysis buffer (30mM Tris HCl pH 7.5, 30mM imidazole, 250 mM NaCl protease inhibitors (1 tablet/50ml), and 100 U/ml DNAaseA). Supernatant was purified in four steps: Ni-NTA affinity chromatography, desalting, histidine cleavage/purification and size exclusion chromatography.

Affinity chromatography was carried out with 1ml and 5ml Histrap FF columns in an ÄKTA Purifier system (GE Healthcare). Columns were washed with 40 CV binding buffer (30mM TrisHCl pH 7.5, 250mM NaCl, 30mM imidazole). Samples were loaded with binding buffer and eluted with elution buffer (30mM TrisHCl pH 7.5, 250mM NaCl, 500mM imidazole) in a 20CV linear gradient from 30 to 500 mM imidazole.

Eluted fractions Ni NTA affinity purification were concentrated up to 10 ml, and loaded onto a HiPrep 26/10 desalting column for buffer exchange equilibrated with desalting buffer (150 mM NaCl, 10mM Tris HCl pH 8, 30 mM imidazole, 1mM DTT).

Samples were collected and digested overnight with TEV protease. Trials with TEV protease were performed at 2 temperatures (4ºC and 20ºC), two ratios TEV:protein (1:50 and 1:100), taking samples after different incubation times (3h and 20h). After the initial trials, tag cleavage was routinely performed for 20h at 4ºC using a ration TEV:protein 1:50.

Cleaved protein was purified to remove uncleaved protein, histidine tags and TEV protease by Ni-NTA affinity purification with a His-trap 1ml column. Sample was recovered with desalting buffer (150 mM NaCl, 10mM Tris HCl pH 8, 30 mM imidazole, 1mM DTT) whereas all other contaminants were eluted in a single step with elution buffer 2 (150 mM NaCl, 10mM NaCl, 500mM imidazole).

Unbound sample was collected, concentrated and further purified by gel filtration. Initial purification trials were performed with Superdex 75 10/300 at 0.5ml/min. Purification was scaled up using Superdex 75 16/60 and Superdex 75 26/60 changing the volume of injected sample and flow rates accordingly.

Heparine affinity purification

ComE(B) cell pellets were resuspended in lysis buffer (10mM sodium phospate pH 7, protease inhibitors (1 tablet/50ml), DNAaseA), sonicated and centrifuged. Supernatants were filtered and loaded onto heparine columns equilibrated with binding buffer (10mM sodium phosphate pH7). Samples were eluted with a 20 column volume gradient from 0 to 1M NaCl.

3.3.1.5 ComA and DprA small scale solubility screens

Small scale solubility screens were performed by solubilizing a cell pellet equivalent to 25ml of cell culture in 2ml of lysis buffer. The following buffers were used:

A. 30mM Tris HCl pH 7.5, 50mM NaCl
B. 30mM Tris HCl pH 7.5, 2M NaCl
C. 30mM Tris HCl pH 7.5, 50mM NaCl, 0.5M urea
D. 30mM Tris HCl pH 7.5, 50mM NaCl, 1M urea
E. 30mM Tris HCl pH 7.5, 50mM NaCl, 2M urea
F. 30mM Tris HCl pH 7.5, 50mM NaCl, 4M urea
G. 30mM Tris HCl pH 7.5, 50mM NaCl, 6M urea
H. 20mM Tris HCl pH 7.5, 50mM NaCl, 0.2% NP40
I. 20mM Tris HCl pH 7.5, 50mM NaCl, 0.2%Triton
J. 20mM Tris HCl pH 7.5, 50mM NaCl, 0.2%Tween20
K. 20mM Tris HCl pH 7.5, 50mM NaCl, 0.2% DDM

Cell suspensions were sonicated, and centrifuged at maximum speed on a table top centrifuge. Samples from the supernatant (soluble fractions) and pellet (insoluble fractions) were loaded on SDS-PAGE gels.

3.3.1.6 ComA and DprA purification trials with large fusion tags

Soluble fractions obtained after expression of protein truncations with fusion tags (GST-tag, MBP-tag, NusA-tag) were purified in two steps: affinity purification and size-exclusion chromatography.

GST-tagged ComA(A), and DprA(A) were expressed in BL21 cells. Cell pellets were lysed in GST lysis buffer (1xPBS, 100U/ml DNAaseA). Samples were loaded onto GST-Trap columns equilibrated with GST binding buffer (1xPBS), and eluted with 10 column volumes GST elution buffer (50mM Tris HCl pH 8, 10mM reduced glutathione). Eluted peaks were purified with a Superdex 200 10/300.

MBP-tagged ComA(A), and DprA(A) were expressed in BL21 cells. Cell pellets were lysed in MBP lysis buffer (20mM Tris HCl pH 7.4, 200mM NaCl, 1mM EDTA, 1mM DTT, 100U/ml DNAaseA). Samples were loaded onto MBP-Trap HP columns equilibrated with MBP binding buffer (20mM Tris HCl pH 7.4, 200mM NaCl, 1mM EDTA, 1mM DTT), and eluted with 10 column volumes MBP elution buffer (20mM Tris HCl pH 7.4, 200mM NaCl, 1mM EDTA, 1mM DTT, 10mM maltose). Eluted peaks were purified with a Superdex 200 10/300.

NusA-tagged ComA(A), and DprA(A) were expressed in BL21 cells. Cell pellets were lysed in NusA lysis buffer (30mM Tris HCl pH 7.5, 30mM imidazole, 250 mM NaCl, 100 U/ml DNAaseA). Samples were loaded onto HisTrap columns equilibrated with binding buffer (30mM Tris HCl pH 7.5, 30mM imidazole, 250 mM NaCl), and eluted with 20 volumes elution buffer (30mM Tris HCl pH 7.5, 500mM imidazole, 250 mM NaCl). Eluted peaks were purified with a Superdex 200 10/300.

3.3.1.7 ComA refolding

ComA(C) pellets were resuspended in denaturing lysis buffer (30mM TrisHCl pH 8, 500mM NaCl, urea 8M, 10mM β-mercaptoethanol, 5mM imidazole, protease inhibitors (1 tablet/50ml), DNAaseA). Soluble fraction were centrifuged, and supernatants were purified by Ni NTA affinity purification under denaturing conditions.

Affinity chromatography was carried out with 1ml and 5ml Histrap FF columns in an ÄKTA Purifier system (GE Healthcare). Columns were washed with 40 CV binding buffer (30mM TrisHCl pH 8, 500mM NaCl, urea 8M, 10mM β-mercaptoethanol, 5mM imidazole). Samples were loaded with binding buffer and eluted with elution buffer (30mM TrisHCl pH 8, 500mM NaCl, urea 8M, 10mM β-mercaptoethanol,

500mM imidazole) in a 20CV linear gradient from 5 to 500 mM imidazole.

Refolding experiments were performed by step-wise dilution in dialysis bags (10KDa cutoff) dialyzing against buffers containing different amounts of urea (8M, 4M, 2M, 1M, 0.5M and 0M) and β-mercaptoethanol (10mM, 5mM, 0M). The effect of L-Arg was analyzed by adding 0.4M L-Arg to the buffers.

After refolding, samples were purified by gel filtration on a Superdex 75 10/300 equilibrated with the buffer used in the last dialysis step. Samples were analyzed by SDS-PAGE gels, native gels and DLS.

3.3.2 The high-throughput approach

3.3.2.1 Erase-a-base overview

Deletion libraries were generated with the commercial kit Erase-a-base (Promega). Erase-a-base allows the generation of unidirectional deletions on linearized DNA. This method has been used for many purposes including DNA sequencing (Badía *et al.*, 1998)(Lodge *et al.*, 1999), mapping of boundaries of regions involved in genetic control (Wu *et al.*, 2007), protein/protein interaction sites (Xie *et al.*, 2006), or protein domains (Cornvik *et al.*, 2006). In this thesis it has been used to generate of a collection of vectors containing random truncations of a particular insert.

The generation of deletion libraries by Erase-a-base relies on the combination of the activity of two enzymes: exonuclease III (ExoIII), and S1 nuclease. ExoIII is a 3′- 5′ exonuclease that catalyzes the removal of nucleotides from short overhangs (3 or less bases), blunt or recessed ends at 3’ (sensitive ends), but does not catalyze the removal of nucleotides if 3’ overhangs are longer than 3 bases (protective ends). Sensitive

ends adjacent to the insert and protective ends adjacent to the vector can be introduced at the cloning step. Since the degradation with ExoIII proceeds at a constant rate and is temperature dependent, the process can be controlled. Deletions in the desired size range can be obtained by removing aliquots of the reaction mixture at appropriate periods of time. Since ExoIII acts just on one of the strands, an additional reaction with S1, an enzyme that catalyzes the removal of nucleotides from single-stranded DNA, is needed to completely remove the nucleotides from the other strand that has been left single-stranded after ExoIII reaction. Following S1 reaction, the resulting material is protected against further nuclease degradations by filling the recessed 3' ends with phosphorothioate deoxynucleotides using the Klenow fragment of DNA polymerase I. The fragments are ligated with T4 ligase.

The colony filtration blot, later referred to as CoFi blot, is a method developed at Karolinska Institute (Cornvik *et al.*, 2005) that allows the identification of *E. coli* clones that express soluble protein. The method relies on a filtration step through a membrane that is permeable to soluble protein but not to inclusion bodies. Soluble protein is transferred to a nitrocellulose membrane, and detected by immunochemical methods. The colony filtration blot has been used to screen the solubility of clones generated by random mutagenesis (Dahlroth *et al.*, 2006), deletion libraries (Cornvik *et al.*, 2006) and complete ORFeomes (Dahlroth *et al.*, 2009). An adaptation of the method has been developed to screen the solubility of membrane proteins in the presence of detergents (Martinez Molina *et al.*, 2008). In this thesis, the colony filtration blot is used for the screening of clones from random deletion libraries obtained from genes encoding a soluble protein (DprA) and a membrane protein (ComA).

3.3.2.2 Primer design

One *dprA* clone (expressing DprA (FL)) and two *comA* clones (expressing ComA(FL) and ComA(CD)) were designed to be substrates for the

generation of N-terminal and C-terminal deletion libraries (Fig. 3.3). For that purpose, a pair of primers was designed for each clone, each containing a sensitive and a protective restriction site.

Erase-a-base protocol provided a list of recommended enzymes that provide sensitive (blunt or short overhangs) and protected ends (overhangs longer than 3 bases), but the optimal combination of enzymes was decided on a case-by-case basis. The restriction pair PstI/NotI was introduced in forward primers and the restriction pairs SacI/NdeI and AatII/NdeI were introduced in reverse primers. Other important features that were taken into consideration in primer design were the conditions of the restriction digest reaction (compatibility and requirements for optimal activity), the cloning method (introduction of LIC linkers for homologous recombination) and the maintenance of the reading frame (addition of nucleotides to keep the frame with the histidine-tag in forward primers, and addition of stop codons in reverse primers). Examples of primer design are shown below. Primers are detailed in Table 3.12.

Forward primer introducing PstI (protective) NotI (sensitive) restriction sites:

Forward DprA (F_DprA) (Tm = 58C)

Linker	PstI	NotI	Sequence
GACGACGACAAGAT	CTGCAG	AGCGGCCGC	ATGACAGAGGACGAACGTTT

Reverse primers introducing SacI (protective) NdeI (sensitive) restriction sites:

Reverse DprA (R1_DprA) (Tm = 58C)

Linker	Stp	Stp	Stp	SacI	NdeI	Sequence
GAGGAGAAGCCCGGT	TCAGTCAGTCA	GAGCTC	ACATATG	TCAAGTTCGGATACGCTGG		

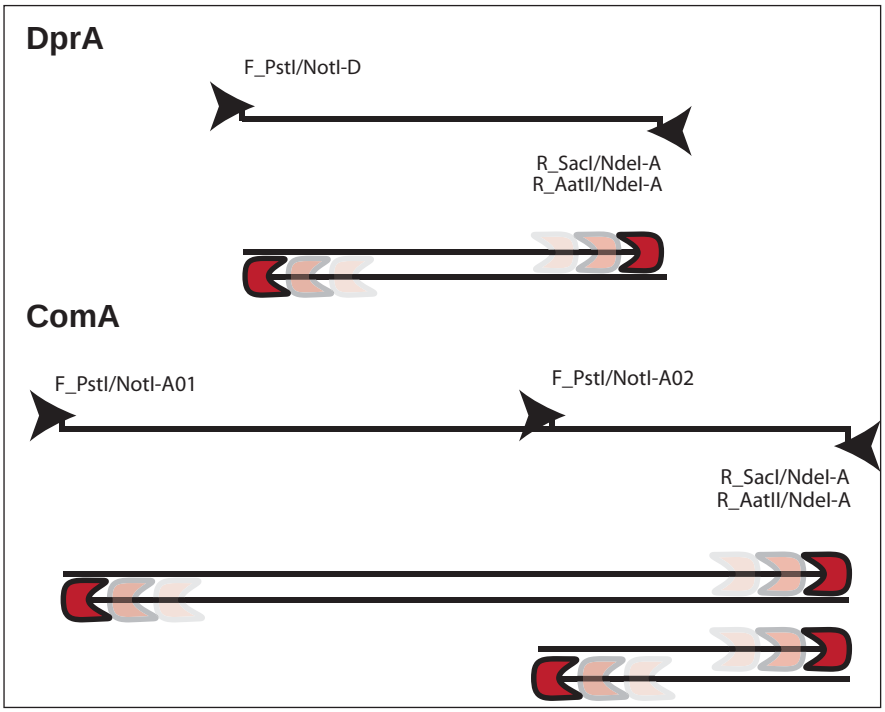


Fig. 3.3 Primers for the amplification of initial clones for the generation of deletion libraries.

Table 3.12 PCR primers for cloning (HTP approach)

ComA primers	
F_PstI/NotI-A01	GACGACGACAAGATCTGCAGAGCGGCCGCATGCTGTGCGTGTTGGCG
F_PstI/NotI-A02	GACGACGACAAGATCTGCAGAGCGGCCGCCCTGAAGCAGTACCTGAAAA
R_SacI/NdeI-A	GAGGAGAAGCCCGGTTCAGTCAGTCAGAGCTCACATATGTCACT-CAAACGGTTTTTTCTG
R_AatII/NdeI-A	GAGGAGAAGCCCGGTTCAGTCAGTCAGACGTCACATATGT-CACTCAAACGGTTTTTTCTG
DprA primers	
F_PstI/NotI-D	GACGACGACAAGATCTGCAGAGCGGCCGCATGACAGAGGACGAACGTTT
R_SacI/NdeI-D	GAGGAGAAGCCCGGTTCAGTCAGTCAGAGCTCACATATGT-CAAGTTCGGATACGCTGG
R_AatII/NdeI-D	GAGGAGAAGCCCGGTTCAGTCAGTCAGACGTCACATAT-GTCACATATGTCAAGTTCGGATACGCTGG

3.3.2.3 PCR and cloning

Insert amplification was carried out by PCR with Phusion Hot-start polymerase (Thermoscientific) by using a protocol adapted from that provided by the manufacturer as indicated in Table 3.13. DNA template was pGFPCR (ComA) for *comA* and *comA(CD)* amplifications and pGFPCR (DprA) for *dprA* amplifications.

Table 3.13 PCR reaction (Phusion Hot-Start)

Reagent	Sample mix (x1)
Plasmid DNA (50 ng/µl)	1 µl
5X HF buffer	10 µl
50 mM $MgCl_2$	1.5 µl
10 mM dNTPs	1 µl
10 µM Forward primer	2.5 µl
10 µM Reverse primer	2.5 µl
Phusion polymerase (2 u/µl)	0.5 µl
ddH_2O	Up to 50 µl
Total volume	50 µl

Number of cycles	Time and temperature
1 cycle	0:30 min at 98ºC
30 cycles	0:10 min at 98ºC 0:30 min at 60ºC 1:20 min at 72ºC
	10:00 min at 72ºC
Hold	∞ at 4ºC

PCR products were purified either directly from the reaction mix or from the gel band by using the QuiaquickPCR purification kit (Qiagen).The length and purity of the amplified products was checked by electrophoresis on 0.7% agarose gels.

Purified PCR products were treated with T4 DNA polymerase and annealed with EK/LIC vector into the vector pET46 EK-LIC (Novagen).

LIC cloning was performed following a protocol with slight modifications of that provided by the manufacturer. PCR products were incubated at room temperature for 30 min in T4 DNA polymerase treatment mix (table 3.14). Treated PCR inserts were annealed with EK/LIC vector by incubating insert:vector in a 2:1 proportion for 30 min at 20ºC. The annealing reaction was inactivated by adding 1µl 25mM EDTA, and heating the sample for 20 min at 75ºC. Annealed products were used to transform competent cells.

Table 3.14 Sample mix for T4 DNA pol treatment for LIC cloning

Reagent	Sample mix (x1)
Purified PCR product	5 µl
10X Phusion T4 DNA pol buffer	1 µl
25 mM dATP	1 µl
10 mM DTT	0.5 µl
T4 DNA polymerase	0.2 µl
ddH_2O	Up to 10 µl
Total volume	10 µl

Annealed products were used to transform MACH1 competent cells (Z-protocol). Cloning outcome was assessed by colony PCR, restriction by the restriction sites provided in the cloning step (pairs PstI/NotI, SacI/NdeI or AatII/NdeI) and sequencing.

3.3.2.4 Erase-a-base step by step

N-terminal and C-terminal deletion libraries of *dprA(FL), comA(FL)* and *comA(CD)* were generated using the Erase-a-Base kit (Promega). The Erase-a-base protocol detailed below is a modified version of that provided by the manufacturer. Steps are summarized in Fig. 3.4.

1. DNA preparation for Erase-a-base: Cloning vectors containing the genes *dprA, comA*, and the C-terminal end of *comA (comA(CD))*

were cut with the restriction enzyme pairs indicated in the cloning step so as to generate a linear molecule with a 5'-overhang adjacent to the gene and a 3'-overhang on the vector side of the molecule. Linearized products were run on a 0.8% agarose gel, and were purified with the QuiaquickPCR purification kit (Qiagen).

2. Exonuclease reaction and S1 reaction: ExoIII mix and S1 mix were prepared following manufacturer's guidelines. S1 mix was placed on ice and ExoIII mix was placed either at room temperature or at 4ºC. Exonuclease reaction trials were performed at two temperatures: 22ºC and at 4ºC. Exonuclease reaction was triggered by adding ExoIII to linearized DNA (ratio ExoIII:vector mix; 1:30). Aliquots were transferred from ExoIII tube to S1 tubes at fixed time-points (30s when performing the digests at 22ºC, 60s when performing the digests at 4ºC). 6 aliquotes were transferred to each S1 tube (6 pools of 6 samples each for *dprA*, 8 pools of 6 samples each for *comA(FL)* and 4 pools of 6 samples each for *comA(CD)*). S1 reactions were incubated at 22ºC for 30min and were subsequently inactivated by adding S1 nuclease stop buffer (ratio S1 mix: stop buffer 10:1) and by heating for 10min at 70ºC.

3. DNA precipitation and Klenow reaction: DNA was precipitated in Ammonium acetate:Ethanol, resuspended in 10 µl resuspension buffer (10mM Tris-HCl pH 8.5), and subjected to the Klenow reaction. Klenow reaction was initiated by adding 1 µl Klenow mix (20µl Klenow buffer, 1µl Klenow enzyme) and 1 µl 1mM dNTPs) to the tubes. Tubes were incubated at 37ºC for 8 min and inactivated at 65ºC for 10 min.

4. Ligation and transformation: Vectors were ligated by adding 40 µl ligation mix (prepared as indicated by the manufacturer's) to the DNA. Ligation reactions were incubated for 1h at 22ºC. 200µl MACH1 competent cells were transformed with 10µl ligation re-

action by using the Z-competent protocol, and plated onto large (d=14cm) LB-plates. Cells were grown overnight at 37ºC.

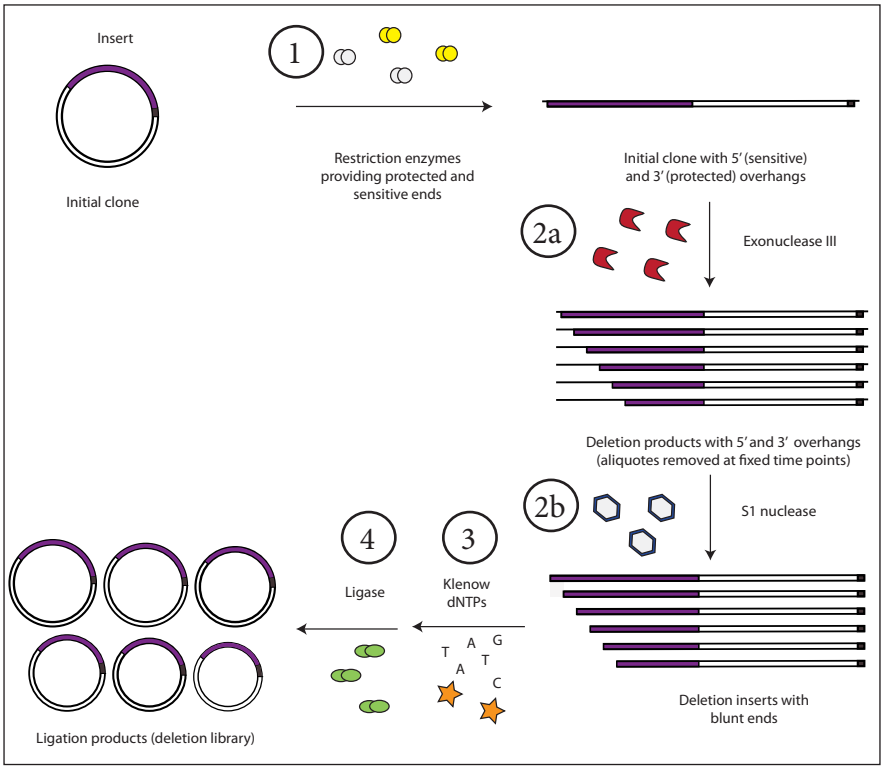


Fig. 3.4 Diagram of the steps for the generation of deletion libraries. 1) Restriction digest 2a) Exonuclease III reaction. 2b) S1 nuclease reaction. 3) Klenow reaction in the presence of dNTPs. 4) Ligation.

5. Library harvesting: Transformed cells were harvested with a scrapper and resuspended in LB liquid media. Plasmid libraries were isolated with a Qiaprep® Miniprep kit (Qiagen).

6. Library quality check: The quality of the libraries was assessed by estimating the number of clones per plate and by performing restriction digests at restriction sites flanking the inserts NotI/XhoI.

3.3.2.5 Colony Filtration blot

The solubility of N-terminal and C-terminal deletion libraries of DprA(FL), ComA(FL) and ComA(CD) was assessed at a colony level by Colony Filtration blot. Positive clones were rechecked by mini colony filtration blot. Colony filtration blot and mini colony filtration protocols detailed below are modified versions of the protocols provided by (Dahlroth *et al.*, 2006), (Martinez Molina *et al.*, 2008). Steps are summarized in Fig. 3.5.

1. Transformation with deletion libraries: Z-competent expression cells (BL21 or C41) were transformed with approximately 100 ng of plasmid library. Transformation mixes were plated onto LB-plates (d=14cm) containing 50µl/ml carbenicillin (growth plates). Growth plates were incubated overnight at 37ºC.

2. Induction: 0.45µm Durapore membranes (Millipore) were placed onto the growth plates to get a footprint containing a representation of all the colonies. Growth plates were wrapped with plastic foil, and kept at 4ºC whereas Durapore membranes were placed with the colonies facing up on fresh LB-plates (d=14cm) containing 50µl/ml carbenicillin and 0.2 mM IPTG (induction plates). Protein expression was induced for 12h at 22ºC.

3. Filter sandwiches preparation and cell lysis: Durapore membranes were placed on top of a nitrocellulose filter and a Whatman 3MM paper (GE Healthcare) forming a "filter sandwich". Fig. 3.6 shows a transversal cut of a filter sandwich and the differential behavior of a folded and an aggregated protein. "Filter sandwiches" were soaked in lysis buffer, and incubated at 22 ºC for 30 min.

    ◊ DprA truncations solubility was screened under native conditions. "Filter sandwiches" containing colonies that had expressed DprA truncations were soaked in 8ml native lysis buffer (20 mM

Tris pH 8, 100 mM NaCl, 0.1 mg/ml lysozyme, 0.75 mg/ml DNAse I, ½ complete EDTA- free protease inhibitor cocktail tablet/10 ml (Roche)).

◊ ComA truncations solubility was screened under two conditions: native buffer (to detect soluble fragments) and buffer containing the detergent FC12 (to detect fragments associated or inserted into the membranes). Durapore membranes with colonies were splitted in two parts and the halves were placed on top of nitrocellulose filters onto Whatman 3MM papers soaked either in 4ml Native lysis buffer (20 mM Tris pH 8, 100 mM NaCl, 0.1 mg/ml lysozyme, 0.75 mg/ml DNAse I, ½ complete EDTA-free protease inhibitor cocktail tablet/10 ml (Roche)) or in 4ml FC12 detergent lysis buffer (20 mM Tris pH 8, 100 mM NaCl, 0.1 mg/ml lysozyme, 0.75 mg/ml DNAse I, ½ complete EDTA-free protease inhibitor cocktail tablet/10 ml (Roche), 1% FC12).

4. Immunodetection: Nitrocellulose membranes were removed from the filter sandwiches and dried overnight at 22ºC. Dry membranes were washed 3x10 min in TBST, incubated with HisProbe-HRP (Pierce) diluted 1:5000 in TBST for 1h, and washed again 3x10 min with TBST. Membranes were developed with SuperSignal® WestDura chemiluminescence reagents (Pierce) following manufacturers' instructors with a Fluor-S MultImager (BioRad).

5. Identification of positive clones: Stronger spots were identified in the pictures of the blotted membranes. The relative position of the spots in the membranes was compared to the relative position of individual colonies in the growth plates. Samples of positive clones were carefully picked with a pipette tip, grown overnight in 96 well-plates containing 1ml LB, and frozen as glycerol stocks.

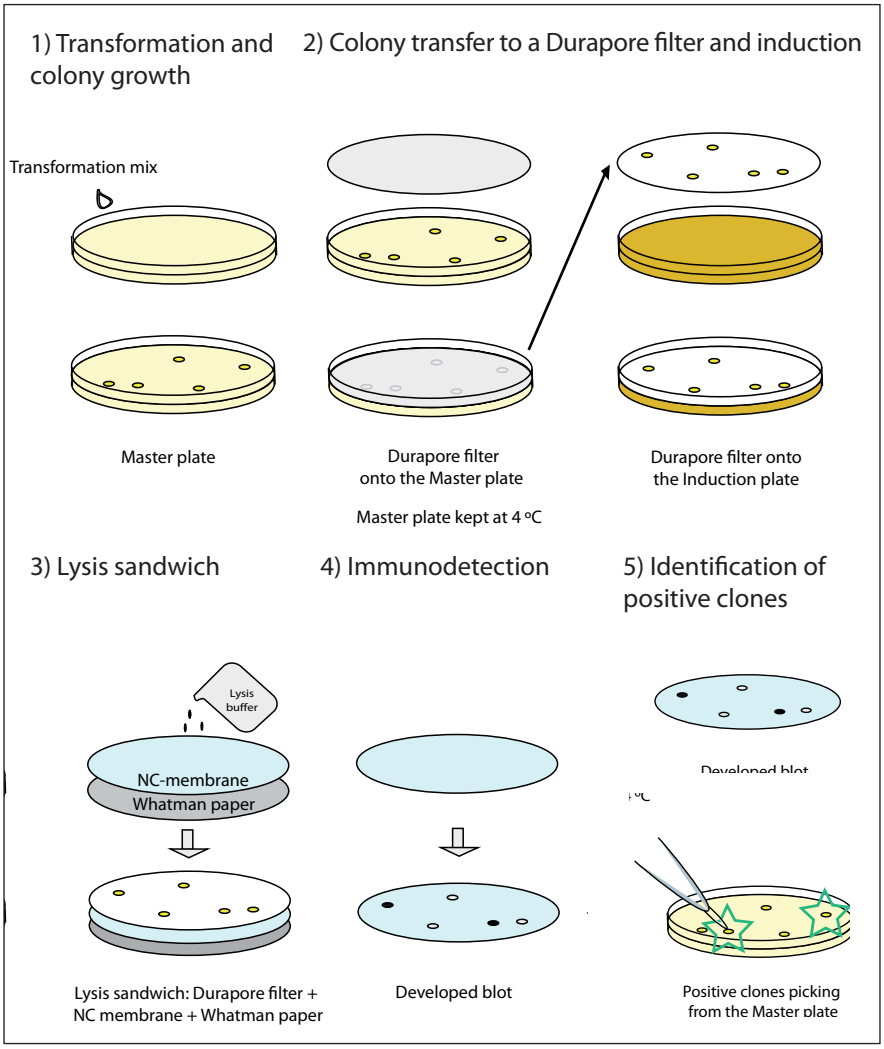


Fig. 3.5 Diagram of the steps for the Colony filtration blot.

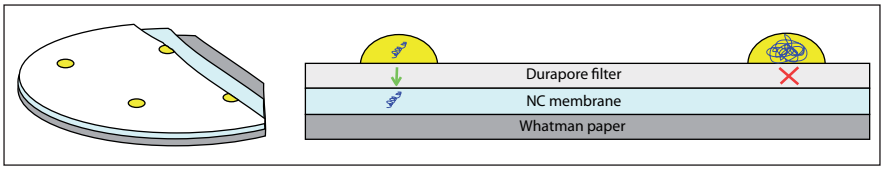


Fig. 3.5. Detail of the Filter Sandwich showing the differential behavior of a colony expressing soluble protein and a colony expressing aggregated protein.

3.3.2.6 Mini Colony Filtration Blot

1. Colony transfer to Durapore membranes: Colonies were picked from frozen stocks and transferred to 96-well plates filled with 200µl LB containing carbenicillin. 1µl from each well was transferred to a rectangular (8x 12 cm) 0.45µm Durapore membrane (Millipore). Durapore membranes were put on the top of LB-plates (d=14cm) containing carbenicillin (growth plates). Growth plates were incubated overnight at 37ºC to promote cell growth.

2. Induction: Durapore membranes were lifted and put onto fresh LB-plates (d=14cm) containing 50µl/ml carbenicillin and 0.2 mM IPTG (induction plates). Protein expression of the colonies was induced for 12h at 22ºC.

3. Preparation of the filter sandwiches and cell lysis was carried out as indicated above for the colony filtration blot. DprA solubility trials were performed using native lysis buffer (20 mM Tris pH 8, 100 mM NaCl, 0.1 mg/ml Lysozyme, 0.75 mg/ml DNAse I, ½ Complete EDTA- free protease inhibitor cocktail (Roche) tablet/10 ml). ComA solubility trials were performed in two buffers: FC12 Detergent lysis buffer (20 mM Tris pH 8, 100 mM NaCl, 0.1 mg/ml Lysozyme, 0.75 mg/ml DNAse I, ½ Complete EDTA- free protease inhibitor cocktail tablet/10 ml (Roche), 1% FC12) and DDM Detergent lysis buffer (20 mM Tris pH 8, 100 mM NaCl, 0.1 mg/ml Lysozyme, 0.75 mg/ml DNAse I, ½ Complete EDTA- free protease inhibitor cocktail (Roche) tablet/10 ml, 1% DDM).

4. Immunodetection and identification of positive clones was carried out identically as described above in steps 4 and 5 from the section “Colony Filtration blot”.

3.3.2.7 Solubility screens in 96-well plates

The expression of soluble protein from all positive clones was screened in liquid culture in 96-well plates. The most important steps are described below.

1. Colony transfer to liquid LB media: A sample from all glycerol stocks from the 96 deep-well plates was picked with a 96-multi-tip picker, and inoculated into another 96 well plate containing 1ml fresh LB media + carbenicilline /well. Cells were grown overnight at 37ºC.

2. Induction: A small volume of cell culture was transferred to a new 96 deep-well plate containing 1ml LB + carbenicilline/well. Cell cultures were grown at 37ºC. Induction was performed when cell cultures reached an OD550 between 0.7 and 1 with 0.2 mM IPTG for 20h.

3. Cell harvesting and lysis: Cell plates were centrifuged at 2000g for 5-10 min and cell pellets were frozen at -80ºC. Cell pellets were resuspended for 1h at 4ºC on a shaker in 100µl native lysis buffer (50mM Tris HCl pH 8, 100mM NaCl, 0.5mg/ml lysozyme, 1 tablet/50ml PI cocktail, 10U/ml benzonase, 0.5 mM TCEP) or detergent lysis buffer (native lysis buffer supplemented with 1% FC12 or 1% DDM). Solutions were filtered by centrifugation (1500g, 15 min) through MABVN filter plates (Millipore).

4. Affinity purification: Small scale purification was performed using Ni-NTA resin (Qiagen). Filtrated lysates were applied onto MSDVN filter plates (Millipore) containing Ni-NTA resin washed with binding buffer (50mM Tris HCl pH 8, 300mM NaCl, 0.5 mM TCEP). Ni-NTA-plates containing the clarified cell lysate were incubated under agitation for 15 min at 4°C and centrifuged for 30 s at 100g. Ni-NTA plates were washed twice with 200µl/well

wash buffer (50mM Tris HCl pH 8, 300mM NaCl, 0.5 mM TCEP, 40mM imidazole) 30s at 100g. Bound protein was eluted with 20µl/well elution buffer (50mM Tris HCl pH 8, 300mM NaCl, 0.5 mM TCEP, 500mM imidazole). Elution buffer was added to the wells, and plates were first incubated at 4 ºC for 5 min, and then centrifuged for 1 min at 100g.

5. Dot blotting and SDS-PAGE: A sample of each well was spotted onto a nitrocellulose membrane. Membranes were air-dried and immunodetection was carried out following the same protocol used for the colony filtration blot, and membranes were developed with WestPico chemiluminescent reagents (Pierce). A 10 µl aliquot of each protein construct was loaded on a Nu-PAGE 4-12% Bis-Tris (Invitrogen) and run according to the manufacturers' instructions.

3.3.2.8 Sequencing

Plasmids were isolated from 96-well plates with the Millipore Montáge miniprep kit and a vacuum manifold following manufacturer's instructions. Samples were eluted in EB buffer, and prepared for in-house sequencing at the MBB department, Karolinska Intitutet (Sweden) in 96-well plates for PCR with the BigDye® Terminator v3.1 Cycle Sequencing Kit (ABI, Life technologies) following manufacturer's guidelines. Sequencing reactions from N-terminal deletions were carried out with T7 promoter primer whereas sequencing reactions from C-terminal deletions were carried out with T7 terminator primer with the reaction mix and program detailed in table 3.15 and table 3.16.

Extension products from the PCR reactions were purified with Sephadex G-50 purification plates. Purified extension products were transferred to PCR plates for sequencing, and sequenced at the MBB sequencing service on an AbiPrism sequencer. Some plasmids from each plate were purified individually, and sent for sequencing to Macrogen (Korea).

Table 3.15 Sample mix for DNA sequencing

Reagent	Sample mix (x1)
Big Dye v.3.1	0.8 µl
Dilution buffer (5x)	4 µl
Primer (8µM)*	0.25 µl
DNA template (50-100ng/µl)	8 µl
ddH_2O	Up to 20 µl
Total volume	20 µl

Table 3.16 DNA sequencing cycling conditions

Cycles (n)	Time / temperature
1 cycle	1:00 min at 96ºC
30 cycles	0:30 min at 96ºC 0:15 min at 51ºC 4:00 min at 60ºC
Hold	∞ at 4ºC

3.3.2.9 DprA medium and large scale purification

Medium scale purification

8 DprA truncations were expressed in 100 ml cultures. BL21(DE3) cells containing the plasmid of interest were induced at 20ºC for 20h with 0.1mM IPTG. Cell pellets were lysed using tip-sonicator in lysis buffer (30mM TrisHCl pH 7.5, 250mM NaCl, 30mM imidazole, 1 tablet/50 ml Protease inhibitors, DNAaseA), and lysed suspensions were centrifuged (1h, 20000g). Supernatants were filtered through a 22µm filter, and purified in two steps: affinity chromatography and size exclusion chromatography.

Affinity chromatography was carried out with 1ml Histrap FF columns in an ÄKTA Purifier system (GE Healthcare). Columns were washed with 40 CV binding buffer (30mM TrisHCl pH 7.5, 250mM NaCl, 30mM imidazole). Samples were loaded with binding buffer and eluted

with elution buffer (30mM TrisHCl pH 7.5, 250mM NaCl, 500mM imidazole) in a 20CV linear gradient from 30 to 500 mM imidazole. Size exclusion chromatography of a sample of each peak was performed using the analytical gel filtration column Superdex 200 3/150 and gel filtration buffer (30mM TrisHCl pH 7.5, 250mM NaCl).

Large scale purification

The expression and purification protocol for the best three clones (later referred to as DprA (FL), DprA (medium) and DprA (short)) was scaled up. Expression was carried out with 650 ml cultures BL21 (DE3) and the same induction conditions used in the medium scale experiments. Cell lysis and purification were carried out as in medium scale experiments but using HisTrap 1ml or 5ml columns and Superdex 200 10/300 and Superdex 200 16/60 columns.

Buffer screen for DprA(FL) purification

The assays to check the effect of buffer composition in protein oligomerization were all performed with a sample coming from a HisTrap affinity purification. Size exclusion chromatography was initially performed on a Superose 6 10/300 column equilibrated with the following buffers:

A) 650mM NaCl, 30mM Tris HCl pH 7.5
B) 500mM NaCl, 30mM Tris HCl pH 7.5
C) 350mM NaCl, 30mM Tris HCl pH 7.5
D) 250mM NaCl, 30mM Tris HCl pH 7.5
E) 150mM NaCl, 30mM Tris HCl pH 7.5
F) 250mM NaCl, 30mM Tris HCl pH 9
G) 250mM NaCl, 30mM MES pH 6.5
H) 250mM NaCl, 30mM MES pH 5.5

Additional trials were performed on a Sephacryl S300 16/60 equilibrated with the following buffers:

A) 500mM NaCl, 30mM Tris HCl pH 7.5
B) 350mM NaCl, 30mM Tris HCl pH 7.5

Samples from each peak were analyzed by DLS.

DprA (FL) purification for biochemical assays and crystallization trials
DprA(FL) for biochemical assays and crystallization trials was routinely purified from large scale cultures of BL21 (DE3) cells induced at 20°C for 20h with 0.1mM IPTG.

Cell pellets were lysed using a tip-sonicator or an Emulsifier in lysis buffer (30mM TrisHCl pH 7.5, 250mM NaCl, 30mM imidazole, 1 tablet/50 ml Protease inhibitors, DNAaseA), and lysed suspensions were centrifuged (1h, 20000g). Supernatants were filtered through a 22µm filter, and purified in two steps: affinity chromatography and size exclusion chromatography.

Gradient affinity chromatographs was performed as explained above with HisTrap 5ml columns (GE Healthcare) and the buffers: binding buffer (30mM TrisHCl pH 7.5, 250mM NaCl, 30mM imidazole), and elution buffer (30mM TrisHCl pH 7.5, 250mM NaCl, 500mM imidazole). Eluted samples were concentrated with Vivaspin devices, and injected onto the gel filtration columns Superdex 200 16/60 or Superdex 200 26/60 equilibrated with gel filtration buffer (30mM TrisHCl pH 7.5, 350mM NaCl).

DprA (short) purification for biochemical assays and crystallization trials
DprA(short) for biochemical assays and crystallization trials was routinely purified from large scale cultures of BL21 (DE3) cells induced at 20°C for 20h with 0.1mM IPTG.

Cell pellets were lysed using a tip-sonicator or an Emulsifier in lysis buffer (30mM TrisHCl pH 7.5, 250mM NaCl, 30mM imidazole, 1 tab-

let/50 ml Protease inhibitors, DNAaseA), and lysed suspensions were centrifuged (1h, 20000g). Supernatants were filtered through a 22µm filter, and purified in two steps: affinity chromatography and size exclusion chromatography.

Affinity chromatographs was performed with HisTrap 5ml columns (GE Healthcare) and the buffers: binding buffer (30mM TrisHCl pH 7.5, 250mM NaCl, 30mM imidazole), and elution buffer (30mM TrisHCl pH 7.5, 250mM NaCl 30mM, 500mM imidazole).

Samples were eluted in a single step. Eluted samples were concentrated with Vivaspin devices, and injected onto gel filtration columns Superdex 75 16/60 or Superdex 75 26/60 equilibrated with gel filtration buffer (30mM TrisHCl pH 7.5, 350mM NaCl).

3.3.2.10 ComA purification

Solubility screens in different detergents

8 ComA truncations were expressed in 100 ml cultures. BL21(DE3) cells contaning the plasmid of interest were induced at 20ºC for 20h with 0.1mM IPTG. Cell pellets were lysed using tip-sonicator in lysis buffers containing different detergents:

A. 50mM TrisHCl pH 8, 300mM NaCl, 1% DDM, 1mM DTT, 1 tablet/50 ml Protease inhibitors
B. 50mM TrisHCl pH 8, 300mM NaCl, 1% Triton X-100, 1mM DTT, 1 tablet/50 ml Protease inhibitors
C. 50mM TrisHCl pH 8, 300mM NaCl, 1% Tween 20, 1mM DTT, 1 tablet/50 ml Protease inhibitors
D. 50mM TrisHCl pH 8, 300mM NaCl, 1% Chaps, 1mM DTT, 1 tablet/50 ml Protease inhibitors

Lysed suspensions were centrifuged (1h, 20000g) and supernatants and resuspended pellets were analyzed on SDS-PAGE gels stained with Coomassie. Samples were also analyzed by Western blot.

Purification in buffers contaning DDM

8 ComA truncations were expressed in 100 ml cultures. BL21(DE3) cells contaning the plasmid of interest were induced at 20ºC for 20h with 0.1mM IPTG. Cell pellets were lysed using tip-sonicator in lysis buffers containing DDM (50mM TrisHCl pH 8, 300mM NaCl, 1% DDM, 1mM DTT, PI cocktail), and were purified in 2 steps: affinity chromatography and size exclusion chromatography.

Gradient affinity chromatographs was performed with HisTrap 1ml columns (GE Healthcare) and the buffers: binding buffer (30mM TrisHCl pH 7.5, 250mM NaCl, 1% DDM, 30mM imidazole), and elution buffer (30mM TrisHCl pH 7.5, 250mM NaCl 30mM, 1% DDM, 500mM imidazole). Size exclusion chromatography of a sample of each peak was performed using the analytical gel filtration column Superdex 200 3/150 and gel filtration buffers containing different amounts of DDM ranging between 0.01% DDM and 1% DDM).

Membrane preparations

8 ComA truncations were expressed in 100 ml cultures. BL21(DE3) cells contaning the plasmid of interest were induced at 20ºC for 20h with 0.1mM IPTG. Cell pellets were lysed using tip-sonicator in 10ml lysis buffer containing DDM (20mM TrisHCl pH 8, 50mM NaCl, 0.5mM DTT, 1:100 PI cocktail, 1mg/ml lysozyme, 10U DNase). Cell debris was removed by centrifugation (10000g, 10min).

Membranes were harvested in an ultracentrifuge (350000rpm, 50min). Supernatants were removed, and membranes were resuspended in a potter homogeniser in solubilization buffer (20mM Na-P pH 7.5, 300mM NaCl, 0.5mM DTT, 1:100 PI cocktail). Membranes were batched and

flash frozen in liquid nitrogen. Samples of all fractions were analyzed on SDS-PAGE gels stained with Coomassie and by Western blot.

3.3.3 ComE, ComA and DprA characterization

3.3.3.1 ComE(B) crystallization trials

Crystallization trials with His-tagged and non-His-tagged protein were performed as indicated in the general methods section at 20ºC and at 4ºC with protein concentrations ranging between 10 and 80 mg/ml with protein in several buffers with different ionic strength, and pH.

Initial crystallization screens were carried out in 96-well plates using general screens at the PCB Crystallography platform. Some conditions were optimized in 24-well plates.

3.3.3.2 ComE(B) 1-D NMR

ComE(B) 1D-NMR experiments were carried out with His-tagged protein at the PCB NMR facilities.

A protein sample at 3mg/ml in gel filtration buffer (30mM Tris HCl pH 7.5, 500mM NaCl) was two-fold and four-fold diluted in ddH_2O. For the DNA binding assays dsDNAextDUS was generated by annealing the primer pairs "F_DNAextDUS + R_DNAextDUS" (Table 3.17) following the protocol for DNA annealing detailed in the general methods. Final dsDNA concentration was 2mM.

Table 3.17 Oligonucleotides for ComE 1-D NMR

Primers	
F_extDUS	GCATCGCCGTCTGAAGCACG
R_extDUS	CGTGCTTCAGACGGCGATGC

3.3.3.3 ComA crystallization trials

Crystallization trials with His-tagged ComA(C) obtained by refolding were performed as indicated in the general methods section at 20ºC and 4ºC with protein concentrations ranging between 3 and 10 mg/ml with protein in several buffers with different ionic strength and pH.

Initial crystallization screens were carried out in 96-well plates using Hampton Research general screens. The most promising conditions were optimized in 24-well plates.

3.3.3.4 DprA limited proteolysis assays

Limited Proteolysis assays were performed with His-tagged DprA(FL) at 1mg/ml.

Reactions with trypsin (Sigma) and chymotrypsin (Sigma) were carried out in ten-fold dilutions of 10X trypsin buffer (0.45M Tris pH8) and 10X chymotrypsin buffer (1M $CaCl_2$, 0.8M Tris pH7.8) at 2 temperatures (22ºC and 37ºC), at three different protease concentrations (1/20, 1/200, 1/2000).

Proteolytic reactions were inactivated at different time points (t=0, t=10min, t=30min, t=90') by adding loading buffer, and boiling the samples. Proteolyzed samples were analyzed on SDS-PAGE gels stained with Coomassie, and by Western blot.

3.3.3.5 DprA cross-linking assays

Crosslinking assays were performed with His-tagged DprA(FL) at 1mg/ml. Since Tris based buffers were incompatible with crosslinking assays, the gel filtration column was equilibrated with a NaP buffer (NaP buffer pH7.5, 300mM NaCl).

Reactions were carried out with two crosslinkers: dimethyl suberimidate•2 HCl (DMS) and dimethyl 3,3´-dithiobispropionimidate•2 HCl (DTBP) (both from Thermoscientific) at 2 temperatures (22ºC and

4ºC), at four crosslinker concentrations (0.2mM, 1mM, 2mM, and 5mM). Reactions were inactivated after one hour (samples at 22ºC), after 2 hours (samples at 4ºC), or after 20 hours (samples at 4ºC) by adding TrisHCl pH 8 to a final concentration of 50mM.

Cross-linked samples were analyzed on SDS-PAGE gels stained with Coomassie, and by Western blot.

3.3.3.6 DprA DNA binding assays

DNA binding assays were performed with His-tagged short DprA (clone 1H9).

Assays were carried out in buffer E (30mM TrisHCl pH 7.5, 50mM NaCl) with dsDNA (annealed FextDUS + RextDUS and annealed FDUS + RDUS), and ssDNA (FextDUS, RextDUS, FDUS, RDUS, ssDNA(20T), ssDNA(20C), ssDNA(10T), and ssDNA(10C)) generated by annealing the primers detailed in table 3.18. Assays were carried out at different protein:DNA ratios.

DNA was incubated with protein for 30 minutes on ice, and samples were loaded on EMSA gels prepared as indicated in the general methods section. Samples ran at 20mA for 1h in TAE buffer. Gels were stained with Coomassie or with SyberGold to see the relative positions of protein and DNA.

Table 3.18 Oligonucleotides for EMSA with DprA (FL)

DprA EMSA	
Poli-T(10)	TTTTTTTTTT
Poli-T(20)	TTTTTTTTTTTTTTTTTTTT
Poli-C(10)	CCCCCCCCCC
Poli-C(20)	CCCCCCCCCCCCCCCCCCCC
F_DUS	GCCGTCTGAA
R_DUS	TTCAGACGGC
F_extDUS	GCATCGCCGTCTGAAGCACG
R_extDUS	CGTGCTTCAGACGGCGATGC

3.3.3.7 DprA crystallization

Crystallization trials were performed as indicated in the general methods section at 20ºC and at 4ºC. Initial crystallization screens were carried out on 96 well plates using Hampton Research general screens. The most promising conditions were optimized in 24-well plates.

In the case of DprA(FL), crystallization trials were performed with protein concentrations ranging between 5 and 20 mg/ml. Crystallization trials with trypsin were carried out by adding trypsin to the protein prior to drop setting, keeping a ratio protease:protein 1:1000.

In the case of DprA (short) crystallization trials were performed with protein concentrations ranging between 5 and 50 mg/ml.

3.4 Materials and methods: Toxin-antitoxin systems

This section describes the steps for the crystallization trials and structural characterization of *S. pneumoniae* RelBE2. Section 3.4.1 describes protocols for sample preparation (expression, purification, annealing and purification of protein-DNA complexes and section 3.4.2 describes crystallization and structural characterization.

3.4.1 Sample preparation

3.4.1.1 RelBE2 expression and purification

All experiments were performed with protein complex expressed and purified in the laboratory of Prof. Manuel Espinosa (CIB-CSIC) according to the protocol summarized below.

S. pneumoniae RelBE2 protein complex was expressed in BL21(DE3) cells transformed with a vector containing a copy of RelE with a C-terminal His-tag (6) and a copy of native RelB.

RelBE2 protein complexes were purified in 2 steps: affinity purification and gel filtration. Affinity purification was performed with nickel columns equilibrated with equilibration buffer (20mM TrisHCl pH 7.6, 1mM EDTA, 5% ethilenglycol, 10mM imidazole, 1mM β-mercaptoethanol, 500mM NaCl), and eluted with elution buffer (20mM TrisHCl pH8, 5% ethilenglycol, 250mM imidazole, 1mM β-mercaptoethanol, 500mM NaCl). Samples were dyalized against gel filtration buffer (20mM TrisHCl pH 7.6, 1mM EDTA, 5% ethilenglycol, 0.1mM DTT, 500mM NaCl), and purified in the same buffer with a Superdex 200. Protein samples were flash frozen in liquid nitrogen, and kept at -80ºC.

3.4.1.2 RelBE2-DNA complex preparation

The putative operator region of *RelBE2* was determined by DNAase footprinting assays and EMSA in the laboratory leaded by Prof. Manuel Espinosa as indicated in (Moreno-Córdoba *et al.*, 2012). According to these results, three pairs of primers covering the operator region were designed: DNA(30), DNA(32), and DNA(34) (table 3.19).

Table 3.19 Primers for crystallization (toxin-antitoxin project)

RelBE2 crystallization	
F_RelBE2*Spn*30	CCTAAAAAGTAATACAATGGTGTTACCATT
R_RelBE2*Spn*30	AATGGTAACACCATTGTATTACTTTTTAGG
F_RelBE2*Spn*32	CCTAAAAAGTAATACAATGGTGTTACCATTAA
R_RelBE2*Spn*32	TTAATGGTAACACCATTGTATTACTTTTTAGG
F_RelBE2*Spn*34	CCTAAAAAGTAATACAATGGTGTTACCATTAAAA
R_RelBE2*Spn*34	TTTTAATGGTAACACCATTGTATTACTTTTTAGG

Primers were diluted in ddH_2O to a final concentration of 2mM. Complementary pairs of primers were mixed. Samples were denatured in a water bath at 80ºC for 30min and annealed overnight by gradual cooling after turning off the water bath.

RelBE2 was slowly added to annealed DNA until the protein/DNA molar ratio suggested by EMSA studies was reached. The proposed protein/DNA ratio was 5/1, but DNA was put in excess (20%) reaching a final Protein/DNA ratio 4/1 to ensure that all protein would bind DNA.

The complex protein-DNA was purified by gel-filtration with a Superdex 200 in gel filtration buffer (20mM TrisHCl pH 7.6, 1mM EDTA, 0.1mM DTT, 500mM NaCl). Samples were concentrated with Vivaspin devices until they reached the appropriate concentration for crystallization.

3.4.2 Crystallization and structural characterization

3.4.2.1 RelBE2 crystallization

Initial crystallization trials were performed with native protein/DNA complex (30 pb, 32 pb and 34 bp) under general crystallization screens. Protein concentration in the initial crystallization drops was 10-12 mg/ml. Nanodrops (0.1µl protein/DNA complex + 0.1µl reservoir) were set in MRC 96-well plates (Hampton) containing 50µl reservoir and were incubated at 20ºC and 4ºC.

Trials with SeMet protein/DNA complex and with native protein alone were performed in MRC 96-well plates (Hampton) containing general crystallization screens. Plates were incubated at 20ºC and 4ºC.

The most promising hits from initial plates were optimized in 24-well plates by changing pH, precipitant and protein concentration up to a final protein concentration of 16mg/ml.

Protein crystals were fished with nylon cryoloops, and momentarily soaked in solutions containing reservoir solution supplemented with variable amounts of cryoprotectant agents. Crystals obtained in cryoprotectant buffers (containing 30-40% MPD) were frozen without supplementing further cryoprotection. Cryoprotected crystals were flash frozen in liquid nitrogen.

3.4.2.2 RelBE2 crystal data collection and processing

All diffraction data sets were collected at the ESRF (Grenoble, France) at the beamlines 23.2, 14.2 (native protein), and 14.4 (Se-Met labeled protein).

Diffraction data were indexed and integrated with iMosflm (Leslie and Powell, 2007) and XDS (Kabsch, 2010) and scaled and merged with Scala (Evans, 2006).

3.4.2.3 RelBE2 structural determination

Initial phasing trials using molecular replacement were carried out with MolRep (Vagin and Teplyakov, 2010) and Phaser (McCoy *et al.*, 2007) using models for the toxin subunit (RelE) from homologous proteins. To that end, poly-alanine models based on RelE structures from *Mycobacterium tuberculosis* (PDB:3G5O) and *Pyrococcus horikoshii* (PDB ID:1wmi) and on models calculated by Swiss-Model [Web10] were generated with CCP4i tools for PDB editing [Web11]. Molecular replacement was performed using data from native crystals that diffracted up to 3.6Å.

Since molecular replacement failed, phasing was performed by using multiple-wavelength anomalous diffraction (MAD) with data collected at two different wavelengths (peak and inflection point wavelengths for the selenium absorption edge) from a single SeMet-labeled RelBE2 crystal that diffracted up to 3Å.

The substructure of Se atoms, experimental MAD electron density maps and an initial model were calculated with AutoRickshaw (Panjikar *et al.*, 2005). 10 selenium sites were identified with ShelxD (Sheldrick, 2010) and refined with Mlphare (Otwinowski, 1991) using data up to 3.5Å. Initial experimental phases were improved by non-crystallographic symmetry averaging and statistical density modification with Pirate and Resolve (Cowtan, 2000). An initial RelBE2 model was built with Buccaneer (Cowtan, 2006).

This model was manually corrected with Coot (Emsley and Cowtan, 2004). Iterative rounds of model building with Coot and refinement with Refmac (Murshudov *et al.*, 2011) were performed using TLS and non-crystallographic symmetry restraints.

The refined SeMet model was then used as a search model for molecular replacement using MolRep (Vagin and Teplyakov, 2010) to determine the structure of native RelBE2 at 2.5Å resolution. Multiple rounds of

refinement with Refmac (Murshudov *et al.*, 2011) followed by model rebuilding were done until Rwork and Rfree converged to 24.5% and 29.2% respectively.

Structural similarity searches were carried out using the Dali server (Holm and Rosenström, 2010). Electrostatic surface analysis was performed with EF-Site (Kinoshita and Nakamura, 2003). All figures were made with Pymol (PyMOL).





Chapter 4
Results & Discussion





4.1 Results and discussion: Natural transformation

This section describes the results of expression, purification, crystallization trials and biochemical characterization of three proteins involved in natural transformation in *N. gonorrhoeae:* ComE (section 4.1.1), ComA (section 4.1.2) and DprA (section 4.1.3).

4.1.1 ComE: results and discussion

4.1.1.1 Bioinformatics analysis

After the bioinformatics analysis, ComE was described as a globular, mainly alpha-helical protein with an N-terminal hydrophobic alpha-helical peptide (residues 1-25), a central alpha-helical folded region (26-81) and a disordered region at the C-terminal end (82-99). These regions are depicted in Fig. 4.1. Throughout the text, ComE full-length is referred to as ComE(FL). ComE truncations are referred according to Fig. 4.1. as ComE(BC) and ComE(B).

Since ComE had been previously described as a periplasmic protein the N-terminal hydrophobic region (region A in Fig. 4.1) was hypothesized to be a signal peptide for periplasmatic export. This hypothesis was confirmed by SOSUI signal.

ComE is a DNA binding protein. Its very basic IP (9.78) due to the relative abundance of lysine residues and the helix-hairpin-helix DNA binding motif class 1 located between residues 59 and 78 would explain ComE affinity for DNA.

ComE homologues have been described in other species. Two examples are ComEA/CelA from *T. thermophilus* and ComEA from *S. pneumoniae*. Homologous proteins are larger (298 residues and 216 residues re-

spectively) and membrane associated. The overall identity and similarity between these proteins and *N. gonorrhoeae* ComE is low, but both proteins contain a C-terminal region that shares more than 30% identity with *N. gonorrhoeae* ComE.

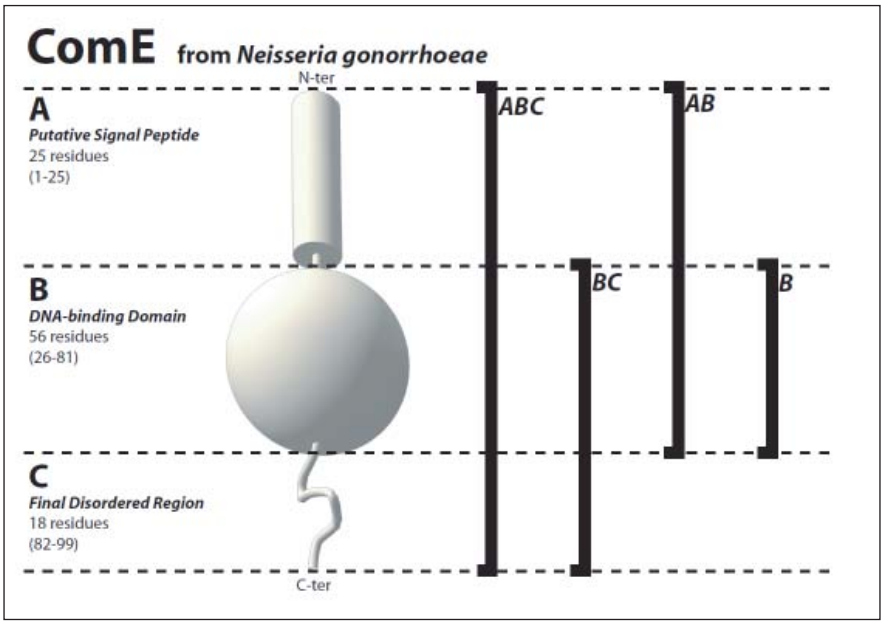


Fig. 4.1 ComE: definition of domain borders for construct design.

4.1.1.2 Cloning

ComE(FL) was successfully cloned into pGFPCR, pPROEXHTa, pET28a, and pBAT4. *ComE(BC)* and *comE(B)* were successfully cloned into pPROEXHTa, pET28a, pBAT4 and pET22b, a vector for periplasmic expression.

4.1.1.3 Expression and purification

ComE (FL) expression was very poor and resulted in the production of insoluble protein. ComE(BC) and ComE(B) expression was good (Fig 4.2) and resulted in the production of soluble protein. Expression yields were significantly better for ComE(B) (Fig 4.3).

Expression with *E. coli* signal peptide for periplasmic export or expression at low temperature did not improve solubility. Best expression/solubility results were obtained after expression in Rosetta cells transformed with pPROEX HTa (ComE(B)) induced for 3h at 37ºC (Fig. 4.2, Fig. 4.3).

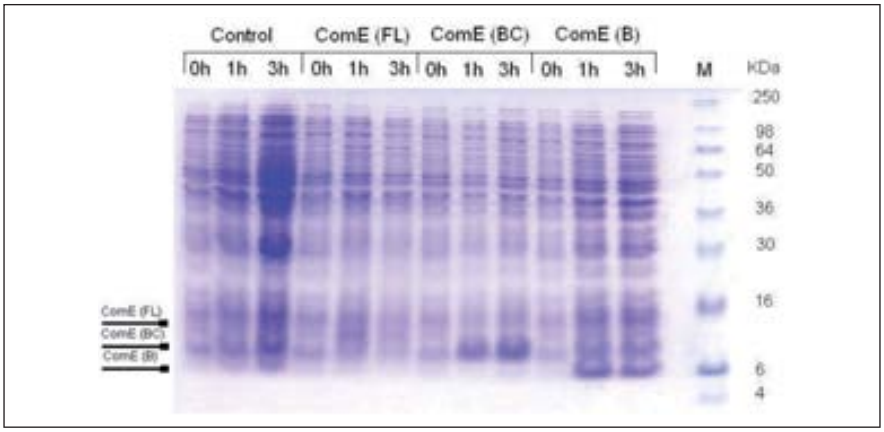


Fig. 4.2 His-ComE expression trials with pPROEXHTa expression vector: M = molecular weight marker, 0h = before induction, 1h = 1h induction, 3h = 3h induction.

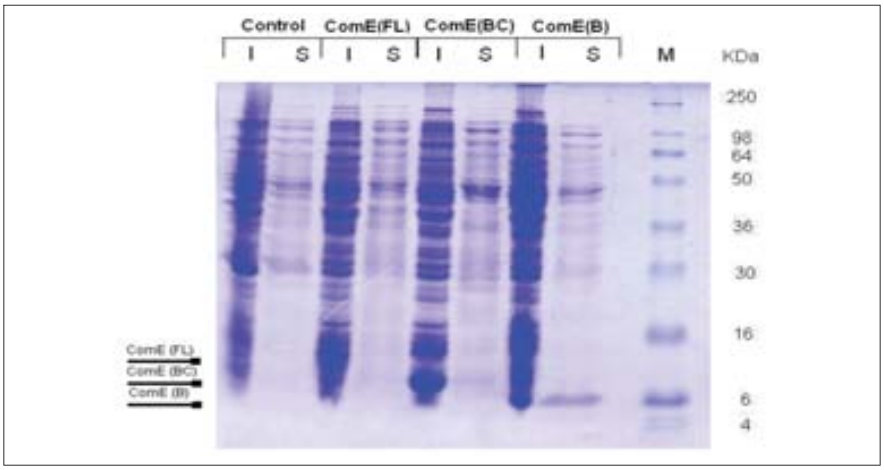


Fig. 4.3 His-ComE solubility trials: M = molecular weight marker, I = insoluble fraction, S = soluble fraction.

As indicated in the methods sections, IMAC and heparin-based columns were compared as initial purification steps of his-tagged. Heparin-based purifications were rejected because the purity of the sample after purification was not satisfactory. Therefore ComE(B) was routinely purified by combining affinity and size exclusion chromatography. IMAC purifications provided a single peak elution that corresponded to a single band in SDS-PAGE (Fig 4.4, Fig. 4.6).

After IMAC, the ratio 260/280 was high, indicating that protein co-purified with substantial amounts of DNA from the host strain. The high isoelectric point would explain why the protein binds so strongly to DNA from the host strain during purification. Since ComE is supposed to bind DNA fragments that enter the periplasm in an unespecific manner, the high affinity for unspecific dsDNA observed during purification is relevant from a biological point of view.

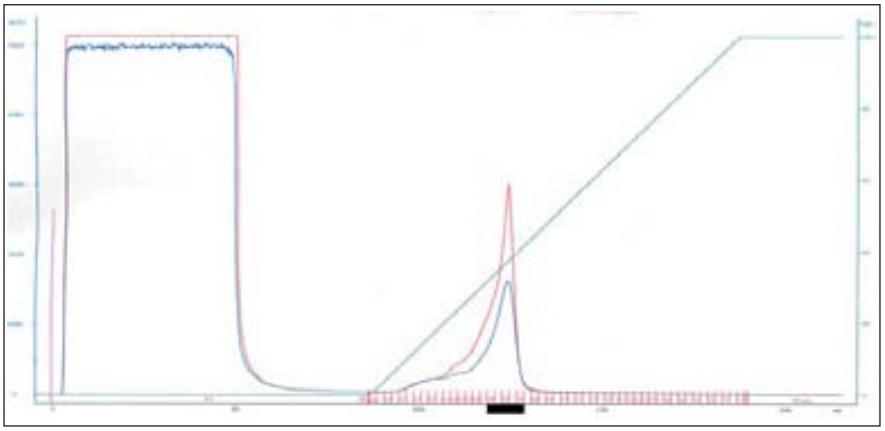


Fig. 4.4 His-ComE IMAC purification: Green line (imidazole gradient), blue line (Abs 280), red line (Abs 260). Black line indicates fractions collected for SEC.

DNA was successfully removed by using high NaCl concentration during gel filtration. After gel filtration, ComE(B) eluted in a single peak corresponding to the size of a monomer (Fig 4.5). Final yield was 3 mg/l cell culture.

ComE(B) was also purified without the N-terminal histidine tag. As explained in the methods section protein was purified in several steps: IMAC, desalting, overnight cleavage with TEV protease, IMAC and SEC. After gel filtration protein eluted in a single peak corresponding to the size of a monomer (Fig 4.7). ComE(B) appeared as a single band on SDS PAGE gels. Final yield was 1mg/l cell culture.

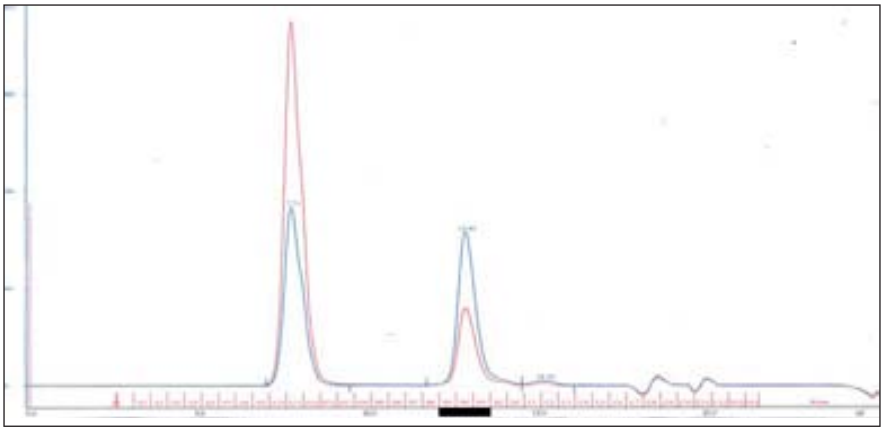


Fig. 4.5 His-ComE SEC purification:, blue line (Abs 280), red line (Abs 260). Black line indicates fractions pooled and concentrated.

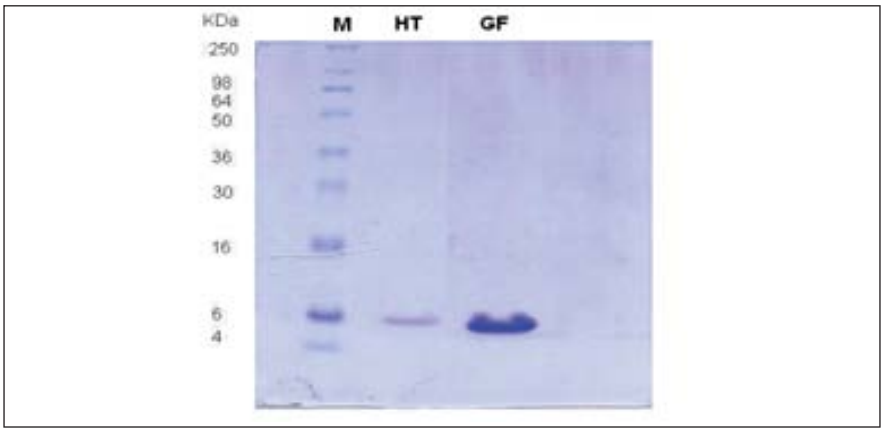


Fig. 4.6 His-ComE purification: M = molecular weight marker, HT = IMAC purified protein, GF = SEC purified protein.

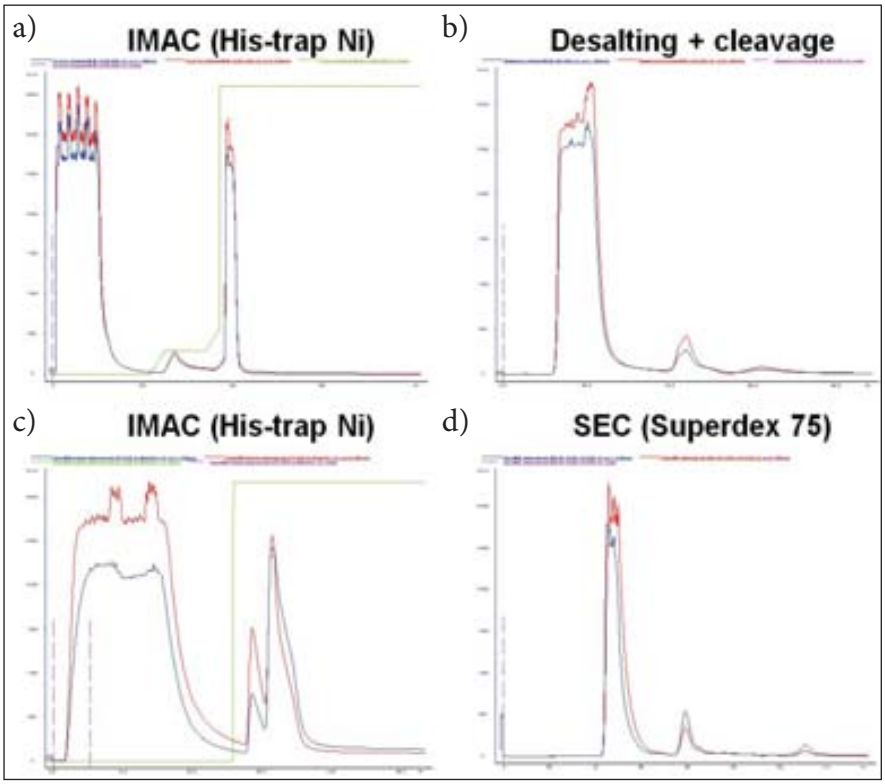


Fig. 4.7 ComE(B) purification: a) 1st IMAC previous to His-tag cleavage, b) Desalting previous to His-tag cleavage, c) 2nd IMAC after His-tag cleavage, d) Final SEC. Green line (imidazole step), blue line (Abs 280), red line (Abs 260).

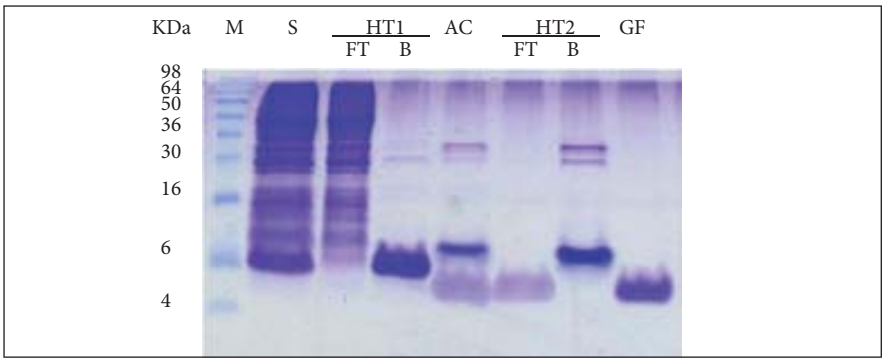


Fig. 4.8 ComE(B) purification steps. M = molecular weight marker, S = soluble fraction, HT1 = 1st IMAC previous to His-tag cleavage, AC = Sample after cleavage, HT2 = 2nd IMAC after His-tag cleavage, GF = gel filtration, FT = flow through, B =bound sample.

4.1.1.4 Crystallization trials

Crystallization trials with ComE(B) at 10mg/ml were unsuccessful. Negative results in initial crystallization screenings suggested to increase protein concentration first up to 20 mg/ml, and lately up to 60 and 80mg/ml. Increasing protein concentration lead to spherulite formation and phase separation, but protein was reluctant to crystallize. Fig. 4.9 shows some hits obtained in 96 well- trials that were optimized without any success. Crystallization trials without the histidine tag gave similar results.

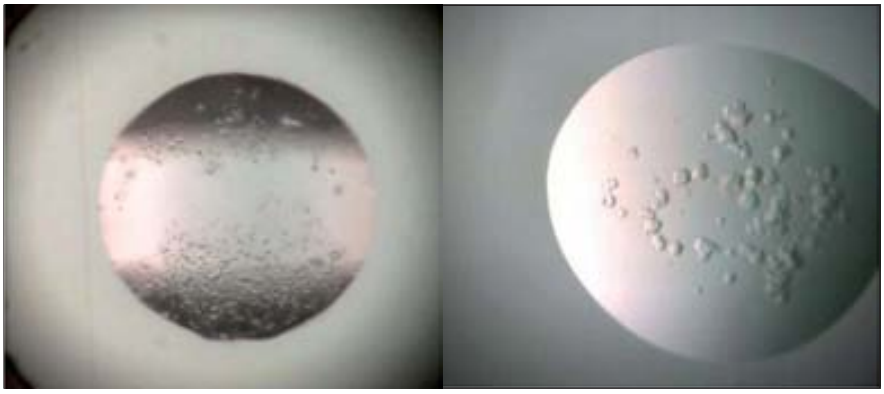


Fig. 4.9 ComE(B) crystallization trials showing promising hits.

4.1.1.5 Unidimensional NMR

Given the difficulties in the crystallization step, we performed 1H-NMR of ComE (B) alone so as to assess the crystallizability of the protein and to investigate the possibility to study the protein by NMR, and ComE(B) in the presence of DNA to study ComE(B) DNA binding behaviour.

In 1H-NMR methyl groups and amide protons from the sample are used to assess the "foldedness" of the protein. As a rule of dumb, in unfolded proteins chemical shifts for methyl groups are between 0.8 and 1.2ppm, and chemical shifts for amide protons are between 7.5 and 8.5 ppm. By contrast, in globular proteins methyl groups are shifted by their closeness to aromatic rings giving signals below 0.5ppm, and

backbone amide protons are up-field shifted through hydrogen bonding and can give signals above 8.5 ppm.

In the 1H-NMR spectrum for ComE(B) alone, the narrow distribution of methyl groups (close to 0.8 ppm), and amide groups (between 7.75 and 8.3 ppm) would be characteristic of an unfolded protein (Fig. 4.10).

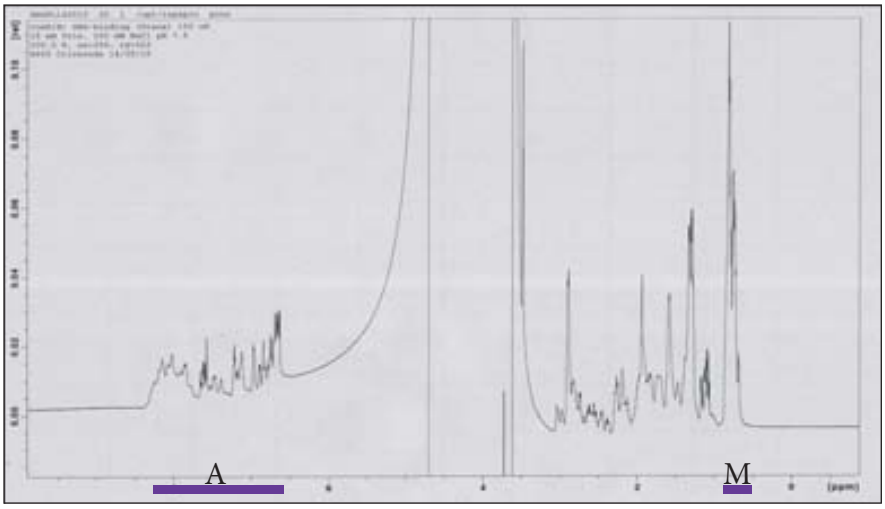


Fig. 4.10 1H-NMR spectrum of ComE(B) alone. The spectrum shows the narrow distribution of methyl groups (M) and amide groups (A).

ComE(B) DNA binding behavior was analyzed by analyzing the chemical shift in 1H NMR spectra due to the presence of double stranded DNA in low and high ionic strength buffers.

1H NMR spectra ComE(B) in presence of dsDNA and ComE(B) alone are shown in Fig 4.11. At high ionic strength, the spectrum for ComE(B) in presence of dsDNA is very similar to that of ComE(B) alone (A line in Fig. 4.11), indicating that in this condition the interaction between protein and DNA would be weak. By contrast, at lower ionic strength there are signals at 0 ppm or even negative values that would correspond

to methyl groups close to aromatic groups or to DNA bases (B line in Fig. 4.11).

Interestingly, at low ionic strength there are also changes in DNA imino protons induced by the protein. These changes are evident in Fig 4.12 which is a superposition of the spectra corresponding to the DNA alone (grey in Fig. 4.12), and to the DNA in the presence of protein (black in Fig. 4.12).

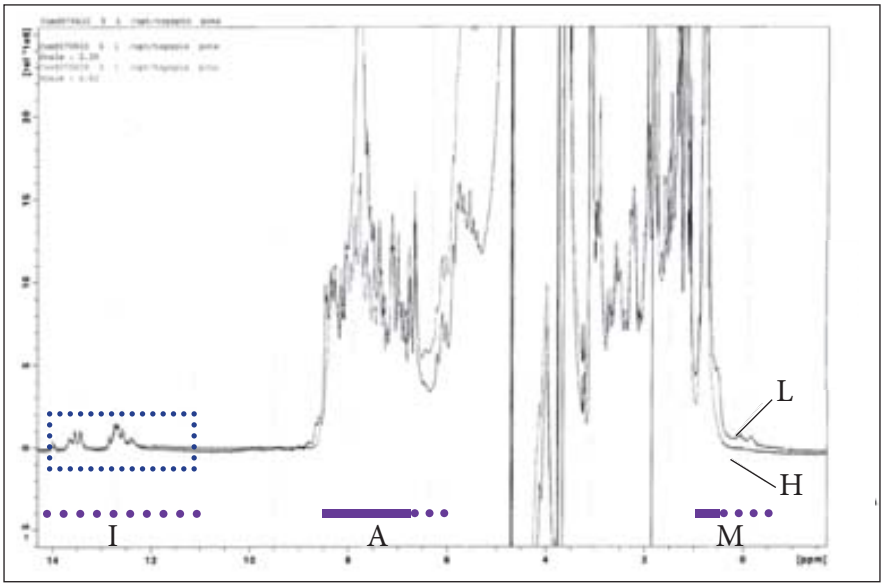


Fig. 4.11 Superimposed 1H-NMR spectra of ComE(B) + DNA in high ionic strength buffer (H line) and in low ionic strength buffer (L line). Spectra show changes in the distribution of methyl groups (M), amide groups (A) and imine groups (I). Changes with respect to the spectrum of ComE(B) alone (Fig. 4.10) are indicated with dotted lines. The region corresponding to that represented in Fig. 4.12 is framed with a dotted square.

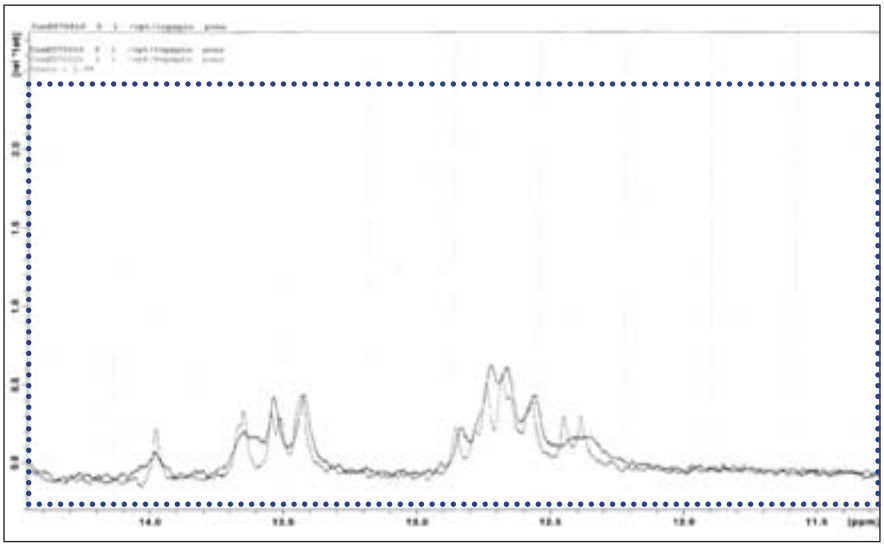


Fig. 4.12 Superimposed 1H-NMR spectra of ComE(B) plus DNA (black) and DNA alone (grey) in low ionic strength buffer.

Altogether these results indicate that there is a clear interaction between protein and dsDNA, but also that dsDNA might have an important effect in ComE folding. This behavior has been described for other DNA binding proteins such as HES-1, a transcription factor of the basic helix-loop-helix (bHLH) family that contains an N-terminal region that undergoes a coil-helix transition upon binding to DNA (Coglievina *et al.*, 2010). DNA binding properties should be further analyzed. Complementary DNA binding assays to assess if DNA binding is totally unspecific or if there is preference for AT or GC rich sequences, and NMR studies with protein marked with 15N or DNA labelled with 31P to determine the residues involved in DNA binding would be extremely useful to better characterize ComE binding to DNA.

In any case, it is worth to keep in mind that ComE(B) is predicted to be a mainly alpha-helical protein with a few aromatic residues. Therefore, chemical shifts due to the formation of secondary structure elements or to the proximity to aromatic rings are expected to be low.

4.1.1.6 Structural analysis

During the course of this project, the C-terminal domain of a homologous protein, ComEA from *T. thermophilus*, was characterized by a structural genomics consortium. ComEA atomic coordinates can be accessed at the Protein Data Bank (2DUY). The high identity between ComE(B) and the C-terminal domain of this homologue (higher than 40%), allowed the building of a confident ComE(B) model. ComE(B) appeared to be a alpha-helical protein with an overall positively charged surface due to the abundance of lysines (Fig 4.13).

In addition, the structural analysis of a model of ComE(B) truncation provided a hypothesis that would explain ComE(B) reluctancy to crystallize. In comparison to *T. thermophilus* ComEA, our engineered version of *N. gonorrhoeae* ComE(B) would lack a short β-strand between N-terminal and C-terminal residues (Fig.4.14). The presence of this secondary element could be crucial for DNA folding. The absence of this critical β-strand in our construct together with the surface charge might explain why the protein was so reluctant to crystallize.

It would be interesting to work with new clones of *N. gonorrhoeae* ComE to see if the missing β-strand is crucial for ComE stability. This could be done by removing a shorter number of residues at N-terminal and keeping the C-terminal end.

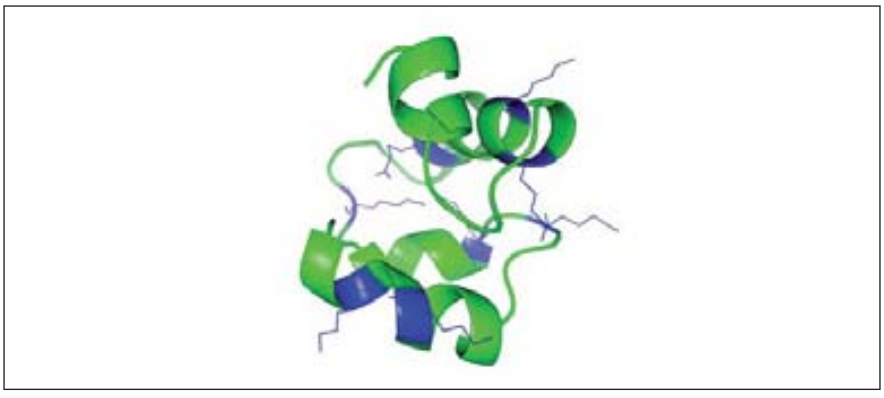


Fig. 4.13 Model of *N. gonorrhoeae* ComE(B) structure highlighting the abundance of lysine residues.

Fig. 4.14 *T. termophilus* **ComEA truncation contains the β–strand missing in** *N. gonorrhoeae* **ComE(B) truncation.**

4.1.2 ComA: results and discussion

4.1.2.1 Bioinformatics analysis

ComA was described as a putative channel for DNA transport across the internal membrane. The very basic isoelectric point (IP = 9.8), calculated from the amino acid sequence, suggested the presence of positively charged surfaces that could bind DNA but sequence analysis did not provide any additional information regarding the presence of DNA binding motifs. Charged residues were distributed unevenly throughout the protein being preferentially located at both termini of the protein.

Given the difficulties associated to the work with membrane proteins, we decided to remove the transmembrane regions from the initial clones and focus on the parts of the protein that were predicted to be soluble. The prediction of transmembrane helices by a number of bioinformatics tools provided several solutions that were in disagreement in the number and location of the helices, but agreed in the definition of a central hydrophobic region composed of 6 to 8 transmembrane helices located between residues 180 and 435 (Fig. 4. 15).

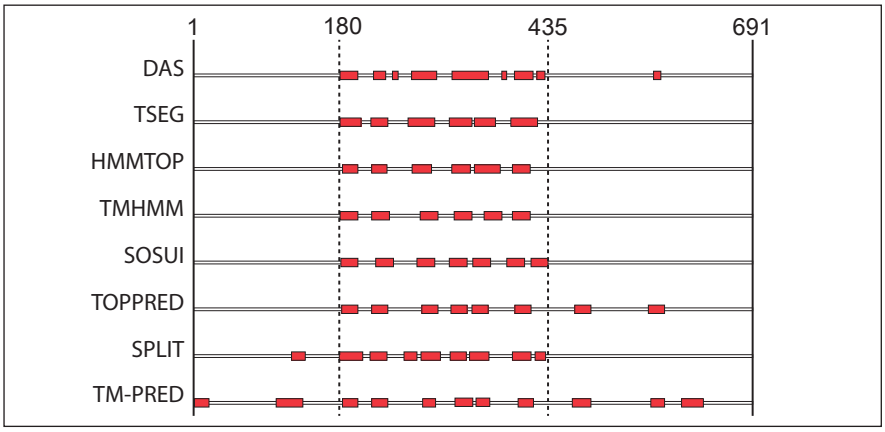


Fig. 4.15 Prediction of ComA transmembrane helices with the bioinformatic tools DAS, TSEG, HMMTOP, TMHMM, SOSUI, TOPPRED, SPLIT and TM-PRED. Transmembrane regions are indicated as red boxes. Position and length of each helix can be inferred from the numbers indicated above (numbers refer to amino acid position relative to the first methionine).

Secondary structure prediction tools indicated the presence of an N-terminal region alternating coiled regions and β-strands, a central α-helical region and a C-terminal region alternating α-helices, β-strands and coiled regions. Globularity and disorder predictions confirmed the presence of a highly ordered central region flanked by two globular regions with slightly higher disorder values.

With all the data coming from the bioinformatic analysis we defined four regions: an N-terminal soluble domain (later referred to as ComA(A)), a central hydrophobic region (later referred to as ComA(B)), a C-terminal soluble domain (later referred to as ComA(C)), and a small disordered C-terminal region (Fig. 4.16). We initially focused on the structural characterization of the domains ComA(A) and ComA(C) assuming that they could be treated as soluble regions.

4.1.2.2 *ComA(A)* and *comA(C)* cloning

ComA(A) and *comA(C)* were successfully cloned in the expression vectors pPROEX HTa, pET28a, pBAT4, pET22b, pETM40, pBADM60 and pBADM40.

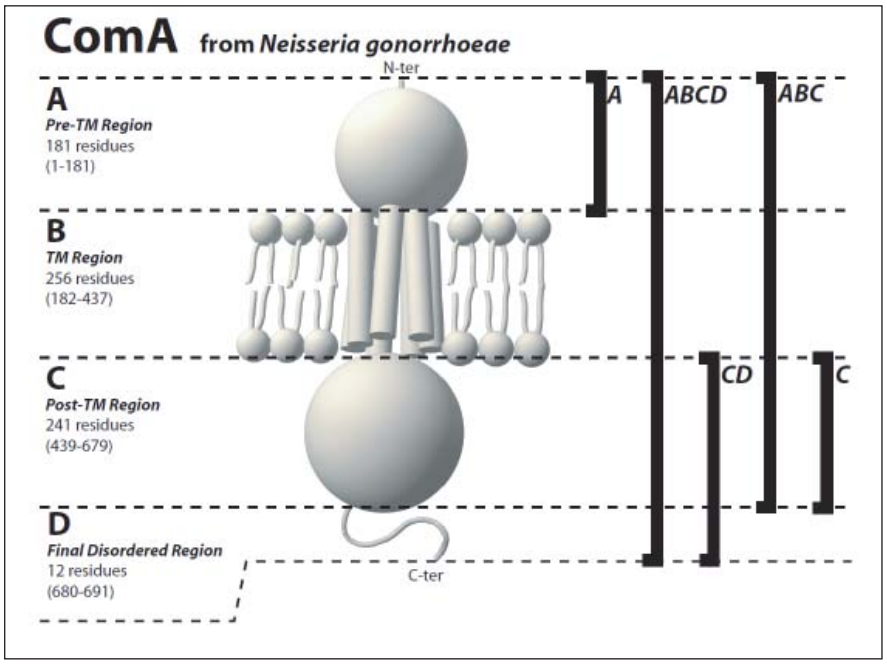


Fig. 4.16 ComA: definition of domain borders for construct design.

4.1.2.3 ComA(A) and ComA(C) expression and purification trials

Expression trials of His-tagged or native ComA(A) and ComA(C) resulted in the production of significant amounts of insoluble protein (Fig. 4.17). The modification of growth temperature, inducing agent (IPTG or arabinose), induction time or inductor concentration did not result in any substantial improvement of solubility.

Solubility screens with small amounts of urea, detergents or high salt concentration did not help in the solubilization of ComA(A). By contrast, urea at non-denaturing concentrations (below 2M) was useful to solubilize small quantities of ComA(C) (Fig. 4.18). Since purification of ComA(C) in buffers containing 2M urea was not successful we decided to try denaturing and refolding protocols. ComA(C) purification under denaturing conditions

resulted in the isolation of a small peak and a large peak (Fig. 4.19). Fractions from the large peak were collected, diluted and step-wise re-folding as explained in the methods section.

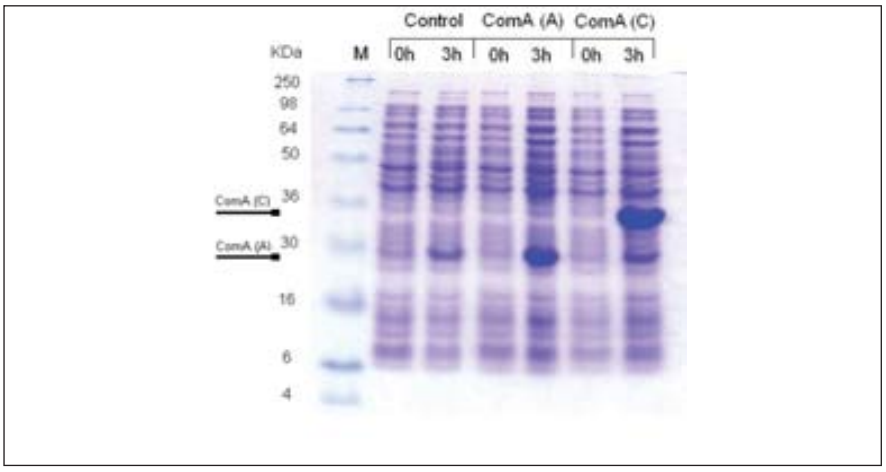


Fig. 4.17 ComA: expression trials of His-ComA(A) and His-ComA(C): M = molecular weight marker, 0h = before induction, 1h = 1h induction, 3h = 3h induction.

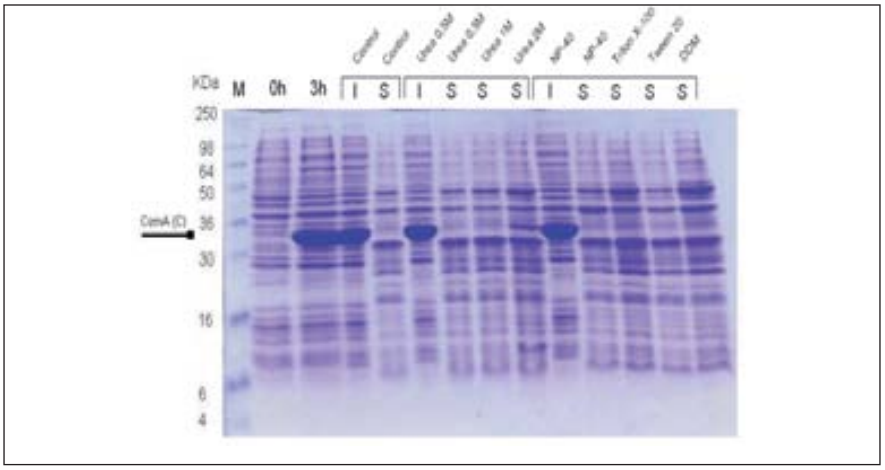


Fig. 4.18 His-ComA(C): solubility screens: M = molecular weight marker, 0h = before induction, 1h = 1h induction, 3h = 3h induction, I = insoluble fraction, S = soluble fraction.

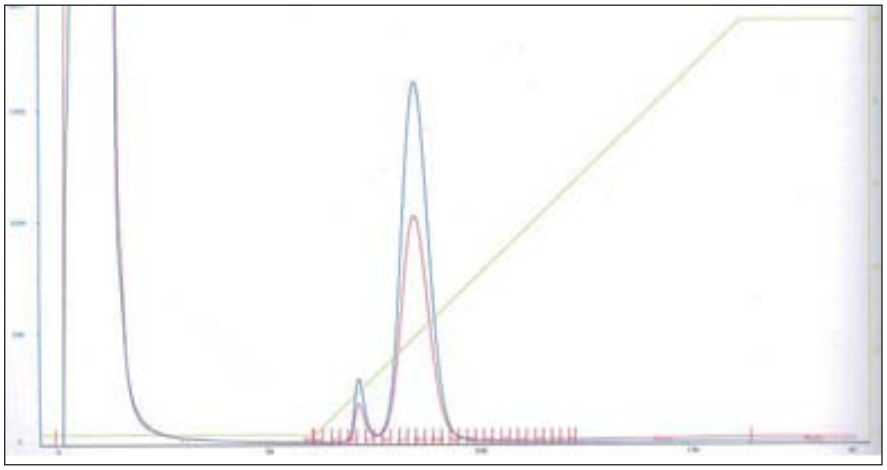


Fig. 4.19 His-ComA(C) IMAC purification under denaturing conditions. Green line (imidazole step), blue line (Abs 280), red line (Abs 260).

During step-wise refolding a substantial amount of protein precipitated resulting in the loss of approximately 90% of initial protein. After gel filtration, there was no sign of aggregation and refolded protein eluted in two peaks: a main peak corresponding to the size of the monomer and a small peak with a maximum at 260 nm that could be easily separated from the main peak (Fig. 4.20). Eluted fractions appeared as a single band on SDS-PAGE gels (Fig. 4.21). Final yield was 4 mg/l cell culture.

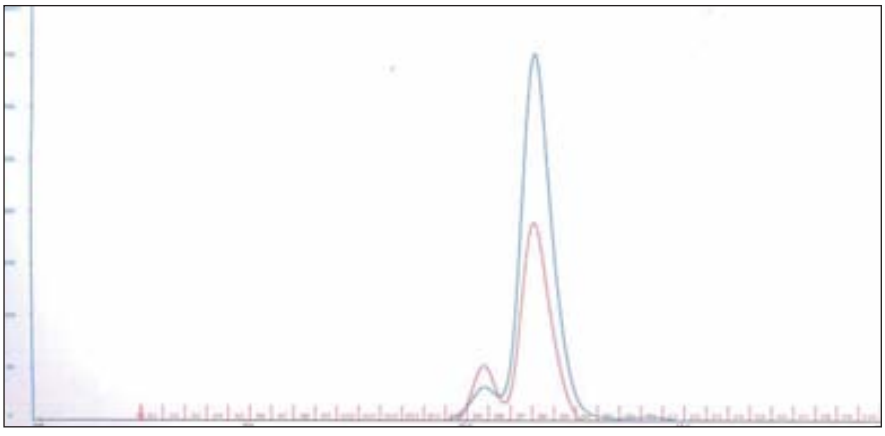


Fig. 4.20 His-ComA(C) SEC purification in native buffer after IMAC purification under denaturing conditions and refolding: blue line (Abs 280), red line (Abs 260).

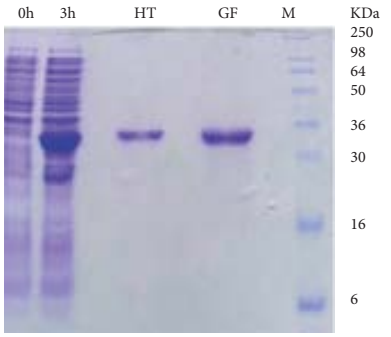


Fig. 4.21 His-ComA(C) purification under denaturing conditions: M = molecular weight marker, HT = IMAC purified protein, GF = SEC purified protein.

Despite sample purity, protein crystallization was unsuccessful probably due to the lack of long-term stability of the protein.

Protein expression with larger solubility tags (GST, NusA and MBP) allowed us to obtain small amounts of both ComA(A) and ComA(C), but most protein came out in the void volume.

4.1.2.4 Deletion libraries

As indicated in the methods section, *comA* deletion libraries were designed so as to generate clones with N-terminal or C-terminal deletions that would cover the whole sequence of full-length ComA, but also the sequence of a C-terminal truncation, later referred to as ComA(CD).

The representativeness or lack of representativeness of each library was initially assessed by the number of colonies obtained after transformation of competent cells with the ligation products for each library. Plates with too few colonies were systematically discarded.

The integrity of the vector and the length of the deleted inserts were assessed by restriction digest with restriction enzymes targeting restriction sequences flanking the insert as indicated in the methods section. Some common problems were:

- Digestions performed at 4ºC were too slow and did not cover the whole ComA sequence (Fig. 4.22a).
- Digestions performed on inserts containing the restriction pairs SacI/NdeI were not successful because SacI did not protect the vector terminal end and deletion was bidirectional (Fig. 4.22b).
- If digestion with the enzyme that would generate a non-protected end was not optimal there was no deletion (Fig. 4.22c).

Best deletion libraries were obtained at room temperature with the restriction site pairs PstI/NotI and AatII/NdeI (Fig. 4.22d). These libraries were used for Colony Filtration Blot experiments.

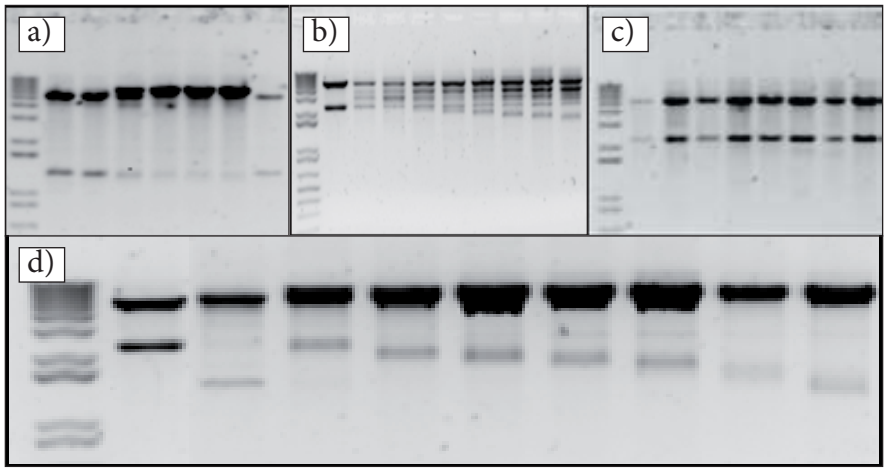


Fig. 4.22 Examples of deletion libraries a) too slow digest b) wrong sense of deletion, c) no deletion, d) good library.

4.1.2.5 Colony filtration blots

Fig. 4.23 shows an example of a plate transformed with a *comA* deletion library, and its corresponding blot in the absence and in the presence of the detergent fos-choline 12 (FC12).

A number of clones were soluble in the detergent FC12, but no clones were soluble in the absence of detergent. Most positive clones were

confirmed after re-screening with FC12 and some of them were found to be slightly soluble in the milder detergent n-Dodecyl-β-maltoside (DDM) (Fig. 4.24).

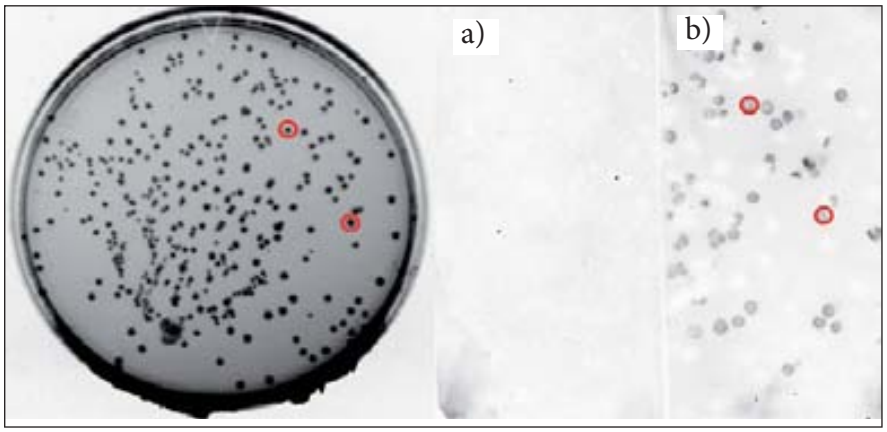


Fig. 4.23 Plate containing cells transformed with a deletion library (a) and its corresponding colony filtration blots in the absence (b) and in the presence of FC12 Red circles are examples of positive colonies.

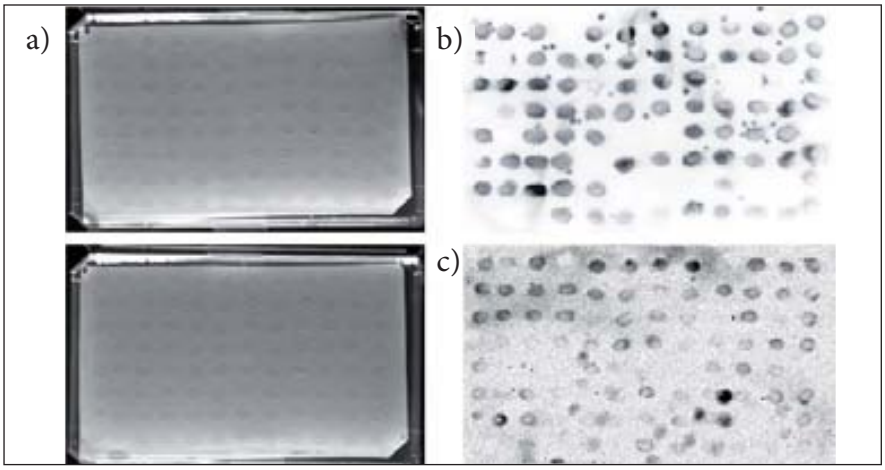


Fig. 4.24 Re-screening of positive clones: a) plates containing colonies to be re-screened by mini-colony filtration blot, b) mini-colony filtration blot in FC12, c) mini-colony filtration blot in DDM.

4.1.2.6 Small scale expression and purification of ComA truncations

Small scale expression experiments in liquid culture confirmed the results obtained at a colony level. All positive clones from the colony filtration blot step turned to express soluble protein in liquid culture as well (Fig. 4.25).

The size of soluble truncations was assessed by SDS-PAGE of Ni-NTA purified samples in a buffer containing the detergent FC12. The range in protein sizes was wide (Fig. 4.26). DNA sequencing of the positive clones confirmed that large clones corresponded to truncations located at ComA C-terminal end.

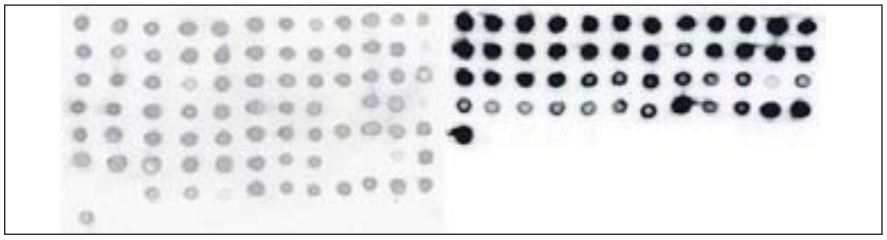


Fig. 4.25 Small-scale expression of ComA truncations. Dot blots of positive clones after small scale expression and purification in FC12.

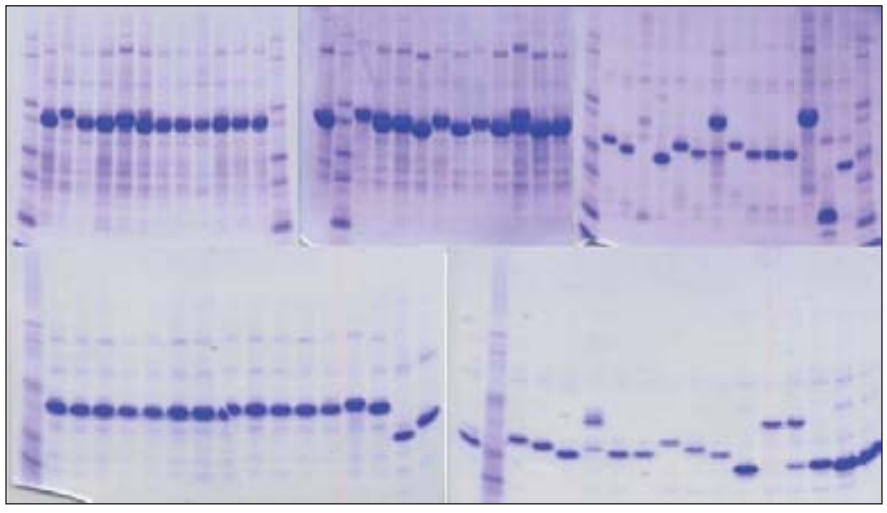


Fig. 4.26 Small-scale expression of ComA truncations. SDS-PAGE of a selection of ComA clones soluble in FC12.

4.1.2.7 Large scale expression and purification of ComA truncations

We scaled-up expression for 8 clones and we managed to solubilize and purify small amounts of some clones in the detergent DDM (Fig. 4.27). In all cases, we observed high expression levels but purification yields were very low.

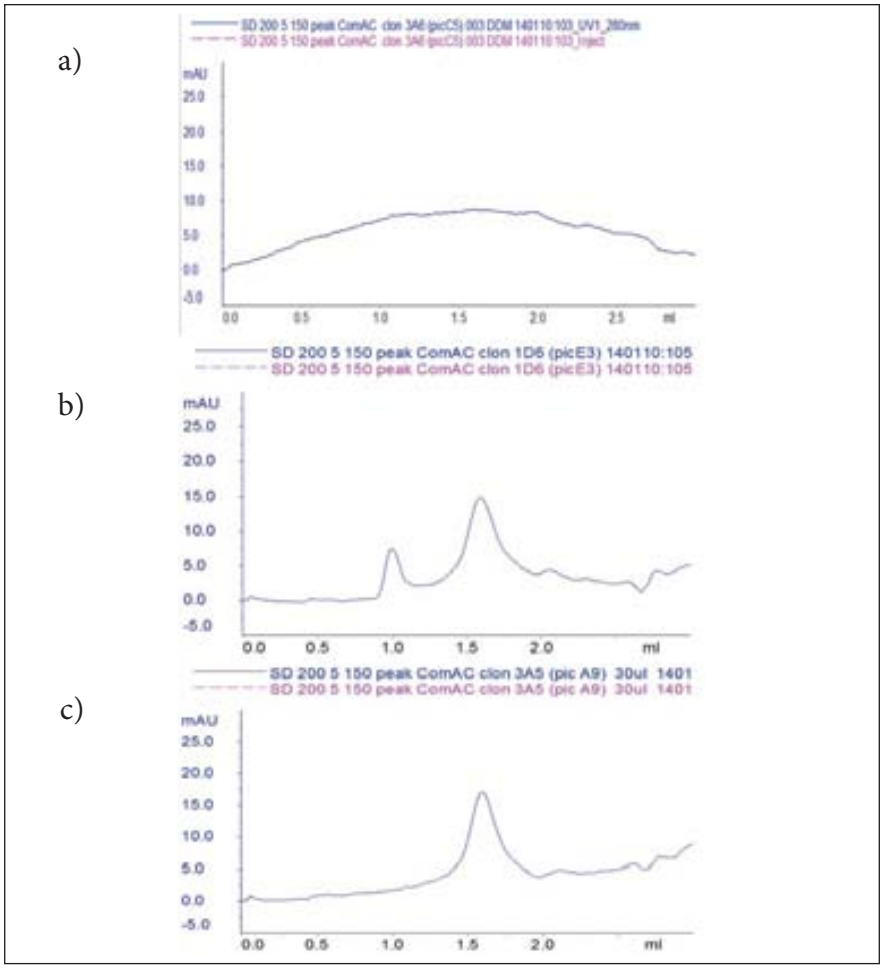


Fig. 4.27 Examples of analytical gel filtration of ComA truncations solubilized in DDM. a) sample containing no protein, b) sample containing aggregated protein and solubilized protein, c) sample containing soluble protein.

When we isolated the membrane fraction from cell debris and inclusion bodies, we realized that most protein was in the form of inclusion bodies. We concluded that the harsh activity of FC12 could have extracted protein from inclusion bodies.

Since we had previously managed to refold and purify the region ComA(C) and most positive clones from the deletion libraries were slight modifications of the C-terminal region that we had previously denatured, refolded and purified, the results were compatible with the hypothesis that FC12 could have had a mild denaturing activity.

Fos-choline detergents are zwitterionic molecules that, due to their structure are very efficient in membrane protein solubilization. However, their utility in crystallization studies is being debated. Even though they often provide good amounts of solubilized protein, to date there are only two examples in the literature of proteins successfully extracted, purified and crystallized only in fos-choline detergents, the *E. coli* mechanoselective channel McsC, obtained in fos-choline 14 (FC14) (Bass *et al.*, 2002; Wang *et al.*, 2008) and the *E. coli* porin OmpF, obtained in FC12 (Kefala *et al.*, 2010). Interestingly, most published membrane protein structures belong to proteins that have been extracted in a few detergents. This fact suggests that despite the highest choice of detergents, the success in membrane protein extraction and purification does not depends on the possibility to perform countless trials in a huge variety of detergents but on a wiser choice of targets and a wiser design of constructs. It seems that, when working with membrane proteins, the most important bottleneck is not at the level of protein solubilization, but at the level of protein expression. Although the expression of correctly folded protein is not *per se* a guarantee of success it is often a *sine qua non* condition.

Membrane protein crystallography still remains an obscure field, mainly guided by trial and error methods. On the one hand, knowing details about the activity of detergents in protein extraction, purification and crystallization stages would provide useful clues for a more intelligent

and economic experimental design. On the other hand, unveiling the mysteries of membrane protein folding and protein insertion into membranes will provide clues for the design of new expression systems that will open the door to a real possibility of structural genomics of membrane proteins. As for now, guarantee of success is very limited. Working with proteins that can be directly extracted from source, working with stable homologues from thermophilic organisms, or working with well-defined soluble domains of membrane proteins seem to be more sensible choices.

4.3.1 DprA: results and discussion

4.3.1.1 Bioinformatics analysis

Similarly as for ComE and ComA, a sequence-based bioinformatics analysis allowed the definition of plausible protein domain boundaries, disordered regions and transmembrane regions.

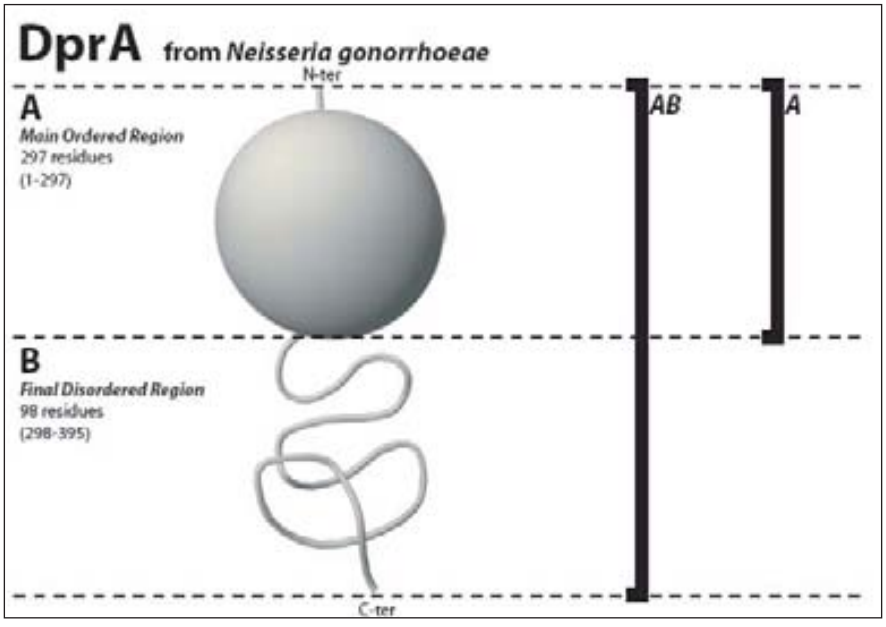


Fig. 4.28 DprA: definition of domain borders for construct design.

DprA was defined as a soluble protein without any membrane-associated region. Globularity and disorder predictions confirmed the presence of a main folded domain and a large disordered region at C-terminal. According to these data we defined two domains: an N-terminal folded domain (later referred to as DprA(A)) and a C-terminal disordered domain (later referred to as DprA(B)) (Fig. 4.28). We initially focused on the structural characterization of DprA(A), the main folded domain.

4.3.1.2 DprA(A) cloning

DprA(A) was successfully cloned in the expression vectors pPROEX HTa, pET28a, pBAT4, pET22b, pETM40, pBADM60 and pBADM40.

4.3.1.3 DprA(A) expression and purification trials

Expression trials of His-tagged or native DprA(A) resulted in the production of large amounts of insoluble protein (Fig. 4.29).

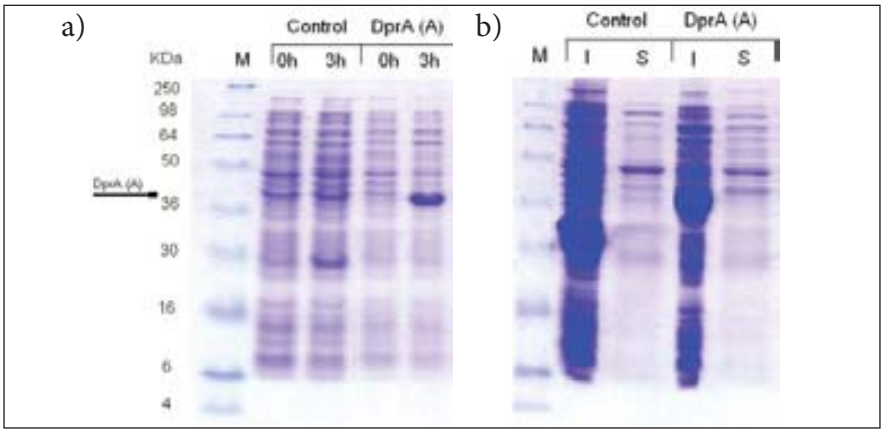


Fig. 4.29 His-DprA(A) expression (a) and solubility (b) trials: M = molecular weight marker, 0h = before induction, 3h = 3h induction, I = insoluble, S = soluble.

The modification of growth temperature, inducing agent (IPTG or arabinose), induction time or inductor concentration didn't result in any substantial improvement of solubility.

Solubility screens with small amounts of urea, detergents or high salt concentration didn't helped in the solubilization of DprA(A), probably

due to the fact that most protein was misfolded and laid down in the form of inclusion bodies.

By contrast, protein expression with larger solubility tags (GST, NusA and MBP) allowed us to obtain small amounts of soluble protein (Fig. 4.30). In any case, most protein turned out to be aggregated (Fig. 4.31), and it was not possible to obtain sufficient amount of protein for crystallization trials.

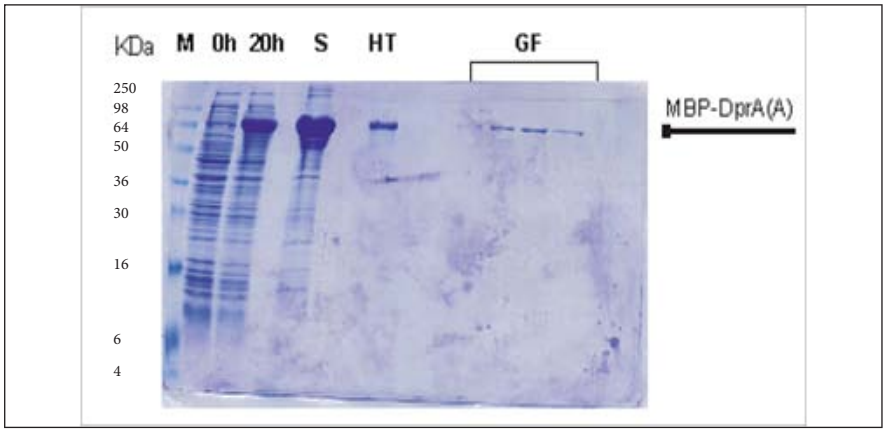


Fig. 4.30 MBP-DprA expression and purification trials: M = molecular weight marker, 0h = before induction, 20h = expression after 20h induction, S = soluble fraction , HT= sample from affinity purification GF = samples from gel filtration purification (void volume).

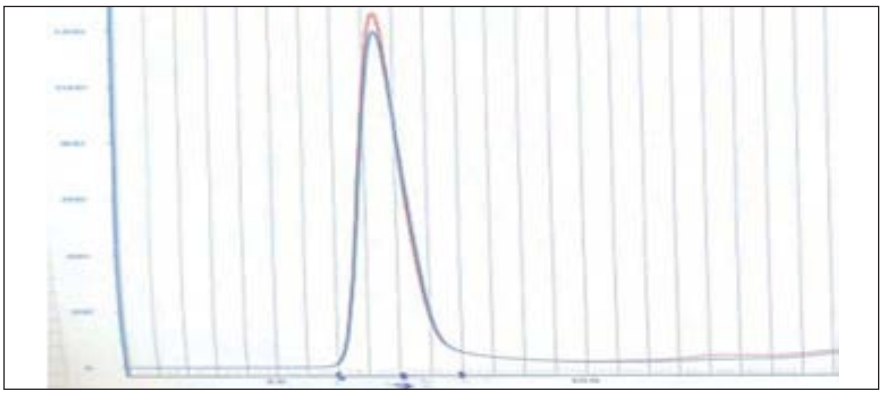


Fig. 4.31 MBP-DprA gel filtration: blue line (Abs 280), red line (Abs 260).

Taken into consideration the difficulties encountered in domain border definition based on *in silico* analysis, we decided to adopt a more experimental approach based on the generation and systematic solubility screen of random deletion libraries.

4.3.1.4 DprA deletion libraries and colony filtration blots

DprA deletion libraries were designed so as to generate clones with N-terminal or C-terminal deletions that would cover the whole *dprA* sequence.

Library representativeness and quality was checked as previously explained for ComA, and we encountered similar problems. In particular, digestions performed on inserts containing the restriction pairs SacI/NdeI were not successful because SacI did not protect the vector terminal end and deletion was bidirectional.

Colony filtration blot screens were performed with deletion libraries obtained at room temperature with the restriction site pairs PstI/NotI and AatII/NdeI.

Positive clone selection was difficult because it depended on the overall quality of the blot and on the uniform spread of colonies. In addition, the difference between low-expressing colonies and non-expressing colonies was not evident in some cases. Fig. 4.32 shows an example of a plate containing *dprA* deletion clones and its respective colony filtration blot.

Interestingly, most positive clones appeared on initial and final plates-*i.e.* plates presumably containing colonies expressing long and short truncations of DprA. Since in the first Colony Filtration blots we preferred oversampling, after re-screening the number of positive clones was reduced by half.

Fig. 4.32 Plate containing cells transformed with a DprA deletion library and its corresponding colony filtration blot. Red circles are examples of low-expressing (L) high-expressing (H) and non-expressing (N) colonies.

4.3.1.5 Small scale expression and purification of DprA truncations

Small scale expression experiments in liquid culture confirmed the results obtained at a colony level. Most positive clones from the colony filtration blot step turned to express soluble protein in liquid culture as well. Interestingly, differences in expression levels were comparable across experiments (Fig. 4.33).

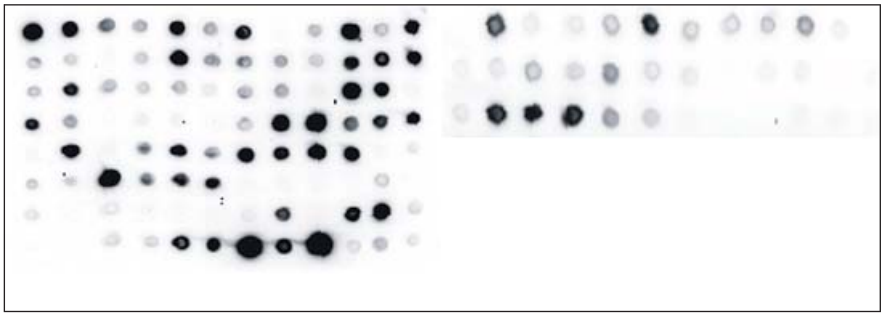


Fig. 4.33 Small-scale expression of DprA truncations. Dot blots of positive clones after small scale expression and purification.

The size of soluble truncations was assessed by SDS-PAGE of Ni-NTA purified samples. Positive clones fell into 3 groups: large, medium and short length (Fig. 4.34). DNA sequencing of the positive clones confirmed that large clones corresponded to the full-length protein, medium-length clones corresponded to a region slightly longer than DprA(A) whereas short clones were all truncations located at DprA C-terminal end.

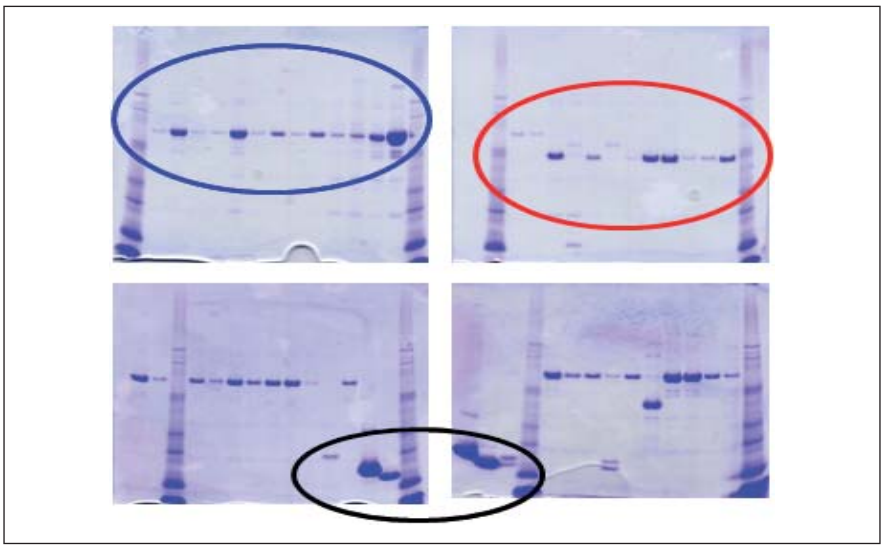


Fig. 4.34 Small-scale expression of DprA truncations. SDS-PAGE gels of a selection of DprA clones. Circles highlight the three groups of clones.

4.3.1.6 Large scale expression and purification of DprA truncations

We were able to scale up protein expression and purification protocols for clones belonging to all three groups.

DprA full-length

By performing affinity chromatography (Fig. 4.35a) and size-exclusion chromatography (Fig. 4.35b) we obtained substantial amounts of soluble DprA(FL). Although it was possible to separate a tetrameric form of DprA(FL) from larger oligomeric species and aggregated sample (Fig. 4.35c), purified protein was a mixture of full-length and degradation products (Fig. 4.35d).

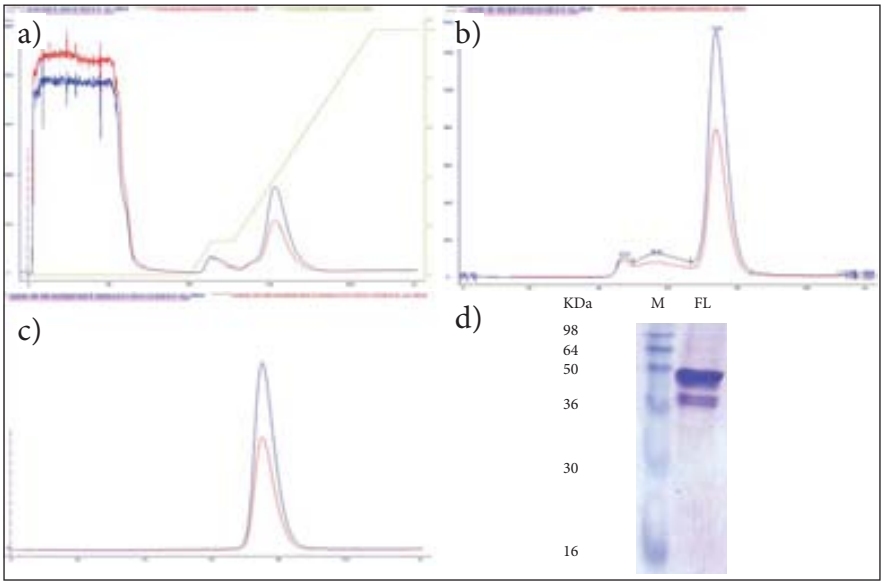


Fig. 4.35 Large-scale expression of DprA (FL). a) IMAC purification, b) SEC purification 1, c) SEC purification 2, d) SDS-PAGE of purified protein after 2nd SEC. M = molecular weight marker, HT = IMAC purified protein, GF = SEC purified protein.

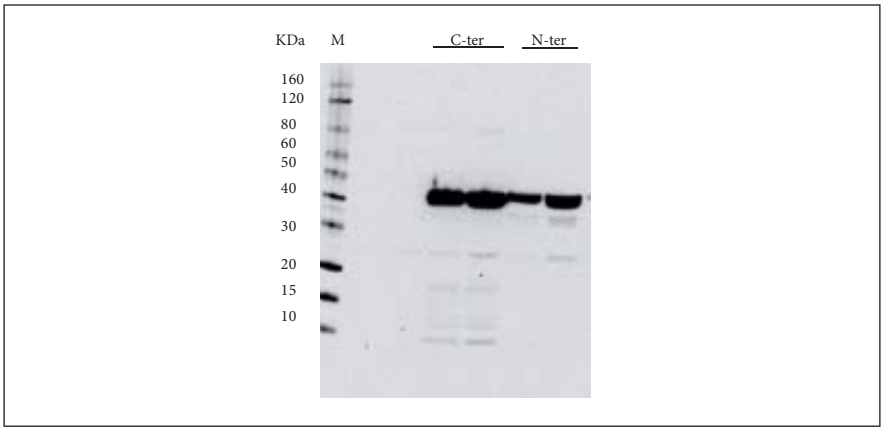


Fig. 4.36 DprA(FL) expression with C-terminal and N-terminal His-tag: Western blot of C-terminal and N-terminal His-tagged DprA(FL). M = molecular weight marker, C-ter = C-terminal histine tag, N-ter = N-terminal histidine tag.

By comparing degradation patterns of N-terminal and C-terminal his-tagged DprA(FL) we concluded that degradation started during expression and that it occurred by the C-terminal end (Fig. 4.36). Protein degradation and aggregation could not be avoided by induction at low temperatures, protease inhibitors, purification at 4ºC or expression with a C-terminal histidine tag. DLS analysis of Dpra(FL) filtered protein samples from gel filtration elution peaks confirmed that samples were monodisperse and that the particle size was bigger than that expected for a monomeric species.

DprA medium length truncations

DprA medium truncations were soluble but aggregated. Interestingly, DprA medium truncations did not show degradation problems and resembled in size DprA(FL) stable N-terminal degradation products (Fig. 4.37) and DprA(A) constructs (Fig. 4.28). DLS analysis of medium-length DprA filtered protein samples confirmed that due to aggregation samples were not monodisperse.

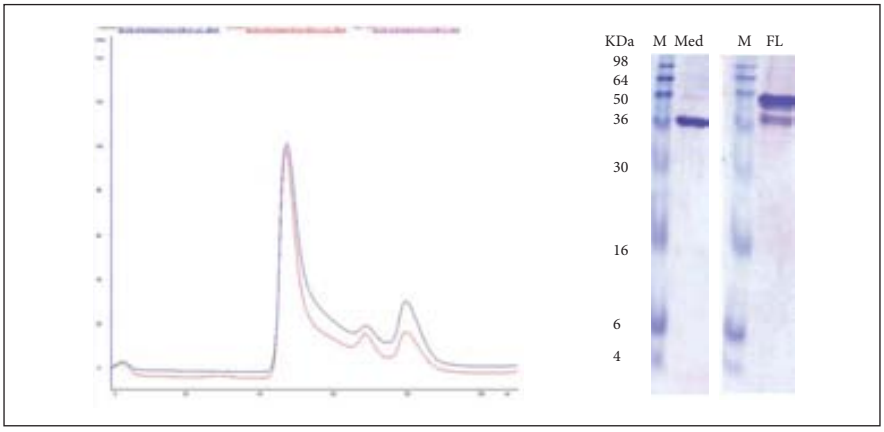


Fig. 4.37 DprA (medium) purification: SEC purification chromatogram and SDS-PAGE. Comparison between purified DprA (medium) and DprA (FL). M = molecular weight marker, Med = SEC purified DprA (medium), FL = SEC purified DprA (FL).

DprA short length truncations

By contrast, DprA C-terminal short length forms with did not present any degradation or aggregation problems (Fig. 4.38).

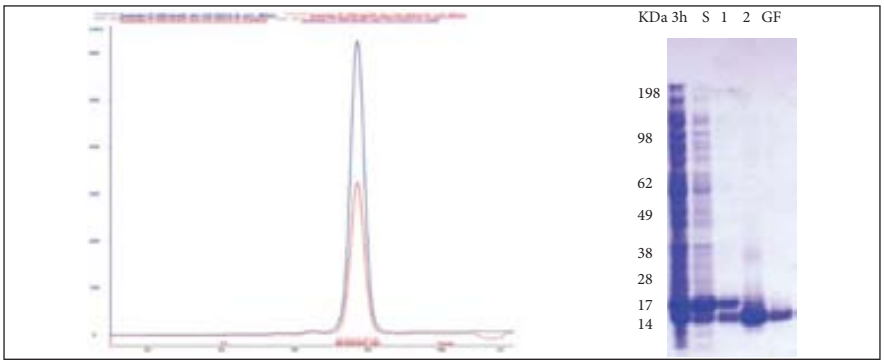


Fig. 4.38 DprA (short) purification: SEC purification chromatogram and SDS-PAGE. M = molecular weight marker, 3h = expression 3h after induction, S = soluble fraction, 1 and 2= samples from affinity purification, GF= SEC purified DprA (short).

4.3.1.7 DprA biochemical characterization and crystallization

DprA (FL) eluted as a molecular species of apparent molecular weight higher than expected for a monomer. Interestingly, oligomerization state was dependant on salt concentration. The minimum observed size was that of a dimer (green line in Fig. 4.39), but at lower salt concentrations we observed that sample eluted at retention times consistent with the formation of larger oligomeric species.

Upon mixing pure DprA (FL) with the crosslinking agent DTBP we observed not only that DprA (FL) formed dimers with itself (MW = 84 KDa), but also that N-terminal spontaneous degradation products maintained the ability to bind full-length DprA or other N-terminal degradation products (Fig. 4.40 a, b).

These results supported the idea that DprA dimerized through a region comprised within the first 300 N-terminal aminoacids. Crosslinking assays at high crosslinker concentrations or long reaction times resulted in

the formation of crosslinked molecules of higher molecular weight (Fig. 40 c). Interestingly, rather than being the result of an unspecific aggregation process, the formation of higher molecular weight species appeared to be the result of an ordered oligomerization process that occurred by pairs (Fig. 4.40 c).

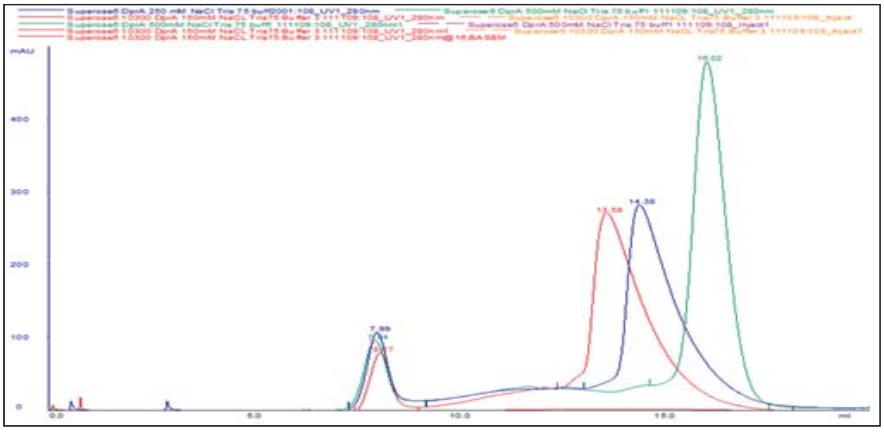


Fig. 4.39 DprA (FL) purification: SEC purification chromatograms of DprA(FL) purified at different NaCl concentrations. Red line (150mM NaCl), blue line (250 mM NaCl), green line (500 mM NaCl). In all cases lines represent Abs at 280 nm.

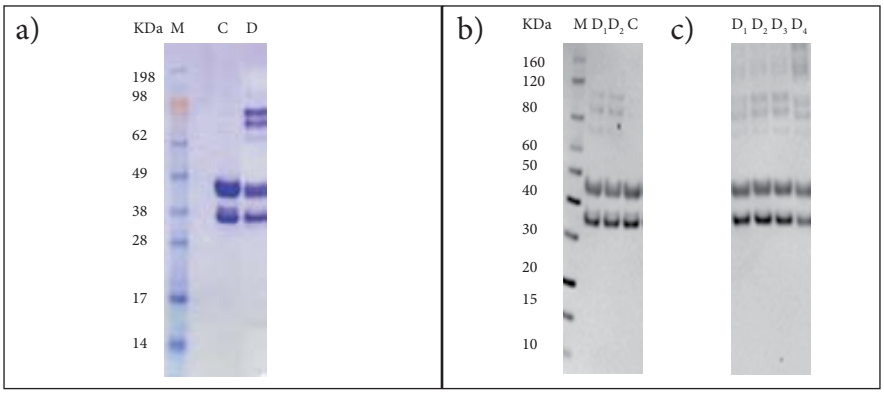


Fig. 4.40 DprA (FL) crosslinking with DTBP: a) SDS-PAGE showing DprA (FL) crosslinking. b) Western blot showing DprA (FL) crosslinking. c) Western blot Western blot showing overnight DprA (FL) at growing [DTBP]. M = molecular weight marker, C = DprA(FL), D = DprA(FL) + DTBP at different concentrations.

Degradation patterns observed upon controlled proteolysis with trypsin were similar to those observed after spontaneous proteolysis. Proteolysis occurred by the linker region between the N-terminal and the C-terminal domain. Even after long digestion times and high trypsin concentration, the N-terminal main folded domain turned to be very resistant to proteolytic digestion (Fig. 4.41). These results suggested that it was worth to perform crystallization trials adding small amounts of trypsin to the crystallization condition.

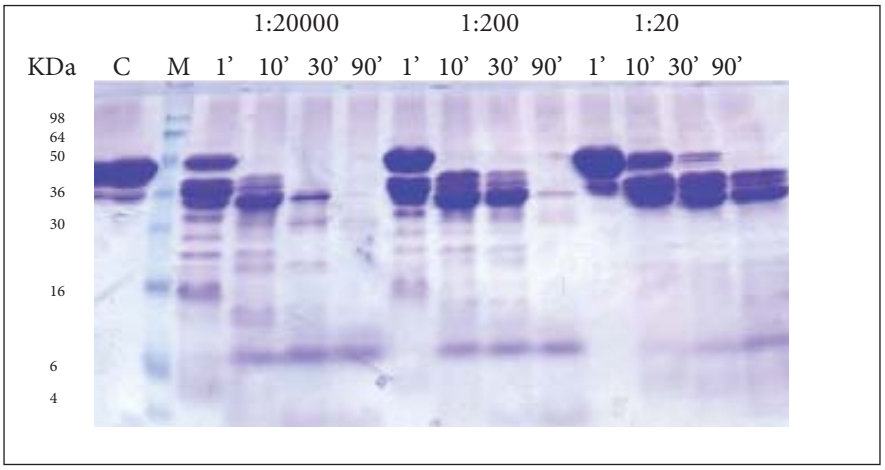


Fig. 4.41 Trypsin proteolysis of DprA (FL) at 37ºC: SDS-PAGE showing proteolytic products upon digestion under different trypsin:protein ratios (1:20000, 1:200 and 1:20) and different times (1', 10' 30' and 90'). C = control, M = molecular weight marker.

DNA binding assays with DprA C-terminal truncations revealed that DprA could bind double stranded DNA through its C-terminal end (Fig. 4.42). Since DprA was expected to bind single stranded DNA, this result was rather surprising.

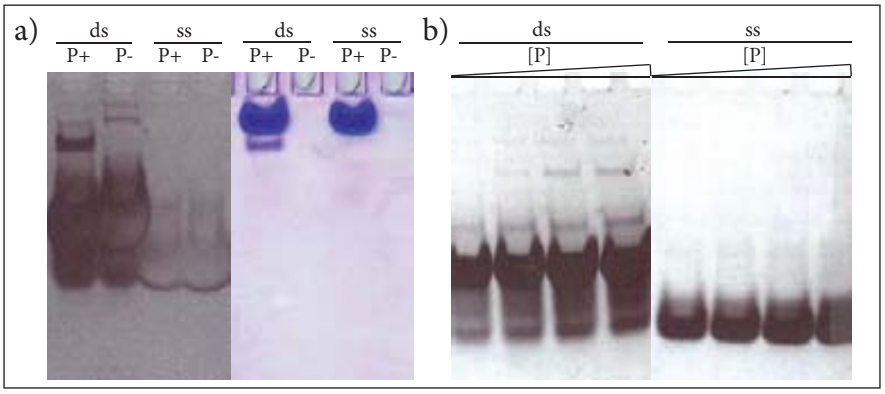


Fig. 4.42 DNA binding assays of DprA (FL): a) The same EMSA gel stained with Syber Gold (left side) and coomassie brilliant blue (right side). b) EMSA gels stained with Syber Gold. ds = extDUS dsDNA, ss = extDUS ssDNA, P+ = with protein, P- = without protein, [P] = protein concentration.

Crystallization trials of full-length DprA were unsuccessful. Although we were able to obtain spherulites and microcrystals under some conditions (Fig. 4.43 a, b), we were unable to optimize them so as to obtain diffraction-quality crystals. We attributed the lack of success to the heterogeneity of the sample due to degradation and aggregation. In an effort to obtain less heterogeneity we decided to speed up the proteolytic process during crystallization by adding small amounts of trypsin to crystallization drops. We observed microcrystals in some conditions containing ammonium sulfate (Fig. 4.43).

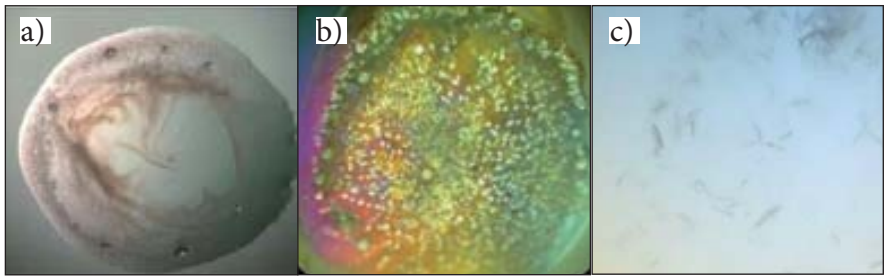


Fig. 4.43 Crystallization trials of DprA (FL): promising crystallization conditions for DprA(FL) in the absence (a, b) and in the presence of trypsin (c).

Despite all the efforts put in the crystallization of a DprA C-terminal truncation we could not obtain diffraction-quality crystals. At low protein concentrations (below 10 mg/ml) most drops appeared to be clear. At higher protein concentrations (up to 60 mg/ml) we observed spherulites and phase separation phenomena (Fig. 4.44).

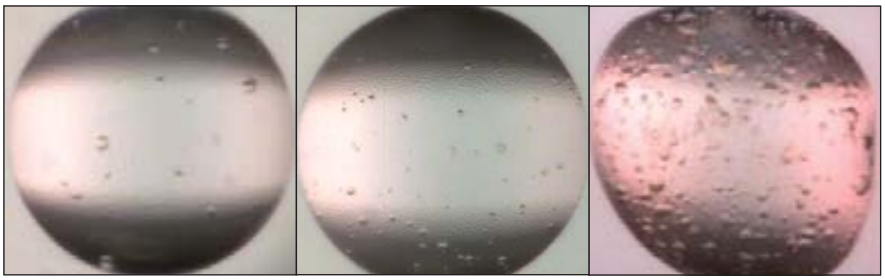


Fig. 4.44 Crystallization trials of DprA (short): promising crystallization conditions for DprA (short).

4.3.1.8 DprA structural analysis

During the course of this project the structure of a close homolog of *N. gonorrhoeae* DprA was determined by a structural genomics group. *Rhodopseudomonas palustris* DprA atomic coordinates were released in May 2010 (PDB: 3MAJ).

Since *R. palustris* DprA shared 45.5% sequence similarity and 31.9% sequence identity with *N. gonorrhoeae* DprA, we assumed that both proteins would have similar folds. Based on this premise, we could infer structural information about our target by combining our biochemical data with structural data provided by the structural analysis of this close homolog (Fig. 4.45).

R. palustris DprA monomer consists of 2 domains: an N-terminal domain (residues 1-302) and a C-terminal domain separated by a flexible linker missing in the structure (303-320) (Fig. 4.45 b).

The N-terminal domain contains an N-terminal alpha-helical region (residues 1-85) followed by a series of 8 parallel β-strands linked to four pairs of α-helices in the topological order β–α–β–α (residues 86-302) (Fig. 4.45 a).

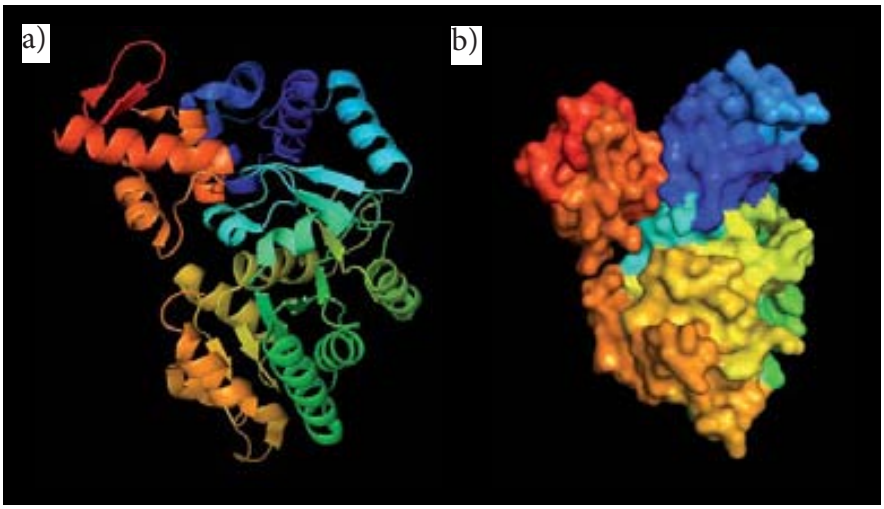


Fig. 4.45 DprA structure: a) model that highlights secondary structure elements, b) model that highlights the separation between the N-terminal and the C-terminal domains.

Whereas residues 1-85 form a four helix bundle that align with all-alpha proteins with unrelated functions, the region comprised between residues 86-302 forms a positively charged surface and resembles a Rossman fold-like motif, a motif commonly found in proteins that bind nucleotides and cofactors. Not surprisingly *R. palustris* DprA structure aligns with other structures with unrelated functions that contain Rossman fold motifs such as the Molybdenum Cofactor (MoCo) carrier protein and a number of lysine decarboxylases (Fig. 4. 46 b). The presence of a Rossman fold-like motif in *N. gonorrhoeae* DprA would explain why the protein was initially classified as a putative lysine decarboxylase by sequence and structure-based functional assignment tools.

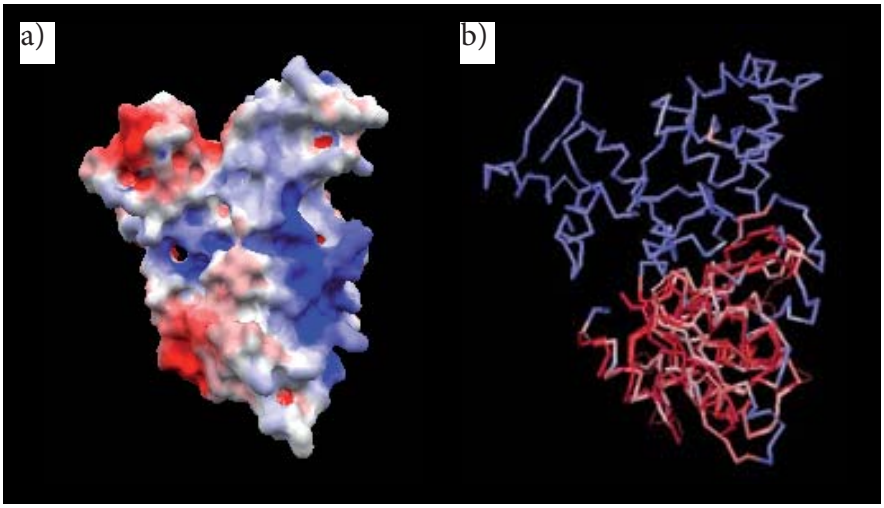


Fig. 4.46 DprA structure showing electrostatic surface charge and alignment of the N-terminal domain: a) model that highlights positively (blue) and negatively (red) charged surfaces, b) model that highlights the alignment between the region of DprA containing the Rossman fold motif and a Molybdenum Cofactor (MoCo) carrier.

Interestingly, *R. palustris* DprA crystallized in a condition containing ammonium sulfate. The fact that sulfate ions were found in most conditions that gave microcrystals of *N. gonorrhoeae* DprA suggest that sulfate ions might be important for crystallization. Indeed sulfate ions occupy conserved positions throughout a positively charged surface of the N-terminal domain (Fig. 4. 48b). We believe that sulfate ions could mimic the biological binding of nucleotide phosphates.

The C-terminal domain contains a winged-helix DNA binding motif (located between residues 320 and 382) (Fig. 4.47). Winged-helix DNA binding motifs are commonly found in proteins that bind double stranded DNA. Many different proteins with diverse biological functions contain a winged helix DNA-binding domain. Among proteins containing winged-helix DNA binding motifs there are transcriptional repressors, transcription factors. endonucleases, histones, transposases, and helicases.

Our DNA binding results support the hypothesis that DprA C-terminal end interacts with dsDNA. We think that the ability to bind both single and double stranded DNA would confer DprA an active role during recombination, and that in addition to its role in single stranded DNA protection, DprA could have additional roles during strand exchange.

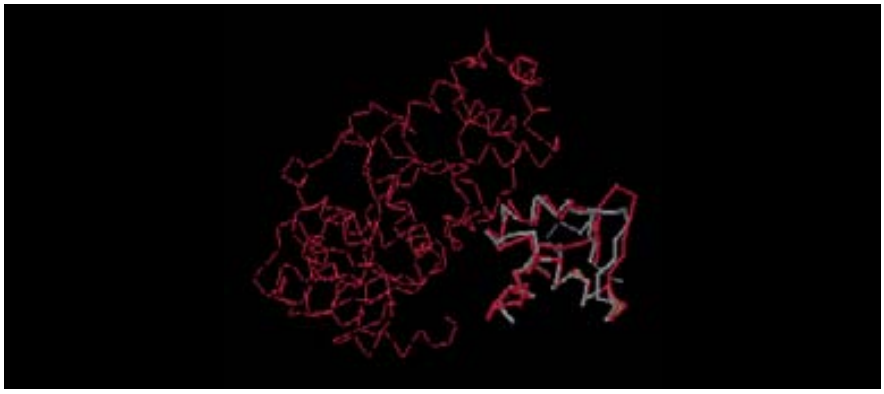


Fig. 4.47 DprA structure showing the alignment of DprA C-terminal domain (red) with a winged-helix DNA binding domain (grey).

The missing region between the N-terminal and the C-terminal domains in the structure is equivalent to the region that in our experiments got degraded. This region is rich in proline aminoacids, a common characteristic of intrinsically disordered regions involved in protein-protein interactions. We suggest that this region plays a role in filament formation by allowing the interaction between DprA and RecA, and by providing the flexibility between DprA N-terminal and C-terminal domains required for the interaction with ssDNA and dsDNA.

The analysis of DprA crystallographic structure also provided hints regarding oligomerization. The biological assembly generated by PISA was a an elongated dimer that resembled a riding saddle (Fig. 4.48 a). In the dimer, interactions between monomers would be guided by salt bridges and hydrogen bonds between basic and acid residues located between residues 283 and 294. Our gel filtration and cross-linking results support this hypothesis.

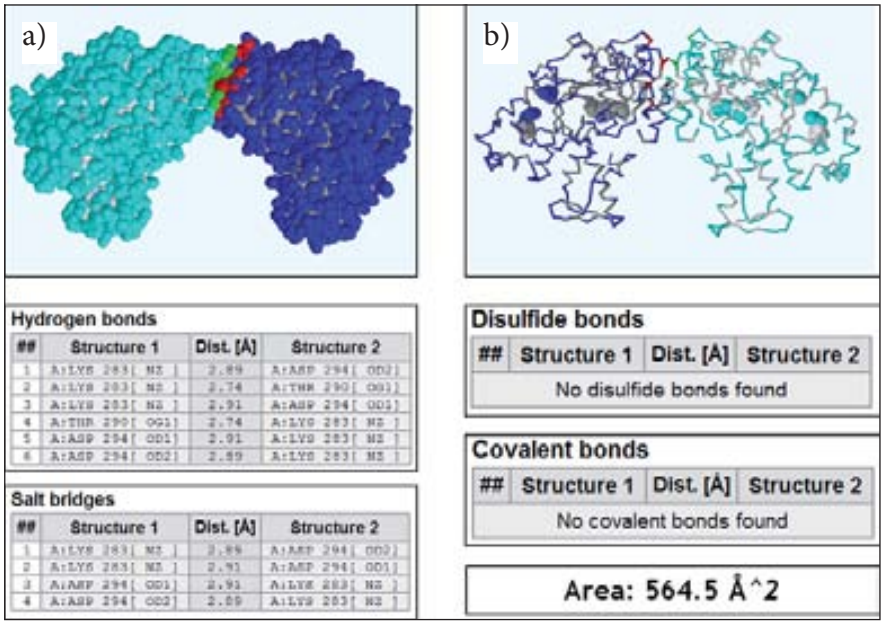


Fig. 4.48 Model of a DprA dimer showing the characteristics of the interaction area (a) and the binding of sulfate ions (b).





4.2 Results and Discussion: Toxin Antitoxin Systems

4.2.1 RelBE2

4.2.1.1 RelBE2 sample preparation

S. pneumoniae RelBE2 expression and purification was performed in the laboratory of Prof. Manuel Espinosa. Annealing and purification of RelBE2-DNA complexes were performed at our laboratory.

RelBE2 complex with DNA eluted as a single species in gel filtration. Sample eluted with an apparent weight of around 80 KDa. Since both proteins had similar sizes, sample purity was assessed by the presence of a single band resulting from the overlap of two bands on SDS-PAGE gels around 10 KDa. A slight band, probably corresponding to a small fraction of a degradation product of RelB, appeared at around 6 KDa (Fig. 4.49).

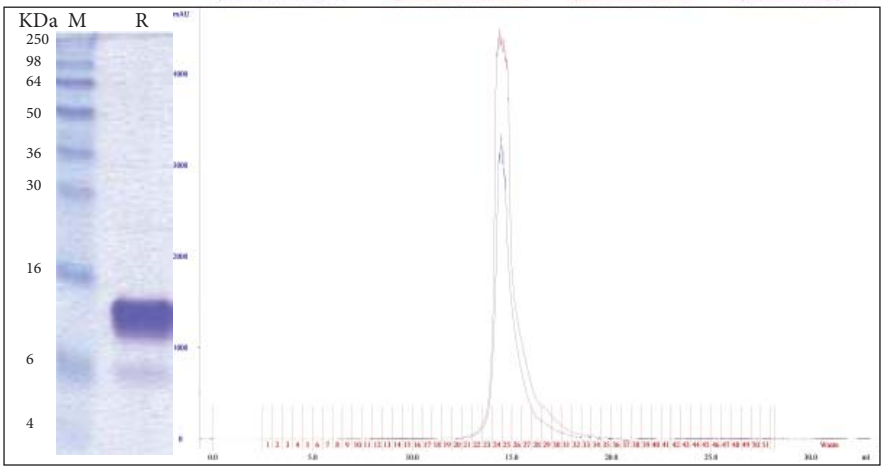


Fig. 4.49 Purification of RelBE2 complexes with 30bp dsDNA: a) SDS-PAGE. b) SEC chromatography. M = molecular weight marker, R = RelBE2 + DNA.

4.2.1.2 Crystallization

Initial hits with native protein were obtained from general crystallization screens PAC1, PAC9 and PAC10 of the Automated Crystallography Platform (Parc Científic de Barcelona). Remarkably, most hits were obtained in conditions with variable amounts of MPD. Initial hits were irregular small plates with round edges (Fig. 4.50 a). The optimization of the most promising conditions gave bigger crystals that could be analyzed at a synchrotron beam-line (Fig. 4.50 b). The best native crystals were obtained in a buffer containing 15% MPD, 0.1% Bis-Tris pH6.5, 0.04M $MgCl_2$, diffracting up to 3.6Å resolution.

SeMet crystals were obtained from the crystallization screens PAC1, PAC9 and PAC10 in similar conditions as used for the native crystals. Derivative crystals were slightly bigger and had more defined edges. The best derivative crystals were obtained in a buffer containing 40% MPD, 0.05M Hepes pH 7, 0.04M $Mg(CH_3COO)_2$ (Fig. 4.50 c).

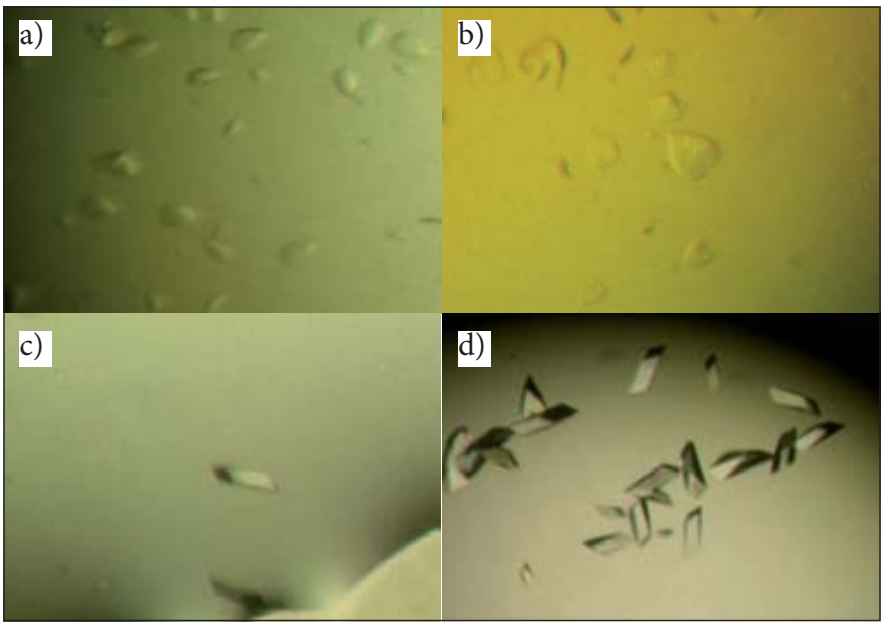


Fig. 4.50 Crystallization of RelBE2. a) Initial hits of native protein, b) optimization of initial hits, c) crystallization of Se-Met labelled protein, d) optimization of native crystals.

Crystals were later optimized by crystallizing RelBE2 without DNA and by increasing protein concentration. In this way, we managed to obtain native crystals that diffracted up to 2.5Å in a crystallization condition containing 39% MPD, 0.05M imidazole pH 7, 0.04M $Mg(CH_3COO)_2$ (Fig. 4.50 d).

4.2.1.3 Structural determination

Since *S. pneunomiae* RelBE2 was homologous to TA systems that had been previously characterized, we initially aimed at solving the structure by molecular replacement.

Fig. 4.51 shows the diffraction pattern of a crystal that diffracted up to 3.6 Å (Fig. 4. 51). Data collection statistics for this crystal are detailed in table 4.1. All analyzed crystals had the same space group and similar cell dimensions.

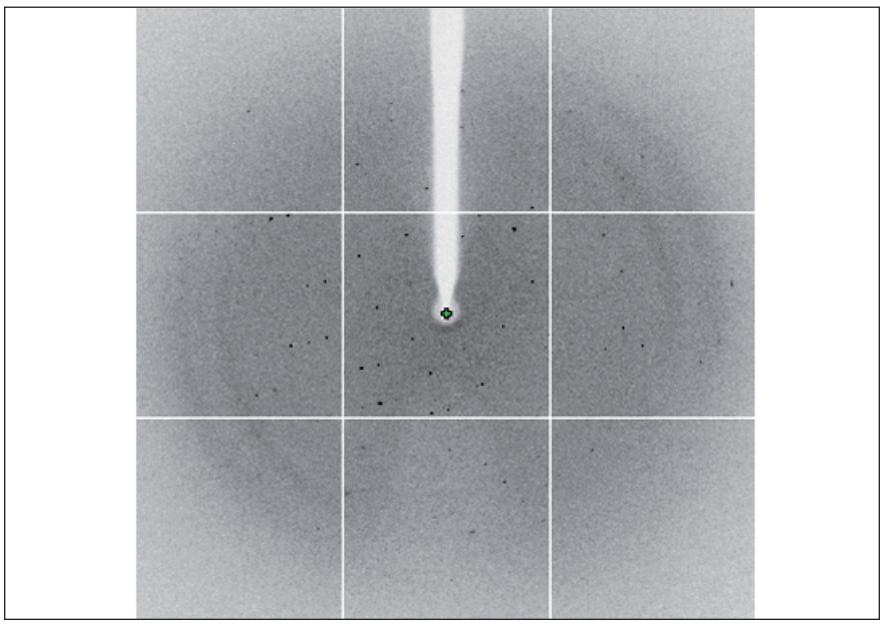


Fig. 4.51 Diffraction pattern of a RelBE2 native crystal that diffracted up to 3.6 Å.

Table 4.1 Data collection statistics Native crystal 1

Beam line	ESRF ID 23-1
Wavelength (Å)	0.87260
Space group	C2
Cell dimensions	
a, b, c (Å)	72.98 85.83 68.70
α, β, γ (°)	90.00 99.64 90.00
Resolution (Å)	45.74 - 3.6
R_{merge} - overall (last shell), (%)	0.178 (0.755)
I/σ(I) - overall (last shell)	8.3 (3.0)
Completeness - overall (last shell), (%)	99.8 (99.9)
Redundancy, (%)	4.0

Native data sets were used for molecular replacement with models from crystallographic structures of close homologues of the toxin RelE. The closest homologue was *M. tuberculosis* RelE, which shared 26.9% identity and 40.7% similarity with *S. pneunomiae* RelE. The second closest homologue was *P. horikoshii* RelE, which shared 19.8% identity and 39,6% similarity with *S. pneumoniae* RelE. Unfortunately, homologous antitoxin partners shared less than 30% identity and could not be used for molecular replacement.

Figuring out the stoichiometry of the complex within the crystal was difficult. According to ultracentrifugation results, *S. pneumoniae* RelBE2 formed hexamers with particle size of 60KDa. The complex between a RelBE2 hexamer and a 30bp DNA would have a particle size of 80KDa, but Mathews' coefficient indicated that the asymmetric unit would contain at most one 60KDa particle. That size would correspond either to the size of one hexamer alone or to the size of one tetramer RelB:RelE plus one DNA molecule.

In order to assess the presence of DNA in the crystals we performed fluorescence scans of RelBE2 crystals stained with Sybersafe but fluorescence results were not conclusive.

Despite all the efforts, molecular replacement using models for the toxin subunit and models for the toxin-antitoxin complex failed presumably due to the low similarity of the models but also to the low quality of diffraction data sets. Resolution was low and radiation damage was significant. Subsequently, we decided to optimize crystallization conditions and to obtain crystals of SeMet-labelled protein so as to determine phases by MAD.

SeMet crystals diffracted better than native crystals, up to 3 Å. Unfortunately, SeMet crystals also suffered from significant radiation damage. Although we aimed at collecting data sets at three different wavelengths (absorbtion peak, inflection point and a high-energy remote wavelength of selenium), we could only use the data sets collected at the peak and inflection point, and useful high-resolution data was up to 3.5 Å. Data collection statistics for this crystal are summarized in table 4.2.

As indicated in the methods section, 10 selenium sites were identified and refined using data up to 3.5Å. Initial electron density maps were used to automatically trace an initial model that contained 309 residues built in 17 fragments. Further manual modelling allowed us to trace must of the protein backbone although some loops could not be traced.

We soon realized that even though the protein complex RelBE2 had been copurified in complex with DNA, crystals contained only protein. Purification conditions and the length of the oligomers would need to be adjusted so as to get crystals of RelBE2 complexed with DNA. Since we believed that DNA could have interfered in the crystallization process and could have somehow hindered crystal quality, we decided to crystallize the protein without DNA and we managed to obtain crystals that diffracted up to 2.5 Å.

Table 4.2 Data collection and refinement statistics Se-Met crystal

Beam line	ESRF ID14-4	
Space group	C2	
Cell dimensions		
a b c, (Å)	72.52 85.99 68.50	
α β γ, (°)	90.00 99.76 90.00	
Wavelength, (Å)	Peak = 0.979	IP = 0.980
Resolution, (Å)	67.51 - 3.50	67.51 - 3.50
R_{merge} - overall (last shell), (%)	0.097 (0.206)	0.204 (1.153)
I/σ(I) - overall (last shell)	13.8 (7.6)	4.7 (1.6)
Anomalous Completeness - overall (last shell), (%)	99.6 (99.6)	99.5 (99.6)
Anomalous Multiplicity, (%)	3.6	3.4

Molecular replacement with data from optimized native crystals that diffracted up to 2.5 A allowed us to improve the quality of the density maps and the model. Most protein backbone could be traced after several rounds of refinement. Rwork and Rfree converged to 24.5% and 29.2% respectively. Tables 4.3 and 4.4 summarize data collection and refinement statistics respectively.

Table 4.3 Data collection statistics Native crystal 2

Beam line	ESRF ID 23-1
Space group	C2
Cell dimensions	
a b c, (Å)	72.35 85.80 68.60
α β γ, (°)	90.00 99.61 90.00
Wavelength, (Å)	0.979
Resolution, (Å)	30 - 2.5
Completeness (overall, last shell), (%)	99.8 (100)
R_{merge} (overall, last shell), (%)	0.057 (0.308)
I/σ (overall, last shell)	14.7 (4.2)
Multiplicity, (%)	5.2

Table 4.4 Refinement statistics Native crystal 2

Resolution, (Å)	30 - 2.5
Number of reflections	13,503
R_{work} / R_{free}	0.2446 / 0.2915
Number of atoms	3,093
Protein	3.056
Water	18
Average B-value, (Å^2)	34.6
R.m.s. deviations	
Bond lengths, (Å)	0.0058
Bond angles, º	0.8762
Ramachandran plot	
Outliers, n (%)	1 (0.28)
Allowed, n (%)	11 (3.04)
Favored, n (%)	350 (96.69)

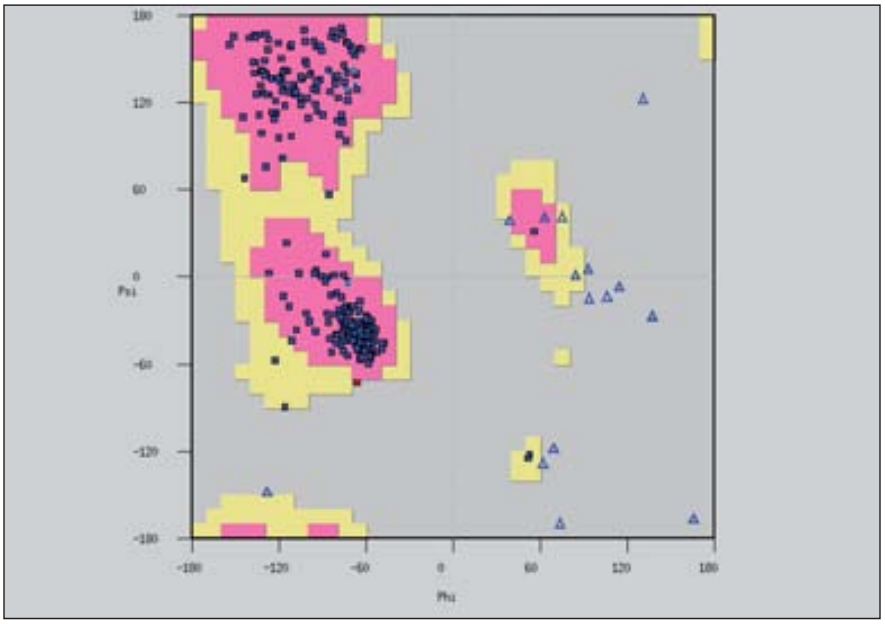


Fig. 4.52 Ramachandran plot of RelBE2. Pink areas are favored regions, yellow areas are allowed regions. Triangles are glycines and squares are any other residue.

4.2.1.4 Crystal contents

The asymmetric unit comprises 5 protein molecules: 2 RelE molecules and 3 RelB molecules (Fig. 4. 53). that form a dimer (RelB:RelE) and a trimer (2RelB:RelE). By crystallographic symmetry, RelB and RelE form tetramers 2(RelB:RelE) and hexamers 2(2RelB:RelE) (Fig.4.54).

According to PISA analysis, the buried area would account for nearly 60% of the total surface area of the hexamer and 46% of the tetramer. There are extensive interactions mediated by hydrogen bonds and salt bridges between antitoxin molecules and between toxin and antitoxin molecules. These results suggest that both conformations are stable in solution and could be the biological units.

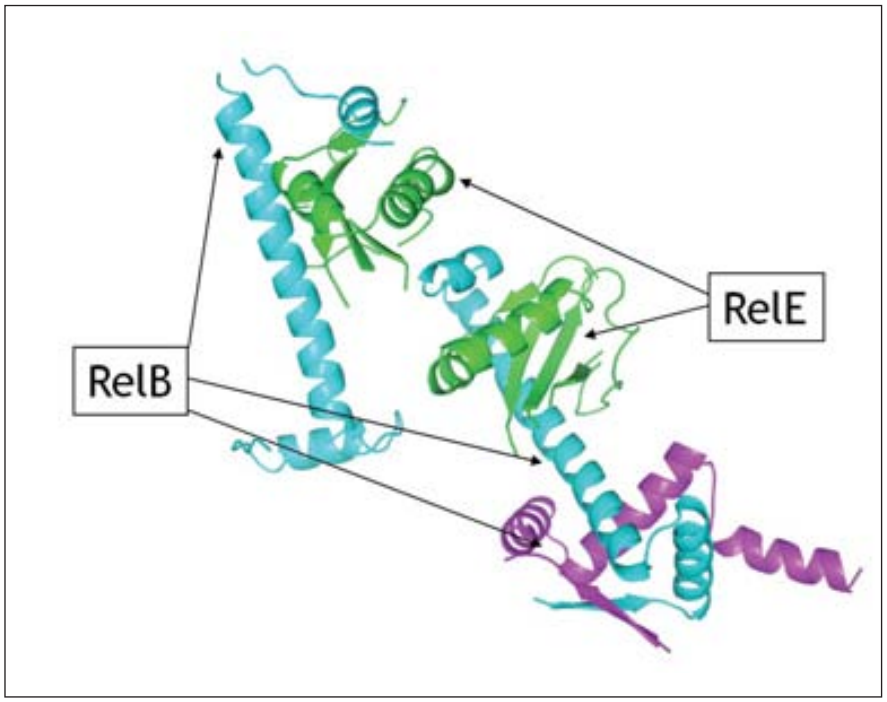


Fig. 4.53 Contents of the asymmetric unit: The asymmetric unit contains 2 RelE molecules and 3 RelB molecules forming a dimer (RelB:RelE) and a trimer (2RelB:RelE).

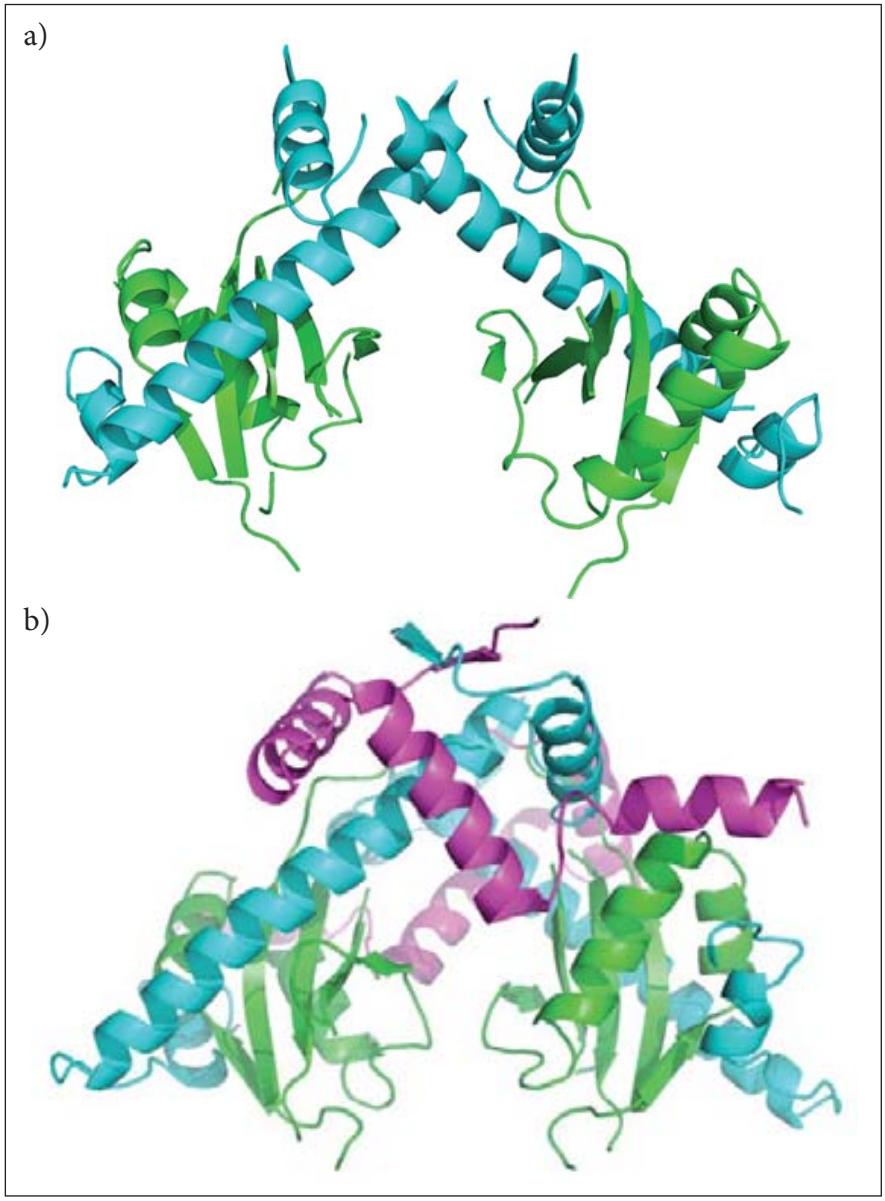


Fig. 4.54 Side view of two alternative oligomerization states of *S. pneumoniae* RelBE2 a) tetramer, b) hexamer.

4.2.1.5 RelE structure and toxicity mechanisms

RelE molecules have an open sandwich fold which is common to homologous toxins belonging to the RelE family (Fig. 4. 55). The fold is characterized by 2 α-helices (α1, α2) and a β-sheet formed by 4 antiparallel β-strands (β1, β2, β3, β4) with a topology β1, α1, α2, β2, β3, β4.

After DALI homology search the highest Z-score was found with *M. tuberculosis* RelE toxin (Z-score=20.1, rmsd=1.7). High Z-score values were also found with RelE toxins belonging to *T. termophilus* (Z-score=11.6, rmsd=1.5), *P. horikoshii* (Z-score=10.7, rmsd=2.5), and *E. coli* (Z-score=10.4, rmsd=1.9), but also with *M. tuberculosis* RelK (Z-score=10.7, rmsd=2.0), *E. coli* YoeB (Z-score=10.6, rmsd=2.0) and *C. crescentus* ParE (Z-score=9.6, rmsd=1.8) toxins.

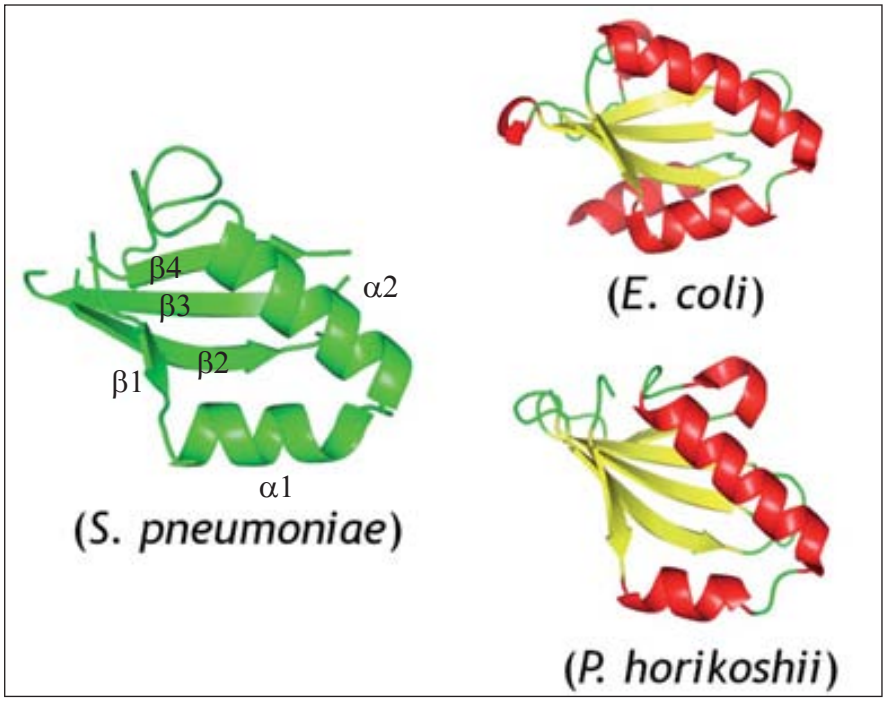


Fig. 4.55 RelE toxin fold. Comparison between *S. pneumoniae*, *E. coli* and *P. horikoshii* RelE folds.

RelE toxicity mechanisms have been characterized in *P. horikoshii* (Takagi *et al.*, 2005) and in *E. coli* (Neubauer *et al.*, 2009). Site-directed mutagenesis experiments performed on *P. horikoshii* and the structural characterization of *E.coli* RelE in complex with *T. thermophilus* ribosome have been discussed in the introduction (chapter 1). The overall high structural similarity and conservation of residues important for RelE activity suggest common mechanisms of action for *S. pneumoniae* RelE.

S. pneumoniae RelE contains two positively charged helices (helices α1 and α2) which similarly to *E. coli* RelE, could bind the ribosomal A-site through extensive contacts with 16S rRNA phosphate backbone (Fig.4.56).

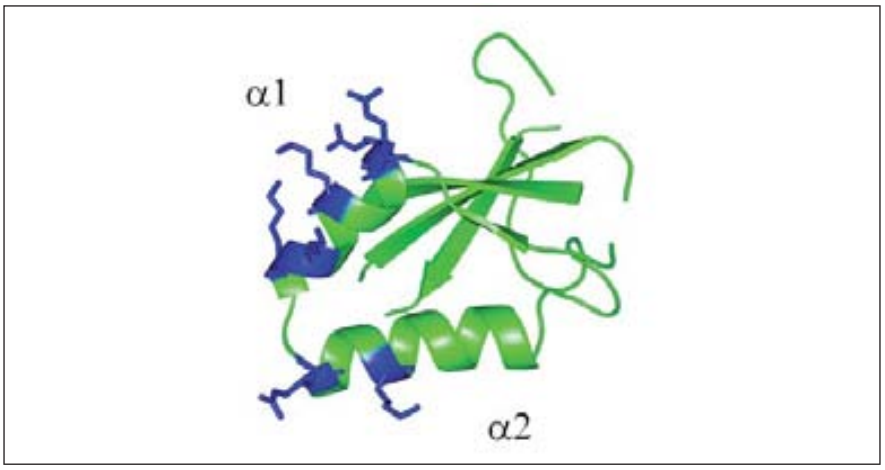


Fig. 4.56 RelE interaction surface with the ribosomal A-site.

Regarding RelE active site, site-directed mutagenesis experiments in *P. horikoshii* RelE suggested that residues Arg40, Leu48, Arg58 and Arg65, and Arg85 played important roles in toxin's activity (Takagi *et al.*, 2005). Equivalent residues in *E. coli* (Arg45, Leu44, Lys 52, Lys 54, Arg 61, Arg 81) do play an important role by binding incoming mRNA that will be cleaved (Neubauer *et al.*, 2009). In *S. pneumoniae* RelE residues

Leu47, Arg 56, Arg 58, and Arg 63 are located in the central β–sheet facing towards a cavity that would accommodate mRNA while Arg 41 is located in a flexible loop (Fig. 4.57).

Residues Arg 83 and Tyr 87 in *S.pneumoniae* RelE could have similar roles to those proposed for equivalent residues in *P. horikoshii* RelE (Arg85 and Tyr89) and *E. coli* RelE (Arg81 and Tyr87). Tyr87 could drive the orientation of incoming mRNA through stacking interactions but also act as a general base accepting a proton from 2'-OH whereas Arg83 would act as a general acid in the reaction. Unfortunately, these residues are not visible in our structure probably due to the inherent flexiblity of RelE C-terminal end.

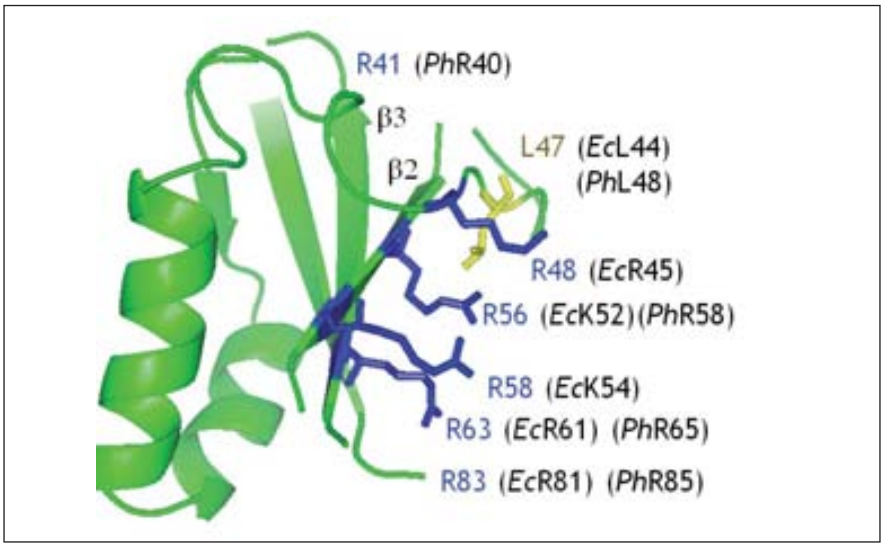


Fig. 4.57 RelE residues involved in mRNA cleavage. Residues occupying equivalent position in the homologues *E. coli* RelE and *P. horikoshii* RelE are indicated in parentheses.

In addition, *S. pneumoniae* RelE active site has additional residues that could explain an intrinsic RNase activity not found in *P. horikoshii* and *E. coli*, but found in farthest homologues such as YoeB and MazF (Christensen-Dalsgaard and Gerdes, 2008). The conservation of resi-

dues His43, Tyr31, Tyr57, Glu38 and Asp 39 in RelE active site would suggest a ribosome-independent RNase activity similar to that observed for ribonuclease T1. Tyr31 and Tyr57 could someone complement the stacking provided by Tyr87 and drive mRNA towards Glu 38/Asp39 and His43. Either Glu 38 or Asp39 would act as a general base and accept a proton from the 2'-OH whereas His43 would act a general acid.

We suggest that RelE would preferentially bind to the ribosomal A-site and would cleave translating mRNA, but that it would also show an intrinsic ribosome-independent nuclease activity. mRNA would be drawn towards RelE active site by electrostatic forces due to the local high concentration of arginine residues and would then be oriented through stacking forces provided by several tyrosine residues located in the active site. mRNA cleavage would occur through a 2'-OH-induced hydrolysis mechanism that would involve a general base (a deprotonated Tyr87 or Glu 38/Asp39) and a general acid (Arg83 or alternatively His43).

However, important molecular details about the catalytic reactions still remain elusive. The characterization of mutants, together with the structural characterization of RelE in complex with RNA substrates would provide invaluable data about the catalytic mechanism.

4.2.1.6 RelB structure and toxicity inhibitory mechanisms

RelB regulates toxin activity by two mechanisms. On the one hand it inactivates RelE by direct binding to it, preventing thus RelE activity. On the other hand it prevents the transcription of both RelE and RelB components by binding to the DNA operator region, acting thus as a repressor of RelE expression. Our study has provided hints for both mechanisms.

RelB has two different conformations: straight and bent (Fig. 4.58). The straight conformation contains a C-terminal β-strand (β1) followed by 3 α-helices (α1, α2, α3), being the second α-helix the longest. The bent

conformation is very much like the straight one with the noticeable difference that in the bent conformation, α2 is split in two shorter helices (α2' and α2"). This change occurs through the extension of the helix turn comprised between the residues Asp44, and Val47 (region circled in Fig. 4.58). The region expected to contain α3 could not be traced in the bent conformation.

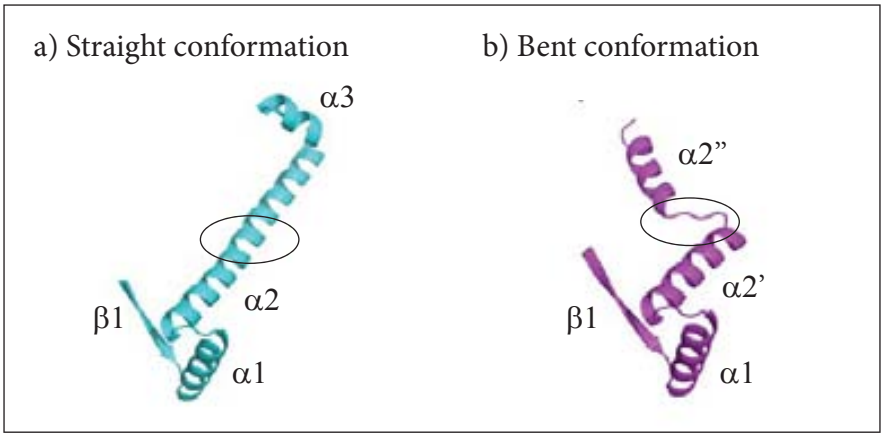


Fig. 4.58 RelB shows two conformations: straight (a) and bent (b). The circled area highlights the stretched turn.

Although sequence homology was low, structural similarity was high and RelB transcription repression mechanism could be inferred from homologous structures. After DALI search for RelB structural similarities the highest Z-score was found with a putative transcription factor from *Vibrio cholerae* (Z-score=6.1, rmsd=6.0). High Z-score values were also found with FitA antitoxin from *N. gonorrhoeae* (Z-score=5.8, rmsd=7.3), PutA repressor from *E. coli* (Z-score=5.1, rmsd=1.2), Arc repressor from bacteriophage p22 (Zscore= 5.1, rmsd=1.3), and CopG repressor from *Streptococcus aglactiae* (Z-score=4.9, rmsd=1.4). All proteins belong to the MetJ/Arc superfamily of transcriptional repressors. Although protein structures superimposed, rmsd values were high due to conformational changes in the N-terminal end. C-terminal end containing the DNA-binding domain was conserved and allowed the superimposition between homologues.

RelB blocks RelE from entering the ribosome

RelB C-terminal end interacts with RelE mainly through electrostatic interactions and sterically prevents RelE from entering the ribosome. Interestingly, residues that are important for RelE toxicity are located at the interface between RelE and RelB, indicating that RelB not only prevents RelE toxicity by preventing its entrance to the ribosome, but also by blocking its active site (Fig. 4. 59).

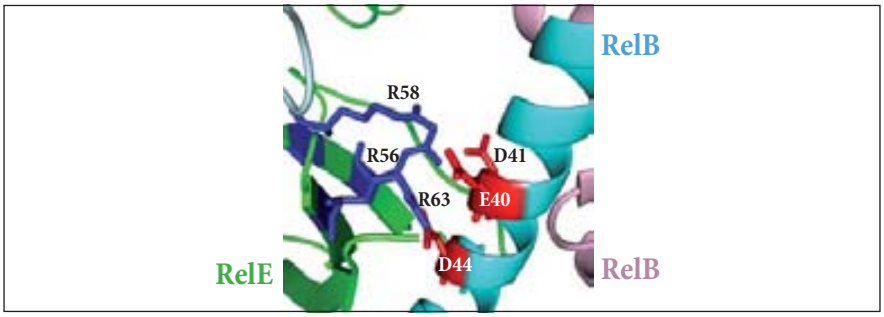


Fig. 4.59 Interaction between RelE and RelB.

Interaction surfaces between RelE and the "straight form" of RelB are equivalent in the hexamer and in the tetramer. In the hexamer, the additional "bent form" of RelB confers further interaction surfaces with RelE. This interaction restricts the RelB flexibility at the C-terminal end, and the flexibility of RelE loops. Lon protease has been described to mediate the degradation of misfolded proteins (Gur and Sauer, 2008). The abundance of coiled coils and the large flexibility of RelB molecules at the C-terminal end would explain the susceptibility of RelB to be degraded by Lon protease. It would be interesting to study if there are any differences in the susceptibility to RelB degradation in tetramers or hexamers.

RelB acts as a transcription repressor

RelB interacts with another copy of RelB through its N-terminal end. Two RelB N-terminal ends form a ribbon-helix-helix (RHH) motif. *S. pneunomiae* RelBE2 hexamers contain two consecutive RHH motifs (Fig. 4.61 a and Fig. 4.61 b). As expected, the electrostatic potential of

the region is positive on the ribbon-helix-helix region in both the tetramer (Fig. 4.61 c) and the hexamer (Fig. 4.61 d).

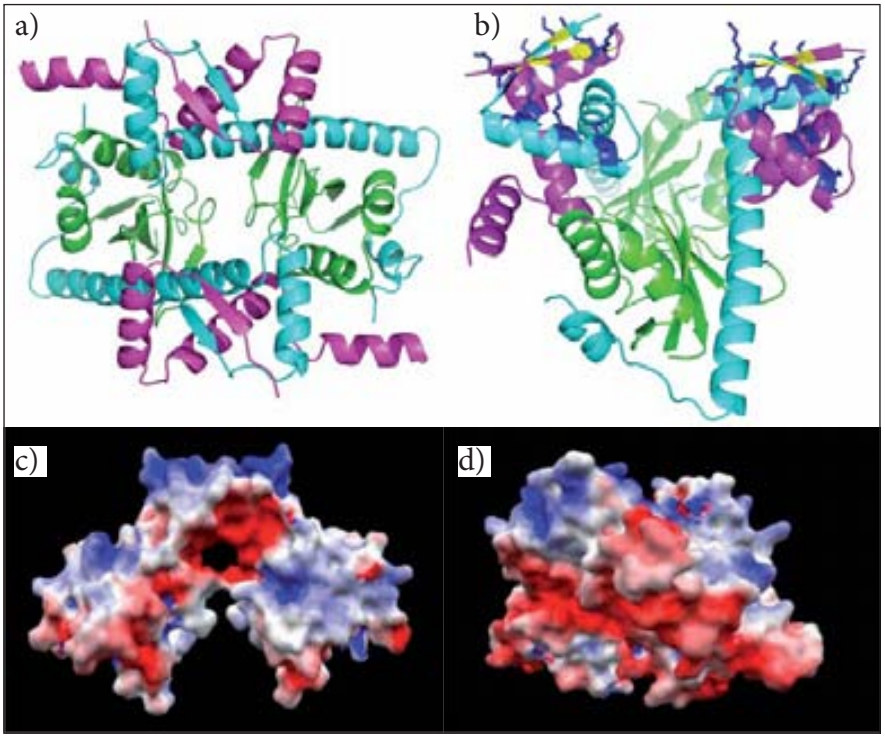


Fig. 4.60 Putative S. pneumoniae RelBE2 interaction surfaces with DNA. a) top view of the hexamer showing 2 RHH motifs, b) side view of the hexamer highlighting positive and apolar residues in the ribbon-helix-helix motifs, c) electrostatic surface of the tetramer, (side view) d) electrostatic surface of the hexamer (side view).

Taking as a model *N. gonorrhoeae* FitAB structure in complex with its operator region (Mattison, 2006) and biochemical data provided by footprinting assays and EMSA assays (Moreno-Cordoba *et al.*, 2012), we propose a DNA binding model for the hexameric form of *S. pneumoniae* RelBE2 (Fig. 4.60). According to our model, hexameric RelBE2 would cover 2 DNA turns. Positive residues located in RelB helices would interact with the DNA phosphate backbone, and residues located in the N-terminal β-sheet would establish contacts with specific bases located

in two consecutive major grooves in *relBE2* operator region. Each DNA binding motif would cover six base pairs.

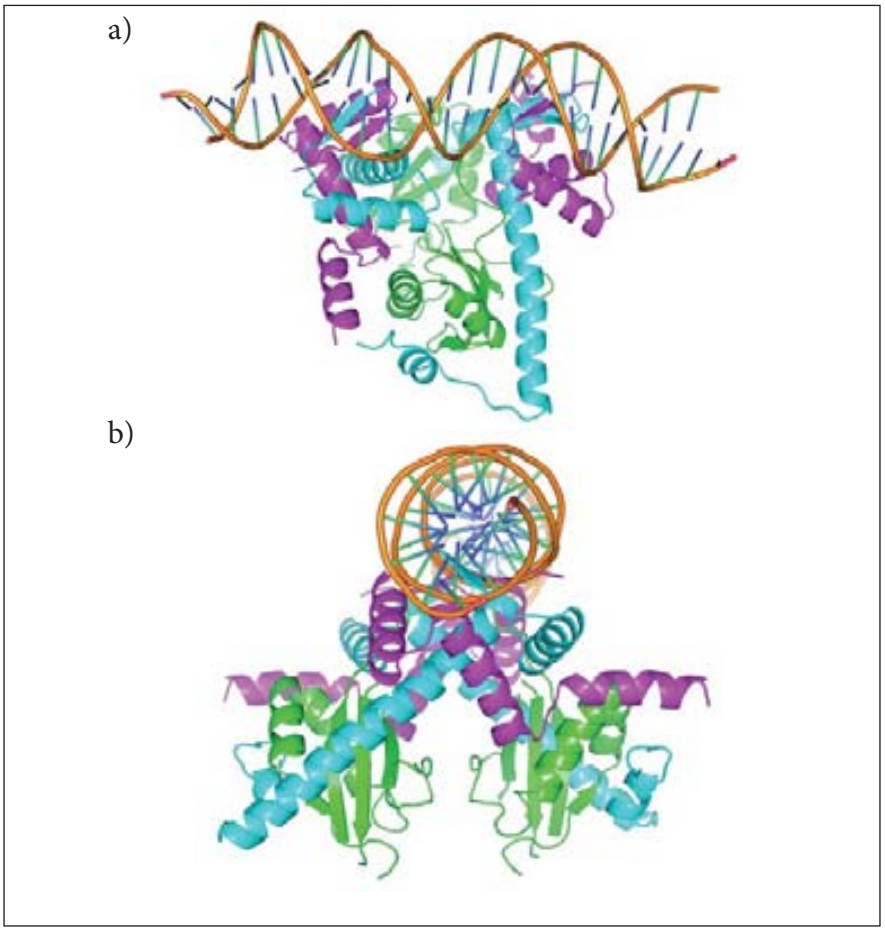


Fig. 4.61 Model of *S. pneumoniae* RelBE2 interaction with DNA. a) side view, b) front view.

4.2.1.7 RelB/RelE ratio as a regulatory mechanism of RelE toxicity

Given the fact that both tetramers and hexamers are stable in solution and mantain the ability to bind DNA and interact with RelE, the regulation of RelE toxicity could depend on RelB/RelE ratio. RelBE2 regulation by RelB/E ratio is summarized in Fig. 4.62.

Depending on RelB availability, RelE would be in three states: free, complexed with one RelB molecule (forming dimers/tetramers) or complexed with two RelB molecules (forming trimers/hexamers).

In the first scenario, free RelE would exert its toxic effects by interacting with the translating ribosome and cleaving mRNAs. In the second scenario, the interaction between one RelE molecule and one RelB molecule would lead to the formation of heterodimers and heterotetramers that would prevent RelE toxicity but would have minor effects on transcription repression. In the third scenario, the interaction between one RelE molecule and two RelB molecules would lead to the formation of heterotrimers and heterohexamers that would not only prevent RelE toxicity but would also repress transcription in an efficient manner by binding to the *relBE2* operator region.

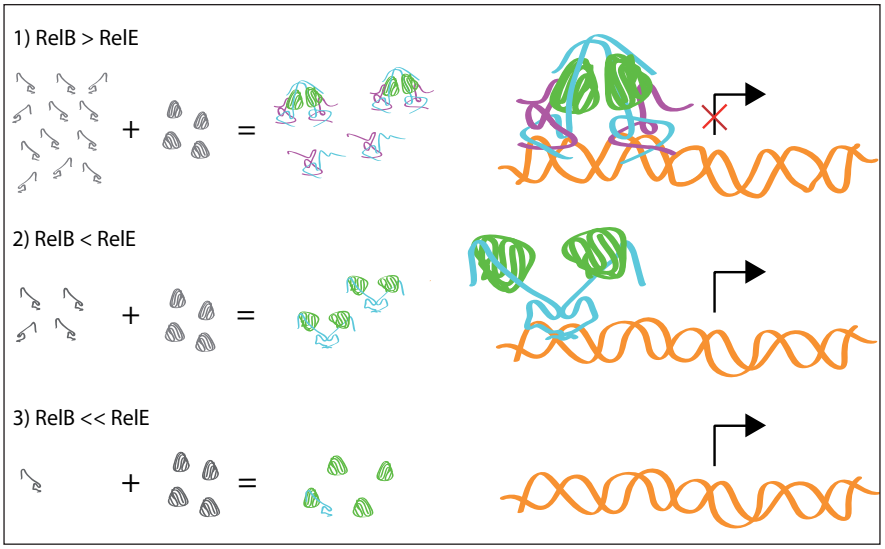


Fig. 4.62 Transcription regulation by the RelBE2 ratio. 1) If RelB > RelE, RelBE forms hexamers that bind to two consecutive operator sites in the relBE2 promoter. Under these circumstances, transcription is repressed, 2) If RelB < RelE, RelBE2 forms tetramers that bind to a single operator site. Under these circumstances, RelE does not exert its toxic action because it is bound to RelE, but transcription is not repressed, 3) If RelB <<RelE, RelE is free and transcription is not repressed.

RelB/RelE ratio is controlled by the degradation-rate of RelB which is determined by Lon protease. In a situation of active growth, RelB level is higher than RelE level. When [RelB] > [RelE], RelBE2 forms hexamers that bind cooperatively to the *relBE2* operator region and repress translation. During nutritional stress, RelB level gets lower as a consequence of the activation of Lon protease. When [RelB] < [RelE], RelBE2 forms tetramers that do not bind cooperatively to the *relBE2* operator region and transcription is activated. The increased transcription-rate of *relBE2* leads to a partial replenishment of the antitoxin pool but also to an increase of toxin concentration. If Lon activity persists, RelB level gets even lower. When [RelB] << [RelE], RelE gets released and it can target the ribosome.







Chapter 5
Conclusions & Future Work





5.1 Conclusions and Future Work: Natural Transformation

This project has explored techniques for the expression, purification and structural characterization of ComE, ComA and DprA, three proteins involved in natural transformation in *N. gonorrhoeae*. Although we have not been able to obtain crystals for any of the targets, this PhD thesis has provided protocols for the expression and purification of truncations of all three targets and some insights into ComE and DprA structures.

5.1.1 ComE: Conclusions and Future Work

We were able to express and purify substantial amounts of a ComE truncation but the protein was reluctant to crystallize. By combining information from 1-H NMR experiments and data coming from structural models we deduced plausible causes of the failure in crystallization.

1-H NMR experiments indicated that the truncation we had been working with was unfolded, and that DNA could promote protein folding. ComE could be translocated as an unfolded substrate across the internal membrane and folded upon DNA binding. Even if unfolded or partially folded, the high affinity for unspecific DNA would be guaranteed by the abundance of lysine residues. The high abundance of lysine residues would explain the need for buffers with a high ionic strength to avoid co-purification with DNA from the host strain, but also the fact that the protein was so reluctant to crystallize.

From a structural model based on a homologous domain from *T. thermophilus* ComEA (PDB:2DUY) we deduced that the protein contained a helix-hairpin-helix domain, but also that the truncation could lack

of a few residues essential for the formation of a β-strand that could be critical for ComE folding. The inclusion of those four residues could have provided a more stable conformation.

It would be interesting to further characterize ComE folding dependence upon DNA binding, and to perform expression and crystallization trials of constructs containing the residues involved in the formation of the missing β-strand so as to determine the effect of those residues in protein folding.

5.1.2 ComA: Conclusions and Future Work

Although we have not been able to crystallize *N. gonorrhoeae* we think that our results have provided important information that will be useful for future works with ComA homologues and with other membrane proteins.

This thesis presents two alternative strategies for the expression and purification of C-terminal truncations of *N. gonorrhoeae* ComA. In addition, it provides information regarding the use of detergents for the solubilization of membrane proteins.

The first strategy consisted of the location of transmembrane helices and the delimitation of soluble domain borders by bioinformatic tools. We delimited two regions predicted to be soluble at both extremes of the protein and were able to obtain substantial amounts of one of them: a C-terminal truncation of ComA, referred throughout the thesis to as ComA(C). We obtained ComA(C) by combining IMAC purification under denaturing conditions, step-wise refolding and SEC chromatography in native buffer. Unfortunately, the protein was not stable in solution and did not crystallize. We believe that the reason why ComA

C-terminal domain forms inclusion bodies and why this domain was so unstable in solution was that it contained a membrane-associated region that promoted aggregation.

In the second strategy, deletion libraries of ComA were screened for the expression of soluble protein using the colony filtration blot technique with buffers containing the harsh detergent FC12 and buffers that did not contain it. All positive results were obtained for ComA C-terminal truncations upon solubilization with FC12. Solubilization in buffers containing the milder detergent DDM was not satisfactory. The fractionation of cell extracts by differential centrifugation confirmed that in all cases, ComA C-terminal truncations were in inclusion bodies fractions, and not in membrane fractions.

Since we had previously managed to refold and purify the region ComA(C) and most positive clones from the deletion libraries were slight modifications of the C-terminal region that we had previously denatured, refolded and purified, the results were compatible with the hypothesis that FC12 could have had a mild denaturing activity. Determining the boundaries of this region, or even better, finding an homologue lacking of this membrane-associated region, could be of great importance to get more stable truncations of ComA.

Besides providing information about ComA, this work has also raised some questions about the idoneity of fos-choline detergents in membrane protein studies, and in general, about the experimental design of projects that deal with the purification of membrane proteins. Indeed, this thesis shows an example of high rates of FC12 solubilization of a missfolded target, and provides support to the idea that fos-cholines high efficiency as solubilizing agents results in part from their ability to extract missfolded proteins. Although this mild denaturing ability is potentially exploitable in refolding techniques, one should be careful in order to avoid misleading positive results in solubilization experiments of membrane proteins.

5.1.3 DprA: Conclusions and Future Work

Even though we have not been able to crystallize and characterize by X-ray crystallography *N. gonorrhoeae* DprA, this study has put some light on DprA function and has raised some questions regarding the manageability of proteins containing large disordered regions.

The soluble constructs obtained by Erase-a-base and Colony Filtration blot correspond with two domains; a large N-terminal domain with a Rossman-fold motif and a C-terminal domain with DNA binding properties; separated by a flexible link. This link could play a role in the formation of RecA/ssDNA filaments.

DprA had been previously described as a ssDNA binding protein. It is particularly striking the dsDNA binding ability of DprA C-terminal domain. It is tantalizing to think that, similarly to RecA, the ability of DprA to interact both with ssDNA and dsDNA has a role in homology recognition and strand exchange for homologous recombination. The characterization of DprA/RecA complexes alone and in combination with ssDNA and dsDNA will provide more clues about the interplay between both proteins and about the exact role of DprA in the integration of incoming DNA in the chromosome.

From a methodological point of view, the results obtained in this project invite reflection on disordered regions. Dealing with proteins containing large disordered regions is a cumbersome issue in X-ray crystallography projects. Introducing disorder often means introducing heterogeneity and reducing the possibility of getting protein crystals. Therefore trimming disordered regions (either at a protein level by using proteases; or at a DNA level by using recombinant techniques) is a common technique in many structural biology projects. But the structural characterization of proteins containing large disordered regions still poses many technical difficulties. Although bioinformatic analysis can provide inter-

esting hints about the location of disordered regions, this information needs to be handled with care. On the one hand, disorder predictions might not be accurate enough. On the other hand trimming regions predicted to be disordered might mean trimming regions that are essential for protein function or for protein stability. Other methods such as the generation of deletion libraries and the controlled digest of protruding loops may reduce uncertainty in the cleavage of disordered region but do not solve the fundamental question regarding the *raison d'etre* of those regions that in most cases are essential for protein function. Actually, since many functionally important protein segments occur outside of globular domains, disorder should not be seen as a problem but as an indicator of protein function.

X-ray crystallography is a powerful method to get structural information about proteins that have a reduced number of stable conformations, but it is less appropriate for the study of proteins that depend on the ability to fold in different ways to be functional. Other techniques such as NMR may be appropriate to complement the crystallographic analysis.





5.2 Conclusions and Future Work: Toxin Antitoxin Systems

The last part of this thesis presents the crystallization and the structural characterization of the RelBE2, a toxin-antitoxin system from *S. pneumoniae* and offers insights into its function and regulation. This project explored techniques for the crystallization and structural characterization of RelBE2, and provided the X-ray structure of a tetrameric and a hexameric form of RelBE2, a model for RelBE2 interaction with DNA and plausible explanations for the regulation of RelE toxicity and RelB antitoxin activity.

5.2.1 RelBE2: Conclusions and Future work

We managed to crystallize native and Se-Met labelled *S. pneumoniae* RelBE2 complex. Initial hits with native protein were small bi-dimensional plates with round edges that diffracted up to 3.6Å. Crystal optimization with Se-Met labelled protein for MAD phasing allowed building of an initial model at 3Å resolution. Crystal optimization provided bigger tridimensional crystals that were isomorphous to the Se-Met derivative crystals and diffracted up to 2.5Å. Using data at this resolution we were able to trace the protein backbone, and to identify waters and rotamers in well ordered areas. Although the quality of the density maps improved considerably, improvement was not uniform probably due to the intrinsic disorder of RelB loops. This disorder could explain the relatively high final R-values. Techniques such as crystal dehydration to reduce solvent content and shrink crystal lattice could be used in the future to promote crystal packing.

Our model supports the hypothesis that RelE toxicity is due to its RNAase activity upon binding to the A-site of a translating ribosome. Similarly to *E. coli* and *P. horikoshii* RelE homologues, *S. pneumoniae* RelE

would attract mRNA through electrostatic forces due to the local high concentration of arginine residues and mRNA cleavage would occur through a 2'-OH-induced hydrolysis mechanism that would involve a general base and a general acid. RelE could also conserve intrinsic RNA-ase activity. Residues that are important for RelE catalytic activity are found to interact with RelB, suggesting that RelB not only prevents RelE from entering the ribosome by steric hyndrance but also by direct contact with the residues required for RelE activity. However, important molecular details about the catalytic reactions still remain elusive. The characterization of mutants, together with the structural characterization of RelE in complex with RNA substrates would provide invaluable data about the catalytic mechanism.

Similarly to other antitoxins, two RelB N-terminal ends form a ribbon-helix-helix DNA binding motif that establishes interactions with a specific operator region located upstream the start codon. It has been described that the antitoxin alone is sufficient to repress transcription of the toxin-antitoxin operon, but that the toxin can dramatically enhance repression.

But repression has been found to depend on the RelB/RelE ratio rather than on absolute values. Interestingly, in our crystals RelBE2 formed both tetramers and hexamer. We propose that both oligomers could play a role in the regulation of the *S. pneumoniae* RelBE2 system and that the formation of hexamers and tetramers could be driven by the relative abundance of both species. In the hexamer, RelE would enhance repression by allowing the formation of a stable complex that positions two DNA binding domains (each domain formed by 2 RelB monomers) at the right distance for the interaction with two DNA binding sites. By contrast, in the tetramer there would be only one DNA binding domain formed by 2 RelB monomers and that condition would not be sufficient for repression.

One of the aims of the project was the structural characterization of the interaction between RelBE2 and dsDNA. Crystallization trials with a 30 bp DNA duplex were unsuccessful, but the DNA binding model proposed in this thesis has provided useful information for further crystallization trials with shorter oligomers that will eventually lead to the structural characterization of the interaction.





Bibliography

Agarwal, D., Gregory, S.T., and O'Connor, M. (2011) Error-Prone and Error-Restrictive Mutations Affecting Ribosomal Protein S12. *J. Mol. Biol.,* **410** (1): 1-9.

Porter, J.R. (1976) Antony van Leeuwenhoek: tercentenary of his discovery of bacteria. *Bacteriol. Rev.,* **40** (2): 260-269.

Ramirez, M.S., and Tolmasky, M.E. (2010) Aminoglycoside modifying enzymes. *Drug Resist Updat.,* **13** (6): 151-171.

Zapun, A., Contreras-Martel, C., and Vernet, T. (2008) Penicillin-binding proteins and β–lactam resistance. *FEMS Microbiol. Rev.,* **32** (2): 361-385.

Aas, F.E., Wolfgang, M., Frye, S., Dunham, S., Lovold, C., and Koomey, M. (2002) Competence for natural transformation in Neisseria gonorrhoeae: components of DNA binding and uptake linked to type IV pilus expression. *Mol. Microbiol.,* **46** (3): 749-760.

Andam, C.P., and Gogarten, P.J. (2011) Biased gene transfer and its implications for the concept of lineage *Biol Direct.,* **6** (47).

Appelbaum, P.C. (1992) Antimicrobial Resistance in Streptococcus pneumoniae: An Overview. *Clin. Infect. Dis.,* **15** (1): 77-83.

Arbing, M.A., Handelman, S.K., Xiao R., Acton, T., Inouye, M., Montelione, G.T. Woychik, N.A. Hunt, J.F. Kuzin A.P. , Verdon, G., Wang, C., Su M., Rothenbacher, F.P., Abashidze, M., Liu, M., and Hurley, J.M. (2010) Crystal structures of Phd-Doc, HigA, and YeeU establish multiple evolutionary links between microbial growth-regulating toxin-antitoxin systems. *Structure,* **18** (8): 996-1010.

Avery, O.T., MacLeod, C.M., and McCarty, M. (1944) Studies on the chemical nature of the substance inducing transformation of pneumococcal types: Induction of transformation desoxyribonucleic acid fraction isolated from pneumococcus type III. *J. Exp. Med.,* **79**: 137-157.

Badía, J., Ibáñez, E., Sabaté, M., Baldomà, L., and Aguilar J. (1998) A rare 920-kilobase chromosomal inversion mediated by IS1 transposition causes constitutive expression of the yiaK-S operon for carbohydrate utilization in Escherichia coli. *J. Biol. Chem.*, **273** (14): 8376-81.

Bailey, S.E., and Hayes, F., (2009) Influence of operator site geometry on transcriptional control by the YefM-YoeB toxin-antitoxin complex. *J. Bacteriol.*, **191** (3):762-72

Balasingham, S.V., Collins, R.F., Assalkhou, R., Homberset, H., Frye, S.A., Derrick J.P., and Tønjum, T. (2007) Interactions between the lipoprotein PilP and the secretin PilQ in Neisseria meningitidis. *J. Bacteriol.*, **189** (15): 5716-5727.

Barh, D., Misra A.N., Kumar, A., and Vasco, A. (2010) A novel strategy of epitope design in Neisseria gonorrhoeae. *Bioinformation.,* **5** (2): 77-85.

Barlow, M. (2009) What antimicrobial resistance has taught us about horizontal gene transfer. *Methods Mol. Biol,.* **532**: 397-411.

Bass, R.B., Strop, P., Barclay, M., and Rees, D.C. (2002) Crystal structure of Escherichia coli MscS, a voltage-modulated and mechanosensitive channel. *Science*. **298** (5598): 1582-1587.

Beaber, J.W., Hochhut, B., and Waldor, M.K. (2004) SOS response promotes horizontal dissemination of antibiotic resistance genes. *Nature,* **427** (6969): 72-74.

Benam, AV., Lang, E., Alfsnes K., Fleckenstein, B., Rowe, A.D. Hovland, E., Ambur, OH, Frye S.A., and Tønjum, T. (2011) Structure-function relationships of the competence lipoprotein ComL and SSB in meningococcal transformation. *Microbiology,* **157** (5): 1329-1342.

Bergfors, T. Protein Crystallization. (2009) 2nd ed. International University Line.

Bley, C., van der Linden M, and Reinert, R.R. (2011) Mef(A) is the predominant macrolide resistance determinant in Streptococcus pneumoniae and Streptococcus pyogenes in Germany. *Int. J. Antimicrob. Agents*, **37** (5): 425-431.

Blow, D. Outline od Cristallography for Biologists. (2002) 1st ed. Oxford University Press, USA.

Bolla , J. M., Alibert-Franco, S., Handzlik, J., Chevalier, J., Mahamoud, A., Boyer, G., Kiec-Kononowicz, K., and Pagès, J.M. (2011) Strategies for bypassing the membrane barrier in multidrug resistant Gram-negative bacteria. *FEBS Lett.,* **585** (11): 1682-90.

Bonifield, H.R., and Hughes, K.T. (2003) Flagellar phase variation in Salmonella enterica is mediated by a posttranscriptional control mechanism. *J Bacteriol.,* **185** (12): 3567-3574.

Borders, D. B., Antibiotics, Survey, Kirk-Othmer Encyclopedia of Chemical Technology (2007) Wiley.

Brock, T.D., Madigan, M.T., and Martinko, J.M. Biology of microorganisms. (2006) 11. ed. Upper Saddle River, NJ: Pearson Prentice Hall.

Brown, B.L., Grigoriu, S., Kim, Y., Arruda J.M., Davenport, A., Wood, T.K., Peti, W., and Page, R. (2009) Three dimensional structure of the MqsR:MqsA complex: a novel TA pair comprised of a toxin homologous to RelE and an antitoxin with unique properties. *PLoS Pathog.,* **5** (12).

Brown, B.L., Wood,T.K., Peti, W., and Page, R. (2011) Structure of the Escherichia coli antitoxin MqsA (YgiT/b3021) bound to its gene promoter reveals extensive domain rearrangements and the specificity of transcriptional regulation. *J. Biol. Chem.*, **286** (3): 2285-2296.

Bush, K. (2010) Bench-to-bedside review: The role of β-lactamases in antibiotic-resistant Gram-negative infections. *Crit. Care,* **14** (3): 224.

Carter, K.C. (1985) Koch's postulates in relation to the work of Jacob Henle and Edwin Klebs. *Med. Hist.,* **29** (4): 353-374.

Chan, W.T., Nieto, C., Harikrishna, J.A., Khoo, S.K., Othman, R.Y., Espinosa M., and Yeo, C.C. (2011) Genetic regulation of the yefM-yoeB toxin-antitoxin locus of Streptococcus pneumoniae. *J. Bacteriol.,* **193** (18): 4612-4625.

Chen, I., and Gotschlich, E.C. (2001) ComE, a competence protein from Neisseria gonorrhoeae with DNA-binding activity. *J. Bacteriol.,* **183** (10): 3160-3168.

Chen, I., and Dubnau D. (2004) DNA uptake during bacterial transformation. *Nat. Rev. Microbiol.*, **2** (3): 241-249.

Chen, I., Christie, P.J. and Dubnau, D. (2005) The ins and outs of DNA transfer in bacteria." *Science,* **310** (5753): 1456-1460.

Chiura, H.X., Velimirov, B., and Kogure, K. (2000) Virus-like particles in microbial population control and horizontal gene transfer in aquatic environments. In: Bell C.R., Brylinsky M., Johnson-Green P., editors. Halifax: Microbial Biosystems: New frontiers.

Christensen, S.K., Mikkelsen M., Pedersen, K., and Gerdes, K. (2001) RelE, a global inhibitor of translation, is activated during nutritional stress. *Proc. Natl Acad Sci. U.S.A.*, **98** (25): 14328-14333.

Christensen-Dalsgaard, M., and Gerdes, K. (2008) Translation affects YoeB and MazF messenger RNA interferase activities by different mechanisms. *Nucleic Acids Res.*, **36**(20): 6472-81.

Ciccarelli, F. D., Doerks, T., von Mering C., Creevey C.J., Snel B., and Bork, P. (2006) Toward Automatic Reconstruction of a Highly Resolved Tree of Life. *Science,* **311** (5765): 1283-1287.

Claros, M.G., and von Heijne, G. (1994) TopPred II: An Improved Software For Membrane Protein Structure Predictions. *CABIOS,* **10**: 685-686.

Claverys, J.P., and Martin, B. (2003) Bacterial "competence" genes: signatures of active transformation, or only remnants?. *Trends Microbiol.,* **11** (4): 161-165.

Claverys, J.P., Martin, B., and Håvarstein, L.S. (2007) Competence-induced fratricide in streptococci. *Mol. Microbiol.,* **64** (6): 1423-1433.

Coglievina, M., Guarnaccia, C., Pintar, A., and Pongor, S. (2010) Different degrees of structural order in distinct regions of the transcriptional repressor HES-1. *Biochim. Biophys. Acta.*, **1804** (12): 2153-2161.

Collins, R.F., Davidsen, L., Derrick, J.P., Ford, R.C., and Tønjum, T. (2001) Analysis of the PilQ secretin from Neisseria meningitidis by transmission electron microscopy reveals a dodecameric quaternary structure. *J. Bacteriol.,***183** (13): 3825-3832.

Collins, R.F., Saleem, M., and Derrick, J.P. (2007) Purification and three-dimensional electron microscopy structure of the Neisseria meningitidis type IV pilus biogenesis protein PilG. *J. Bacteriol.,* **189** (17): 6389-6396.

Cornvik, T., Dahlroth, S.L., Magnusdottir, A., Flodin, S., Engvall, B., Lieu, V., Ekberg, M., and Nordlund P. (2006) An efficient and generic strategy for producing soluble human proteins and domains in E. coli by screening construct libraries. *Proteins,* **65** (2): 266-273.

Cornvik, T., Dahlroth, S.L., Magnusdottir, A., Herman, M.D., Knaust, R., Ekberg, M., and Nordlund, P. (2005) Colony filtration blot: a new screening method for soluble protein expression in Escherichia coli. *Nat. Methods.* **2** (7): 507-509.

Cowtan, K. (2000) General quadratic functions in real and reciprocal space and their application to likelihood phasing. *Acta Cryst. D,* **56**: 1612-1621.

Cowtan, K. (2006) The Buccaneer software for automated model building *Acta Cryst. D,* **62**: 1002-1011.

Craig, L., Volkmann, N., Arvai, A.S., Pique, M.E., Yeager, M., Egelman, E.H., and Tainer, J.A. (2006) Type IV pilus structure by cryo-electron microscopy and crystallography: implications for pilus assembly and functions. *Mol. Cell*, **23** (5): 651-662.

Criss, A.K., Kline, K.A., and Seifert H.S. (2005) The frequency and rate of pilin antigenic variation in Neisseria gonorrhoeae. *Mol. Microbiol.,* **58** (2): 510-519.

Cserzo, M., Wallin, E., Simon, I., von Heijne, G., and Elofsson, A. (1997) Prediction of transmembrane alpha-helices in prokaryotic membrane proteins: the Dense Alignment Surface method. *Prot. Eng.,* **10** (6): 673-676.

Dal Conte, I., Starnino, S., Di Perri, G., and Stefanelli, P. (2006) Disseminated gonococcal infection in an immunocompetent patient caused by an imported Neisseria gonorrhoeae multidrug-resistant strain. *J. Clin. Microbiol.,* **44** (10): 3833-384.

Dale, J., and Park, S. Molecular genetics of bacteria. (2004) 4th ed. Chichester, West Sussex, England: John Wiley & Sons.

Dahlroth, S.L., Lieu, V., Haas, J., and Nordlund P. (2009) Screening colonies of pooled ORFeomes (SCOOP): a rapid and efficient strategy for expression screening ORFeomes in Escherichia coli. *Protein. Expr. Purif.,* **68** (2): 121-127.

Dahlroth, S.L., Nordlund, P., and Cornvik, T. (2006) Colony filtration blotting for screening soluble expression in Escherichia coli. *Nat. Protoc.,* **1** (1): 253-258.

Dalton, K.M., Krosson, S. (2009) A conserved mode of protein recognition and binding in a ParD-ParE toxin-antitoxin complex. *Biochemistry,* **49**(10): 2205-15.

Dao-Thi, M.H., Van Melderen, L., De Genst, E., Afif, H., Buts, l., Wyns, L., and Loris, L. Molecular basis of gyrase poisoning by the addiction toxin CcdB. (2005) *J. Mol. Biol.,* **348** (5): 1091-1102.

Daubin, V., Moran, N. A., and Ochman, H. (2003) Phylogenetics and the Cohesion of Bacterial Genomes. *Science,* **301** (5634): 829-832.

Davidsen, T., Rødland, E.A., Lagesen, K., Seeberg, E., Rognes, T., and Tønjum, T. (2004) Biased distribution of DNA uptake sequences towards genome maintenance genes. *Nucleic Acids Res.,* **32** (3): 1050-1058.

Dawkins, R.. The selfish gene. (1989) New ed. Oxford: Oxford University Press.

De Jonge, N., Hohlweg, W., Garcia-Pino, A., Respondek, M., Buts, L., Haesaerts, S., Lah, J., Zangger, K., and Loris, R. (2010) Structural and thermodynamic characterization of Vibrio fischeri CcdB. *J. Biol Chem.,* **285** (8): 5606-5613.

Denamur, E., and Ivan, M. (2006) Evolution of mutation rates in bacteria. *Mol. Microbiol.,* **60** (4): 820-827.

Dillard, J.P., and Seifert, H.S. (1997) A peptidoglycan hydrolase similar to bacteriophage endolysins acts as an autolysin in Neisseria gonorrhoeae. *Mol. Microbiol.,* **25** (5): 893-901.

Donlan, R.M. (2001) Biofilms and device-associated infections. *Emerg. Infect. Dis.* **7** (2): 277-281.

Dos Vultos, T., Mestre, O., Tønjum, T., and Gicquel, B. (2009) DNA repair in Mycobacterium tuberculosis revisited. *FEMS Microbiol. Rev.,* **33** (3): 471-487.

Draskovic, I., and Dubnau, D. (2005) Biogenesis of a putative channel protein, ComEC, required for DNA uptake: membrane topology, oligomerization and formation of disulphide bonds. *Mol. Microbiol.,* **55** (3): 881-896.

Duffin, P.M., and Seifert, H.S. (2010) DNA uptake sequence-mediated enhancement of transformation in Neisseria gonorrhoeae is strain dependent. *J. Bacteriol.,* **192** (17): 4436-4444.

Elkins, C., Thomas, C.E., Seifert, H.S., and Sparling, P.F. (1991) Species-specific uptake of DNA by gonococci is mediated by a 10-base-pair sequence. *J. Bacteriol.,* **173** (12): 3911-3913.

Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D. Biol. Crystallogr.,* **60** (Pt 12 Pt 1): 2126-2132.

Engelberg-Kulka, H., and Glaser, G. (1999) Addiction modules and programmed cell death and antideath in bacterial cultures. *Annu. Rev. Microbiol.,* **53**: 43-70.

Engelberg-Kulka, H., Amitai, S., Kolodkin-Gal, I., and Hazan, R. (2006) Bacterial programmed cell death and multicellular behavior in bacteria. *PLoS Genet,.* **2** (10).

Errington, J. (2003) Regulation of endospore formation in Bacillus subtilis. *Nat. Rev., Microbiol.,* **1** (2): 117-126.

Evans, P. (2006) Scaling and assessment of data quality. *Acta Crystallogr. D. Biol. Crystallogr.,* **62** (Pt 1): 72-82.

Facius, D., and Meyer, TF. (1993) A novel determinant (comA) essential for natural transformation competence in Neisseria gonorrhoeae and the effect of a comA defect on pilin variation. *Mol. Microbiol.,* **10** (4): 699-712.

Faguy, DM. (2003) Lateral gene transfer (LGT) between Archaea and Escherichia coli is a contributor to the emergence of novel infectious disease. *BMC Infect. Dis.,* **3** (13).

Fineran, P.C., Blower, TR. . Foulds, IJ Humphreys, DP. Lilley, KS. and Salmond, GP.. (2009) The phage abortive infection system, ToxIN, functions as a protein-RNA toxin-antitoxin pair. *Proc. Natl. Acad. Sci. U.S.A.,* **106** (3): 894-899.

Flemming, H.C., and Wingender, J. (2010) The biofilm matrix. *Nat. Rev., Microbiol.,* **8** (9): 623-633.

Fox, M.S. (1960) The fate of transforming deoxyribonucleate following fixation by transformable bacteria. *Nature,* **187**: 1004-1006.

Fozo, E.M., Hemm, M.R., and Storz, G. (2008) Small toxic proteins and the antisense RNAs that repress them. *Microbiol. Mol. Biol. Rev.,* **72** (4): 579-589.

Francuski, D., and Saenger, W. (2009) Crystal structure of the antitoxin-toxin protein complex RelB-RelE from Methanococcus jannaschii. *J. Mol. Biol.,* **393** (4): 898-908.

Fussenegger, M., Kahrs, A.F., Facius, D., and Meyer, T.F. (1996) Tetrapac (tpc), a novel genotype of Neisseria gonorrhoeae affecting epithelial cell invasion, natural transformation competence and cell separation. *Mol. Microbiol.,* **19** (6): 1357-1372.

García-Ortíz, M.V., Marsin, S., Arana, M.E., Gasparutto, D., Guérois, R., Kunkel, T.A., Radicella, J.P. (2011) Unexpected Role for Helicobacter pylori DNA Polymerase I As a Source of Genetic Variability. *PLoS Genet.,* **7** (6).

Gardner, A., West, S.A., and Griffin, A.S. (2007) Is bacterial persistence a social trait? *PLoS One,* **2** (8).

Gerdes, K., Larsen, J.E., and Molin, S. (1985) Stable inheritance of plasmid R1 requires two different loci. *J. Bacteriol.,* **161** (1): 292-298.

Gerdes, K., Rasmussen, P.B., and Molin, S. (1986) Unique type of plasmid maintenance function: postsegregational killing of plasmid-free cells. *Proc. Natl. Acad. Sci. U.S.A.,* **83** (10): 3116-3120.

Gerdes, K., Gultyaev, A.P., Franch, T., Pedersen, K., and Mikkelsen, N.D. (1997) Antisense RNA-regulated programmed cell death. *Annu. Rev. Genet.,* **31**: 1-31.

Gerdes, K., Christensen, S.K., and Løbner-Olesen, A. (2005) Prokaryotic toxin-antitoxin stress response loci. *Nat. Rev., Microbiol.,* **3** (5): 371-482.

Gil-Prieto, R., García-García, L., Alvaro-Meca, A., González-Escalada, A., Viguera Ester, P., and Gil De Miguel, A. (2011) The burden of hospitalizations for meningococcal infection in Spain (1997-2008). *Vaccine,* **29** (34): 5765-5770.

Gotfredsen, M., and Gerdes, K. (1998) The Escherichia coli relBE genes belong to a new toxin-antitoxin gene family. *Mol. Microbiol.,* **29** (4): 1065-1076.

Griffith, F. (1928) The Significance of Pneumococcal Types. *J. Hyg. (Lond).,* **27** (2): 113-159.

Grkovic, S., Brown, M.H., and Skurray, R.A. (2002) Regulation of Bacterial Drug Export Systems. *Microbiol. Mol. Biol. Rev.*, **66** (4): 671-701.

Grønlund, H., Gerdes, K. (1999) Toxin-antitoxin systems homologous with relBE of Escherichia coli plasmid P307 are ubiquitous in prokaryotes. *J. Mol. Biol.,* **285** (4): 1401-1415.

Gur, E., Vishkautzan, M., and Sauer, R.T. (2012) Protein unfolding and degradation by the AAA+ Lon protease. *Protein Sci.*, **21**(2):268-78.

Hall-Stoodley, L., Costerton, J.W., and Stoodley, P. (2004) Bacterial biofilms: from the natural environment to infectious diseases. *Nat. Rev., Microbiol.,* **2** (2): 95-108.

Hamilton, H.L., and Dillard J.P. (2006) Natural transformation of Neisseria gonorrhoeae: from DNA donation to homologous recombination. *Mol. Microbiol.,* **59** (2): 376-385.

Hamilton, H.L., Domínguez, N.M., Schwartz, K.J., Hackett, K.T., and Dillard J.P. (2005) "Neisseria gonorrhoeae secretes chromosomal DNA via a novel type IV secretion system." *Mol. Microbiol.,* **55** (6): 1704-1721.

Hanage, W.P., Fraser, C., Tang, J., Connor, T.R., and Corander, J. (2009) "Hyper-recombination, diversity, and antibiotic resistance in pneumococcus." *Science,* **324** (5933) : 1454-1457.

Hastings, P.J., Rosenberg, S.M., and Slack, A. (2004) "Antibiotic-induced lateral transfer of antibiotic resistance." *Trends Microbiol.,* **12** (9): 401-404.

Heath, P.T. (2011) "An update on vaccination against group B streptococcus." *Expert Rev. Vaccines,* **10** (5): 685-694.

Hill, S.A., and Davies, J.K. (2009) "Pilin gene variation in Neisseria gonorrhoeae: reassessing the old paradigms." *FEMS Microbiol. Rev.,* **33** (3): 521-530.

Hurley, J.M., and Woychik, N.A. (2009) "Bacterial toxin HigB associates with ribosomes and mediates translation-dependent mRNA cleavage at A-rich sites." *J. Biol Chem.,* **284** (28): 18605-18613.

Iwasaki, W., and Takagi, T. (2009) "Rapid pathway evolution facilitated by horizontal gene transfers across prokaryotic lineages." *PLoS Genet.,* **5** (3).

Jain, S., Mościcka, K.B., Bos, M.P., Pachulec, E., Stuart, M.C., Keegstra, W., Boekema, E.J., van der Does, C. (2011) "Structural characterization of outer membrane components of the type IV pili system in pathogenic Neisseria." *PLoS ONE,* **6** (1).

Jernberg, C., Löfmark, S., Edlund, C., Jansson, J.K. (2010) "Long-term impacts of antibiotic exposure on the human intestinal microbiota." *Microbiology (Reading, England),* **156** (11): 3216-3223.

Jevitz Patterson, M. "Chapter 13, Streptococcus." Medical Microbiology. (1996) 4th ed. Galveston, Texas: University of Texas Medical Branch at Galveston.

Jiang, S.C., and Paul, J.H. (1998) Gene Transfer by Transduction in the Marine Environment *Appl. Environ. Microbiol.,* **64** (8): 2780-2787.

Jiang, Y., Pogliano, J., Helinski, D.R., and Konieczny, I. (2002) "ParE toxin encoded by the broad-host-range plasmid RK2 is an inhibitor of Escherichia coli gyrase." *Mol. Microbiol.,* **44** (4): 971-979.

Johnsborg, O., Eldholm, V., Håvarstein, L.S. (2007) "Natural genetic transformation: prevalence, mechanisms and function." *Res. Microbiol.,* **158** (10): 767-778.

Jolivet-Gougeon, A., Kovacs, B., Le Gall-David, S., Le Bars, H., Bousarghin, L., Bonnaure-Mallet, M., Lobel, B., Guillé F, Soussy, C.J., and Tenke, P. (2011) "Bacterial hypermutation: clinical implications." *J. Med. Microbiol.,* **60** (5): 563-573.

Jones, D.T. (1999) Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.,* **292**: 195-202.

Juretic, D., Zoranic, L., and Zucic, D. (2002) Basic charge clusters and predictions of membrane protein topology. *J. Chem. Inf. Comput. Sci.,* **42**: 620-632.

Hirokawa, T., Boon-Chieng, S., and Mitaku, S., (1998) SOSUI: classification and secondary structure prediction system for membrane proteins. *Bioinformatics.,* **14**: 378-379.

Hofmann, K. and Stoffel, W. (1993) TMbase - A database of membrane spanning proteins segments. *Biol. Chem. Hoppe-Seyler,* **374**: 166.

Holm, L., and Rosenström, P. (2010) Dali server: conservation mapping in 3D. *Nucleic Acids Res.,* **38** (Web Server issue): W545-9.

Kabsch, W. (2010) XDS. *Acta Crystallogr. D. Biol. Crystallogr.,* **66** (Pt 2):125-132.

Kamada, K., and Hanaoka, F. (2005) "Conformational change in the catalytic site of the ribonuclease YoeB toxin by YefM antitoxin." *Mol. Cell,* **19** (4): 497-509.

Kamphuis, M.B., Bonvin, A.M., Monti, M.C., Lemonnier, M., Muñoz-Gómez, A., van den Heuvel, R.H., Díaz-Orejas, and R., Boelens, R. (2006) "Model for RNA binding and the catalytic site of the RNase Kid of the bacterial parD toxin-antitoxin system." *J. Mol. Biol.,* **357** (1): 115-126.

Kefala, G., Ahn, C., Krupa, M., Esquivies, L., Maslennikov,I., Kwiatkowski, W., and Choel, S. (2010) Structures of the OmpF porin crystallized in the presence of foscholine-12. *Protein Sci.*, **19** (5): 1117–1125.

Kendrew, J.C., Bodo, G., Dintzis, H.M., Parrish, R.G., Wyckoff, H.M., and Phillips, D.C. (1958) A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature,* **181** (4610): 662-666.

Khoo, S.K., Loll, B., Chan, W.T., Shoeman, R.L., Ngoo, L., Yeo, C.C., and Meinhart, A. (2007) "Molecular and structural characterization of the PezAT chromosomal toxin-antitoxin system of the human pathogen Streptococcus pneumoniae." *J. Biol Chem.,* **282** (27): 19606-19618.

Kihara, D., Shimizu,T., and Kanehisa, M. (1998) Prediction of Membrane Proteins Based on Classification of Transmembrane Segments. *Protein Eng.,* **11**: 961-970.

Kinoshita, K., and Nakamura, H. (2003) Identification of protein biochemical functions by similarity search using the molecular surface database eF-site. *Protein Sci.,* **12** (8): 1589-1595.

Kokjohn, T. A. (1989). Transduction: mechanism and potential for gene transfer in the environment. In: Gene Transfer in the Environment (S. B. Levy and R. V. Miller, Eds): 73-97.

Kramer, N., Hahn, J., and Dubnau, D. (2007) "Multiple interactions among the competence proteins of Bacillus subtilis." *Mol. Microbiol.,* **65** (2): 454-464.

Kresge, N., Simoni, R.D., and Hill, R.L. (1942) "Selman Waksman: the Father of Antibiotics." *J. Biol. Chem.,* **142**: 519-528.

Krogh A, Larsson B, von Heijne G, Sonnhammer EL. (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.,* **305** (3): 567-580.

Kroll, J.S., Wilks, K.E., Farrant, J.L., and Langford, P.R. (1998) "Natural genetic exchange between Haemophilus and Neisseria: intergeneric transfer of chromosomal genes between major human pathogens." *Proc. Natl. Acad. Sci. U.S.A.,* **95** (21): 12381-12385.

Kumar, A., and Schweizer, H.F. (2005) "Bacterial resistance to antibiotics: active efflux and reduced uptake." *Adv. Drug Deliv. Rev.,* **57**: 1486-1513.

Kunin, V., Goldovsky, L., Darzentas, N., and Ouzounis, C.A. (2005) "The net of life: Reconstructing the microbial phylogenetic network." *Genome Res.*, Cold Spring Harbor Laboratory Press **15**: 954-959.

Kussell, E., Kishony, R., Balaban, N.Q., and Leibler, S. (2005) "Bacterial persistence: a model of survival in changing environments." *Genetics,* **169** (4): 1807-1814.

Lee, H.H., Molla, M.N., Cantor, C.R., and Collins, J.J. (2010) "Bacterial charity work leads to population-wide resistance." *Nature,* **467** (7311): 82-85.

Leplae, R., Geeraerts, D., Hallez, R., Guglielmini, J., Drèze, P., and Van Melderen, L. (2011) "Diversity of bacterial type II toxin-antitoxin systems: a comprehensive search and functional analysis of novel families." *Nucleic Acids Res.,* **39** (13): 5513-5525.

Levin, B.R., and Rozen, D.E. (2006) "Non-inherited antibiotic resistance." *Nat. Rev., Microbiol.,* **4** (7): 556-562.

Lee Ligon, B. (2004) "Penicillin: its discovery and early development." *Semin. Pediatr. Infect. Dis.,* **15** (1): 52-57.

Lee Ligon, B. (2005) "Albert Ludwig Sigesmund Neisser: discoverer of the cause of gonorrhea." *Semin. Pediatr. Infect. Dis.,* **16** (4): 336-341.

Leslie, A.G.W., and Powell, H.R. (2007) Evolving Methods for Macromolecular Crystallography. ISBN 978-1-4020-6314-5 **245**: 41-51.

Letunic I, Doerks T, Bork P. (2012) SMART 7: recent updates to the protein domain annotation resource. *Nucleic Acids Res.,* **40** (Database issue): D302-5.

Lin, T., Gao, L. Edmondson, D.G., Jacobs, M.B., Philipp, M.T., and Norris, S.J. (2009) "Central role of the Holliday junction helicase RuvAB in vlsE recombination and infectivity of Borrelia burgdorferi." *PLoS Pathog.,* **5** (12).

Linding, R., Russell, R.B., Neduva, V., and Gibson, T.J. (2003) GlobPlot: Exploring protein sequences for globularity and disorder. *Nucleic Acids Res.,* **31** (13): 3701-3708

Lioy, V.S., Martín, M.T., Camacho, A.G., Lurz, R., Antelmann, H., Hecker, M., Hitchin, E., Ridge, Y., Wells, J.M., and Alonso, J.C. (2006) "pSM19035-encoded zeta toxin induces stasis followed by death in a subpopulation of cells." *Microbiology,* **152** (8): 2365-2379.

Liu, M., Zhang, Y., Inouye, M., and Woychik, N.A. (2008) "Bacterial addiction module toxin Doc inhibits translation elongation through its association with the 30S ribosomal subunit." *Proc. Natl. Acad. Sci. U.S.A.,* **105** (15): 5885-5890.

Llosa, M., Gomis-Rüth, F.X., Coll M., and de la Cruz Fd, F. (2002) "Bacterial conjugation: a two-step mechanism for DNA transport." *Mol. Microbiol.,* **45** (1): 1-8.

Lodge, A.P., Walsh, A., McNamee, C.J., and Moss, D.J. (1999) Identification of chURP, a nuclear calmodulin-binding protein related to hnRNP-U. *Eur. J. Biochem.,* **261** (1): 137-47.

Macià, M.D., Pérez, J.L., Molin, S., and Oliver, A. (2011) "Dynamics of mutator and antibiotic-resistant populations in a pharmacokinetic/pharmacodynamic model of Pseudomonas aeruginosa biofilm treatment." *Antimicrob. Agents Chemother.,* **55** (11): 5230-5237.

Maciá, M.D., Blanquer, D., Togores, B., Sauleda, J., Pérez, J.L., and Oliver, A. (2005) "Hypermutation Is a Key Factor in Development of Multiple-Antimicrobial Resistance in Pseudomonas aeruginosa Strains Causing Chronic Lung Infections." *Antimicrob. Agents Chemother.,* **49** (8): 3382-3386 .

Madl, T., Van Melderen, L., Mine, N., Respondek, M., Oberer, M., Keller, W., Khatai, L., and Zangger, K. (2006) "Structural basis for nucleic acid and toxin recognition of the bacterial antitoxin CcdA." *J. Mol. Biol.,* **364** (2): 170-185.

Maier, B., Koomey, M., and Sheetz, M.P. (2004) "A force-dependent switch reverses type IV pilus retraction." *Proc. Natl. Acad. Sci. U.S.A.,* **101** (30): 10961-10966.

Maiques, E., Ubeda, C., Campoy, S., Salvador, N., Lasa, I., Novick, R.P., Barbé, J., and Penadés, J.R. (2006) "beta-lactam antibiotics induce the SOS response and horizontal transfer of virulence factors in Staphylococcus aureus." *J. Bacteriol.,* **188** (7): 2726-2729.

Maisonneuve, E., Shakespeare, L.J., Jørgensen, M.G., and Gerdes, K. (2011) "Bacterial persistence by RNA endonucleases." *Proc. Natl. Acad. Sci. U.S.A.,* **108** (32): 13206-13211.

Marsden, R.L., McGuffin, L.J., and Jones, D.T. (2002) Rapid protein domain assignment from amino acid sequence using predicted secondary structure. *Protein Sci.,* **11** (12): 2814-2824.

Martinez Molina, D., Cornvik, T., Eshaghi, S., Haeggström, J.Z., Nordlund, P., and Sabet, M.I. (2008) Engineering membrane protein overproduction in Escherichia coli. *Protein Sci.,* **17** (4): 673-680.

Mattison, K., Wilbur, J.S., So, M., and Brennan, R.G. (2006) "Structure of FitAB from Neisseria gonorrhoeae bound to DNA reveals a tetramer of toxin-antitoxin heterodimers containing pin domains and ribbon-helix-helix motifs." *J. Biol Chem.,* **281** (49): 37942-37951.

McArthur, J.D., Cook, S.M., Venturini, and C., Walker, M.J. (2012) "The Role of Streptokinase as a Virulence Determinant of Streptococcus pyogenes - Potential for Therapeutic Targeting." *Curr. Drug Targets,* **13** (3): 297-307.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.,* **40** (Pt 4): 658-674.

"Meningococcal vaccines: WHO position paper." (2011) *Wkly Epidemiol. Rec.,* 18 Nov. : 521-540.

Meyer, E.E. (1997) The first years of the Protein Data Bank. *Protein Science,* **6** (7): 1591–1597.

Meyers, L.A., Levin, B.R., Richardson, A.R., and Stojiljkovic, I.(2003) "Epidemiology, hypermutation, within-host evolution and the virulence of Neisseria meningitidis." *Proc. Biol. Sci.,* **270** (1525): 1667-1677 .

Miao, R., and Guild, W.R. (1970) "Competent Diplococcus pneumoniae Accept Both Single- and Double-Stranded Deoxyribonucleic Acid." *J. Bacteriol.,* **101** (2): 361-364.

Michel, B. (2005) "After 30 years of study, the bacterial SOS response still surprises us." *PLoS Biol.,* **3** (7).

Miller, C., Thomsen, L.E., Gaggero, C., Mosseri, R., Ingmer, H., and Cohen, S.N. (2004) "SOS response induction by beta-lactams and bacterial defense against antibiotic lethality." *Science,* **305** (5690): 1629-1631.

Miller, K., O'Neill, A.J., and Chopra, I. (2002) "Response of Escherichia coli hypermutators to selection pressure with antimicrobial agents from different classes." *J. Antimicrob. Chemother.,* **49** (6): 925-934.

Mochizuki, A., Yahara, K., Kobayashi, I., and Iwasa, Y. (2006) "Genetic addiction: selfish gene's strategy for symbiosis in the genome." *Genetics,* **172** (2): 1309-1323.

Monfort, L., Caro, V., Devaux, Z., Delannoy, A.S., Brisse, S., and Sednaoui, P. (2009) "First neisseria gonorrhoeae genotyping analysis in france: identification of a strain cluster with reduced susceptibility to Ceftriaxone." *J. Clin. Microbiol.,* **47** (11): 3540-3545.

Mook-Kanamori, B.B., Geldhoff, M., van der Poll, T., and van de Beek, D. (2011) "Pathogenesis and pathophysiology of pneumococcal meningitis." *Clin. Microbiol. Rev.,* **24** (3): 557-591.

Moreno-Córdoba, I., Diago-Navarro, E., Barendregt, A., Heck, A.J., Alfonso, C., Díaz-Orejas, R., Nieto, C., and Espinosa, M. (2012) The toxin-antitoxin proteins relBE2Spn of Streptococcus pneumoniae: Characterization and association to their DNA target. *Proteins.,* **80**(7):1834-46.

Morrison, D,A., and Mannarelli, B. (1979) "Transformation in pneumococcus: nuclease resistance of deoxyribonucleic acid in the eclipse complex." *J. Bacteriol.,* **140** (2): 655-665.

Morse, S.A. (1996) "Neisseria, Moraxella, Kingella and Eikenella." In: Baron S, editor. *Medical Microbiology.* 4th edition. Galveston (TX): University of Texas Medical Branch at Galveston; 1996. Chapter 14.

Mortier-Barrière, I., Velten, M., Dupaigne, P., Mirouze, N., Piétrement, O., McGovern, S., Fichant, G., Martin, B., Noirot, P., Le Cam, E., Polard, P., and Claverys, J.P. (2007) "A key presynaptic role in transformation for a widespread bacterial protein: DprA conveys incoming ssDNA to RecA." *Cell*, **130** (5): 824-836.

Mudany, M.A., Kikuchi, K., Totsuka, K., and Uchiyama, T. (2003) "Evaluation of a new serotyping kit for Streptococcus pneumoniae." *J. Med. Microbiol.,* **52** (11): 975-980.

Murshudov, G.N., Skubák, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A., Winn, M.D., Long, F., and Vagin, A.A. (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D. Biol. Crystallogr.,* **67** (Pt 4): 355-367.

Mutschler, H., Reinstein, J., and Meinhart, A. (2010) "Assembly dynamics and stability of the pneumococcal epsilon zeta antitoxin toxin (PezAT) system from Streptococcus pneumoniae." *J. Biol Chem.,* **285** (28): 21797-21806.

Mutschler, H., Gebhardt, M., Shoeman, R.L., and Meinhart, A. (2011) "A novel mechanism of programmed cell death in bacteria by toxin-antitoxin systems corrupts peptidoglycan synthesis." *PLoS Biol.,* **9** (3).

Nadell, C.D., Xavier, J. B., Levin, S.A., and Foster, K.R. (2008) "The evolution of quorum sensing in bacterial biofilms." *PLoS Biol.,* **6** (1).

Nariya, H., and Inouye, M. (2008) "MazF, an mRNA interferase, mediates programmed cell death during multicellular Myxococcus development." *Cell,* **132** (1): 55-66.

Neubauer, C., Gao, Y.G., Andersen, K.R., Dunham, C.M., Kelley, A.C., Hentschel, J., Gerdes, K., Ramakrishnan, V., Brodersen, D.E. (2009) "The structural basis for mRNA recognition and cleavage by the ribosome-dependent endonuclease RelE." *Cell,* **139** (6): 1084-1095.

Nieto, C., Pellicer, T., Balsa, D., Christensen, S.K., Gerdes, K., and Espinosa, M. (2006) "The chromosomal relBE2 E2Spn toxin-antitoxin system of Streptococc toxin-antitoxin locus of Streptococcus pneumoniae: characterization and use of a bioluminescence resonance energy transfer assay to detect toxin-antitoxin interaction." *Mol. Microbiol.,* **59** (4): 1280-1296.

Nieto, C., Cherny, I., Khoo, S.K., de Lacoba, M.G., Chan, W.T., Yeo, C.C., Gazit, E., and Espinosa, M. (2007) "The yefM-yoeB toxin-antitoxin systems of Escherichia coli and Streptococcus pneumoniae: functional and structural correlation." *J. Bacteriol.,* **189** (4): 1266-1278.

Nieto, C., Sadowy, E., de la Campa, A.G., Hryniewicz, W., and Espinosa, M. (2010) "The relBE2Spn toxin-antitoxin system of Streptococcus pneumoniae: role in antibiotic tolerance and functional conservation in clinical isolates." *PLoS ONE,* **5** (6).

Nossal, G.J. (2011) "Vaccines of the future." *Vaccine* ,29 (Suppl. 4).

Ogura, T., and Hiraga, S. (1983) "Mini-F plasmid genes that couple host cell division to plasmid proliferation." *Proc. Natl. Acad. Sci. U.S.A.,* **80** (15): 4784-4788.

Oliver, A., and Mena, A. (2010) "Bacterial hypermutation in cystic fibrosis, not only for antibiotic resistance." *Clin. Microbiol. Infect.*, **16** (7): 798-808.

Otwinowski, Z. (1991) In Wolf, W., Evans, P.R. and Leslie, A.G.W. (ed.) Isomorphous replacement and anomalous scattering, Daresbury Laboratory, Warrington: 80.

Overgaard, M., Borch, J., Jørgensen, M.G., and Gerdes, K. (2008) "Messenger RNA interferase RelE controls relBE transcription by conditional cooperativity." *Mol. Microbiol.,* **69** (4): 841-857.

Overgaard, M., Borch, J., and Gerdes, K. (2009) "RelB and RelE of Escherichia coli form a tight complex that represses transcription via the ribbon-helix-helix motif in RelB." *J. Mol. Biol.,* **394** (2): 183-196.

Maskell, J.P., Sefton, A.M., and Hall, L.M. (2001) "Multiple Mutations Modulate the Function of Dihydrofolate Reductase in Trimethoprim-Resistant Streptococcus pneumoniae." *Antimicrob. Agents Chemother.,* **45** (4): 1104-1108 .

Palchevskiy, V., and Finkel, S.E. (2006) "Escherichia coli competence gene homologs are essential for competitive fitness and the use of DNA as a nutrient." *J. Bacteriol.,* **188** (11): 3902-3910.

Pandey, D.P., and Gerdes, K. (2005) "Toxin-antitoxin loci are highly abundant in free-living but lost from host-associated prokaryotes." *Nucleic Acids Res.,* **33** (3): 966-976.

Panjikar, S., Parthasarathy, V., Lamzin, V. S., Weiss, M. S. & Tucker, P. A. (2005). Auto-Rickshaw - An automated crystal structure determination platform as an efficient tool for the validation of an X-ray diffraction experiment. *Acta Crystallogr D Biol Crystallogr.,* **61**(Pt 4):449-57.

Panoff, J.M, and Chuiton, C. (2004) "Horizontal Gene Transfer: A Universal Phenomenon." *Human and Ecological Risk Assessment,* **10** (5): 939-943.

Partridge, S.R. (2011) "Analysis of antibiotic resistance regions in Gram-negative bacteria." *FEMS Microbiol. Rev.,* **35** (5): 820-855.

Plant, L.J., and Jonsson, A.B. (2006) "Type IV pili of Neisseria gonorrhoeae influence the activation of human CD4+ T cells." *Infect. Immun.,* **74** (1): 442-448.

Prescott, L.M., Klein, D.A., and Harley, J.P. Microbiology. (2002) 5th ed. McGraw-Hill.

Prilusky, J., Felder, C.E., Zeev-Ben-Mordehai, T., Rydberg, E.H., Man, O., Beckmann, J.S., Silman, I., and Sussman, J.L. (2005) "FoldIndex: a simple tool to predict whether a given protein sequence is intrinsically unfolded." *Bioinformatics.,* **21** (16): 3435-3438.

Provvedi, R., and Dubnau, D. (1999) "ComEA is a DNA receptor for transformation of competent Bacillus subtilis." *Mol. Microbiol.,* **31** (1): 271-280.

PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC.

Ramsey, M.E., Woodhams, K.L., and Dillard, J.P. (2011) "The Gonococcal Genetic Island and Type IV Secretion in the Pathogenic Neisseria." *Front. Microbiol.,* **2** (61).

Rhodes, G. Cristallography Made Crystal Clear, Third Edition: A guide for Users of Macromolecular Models. (2006) Academic Press.

Saavedra De Bast, M., Mine, N., Van Melderen, L. (2008) "Chromosomal toxin-antitoxin systems may act as antiaddiction modules." *J. Bacteriol.*, **190** (13): 4603-4609.

Salgado-Pabón, W., Du, Y., Hackett, K.T., Lyons, K.M., Arvidson, C.G., and Dillard J.P. (2010) "Increased expression of the type IV secretion system in piliated Neisseria gonorrhoeae variants." *J. Bacteriol.,* **192** (7): 1912-1920.

Saunders, N.J., Moxon, E.R., and Gravenor, M.B. (2003) "Mutation rates: estimating phase variation rates when fitness differences are present and their impact on population structure." *Microbiology,* **149** (2): 485-495.

Sbordone, L., and Bortolaia, C. (2003) "Oral microbial biofilms and plaque-related diseases: microbial communities and their role in the shift from oral health to disease." *Clin. Oral Investig.,* **7** (4): 181-188.

Schumacher, M.A., Piro, K.M., Xu, W., Hansen, S., Lewis, K., and Brennan, R.G. (2009) "Molecular mechanisms of HipA-mediated multidrug tolerance and its neutralization by HipB." *Science,* **323** (5912): 396-401.

Schwarz, S., Kehrenberg, C., Doublet, B., Cloeckaert, A. (2004) "Molecular basis of bacterial resistance to chloramphenicol and florfenicol." *FEMS Microbiol. Rev.,* **28** (5): 519-542.

Shao, Y., Harrison, E.M., Bi, D., Tai, C., He, Xk, Ouk, H.Y., Rajakumar, K., and Deng, Z. (2011) "TADB: a web-based resource for Type 2 toxin-antitoxin loci in bacteria and archaea." *Nucleic Acids Res.,* **39**.

Sheldrick, G.M. (2010) "Experimental phasing with SHELXC/D/E: combining chain tracing with density modification." *Acta Cryst. D,* **66**: 479-485.

Sinha, S., Ambur, O.H., Langford, P.R., Tønjum, T., and Kroll, J.S. (2008) "Reduced DNA binding and uptake in the absence of DsbA1 and DsbA2 of Neisseria meningitidis due to inefficient folding of the outer-membrane secretin PilQ." *Microbiology (Reading, England),* **154** (1): 217-225.

Smith, J.M., Smith, N.H., O'Rourke, M., and Spratt, B.G. (1993) "How clonal are bacteria?." *Proc. Natl. Acad. Sci. U.S.A.,* **90** (10): 4384-4388.

Snyder, L.A., Saunders, N.J., and Shafer, W.M. (2001) "A putatively phase variable gene (dca) required for natural competence in Neisseria gonorrhoeae but not Neisseria meningitidis is located within the division cell wall (dcw) gene cluster." *J. Bacteriol.,* **183** (4): 1233-1241.

Steer, A.C., Batzloff, M.R., Mulholland, K., Carapetis, J.R. (2009) "Group A streptococcal vaccines: facts versus fantasy." *Curr. Opin. Infect. Dis.,* **22** (6): 544-552.

Stefanelli, P. (2011) "Emerging resistance in Neisseria meningitidis and Neisseria gonorrhoeae." *Expert Rev. Anti. Infect. Ther.,* **9** (2): 237-244.

Syvanen, M., and Clarence, I.K. Horizontal gene transfer. (2002) 2nd ed. San Diego: Academic Press.

Szeto, T.H., Dessen, A., and Pelicic, V. (2011) "Structure/function analysis of Neisseria meningitidis PilW, a conserved protein that plays multiple roles in type IV pilus biology." *Infect. Immun.,* **79** (8): 3028-3035.

Sørensen, S.J., Bailey, M., Hansen, L.H., Kroer, N., and Wuertz, S. (2005) "Studying plasmid horizontal transfer in situ: a critical review." *Nat. Rev., Microbiol.,* **3** (9): 700-710.

Takagi, H., Kakuta, Y., Okada, T., Yao, M., Tanaka, I., and Kimura, M. (2005) "Crystal structure of archaeal toxin-antitoxin RelE-RelB complex with implications for toxin activity and antitoxin effects." *Nat. Struct. Mol. Biol.,* **12** (4): 327-331.

Torres-Cruz, J., and van der Woude, M.W. (2003) "Slipped-strand mispairing can function as a phase variation mechanism in Escherichia coli." *J. Bacteriol.,* **185** (23): 6990-6994.

Tusnády, G.E., and Simon, I. (2001) The HMMTOP transmembrane topology prediction server. *Bioinformatics,* **17** (9): 849-850.

Tyndall, C., Lehnherr, H., Sandmeier, U., Kulik, E., and Bickle, T.A. (1997) "The type IC hsd loci of the enterobacteria are flanked by DNA with high homology to the phage P1 genome: implications for the evolution and spread of DNA restriction systems." *Mol. Microbiol.,* **23** (4): 729-736.

Unemo, M., Fasth, O., Fredlund, H., Limnios, A., and Tapsall, J. (2009) "Phenotypic and genetic characterization of the 2008 WHO Neisseria gonorrhoeae reference strain panel intended for global quality assurance and quality control of gonococcal antimicrobial resistance surveillance for public health purposes." *J. Antimicrob. Chemother.,* **63** (6): 1142-1151.

Vagin, A., and Teplyakov, A. (2010) Molecular replacement with MOLREP. *Acta Crystallogr. D Biol Crystallogr.,* **66** (Pt 1): 22-25.

Van der Woude, M.W., and Henderson, I.R. (2008) "Regulation and function of Ag43 (flu)." *Annu. Rev. Microbiol.,* **62**: 153-169.

Van der Woude M.W. (2011) "Phase variation: how to create and coordinate population diversity." *Curr. Opin. Microbiol.,* **14** (2): 205-211.

Venema, G., Pritchard, R.H., and Venema-Schroeder, T. (1965) "Fate of transforming deoxyribonucleic acid in Bacillus subtilis" *J. Bacteriol.*, **89**: 1250-5.

Van Melderen, L., and Saavedra De Bast M. (2009) "Bacterial toxin-antitoxin systems: more than selfish entities?" *PLoS Genet.,* **5** (3).

Van Melderen, L. (2010) "Toxin-antitoxin systems: why so many, what for?" *Curr. Opin. Microbiol.,* **13** (6): 781-785.

Villarreal, L.P. (2011) "Viral ancestors of antiviral systems." *Viruses*, **3** (10): 1933-1958.

Vink, C., Rudenko, G., and Seifert, H.S. (2011) "Microbial antigenic variation mediated by homologous DNA recombination." *FEMS Microbiol. Rev.,* doi: 10.1111/j.1574-6976.2011.00321.x.

Waksman, S.A., and Woodruff, H.B. (1942) "Selective antibiotic action of various substances of microbial origin." *J. Bacteriol.,* **44** (3): 373-384.

Walker, C.K., and Sweet, R.L. (2011) "Gonorrhea infection in women: prevalence, effects, screening, and management." *Int. J. Womens Health,* **3**: 197-206.

Wang, X., Kim, Y., Hong, S.H., Ma, Q., Brown, B.L., Pu, M., Tarone, A.M., Benedik, M.J., Peti, W., Page, R., and Wood, T.K. (2011) "Antitoxin MqsA helps mediate the bacterial general stress response." *Nat. Chem. Biol.,* **7** (6): 359-366.

Wang, W., Black, S.S., Edwards, M.D., Miller, S., Morrison, E.L., Bartlett, W., Dong, C., Naismith, J.H., and Booth, I.R. (2008) The structure of an open form of an E. coli mechanosensitive channel at 3.45 A resolution. *Science*, **321** (5893): 1179-1183.

Ward, J.J., Sodhi, J.S., McGuffin, L.J., Buxton, B.F., and Jones, D.T. (2004) Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J. Mol. Biol.,* **337** (3): 635-645

Wilson, L.G. (1987) "The early recognition of streptococci as causes of disease." *Medical History,* **31** (4): 403-414.

Winther, K.S., and Gerdes, K. (2011) "Enteric virulence associated protein VapC inhibits translation by cleavage of initiator tRNA." *Proc. Natl. Acad. Sci. U.S.A.,* **108** (18): 7403-7407.

Woese, C.R., and Fox, G.E. (1977) "The concept of cellular evolution." *J. Mol. Evol.,* **10** (1): 1-6.

Woods, C.R. (2005) "Gonococcal infections in neonates and young children." *Semin. Pediatr. Infect. Dis.,* **16** (4): 258-270.

Wright, G.D. (2007) "The antibiotic resistome: the nexus of chemical and genetic diversity." *Nat. Rev., Microbiol.,* **5** (3): 175-186.

Wu, J., Apontes, P., Song, L., Liang, P., Yang, L., and Li, F. (2007) Molecular mechanism of upregulation of survivin transcription by the AT-rich DNA-binding ligand, Hoechst33342: evidence for survivin involvement in drug resistance. *Nucleic Acids Res.,* **35** (7): 2390-2402.

Xie, B., Tassi, E., Swift, M.R., McDonnell, K., Bowden, E.T., Wang, S., Ueda, Y., Tomita, Y., Riegel, A.T., and Wellstein, A. (2006) Identification of the fibroblast growth factor (FGF)-interacting domain in a secreted FGF-binding protein by phage display. *J. Biol. Chem.,* **281** (2): 1137-1144.

Yamaguchi, Y., Park, J.H., and Inouye, M. (2011) "Toxin-antitoxin systems in bacteria and archaea." *Annu. Rev. Genet.,* **45**: 61-79.

Yamaguchi, Y., and Inouye, M. (2011) "Regulation of growth and death in Escherichia coli by toxin-antitoxin systems." *Nat. Rev., Microbiol.,* **9** (11): 779-790.

Yang, L., Haagensen, J.A., Jelsbak, L., Johansen, H.K., Sternberg, C., Høiby, N., and Molin, S. (2008) "In situ growth rates and biofilm development of Pseudomonas aeruginosa populations in chronic lung infections." *J. Bacteriol.,* **190** (8): 2767-2776.

Yang, M., Gao, C., Wang, Y., Zhang, H., and He, Z.G. (2010) "Characterization of the interaction and cross-regulation of three Mycobacterium tuberculosis RelBE modules." *PLoS ONE,* **5** (5).

Yogev, R., and Tan, T. (2011) "Meningococcal disease: the advances and challenges of meningococcal disease prevention." *Hum. Vaccin.,* **7** (8): 828-837.

Zalmanovici Trestioreanu, A., Fraser, A., Gafter-Gvili, A., Paul, M., and Leibovici, L.(2011) "Antibiotics for preventing meningococcal infections." *Cochrane Database Syst. Rev.,* **10** (8).

Zhang, L., and Ma, S. (2010) "Efflux pump inhibitors: a strategy to combat P-glycoprotein and the NorA multidrug resistance pump." *ChemMedChem,* **5** (6): 811-822.

Zhang, Y., Zhang, J., Hara, H., Kato, I., and Inouye, M. (2005) "Insights into the mRNA cleavage mechanism by MazF, an mRNA interferase." *J. Biol. Chem.,* **280** (5): 3143-3150.

Web references

Web1 Centers for Disease Control and Prevention, *Centers for Disease Control and Prevention*. <http://www.cdc.gov/>.

Web2 World Health Orgainzation, *World Health Organization*, <http://www.who.int/>.

Web3 Antibiotics, Survey, *Scribd*, <http://es.scribd.com/doc/30136935/Antibiotics-Survey>.

Web4 Ryan, J. Biofilms and their role in human health, *Webbetraining*, <www.webbertraining.com/files/library/docs/57.pdf>.

Web5 Pathogenicity - Gonorrhea - STD Information from CDC, *Centers for Disease Control and Prevention*, <http://www.cdc.gov/std/gonorrhea/lab/pathogenicity.htm>.

Web6 Meningococcal vaccines: WHO position paper, November 2011 *Weekly epidemiological record, World Health Organization*, < http://www.who.int/wer/2011/wer8647.pdf >.

Web7 Meningitis Vaccine, *Meningitis & Septicaemia, Meningitis Research Foundation*, <http://www.meningitis.org/disease-info/vaccines>.

Web8 Identification of Streptococcus Species, Enterococcus Species and Morphologically similar organisms, *Standards Unit, Evaluations and Standards Laboratory, Centre for Infections, Health Protection Agency, BSOP ID 4i2.2*, <http://www.hpa.org.uk/webc/HPAwebFile/HPAweb_C/1317132925301>.

Web9 ExPASy Server, *SIB Bioinformatics Resource Portal*, <http://expasy.org/>.

Web10 Swiss-Model, <http://swissmodel.expasy.org/>.

Web11 CCP4i tools for PDB editing, <http://www.ccp4.ac.uk/>.